Precise information retrieval in semantic scientific digital libraries

DE RIBAUPIERRE, Hélène

Abstract

When scientists are looking for information in document collections, they generally have a precise objective in mind. They try to answer specific needs such as finding the definition of a concept, checking whether an idea has already been tested, or comparing the scientific conclusions of several documents. To build better information system models, it is important to understand the needs of scientists in terms of information and their seeking information behavior. The contributions of this thesis are to a) offer a user model; b) propose a generic annotation model (SciAnnotDoc) for scientific documents; c) demonstrate that the annotation model is realistic enough to be used by both manual and automatic methods of annotation; d) build a faceted search interface (FSAD) and e) evaluate the model with scientists and show that the FSAD interface is more effective than a keywords-search, notably in terms of user preference, recall and F1 measure.

Reference


URN : urn:nbn:ch:unige-431654
DOI : 10.13097/archive-ouverte/unige:43165
Precise information retrieval in semantic scientific digital libraries

Hélène de Ribaupierre
PRECISE INFORMATION RETRIEVAL IN SEMANTIC SCIENTIFIC DIGITAL LIBRARIES

Thèse présentée à la Faculté des sciences économiques et sociales de l'Université de Genève

Par Hélène de Ribaupierre

pour l'obtention du grade de
Docteure ès sciences économiques et sociales
mention : Système d'Information

Membres du jury de thèse :

Prof. Gilles FALQUET, directeur de thèse, Université de Genève
Prof. Giovanna Di MARZO SERUGENDO, présidente du jury, Université de Genève
Prof. Fabio CRESTANI, Università della Svizzera italiana USI, Lugano
Prof. Henning MÜLLER, Université de Genève et Institut informatique de gestion, HES-SO, Valais

Thèse N° 869
Genève, le 1er décembre 2014
La Faculté des sciences économiques et sociales, sur préavis du jury, a autorisé l'impression de la présente thèse, sans entendre, par là, n'émettre aucune opinion sur les propositions qui s'y trouvent énoncées et qui n'engagent que la responsabilité de leur auteur.

Genève, le 1\textsuperscript{er} décembre 2014

Le vice-recteur  
Denis HOCHSTRASSER

Impression d'après le manuscrit de l'auteur
Precise information retrieval in semantic scientific digital libraries

THÈSE

présentée à la Faculté des sciences économiques et sociales
de l’Université de Genève
par

Hélène de Ribaupierre

Sous la direction de

Prof. Gilles Falquet

pour l’obtention du grade de

Docteure ès sciences économiques et sociales
mention systèmes d’information

Membres du jury de thèse:

M. Fabio CRESTANI, Professeur, Università della Svizerra Italiana
Mme. Giovanna DI MARZO SERUGENDO, Professeure, présidente du jury
M. Henning MULLER, Professeur, HES-SO Valais et Faculté de Medecine,
Université de Genève

Thèse n°869
Carouge, le 1er décembre 2014
When scientists are looking for information in document collections, or on the web, they generally have a precise objective in mind. Instead of looking for documents "about a topic T ", they rather try to answer specific needs such as finding the definition of a concept, finding results for a particular problem, checking whether an idea has already been tested, or comparing the scientific conclusions of several documents.

The first contribution of this thesis, is to understand better the needs of scientists in terms of information and their seeking information behaviour and to build a user model from these observations. The methodology used were interviews and a survey. One of the results was that scientists focus in chunks of texts (discourse elements) when they are seeking and reading scientific documents, and not at the entire document.

The second contribution of this thesis is a generic annotation model for scientific documents using Semantic Web technologies (SciAnnotDoc model) based on the discourse elements that are most frequently used by scientists according to the user model: findings, hypothesis, definition, methodology and related work. The SciAnnotDoc model is based on four dimensions or facets: the metadata dimension, the conceptual dimension (terms and concepts), the discourse dimension and the relationships dimension (relation between discourse elements). The different dimensions are modelled into separate ontologies.

The third contribution of this thesis is an automatic annotation process based on these ontologies. We annotated automatically 1’410 document in gender studies with the SciAnnotDoc model using syntactic patterns to discover the discourse elements at the sentence level.

The fourth contribution of this thesis is the evaluation of the annotation model with "real users". We built a faceted-search based (FSAD) on the four dimensions/facets of the SciAnnotDoc model. The evaluation by scientists showed that they seem to prefer and find the results obtained using the FSAD more relevant and with a better recall than result provided by a keywords-search interface using a terms indexing.

**Keywords:** Semantic publishing, Information retrieval, Ontology, Knowledge management
Lorsqu’ils effectuent des recherches de documents scientifiques ou techniques, les scientifiques ont des objectifs précis en tête. Avec l’augmentation des documents scientifiques disponibles aujourd’hui, répondre à ce besoin d’information est de plus en plus difficile pour les scientifiques que cela soit en termes de temps ou de détection des documents pertinents dans cet océan de littérature. Il devient donc de plus en plus important d’avoir de bon moteurs de recherche qui répondent aux besoins des scientifiques.

Pour ce faire, nous avons d’abord mené des interviews et une enquête auprès de scientifiques pour comprendre plus précisément comment ils recherchent leurs informations et travaillent avec les documents trouvés. Nous avons observé que les scientifiques recherchent leurs informations dans des éléments de discours précis qui peuvent varier suivant la tâche qu’ils ont à effectuer, et non pas toujours dans le document en entier.

A partir de ce modèle d’utilisateur, nous avons ensuite créé un modèle d’annotation (SciAnnotDoc) prenant en compte ces éléments de discours, de même que les différentes informations contenues dans le document scientifique. Ce modèle est composé de quatre dimensions ou facettes: les méta-données, la dimension conceptuelle (les termes et les concepts), la dimension discursive et les relations de citations ou de références entre les différents éléments de discours d’un documents scientifique. Le modèle a été implémenté en OWL et différentes ontologies ont été créées pour répondre aux différents besoins d’annotation des quatre dimensions.

Dans une troisième phase, plusieurs méthodes d’annotation manuelle et automatique ont été évaluées. GATE a été choisi et utilisé pour annoter automatiquement 1400 documents scientifiques en études genre. Des règles JAPE ont été créées basées sur la syntaxe grammaticale et des mots-clés pour détecter automatiquement les cinq éléments de discours (findings, définition, hypothèse, méthodes et travaux référencés).

Finalement, ces annotations et l’efficacité d’un tel système d’annotation ont été évaluées avec des scientifiques. Une interface en facette (FSAD) basée sur les quatre dimensions du modèle d’annotation a été créée pour cette évaluation et comparée à un système de recherche par mots-clés. Les observations semblent montrer que les scientifiques préfèrent le système (FSAD) basé sur SciAnnotDoc et trouvent que les résultats obtenus sont plus utilisables et ont un meilleur taux de rappel que ceux fournis par le système de recherche par mot-clés.

**Mots-clés:** Publication sémantique, Recherche d’information, Ontologie, Représentation des connaissances
Je remercie mon directeur de thèse et professeur Gilles Falquet, pour m’avoir donné la chance de me lancer dans une thèse en m’engageant sur un projet du Fonds National\textsuperscript{1}. Je le remercie pour ces merveilleuses heures passées à discuter de ma thèse, ou d’autres sujets autour d’un thé, sous un chêne ou encore sur un téléskiage. Il m’a toujours encouragée à aller plus loin, même quand j’étais un peu décourageée; à explorer d’autres pistes et à trouver mes propres chemins. Je le remercie aussi pour sa patience et ses conseils.

Je remercie les membres de mon jury. Le professeur Fabio Crestani d’avoir accepté d’être membre de mon jury, et d’avoir lu et corrigé ma thèse; et de m’avoir donné de bonnes pistes de recherche future. Je remercie le professeur Henning Müller pour les mêmes raisons, de même que pour quelques conseils de skis, et de cueillette de champignons.

Je remercie aussi ma présidente de jury la professeure Giovanna Di Marzo Serugendo pour avoir accepté de jouer ce rôle et pour ses conseils avisés.

Je remercie Claudine Métral qui a eu la bonté de m’accueillir dans son bureau à l’ombre de ses plantes vertes pendant toutes ces années, et qui m’a donné de bons conseils, que ce soit sur ma thèse ou mon travail d’assistante, et avec qui j’ai eu de bons éclats de rire.

Je remercie aussi Nizar Ghoula d’avoir bien voulu relire certains passages de ma thèse, et d’avoir été un collègue et ami sur qui j’ai pu m’appuyer pendant ces quelques années. Je remercie bien sûr le reste de l’équipe, mes collègues, les membres du CUI pour leurs point de vue et leur aide sur ma thèse. Marie-France mérite un remerciement particulier pour son soutien et sa patience avec certains de mes formulaires pas toujours bien remplis.

Je remercie évidemment ma mère pour tout, mais particulièrement pour sa relecture attentive de ma thèse, son aide et son soutien infaillible. Je remercie ma soeur pour ses conseils, ses relectures, ses corrections et son soutien; et bien sûr le reste de la famille pour son support tout au long de ce processus. Je remercie Roy pour la discussion assis sur les télésièges de Marble Mountain (Terre-Neuve) portant sur les différents techniques de traitements naturels du langage.

Je remercie mes amis pour leur patience, même quand cela n’était pas toujours facile. Pour avoir supporté mes coups de gueule et mon humeur pas toujours rose. Je remercie plus particulièrement Raluca pour m’avoir donné le petit coup de pouce qu’il fallait quand je me décourageais; Lucile pour avoir lu mes complaints (via What’s up ou email) et avoir partagé un grand nombre de mes expériences de

\footnote{Ce travail a été effectué grâce à deux subsides successifs du Fonds National (Ontology-Based Precise Multilingual Information Retrieval, n° 122061; Precise information retrieval in semantic scientific digital libraries, n° 138252)}
doctorant (conférences, mentorat, et j’en passe); Wanda pour m’avoir sustenté et m’avoir permis de me défouler avec une raquette quand le besoin était trop grand; Lonneke pour m’avoir aidé à construire mon premier modèle d’annotation; Sandra pour une plume jaune qui a servit à écrire de longues pages de ma thèse et qui ne me quitte plus; ma mentor Franziska pour nos longues discussions et ses conseils avisés; Jelena, pour m’avoir dit de foncer et que la thèse serait mes plus belles années (j’en doute encore un peu); je remercie Ilinca pour ses sourires et me rappeler qu’une thèse c’est beau, mais il n’y a pas que cela. Je remercie encore pour leur patience, leur écoute et leur soutien tous mes proches et amis dont faire l’éloge ici serait trop longue.

Finalement, je remercie tous mes amis, proches et collègues qui ont participé avec patience et bonne volonté aux différents interviews et évaluation du système: Alain, Anik, Caroline, Céline, Delphine, Dorina, Jean-Pierre, Kate, Lonneke, Lucile, Mélanie, Nizar, Sandra, Sandrine, Vanessa, Wanda, et les participants à l’enquête. Sans eux cette thèse n’aurait pas été possible.

Je vous remercie tous en espérant n’avoir oublié personne.
## Contents

1 Introduction ................................. 1

2 User model .................................. 7
   2.1 Introduction .......................... 8
   2.2 State of the art ....................... 9
       2.2.1 Information-seeking behaviour models ............ 9
       2.2.2 Scientific inquiry models, nature of the needs of information, new forms of behaviour .......... 17
   2.3 Survey .................................. 24
       2.3.1 Methodology ....................... 24
       2.3.2 Results ........................... 25
       2.3.3 Survey Discussion and Conclusion ............... 36
   2.4 Interview ................................ 36
       2.4.1 Methodology ....................... 37
       2.4.2 Results ........................... 38
       2.4.3 Interview, Discussion and Conclusion ............. 46
   2.5 Discussion and Conclusion ................ 49

3 Annotation models for scientific documents ............ 53
   3.1 Introduction .......................... 53
   3.2 State of the art ....................... 54
       3.2.1 Conceptual dimensions (1) ............... 54
       3.2.2 Meta-data dimension (2) ................ 55
       3.2.3 Structural dimension (3) ................ 57
       3.2.4 Rhetorical or discursive dimension (4) ............ 58
       3.2.5 Relationships dimension (5) ............... 69
   3.3 Our model: SciAnnotDoc ................. 74
       3.3.1 Components of the model ................ 77
       3.3.2 Formalisation of SciAnnotDoc ............... 88
   3.4 Discussion and Conclusion ............... 100

4 Manual and automatic corpus annotation ............... 101
   4.1 Introduction .......................... 101
   4.2 Manual annotation .................... 102
   4.3 Crowdsourcing annotation ............... 106
       4.3.1 Introduction ....................... 106
       4.3.2 Scientists’ annotation study ............... 106
       4.3.3 Amazon Mechanical Turk study .......... 109
       4.3.4 Conclusion ........................ 113
   4.4 Automatic annotation .................. 114
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.4.1 Introduction</td>
<td>114</td>
</tr>
<tr>
<td>4.4.2 Implementation</td>
<td>116</td>
</tr>
<tr>
<td>4.4.3 Corpus statistics</td>
<td>125</td>
</tr>
<tr>
<td>4.4.4 Annotation evaluation</td>
<td>128</td>
</tr>
<tr>
<td>4.4.5 Discussion</td>
<td>128</td>
</tr>
<tr>
<td>4.5 Conclusion</td>
<td>129</td>
</tr>
<tr>
<td>5 Query model and Faceted search interface</td>
<td>131</td>
</tr>
<tr>
<td>5.1 Introduction</td>
<td>131</td>
</tr>
<tr>
<td>5.2 Query model</td>
<td>132</td>
</tr>
<tr>
<td>5.2.1 Simple queries</td>
<td>135</td>
</tr>
<tr>
<td>5.2.2 Queries with multiple elements</td>
<td>136</td>
</tr>
<tr>
<td>5.2.3 Examples of queries in SPARQL extracted from interviews</td>
<td>138</td>
</tr>
<tr>
<td>5.3 Faceted search on the SciAnnotDoc (FSAD)</td>
<td>140</td>
</tr>
<tr>
<td>5.3.1 Meta-data facet or dimension</td>
<td>142</td>
</tr>
<tr>
<td>5.3.2 Conceptual facet/dimension</td>
<td>142</td>
</tr>
<tr>
<td>5.3.3 Discursive facet/dimension</td>
<td>143</td>
</tr>
<tr>
<td>5.3.4 Relationships facet/dimension</td>
<td>144</td>
</tr>
<tr>
<td>5.3.5 Result’s visualisation</td>
<td>145</td>
</tr>
<tr>
<td>5.4 Conclusion</td>
<td>145</td>
</tr>
<tr>
<td>6 User evaluation</td>
<td>147</td>
</tr>
<tr>
<td>6.1 Introduction</td>
<td>147</td>
</tr>
<tr>
<td>6.2 Methodology</td>
<td>148</td>
</tr>
<tr>
<td>6.3 Results</td>
<td>152</td>
</tr>
<tr>
<td>6.3.1 Questions answered after each task</td>
<td>152</td>
</tr>
<tr>
<td>6.3.2 Questions asked after completion of the three tasks</td>
<td>157</td>
</tr>
<tr>
<td>6.3.3 Logs</td>
<td>159</td>
</tr>
<tr>
<td>6.4 Discussion and Conclusion</td>
<td>165</td>
</tr>
<tr>
<td>7 Conclusion</td>
<td>169</td>
</tr>
<tr>
<td>7.1 User model</td>
<td>169</td>
</tr>
<tr>
<td>7.2 Annotation models for scientific documents</td>
<td>169</td>
</tr>
<tr>
<td>7.3 Corpus annotation</td>
<td>172</td>
</tr>
<tr>
<td>7.4 Query model and FSAD</td>
<td>173</td>
</tr>
<tr>
<td>7.5 User evaluation</td>
<td>174</td>
</tr>
<tr>
<td>Bibliography</td>
<td>177</td>
</tr>
<tr>
<td>Glossary</td>
<td>195</td>
</tr>
<tr>
<td>A Survey Questions</td>
<td>199</td>
</tr>
<tr>
<td>B Semi-direct interview question</td>
<td>201</td>
</tr>
</tbody>
</table>
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>C SciDeo ontology</td>
<td>205</td>
</tr>
<tr>
<td>D Methodology ontology</td>
<td>209</td>
</tr>
<tr>
<td>E Scientific object ontology</td>
<td>213</td>
</tr>
<tr>
<td>F Gender studies ontology</td>
<td>217</td>
</tr>
<tr>
<td>G Jape Rules</td>
<td>235</td>
</tr>
<tr>
<td>G.1 Macro common to all Jape rules</td>
<td>235</td>
</tr>
<tr>
<td>G.2 Jape rules Findings</td>
<td>236</td>
</tr>
<tr>
<td>G.3 Jape rules Definition</td>
<td>245</td>
</tr>
<tr>
<td>G.4 Jape rules Methodology</td>
<td>258</td>
</tr>
<tr>
<td>G.5 Jape rules Hypothesis</td>
<td>264</td>
</tr>
</tbody>
</table>
List of Tables

2.1 "Why do you read scientific documents" (in percent, n=91) ....... 25
2.2 "Why do you read scientific documents by area of research" (in percent n=91), "soft" science (SS), "hard" science (HS) and other (O)  
$p = \text{probability of significance in a } \chi^2 \text{ test} \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldot
4.11 Cross-table between Workers A and scientist 1 for the five types of discourse elements .................................................. 111
4.12 Fleiss’ kappa results for Workers B and scientist 1 (on 158 sentences) ................................................................. 112
4.13 Cross-table between worker B and scientist 1 for the five types of discourse elements .................................................. 112
4.14 Precision/recall for the two annotators (GS=golden standard) ........... 113
4.15 Number of articles by journals .............................................................. 119
4.16 Annotated corpus statistics ................................................................. 125
4.17 Annotated corpus statistics by Discourse elements .......................... 125
4.18 Precision/recall values ...................................................................... 128

6.1 Design experiment (Latin square rotation of tasks) ............... 149
6.2 Sample Population (the number of the question related to the socio-demographic questionnaire) ........................................... 151
6.3 Relevance of the set of results .............................................................. 152
6.4 Relevance of the set of results in percent ............................................. 153
6.5 Time to obtain information ................................................................. 155
6.6 Number of requests ........................................................................... 155
6.7 Satisfaction for each query ................................................................. 156
6.8 Satisfaction with the overall results ..................................................... 156
6.9 Frustration with the set of results ....................................................... 156
6.10 Difficulty of the task ....................................................................... 157
6.11 Average time (minutes) spent on a set of results and on an answer (the results for the FSAD participants are shown in bold) ........ 160
6.12 Average time (minutes) spent on a set of results and on an answer (the results for the FSAD participants are shown in bold) ........ 161
6.13 Average time (minutes) spent on a set of results and on an answer (the results for the FSAD participants are shown in bold) ........ 162
List of Figures

2.1 A stage process version of Ellis’s behavioural framework (1993) [Ellis et al. 1993] ........................................ 9
2.7 Garvey’s table for the nature of information needed at each stage by the majority of scientists involved in that stage of scientific work ([Garvey et al. 1974, p.120]) .......................... 19
2.8 "Why do you read scientific documents by area of research by area of research"; often answer .......................... 27
2.9 "Why do you read scientific documents by area of research by years of birth"; often answer .......................... 27
2.10 "Where do you generally find your answers by area of research"; often answer .......................... 30
2.11 Relationship between task and what parts are read .......................... 35
2.12 "Would you like a retrieval system that would be able to find articles by their conclusions" .......................... 34
2.13 Fragment read by person .......................... 43
2.14 Fragments on which users focus as a function of task .......................... 45
2.15 Fragment by tasks .......................... 48
2.16 Fragment; comparison survey/interview .......................... 49
2.17 Relationship between tasks and parts read, as indicated by interview and survey .......................... 51
3.2 Class structure of the scholarly discourse ontology (ScholOnto [Shum et al. 2007, p.9]) .......................... 61
3.3 Scientific annotation models summarised .......................... 73
3.4 Method for the creation of the model .......................... 77
3.5 Conceptual model .......................... 78
3.6 An example of relationships between discourse elements: Correll, England and colleagues and human capital theory .......................... 85
3.7 Lineage of idea or "fil d’Ariane" .......................... 86
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.8</td>
<td>SciAnnotDoc model ontology; orange classes are the SciDeo ontology</td>
<td>88</td>
</tr>
<tr>
<td>3.9</td>
<td>SciAnnotDoc ontology in Protege</td>
<td>90</td>
</tr>
<tr>
<td>3.10</td>
<td>Methodology ontology: high-class level</td>
<td>91</td>
</tr>
<tr>
<td>3.11</td>
<td>Methodology ontology: all classes and properties (to give an idea of the size and the structure of the ontology)</td>
<td>92</td>
</tr>
<tr>
<td>3.12</td>
<td>Scientific object ontology: high-class level</td>
<td>93</td>
</tr>
<tr>
<td>3.13</td>
<td>Scientific object ontology: A summary view (to give an idea of the size and the structure of the ontology)</td>
<td>95</td>
</tr>
<tr>
<td>3.14</td>
<td>Gender studies ontology: high-class level</td>
<td>96</td>
</tr>
<tr>
<td>3.15</td>
<td>The GendStud ontology: A summary view (to give an idea of the size and the structure of the ontology)</td>
<td>97</td>
</tr>
<tr>
<td>3.16</td>
<td>Definition, Definiendum, Definiens concepts in the SciAnnotDoc</td>
<td>98</td>
</tr>
<tr>
<td>3.17</td>
<td>Inference between discourse elements</td>
<td>99</td>
</tr>
<tr>
<td>4.1</td>
<td>Manual annotation of three paper documents</td>
<td>104</td>
</tr>
<tr>
<td>4.2</td>
<td>Automated annotation process</td>
<td>120</td>
</tr>
<tr>
<td>4.3</td>
<td>Gate Pipeline</td>
<td>121</td>
</tr>
<tr>
<td>4.4</td>
<td>Annotation algorithm model</td>
<td>122</td>
</tr>
<tr>
<td>4.5</td>
<td>Inference for the &quot;Lineage of Idea&quot;</td>
<td>123</td>
</tr>
<tr>
<td>4.6</td>
<td>An example for the inference &quot;Lineage of Idea&quot;</td>
<td>124</td>
</tr>
<tr>
<td>4.7</td>
<td>Percentage discourse elements by journals (1= Gender and Society; 2= Feminist Studies; 3=Gender Issues; 4=Signs; 5=American Historical Review; 6= American Journal Of Sociology; 7=Feminist economist)</td>
<td>127</td>
</tr>
<tr>
<td>4.8</td>
<td>Distribution of the concepts in the corpus</td>
<td>127</td>
</tr>
<tr>
<td>4.9</td>
<td>Number of times the concepts &quot;gender&quot;, &quot;sex&quot; or &quot;discrimination&quot; are evoked by year (normalised by the number of concepts by year)</td>
<td>127</td>
</tr>
<tr>
<td>5.1</td>
<td>Query model</td>
<td>134</td>
</tr>
<tr>
<td>5.2</td>
<td>A instance of the query model (simple query)</td>
<td>135</td>
</tr>
<tr>
<td>5.3</td>
<td>Transformation of a simple query in SPARQL with an AND between the discourse elements</td>
<td>135</td>
</tr>
<tr>
<td>5.4</td>
<td>Transformation of a simple query in SPARQL with an OR between the discourse elements</td>
<td>136</td>
</tr>
<tr>
<td>5.5</td>
<td>A instance of the query model (complex query)</td>
<td>136</td>
</tr>
<tr>
<td>5.6</td>
<td>Transformation of a complex query in SPARQL with an AND between the discourse elements</td>
<td>137</td>
</tr>
<tr>
<td>5.7</td>
<td>Representation of Q1 in SPARQL</td>
<td>138</td>
</tr>
<tr>
<td>5.8</td>
<td>The representation of Q2 in SPARQL</td>
<td>139</td>
</tr>
<tr>
<td>5.9</td>
<td>Representation of Q3 in SPARQL</td>
<td>139</td>
</tr>
<tr>
<td>5.10</td>
<td>Adaptative-faceted search interface (with the four dimensions /facets)</td>
<td>141</td>
</tr>
<tr>
<td>6.1</td>
<td>Keyword search interface</td>
<td>149</td>
</tr>
<tr>
<td>6.2</td>
<td>Relevance of the set of results by interface</td>
<td>153</td>
</tr>
</tbody>
</table>
6.3 Irrelevance of the set of results by interface (some scores have been
inverted, so that all are in the same direction (1= irrelevant or useless;
5 or 7=very useful or relevant) ........................................ 154
6.4 Total time (in minutes) spent on query by group and by task ........ 163
6.5 Estimated precision and recall for the task 1 .......................... 164
6.6 Comparison of the number of results in a set of results and the irrele-
vance of the set of result, task 2 ........................................ 165
6.7 Number of queries and feeling of adequacy by interface for task 1
(adequacy; 1= very adequate; 2 = not at all adequate) ............. 166
6.8 Number of queries and feeling of adequacy by interface for task 2
(adequacy; 1= very adequate; 2 = not at all adequate) ............. 167
6.9 Number of queries and feeling of adequacy by interface for task 3
(adequacy; 1= very adequate; 2 = not at all adequate) ............. 167
This research thesis aims to develop models and tools for precise information retrieval and strategic reading in large scientific document collections. Recently, systems supporting electronic scientific publications have introduced some improvements to the simple search of printed edition indices, but the revolution heralded by electronic publishing has not yet occurred.

The construction of scientific knowledge is a long process that began hundreds of years ago. With the explosion of the number of publications (e.g., Medline has a growth rate of 0.5 million items per year [Nováček et al. 2010]), it has become very difficult for scientists to find information that is relevant to precise questions. As John of Salisbury stated, "Bernard of Chartres used to compare us to [puny] dwarfs perched on the shoulders of giants. He pointed out that we see more and farther than our predecessors, not because we have keener vision or greater height, but because we are lifted up and borne aloft on their gigantic stature."\(^1\) [Wikipedia 2014b]. Science is built on the knowledge that other scientists have brought and continue to bring to the equation. Today, the shoulders of giants have become difficult to reach. One of the reasons for this is the difficulty in finding the right information. Indeed, the communication of scientific information is not as efficient and effective as it could be.

In 1945, Bush [Bush 1945] argued that too many publications can be a problem because the information contained in these publications cannot reach other scientists. Bush expounded his argument using the example of Mendel’s laws of genetics, which were lost to the world for a generation because his publication did not reach the few people who were capable of understanding and extending this concept. The difficulty, according to Bush, is not the sheer number of publications but the fact that publications have been extended far beyond our present ability to make real use of these publications. He also argued that classification by alphabetical or numerical order is not an ideal solution; instead, classification should follow the same process as the human brain by the association of ideas, through "selection by association rather than indexing". Bush proposed the idea of the "Memex", in which individuals could store all of their records, books and communications on microfilm. With the "Memex", individuals could create a trail of their readings or consultations by joining items, and scientists could use their trails to navigate the literature.

This observation is still fully relevant. Today, scientists have better tools to man-\(^1\)"Dicebat Bernardus Carnotensis nos esse quasi nanos, gigantium humeris insidentes, ut pos- simus plura eis et remotiora videre, non utique proprii visus acumine, aut eminentia corporis, sed quia in altum subvenimur et extollimur magnitudine gigantea."
age their personal data (photos and video) than are available to manage or search their professional data [Hannay 2010]. The publication model has not improved or evolved as rapidly as it could have. The process of searching for scientific documents has just begun to incorporate new technologies such as Semantic Web. Moreover, the format of publications has not evolved significantly as the electronic format has become more popular.

Currently, scientists in the domain of knowledge representation, information retrieval, the Semantic Web and semantic publishing continue to work on some of the models that Bush proposed. Of course, technology has improved since Bush’s "Memex", but the idea of being able to follow a trail of ideas, for example, has not yet been realised. It is possible to follow the trail that a publication leaves behind and to follow the references of an article, but it is not yet possible to follow the trail of a precise idea or the reproduction of a result. For instance, if scientists want to search for all publications related to the effect that protein X could have on protein Y, they would have to find not only all the publications related to these two proteins but also the correct passages in the documents, when this interaction is proven or not proven.

In [Salton et al. 1993], Salton proposed an approach based on passage retrieval rather than considering a whole document. The author argues that it is not necessarily in the best interest of users to retrieve the entirety of a document, especially when the document is long, but it may be interesting to identify the most relevant passages.

Traditional search engines can find documents based on their metadata (e.g., title, author, year of publication) or by the words they contain (full-text search or keywords search). The first search mode is effective only when the user knows at least one metadata element, such as the title or author. However, scientists do not search only for documents whose metadata they know. Their need for information is often more complex, as are the questions they formulate in their cognitive processes.

Existing scientific information retrieval systems cannot handle (at least) two types of cases representing different information needs:

1. cases in which the information sought is composed of small pieces of information found in different documents, such as "collect all the definitions of term X in a given domain or school of thought" or "find all techniques to graphically represent knowledge bases in 3D";

2. cases in which the expression of the information needed is complex and the information cannot be expressed by a list of keywords, such as "find documents where author X disagrees with author Y about using methodology Z".

In both cases, typical document indexing schemes are inefficient because 1) they represent each document as a whole and do not take their subparts into account and 2) they fail to represent knowledge about the semantic or rhetorical role of document elements. One way to circumvent these limitations is to semantically enrich documents by annotating their content with meta-information or to provide
meta-information not only at the document level but at a finer level, such as the section, paragraph, sentence, or word level.

In addition, conventional information retrieval systems are based on the idea that each document has a certain degree of relevance to the user’s query, and only the most relevant documents must be selected. For scientists, the goal is often to find all the documents that are associated with a very specific issue of interest. Moreover, even if scientists want to find all such documents, they do not have time to read the retrieved documents in their entirety. Therefore, a search system must provide methods and tools for strategically reading these documents (i.e., for highlighting or selecting the passages that actually contribute to answering the user’s query).

Various initiatives have been launched to improve the dissemination of knowledge and the way scientists discover new facts, claims, definitions, methodologies, and results. Two types of initiatives can be identified: e-science, which aims to construct a network of data to support new scientific discoveries, and [Hunter & Liu 2010, Bechhofer et al. 2013], Semantic Publishing [Shotton 2009b], which aims to improve new scientific discoveries by facilitating access to a diversity of publications. The present thesis belongs to the second category.

The research presented herein investigated new paths that should reduce some of the difficulties scientists encounter in obtaining correct, relevant information. The first step consisted of attempting to understand the needs of scientists when they initiate a process of information-seeking behaviour by questioning them with respect to the information need that triggers their search process. In this first step, quantitative and qualitative methods were used. The result of this first step was a user model based on the review of literature, the survey and the interviews that helped to analyse scientists’ searching and reading behaviour for scientific documents.

The second step was to create an annotation model of scientific documents that not only responds to the needs of scientists but is also realistic for use by an automatic annotation system. To paraphrase Bernard de Chartres, this thesis was built on the knowledge of scientists. In other words, the annotation model described herein is based on the user model as well as other annotation models proposed in the literature. The construction of this model was made in several iterations, with changes depending of the results observed throughout the survey and interviews.

The third step was to test the annotation model by annotating a corpus. Several annotations methods were tested, including manual and automatic annotations. Manual annotations were performed using scientist and non-scientist annotators. The automatic annotation system was implemented using Natural Language Processing.

The fourth step was to implement a faceted information retrieval system to query the scientific documents annotated using the annotation model and to evaluate it with scientists against another information retrieval system, specifically, one based on keywords.

The contributions of this thesis are to a) offer a user model built on the basis of a survey and interviews with scientists in order to understand their specific information-seeking behaviour needs; b) propose a generic annotation model (SciAn-
notDoc) for scientific documents based on the user model, c) demonstrate that the annotation model is realistic enough to be used by both manual and automatic methods of annotation, d) build a faceted search interface (FSAD) based on the SciAnnotDoc and e) evaluate the model by obtaining the feedback of end users (scientists) and show that the FSAD interface is more effective than a keywords-search, notably in terms of recall and F1 measure.

The thesis is structured as follows. Chapter 2 discusses different models of information behaviour and information-seeking behaviour and presents studies that were conducted to understand scientists’ information behaviour and the nature of the needs of information. Chapter 3 discusses the scientific document annotation models reported in the literature and presents our model, the SciAnnotDoc model, a complete annotation model that is based on the aforementioned user model and is composed of different ontologies that allow for the annotation of scientific documents along a number of dimensions. Chapter 4 discusses automatic annotation frameworks and presents the methodology we used to annotate the corpus. Chapter 5 discusses query models and a faceted search interface and presents the interface we developed to query annotated scientific documents. Chapter 6 presents the results obtained from the evaluation of the system based on scientists’ feedback. Finally, conclusions and ideas for future studies are presented in Chapter 7. A glossary at the end of the chapters provides some definitions of the terms.

Publications related to this thesis

Some of the results presented in this thesis were published in different conferences and workshops:


Hélène de Ribaupierre and Gilles Falquet. *New trends for reading scientific documents*. In Proceedings of the 4th ACM workshop on Online books, complementary social media and crowdsourcing, BooksOnline ’11, pages 19-24, New York, NY, USA,
## Chapter 2

### User model

<table>
<thead>
<tr>
<th>Contents</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2.1 Introduction</strong></td>
<td>8</td>
</tr>
<tr>
<td><strong>2.2 State of the art</strong></td>
<td>9</td>
</tr>
<tr>
<td>2.2.1 Information-seeking behaviour models</td>
<td>9</td>
</tr>
<tr>
<td>2.2.2 Scientific inquiry models, nature of the needs of information,</td>
<td>17</td>
</tr>
<tr>
<td>new forms of behaviour</td>
<td></td>
</tr>
<tr>
<td><strong>2.3 Survey</strong></td>
<td>24</td>
</tr>
<tr>
<td>2.3.1 Methodology</td>
<td>24</td>
</tr>
<tr>
<td>2.3.2 Results</td>
<td>25</td>
</tr>
<tr>
<td>2.3.2.1 Reasons to read scientific documents</td>
<td>25</td>
</tr>
<tr>
<td>2.3.2.2 Do you read scientific papers to respond to a precise need</td>
<td>29</td>
</tr>
<tr>
<td>2.3.2.3 Where do you generally find the answers</td>
<td>29</td>
</tr>
<tr>
<td>2.3.2.4 Do you often conduct a review of the literature</td>
<td>32</td>
</tr>
<tr>
<td>2.3.2.5 Do you easily find papers/books you are looking for without</td>
<td>33</td>
</tr>
<tr>
<td>too much noise (e.g., a paper that is out of the field in which you are</td>
<td></td>
</tr>
<tr>
<td>searching</td>
<td></td>
</tr>
<tr>
<td>2.3.2.6 Do you use Google Scholar more often than a specialised database</td>
<td>33</td>
</tr>
<tr>
<td>(in your domain)</td>
<td></td>
</tr>
<tr>
<td>2.3.2.7 Would you like a retrieval system that would be able to find</td>
<td>33</td>
</tr>
<tr>
<td>articles by the content of their conclusions (i.e., all articles stating</td>
<td></td>
</tr>
<tr>
<td>that rain is wet)</td>
<td></td>
</tr>
<tr>
<td><strong>2.3.3 Survey Discussion and Conclusion</strong></td>
<td>36</td>
</tr>
<tr>
<td><strong>2.4 Interview</strong></td>
<td>36</td>
</tr>
<tr>
<td>2.4.1 Methodology</td>
<td>37</td>
</tr>
<tr>
<td>2.4.2 Results</td>
<td>38</td>
</tr>
<tr>
<td>2.4.2.1 Reasons to search for and read a scientific document</td>
<td>38</td>
</tr>
<tr>
<td>2.4.2.2 How are you searching a scientific document</td>
<td>39</td>
</tr>
<tr>
<td>2.4.2.3 What are the databases or search engines you use to search for</td>
<td>39</td>
</tr>
<tr>
<td>documents</td>
<td></td>
</tr>
<tr>
<td>2.4.2.4 Do you search documents for a precise question</td>
<td>39</td>
</tr>
<tr>
<td>2.4.2.5 For each task, do you search for documents differently or with</td>
<td>41</td>
</tr>
<tr>
<td>a different search engine or database</td>
<td></td>
</tr>
<tr>
<td>2.4.2.6 Does the knowledge you have in a domain change the way you</td>
<td>43</td>
</tr>
<tr>
<td>search for documents</td>
<td></td>
</tr>
</tbody>
</table>
2.1 Introduction

The number of scientific documents available today has made the information-seeking process more difficult than ever. In 2011, the number of peer-reviewed journals in existence was estimated to be close to 26,746\(^1\). Thus, finding relevant information among this tremendous number of publications proves to be a difficult task, especially for scientists who are searching for an answer to a very precise question, such as "Who has applied the theory of pertinence to translation?" Research in information science (as will be discussed in this section) shows that taking into account the difference between intra-individual variations\(^2\) and inter-individual variations\(^3\) can be important in the design of digital libraries and in the design of information retrieval systems (IRs). To build better scientific IRs, we must consider scientists' need for information, the reason they are searching for information, the difficulties they encounter and the strategies they have created to avoid some of these difficulties. In [Wilson 1999], Wilson argues that if the designer understands the implications of user information-seeking behaviour, it should be possible to build better IRs. Studies about the information-seeking process have been conducted since 1960. A review of all such studies is well beyond the scope of this thesis. In this study, only the most well-known models will be described to provide the reader with a better understanding of the information-seeking behaviour process and the needs of users and to provide the reader with the ability to apply the lens of different information-seeking behaviour models developed by scientists. A more complete review of these studies can be found in publications on the state of the art in [Case 2012, Ingwersen & Järvelin 2006, Wilson 1999]. The first of these studies presents meta models for information-seeking behaviour, whereas the second presents studies on information problems that impact scientific and scholarly

---

\(^1\)site web: http://pages.cmns.sfu.ca/heather-morrison/appendix-c-how-many-active-scholarly-peer-reviewed-journals/.

\(^2\)Intra-individual variations: changes which occur within individual scientists as their scientific work progresses.\(^4\) [Garvey et al. 1974, p.115]

\(^3\)Inter-individual variations: differences between individual scientists (traditionally referred to as ‘individual differences’).\(^5\) [Garvey et al. 1974, p.115]
inquiry, the nature of information needs, the objectives of searches performed by users, and the new behaviour introduced by electronic documents or the Web.

2.2 State of the art

2.2.1 Information-seeking behaviour models

Ellis’s model: Ellis’s model is based on the first investigation conducted by Ellis et al. on empirical studies covering a population of social scientists [Ellis 1989]. A second study was conducted in [Ellis et al. 1993], in which the authors compared scientists in the physical sciences to those in the social sciences. In [Ellis & Haugan 1997], the authors compared engineers and research scientists in an oil company. The methodology used to interview the scientists was semi-directive interviews. In the first study [Ellis 1989], the authors modelled six features (starting, chaining, browsing, differentiating, monitoring and extracting). In the 1993 study [Ellis et al. 1993], the authors added two more features (verifying and ending). The extended (last) model (see figure 2.1) described eight features of the information-seeking patterns of scientists:

Figure 2.1: A stage process version of Ellis’s behavioural framework (1993) [Ellis et al. 1993]

- **Starting**: Identifying references that could serve as a starting point of the research cycle; asking colleagues and consulting literature reviews, online catalogues, indexes and abstracts.
- **Chaining**: Following footnotes and citations in known material, the direction of which can be forward or backward.
- **Browsing**: Looking for information in areas of potential interest.
- **Differentiating**: Using known differences in information sources, such as authors’ names and journal hierarchies, to filter information.
- **Monitoring**: Keeping knowledge about users’ area of research up to date by regularly following particular sources.
Chapter 2. User model

- **Extracting:** Selecting and identifying relevant material.
- **Verifying:** Checking that the information is correct.
- **Ending:** Activities that complete the information-seeking process.

Ellis suggests that the interaction and the order of the eight features can vary. This model can be useful in providing a set of categories that allows for the analysis of information seeking at the individual level. In [Ellis & Haugan 1997], the authors conclude that even if sources of information differ between different fields of research, the behavioural characteristics remain similar. The authors consider the behavioural model to be quite robust in relation to the information-seeking patterns of physical scientists, engineers and social scientists.

**Kuhlthau model:** Based on different longitudinal\(^4\) empirical studies, Kuhlthau [Kuhlthau 1991] proposed a model composed of six stages (see figure 2.2). The first study concerned high-school students who had to write an essay. The model developed by Kuhlthau suggests that people search and use information differently depending on the stages of the search process, which are defined as follows:

- **Initiation:** Users become aware of the need for information when they face a problem.
- **Selection:** Users identify and select the general topic for seeking information.
- **Exploration:** Users seek and investigate information on the general topic of interest.
- **Formulation:** Users fix and structure the problem to be solved.
- **Collection:** Users gather pertinent information for the topic of interest.
- **Presentation:** Users complete their search and implement the results of their search.

This model, in contrast to Ellis’s model, proposes features in an order that cannot be changed. This difference between the two models may derive from the differences in the populations studied and in the tasks performed to retrieve information. Kuhlthau’s model also acknowledges the subjective characteristics (feelings and thoughts) of the user at each step.

In his article, Wilson [Wilson 1999] compares Ellis’s model (see figure 2.1) and Kuhlthau’s model (see figure 2.2) to determine the similarities between them (see figure 2.3) and to merge the models, although the hypotheses underlying the two models are slightly different.

\(^4\)The study was conducted over several years.
2.2. State of the art

Figure 2.2: Information Search Process (ISP), Kuhlthau’s model, 1991 [Kuhlthau 1991, p.367]

<table>
<thead>
<tr>
<th>Stages in ISP</th>
<th>Feelings Common to Each Stage</th>
<th>Thoughts Common to Each Stage</th>
<th>Actions Common to Each Stage</th>
<th>Appropriate Task According to Kuhlthau Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Initiation</td>
<td>Uncertainty</td>
<td>General/ Vague</td>
<td>Seeking Background Information</td>
<td>Recognize</td>
</tr>
<tr>
<td>2. Selection</td>
<td>Optimism</td>
<td></td>
<td></td>
<td>Identify</td>
</tr>
<tr>
<td>3. Exploration</td>
<td>Confusion/ Frustration/ Doubt</td>
<td></td>
<td>Seeking Relevant Information</td>
<td>Investigate</td>
</tr>
<tr>
<td>4. Formulation</td>
<td>Clarity</td>
<td>Narrowed/ Clearer</td>
<td></td>
<td>Formulate</td>
</tr>
<tr>
<td>5. Collection</td>
<td>Sense of Direction/ Confidence</td>
<td>Increased Interest</td>
<td>Seeking Relevant or Focused Information</td>
<td>Gather</td>
</tr>
<tr>
<td>6. Presentation</td>
<td>Relief/ Satisfaction or Disappointment</td>
<td>Clearer or Focused</td>
<td></td>
<td>Complete</td>
</tr>
</tbody>
</table>

Figure 2.3: A comparison of Ellis’s model and Kuhlthau’s stage process model in [Wilson 1999, p.256]

Stage: Initiation Selection/exploration Formulation Collection Presentation

Starting — Chaining — Differentiating — Extracting — Verifying — Ending

Activity: Recognize Identify/formulate Gather Complete
Wilson’s model Wilson began developing his model in 1981 (see figure 2.4) [Wilson 1981]. The aim of his model was not to describe information behaviour but to draw attention to the interrelationships among concepts used in the field. He suggests that information behaviour results from users’ recognition of needs. This behaviour can take several forms: 1) through a formal system, such as an information system or other information sources, or 2) through information exchange. In both cases, failure or success can be tested. When users are successful (i.e., when they believe that the information retrieved can fully or partially satisfy their needs), the search process can end. However, if the information does not satisfy the user’s needs, the user performs the search process again.

Figure 2.4: Wilson’s model of information behaviour (1981) [Wilson 1999, p.251]

In the same paper, Wilson presents a second (see figure 2.5) model for information-seeking behaviour that takes into account basic needs:

- Physiological states, such as the need for food, water, and shelter;
- Affective states (sometimes called psychological or emotional needs), such as the need for attainment or domination;
- Cognitive states, such as the need to plan or to learn a skill.

Wilson argues that these three types of needs are interrelated. When an individual engages in information-seeking behaviour, these needs may be fulfilled or may remain unfulfilled. The other factors that may intervene are the role of the person and the environment within which the person is living (see Figure 2.5). This
model, as described by Wilson, is a macro-model that describes information-seeking behaviour and suggests how information arises and what may prevent or aid the actual search for information.

Wilson’s 1996 model [Wilson 1999] (see figure 2.6) represents a major revision of Wilson’s second model of 1981 (see figure 2.5), although the basic framework persists. The barriers are represented as "intervening variables", which, in the previous model, could be preventive or supportive of information-seeking behaviour. Wilson incorporates some concepts that activate information-seeking behaviour: 1) risk/reward theory, which may help to explain which sources of information may be used more often than others, and 2) social learning theory, which incorporates the concept of self-efficacy.

The incorporation of different variables that may or may not help the user to initiate searching behaviour is a good start to account for the barriers that can arise in users’ information-seeking process, which should be considered in the construction and evaluation of IRs. These "intervening variables" can have implications for the decisions involved in the construction of IRs and the methods and variables measured for evaluation.

---

5“Perceived self-efficacy is defined as people’s beliefs about their capabilities to produce designated levels of performance that exercise influence over events that affect their lives.” [Bandura 1994, p.71].
Figure 2.6: Wilson’s model of information behaviour (1986) [Wilson 1999, p.257]
The Ellis model has been used in other studies, such as the study by [Meho & Tibbo 2003], in which the authors addressed the change introduced by the Web in information-seeking behaviour. The authors interviewed by email social scientists who were working in the same field but in different areas of the world. The authors’ findings are based on the eight features of the Ellis model and two additional features they identified for different purposes:

- **Starting:** Scientists first search through their own personal collection. They also search online catalogues, indexes and abstracts. Scientists primarily use literature searches as a way to determine what has been published on a given research topic or to find background information on a certain topic.

- **Chaining:** Chaining is used to identify new sources of information or to satisfy a potential information need. Chaining is performed by following the references obtained through reading and personal content (literature already known and owned by scientists in their own database and files). The resulting chains of citations are primarily based on topical relevance, the importance of the authors’ research, novelty, the publisher’s reputation, the time it takes to locate the information/material, or citation frequency). The authors show that the most important features for the chains of citations are topical relevance, authors’ reputation, and availability.

- **Browsing:** All of the interviewed scientists were engaged in this part of the process, and it was an important aspect for them. Authors divided this part of the process into two different processes: 1) browsing journals by searching through new issues of journals and the table of contents of relevant books and 2) browsing online catalogues, indexes and abstracts.

- **Monitoring:** In their results, the authors demonstrate the different ways scientists maintain awareness of research developments in their topic of interest. One of the ways that scientists maintain this awareness is by consulting lists of distribution, journals (based on their own subscriptions), conferences and colleagues.

- **Accessing:** The authors show that accessing information can be an issue; depending on the source, access can be difficult to obtain.

- **Differentiating:** The scientists interviewed in the [Meho & Tibbo 2003] study evaluate and judge sources according to their nature, quality, relative importance and usefulness. They also use different sources to overcome problems of bias.

- **Extracting:** Scientists identify and select relevant material from selected sources. The authors of the study analysed two types of extraction activities: those applied to direct sources (e.g., books and journal articles) and those applied to indirect sources (e.g., bibliographies, indexes and abstracts, and online catalogues).
• **Verifying:** The scientists interviewed in this study were preoccupied with the reliability and accuracy of their sources because of the nature of their research. In fact, the extent of this preoccupation can vary with the nature of the research topic and the nature of the relation that scientists may have with national and government organisations.

• **Networking:** Scientists communicate and maintain a close relationship with a broad range of people, such as friends, colleagues, and intellectuals working on similar topics, members of national organisations, government officials, and booksellers.

• **Information managing:** In the interviews, scientists expressed the need for systems that help them filter, archive and organise the information they have collected.

Meho and Tibbo find that the personal collection is the most important source of information for scholars. The authors also show that the activities in which scientists are involved are not necessarily or entirely sequential, as Ellis argued, but they can be generally divided into four categories: 1) searching, 2) accessing, 3) processing, 4) ending.
2.2.2 Scientific inquiry models, nature of the needs of information, new forms of behaviour

Scientists have created strategies to read documents rapidly. In [Tenopir et al. 2009], the authors analysed the reading patterns of science faculty members from 1977 through 2006 and found that the amount of reading increased, but the time spent reading an article diminished from 48 minutes in 1977 to 31 minutes in 2005. However, because scientists read more articles, they spent 120 hours annually reading in 1977 versus 144 hours annually in 2005. The authors found that the amount of reading varied with the field of research; for example, medical faculty read 414 articles per year versus 331 read by faculty in the natural sciences and 233 read by faculty in the social sciences. In [Nicholas et al. 2007], the authors found that users who tended to visit a Web site just once tended to publish the fewest papers. The authors found that the average time spent online for each full text item was 20-30 sec, which suggests that participants were checking documents online but not reading them online. The authors also found that the longer an article was, the more likely participants were to read the abstract or the summary and the less likely they were to look at the full text. Participants also spent more time online (42 sec) reading short articles (4-10 pages) than reading large articles (32 sec). The authors showed that search behaviour can vary as a function of academic field and the amount of experience users have in their research field. The study shows that scientists have developed different strategies to read faster and to avoid unnecessary reading. It also appears that scientists tend to avoid reading long articles. By analysing these types of strategies, we can assume that the more targeted information is, the more useful it will be to scientists. Indeed, scientists need techniques that will help them directly target the information they are seeking. Consequently, IR designers must understand the nature of the information needed.

In the literature, we can find several studies that describe the nature of the information needs of scientists.

In [Garvey et al. 1974], the authors defined the nature of information needed by scientists as follows:

- To aid in the perception or definition of a problem
- To formulate a scientific or technical solution
- To place work in its proper context with similar work that has already been completed
- To relate work to ongoing work in an area
- To select a design/strategy for data collection
- To select a data-gathering technique
- To design equipment or apparatus
- To choose a data analysis technique
- To enable full interpretation of collected data
• To integrate findings into the current state of knowledge in an area

and the stages of research in which scientists may be involved:

1. Preliminary planning (general)
2. Specific planning: Theoretical/conceptual
3. Preparation of written research proposal
4. Preliminary experimentation/field trials or mockups
5. Calibration, pretesting, etc.
6. Design and development of equipment/apparatus
7. Formulation of experimentation/study design
8. Collection of data
9. Analysis of data
10. Interpretation of results
11. Preparation for report of work

In [Garvey et al. 1974], the authors also discuss the interaction between the nature of the information needs of scientists and the stage of research scientists have reached. For example, we can observe (see figure 2.7) that in the preliminary planning stage, the nature of the information needs of scientists involve the perception or definition of a problem, the formulation of a scientific or technical solution, the placement of their work in its proper context with similar work that has already been completed (this nature of information need is involved in almost all stages except for the design and development of the equipment or apparatus) and the relation of their work to ongoing work in an area.
Figure 2.7: Garvey’s table for the nature of information needed at each stage by the majority of scientists involved in that stage of scientific work ([Garvey et al. 1974, p.120])

<table>
<thead>
<tr>
<th>Nature of information needed</th>
<th>A. Preliminary planning</th>
<th>B. Specific planning/ Theoretical/conceptual</th>
<th>C. Preparation of written research proposal</th>
<th>D. Execution of experimental/ field trials or mockups</th>
<th>E. Calculation, testing, etc.</th>
<th>F. Design and development of equipment</th>
<th>G. Formulation of experimental/ study design</th>
<th>H. Collection of data</th>
<th>I. Analysis of data</th>
<th>J. Interpretation of results</th>
<th>K. Preparation of report</th>
</tr>
</thead>
<tbody>
<tr>
<td>To aid in perception or definition of problem</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>To formulate a scientific or technical solution</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>To place work in proper context with similar work already completed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>To relate work to ongoing work in area</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>To select a design/strategy for data collection</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>To select a data-gathering technique</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>To design equipment or apparatus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>To choose a data analysis technique</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>To enable full interpretation of collected data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>To integrate findings into current state of knowledge in area</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
In a thorough and wide-ranging study on changes in scholars’ article-seeking and reading patterns, Tenopir et al. 2009 showed that scientists in "hard" sciences in the US (n=880) read scientific documents for the following purposes: research (48.5%), teaching (22.5%), writing articles, proposals, or reports (10.8%), current awareness, keeping up to date (8.0%), continuing education (3.0%), advising others (2.9%), and other reasons (4.2%). The authors also asked scientists how the last article they had read had affected their ways of thinking: 54.5% were inspired with new ways of thinking/new ideas, 39.9% obtained improved results, 26.6% narrowed/broadened/changed their focus, 12.4% saved time or other resources, 11.6% resolved technical problems, 7.4% were able to complete their work more quickly, 6.0% resulted in collaboration/joint research, and 5.6% provided other responses.

In Palmer et al. 2007, the authors interviewed subjects and found that scientists review documents for different purposes: to assess their own work and the work of others, to solve short-term instrumental problems, to consult and talk shop and to form complex intellectual relationships among the vast body of neuroscience findings.

In contrast with Ellis’s results Ellis & Haugan 1997, which show no differences between academic fields in information-seeking behaviour, other studies show that information-seeking behaviour can vary depending on the field of research or can be affected by the personality of the seeker. This difference may be due to the fact that the more recent studies focus on a different level of analysis from Ellis’s study. Whereas the Ellis model analyses information-seeking behaviour at the meta-level, more recent studies analyse this behaviour at a lower meta-level, taking into account what type of information scientists seek and how they do so.

In Garvey et al. 1974, the authors found that depending on the field of research, scientists’ experience in research, and scientists’ experience in their exact field of research, scientists do not share the same information needs. In fact, these needs vary depending on the stage of research in which scientists are involved.

In Bates 1996, the authors assume that the information-seeking process should be even more complex for interdisciplinary studies than for "classical academic disciplines". In Whitmire 2002, the authors study the difference in information-seeking behaviour between different academic disciplines. The authors classify academic fields of research into the three dimensions described by Biglan6. The authors surveyed a total of 5176 undergraduate students in four-year institutions and found that differences exist between the "hard" and "soft" disciplines, between life and non-life sciences, and between pure and applied science. Students in "soft" disciplines and "life" and "pure" science used online catalogues to a greater extent, asked librarians for help, read reference sections more often, used indexes more often, developed their bibliographies more extensively, checked for citations more often, and checked out books more often than students in the "hard", "applied" and "non-life" sciences did.

6Biglan [Biglan 1973]: 1) discipline-level paradigm (hard science and soft science); 2) practical application of research (pure and applied science); and 3) degree of involvement with the living or organic object of study (life versus non-life)
In [Kembellec 2012], the authors also detected some differences between different fields of research (196 PhD students and researchers in the human sciences, social sciences and hard sciences were surveyed). The authors found that 89% of searches were performed using a commercial information retrieval system, with an overall preference for Google Scholar. For the hard sciences, the authors found no differences between PhD students in the beginning stages of their programmes, PhD students at the end of their programmes, and researchers. For the social sciences and human sciences, searches were also performed using other systems, such as OPAC. The authors found differences between PhD students who were beginning to write their theses and those in the final stages of writing their theses, with the number of PhD students using systems such as OPAC decreasing from (44%) to (33%). Among researchers, 44% used OPAC.

In [Nicholas et al. 2007], the authors address the bouncing behavior of users of ScienceDirect. The authors analysed the logs of 800 senior academics in the life sciences, mathematics and medicine. [Nicholas et al. 2007] found that bouncing behavior varies different between fields of research. The authors also found that the number of times an item is viewed during a session changes as a function of the field of research. Medical scientists tend to show the strongest bouncing behaviour. The authors also found that graduate students exhibit more significant bouncing behaviour than senior scientists do.

Heinström [Heinström 2005] defined three different types of patterns for surfing the Web. She analysed the relations between information-seeking behaviour and personality traits and found that psychological features had a stronger influence on student information-seeking behaviour than students’ academic background or their stage of thesis writing. The author categorised the participants into three groups:

- **Fast-surfing students**: Students who wanted information available quickly and easily with the least effort. Depth of information was not a concern. Students judged the relevance of the documents by their look, the type of document and whether documents were easily available.

- **Broad-scanning student**: Students with exhaustive and flexible information-seeking behaviour over a wide range of sources. These students developed their search gradually rather than planning it.

- **Deep-diving student**: These students went beneath the surface in seeking information. Students of this type showed quality and conscientiousness in their searches and worked hard to obtain high-quality information. These students sought greater quality than quantity.

This study shows that it is not necessarily the academic field that creates the variation in information-seeking behaviour between persons; psychological features

---

7"Bouncing describes a form of behaviour in which users view only a few web pages from the vast number available and generally do not return to the same website very often, if at all." [Nicholas et al. 2007, p.1086]
(e.g., neuroticism, extraversion openness, agreeableness) may change this behaviour. Considering more than the differences between academic fields could be important in the design phase of IRs, but it could also be important to consider psychological features during the evaluation phase, as we will discuss in the evaluation chapter.

Scientists have created different strategies to avoid unnecessary reading of articles in the literature and to seek more efficient ways to obtain information. As demonstrated above, some of these strategies can vary depending on the academic field in which scientists are engaged or the personality traits of scientists. However, the difficulties that scientists may encounter in attempting to gather the information they need may arise due to other factors. In [Palmer et al. 2007], the authors sought to determine the effect of characteristics of a problem to be solved (weak and strong) on information-seeking behaviour in the field of neuroscience. The authors discussed two concepts derived from Simon’s theory (1986): weak and strong problem solving. Weak problem solving occurs when a problem space is ill structured, users have low domain knowledge and users take unclear or unsystematic steps in their research. These problems make the information-seeking process more difficult and more time consuming. In [Palmer et al. 2007], the authors indicate that data- and literature-mining techniques are relatively weak, compared to data-driven searches and are often correlated with an ill-defined problem focus. Palmer et al. also found that users use weaker techniques to search for new collaborations or strategic alliances, explore new domains and techniques, and gather new ideas and techniques. Users use stronger approaches to address problems when they already have a theory in mind. In this case, users have a high level of knowledge in the domain and proceed with clear and calculated steps. Palmer et al. showed that weak and strong approaches are both important for users seeking information. The authors suggest that weaker practices have a greater potential for promoting innovation and new discoveries. Taking into account the impact of weak or strong problem solving could introduce new perspectives to the design of IRs and the evaluation of such systems.

In [Renear & Palmer 2009a], the authors argue that scientists have problems locating very problem-specific information (methods, protocols). They also argue that scientists do not simply read individual articles in sequence but handle many articles simultaneously to search, filter, compare, arrange, link, annotate and analyse fragments of content to gather information as efficiently as possible. The authors call this behaviour "Strategic reading". The authors also argue that with the emergence of online indexing and navigation systems such as PubMed, Scopus, Google Scholar, and Citeseer, practices of strategic reading have intensified. Scientists search and browse literature as if they are playing video games, by rapidly moving through resources, changing search strings, chaining references, and accumulating bits of information, such as findings, equations, protocols and data. Scientists make rapid judgements across these documents.

In [Bishop 1999], the authors showed that scientific writing is "disaggregated and reaggregated" in the process of constructing new knowledge in science and engineering. In [Schatz et al. 2002, Bishop 1999], the authors used the DELIVER (DESktop LInk to virtual Engineering Resources) system to analyse the behaviour of scientists
using this system. DELIVER was a system deployed at the University of Illinois at Urbana-Champaign within the scope of the Digital Libraries Initiatives project (1994-1998). The system indexed 66,000 articles from five professional societies in physics, engineering and computer science. DELIVER used the document structure to index parts of documents, such as tables, figures and article titles; tags were generated in SGML. The authors found that scientists use different parts of the document structure, such as abstracts, figures, equations, and bibliographic citations, in their own work for several purposes:

- Identify a document of interest
- Address the relevance of an article before retrieving or reading it
- Create customised document surrogates after retrieval, for example, by combining bibliographic and other data
- Provide specific pieces of information
- Convey knowledge that is not easily rendered by words through figures and tables

The authors also found that the participants read the title and abstract of a document first and then skimmed section headings. More recently, Bishop et al. examined lists, summary statements, definitions and illustrations after key sections and observed that scientists read the conclusion and skimmed the references. This study showed that users may use a search engine to retrieve specific components of a document. However, the study was restricted to researchers in science and engineering.

In sum, in these studies, the authors tended to find that scientists search and read scientific documents differently depending on their field of research, their psychological traits, and the difficulty of their tasks. The authors also found that scientists avoid having to read documents unnecessarily and that scientists attempt to concentrate their reading in certain parts of documents. However, despite the large number of studies about how scientists search for information of interest, there is a need for further studies on this subject. In this thesis, we sought to delve deeper into the reasons scientists read scientific documents and the parts of documents on which they focus.
2.3 Survey

Scientists have some specificity in their information-seeking behaviour, as demonstrated above. They have specific needs, and they construct certain strategies to search for information. However, although the literature regarding the information-seeking behaviour of scientists is abundant, a more complete approach is necessary to address certain questions and to build a better information retrieval system for scientific literature. We designed an on-line survey at a very early stage of this research. The survey was constructed based on three hypotheses. The first hypothesis is that scientists seek and read scientific documents for different reasons. This hypothesis is based on the results of [Garvey et al. 1974, Palmer et al. 2007, Tenopir et al. 2009]. The second hypothesis is that scientists focus on different parts of a document rather than the whole document. These parts are the hypothesis, definition, methodology, results, opinion/critique and conclusion. The third hypothesis is that depending on the field of research, scientists do not seek information from the same parts of documents. For example, a researcher in the social sciences may read the same document as a researcher in medicine but may focus on different parts of the document.

2.3.1 Methodology

The survey (see appendix A) features 15 sets of questions in English (see appendix A). The first part is socio-demographic (age and sex) and concerns the domain of research (arts, social science, economic science, law, engineering, science, psychology and science of education, translation, theology, medicine and other). The next six questions are about the purpose of reading a document (to stay informed, to find all/some other studies in your research domain, to analyse how studies are conducted in your field, to compare studies, to find references for writing your own papers, to find ideas for your own research); the possible answers to these questions are yes, no or uncertain. The next set of six questions asks where users generally find the answers to their questions (hypothesis, definition, methodology, results, opinion/critique, conclusion); the possible answers to these questions are distributed on a Likert scale from always to never. The next question is "Do you often do a review of literature?", with the answers to the question limited to yes, no or uncertain. The next question is about the ease of finding the document for which the scientist is searching; the possible answers correspond to a Likert scale ranging from strongly agree (it is easy to find the document) to strongly disagree. The next question is about the use of a specialised database or search engine, such as PubMed or Citeseer, or the use of Google Scholar. This question asks which scientists use more often, Google Scholar or other specialised databases. Again, answers are measured on a Likert scale ranging from strongly agree to strongly disagree. The last question asks whether scientists want a retrieval system that would be able to find articles based on the content of their conclusions (i.e., all

\footnote{Some of the results presented in this section were published in [Ribaupierre & Falquet 2011, Ribaupierre 2012]}
articles stating that rain is wet), with answers measured on a Likert scale ranging from strongly agree to strongly disagree. The survey was sent by e-mail to 1,600 persons at the University of Geneva and other universities; only 91 responses were received (5.6% of answer), with 77 providing full answers.

2.3.2 Results

We analysed our data using SPSS. The sample was composed of 34 females and 57 males (with an average age of 39 years old). The population was not well distributed among different research areas: science (28), engineering (19), law (12), psychology and science of education (11), medicine (7), other (6), social science (5), economic science (1) and art (1). Therefore, for some of the hypotheses we posed, the analysis is difficult, and results could be biased because of the small number of persons in some groups. Using the Biglan [Biglan 1973] classification, the research areas were grouped into three categories: "hard" sciences (medicine, engineering and science), "soft" sciences (psychology, social science, economic science and art) and other.

2.3.2.1 Reasons to read scientific documents

Participants in this survey agreed with all the reasons that were proposed for reading an article (see table 2.1).

Table 2.1: "Why do you read scientific documents" (in percent, n=91)

<table>
<thead>
<tr>
<th>Why do you read scientific documents</th>
<th>yes</th>
<th>no</th>
<th>uncertain</th>
</tr>
</thead>
<tbody>
<tr>
<td>To find other studies in your research domain</td>
<td>94.5</td>
<td>2.2</td>
<td>3.3</td>
</tr>
<tr>
<td>To find references for writing your own papers</td>
<td>90.1</td>
<td>5.5</td>
<td>4.4</td>
</tr>
<tr>
<td>To stay informed</td>
<td>83.5</td>
<td>7.7</td>
<td>8.8</td>
</tr>
<tr>
<td>To compare studies</td>
<td>81.3</td>
<td>8.8</td>
<td>9.9</td>
</tr>
<tr>
<td>To analyse how studies are conducted in your field</td>
<td>72.5</td>
<td>9.9</td>
<td>17.6</td>
</tr>
<tr>
<td>To find ideas for your own research</td>
<td>71.4</td>
<td>13.2</td>
<td>15.4</td>
</tr>
</tbody>
</table>
We found few differences between the domains of research regarding the reasons why scientists read scientific documents (see table 2.2). We can observe that respondents in "hard" sciences seem to be more interested in reading to analyse how studies are conducted than are respondents in "soft" sciences or others. The respondents in "soft" sciences seem to read to find references for their own writing more than the respondents in "hard" sciences and "other" do. The respondents in "hard" sciences seem to read more to find new idea for their own research than do the respondents in "soft" sciences or other (see figure 2.8).

Table 2.2: "Why do you read scientific documents by area of research" (in percent n=91), "soft" science (SS), "hard" science (HS) and other (O)

<table>
<thead>
<tr>
<th>Why do you read scientific documents by area of research</th>
<th>Yes</th>
<th>No</th>
<th>Uncertain</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>To stay informed</td>
<td>SS</td>
<td>HS</td>
<td>O</td>
<td></td>
</tr>
<tr>
<td></td>
<td>87.1</td>
<td>81.5</td>
<td>83.3</td>
<td>9.7</td>
</tr>
<tr>
<td></td>
<td>5.6</td>
<td>16.7</td>
<td>3.2</td>
<td>10.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0</td>
<td></td>
<td>.424</td>
</tr>
<tr>
<td>To analyse how studies are conducted in your field</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>64.5</td>
<td>79.6</td>
<td>50.0</td>
<td>12.9</td>
</tr>
<tr>
<td></td>
<td>7.4</td>
<td>16.7</td>
<td>22.6</td>
<td>13.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>33.3</td>
<td>16.7</td>
<td>.418</td>
</tr>
<tr>
<td>To find all/some other studies in your research domain</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>93.5</td>
<td>96.3</td>
<td>83.3</td>
<td>3.2</td>
</tr>
<tr>
<td></td>
<td>0.0</td>
<td>16.7</td>
<td>3.2</td>
<td>3.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0</td>
<td></td>
<td>.117</td>
</tr>
<tr>
<td>To compare studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>80.6</td>
<td>83.3</td>
<td>66.7</td>
<td>9.7</td>
</tr>
<tr>
<td></td>
<td>7.4</td>
<td>16.7</td>
<td>9.7</td>
<td>9.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>16.7</td>
<td>16.7</td>
<td>.901</td>
</tr>
<tr>
<td>To find references for writing your own papers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>96.8</td>
<td>87.0</td>
<td>83.3</td>
<td>3.2</td>
</tr>
<tr>
<td></td>
<td>5.6</td>
<td>16.7</td>
<td>0.0</td>
<td>7.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0</td>
<td></td>
<td>.327</td>
</tr>
<tr>
<td>To find ideas for your own research</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>67.7</td>
<td>75.9</td>
<td>50</td>
<td>19.4</td>
</tr>
<tr>
<td></td>
<td>7.4</td>
<td>33.3</td>
<td>12.9</td>
<td>16.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>16.7</td>
<td>16.7</td>
<td>.298</td>
</tr>
</tbody>
</table>

Year of birth did not have a significant effect, but it appears that a difference exists between older scientists and younger scientists with respect to staying informed (see Table 2.3). It appears that older scientists are more likely to read scientific documents with the aim of staying informed than younger scientists are. No significant correlation was observed between age and research domain ($r = .101, N = 91, p.340$). It also appears that older and the youngest scientist are more likely to read scientific documents with the aim of funding references for writing their own papers (see figure 2.9).
Figure 2.8: "Why do you read scientific documents by area of research by area of research"; often answer

Figure 2.9: "Why do you read scientific documents by area of research by years of birth"; often answer
Table 2.3: "Why do you read scientific documents by area of research" (in percent n=91), year of birth

<table>
<thead>
<tr>
<th>Why do you read scientific documents by year of birth</th>
<th>Yes</th>
<th></th>
<th></th>
<th>No</th>
<th></th>
<th></th>
<th>Uncertain</th>
<th></th>
<th></th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year of birth</td>
<td>40-69</td>
<td>70-79</td>
<td>80-...</td>
<td>40-69</td>
<td>70-79</td>
<td>80-...</td>
<td>40-69</td>
<td>70-79</td>
<td>80-...</td>
<td></td>
</tr>
<tr>
<td>To stay informed</td>
<td>96.7</td>
<td>74.0</td>
<td>79.4</td>
<td>3.3</td>
<td>11.1</td>
<td>8.8</td>
<td>0</td>
<td>14.8</td>
<td>11.8</td>
<td>.179</td>
</tr>
<tr>
<td>To analyse how studies are conducted in your field</td>
<td>70.0</td>
<td>74.1</td>
<td>73.5</td>
<td>16.7</td>
<td>11.1</td>
<td>2.9</td>
<td>13.3</td>
<td>14.8</td>
<td>23.5</td>
<td>.374</td>
</tr>
<tr>
<td>To find all/some other studies in your research domain</td>
<td>93.3</td>
<td>92.6</td>
<td>91.1</td>
<td>3.3</td>
<td>3.7</td>
<td>0</td>
<td>3.3</td>
<td>3.7</td>
<td>2.9</td>
<td>.867</td>
</tr>
<tr>
<td>To compare studies</td>
<td>83.3</td>
<td>77.8</td>
<td>82.4</td>
<td>10.0</td>
<td>11.1</td>
<td>5.9</td>
<td>6.7</td>
<td>11.1</td>
<td>11.8</td>
<td>.897</td>
</tr>
<tr>
<td>To find references for writing your own papers</td>
<td>93.3</td>
<td>85.2</td>
<td>91.2</td>
<td>6.7</td>
<td>11.1</td>
<td>0</td>
<td>0</td>
<td>3.7</td>
<td>8.8</td>
<td>.167</td>
</tr>
<tr>
<td>To find ideas for your own research</td>
<td>73.3</td>
<td>66.7</td>
<td>73.5</td>
<td>20.0</td>
<td>14.8</td>
<td>5.9</td>
<td>6.7</td>
<td>18.5</td>
<td>20.6</td>
<td>.302</td>
</tr>
</tbody>
</table>
2.3. Survey

2.3.2.2 Do you read scientific papers to respond to a precise need

With regard to this issue, 83.5% of respondents read scientific documents for a specific purpose, whereas 16.5% do not.

2.3.2.3 Where do you generally find the answers

Responses (see table 2.4) were grouped into three categories: "always" and "very frequently" responses were grouped into the category Often; "occasionally" and "rarely" responses were grouped into the category Sometimes; and "very rarely" and "never" responses were grouped into the category Never.

Table 2.4: "Where do you generally find your answers" (in percent, n=77)

<table>
<thead>
<tr>
<th></th>
<th>always</th>
<th>sometimes</th>
<th>never</th>
</tr>
</thead>
<tbody>
<tr>
<td>In results</td>
<td>75.0</td>
<td>23.7</td>
<td>1.3</td>
</tr>
<tr>
<td>In methodology</td>
<td>59.7</td>
<td>36.4</td>
<td>3.9</td>
</tr>
<tr>
<td>In conclusion</td>
<td>56.6</td>
<td>35.5</td>
<td>7.9</td>
</tr>
<tr>
<td>In definition</td>
<td>50.6</td>
<td>44.2</td>
<td>5.2</td>
</tr>
<tr>
<td>In opinion/critic</td>
<td>42.1</td>
<td>46.1</td>
<td>11.8</td>
</tr>
<tr>
<td>In hypothesis</td>
<td>40.3</td>
<td>46.8</td>
<td>13</td>
</tr>
</tbody>
</table>

We tested whether the domains of research differed with respect to the parts of documents on which scientists focus (see table 2.5). All domains indicated interest in the results section. Respondents in the field of "hard" sciences appeared to be more interested in the methodology section than those in "soft" sciences and other. Respondents in other and "soft" sciences appeared to be more interested in the conclusion section than those in "hard" sciences. No real differences were observed between the domains for the definition section. Respondents in "soft" sciences appeared to be more interested in the opinion/critique section. No differences were observed between the domains with respect to the hypothesis section (see figure 2.10). The only significant differences observed were for the opinion/critique section, where $p = 0.41$, and the methodology section, where the difference was marginally significant with $p = 0.63$.

We tested also whether the age of the respondents differed with respect to the parts of the documents (see table 2.6 and figure 2.11). We observed that young scientists are focusing more for the methodology with a significant difference ($p = 0.047$). We also found a small difference for the hypothesis, where older scientists seems to be more interested, the difference is not significant ($p = 0.090$).
Table 2.5: "Where do you generally find your answers by area of research" (in percent, n=77) "soft" science (SS), "hard" science (HS) and other (O)

<table>
<thead>
<tr>
<th></th>
<th>Often</th>
<th>Sometimes</th>
<th>Never</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SS</td>
<td>HS</td>
<td>O</td>
<td>SS</td>
</tr>
<tr>
<td>Opinion/critic</td>
<td>64.3</td>
<td>30.2</td>
<td>20.0</td>
<td>25.0</td>
</tr>
<tr>
<td>Hypothesis</td>
<td>46.4</td>
<td>38.6</td>
<td>20.0</td>
<td>35.7</td>
</tr>
<tr>
<td>Definition</td>
<td>53.6</td>
<td>50.0</td>
<td>40.0</td>
<td>35.7</td>
</tr>
<tr>
<td>Methodology</td>
<td>46.4</td>
<td>70.5</td>
<td>40.0</td>
<td>42.9</td>
</tr>
<tr>
<td>Result</td>
<td>71.4</td>
<td>74.4</td>
<td>100</td>
<td>25.0</td>
</tr>
<tr>
<td>Conclusion</td>
<td>64.3</td>
<td>48.8</td>
<td>80.0</td>
<td>25.0</td>
</tr>
</tbody>
</table>

Table 2.6: "Where do you generally find your answers" by year of birth (in percent, n=77)

<table>
<thead>
<tr>
<th>Year of Birth</th>
<th>Often</th>
<th>Sometimes</th>
<th>Never</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Opinion/critic</td>
<td>47.8</td>
<td>39.1</td>
<td>40.0</td>
<td>43.4</td>
</tr>
<tr>
<td>Hypothesis</td>
<td>60.9</td>
<td>26.1</td>
<td>35.5</td>
<td>26.1</td>
</tr>
<tr>
<td>Definition</td>
<td>43.5</td>
<td>52.2</td>
<td>54.8</td>
<td>43.5</td>
</tr>
<tr>
<td>Methodology</td>
<td>39.1</td>
<td>78.3</td>
<td>61.3</td>
<td>52.2</td>
</tr>
<tr>
<td>Result</td>
<td>78.3</td>
<td>69.9</td>
<td>57.9</td>
<td>17.4</td>
</tr>
<tr>
<td>Conclusion</td>
<td>60.9</td>
<td>56.5</td>
<td>53.3</td>
<td>30.4</td>
</tr>
</tbody>
</table>

Figure 2.11: "Where do you generally find your answers by years of birth"; often answer
Figure 2.10: "Where do you generally find your answers by area of research"; often answer

![Graph showing the distribution of answers by area of research.](image)

Legend:
- Soft science
- Hard science
- Other
2.3.2.4 Do you often conduct a review of the literature

In response to the question, "Do you often conduct a review of the literature", 61.5% answered yes and 38.5% answered no. Table 2.7 shows that respondents in "soft" sciences review the literature less often than do those in "hard" sciences, with the difference being marginally significant ($p = .691$). Regarding the section in which respondents are most interested (see table 2.8), the only significant difference observed was for the results section ($p = .019$).

Table 2.7: Do you often conduct a review of the literature (in percent, $n = 77$)

<table>
<thead>
<tr>
<th></th>
<th>&quot;soft&quot; science</th>
<th>&quot;hard&quot; science</th>
<th>other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>32.1</td>
<td>62.5</td>
<td>5.4</td>
</tr>
<tr>
<td>No</td>
<td>37.1</td>
<td>54.3</td>
<td>8.6</td>
</tr>
</tbody>
</table>

Table 2.8: "Do you often conduct a review of the literature?" by "Where do you generally find answers" ?[Result] (in percent, $n = 77$)

<table>
<thead>
<tr>
<th>Do you often conduct a review of the literature?</th>
<th>Do you find generally your answer in the result part?</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Often</td>
<td>Sometime</td>
<td>Never</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>84.0</td>
<td>14.0</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>75.9</td>
<td>23.7</td>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>75.0</td>
<td>18.0</td>
<td>1.3</td>
<td></td>
</tr>
</tbody>
</table>
2.3. Survey

2.3.2.5 Do you easily find papers/books you are looking for without too much noise (e.g., a paper that is out of the field in which you are searching)

Regarding the question "Do you easily find papers/books you are looking for without too much noise", 60.4% agreed, 16.5% strongly agreed, 15.4% neither agreed nor disagreed, 6.6% disagreed and 1.1% strongly disagreed. No significant difference was observed between the domains of research or between years of birth.

2.3.2.6 Do you use Google Scholar more often than a specialised database (in your domain)

Regarding the question "Do you use Google Scholar more often than a specialised database", 19.8% strongly agreed, 20.9% agreed, 16.5% neither agreed nor disagreed, 29.7% disagreed and 13.7% strongly disagreed. No significant difference was observed between the domains of research or between years of birth.

2.3.2.7 Would you like a retrieval system that would be able to find articles by the content of their conclusions (i.e., all articles stating that rain is wet)

Regarding the final question, "Would you like a retrieval system that would be able to find articles by the content of their conclusions", 26.4% strongly agreed, 30.8% agreed, 26.4% neither agreed nor disagreed, 8.8% disagreed and 7.7% strongly disagreed. No significant difference was observed concerning the domains of research, but a significant difference (see figure 2.12) was observed with respect to the year of birth ($p = .033$). The percentage of young persons indicating interest in this type of system was greater than the percentage of older persons indicating the same. No significant difference was observed concerning the difficulty scientists encounter in finding their documents.

We can observe in the cross-table (see Figure 2.13) between "What is the reason(s) you are reading scientific documents" and the parts of documents scientists read that there are significant positive relationships between "to stay informed" and "results"; "to find references for writing your own papers" and "conclusion"; "to analyse how studies are conducted in your field" and "definition" and "methodology"; and "to find ideas for your own research" and "definition" and "hypothesis."
Figure 2.12: "Would you like a retrieval system that would be able to find articles by their conclusions"
### Figure 2.13: Relationship between task and what parts are read

<table>
<thead>
<tr>
<th>What is/are the reason(s) why you are reading scientific</th>
<th>Hypothesis</th>
<th>Definition</th>
<th>Methodology</th>
<th>Result</th>
<th>Conclusion</th>
<th>Opinion/Criticism</th>
</tr>
</thead>
<tbody>
<tr>
<td>To stay informed</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Yes</td>
<td>78.8%</td>
<td>21.2%</td>
<td>87.9%</td>
<td>12.1%</td>
<td>83.3%</td>
<td>16.7%</td>
</tr>
<tr>
<td>No</td>
<td>75.0%</td>
<td>25.0%</td>
<td>100.0%</td>
<td>0.0%</td>
<td>75.0%</td>
<td>25.0%</td>
</tr>
<tr>
<td>Uncertain</td>
<td>71.4%</td>
<td>28.6%</td>
<td>85.7%</td>
<td>14.3%</td>
<td>100.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>To find all/some other studies in your research domain</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Yes</td>
<td>78.4%</td>
<td>21.6%</td>
<td>89.2%</td>
<td>10.8%</td>
<td>83.8%</td>
<td>16.2%</td>
</tr>
<tr>
<td>No</td>
<td>40.0%</td>
<td>60.0%</td>
<td>40.0%</td>
<td>60.0%</td>
<td>60.0%</td>
<td>40.0%</td>
</tr>
<tr>
<td>Uncertain</td>
<td>66.7%</td>
<td>33.3%</td>
<td>66.7%</td>
<td>33.3%</td>
<td>100.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>To analyse how studies are conducted in your field</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Yes</td>
<td>80.7%</td>
<td>19.3%</td>
<td>91.2%*</td>
<td>8.8%</td>
<td>89.5%*</td>
<td>10.5%</td>
</tr>
<tr>
<td>No</td>
<td>40.0%</td>
<td>60.0%</td>
<td>40.0%</td>
<td>60.0%</td>
<td>60.0%</td>
<td>40.0%</td>
</tr>
<tr>
<td>Uncertain</td>
<td>80.0%</td>
<td>20.0%</td>
<td>93.3%</td>
<td>6.7%</td>
<td>73.3%</td>
<td>26.7%</td>
</tr>
<tr>
<td>To compare studies</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Yes</td>
<td>78.8%</td>
<td>21.2%</td>
<td>87.9%</td>
<td>12.1%</td>
<td>84.8%</td>
<td>15.2%</td>
</tr>
<tr>
<td>No</td>
<td>75.0%</td>
<td>25.0%</td>
<td>75.0%</td>
<td>25.0%</td>
<td>75.0%</td>
<td>25.0%</td>
</tr>
<tr>
<td>Uncertain</td>
<td>71.4%</td>
<td>28.6%</td>
<td>100.0%</td>
<td>0.0%</td>
<td>85.7%</td>
<td>14.3%</td>
</tr>
<tr>
<td>To find references for writing your own papers</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Yes</td>
<td>78.4%</td>
<td>21.6%</td>
<td>87.8%</td>
<td>12.2%</td>
<td>83.8%</td>
<td>16.2%</td>
</tr>
<tr>
<td>No</td>
<td>0.0%</td>
<td>100.0%</td>
<td>0.0%</td>
<td>100.0%</td>
<td>0.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Uncertain</td>
<td>100.0%</td>
<td>0.0%</td>
<td>100.0%</td>
<td>0.0%</td>
<td>100.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>To find ideas for your own research</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Yes</td>
<td>84.2%</td>
<td>15.8%*</td>
<td>93.0%*</td>
<td>7.0%</td>
<td>86.0%</td>
<td>14.0%</td>
</tr>
<tr>
<td>No</td>
<td>33.3%</td>
<td>66.7%</td>
<td>55.6%</td>
<td>44.4%</td>
<td>66.7%</td>
<td>33.3%</td>
</tr>
<tr>
<td>Uncertain</td>
<td>81.8%</td>
<td>18.2%</td>
<td>90.9%</td>
<td>9.1%</td>
<td>90.9%</td>
<td>9.1%</td>
</tr>
</tbody>
</table>

*p<0.05
**p<0.1
2.3.3 Survey Discussion and Conclusion

The reasons scientists read scientific documents appear to be the same as those indicated in [Garvey et al. 1974]. It appears that scientists in "hard" sciences more often read to find how a study was conducted in their field and to find new ideas. In contrast, scientists in "soft" sciences appear to read to find references to cite in their own papers. Few differences between the social sciences and physics in the study of [Garvey et al. 1974] were found regarding the nature of the information needed. In addition, older scientists appear to read more to stay informed and older and youngest scientists seem to read to find references for their own papers than the middle age categories do.

It can be observed that scientists are more likely to find their answers in certain types of sentences or paragraphs than in others. Among the proposed parts of documents, scientists focus primarily on the "results" section of scientific documents. Certain significant differences can be observed: scientists in the "soft" sciences more often find their answers in the "opinion/critique" section, which is due to the fact that such content ("opinion/critique") is observed more frequently in the "soft" sciences than in the "hard" sciences. In contrast, scientists in the "hard" sciences appear to read the "methodology" section more often than do scientists in the "soft" sciences. Older scientists read the "hypothesis" section more often than younger scientists do, but this difference is not significant. In contrast, younger scientists appear to read the "methodology" section more often than older scientists do. Scientists who review the literature more often appear to find their answers in the "results" section more often than other scientists do.

Because this study did not answer all of the questions we proposed, especially questions that require more precise information regarding how scientists seek or read information, we conducted interviews to better understand these processes.

2.4 Interview

The questionnaire was based on several hypotheses:

- Scientists search through scientific documents to find an answer to a specific question
- Scientists search through specific chunks of scientific documents
- Scientists search through specific chunks of scientific documents depending of the task they have to fulfil
- Scientists read scientific documents for a specific purpose
- Scientists read specific fragments to answer to specific questions
- Scientists read different types of fragments depending on the tasks they intend to complete

Some of the results presented in this section were published in [Ribaupierre 2012, Ribaupierre & Falquet 2013]
2.4.1 Methodology

We conducted semi-directive interviews with scientists from different areas of research and with different levels of research expertise. The interviews were conducted in English or in French.

The sample consisted of 10 participants (see table 2.9). A large part of the interview was devoted to determining the sections on which the scientists focused.

Table 2.9: Sample Population

<table>
<thead>
<tr>
<th>Subject</th>
<th>Year of birth</th>
<th>Gender</th>
<th>Field of research</th>
<th>Highest education degree, if PhD year</th>
<th>Nb of article written as first author</th>
<th>Total nb of article published</th>
<th>Nb of article read by month</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>1983</td>
<td>M</td>
<td>Computer science</td>
<td>Master</td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>P2</td>
<td>1978</td>
<td>F</td>
<td>Computer science</td>
<td>2009</td>
<td>15</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>P3</td>
<td>1976</td>
<td>F</td>
<td>Linguistic/Computer science</td>
<td>2008</td>
<td>17</td>
<td>29</td>
<td>25</td>
</tr>
<tr>
<td>P4</td>
<td>1961</td>
<td>M</td>
<td>Computer science/health</td>
<td>Master</td>
<td>10</td>
<td>20</td>
<td>15</td>
</tr>
<tr>
<td>P5</td>
<td>1976</td>
<td>F</td>
<td>Information System</td>
<td>Master</td>
<td>1</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>P6</td>
<td>1946</td>
<td>F</td>
<td>Psychology</td>
<td>1975</td>
<td>68</td>
<td>141</td>
<td>6</td>
</tr>
<tr>
<td>P7</td>
<td>1957</td>
<td>M</td>
<td>Physics/Science History</td>
<td>1987</td>
<td>15</td>
<td>60</td>
<td>50</td>
</tr>
<tr>
<td>P8</td>
<td>1985</td>
<td>F</td>
<td>Traduction &amp; communication</td>
<td>Master</td>
<td>2</td>
<td>3</td>
<td>20</td>
</tr>
<tr>
<td>P9</td>
<td>1983</td>
<td>F</td>
<td>Social Science, Gender studies</td>
<td>Master</td>
<td>1</td>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>P10</td>
<td>1974</td>
<td>F</td>
<td>Medecine</td>
<td>2000</td>
<td>20</td>
<td>28</td>
<td>25</td>
</tr>
</tbody>
</table>

We constructed a questionnaire composed of 23 questions (see Appendix B), which formed the basis of the interview. The interview lasted one to two hours. The first 13 questions were about socio-demographic data and the researchers’ level of expertise. The second part focused on search patterns, what type of IR system the researchers used, what type of information the researchers searched for, and what types of questions the researchers asked themselves before querying an IR system.

The third part concerned the reading behaviour of the researchers, such as the specific part of a document (fragment) on which the researchers focused when they did not read the entire article. Specific questions were asked to investigate whether
readers looked at different fragments depending on their ultimate task. We defined a list of tasks that scientists can achieve while reading a document. This list was discussed between people within the research group and validated during the interviews.

2.4.2 Results

2.4.2.1 Reasons to search for and read a scientific document

We established a list of reasons why scientists seek and read scientific documents. During the interviews, we discussed this list with the interviewees, and they were able to add new reasons during the interviews. The reasons that the interviewees agreed most strongly with were (in descending order):

- to enhance their knowledge
- to write papers
- to find a new idea/inspiration
- to discover another point of view
- look at a new idea
- to stay informed
- because they were advised by someone
- to compare documents
- to write a state-of-the-art review
- to find a specific definition
- curiosity
- to provide a new point of view
- to compare methodologies
- to provide expertise/review
- to summarise their knowledge
- to be able to take a position in a document
- because it was referenced by another document
- to resolve a problem
- to build a model
- to determine whether an idea was new or whether someone else had already worked on it.
2.4.2.2 How are you searching a scientific document

The question posed concerned how researchers search for documents. The most common answer was searching by entering keywords and metadata into a search engine or a specialised database. It was also observed that, in general, the researchers had to reformulate their queries. Some interviewees elaborated on their responses. For example, a researcher answered that she was formulating a research hypothesis in her mind before processing it into a query with keywords. To find keywords, an interviewee stated that she initially used keywords she knew and reformulated her query with the keywords she acquired in the previous iteration of her query. The researchers also indicated that they encountered difficulties in finding the correct keywords. An interviewee noted that she used keywords that experts around her provided when she was stuck in a search session. Another way to search for information is to examine the references in a document that the researcher has already read or is reading. Another respondent stated that he followed the authors or the references in documents and that the search process was like "unwinding a ball of wool". Other methods indicated by some interviewees included obtaining a colleague's advice on what to read, such as receiving information through mailing lists or by speaking directly to someone, and focusing on reading articles only within a list of pre-selected journals. Researchers can also follow authors by searching their websites, by entering their name as keywords or by searching the proceedings of a conference. One interviewee reported usually searching journals with science popularisation articles. In general, the researchers had more than one strategy at their disposal to search for information, and they combined these strategies to fulfill their information needs.

2.4.2.3 What are the databases or search engines you use to search for documents

The interviewees mostly used Google Scholar to search for documents, followed by the databases of publishers such as Elsevier, Springer, ACM, and IEEE. Interviewees also mentioned certain specialised databases or search systems, such as Academia, TEL, RERO, the electronic archive of the University of Geneva, and Nebis. A participant in the medical field answered that she most often searched for articles in PubMed. Again, the interviewees generally mixed their sources of information depending on where they could most easily find information and the access available from their workplace.

2.4.2.4 Do you search documents for a precise question

The researchers all stated that they searched for documents for a specific purpose. We asked them to provide one or two examples (see table 2.10).
Table 2.10: Specific purpose for searching for information

<table>
<thead>
<tr>
<th>Person</th>
<th>Precise question</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>See how data knowledge warehouse resources store and reason. Find the implementation of these data warehouses, all the details on it and the reasons for the choice of platforms and techniques. Find all rigorous discussions of the representation of visualisation techniques with access to the model and that ontologies are accessible.</td>
</tr>
<tr>
<td>P2</td>
<td>Does it work to transfer the annotations in a parallel corpus when some parsing has already been done? I want the methodology and the results</td>
</tr>
<tr>
<td>P3</td>
<td>I want the state of the art on bedsores and bandaging techniques The definition and the utilisations of the archetypes</td>
</tr>
<tr>
<td>P4</td>
<td>How do we produce the different facets of a Information System, and the definition I want to know if there is a definition for semantic homogeneity and if we can calculate it</td>
</tr>
<tr>
<td>P5</td>
<td>Intra-individual variability in terms of the brain and behaviour. Method of analysis, which are the indices of variability. Brain imaging method; what are the methods and how the data were collected</td>
</tr>
<tr>
<td>P6</td>
<td>Who has written something about Gustave Juvet I have not read Who are the authors who were climate credulous</td>
</tr>
<tr>
<td>P7</td>
<td>What is the state of the art on mixity in education I want to find all the articles where Christine Delphy disagrees with Patricia Roux</td>
</tr>
<tr>
<td>P8</td>
<td>Who has applied the theory of pertinence to translation Who are the authors who have applied discourse analysis of journalistic translation, and how have they applied the methodology? Who wrote about the translation of journalistic texts in the last 10 years?</td>
</tr>
<tr>
<td>P9</td>
<td>Gender, historicity of the concept, its various definitions and how it evolves and anchors culturally. Highlight the main controversy Social transactions, historicity of the concept, its various definitions and how it evolves and anchors culturally. Highlight the main controversy</td>
</tr>
<tr>
<td>P10</td>
<td>Variability within individual tasks in working memory Vulnerability among the elderly</td>
</tr>
</tbody>
</table>
2.4. Interview

2.4.2.5 For each task, do you search for documents differently or with a different search engine or database

All of the researchers answered that they searched for information differently depending on the task they had to complete (see table 2.11). In general, when the interviewees stated that they searched by entering keywords or metadata, they meant that they performed queries using their preferred search engines or databases. For some tasks, the researchers used the same strategy they used when they did not have a specific task to complete, but for certain other tasks, they used different strategies or adapted their initial strategies. Some participants answered that they sometimes encountered difficulties in finding documents when, for example, they were searching for definitions or when their aim was to write a state-of-the-art review for one of two reasons: they did not have a system that allowed them to search for definitions or for specific passages or it was difficult to perform an exhaustive search.
Table 2.11: How people search for information as a function of task

<table>
<thead>
<tr>
<th>No</th>
<th>Task</th>
<th>Interview Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Keeping track</td>
<td>following references in the document, website of researcher in the area of research, proceedings of conferences, browsing, keywords and dates, Wikipedia, popularisation website or journal</td>
</tr>
<tr>
<td>2</td>
<td>Discover new area of research</td>
<td>specialised databases, website of researcher in the area of research, following references in the document, recommendation from colleagues, keywords, interaction with specialised person, popularisation website</td>
</tr>
<tr>
<td>3</td>
<td>Writing a state-of-the-art review</td>
<td>personnel databases, website of researcher in the area of research, proceedings of conferences, recommendation from colleagues, keywords, Wikipedia, popularisation journal, following authors publications, documents that are already surveyed</td>
</tr>
<tr>
<td>4</td>
<td>Finding new results/findings</td>
<td>specialised database, following authors publications, proceedings of conferences, metadata, email list, keyword+findings, popularisation journal, following references in the document, recommendation</td>
</tr>
<tr>
<td>5</td>
<td>Finding a definition</td>
<td>looking into all documents: theoretical part, keywords, proceedings of conferences, dictionary, keywords+ term &quot;definition&quot;, Wikipedia, library services, define in Google, keywords</td>
</tr>
<tr>
<td>6</td>
<td>Finding a hypothesis</td>
<td>specialised database, following authors publications, library services, proceedings of conferences, keywords, keywords+ term &quot;hypothesis&quot;, popularisation journal, I don’t look at that</td>
</tr>
<tr>
<td>7</td>
<td>Finding a methodology</td>
<td>personnel databases, following authors publications, reference books, library services, keywords+ term &quot;methodology&quot;, keywords</td>
</tr>
<tr>
<td>8</td>
<td>Finding new ideas</td>
<td>recommended papers, interaction with colleagues, browsing, user, Tweets, Ted talks</td>
</tr>
<tr>
<td>9</td>
<td>Learning to build a background of knowledge</td>
<td>recommendation, library services, keywords + metadata, proceedings of conferences, looking for tutorial, books instead of articles, interaction with colleagues, following author publications, browsing</td>
</tr>
<tr>
<td>10</td>
<td>Writing new project</td>
<td>specialised databases, browsing, project that I have already written, website of researcher in the area of research, keywords, proceedings of conferences, popularisation journal, recommendation, following author publications</td>
</tr>
</tbody>
</table>
2.4.2.6 Does the knowledge you have in a domain change the way you search for documents

Regarding the question "Does the knowledge you have in a domain change the way you are searching documents", all researchers responded affirmatively. For example, subject P4 answered as follows: "I refine my request when I know the area better; if not, I have more general keywords". P7 answered as follows: "The procedure is easier in an area I know well". P1 answered as follows: "I am not looking into the same networks, but my strategy is the same".

2.4.2.7 Do you focus on specific fragments or parts of a document, and which ones do you prioritise

The responses provided most frequently were the following: findings, hypothesis, methodology, definition and abstract (see figure 2.14). To consider the importance of these fragments to the scientists, we examined the number of times a person cited one fragment during the interview, and we divided that figure by the number of times they cited any fragment, which allowed us to compare people who spoke more with those who spoke less. We were therefore able to observe some differences between the direct question the researchers sought to answer and to more deeply analyse the interview responses. We could also determine what was most important for the users.

Figure 2.14: Fragment read by person
2.4.2.8 Do you focus on different parts of a document depending on the task you must complete

Depending on the task, scientists do not read the same types of fragments (see Table 2.15). Some tasks appear to require researchers to search more than a single fragment to find the information that allows them to complete these tasks.
### Figure 2.15: Fragments on which users focus as a function of task

<table>
<thead>
<tr>
<th></th>
<th>Keeping up-to-date</th>
<th>Discovering new area of research</th>
<th>Writing a state-of-the-art review</th>
<th>Finding new findings</th>
<th>Finding a definition</th>
<th>Finding a hypothesis</th>
<th>Finding a methodology</th>
<th>Finding a new idea</th>
<th>Learning</th>
<th>Writing a new project</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Findings</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>36</td>
</tr>
<tr>
<td>Whole article</td>
<td>8</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>36</td>
</tr>
<tr>
<td>Abstract</td>
<td>3</td>
<td>1</td>
<td>8</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>17</td>
</tr>
<tr>
<td>Methodology</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>Introduction</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>Background</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>Hypothesis</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>Definition</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Central problematic</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Positioning</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Discussion</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Title</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>New idea</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Aim/objective</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>References</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Image</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
2.4.2.9 Do you read documents differently depending on the knowledge you have in the domain, and do you read specific sections

All researchers answered yes. One interviewee answered that when he knows the domain well, he stops reading the document sooner if it does not correspond to what he is looking for. If he does not know the domain well, he reads the introduction first, and if it seems interesting, he reads the whole article linearly. Another interviewee stated that she reads a document more carefully when she does not know the domain; she quickly reads the methodology section but carefully reads the introduction, results and conclusion. Another interviewee answered that she focuses less on the methodology when she is not as familiar with a domain and focuses more on the conclusion, the hypothesis and the results but not the raw results. In contrast, when she knows the domain well, she focuses more attentively on the methodology and the raw results.

2.4.3 Interview, Discussion and Conclusion

The results of this study demonstrate that scientists search for documents using different strategies and that strategies can vary depending on the tasks researchers must accomplish using the document they are seeking. The researchers described encountering some difficulties in finding the most appropriate documents, especially when they wanted to find a specific methodology, definition or hypothesis and when they performed a literature search to write a state-of-the-art review. The researchers also answered that their knowledge of the area of research changed the way they searched for documents and that it was easier to search for documents when they knew the domain well. Most of the researchers used multiple databases (e.g., ACM, IEEE, and Nebis) specialised in their domain as well as Google Scholar, and they varied with respect to which database they used as a function of the tasks they had to complete. This behaviour differed from that reported by the medical expert, who mostly used PubMed. We were also able to validate the list of tasks that scientists wanted to achieve when reading scientific documents.

The results show that scientists read articles for different purposes: to write an article, to discover new points of view, because they were advised by someone or to enhance their knowledge. The research areas of the interviewees did not appear to make a difference in the researchers’ responses. In some interviews, the subjects appeared to have more difficulty answering the question regarding why they read scientific documents. We believe that this difficulty might arise because the behaviour is so embedded in their practice. These results are consistent with those of Reanear et al. [Renear & Palmer 2009b].

Scientists sometimes search for and read only fragments of documents. With respect to their purposes for reading, the area of research did not seem to make a difference. Fragments are prioritised as follows: findings and hypothesis, methodology, definition, and abstract. We found a gap between these five fragments and others, specifically, introduction, aim, background, positioning, figures, model and
2.4. Interview

Scientists also focus on different fragments depending on the task they want to accomplish. To keep up to date with other research, scientists focus on findings without looking at other parts of a document. When they set out to discover a new area of research, the interviewees indicated that their focus was more widely distributed between fragments. By taking into account only fragments that more than one person uses as a function of task, we were able to simplify the table (see Figure 2.16). When reading to keep track of other studies, researchers read the findings, abstract or the methodology. When reading with the aim of writing an article, researchers are more likely to read whole articles rather than only parts. In contrast, to find new ideas, the interviewees' answers varied between reading fragments and whole articles. To acquire new knowledge (learning), the researchers typically read whole articles or sometimes focused on the findings, methodology and background. To discover new areas of research, the researchers' answers varied between findings, background, hypothesis, references, definition, and the whole document. Furthermore, when planning to start a new project, although not all interviewees indicated that they performed this task, most also read whole articles.

It was also observed that sometimes the differences between fragments were more difficult for the respondents to identify because of the overlap between structural fragments and semantic fragments. For example, the introduction is considered more of a structural fragment than a semantic one. When an article has an introduction section, it is composed of multiple semantic fragments, such as background, aim of the research, and central problem.
<table>
<thead>
<tr>
<th>Figure 2.16: Fragment by tasks</th>
<th>Definition</th>
<th>Reference</th>
<th>Hypothesis</th>
<th>Background</th>
<th>Methodology</th>
<th>Abstract</th>
<th>Findings</th>
<th>Whole article</th>
</tr>
</thead>
<tbody>
<tr>
<td>Writing new</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learning</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Finding new idea</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Writing a article</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Research of new area</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Keeping track</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2.5 Discussion and Conclusion

The results of the studies demonstrate that scientists seek and read documents differently depending on the tasks they must complete. The results also show that to be efficient in completing certain tasks, scientists do not read entire documents but only read some parts of documents. They also search for documents according to those parts and depending on the tasks they must accomplish using the information contained in the documents. Figure 2.17 shows a comparison between the interview and survey results regarding the parts of documents that scientists read most often. By comparing the two studies, it can be observed that some parts of scientific documents appear to be more important. The difference in the ranking of the most important parts may be due to the differences among the population, the number of persons that participated in the studies, or the fact that by proposing an answer to the scientists instead of allowing them to provide an answer freely, the responses were affected. Although these pilot studies could have been conducted with more participants and using different questions, the results support some of our hypotheses: scientists read some parts of documents more than they do others, but they also seek information by incorporating those parts into their query parameters.

Figure 2.17: Fragment; comparison survey/interview

As indicated by the results of the two studies, scientists focus on different fragments of documents depending on the task they want to accomplish. The parts indicated by the respondents were not exactly the same between the two studies. The manner in which answers were provided, open (interview) vs. closed (survey), may have affected the results. Figure 2.18 shows the relationship between tasks and the parts of documents read by researchers as indicated by the interviews and surveys. It can be observed that some answers overlap between the two studies. The
results of both studies confirm that scientists do not focus on entire documents for every task they must perform using the documents they seek. This behaviour may be a factor to consider in the construction of models concerning indexing, queries, the display of information, and evaluation.

The size of the sample used in each study was small, and both studies should be extended to be more representative. For the interviews, larger sampling is required (40 person minimum) to be more representative. The limitation of the survey was not the size of the sample but the representation of the research domain area in each group. Consequently, to evaluate whether differences exist between each area of research, the number of participants should also be increased. Although the sample is not representative, the results of the two studies provide a sufficiently solid foundation for developing an empirical annotation model. In future studies, the use of a larger sample for the interviews would be interesting.
## 2.5. Discussion and Conclusion

Figure 2.18: Relationship between tasks and parts read, as indicated by interview and survey.
3.1 Introduction

To search for a scientific document or for a particular fragment of information in a scientific document, metadata, annotation or indexation is needed. In the first part of this chapter, we will discuss some of the different models and studies reported
in the literature that were conceptualised or implemented to build better information retrieval systems, summarisation systems, and digital libraries. To easily understand the contributions of the various studies, we have classified the different models proposed into five dimensions. Based on this classification, we can identify several models or studies that overlap along more than one dimension.

1. Conceptual dimension: Ontology or controlled vocabularies that describe scientific terms or concepts (conceptual indexing)


3. Structural dimension: description of the structural part of a document or logical structure (e.g., paragraph, chapter, section, introduction, abstract)

4. Rhetorical or discursive dimension: description of discursive or rhetorical elements

5. Relationships dimension: description of the citations and relationships between documents

The second part of this chapter describes the model (SciAnnotDoc) that we have proposed and developed in this thesis and the different ontologies linked to this model. To illustrated the model, we chose gender studies documents as use cases. We chose this domain because it consists of very heterogeneous written documents ranging from highly empirical studies to "philosophical" texts.

3.2 State of the art

3.2.1 Conceptual dimensions (1)

The first dimension concerns the use of knowledge bases to represent the terms contained in scientific documents by the ID of the concepts instead of ambiguous words. In the literature, this dimension is often referred to as conceptual indexing\(^1\) or semantic indexing [Biemann 2005, Mihalcea & Moldovan 2000, Baziz 2005, Radhouani 2008, Radhouani & Falquet 2006]. In conceptual indexing, a document is represented by a set of concepts that belong to one or several general or domain ontologies.

Multiple ontologies are available on the Web\(^2\). In 2012, [Aquín & Noy 2012] reported that a total of 2,937 ontologies were available on different portals. Several "hard" science domain ontologies are available\(^3\), such as GO\(^4\) and OBO\(^5\). EXPO

\(^1\) The conceptual indexing consists in representing documents (queries) by concepts instead of words. Thus, during the retrieval process, the matching between a query and a document is done based on a non-ambiguous vocabulary (concepts).\(^6\)[Radhouani et al. 2008, p.3]

\(^2\) with services such as Watson or Swoogle

\(^3\) Ontologies in research areas such as sociology, psychology, gender studies were non-existent, to the best of our knowledge, at the beginning of this thesis

\(^4\) http://www.geneontology.org

\(^5\) http://www.obofoundry.org
[Soldatova & King 2006] is a complex ontology that describes experiments in "hard science"; it is very formal and is not compatible with sciences that are less formal, such as sociology or gender studies. DBpedia\(^6\) is a generic ontology that has formalised the information provided by Wikipedia, but it may be too generic to be used for scientific documents.

One of the most frequently used information retrieval systems for scientific documents that operates with semantic indexing is PubMed\(^7\). PubMed uses UMLS\(^8\), which, in turn, uses controlled vocabularies or ontologies such as MeSH\(^9\) and SNOMED\(^10\).

Depending on the complexity of a knowledge base and the structure of text, studies have shown that it is possible to build an ontology semi-automatically or automatically [Khan & Luo 2002, Alani et al. 2003, Jilani et al. 2009, Osborne et al. 2013, Osborne & Motta 2012]. In [Falquet et al. 2004], the authors propose automatically populating an ontology via a classification algorithm using the terms found in documents.

Conceptual indexing has some limitations: 1) the pragmatic of the knowledge is unknown, and 2) concepts are presented as a "bag of concepts", and this results in a lack of "deep" semantics of concepts.

### 3.2.2 Meta-data dimension (2)

The second dimension describes bibliographic data. Some generic models have been proposed that do not focus solely on scientific documents. The use of bibliographic data can improve the strategic search for information and may allow for the definition of a ranking system for classifying documents.

MARC21 was developed by the Library of Congress\(^11\). MARC21 provides a complete but complex description of bibliographic metadata using code numbers to describe data. In [Weinstein & Birmingham 1998], the authors propose a procedure for converting MARC records to an ontological knowledge base. In [Styles et al. 2008], the authors formalise MARC21 as an RDF vocabulary for use in Semantic Web applications. However, to the best of our knowledge, this RDF vocabulary was never used in such applications. MARC21 presents some inconveniences, such as its high complexity and its inability to be easily read by humans.

Dublin Core was developed in 1995 as a result of a workshop sponsored by the Online Computer Library Center (OCLC) and the National Center for Supercomputing Applications (NCSA)\(^12\). The aim of this workshop was to model a set of metadata elements to describe networked resources. The DCMI (International Conference on Dublin Core and Metadata Applications) is held each year, and the

\(^6\)http://dbpedia.org/About
\(^7\)http://www.nlm.nih.gov/pubs/factsheets/pubmed.html
\(^8\)http://www.nlm.nih.gov/research/umls/
\(^10\)http://www.nlm.nih.gov/research/umls/Snomed/snomed_main.html
\(^11\)http://www.loc.gov/marc/bibliographic/ecbdhome.html
\(^12\)http://dublincore.org/workshops/dc1/report.shtml
specifications of the Dublin Core are evolving. The tags in Dublin Core are too
generic for scientific documents; for example, a tag for journal name does not exist,
or the tag creator is too generic. Dublin Core is formalised in RDF.

FOAF (Friend of a Friend\(^\text{13}\)) has been used to describe authors
and other characteristics, such as the publisher, in various studies
De Roure et al. 2008, Ciccarese et al. 2011, Osborne & Motta 2012].

FRBR (functional requirements for bibliographic records) is a concep-
tual model that is based on entities that are divided into three groups
[IFLA Study Group on the Functional Requirements for Bibliographic Records 1998].
The first group includes four concepts: work, expression, manifestation and item
model. The model describes intellectual or artistic products. A work is a distinct
intellectual or artistic creation, expression is the realisation of this work, manifesta-
tion is the physical embodiment of an expression of a work, and an item is a single
example of a manifestation. The second group includes the entities responsible
for the intellectual or artistic content. The third group includes an additional set
of entities that describe intellectual or artistic creations, such as concept, object,
event and place.

BIBO [D’Arcus & Giasson 2009] is an ontology formalised in RDF that can be
mixed with other vocabularies, such as FOAF\(^\text{14}\) or Dublin Core. This ontology
allows for the description of the type of documents destined for publication on
the Semantic Web. This ontology is used in various projects [Osborne et al. 2013,
Aquin et al. 2013].

FaBiO (FRBR-aligned Bibliographic Ontology) [Peroni 2012] incorporates var-
ious ontologies (DC terms, PRISM, FRBR, SKOS) to overcome the limitations of
each of those ontologies. FaBiO allows for the semantic description of a variety of
bibliographic objects, such as research articles, journal articles, and journal volumes,
to clearly separate each part of the publishing process, the people involved in the
publication process, and the various versions of documents (electronic or physical).
FaBiO was implemented in OWL 2 DL.

In [Hernandez 2005], the author uses an ontology to improve strategic search
monitoring in astronomy using concepts such as laboratory, affiliation, and the au-
thors’ country of residence.

In [Jörg et al. 2012], the authors propose an ontology based on CERIF (the
Common European Research Information Format) to describe bibliographic data,
research activity, methods and datasets. This ontology describes research entities
(e.g., person, project, organisation), link entities (relationships) and multilingual
entities and features a small research methods ontology. The latest release of CERIF
was in 2008.

In [Yaginuma et al. 2004], the authors propose a metadata layer based on Dublin
Core (DC) in which they add 27 elements not described in DC, such as the article

\(^{13}\)http://www.foaf-project.org/
\(^{14}\)http://xmlns.com/foaf/spec/
type, keyword list, copyright, supplier, edition, and target audience for an article. The study was part of the European OmniPaper project, which proposed to support conceptual search mechanisms.

In the Inspire Project [Ivanov & Raae 2010], metadata associated with an author and affiliation are used to summarise the author page.

3.2.3 Structural dimension (3)

The structural parts of a document include the title, sections, divisions, and headers. At a finer level of granularity, the parts can be sentences and paragraphs. Several models exist for describing logical structure of scientific content; the most commonly used is the IMRaD structural model.

IMRaD [Sollaci & Pereira 2004] structures the headings of scientific documents, which include the introduction, methodology, results and discussion. The introduction section defines the goal, aim, and hypothesis of the study and the research questions addressed. The methods describes when, where, and how the study was performed, what materials were used, and who participated. The results section is where the results are presented and indicates whether the hypothesis was supported. Finally, the discussion section describes the implications of the results, indicates whether the results are coherent with other research, and suggests future work. This model is used for the whole article and/or the abstract (structured abstract).

SQUIRE (Standards for Quality Improvement Reporting Excellence) [http://squire-statement.org] provides guidelines that are used in some journals, particularly medical journals. The guidelines provide standards for headings, such as the title, abstract, introduction (background, local problem, intended improvement, study question), methods (ethical issues, settings, planning the intervention, planning the study of the intervention, methods of evaluation, analysis), results (outcomes), and discussion (summary, relation to other evidence, limitations, interpretation, conclusions).

DOCO (Document components ontology) [http://vocab.ox.ac.uk/doco, http://purl.org/spar/doco] allows for the description of the structural parts (paragraph, section, chapter) and the logical and rhetorical parts (introduction, discussion, background, reference) of a document. This ontology imports other ontologies, including the discourse elements ontology, the document structural patterns ontology, and some parts of the SALT ontology [We will discuss this ontology in the next section.]

In [Angrosh et al. 2011], the authors propose modelling the contexts associated with a citation sentence [We will also discuss this work in the section on the relationships between fragments or document] using a more general model (the Sentence Context Ontology). The content of a document is divided into three categories: Introduction Block (introduction, related work, overview, motivation), Conclusion Block (conclusion and future work), and Body Block (other sections). The authors use a

[18]We will discuss this ontology in the next section.
[19]We will also discuss this work in the section on the relationships between fragments or document.
probabilistic model to achieve the automatic classification of sentences. Metadata indexation is important for describing the scientific document, but it is not enough when the query that users want to create become to be complex or when the users don’t know the metadata of the documents they are seeking.

3.2.4 Rhetorical or discursive dimension (4)

Some authors propose the use of rhetorical structure or argumentative or discursive categories to manually or automatically annotate scientific writing to improve summarisation or information retrieval systems. These studies focus mainly on the "hard" sciences, such as biology, medicine, physics, and bio-computation, in which there is relatively little variation in describing results, methodologies, hypotheses, and conclusions.

3.2.4.1 Model used in manual annotation

In [Harmsze & Kircz 1998, Harmsze et al. 1999, Harmsze 2000], the author proposes a very complete and complex model for decomposing scientific documents into modular structures. The aim of the author’s thesis [Harmsze 2000] is that authors of scientific documents publish their research using a modular structure when the elements are independent and are linked by meaningful relations. She defines a module as follows: "A module is a uniquely characterised, self-contained representation of a conceptual information unit, which is aimed at communicating that information. This definition can be applied to any coherent body of information."[Harmsze 2000, p.39]. She also defines a modular structure as follows: "A modular structure is a pattern of modules and explicit links between modules"[Harmsze 2000, p.41] and links (hyperlink) as follows: "A link is a uniquely characterised explicit, direct connection, between entire modules or particular segments of modules, that represent one or more different kinds of relevant relations"[Harmsze 2000, p.42]. The author proposes six modules (see figure 3.1): meta-information (bibliographic information, list of domain, maps of contents, abstract, list of references, acknowledgements), positioning (situation, central problem), methods (experimental methods, numerical methods, theoretical methods), results (raw data and treated results), interpretation (qualitative interpretation and quantitative interpretation) and outcome (findings and leads for further research). She characterises these modules by their range: microscopic, mesoscopic and macroscopic. Microscopic modules represent information that belongs only to one particular article. Mesoscopic modules define information at the level of a research project, and macroscopic modules represent information that is more general than one article or one research project. To link those modules, the model proposes two types of relations. The first is organisational links (hierarchical, proximity, range-based, administrative sequential (complete-path and essay-path), representational), which address the organisational structure of a modular article. The second type of link addresses the content of modules. Scientific discourse relations are divided into two categories: relations...
3.2. State of the art

based on the communicative function (elucidation (clarification (definition and specification) and explanation) and content relations (dependency in the problem-solving process (transfer), elaboration (resolution, context), similarity, synthesis (generalisation, aggregation) and causality). [Harmsze 2000] highlights some differences between expressing the internal and external links of documents. The author manually annotates scientific documents in physics, one of which is accessible through the demo she developed\(^\text{20}\). Because this model is rather complicated, it has never been completely adopted by any type of system.

Figure 3.1: Overview of the Harmsze model [Harmsze 2000, p.174]

\(^{20}\)http://www.science.uva.nl/projects/commphys/papers/thesisfh/demo.html#intro
The ScholOnto [Shum et al. 1999] project proposes the decomposition of scientific publication into elementary discourse knowledge items and meaningful relationships. The authors developed an ontology to represent scientific documents. The main concepts are analysis, approach, data, idea, language, methodology, problem, software and theory/model. The proposed relationships are addresses, analyses, modifies/extends, predicts, uses/applies, supports, raises issues with, and refutes. The authors of the project proposed a preliminary prototype user interface for submitting an article’s description to the repository and a formed-based interface to automatically generate queries based on the ScholOnto ontology and demonstrated the potential of their approach by manually annotating one document using the different concepts and relationships between the concepts. The second interface proposed in the SchoolOnto project was the ClaiMaker [Shum et al. 2007], which is a web application that provides a user interface for building a semantic hypertext network based on the relationships proposed in the ScholOnto ontology between two nodes. The authors re-defined the class structure and the relationships of ScholOnto using a structure that is more precise. Each relation is redefined with finer granularity and features polarity and weight as properties (see Figure 3.2). In this project, various tools were developed to annotate documents: a Word plug-in and a sketch interface that allows users to sketch their claims and the relations between the claims based on the Compendium visual hypertext system\(^{21}\). To navigate and search through the annotated claims, the authors created a user search interface (ClaimFinder). The interface generates interactive visualisations of relationships and claims. ClaiMaker and ClaiFinder provide tools for finding the lineage of an idea, or, as the authors call it, "the 'spaghetti' of claims", which represents the flow of ideas that can be picked from different documents to build some knowledge on a subject. The idea behind finding the "spaghetti" of claims, or the "fil d’Ariane", is to understand the birth of an idea and follow the path of this idea through different studies. The COHERE project [De Liddo & Shum 2010, Shum 2008] is an extension of ClaiMaker that exploits the advantages of Web 2.0, such as networking and interactive web user interfaces. Instead of using an independent system, COHERE is integrated into Firefox as an extension, where the aim is to connect ideas using link types proposed in ScholOnto and COHERE.

\(^{21}\)http://compendium.open.ac.uk/institute/
3.2. State of the art

In [Waard et al. 2009b], the authors propose an alternative view of research articles in biology as a network connecting "hypotheses and evidence". Their model is based on a list of claims or hypotheses, the related evidence, and the relationships that connect them. This network of claims/hypotheses, connected to experimental evidence by argumentative relationships, is called the Hypotheses, Evidence and Relationships (HypER) model. To the best of our knowledge, this model has never been implemented.

In [Waard & Tel 2006], the authors propose an open-standard format, the ABCDE format, to explicitly create a semantic structure for articles during writing. They provide commands in \LaTeX such as \texttt{\{background\}} to declare that the material following this command belongs to a certain category. The authors defined five categories:

- **Annotations**: metadata (in Dublin Core Standard), minimal fields are title, creator, identifier and date
- **Background**: describes the positioning of the research, ongoing issues and central research question
- **Contribution**: describes the work the authors have done
- **Discussion**: contains the discussion of the work, comparison with other work, implications and next steps
- **Entities**: references, personal names, project websites
The authors’ aim was to provide this stylesheet to conferences, journals, and websites to interconnect pages according to the aforementioned categories. It appears that the stylesheet was never adopted. In [Waard et al. 2009a], the authors propose a new model for extracting seven epistemic segment types:

**Fact:** statement that should have been accepted as true by the community

**Hypothesis:** possible explanation for a set of phenomena

**Implication:** the interpretation of results

**Method:** ways in which the experiment was performed

**Problem:** unknown aspects

**Goal:** the goal or aim of a study

**Result:** the results of measurements

In an experiment, the segments were manually categorised into one of those categories using cue sentences and the understanding of the context of a segment. [Waard et al. 2009a] used the defined set of markers to automatically identify the type of segments in 1,721 biomedical abstracts. The authors evaluated their system on 100 randomly selected sentences. The segments were evaluated by the first author of this article [Waard et al. 2009a], who found that only 30 of 100 were not correctly identified. [Waard et al. 2009a] demonstrated that verb tense is a good indicator for classification and that sentences containing methods, results and implications sections are easily detected. In [Waard & Maat 2012], the authors explored the implications of verb tense for the classification between seven discourse segment categories: facts, results, hypothesis, problem, implications, goals and methods and changing the verb tenses of facts, hypothesis, results and methods. Generally, facts are described in the present tense, results in the past tense, hypotheses in modal auxiliaries, problems with negations and the present tense, implications with present and matrix clauses, goals with the infinitive verb tense and methods in past tense. The authors tested the effect of swapping verb tense on 12 people and observed that this swap altered the interpretation of the sentence (e.g., a factual statement in the past tense is perceived as a result, and a hypothesis in the past tense is perceived as a result). Verb tense functions as a marker for discriminating between types of scientific discourse, but it is not enough to identify the nature of scientific discourse. The corpus used to perform this study was adopted from biology, in which sentences are more formalised than in other domains. In addition, as the written scientific language become to be more and more English, not every authors are native English speaking, and the grammatical rules of writing are often not well respected.

In [Waard & Kircz 2008], the authors propose an improvement of the ABCDE format based on the Harmzse model. The aim is to benefit from the modularity of the Harmzse model and the linearity of the ABCDE format to propose a framework as a \LaTeX stylesheet. The authors argue that scientific documents are an attempt to convey meaning and convince peers of the validity of a specific claim using research
data. To optimally represent a scientific paper, the model should emphasise the relations that link the claims to the data. They propose a new model that features three types of modules:

- **Content module**
  - Front matter/metadata
    - * Introduction
    - * Positioning
    - * Central Problem
    - * Hypothesis
    - * Summary of results
  - Experiments, containing the following discourse unit types:
    - * Facts, Goals, Problem, Methods, Results, Implications, Experimental Hypothesis
  - Discussion, containing the following subsections:
    - * Evaluation
    - * Comparison
    - * Implications
    - * Next steps

- **Entities** (contained within the (sub)sections described above), consisting of the following:
  - Domain-specific entities, such as genes, proteins, and anatomical locations.
  - Figures and tables
  - Bibliographic references

- **Relations**, consisting of the following:
  - Relations within the document from entity representations to entities (figures, table and bibliographic references)
  - Relations out of the document, from entities to external representation (e.g., from a protein name to its unique identifier in a protein databank)
  - Relations between moves within documents
  - Relations between moves between documents

The authors manually annotate 13 documents in cell biology using Cohere, through which documents are decomposed into a single sentence or fragment of text.

The authors of the SALT (Semantically Annotated \LaTeX) project [Groza et al. 2007b, Groza et al. 2007a, Groza et al. 2007c] propose a \LaTeX annotation framework for creating Semantic PDF documents. The aim is to allow authors
to directly structure a document during the writing process. The framework features three ontologies: a document ontology (sections, paragraphs, text chunks), a rhetorical ontology (the argumentation and rhetorical structure of publications), and an annotation ontology (connects the rhetorical structure within a publication to its textual representation and the metadata of the document). The rhetorical ontology consists of two axes: rhetorical relations (antithesis, circumstance, concession, condition evaluation, evidence, means, preparation, purpose, restatement, solution) and rhetorical blocks (Abstract, Background, Motivation, Scenario, Contribution, Discussion, Conclusion and Entities). The authors extend the ABCDE model of [Waard & Tel 2006]. In [Groza et al. 2007a], the authors propose a search engine for querying publications in which the claims are generated as metadata during the generation process of LaTeX documents.

The SWAN project (Semantic Web application in Neuromedicine) [Passant et al. 2009, Ciccarese et al. 2008] focuses on developing a semantic framework for representing biomedical discourse based on an ontology. This project was initially applied to research on Alzheimer’s disease. The main elements of this ontology are

People, groups and organisations: using the FOAF vocabulary.

Discourse elements: Research statements (a claim or a hypothesis), research questions (topics under investigation), structured comments: the structured representation of a comment publisher in a digital resource). Discourse elements can have relationships between them: alternativeTo, consistentWith, inconsistentWith and discusses. Research statements can also cite as evidence other digital resources (citesAsEvidence) or life science entities (citesLifeScienceEntity).

Bibliographic records and citations (metadata properties)

Life science entities (genes, proteins, organisms and relationships between them)

Tags, qualifiers, and vocabularies

Versions and provenance

Two software applications use SWAN. The Workbench allows scientists and scientific editors to instantiate an ontology. This application was implemented using AHAY and stored in the ANZO triple-store. The browser allows scientists to manipulate and download information from the SWAN knowledge base. A faceted browser is included to navigate through sets of research statements and their evidence. The SWAN project is part of the HCLSIG/SWANSIOC\textsuperscript{22} (Scientific Discourse Task Force) in the W3C, which uses ontologies such as the CiTO, PRISM, and DOCO. This group is co-chaired by Tim Clark and Anita de Waard.

In [Marcondes 2011], the authors propose a semantic model for annotating biomedical articles that extends bibliographic notices and discourse elements, such

\textsuperscript{22}http://www.w3.org/wiki/HCLSIG/SWANSIOC
as the **problem**, **hypothesis**, **experiment** (indicative, exploratory, deductive), **conclusions**, and the **relationships between phenomena**. The annotation is manual and performed by the authors.

### 3.2.4.2 Models used in automatic annotation

In [Teufel 1998, Teufel & Moens 1999a, Teufel 1999, Teufel & Moens 1999b, Teufel & Moens 2002, Teufel 2010], the authors propose Argumentative Zoning, the aim of which is to annotate rhetorical moves in research articles to summarise research articles. The authors use the CARS (Creating a Research Space) model of Swales (1990) (as cited in [Teufel 1999]) as a starting point. They define the Rhetorical Document Profile (RDP) that describes the typical information needs of new readers. The RDP is defined by seven elements: **Background** (generally accepted background knowledge), **Other** (specific other work), **Own** (own work: method, results, future work), **Aim** (specific research goal), **Textual** (textual section structure), **Contrast** (contrast, comparison, weakness of other solution) and **Basis** (other work that provides a basis for own work). A basic schema is also defined featuring Background, Other and Own.

To help the human annotator annotate a document, the authors created a decision tree for the full annotation scheme. In her PhD thesis [Teufel 1999], Teufel automatically annotates documents on computer linguistics and medicine using statistical models based on a Naive Bayesian model and an n-gram model operating on sentences, taking the hand-annotated corpus as training material. Precision is between 26% for the Basis and 81% for Own, and recall is between 21% for Contrast and 91% for Own. Teufel obtains better results with the statistical model using Naive Bayes and Bigram than when using only Naive Bayes. In [Copestake et al. 2006], the authors adapt the model for a chemistry corpus. They argue that if the argumentation patterns are generally similar across the disciplines, some factors change, such as the use of citations, passive voice or cue phrases.

In [Sándor et al. 2006, Sándor & Vorndran 2009, Sándor & Waard 2012], the authors use the Xerox Incremental Parser (XIP), a rule-based dependency parser, to extract discourse sentences from text. In [Sándor et al. 2006], the authors propose two systems. The first system detects the discourse function in scientific argumentation, such as **background knowledge**, **logical contradiction**, **element insufficiently or not known**, **research trend**, **summary sentence**, **contrast with past findings** and **substantially new findings**. The second system detects sentences that contain bibliographical references and the characterisation of the relationships between the authors and the referred documents (**background knowledge**, **based on**, **comparison and assessment**). The authors test their system on biomedical documents.

In [Sándor & Vorndran 2009], the authors extract **research problems**, **purposes** and **conclusions** from a corpus of educational science documents. These sentences receive two types of labels: a Summary or Problem or both tag. The features of key sentences are determined by the argumentative expressions of the
sentences. In [Sándor & Waard 2012], the authors propose to define and extract **claimed knowledge** updates (CKUs) in biomedical research articles. They define CKUs as fragments that convey the most important new factual information. By listing them, they can easily highlight the new main knowledge provided by a document. These CKUs can be automatically detected in sentences featuring phrases such as "our results demonstrate". The analysis is conducted by XIP. The evaluation of two articles shows CKU coverages of 81% and 80% and precision levels of 62% and 51%. In [De Liddo et al. 2012], the authors use XIP to automatically annotate issue for, summary for, and research questions for and recover the automatic annotation in Cohere, in which a user can create links between segments or documents and validate the automatic annotation of XIP. The evaluation is performed using 41 annotation templates; the manual annotation results in 161 sentences in the three fields (Objectives, Issues and Questions), and XIP highlights 797 sentences.

In [Ibekwe-SanJuan 2006, Ibekwe-SanJuan et al. 2011], the authors annotate abstracts with meta-semantic tags. In [Ibekwe-SanJuan 2006], the author sets out to identify patterns regarding novelty or changes in the titles and abstracts of documents in biology. The author is interested in original parts written by the authors. **Novelty** is divided into three categories: object of, contribution/results and conclusion. The author examines 50 documents to find patterns. She finds, as does [Teufel & Moens 2002], that one sentence can convey more than one pattern. The patterns observed are fragments of sentences featuring a specific word or specific words. She tags the corpus using a rule-pattern system in XML with elements such as New, Aim, Results, and Conclusion. Aim, Results and Conclusion can include the element NEW. In [Ibekwe-SanJuan et al. 2011], the authors extend the first model [Ibekwe-SanJuan 2006] using the same methods to tag seven categories: Objective, Results, New, Hypothesis, Related Work, Conclusion and Future Work. They use WordNet to extend their patterns using synonyms. The authors evaluate the precision of their XML annotation and obtain high precision in several categories on a computer science corpus. They argue that they are able to obtain 100% precision for the objectives and the results due to the very technical nature of their corpus.

In [Ruch et al. 2007, Gobeill et al. 2008], the authors propose classifying sentences from structured abstracts by Latent Argumentative Structuring into four classes: Purpose, Methods, Results and Conclusion. The authors use structured abstracts in MEDLINE to train their Bayesian classifier. It appears that the classification and the evaluation are based on the hypothesis that all sentences that are written under the headings belong to the categories of the headings (i.e., all the sentences that are written under the heading "results" are results sentences).

In [Ou & Kho 2008], the authors propose a method for aggregating the search results of research abstracts returned by an information retrieval system. The system then organises and presents the results to users. The authors classify sentences extracted from structured abstracts into two categories, research-objective and research-results, and extract the concept terms and the contextual relations from these sentences. They use a decision tree classifier for sentence categorisation and
extract relationships between concepts with linguistic patterns (regular expression). The authors evaluate their system with 40 users who indicate a preference for using the concept-based system. They argue that the structured abstracts provide a good summary of the research findings.

In [Tutin et al. 2009], the authors use a corpus of heterogeneous scientific documents written in French and in English that were previously manually annotated with TEI\(^{23}\) (Text Encoding Initiative) to define the **positions of authors** (the authors of the study define position as the manner in which an author subscribes to a community, how he/she evaluates and is evaluated by his/her peers and what propositions he/she puts forward\(^{24}\)[Tutin et al. 2009, p.338]. The authors use cue verbs that indicate opinions or points of view, choice and intention or verbs that indicate a specific value that the author holds. They find that some verbs are more frequently used in some domains than others. For example, linguistics uses more verbs that indicate an opinion or point of view than psychology and education science do. The authors develop a search interface that allows users to request scientific documents using certain elements of their TEI schema of annotation. They also extract the **hypothesis** of documents using cue sentences that contain the word "hypothesis" and a verb (e.g. "Faire Hypothèse", "Tester Hypothèse", "Confirmer Hypothèse").

In [Groza et al. 2010], the authors automatically extract **epistemic items** (antithesis, cause, circumstance, concession, condition, elaboration, evidence, means, preparation, purpose, restatement, and power (an item that is not part of one of the other items) from scientific documents. The authors analyse 130 computer science publications to extract 75 cue phrases that signal rhetorical relations. The authors identify the positions of the cue phrases in sentences and the positions of the cue phrases according to the neighbouring text spans and the surrounding punctuation. Rules based on the cue phrases are constructed in JAPE in GATE. The methodology is evaluated by consulting the authors of the extracted publications. Groza et al. argue that the most reliable annotator for a scientific publication is its author. The authors evaluate inter-annotator agreement and find that annotators agree more on antithesis than on other items. To test precision and recall, the authors set up an evaluation with two corpora, the first written by the annotator (15 annotators for 15 articles) and the second composed of 5 papers annotated by the same 15 persons. Annotators were asked to freely annotate the text spans that they considered to play the role of epistemic items. Precision was 0.5 and recall was 0.3 for the documents annotated by their own authors, and for the second corpus, precision was 0.43 and recall was 0.32.

In [Liakata 2010], the authors propose the use of Core Scientific Concepts (CoreSCs), which describe, at the sentence level, the core components that constitute a scientific investigation. CoreSCs is an ontology derived from CISP metadata [Soldatova & Liakata 2007]. CISP is composed of eight concepts: **Motiva-**
tion, Goal, Objective, Methods, Experiments, Observations, Results and Conclusion. These elements were validated using an on-line survey completed by 33 researchers. CoreSCs also implements three other concepts: Hypothesis, Model and Background. These 11 concepts can have properties such as new or old and advantage or disadvantage. Using SAPIENT, a semantic annotation tool [Liakata et al. 2009], Liakata 2010 annotate 225 papers from physical chemistry and biochemistry. In [Liakata et al. 2012], the authors use 265 documents manually annotated by chemistry experts to train two types of machine-learning classifiers (support vector machine and continual random field). The most discriminating features are unigrams, bigrams and grammatical dependencies, such as section headings. The categories in which the authors obtain the best precision/recall are the background, experiments and model, followed by the conclusion, observations, results, objective, goal and hypothesis. The categories that obtain good inter-annotator agreement are the conclusion, methodology and objective. In [Liakata et al. 2010], the authors compare the AZ (argumentative zoning, see above) model of Teufel [Teufel 1999] with the CoreSCs schema. The authors find that the two annotation schemes are complementary and suggest that it is beneficial to use both schemas. CoreSCs is more suitable for the identification of potential zones of interest, whereas AZ is more suitable for monitoring the relations between current work and cited work.

In [Blake 2010], the authors propose to automatically extract the claims of biomedical texts. The authors classify claims into five categories, explicit claims, implicit claims, correlations, comparisons, and observations, and define a framework for their identification. The first study involved the manual annotation of 29 articles in the Genomics Text Retrieval (TREC) collection by a student using the Claim Framework. The most frequently annotated type of claim was explicit claims, which constituted 77.11% of all annotated claims. The authors also examine the distribution of claim types among major article sections (IRMaD model) and observe that the majority of claims are made in the discussion section, followed by the introduction. Only 7.84% of claims are described in the abstract. To realise automatic annotation, the authors use the annotations made for explicit claims. The authors’ approach combines the semantics and lexico-syntactic provided by the Stanford dependency grammar parser. The authors find that by emphasising the role of verbs rather than only concepts in a sentence, the automatic annotation model exhibits high precision/recall by using semantic and syntactic features, and it should be possible to generalise this approach to domains other than biomedical text. In [Park & Blake 2012], the goal of the authors is to automatically identify comparative claims using a set of semantic and syntactic features that characterise a sentence. The authors argue that comparison is an important discourse element because authors can compare research hypotheses, data collection, methods, subject groups and findings in their articles. The authors implement 35 features that capture both the semantic and syntactic characteristics of a sentence. They use these features with three different classifiers, Naïve Bayes, Support Vector machines and Bayesian Networks, and find that the highest accuracy and F1 scores are
obtained using Bayesian Networks. Teufel et al. [Teufel 1999, Ruch et al. 2007] also use Bayesian Networks for their classifiers.

### 3.2.5 Relationships dimension (5)

The fifth dimension describes relationships with external documents (inter-document, citations, references) or with other parts of the same document (intra-document). In 1965, Garfield [Garfield 1965] identified fifteen different reasons for citation:

- Paying homage to pioneers
- Giving credit for related work (homage to peers)
- Identifying, for example, methodology equipment
- Providing background reading
- Correcting one’s own work
- Correcting the work of others
- Criticizing previous work
- Substantiating claims
- Alerting to forthcoming work
- Providing leads to poorly disseminated, poorly indexed, or uncited work
- Authenticating data and classes of fact, physical constants, etc.
- Identifying original publications in which an idea or concept was discussed
- Identifying original publication or other work describing an eponymic concept or term, such as Hodgkin’s disease, Pareto’s law, or Friedel-Crafts reactions
- Disclaiming the work or ideas of others (negative claims)
- Disputing priority claims of others (negative homage)

At this early stage of the automation of citation indexing, Garfield had already addressed the problem of identifying all pertinent references. Moreover, he argued that it would be interesting "to produce a useful 'marker' which would describe briefly the kind of relationship that exists between the citing and cited documents" [Garfield 1965, p.191].

In [Bertin et al. 2006], the authors use EXCOM (EXploration COntexuel Multilingue) [Djioua et al. 2006] to provide information about the nature of citations and allow for a qualitative approach to bibliometric methods rather than a quantitative approach (e.g., impact factor is a quantitative approach for bibliometrics). They use linguistic cues (words or fragments of sentence) to determine five categories of citations in a French corpus: point of view (position), comparison
(resemblance, disparity), information (hypothesis, analysis, result, method, citation, counterexample), definition, and appreciation (agree, disagree). The citations are extracted with regular expressions, and the categories of the citation are extracted with syntactic rules implemented in EXCOM, but the authors do not evaluate their annotations. In [Bertin et al. 2009b], the authors propose to extract referenced sentences based on definition type. They use the same methodology used in [Bertin et al. 2006] to extract referenced sentences and definitions, but they only take into account the definition near a referenced sentence to reduce the risk of noise. The precision/recall test for the extraction of the citations is high (500 sentences, recall of 91% and a precision of 98%). The authors extract 50 definition sentences for their annotations and ask two judges to determine whether each is a definition. The authors use a Kappa test to determine the inter-judgement measure. They obtain $\kappa = 0.65$, which, according to the scale of the Kappa measure, is substantial agreement. In [Bertin et al. 2009a], the authors elaborate a method that automatically annotates scientific articles to identify and categorise authors’ judgement in texts. The authors of the study argue that sentences indicating judgement in the referenced studies will be close to the indexed references. The authors use regular expressions to find the references and use EXCOM to detect the categories they have defined (definition, point of view, appreciation, comparison and information). Two hundred scientific papers and six PhD theses in French were evaluated. To evaluate the precision of the authors’ automatically assigned categories, two judges evaluated 100 sentences, and excellent agreement between judges was obtained ($\kappa = 0.83$ (Kappa coefficient)), but the judges were unable to identify the authors’ judgements in the referenced documents.

In [Teufel et al. 2006a, Teufel et al. 2006b], the authors propose an annotation scheme for citation function. They define 12 categories marking relationships with other works, with four top-level categories:

**Explicit** statement of weakness

**Contrast or comparison** with other work (four categories)

**Agreement/usage/compatibility** with other work (six categories)

**A neutral** category

Each citation has only one category. The authors use a machine-learning system to classify 2,892 sentences from 116 articles (computational linguistics) into the defined categories. The inter-judge agreement is $\kappa = .57$ for the full scheme and $\kappa = .58$ for the three-way classification (Weak, Positive, Neutral).

In [Sendhilkumar et al. 2013], the authors propose to classify citations using SentiWordNet to identify the sentiment of citations as positive, negative or neutral. They also use the classification scheme of Teufel et al. [Teufel et al. 2006b] to classify the extracted citation into eight categories. The authors evaluate their classifier for sentiment analysis using 80 sentences from seven articles and find low recall but high precision.
In [Angrosh et al. 2010, Angrosh et al. 2011], the authors argue that they can classify sentences into two categories, citation sentences and non-citation sentences. They define citation sentences as follows: "Citation sentences are defined as those sentences that have a reference to a published or unpublished source" [Angrosh et al. 2011, p.5]. Non-citations are all other sentences. The authors assume that citation sentences are sufficient to provide a good representation of an article. Each sentence can be associated with a context: issue sentence, shortcoming, description, methodology, current work, (future work, current work shortcoming, current work outcome), compares cited works, extends cited work, overcomes gaps, identifies gaps, related to issues, shortcomings in cited works, gaps in cited works, related to subject issuers, uses outputs in cited work and relates results with cited work. The authors propose an ontology, SENTCON (Sentence Context Ontology), in OWL that describes the context of sentences, with a specific focus on citation sentences. This ontology uses BIBO. The authors train their system with 20 articles extracted from the LNCS collection of Springer and classify 1,162 sentences into 17 categories. The precision and recall test results are good, with an accuracy of 93.37%.

In [Shotton 2009a, Shotton 2010, Peroni & Shotton 2012], the authors propose CiTO (Citation Typing Ontology) (see Table 3.1) to describe the nature of reference citations in scientific documents. CiTO is written in OWL. The authors describe 23 relationships between citing and cited documents. The ontology also describes the citation frequency using 10 classes. CiTO uses FRBR, which is restricted to works that cite or are cited and allows for the description of the type of work (e.g., clinical case report, clinical trial report, algorithm, and catalogue) of the citing or cited study. CiTO also imports BIBO to describe bibliographic citations. This ontology is in its second version and operates with Creative Commons licences.

Table 3.1: The 23 relationships between citing and cited documents in CiTO [Shotton 2009a]

<table>
<thead>
<tr>
<th>cites</th>
<th>citesAsAuthority</th>
<th>citesAsMetadataDocument</th>
</tr>
</thead>
<tbody>
<tr>
<td>citesAsSourceDocument</td>
<td>citesForInformation</td>
<td>isCitedBy</td>
</tr>
<tr>
<td>obtainsBackgroundFrom</td>
<td>sharesAuthorsWith</td>
<td>usesDataFrom</td>
</tr>
<tr>
<td>usesMethodIn</td>
<td>confirms</td>
<td>credits</td>
</tr>
<tr>
<td>extends</td>
<td>obtainsSupportFrom</td>
<td>supports</td>
</tr>
<tr>
<td>updates</td>
<td>corrects</td>
<td>discusses</td>
</tr>
<tr>
<td>critiques</td>
<td>reviews</td>
<td>disagreesWith</td>
</tr>
<tr>
<td>qualifies</td>
<td>refutes</td>
<td></td>
</tr>
</tbody>
</table>

The foregoing review of the literature demonstrates the existence of a large number of models for scientific document annotation. Our review of existing annotation models of scientific documents indicates that research is very active. The number of

25 http://purl.org/spar/cito
studies on the annotation of scientific documents shows that this area of research is important, and the annotation of scientific documents can be useful for applications ranging from information retrieval to the humanities. The models used for manual annotation are generally more complex than the ones used for automatic annotation.

Why propose a new model?

• To the best of our knowledge, none of the above models is based on empirical studies that aim to understand what type of information users want to retrieve or to use in selection criteria.

• As demonstrated in the previous chapter, scientists might be interested in very specific parts of documents, particularly when they do not want to read documents in their entirety, and these parts or elements are rarely present in the models presented above (e.g., definition). To compare these studies with the user model we propose, we summarise in the figure 3.3 different models that annotate the same discourse elements that we observed in the user study. We group the different models following the five discourse elements that scientists seek (findings, hypothesis, methods, definition and related work). Findings is the element that includes the largest number of annotation models. For instance, the only models that take the element "definition" into consideration are [Harmsze & Kircz 1998, Bertin et al. 2006, Bertin et al. 2009b]26. The study that aggregates the most discourse elements is the model of [Waard et al. 2009a, Waard & Kircz 2008], which identifies hypothesis, methods, results and related work, but didn’t identify neither definition or relationships between documents.

• Most of the models focus on the "hard" sciences, such as biology, in which there is relatively little variation in the way the results, hypothesis, and conclusions are described (especially for automatic annotation). The sentences are generally short, even formalised, when describing the results and conclusion. It should be possible to build a model that describes the different areas of scientific knowledge.

26In our review of the literature, we did not consider all studies that have been conducted based on the element "definition". We restricted our search to larger models and examined whether they included "definition" as part of their methodology.
3.2. State of the art

Figure 3.3: Scientific annotation models summarised

- **Definition**
  - [Bertin et al. 2006]
- **Hypothesis**
  - (Tutin et al. 2009)
  - [Bekwe-SanJuan 2006, Bekwe-SanJuan et al. 2011]
  - [Marcondes 2011]
  - [Swan [Passant et al. 2009]]
- **Background/Related Work**
- **Discussion/Conclusion/Results/Data**
  - SALT [Groza et al. 2007b, Groza et al. 2007c]
  - [Ou & Kho 2008]
  - XIP [Sándor et al. 2006, Sándor & Vorndran 2009, Sándor & De Waard 2012]
  - [Ruch et al. 2007, Robell et al. 2008]
  - [Liakata 2010, Solórová & Liakata 2007]
  - [Buckingham Shum et al. 2007]
3.3 Our model: SciAnnotDoc

In the scientific literature, scientists report all types of discoveries, empirical results, assumptions, and methodologies they have used, concepts they have defined or redefined, claims, aims of their research, and studies that have inspired them and to which they compare their work. As demonstrated in the previous chapter, scientists would like to use the same entries when seeking information or to strategically read documents to avoid wasting time.

Scientific papers are more or less structured depending on the field. Scientific documents have changed since their inception in 1655 [Kronick 1976]. In those early years, the writing style was polite and addressed several subjects at the same time. Scientific articles described experimental reports and described events in chronological order.

In the early twentieth century, scientific documents began to be more standardised, and the IMRaD (introduction, methods, results and discussion) structure was adopted in various domains, especially medicine [Sollaci & Pereira 2004]. Although the IMRaD structure is the most frequently used, it is not the only model indicating how a scientific report should be structured (e.g., the IBC model: introduction, body, conclusion). Moreover, the IMRaD structure is not sufficient for annotating a document using only the headings for the following reasons: 1) this structure is not adopted in all scientific documents and by all domains of research; 2) authors may adopt IMRaD in the construction of their narrative, but the headings will not necessarily reflect that structure (e.g., in computer science, the part corresponding to implementation may correspond to the methodology and to the results); 3) the structure describes rhetorical parts or discourse elements but neglects others (e.g., background, aim, definition); 4) it is possible to find a sentence that describes other discourse elements than the one declared by a particular heading (e.g., a sentence discussing methodology below the heading "Results"). In [Ibekwe-SanJuan 2006], the author argues that the headings of sections are too generic and are not strictly delimited. The same heading can convey information of different natures. The author argues that precise information retrieval requires the use of lower-level indices, such as paragraphs or sentences.

IMRaD is still the most widely used structure, but its adoption as a basis for an IRS may lead to the exclusion of a number of scientific documents. To provide an example, we examine two journals in the domain of gender studies and sociology.

Out of 24 scientific documents extracted from the journal Feminist Studies in 2007, only 9 contain titles for paragraphs or sections, and the titles do not always correspond to the IMRaD model (see table 3.2). Out of 18 articles extracted from the American Journal of Sociology in 2013 (a journal with more empirical studies than Feminist Studies), 10 articles have a complete IMRaD structure, but 8 articles present only a partial or no IMRaD structure. These articles also contain other names for headings, such as Theory and Theoretical Framework, as well as more discursive titles (see table 3.3). An information retrieval system based only on the IMRaD structure would not have identified the sections pertaining to the results or
methodology in these articles.

Table 3.2: Example of the logical structure of an article in Feminist Studies

<table>
<thead>
<tr>
<th>Title</th>
<th>Permission to Rebel: Arab Bedouin Women’s Changing Negotiation of Social Roles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Section</td>
<td>The Dual Marginality Of Bedouin Women</td>
</tr>
<tr>
<td>Subsection</td>
<td>Gender Marginality in Bedouin Society</td>
</tr>
<tr>
<td>Subsection</td>
<td>The Intersection Of Ethnic And Gender Marginality</td>
</tr>
<tr>
<td>Section</td>
<td>Women’s Struggle In The Middle East And North Africa</td>
</tr>
<tr>
<td>Section</td>
<td>The Study</td>
</tr>
<tr>
<td>Section</td>
<td>Coping Strategies: Inside And Against Boundaries</td>
</tr>
<tr>
<td>Subsection</td>
<td>Grandmother’s Voices: Dominant Traditional Orientation</td>
</tr>
<tr>
<td>Subsection</td>
<td>Mothers’ Voices: Between Two Spaces.</td>
</tr>
<tr>
<td>Subsection</td>
<td>Daughters’ Voices: Challengers within Patriarchal Boundaries-Explicit and Implicit Resistance</td>
</tr>
<tr>
<td>Subsection</td>
<td>The Schoolgirls</td>
</tr>
<tr>
<td>Subsection</td>
<td>Girls Who Dropped Out of School</td>
</tr>
<tr>
<td>Section</td>
<td>Discussion</td>
</tr>
<tr>
<td>Section</td>
<td>Notes</td>
</tr>
</tbody>
</table>
Table 3.3: Example of the logical structure of an article in American Journal of Sociology

<table>
<thead>
<tr>
<th>Title</th>
<th>Relative Deprivation and Internal Migration in the United States: A Comparison of Black and White Men</th>
</tr>
</thead>
<tbody>
<tr>
<td>Section</td>
<td>Trends In U.S. Internal Migration</td>
</tr>
<tr>
<td>Section</td>
<td>Theoretical Background</td>
</tr>
<tr>
<td>Subsection</td>
<td>Migration and Social Mobility</td>
</tr>
<tr>
<td>Subsection</td>
<td>Relative Deprivation and Migration</td>
</tr>
<tr>
<td>Subsection</td>
<td>Reappraising the Social Mobility Consequences of Migration</td>
</tr>
<tr>
<td>Section</td>
<td>Data And Methods</td>
</tr>
<tr>
<td>Subsection</td>
<td>Model Specification</td>
</tr>
<tr>
<td>Subsection</td>
<td>Dependent Variables: Absolute and Relative Deprivation Outcomes</td>
</tr>
<tr>
<td>Subsection</td>
<td>Independent Variables</td>
</tr>
<tr>
<td>Section</td>
<td>Results</td>
</tr>
<tr>
<td>Section</td>
<td>Selection Into Migration And Employment</td>
</tr>
<tr>
<td>Section</td>
<td>Discussion And Conclusions</td>
</tr>
<tr>
<td>Section</td>
<td>Appendix</td>
</tr>
<tr>
<td>Section</td>
<td>References</td>
</tr>
</tbody>
</table>
To ensure a realistic annotation model for all scientific documents, the first step was to build a user model based on a survey and interviews (see Chapter 2) and the review of the literature. In the user model, we observed that scientists want to find information depending on different types of discourse elements, such as findings, definition, hypothesis, methodology and related work. We also observed that depending on the task they have to accomplish, the pieces of text in which they are interested differ. The analysis of the interviews shows that when scientists mention that they are interested in the "results" of the document, they are not necessarily interested in the part of the document under the heading "results" but rather the chunk of text that conveys the results of the study of related work. The model was constructed in several iterations following the schema shown in Figure 3.4. In the iterations, we considered the analysis of the review of the literature and incorporated ideas and models from various studies into the model.

![Figure 3.4: Method for the creation of the model](image)

### 3.3.1 Components of the model

In the obtained conceptual model\(^\text{27}\), the scientific document is decomposed into different chunks of text that can contain one or more sentences. The text is the concatenation of all chunks of text. The model assumes that the most efficient way to decompose a scientific document consists of four dimensions (see figure 3.5), which can also be used in combination to query a document. These dimensions can be viewed as facets:

1. Meta-data dimension
2. Conceptual dimension (terms and concepts)

---

\(^{27}\)Some of the results presented in this section were published in [Ribaupierre & Falquet 2010, Ghoula et al. 2011, Ribaupierre 2012, Ribaupierre & Falquet 2013, Ribaupierre & Falquet 2014a, Ribaupierre & Falquet 2014b]
3. Discourse dimension (hypothesis, definition, etc.)

4. Relationships dimension

Figure 3.5: Conceptual model

### 3.3.1.1 Meta-data dimension

The first dimension consists of standard meta-data in the field of scientific documents (bibliographic data), such as the names of authors, titles of articles, names of journals or publishers, date of publication, sources, and DOI. We observed during the interviews (see chapter 2 table 2.11) that the metadata are not sufficient to respond to precise question. Yet the metadata remain essential because scientists are following the publications of a number of authors and because they make identification of a particular document possible. The bibliographic data are independent of the domain of research.
3.3.1.2 Conceptual dimension (terms and concepts)

The second dimension consists of the representation of terms. In the interviews, we observed that the search by keywords was important for scientists, even if it was sometime difficult to identify the right keywords. In order to obtain relevant results, they have to try several combinations of keywords and synonyms (see chapter 2 section 2.4.2.2 and table 2.11). Each chunk of text can contain one or more scientific terms. In SciAnnotDoc, terms are semantically represented by concept IDs based on auxiliary ontologies. It is possible to add as many ontologies as necessary. Terms can be very domain specific, such as protein, gender, and salary gap, or they can be more generic and domain independent, such as types of methods, types of variables, tools, and tests. Terms can also define scientific objects. A scientific object is an object used in different domains, such as equations, models, algorithms, and theorems. Different domain ontologies can describe these terms. Indexing the terms contained in the fragments by a concept’s ID and not only by keywords allows for the resolution of problems such as the ambiguity of a word, homonymy and synonymy and, as we will demonstrate in the following chapter, allows for new inferences to be made.

3.3.1.3 Discourse dimension

The third dimension is associated with discourse elements, the core of SciAnnotDoc. A discourse element is a chunk of text. This dimension is not research-domain dependent; although the example has been extracted from an annotated corpus in gender studies, the model can be applied to any scientific document. The type of discourse elements we selected are those that scientists use most frequently (see chapter 2 figure 2.17). Depending on the decomposition of the discourse, elements can have sub-elements. Each discourse element can be associated with one or more concepts/terms of one or more of the domain ontologies. Each element is also associated with bibliographic data.

**Definition**  Definition was often mentioned as an element sought by interviewees. For example, one interviewee was searching for all the definitions of "social transactions", and another was looking for the definitions of "semantic homogeneity". In certain domains of research, especially when a definition shows no consensus, the search for different definitions is an important task. This task can also be crucial for scientists who enter a new domain of research (e.g., PhD students or researchers who change their research domain) or those in interdisciplinary studies.

In linguistics, a definition is composed of the definiens and the definiendum. Consider the following definition: "Cat (Felis catus), also known as the domestic cat or housecat to distinguish it from other felines and felids, is a small, furry, domesticated, carnivorous mammal that is valued by humans for its companionship and for its ability to hunt vermin and household pests". The definiens is the

---

sentence or the word that defines the definiendum: "also known as the domestic cat or housecat to distinguish it from other felines and felids, is a small, furry, domesticated, carnivorous mammal that is valued by humans for its companionship and for its ability to hunt vermin and household pests". The definiendum is the term, in this example: "cat". In addition to providing a more precise indexing, separating the definiens and the definiendum is a way to solve homonymy and synonymy problems. The definiendum can be associated with several linguistic forms (synonyms) and to a precise concept in a domain ontology. This decomposition allows for specific queries such as Find all the definitions of the term "gender" where the definiens contains "social construction".

Some examples of definitions include the following:

"On the other hand, gender has increasingly become used to refer to any social construction having to do with the male/female distinction, including those constructions that separate "female bodies" from "male" bodies". [Nicholson 1994, p.79]

"Persons at work (1) are defined as persons who during the reference period performed at least one hour of work for a wage or salary, or for profit or family gain, in cash or in kind (International Labour Organization [ILO] 1994)". [De Henau et al. 2010, p.67]

**Findings**  Findings regroup all research results, observations, discussions, claims, outcomes, interpretations and conclusions of a document. As indicated in the literature, it can be difficult to differentiate between these elements, and the meanings of these terms can vary among research areas or even scientists. Findings are subdivided into two sub-categories:

**Raw findings** data description, interview reported, etc.

**Interpreted findings** findings that have been analysed.

The following sentence could be an example of raw findings, in which the data are presented but not analysed:

"A 1909 survey of recent graduates from Los Angeles High School found that only 34 percent of the girls were in college, compared with 64 percent of the boys; and 44 percent of the girls were in normal school, compared with just 2 percent of the boys". [Brown 1990, p.504]

The following sentences are examples of findings that have been analysed (discussed).

"In a macroeconomic study of South Korea and Taiwan, Seguino (2000b) finds a positive correlation between total FDI (inward plus outward FDI) and the gender-wage gap in Taiwan but no similar relationship in Korea".[Braunstein & Brenner 2007, p.215]
3.3. Our model: SciAnnotDoc

"AMM, GS: Studies show that women faculty stay longer in each rank before being promoted and are less likely to obtain full professor than their male counterparts".[May & Summerfield 2012, p.32]

However, the distinction between raw and analysed findings may be difficult to establish. Depending on the context, the following sentence could be annotated as a raw finding or as an analysed finding.

"Worldwide, over 8.6 million women die from cardiovascular disease (CVD) each year [1]".[Schenck-Gustafsson 2009, p.186]

In this case, the fragment will have both types.

Hypothesis The hypothesis elements typically propose answers to open questions. They do not necessarily exist in all scientific documents. There are many studies that do not describe or even propose a research hypothesis, especially in research using an inductive approach. One can find very strict definitions of a hypothesis, such as, "Scientific hypothesis, an idea that proposes a tentative explanation about a phenomenon or a narrow set of phenomena observed in the natural world. The two primary features of a scientific hypothesis are falsifiability and testability, which are reflected in an 'If...then' statement summarizing the idea and in the ability to be supported or refuted through observation and experimentation. The notion of the scientific hypothesis as both falsifiable and testable was advanced in the mid-20th century by Austrian-born British philosopher Karl Popper"[Encyclopaedia Britannica Online 2014b]. However, from a more general perspective, can we consider questions to be hypotheses? For example, one of the interviewees in our survey said that she was interested in determining whether it was possible to transfer the annotations in a parallel corpus when some parsing had already been performed. This conjecture can be viewed as a hypothesis if we consider the broad definition of hypothesis as an element that proposes to answer an open question.

"Are there specific contexts where you find that women or women’s groups approach the governance of the commons differently from men?"[May & Summerfield 2012, p.31]

"Second, we hypothesized that there would be regional differences in the emphasis placed on various gender-specific norms".[Suitor & Carter 1999, p.89]

"One possible reason is that the dimensions of the coronary arteries and lungs are smaller in women, and therefore the same level of smoking might be more harmful to female than to male organs."[Schenck-Gustafsson 2009, p.188]
Methodology  Methodology elements describe the different steps taken in an experimental study or the methodology used to interview a sample of a population, such as the choice of sample or the test used. For example, in our interview, one interviewee was interested in finding the methods used to analyse the indices of variability in terms of the brain and behaviour. Another interviewee indicated that he was interested in finding all the implementations of a data store, the details of the implementation and the choice of platforms and techniques.

The following sentences provide examples of methodology elements:

"Carla Freeman’s (2000) study of women office workers in multinational corporations in the Caribbean provides a case study in which employer-dictated dress codes symbolize both gender and class in the culture at large".[Hopkins 2007, p.299]

"Nevertheless, the four items analyzed here are more detailed and sophisticated than the other GSS gender-role items, and are similar to items used in many other surveys (Mason et al. 1976)".[Mason & Lu 1988, p.43]

"The reports of the statistical bureau include monthly statistical reports, giving a survey of the condition of crops and special statistical papers on such subjects as freight rates, the production of farm products in foreign countries, etc.; monthly crop synopses, being a four-page summary of the conditions, prospects, yield, price, distribution, and consumption of crops, and the number and value of farm animals".[Wright 1895, p.267]

"Similarly, the four focus groups were to have been organised and run by the research organisation - this was now changed so that there was input in the form of developing areas of discussion from the two women sociologists who would, in addition, facilitate one of the focus groups and take responsibility for analysing and writing up the interview and focus group data".[Logan & Huntley 2001, p.625]

Related work  The analysis of the interviews with scientists made us realise that scientific documents should be annotated from a "universal knowledge" perspective and not from a perspective centred on the author. That is, the interesting definitions and findings from a reader’s point of view are not necessarily those created by the author but can be those reported by an author in his/her document. For example, when an interviewee mentioned finding "the different articles that deal with the evaluation of surgical simulators", this scientist indicated that she was interested in finding all documents on this subject, irrespective of the author. In her practice, she begins by looking for surveys on techniques for assessing surgical simulators. If such surveys do not exist, she begins reading various articles on the subject and looking for passages that address these techniques, regardless of whether the passages of
interest pertain to work by the author of the article or provide references to other articles.

Related work elements are always associated with another type of element (definition, findings, methodologies, or hypothesis).

References to other authors can be made in a very formalised way (as in this thesis), but they can also be made simply by indicating a name[^29] or even a school of thought. In the following sentence, the name "Butler" is an author, and the term "Foucauldian" refers to the author Foucault. In the model, the following example should be indicated as a definition (of the concept "sexual identity") and as related work with two documents.

"As an antidote to the production and reinforcement of prevailing notions of sexual identity, Butler argues that homosexuality and heterosexuality-like gender-exist as enactments of cultural and aesthetic performances; even as these identities are performed and repeated, they are (in true Foucauldian form) being contested and unraveled".[Deveaux 1994, p.239]

In the model, the following sentence should be annotated as both a methodology and related work.

"Cherlin and Walters (1981) and Ransford and Miller (1983), for example, examined attitudes about the suitability and appropriateness of women's presence in the public worlds of work and politics, while Porter Gump (1975), Randon Hershey (1978), and Hermons (1980) measured women's identification with the values of "femininity" and the private roles of wife and mother".[Dugger 1988, p.429]

A chunk of text can be annotated with more than one discourse type because it may belong to more than one category [Teufel & Moens 2002, Ibekwe-SanJuan 2006].

The following example could be annotated as both a definition and a finding[^30]:

"We find, for example, that when we use a definition of time poverty that relies in part on the fact that an individual belongs to a household that is consumption poor, time poverty affects women even more, and is especially prevalent in rural areas, where infrastructure needs are highest." [Bardasi & Wodon 2010, p.70].

The following example could belong to three discourse types: a finding, a methodology and a definition. It can be classified as a definition because of the reference to a definition in the sentence; although the term "work" is not defined

[^29]: Usually, this is the case when an author's work has almost become part of everyday knowledge.

[^30]: The text in blue is a finding, the text in red is a definition, the text in orange is a methodology and the text in violet is a hypothesis.
in the sentence, the reference will indicate that a definition of this word is located somewhere in the text. Because the aim of this annotation model is to be integrated into an indexing model for an information retrieval system and because it does not constitute a summarising system, the reference of the discourse element sought by users should be a sufficient indication for users to look at the document. This is what we call an "indirect definition".

"When the proportion of children aged 6-14 reported to be working is tabulated against a variable indicating whether there was a proxy respondent, we find that there is no significant difference in activity rates by type of respondent for either boys or girls when we use the market definition of work (defined below)". [Assaad et al. 2010, p.87].

The following sentence could be annotated as a finding and a hypothesis.

"The finding that a majority of women who have reached positions of the greatest power and authority in their respective professions continue to fight against gender discrimination provides home support for the hypothesis that discrimination continues to inhibit women’s access to power". [Black & Rothman 1998, p.117]

Each element of discourse can be related to another discourse element or to a document.

3.3.1.4 Relationships dimension

The fourth dimension consists of all explicit references established between a document or discourse element and another document or discourse element (inter-document). The aim of element relationships is to bind two documents together using meaningful links. For example, one of the interviewees indicated that she was interested in finding all documents in which one author disagrees with another author: "I want to find all the articles where Christine Delphy disagrees with Patricia Roux" (see chapter 2 table 2.10). The relationships between both documents must be indicated with a semantic relationship. Another interviewee wanted to search for all the definitions of gender, the historicity of the definitions and the controversy associated with different definitions. In this case, it is necessary to not only link two documents but also to precisely link the different discourse elements with a specific relationship.

In the model, the following sentence should be annotated as both a definition of the concept "gender" and related work presenting a positive relationship with the document of Joan Scott [Scott 1999].

"Joan Scott provides an eloquent description of this second understanding of gender in which the subsumption of sex under gender is made clear: "It follows then that gender is the social organization of sexual difference". [Nicholson 1994, p.79]
The following sentence (see Figure 3.6) should be annotated both as a finding and as related work containing a negative relationship between England and colleagues, on the one hand, and between England and human capital theory, on the other. In this case, three discourse elements should be created: the first represents the information extracted from the document of Correll, the second represents England, and the third represents human capital theory. The relationships will be between England and human capital theory.

"However, England and colleagues (1984; 1988) demonstrate that, contrary to the predictions of human capital theory, women employed in male-dominated occupations actually have higher lifetime earnings".[Correll 2004, p.94]

Figure 3.6: An example of relationships between discourse elements: Correll, England and colleagues and human capital theory

Another aim of identifying element relationships is to be able to infer new information from those relationships (see Figure 3.7). Certain relations between documents are never made explicitly, for instance, because two publications are published at the same time, because the authors are not aware of the other researchers’ work, or because the authors do not want to refer to another author for various reasons (e.g., difference of opinion). The reader may want to find the "missing" reference. To help readers navigate the literature and find documents that may be related to an author, it could be interesting to infer certain relations. For example, if author A cites author B, author B cites author C, and all define the terms "social constructivism", it is likely interesting for other researchers to access all three definitions in relation to each other. To this end, we must infer certain relations between A and C. By using the type of relationships between fragments, the "fil d’Ariane" becomes even more precise, and it is possible to create a network of relations between fragments that can be assembled, for example, to form a school of thought or to emphasise incoherence in the literature.
Figure 3.7: Lineage of idea or "fil d'Ariane"
The second type of relationship is an intra-document relationship. Some discourse elements refer implicitly or explicitly to another discourse element within the same document. In Harmsze et al. 1999, the author proposes several types of links between the modules extracted from a document. In this model, we only propose one type of link, which is defined by all the different types of links that can be represented within a document. The aim of this link is to be able to show users a connection between different discourse elements but not necessarily to detail them. Because the aim of this study is to query a corpus precisely and not to summarise the content of documents, it is not necessary to have detailed links in the same document. When the objective is to summarise, the level of detail should be more important to understand the connections between a finding, a methodology and a hypothesis.
3.3.2 Formalisation of SciAnnotDoc

The SciAnnotDoc annotation (see Figure 3.8) model has been formalised as a set of OWL ontologies. We have defined four ontologies that are described below.

We have maintained a clear separation between the following dimensions: textual content, discourse elements and element relationships. Maintaining this distinction not only clarifies the reading of the ontology but also simplifies the addition of new ontologies or new concepts to an existing ontology.

3.3.2.1 The SciDeo ontology

The SciDeo ontology\(^{31}\) contains 22 classes, 98 object properties (including CiTO) and 13 data properties. The SciDeo ontology is composed of the following classes and properties:

A **document** is a piece of text. It can have as data properties a *title*, a *journal-Name*, and a *year_of_publication*, can contain some **Fragments** and is written_by a person.

A **Fragment** is a piece of text that is part_of a **Document**.

A **Discourse element** is a smaller piece of text that belongs to a **Fragment**. It can refer to another **DiscourseElement** that is part of the same **Document** and can CiTO:cites another **DiscourseElement** or **Document**. It can refer to another **DiscourseElement** within the same **Document**. It also uses the

---

\(^{31}\) All ontologies are available at http://cui.unige.ch/~deribauh/Ontologies/, and see appendix C.
3.3. Our model: SciAnnotDoc

concepts of different ontologies to describe the methods, the scientific object or the domain concept.

A Methodology is a subclass of DiscourseElement that defines the methods discourse element.

A Hypothesis is a subclass of DiscourseElement that defines the hypothesis discourse element.

A Finding is a subclass of DiscourseElement that defines the findings discourse element and has two subclasses: rawFindings and AnalysedFindings.

A RelatedWork is a subclass of DiscourseElement that defines all the discourse elements that CiTO:cites with another discourseElement or a Document.

A Definition is a subclass of DiscourseElement that defines the definition discourse element.

A Definiens is the sentence that defines the definiendum and is part_of a Definition

A Definiendum is the class that defines the concept of a definition and is part_of of a Definition. It can define a concept of the domain ontology and is different from the relation uses between the class discourse element and the domain concept in the sense that this relation pertains only to the definiendum of the definition and not all concepts contained in a definition.

A Person is generally the author of a Document. This class can be equivalent to another class in another ontology if the second ontology defines the researcher in the domain of research.

For the element relationships, we use the CiTO ontology [Shotton 2010] (see Table 3.1) and extend it to represent citations between documents as well as between discourse elements because a large proportion of citations do not refer to a whole document but rather to a short part (e.g., an equation, a sentence, a definition, a paragraph). Each related work element is associated with other elements (discourse elements or document) through one of the relationships defined in CiTO. This is necessary to answer precise queries such as "Find all the elements containing a finding about the difference between girls and boys in school and referencing a findings of Zazzo". It also becomes possible to perform more detailed analyses of the network of citations depending on the types of citing or cited elements. We keep the numerous types of citations defined in CiTO, but we group them in three upper-level citation types: positive (agreesWith, confirms, …), neutral (discusses, extends, reviews, …), and negative (corrects, critiques, disputes, …). CiTO is imported into the SciDeo ontology.

The Figure 3.9 is a representation of SciAnnotDoc in Protege\textsuperscript{32}, which was used to create the ontology.

\textsuperscript{32}http://protege.stanford.edu
The bibliographic data was not defined as a separate ontology in the present thesis because the only bibliographic information needed is the name of the author, the title of the document, the name of the journal and the year of the publication. These data are simply added to the discourse element ontology (SciDeo).

For textual content, when the research on which this thesis is based was initiated, there was no methodology, scientific object or concept domain available for annotating documents in gender studies or sociology. An ontology such as DBpedia\textsuperscript{33} is not sufficiently precise to annotate scientific documents.

3.3.2.2 The SciMeth ontology

The methodology ontology (SciMeth) (see Figure 3.10 and appendix D) was developed with the goal of creating an independent domain ontology. The aim of this ontology is not to describe the steps that the author has taken during his/her study but to describe the different types of methods that exist to make it possible to annotate documents to answer queries such as "all the findings about the gender difference in high-school on mathematical ability using a snow-ball sampling and using a MANOVA".

The general classes are domain independent and can be used to describe methods. Some subclasses may be more domain specific. The methodology ontology was developed within the same domain of research as the corpus of documents (i.e., gender studies and, more generally, human sciences) and in several iterations depending on the concepts found in the scientific documents annotated. The complete ontology contains 105 classes, 14 object properties, and 108 individuals (see Figure\textsuperscript{34} 3.11).

\textsuperscript{33}http://wiki.dbpedia.org/About

\textsuperscript{34}View with SOVA, http://protegewiki.stanford.edu/wiki/SOVA
At the first level, the methodology ontology is composed of the following:

**Methodology ontology: high-class level**

- **Method** this class is the central class of this ontology and has two sub-classes, *Qualitative_method* and *Quantitative_method*. It can use_tools to describe the different tools the methodology uses. It can measure_some_variables and use_statistical_tests that may test_some_variable. The class can use_techniques to describe the different types of methods a study uses.

- **Methodology Tools** this class describes all the different tools or artefacts that can be used in a methodology (e.g., questionnaire, IRM, or paper questionnaire).

- **Variable** this class describes all the different types of variables that can be used in a study and the different vocabularies that can be used to describe the same variable (e.g., a dependent variable can also be referred to as an experimental variable, an extraneous variable or a control variable, and an independent variable can also be referred to as an explanatory variable).

- **Sampling_techniques** this class describes all the different sampling techniques (e.g., accidental sampling and ad hoc quotas).

- **Statistical_tests** this class describes the different statistical tests that exist (e.g., MANOVA, Cohen’s kappa, and Student’s t-test). Statistical_tests rely on a Variable with the object properties tests_some_variable.

- **Methods_technique** this class describes the different types of methods used in different domains of research (e.g., experimental design, historical comparison, and observation).
Figure 3.11: Methodology ontology: all classes and properties (to give an idea of the size and the structure of the ontology)
3.3.2.3 The SciObj ontology

The scientific object ontology (SciObj) (see Figure 3.12 and appendix E) was developed to describe the different scientific objects that authors can discuss in their documents. The term scientific object in this thesis defines the different objects that scientists might use and describe in their studies and that have been used more than once. For example, an algorithm is a scientific object and has subclasses and individuals. The aim of this ontology is to be able to annotate documents to answer queries such as "all the findings about the gender difference in high-school in mathematical ability a test of achievement such as the Kaufman Test of Educational Achievement". In this thesis, we describe general concepts as scientific objects, such as for the methodology ontology. We have adapted this ontology to the domain of the documents to be annotated. This ontology contains 74 classes, 4 object properties, 1 data property and 72 individuals. In the following description, we only describe the upper-level classes and some of the properties (see Figure 3.13).

Figure 3.12: Scientific object ontology: high-class level

- **Scientific object** this class describes the different scientific objects.
- **Creator** this class describes the creator that has developed the scientific object (it can be one person, or a group of person)
- **Research project** this class describes the different projects within the scientific object is used.
- **Test** is a subclass of Scientific object and describes the different test that can be used in sociological and psychological research. It can measure some Measured variable.
- **Achievement test** is a subclass of Test and measures Academic achievement.
- **Afrobarometer** is a subclass of Test and measures the behaviour pattern and the people attitudes.
IQ_test is a subclass of Test and measures intelligence.

Measured_variable is a subclass of Scientific_object and describes the different variables that a Test can measure. For example, a Kaufman test of educational achievement measures academic achievement.

model is a subclass of Scientific_object and describes the different models used in more than one study.

Algorithm is a subclass of Scientific_object and describes the different algorithms used in more than one study.

Equation is a subclass of Scientific_object and describes the different equations used in more than one study.

Theory is a subclass of Scientific_object and describes the different theories used in more than one study.
3.3. Our model: SciAnnotDoc

Figure 3.13: Scientific object ontology: A summary view (to give an idea of the size and the structure of the ontology)
3.3.2.4 The GenStud ontology

The domain concept ontology (see Figure 3.14 and appendix F) was developed for the domain of gender studies (genStud). The aim of this ontology is to describe scientific concepts within gender studies. We developed an ontology, but a taxonomy could be sufficient. This ontology has 798 classes, 10 object properties and data properties (see Figure 3.15).

Figure 3.14: Gender studies ontology: high-class level

![Ontology Diagram]

- **Variable analysed** this class describes the different variables analysed in the study and studied by one or several **Field of research**.
- **Person** this class describes the person that may represent the domain of research. This class is aligned with the class Person in the SciDeo ontology.
- **Field of research** this class describes the different research domain where a person conducts research.
- **School of Thought** this class describes the different schools of thought that exist. A Person belongs to one or several School of Thought

The SciMeth, SciObj and GenStud populations were created by iteration at the same time the annotation process was run. A JAVA program was implemented to automatically assign properties and classes to a list of terms.
3.3. Our model: SciAnnotDoc

Figure 3.15: The GendStud ontology: A summary view (to give an idea of the size and the structure of the ontology)
3.3.2.5 Inference rules

For the relationships between discourse elements or documents, several rules were created and implemented in SWRL\textsuperscript{35}. Two rules were created to simplify the annotation process:

The relations *defines* and *is* \_\_ *defined* between the definiens and the definiendum (see Figure 3.16) are automatically generated by a SWRL rule:

\[
\text{Definiendum}(?x), \text{Definiens}(?y), \text{Definition}(?z), \\
\text{part}_{-}\text{of}(?x,?z), \text{part}_{-}\text{of}(?y,?z) \implies \text{is}_{-}\text{defined}(?y,?x)
\]

Figure 3.16: Definition, Definiendum, Definiens concepts in the SciAnnotDoc

\begin{center}

\begin{tikzpicture}

\node (definition) at (0,0) {Definition};
\node (definiendum) at (-3,-2) {Definiendum};
\node (definiens) at (3,-2) {Definiens};
\node (domain_concept) at (0,-4) {Domain Concept};

\draw [dotted] (definiendum) -- (definiens); 
\draw [dashed] (definiendum) -- (definition); 
\draw [dashed] (definiens) -- (definition); 
\draw [->] (definiendum) -- (domain_concept) node [midway, above] {uses};
\draw [->] (definiens) -- (domain_concept) node [midway, below] {define};
\end{tikzpicture}
\end{center}

The class *RelatedWork* is generated automatically by a SWRL rule:

\[
\text{DiscourseElement}(?x), \text{DiscourseElement}(?y), \\
\text{cites}(?x, ?y) \implies \text{RelatedWork}(?x)
\]

The number of inferences possible based on the SciAnnotDoc model is high, but the risk of adding more noise versus adding information that is useful to the reader is significant. The following rule was created to infer "the lineage of idea". To avoid adding noise, we only inferred relations between discourse elements of the same type and that cite the same author/discourse element but do not cite each other (see Figure 3.17):

\textsuperscript{35}see appendix C for the complete SWRL rules
3.3. Our model: SciAnnotDoc

Figure 3.17: Inference between discourse elements
Chapter 3. Annotation models for scientific documents

The relation *inferredReference* is inferred from the following SWRL rules:

\[
\text{Definition}(?x), \text{Definition}(?y), \text{Definition}(?z), \\
\text{cites}(?x,?y), \text{cites}(?y,?z)), \rightarrow \text{inferredReference}(?x,?z)
\]

\[
\text{Findings}(?x), \text{Findings}(?y), \text{Findings}(?z), \text{cites}(?x,?y), \\
\text{cites}(?y,?z)), \rightarrow \text{inferredReference}(?x,?z)
\]

\[
\text{Hypothesis}(?x), \text{Hypothesis}(?y), \text{Hypothesis}(?z), \\
\text{cites}(?x,?y), \text{cites}(?y,?z)), \rightarrow \text{inferredReference}(?x,?z)
\]

\[
\text{Methodology}(?x), \text{Methodology}(?y), \text{Methodology}(?z), \\
\text{cites}(?x,?y), \text{cites}(?y,?z)), \rightarrow \text{inferredReference}(?x,?z)
\]

3.4 Discussion and Conclusion

In this chapter, we presented different models of annotation for scientific documents that have been proposed and their limitations. In the second part, we presented SciAnnotDoc, a new model that represents one of the main topics of the present thesis, and we described the dimensions of a scientific document: bibliographic data, textual content (terms and concepts), discourse elements, and element relationships. This model is focused on the needs of the user and was constructed based on the discourse elements that are most frequently used by scientists (see chapter 2). As Semantic Web technologies undergo constant change, some ontologies that did not exist or were not available when the research for this thesis was initiated are now available, such as ontologies for bibliographic data based on Dublin Core. SciAnnotDoc is a flexible model, and the integration of new ontologies in the near future is possible. The SciDeo ontology is the core of SciAnnotDoc, and the other ontologies contained in the SciAnnotDoc represent the two other dimensions of the conceptual model (CiTO ontology is the relationships dimension, and GenStudy ontology, the SciMeth ontology and the SciObj ontology represent the conceptual dimension). As we will demonstrate in the next chapter, the annotation was performed principally in the domain of gender studies, which is why some subclasses of the independent domain ontologies were centred on this domain. As mentioned, if the domain of the corpus changes, the subclass can be defined.

We will also see that the inferences proposed in the SciAnnotDoc model have been tested on a corpus that was manually annotated. In the future, it would be interesting to study inferences as a function of the citation types defined in CiTO (e.g., AgreesWith and DisagreesWith). In the present thesis, we grouped these relations into three categories: negative, positive and neutral. Thus, it should be possible to determine whether, when element A positively cites element B and element B positively cites element C, element A also positively cites element C. This type of transitive inference should provide more information to scientists in following the lineage of an idea or in monitoring the consistency or inconsistency between studies.
4.1 Introduction

In this chapter, we present the annotation processes we used to annotate scientific documents using the SciAnnotDoc model. We tested three different approaches for the annotation of scientific documents. The first was a manual approach used to test whether the SciAnnotDoc model was realistic. During this first phase, we introduced some improvements to the model. The annotations were performed by the author of this thesis. The second approach involved assessing whether crowdsourcing could be a realistic approach for this type of annotation. The third approach was an automatic approach that incorporated certain natural language processing (NLP)
techniques. This approach required some changes to SciAnnotDoc to be able to process the annotations automatically. These three approaches are described below.

The SciAnnotDoc annotation model is a generic model for scientific documents. As mentioned previously, the test corpus is composed of gender studies documents. We chose this domain because it consists of very heterogeneous written documents ranging from highly empirical studies to "philosophical" texts, and these documents rarely use the IMRaD model (introduction, methods, results and discussion). Of course, this corpus is more difficult to annotate than a more standardised corpus, such as a corpus of medical documents, but it also presents more challenges (see below). We argue that if our annotation model can be applied to such a heterogeneous corpus, it should also apply to other, more homogeneous types of papers. Therefore, the model should be generalisable to other domains.

For a review of the literature on manual and automatic annotation, the reader can refer to the previous chapter (see Chapter 3). In this chapter, we discuss the advantages and disadvantages of manual annotation versus the automatic annotation process, the precision of the crowdsourcing classification task and automatic annotation processes.

4.2 Manual annotation

The first attempt to manually annotate scientific documents was made by the author of this thesis. At the beginning of this thesis, the choice of the research domain was not obvious. The purpose of this thesis was not to simply improve the automatic process. The criteria for choosing the research domain were as follows: the research domain must present enough variety in writing style and the author of this thesis must understand the meaning of the documents to annotate. Of course, interest in the domain was also important. The first annotated documents were in three domains: computer science, physics history, and gender studies. The first step was to manually identify the different discourse elements and the relationships between the discourse elements. Twelve documents were manually annotated using codes and colour in the margin of the printed documents (see Figure 4.1). From among these 12 documents, 4 documents were fully annotated using the SciDeo, the sciMeth, the SciObj and, finally, the GendStud ontology. The annotations were stored in an OWL ontology using Protégé (see Table 4.1 for the coverage of the annotations). These annotated documents were used, first, to test the feasibility of annotating scientific documents using the SciAnnotDoc model and, second, to improve the different ontologies, such as the SciDeo, the SciMeth, the SciObj and the GendStud ontologies. The process was iterative. Each iteration allowed us to populate not only the different domain concept ontologies that were missing but also the SciDeo ontology, allowing us to use and test the SciDeo for different types of documents with different relations between discourse elements or documents. This methodology allowed us to obtain a more complete and realistic domain concept ontology and a more realistic annotation model.
The manual annotation of one article took approximately 4 hours (reading and annotating). It soon appeared necessary to find a way to accelerate the annotation process to constitute a larger corpus, which could then be evaluated with users. The goal was to produce a sufficiently large annotated corpus and use it to evaluate the IR tasks. The first solution we adopted to improve the speed of annotation was crowdsourcing.
Figure 4.1: Manual annotation of three paper documents

---

1. Introduction

The question of assigning information seekers in locating a specific category (facet) of information has rarely been addressed in the IR community due to the inherent difficulty of such a task. Indeed, efficiency and effectiveness have been the main guiding principles in building IR models and tools. Our aim here is to delve into the problem of how to assist a researcher or a specialist in rapidly accessing a specific category or facet of information in scientific texts. For this, we need annotated corpora where relevant sentences are marked up with the type of information they are pertinently carrying. We identified eight categories of information in abstracts, which can be useful in the framework of information-centric design IR: OBJECTIVE, RESULT, NEWTHING, HYPOTHESIS, FINDINGS, RELATED WORK, CONCLUSION, FUTURE WORK. These categories enable the user to identify what a paper is all about and what the contribution of the author is to higher field. We adopted a surface linguistic analysis using lexico-syntactic patterns that are generic in a given language and rely on surface cues to perform sentence annotation from scientific abstracts. Once annotated, the corpus is fed into an automatic summarizer which takes into account the different semantic annotations for query-oriented document ranking and automatic summarization. The automatic summarizer used here is E3m-Text developed by LIA team at the University of Avignon (Fernandez et al., 2007). E3m-Text is based on neural networks (ANN) inspired by statistical physics to study fundamental problems in Natural Language Processing, like automatic summarization and topic segmentation.

In this paper, we will present some preliminary experiments on abstract ranking and automatic summarization using the semantic annotations resulting from our sentence categorization scheme. The plan of this paper is as follows: section 2 recalls relevant related work; section 3 describes the sentence categorization method. Section 4 describes the query-oriented abstract ranking and automatic summarization experiments using the semantic annotations. Section 5 discusses difficulties inherent in this task as well as earlier
Table 4.1: Coverage of the four manually annotated documents (Ref = references to other works)

<table>
<thead>
<tr>
<th>Doc No.</th>
<th>Total Sentences</th>
<th>Coverage</th>
<th>Ref total</th>
<th>Ref</th>
<th>Total</th>
<th>Ref total</th>
<th>Ref</th>
<th>Total</th>
<th>Ref total</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>97</td>
<td>60.8%</td>
<td>58.1%</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2.1%</td>
<td>50%</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>283</td>
<td>42.0%</td>
<td>29.2%</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3.2%</td>
<td>60%</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>355</td>
<td>55.6%</td>
<td>27.9%</td>
<td>18.3%</td>
<td>9.9%</td>
<td>18.3%</td>
<td>7.0%</td>
<td>66%</td>
<td>19.2%</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>355</td>
<td>64.2%</td>
<td>42.0%</td>
<td>30.2%</td>
<td>7.3%</td>
<td>42.0%</td>
<td>14.9%</td>
<td>4%</td>
<td>4.2%</td>
<td>0</td>
</tr>
</tbody>
</table>
Chapter 4. Manual and automatic corpus annotation

4.3 Crowdsourcing annotation

4.3.1 Introduction

The idea behind crowdsourcing is to use a large group of people to perform an online task. Crowdsourcing is used for various activities, such as the analysis of the content of pictures (search for a missing person or a missing plane), the classification of galaxies, and the annotation of textual content\(^1\). As the number of projects listed in the Wikipedia page shows, the use of this collective "intelligence" has become increasingly important and easier to access. Platforms such as Amazon Mechanical Turk (MTurk)\(^2\) provide access to a large number of participants ("workers") around the world to achieve small tasks for a small amount of money. The literature reveals several studies that use MTurk to complete certain classification tasks or annotations for textual documents, but, to the best of our knowledge, only scientists or experts have been asked to manually annotate the discourse elements extracted from scientific documents. In [Snow et al. 2008], the authors propose the evaluation of annotations of natural language tasks created by workers using MTurk. The authors observe a high level of agreement between MTurk non-expert workers and the existing gold standard provided by expert labellers. In [Callison-Burch & Dredze 2010], the authors relate various experiments performed to create speech and language data. In [Kong et al. 2014], the authors propose a task of selecting sentences that correspond to visual markings in a chart. The authors evaluate the information extracted by workers against a gold standard produced by an expert in the field and report that the distance between the raw set produced by a worker and the gold standard is 0.54 (out of a maximum of 1.0 indicating no concordance). In this thesis, we used two platforms to perform crowdsourcing with two different types of people. The first involved scientists, and we used LimeSurvey to perform the annotation. The second involved a general population using Amazon Mechanical Turk.

4.3.2 Scientists’ annotation study

We asked three scientists to annotate 367 sentences extracts from a scientific document in the domain of gender studies. Scientists one and two answered 4 on a scale of 5 points to a question asking about their level of knowledge in gender studies, and the third scientist answered 2. By consequence, scientists 1 and 2 are considered as expert in gender studies. The task was to classify the 367 sentences (extracts) into only one of five categories (Definition, Findings, Methodology, Hypothesis and Other). We used LimeSurvey to construct the survey. The instructions defined the different categories as follows:

A definition is a statement that explains the meaning of a term (a word, phrase, or other set of symbols).

---

\(^1\)A list of crowdsourcing projects can be found on Wikipedia: http://en.wikipedia.org/wiki/List_of_crowdsourcing_projects

\(^2\)https://www.mturk.com/mturk/welcome.
4.3. Crowdsourcing annotation

A **finding** is a result, a conclusion, a discussion, etc. of a study or an investigation presented or referenced by the authors of the document.

A **methodology** is a collection of methods, practices, procedures and rules used in a study.

A **hypothesis** is a proposed explanation or a question for a phenomenon.

**Other** Any sentence that is not a definition, finding, methodology or hypothesis.

The average time to answer the 367 questions was 2h00. The scientists had no training in this task.

We tested the inter-agreement judgment with Fleiss’ kappa\(^3\). To interpret the score given by Fleiss’ kappa, we followed the interpretation in the Table 4.2.

<table>
<thead>
<tr>
<th>Fleiss’ Kappa measure, significance [Wikipedia 2013]</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0-0.2</td>
<td>slight agreement</td>
</tr>
<tr>
<td>0.21-0.4</td>
<td>fair agreement</td>
</tr>
<tr>
<td>0.41-0.6</td>
<td>moderate agreement</td>
</tr>
<tr>
<td>0.61-0.8</td>
<td>substantial agreement</td>
</tr>
<tr>
<td>0.81-1.0</td>
<td>almost perfect agreement</td>
</tr>
</tbody>
</table>

The results regarding Fleiss’ kappa (see Table 4.3), indicating agreement between the three annotators, were not very high for any of the discourse elements. The annotators appeared to agree more on the methodology.

<table>
<thead>
<tr>
<th>Table 4.3: Fleiss’ kappa results for the three scientific annotators by discourse element type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definition</td>
</tr>
<tr>
<td>Findings</td>
</tr>
<tr>
<td>Methodology</td>
</tr>
<tr>
<td>Hypothesis</td>
</tr>
<tr>
<td>Other</td>
</tr>
</tbody>
</table>

Fleiss’ kappa results obtained between the two annotator "experts" (see Table 4.4), that is, those who declared their level of knowledge in gender studies to be equal to 4 on the 5-point scale, were higher for the definition and the methodology categories than those obtained for the three annotators.

\(^3\)Cohen’s kappa is designed only for two judges; http://en.wikipedia.org/wiki/Fleiss’_kappa.
Table 4.4: Fleiss' kappa results for the two "expert" annotators in gender studies

<table>
<thead>
<tr>
<th>Definition</th>
<th>0.697</th>
</tr>
</thead>
<tbody>
<tr>
<td>Findings</td>
<td>0.388</td>
</tr>
<tr>
<td>Methodology</td>
<td>0.719</td>
</tr>
<tr>
<td>Hypothesis</td>
<td>0.384</td>
</tr>
<tr>
<td>Other</td>
<td>0.375</td>
</tr>
</tbody>
</table>

The cross-table between annotator 1 and annotator 2 (see Table 4.5) shows the discourse elements on which the annotators disagreed and, more precisely, which discourse elements were switched over. The hypothesis and findings appeared to be the discourse elements that were most frequently switched over.

Table 4.5: Cross-table between scientist 1 and scientist 2 for the five types of discourse elements

<table>
<thead>
<tr>
<th>Scientist 1</th>
<th>Def.</th>
<th>Find.</th>
<th>Meth.</th>
<th>Hypo.</th>
<th>Other</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Def.</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td>Find.</td>
<td>4</td>
<td>84</td>
<td>5</td>
<td>54</td>
<td>35</td>
<td>182</td>
</tr>
<tr>
<td>Meth.</td>
<td>0</td>
<td>3</td>
<td>50</td>
<td>2</td>
<td>16</td>
<td>71</td>
</tr>
<tr>
<td>Hypo</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>37</td>
<td>7</td>
<td>45</td>
</tr>
<tr>
<td>Other</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>7</td>
<td>39</td>
<td>56</td>
</tr>
<tr>
<td>Total</td>
<td>18</td>
<td>91</td>
<td>59</td>
<td>101</td>
<td>98</td>
<td>367</td>
</tr>
</tbody>
</table>

In the Table 4.6, we have computed the precision/recall for scientist 2 and used the annotation of scientist 1 as the gold standard. We can observe that the precision and recall are high, except for those associated with the methodology and hypothesis.

Table 4.6: Precision/recall for scientist 2 (gold standard (GS)= scientist 1)

<table>
<thead>
<tr>
<th></th>
<th>GS</th>
<th>Relevant Scientist 2</th>
<th>Total Scientist 2</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Def.</td>
<td>13.0</td>
<td>11.0</td>
<td>18.0</td>
<td>0.6</td>
<td>0.8</td>
<td>0.7</td>
</tr>
<tr>
<td>Find.</td>
<td>182.0</td>
<td>84.0</td>
<td>91.0</td>
<td>0.9</td>
<td>0.5</td>
<td>0.6</td>
</tr>
<tr>
<td>Hypo</td>
<td>45.0</td>
<td>37.0</td>
<td>101.0</td>
<td>0.4</td>
<td>0.8</td>
<td>0.5</td>
</tr>
<tr>
<td>Metho.</td>
<td>71.0</td>
<td>50.0</td>
<td>59.0</td>
<td>0.8</td>
<td>0.7</td>
<td>0.8</td>
</tr>
<tr>
<td>Other</td>
<td>56.0</td>
<td>39.0</td>
<td>98.0</td>
<td>0.4</td>
<td>0.7</td>
<td>0.5</td>
</tr>
</tbody>
</table>
4.3.2.1 Discussion

The results of this experiment were not very conclusive. First, the inter-judge agreement between the three annotators was not high. One reason for this relatively low score could be the number of categories. The more categories that are involved, the more the annotators may differ in their judgements. Another reason might be the interpretation of the instructions and, as a consequence, the misinterpretation of the sentences. After discussion with the second annotator, we discovered that she annotated most of the findings that were referred to as findings as something other than findings (general hypotheses). This type of error could be rectified by improving the precision of the instructions or by providing training. The definition and methodology categories appear to have been less confusing than the hypothesis and findings categories. This experiment was a good way to develop a gold standard to test whether non-expert crowdsourcing could function with a task that requires a relatively high level of understanding of the content of a scientific document and to evaluate expert workers against non-expert workers.

4.3.3 Amazon Mechanical Turk study

Amazon Mechanical Turk (MTurk) and other platforms of this type provide the opportunity to quickly implement an experiment. In this thesis, we wanted to determine whether an annotation is sufficiently precise when non-scientists annotate discourse elements extracted from scientific documents. We created a task in MTurk with the same sentences used in the previous experiment (scientists). This study was performed in parallel with the scientist annotation study. We paid workers 0.01$ by sentence, and we provided the same set of instructions. The batch has been on-line since 12 March 2013, and 15 workers have completed 158 of 367 sentences. MTurk gathered the 15 workers into two groups (Workers A and Workers B). The average time required to complete one sentence was 1 min 14 sec. The batch was established such that each sentence had to be judged by two workers.

As in the case of the scientists, we tested the inter-judge agreement between both groups of workers (see Table 4.7).

Table 4.7: Fleiss’ kappa results for the two groups of workers in Amazon Mechanical Turk

<table>
<thead>
<tr>
<th></th>
<th>Fleiss’ Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definition</td>
<td>0.316</td>
</tr>
<tr>
<td>Findings</td>
<td>0.487</td>
</tr>
<tr>
<td>Methodology</td>
<td>0.564</td>
</tr>
<tr>
<td>Hypothesis</td>
<td>0.609</td>
</tr>
<tr>
<td>Other</td>
<td>0.175</td>
</tr>
</tbody>
</table>

As in the case of the scientists, the Fleiss’s kappa was not very high for any
of the discourse elements, but it was not particularly low. The hypothesis was the element that obtained the best inter-judge score, followed by the methodology.

The cross-table between Workers A and Workers B (see Table 4.8) shows that the workers did not appear to confuse the definition and hypothesis categories or the hypothesis and methodology categories.

Table 4.8: Cross-table between Workers A and Workers B for the five types of discourse elements

<table>
<thead>
<tr>
<th>Worker A</th>
<th>Worker B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Def.</td>
<td>4</td>
</tr>
<tr>
<td>Find.</td>
<td>0</td>
</tr>
<tr>
<td>Hypo.</td>
<td>0</td>
</tr>
<tr>
<td>Meth</td>
<td>0</td>
</tr>
<tr>
<td>Other</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>4</td>
</tr>
</tbody>
</table>

We combined the two sets of data (scientists and workers), taking from the scientists’ set of data only the same sentences that were annotated by the workers. First, we ran a Fleiss’ kappa test on the five annotators’ results for the 158 sentences that had been evaluated by each group (see Table 4.9). Findings and hypothesis appeared to have the highest scores.

Table 4.9: Fleiss’ kappa results for the five annotators (scientists and workers) on 158 sentences

<table>
<thead>
<tr>
<th>Category</th>
<th>Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definition</td>
<td>0.353</td>
</tr>
<tr>
<td>Findings</td>
<td>0.550</td>
</tr>
<tr>
<td>Methodology</td>
<td>0.226</td>
</tr>
<tr>
<td>Hypothesis</td>
<td>0.473</td>
</tr>
<tr>
<td>Other</td>
<td>0.111</td>
</tr>
</tbody>
</table>

In a second phase, we hypothesised that the responses of scientist 1 (expert) were correct and established her results as a gold standard with which to compare the responses of the group of workers. The inter-judge agreement score was the highest for hypothesis, followed by the score associated with the methodology (see Table 4.10).
4.3. Crowdsourcing annotation

Table 4.10: Fleiss’ kappa results for Workers A and scientist 1 (on 158 sentences)

<table>
<thead>
<tr>
<th>Definition</th>
<th>0.316</th>
</tr>
</thead>
<tbody>
<tr>
<td>Findings</td>
<td>0.487</td>
</tr>
<tr>
<td>Methodology</td>
<td>0.565</td>
</tr>
<tr>
<td>Hypothesis</td>
<td>0.609</td>
</tr>
<tr>
<td>Other</td>
<td>0.175</td>
</tr>
</tbody>
</table>

Another cross-table (see Table 4.11) was established between Workers A and scientist 1. The table shows that Workers A annotated twice as many definitions as did scientist 1, but all six definitions annotated by scientist 1 were also annotated by Workers A. With respect to the findings, scientist 1 annotated 70 findings versus the 83 annotated by Workers A, among which 55 findings were common to both groups. The 15 other sentences were distributed between the definition (4) and hypothesis (11). For the methodology, out of 38 sentences, scientist 1 and Workers A shared 24 sentences. Workers A categorised the 14 other sentences as findings (10), definition (3) and other (1). For the hypothesis, out of 25 sentences, scientist 1 and Workers A shared 13 sentences in common. Workers A categorised the 12 other sentences as findings (8), methodology (2), definition (1) and other (1).

Table 4.11: Cross-table between Workers A and scientist 1 for the five types of discourse elements

<table>
<thead>
<tr>
<th>Scientist 1</th>
<th>Def.</th>
<th>Find.</th>
<th>Hypo.</th>
<th>Meth.</th>
<th>Other</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Worker A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Def.</td>
<td>6</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>18</td>
</tr>
<tr>
<td>Find.</td>
<td>0</td>
<td>55</td>
<td>8</td>
<td>10</td>
<td>10</td>
<td>83</td>
</tr>
<tr>
<td>Hypo.</td>
<td>0</td>
<td>11</td>
<td>13</td>
<td>0</td>
<td>1</td>
<td>25</td>
</tr>
<tr>
<td>Meth.</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>24</td>
<td>4</td>
<td>30</td>
</tr>
<tr>
<td>Other</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>6</td>
<td>70</td>
<td>25</td>
<td>38</td>
<td>19</td>
<td>158</td>
</tr>
</tbody>
</table>

We performed the same analysis to compare scientist 1 and Workers B. For the inter-judge agreement (see Table 4.12), the highest score was obtained for methodology, followed by that obtained for definition.
Table 4.12: Fleiss’ kappa results for Workers B and scientist 1 (on 158 sentences)

<table>
<thead>
<tr>
<th></th>
<th>Scientist 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definition</td>
<td>0.587</td>
</tr>
<tr>
<td>Findings</td>
<td>0.486</td>
</tr>
<tr>
<td>Methodology</td>
<td>0.590</td>
</tr>
<tr>
<td>Hypothesis</td>
<td>0.349</td>
</tr>
<tr>
<td>Other</td>
<td>0.197</td>
</tr>
</tbody>
</table>

A cross-table (see Table 4.13) was established for Workers B and scientist 1. We can observe that the group of workers annotated fewer definitions than scientist 1, and only three annotations showed consensus. Scientist 1 annotated 70 findings versus 57 for Workers B, among which 44 findings showed consensus. The 26 other sentences were distributed between the hypothesis (14), methodology (9) and other (3). Out of 38 sentences concerning the methodology, 31 sentences were shared by scientist 1 and Workers B. The 7 other sentences were distributed between the other categories without any preference. Out of 25 sentences concerning the hypothesis, 13 sentences were shared by scientist 1 and Workers B. The 12 other sentences were distributed between the findings (5), other (5) and methodology (2).

Table 4.13: Cross-table between worker B and scientist 1 for the five types of discourse elements

<table>
<thead>
<tr>
<th></th>
<th>Scientist 1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Def.</td>
</tr>
<tr>
<td>Worker B</td>
<td>Def.</td>
</tr>
<tr>
<td></td>
<td>Find.</td>
</tr>
<tr>
<td></td>
<td>Hypo.</td>
</tr>
<tr>
<td></td>
<td>Meth.</td>
</tr>
<tr>
<td></td>
<td>Other</td>
</tr>
<tr>
<td></td>
<td>Total</td>
</tr>
</tbody>
</table>

In the Table 4.14, we have computed the precision and the recall of the two worker groups (A and B). We can observe that the precision and the recall for each discourse element differ between the worker groups (i.e. the precision for definition for Workers A is 0.3 versus 0.8 for Workers B).

4.3.3.1 Discussion

Regarding the question of whether non-expert workers responding to MTurk who have no knowledge of gender studies\(^4\) can correctly classify sentences extracted from

\(^4\)Because we did not know the level of education of the workers in MTurk, we assumed that they did not have any knowledge of gender studies. In [Ipeirotis 2010], the author developed a
### 4.3. Crowdsourcing annotation

Table 4.14: Precision/recall for the two annotators (GS=golden standard)

<table>
<thead>
<tr>
<th></th>
<th>GS A (relevant)</th>
<th>Prec.</th>
<th>Rec.</th>
<th>F1</th>
<th>Workers B (relevant)</th>
<th>Prec.</th>
<th>Rec.</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Defi.</td>
<td>6</td>
<td>18(6)</td>
<td>0.3</td>
<td>1.0</td>
<td>4(3)</td>
<td>0.8</td>
<td>0.5</td>
<td>0.6</td>
</tr>
<tr>
<td>Find.</td>
<td>70</td>
<td>83(55)</td>
<td>0.7</td>
<td>0.8</td>
<td>0.7</td>
<td>0.8</td>
<td>0.6</td>
<td>0.7</td>
</tr>
<tr>
<td>Hypo.</td>
<td>25</td>
<td>25(13)</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.4</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Meth</td>
<td>38</td>
<td>30(24)</td>
<td>0.8</td>
<td>0.6</td>
<td>0.7</td>
<td>0.6</td>
<td>0.8</td>
<td>0.7</td>
</tr>
<tr>
<td>Other</td>
<td>19</td>
<td>2(0)</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
</tr>
</tbody>
</table>

Scientific documents into five categories, we did not observe a large difference between the scientists and the workers concerning the inter-judge agreement score. The scientists and workers appeared to have encountered some difficulties in agreeing on the categorisation. Only the definition and methodology (due to the confusion between the findings and hypothesis for scientist 2) yielded a higher score of agreement for scientists. By considering the responses of scientist 1 as the gold standard, we observed that the confusion between the categories was more significant for workers than for scientists, particularly for the methodology and definition categories. This finding lends some support to our hypothesis, according to which scientists should be better at recognising the types of scientific discourse with which they are familiar (cf. cross-table (see Table 4.13, 4.11, 4.5). The precision and recall for the scientists was slightly higher than it was for the workers, especially for the methodology and definition categories.

Depending on the income offered and the number of tasks, it is necessary to provide a small number of documents in a batch.

### 4.3.4 Conclusion

The categorisation of sentences into five categories is not an easy task, even for experts in the domain of interest. The inter-judge agreement was not very high, and confusion exists between these five categories. An explanation for the switch between the methodology and hypothesis categories could be the similarities between the two discourse elements. In [Waard & Maat 2012], the authors show that changing the verb tense of a hypothesis to the past tense makes the hypothesis appear to be a result. Even though in the present annotation process no verb tense was changed, the results show some similarities between the hypothesis and findings categories.

Survey in MTurk to study the reasons for worker participation as well as their weekly revenue, the time spent, and demographic information, such as country of origin, gender, age, education level, and household income. The authors found that workers were mostly from the US or India. Workers used MTurk as a second source of income or as a hobby. The workers reported that they used MTurk as a way to spend their free time fruitfully rather than watching TV. According to the report of [Ipeirotis 2010], only 50% of workers had a college education.
Using workers from Amazon Mechanical Turk appears to be more difficult for a task that requires some comprehension of the domain of interest and a large knowledge base regarding the scientific literature in general. Regarding the ratio between the speed and the number of error, Amazon Mechanical Turk and other crowdsourcing platforms may not be worth the cost. On average, MTurk works faster than experts (when finishing a batch), but the number of classification errors is too high. To provide the same platforms as MTurk provides to experts may increase the time needed by an expert to annotate a sentence, but it is not certain that even experts in the field of research of the document would be faster than scientists in a domain of research other than that of the document they have to annotate. In this test, we also provide only to annotators the possibility to classify the sentence in only one category, but as the analyse of the scientific documents shows, and [Teufel & Moens 2002, Ibekwe-SanJuan 2006] observe, a sentence can be classify into several categories. We simplify the test to try to avoid the confusion for the annotators of several categories, but in a real case, this possibility has to be taken into account.

The number of documents annotated should be sufficient to 1) test the feasibility of the annotation model and 2) evaluate the annotation with "real users". Whether for testing the feasibility of the annotation model the number of manual annotated documents is sufficient, for the evaluation with "real users", the number of manually annotated document is not enough. It is for this main reason we also test to automatically annotate document.

4.4 Automatic annotation

4.4.1 Introduction

The literature reveals a progression from the first complex annotation models, in which the aim was for authors to manually annotate their own articles, to less complex annotation models, in which the aim was to use Natural Language Processing (NLP) or machine learning systems to automatically annotate documents. Defining the level of complexity of a document model is not an easy task. It is interesting to access the maximum amount of information about the content of a document because the more precise the information is, the more information can be found or inferred. However, the greater the complexity of the model used, the more complex the annotation process will become. Model complexity is an issue in the three approaches used for annotation (manual by authors, crowd-sourcing and automatic). For the manual approach, the complexity of the models may lead to low consistency of annotation between different authors or even between different articles written by the same author. Authors must fully understand all components of the model used. A user guide containing a large amount of information (see chapter 3) and some training are proposed. Moreover, the annotation process is very time consuming. The same problem arises when documents are annotated using crowdsourcing, as demonstrated in the previous section. The use of a stylesheet in \LaTeX, such as in
the SALT project or [Waar & Kircz 2008], could be a good solution for the writer but may create additional difficulties, such as a lack of familiarity with \LaTeX. Even if manual annotation by the author were adopted, it would not be sufficient because it does not integrate all of the existing literature. Further, in [Divoli et al. 2012], the authors compare the difference between an abstract written by the authors of a scientific document and the content of sentences citing the aforementioned document written by a peer. They observe that authors in their abstracts and peers in their referenced sentence of same document do not focus on the same information about the document. This type of observation could be a indication that the "best" annotator for a scientific document may not be the author but someone else.

Another solution is to automatically annotate scientific documents. In the literature, we can find different solutions; models are generally less complex and complete than those based on manual annotation. A consequence of simplifying the models, which is necessary when using automatic annotation, is clearly a decrease in the complexity of information.

To assess the relevance of the SciAnnotDoc model, we must evaluate the extent to which it is possible to automatically annotate documents and evaluate the system with end-users. We require an annotated corpus that is large enough to compare the set of documents indexed with SciAnnotDoc against the same set of documents indexed only with keywords.

In the literature, there are three types of systems for the automatic annotation or classification of text documents. The first type is rule based. This type of system is based on the detection of general patterns in sentences [Groza et al. 2010, Tutin et al. 2009, Sándor & Vorndran 2009]. Several systems are freely available, such as XIP\textsuperscript{5}, EXCOM\textsuperscript{6}, and GATE\textsuperscript{7}. The second system is based on machine learning and requires a training corpus, such as the systems described by [Park & Blake 2012, Teufel 1999, Ruch et al. 2007, Liakata et al. 2012]. Several classifiers are available, such as the Stanford Classifier\textsuperscript{8}, Weka\textsuperscript{9}, and Mahout\textsuperscript{10}, based on different algorithms (Decision Trees, Neural Networks, Naïve Bayes, etc.,\textsuperscript{11}). The third system is a hybrid between the two aforementioned systems [Ou & Kho 2008] (see the previous chapter 3 for a more complete review of literature on the automatic annotation for scientific documents).

In this thesis, we opted for a rule-based system because we did not have a training corpus\textsuperscript{12}. Because documents in the human sciences are generally less formalised than in other domains, it may be difficult to have sufficient features with which to distinguish the different categories. Several semantic annotation tools exist. In

\textsuperscript{5}https://open.xerox.com/Services/XIPParser
\textsuperscript{6}http://www.excom.fr
\textsuperscript{7}http://gate.ac.uk
\textsuperscript{8}http://nlp.stanford.edu/software/classifier.shtml
\textsuperscript{9}http://www.cs.waikato.ac.nz/ml/weka/
\textsuperscript{10}https://mahout.apache.org/users/basics/algorithms.html
\textsuperscript{11}see [Kotsiantis 2007] for a complete review of the different algorithms
\textsuperscript{12}The number of sentences annotated by the author of this thesis, the scientists and the workers was not sufficient to create a training corpus.
[Uren et al. 2006], the authors provide an inventory (this survey is slightly dated, and not all tools proposed are still available). We attempted to implement some of these tools, but they were not flexible enough and did not offer the ability to create rules associated with parts of speech. We reviewed other scientific document models and studies that used rule-based systems and tried XIP and GATE. Because GATE is used by a very large community and several plug-ins are available, we ultimately opted to use this system.

### 4.4.2 Implementation

#### 4.4.2.1 Syntactic patterns

To examine the syntactic patterns of the discourse elements, we extracted sentences that corresponded to the different discourse elements from scientific documents in two areas of research: computer science and gender studies. The selected documents were very heterogeneous, ranging from empirical studies to more philosophical and/or discursive studies. To obtain the syntactic structures of these sentences, we analysed these sentences produced by GATE using the components included in ANNIE (in GATE). ANNIE is composed of different components, such as a sentence splitter and a part-of-speech tagger. We create a pipeline in GATE that was composed of the ANNIE Tokeniser component, the ANNIE sentence splitter component and the ANNIE part-of-speech component. Some modifications were made for the sentence splitter to take into account special cases of reference patterns. A brief example is provided in the following XML extract (the XML tags were simplified for readability):

```xml
<?xml version="1.0" encoding="UTF-8"?>
<GateAnnotation>
  <Token category="DT" string="The">The</Token>
  <Token category="NN" string="process">process</Token>
  <Token category="IN" string="of">of</Token>
  <Token category="VBG" string="adding">adding</Token>
  <Token category="NN" string="metadata">metadata</Token>
  <Token category="TO" string="to">to</Token>
  <Token category="NNP" string="Web">Web</Token>
  <Token category="NNS" string="pages">pages</Token>
  <Token category="VBZ" string="is">is</Token>
  <Token category="VBN" string="called">called</Token>
  <Token category="JJ" string="semantic">semantic</Token>
  <Token category="NN" string="annotation">annotation</Token>
  <Token category="," kind="punctuation" string=",”>
  <Token category="PRP" string="it">it</Token>
  <Token category="VBZ" string="involves">involves</Token>
  <Token category="DT" string="the">the</Token>
  <Token category="NN" string="decoration">decoration</Token>
  <Token category="IN" string="of">of</Token>
  <Token category="VBG" string="existing">existing</Token>
  <Token category="NNS" string="data">data</Token>
</GateAnnotation>
```
4.4. Automatic annotation

By analysing the tagged sentences, we created JAPE\textsuperscript{13} rules (20 rules for findings, 34 rules for definitions, 11 rules for hypothesis and 19 rules for methodologies\textsuperscript{14} and 10 rules for the referenced sentences). The methodology used to create the rules was simple. First, we created rules by precisely following the tags given by the part-of-speech tagger for each type of discourse element. Second, we simplified those rules and merged some of them to obtain more generic rules. Between each iteration, we ran these rules in GATE to evaluate them and to ensure that, by generalising them, the rules did not tag all the sentences. To simplify these rules, we also defined macros. It is possible to create a macro in JAPE to create patterns that can be used numerous times in grammar. (NOUN) is one of these patterns; it groups all the tags that concern nouns. The following is an example of a simple rule for a definition in JAPE\textsuperscript{15}:

\[
(\{\text{Token.kind==word, Token.category == "JJ"}\} \\
\text{(NOUN)} \\
\{\text{Token.kind==word, Token.category == "WDT"}\} \\
\{\text{Token.kind==word, Token.category == "MD"}\} \\
\{\text{Token.kind==word, Token.category == "VB"}\} \\
\{\text{Token.kind==word, Token.category == "JJ"}\} \\
\{\text{Token.kind==word, Token.category == "IN"}\} \\
\{\text{Token.kind==word, Token.category == "JJ"}\} \\
\text{(NOUN)} \\
)\]

To be more precise, we also added typical terms that appear in each type of discourse element. For example, the term paper followed, at a short distance, by the verb show likely indicates a finding, or the term "result" followed at a short distance by the term "consistent" likely indicates a finding.

"This result would be consistent with research showing that individuals are more prone to cognitive biases that are self-serving (Markus and Wurf 1987)"\textsuperscript{16}.[Correll 2004, p.108]

\[
(\text{(RESULT)} \\
\text{(RESULT)}[0,2] \\
\text{(CONSISTENT)} \\
\{\text{Token.kind==word, Token.category==IN}\} \\
\text{(NOUN | ADJECTIVE)} \\
)\]

\textsuperscript{13}http://gate.ac.uk/sale/tao/splitch8.html#x12-2130008
\textsuperscript{14}see Appendix G for the Jape rules
\textsuperscript{15}the definition of the part-of-speech tags can be found at http://gate.ac.uk/sale/tao/splitap7.html#x39-789000G
The following sentence is an example for a JAPE detection cue for a definition where the term "refer" appear.

"On this usage, gender is typically thought to refer to personality traits and behavior in distinction from the body." [Nicholson 1994, p.79]

(NOUN)
(VERBE)
{Token.kind==word,Token.category==RB}
{Token.kind==word,Token.category==VBN}
({Token})?
{Token.kind==word,Token.category==TO}
(REFER)
{Token.kind==word,Token.category==TO}

Because the syntactic structure of a definition might be very generic, we also uploaded the domain concept ontology to make it possible to tag sentences such as:

"Gender is the range of characteristics pertaining to, and differentiating between, masculinity and femininity"[Wikipedia 2014a].

This sentence produces part-of-speech tags such as

category= NNP (proper noun)
category= VBZ (verb - 3rd person singular present)
category= DT (determiner)
category= NN (noun - singular or mass)
category= IN (preposition or subordinating conjunction)
category= NNS (noun - plural)
...

This syntactic structure can be found in different discourse elements, not only in definition. However, this is also a very common way to define a term. A solution to tagging this type of sentence without adding too much noise to the annotation consists of using the domain ontology as a filter for the first term (NNP).

({Lookup.classURI=="http://cui.unige.ch/~deribauh/genderStudies#concept"})
(PUNCT)?
(VERBE)
(DETERMINER)
)

To detect the referenced sentences, 10 rules were created. The following is an example of one of these referenced rules. This rules can detect pattern for references such as: Hélène de Ribaupierre et al. 2004)
4.4. Automatic annotation

4.4.2.2 Automated annotation

To simplify automatic annotation, we restricted each discourse element to one sentence and each fragment to one paragraph. We automatically annotated 1,410 documents in English from various journals in gender and sociological studies (see Table 4.15). The full process is illustrated in Figure 4.2.

Table 4.15: Number of articles by journals

<table>
<thead>
<tr>
<th>Journal Name</th>
<th>Number of documents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender and Society</td>
<td>202</td>
</tr>
<tr>
<td>Feminist Studies</td>
<td>457</td>
</tr>
<tr>
<td>Gender issues</td>
<td>60</td>
</tr>
<tr>
<td>Signs</td>
<td>144</td>
</tr>
<tr>
<td>American Historical Review</td>
<td>15</td>
</tr>
<tr>
<td>American Journal Of Sociology</td>
<td>322</td>
</tr>
<tr>
<td>Feminist economist</td>
<td>214</td>
</tr>
<tr>
<td>Total</td>
<td>1414</td>
</tr>
</tbody>
</table>

The first step consisted of transforming a PDF file into a raw text. PDF is the most frequently used format to publish scientific documents, but it is not the most convenient one to transform into raw text. The JAVA program implemented to transform the PDF into raw text used the PDFbox\(^\text{16}\) API and regular expressions to clean the raw text.

\(^{16}\)https://pdfbox.apache.org/
Figure 4.2: Automated annotation process

PDF \rightarrow PDFBox \rightarrow Text
\downarrow
Annie + JAPE rules

PDF representation \leftarrow Java Application
\downarrow
AllegroGraph import

RDF triple store

Discourse Elements (XML)
4.4. Automatic annotation

Second, we applied the GATE pipeline (see figure 4.3) defined using ANNIE (see above) and the rules in the defined JAPE and ontologies\(^\text{17}\) modules to detect the different types of discourse elements in documents as well as the domain concepts (in XML). We imported the different ontologies that we created\(^\text{18}\): the gender studies ontology (GenStud), the scientific ontology (SciObj) and the methodology ontology (SciMeth).

Figure 4.3: Gate Pipeline

The ontologies were used not only for the JAPE rules but also to annotate the concepts in the text. The OntoRoot Gazetteer will 1) tokenises each term; 2) assign part-of-speech; 3) applies Morpher and 4) indicates lemma information to each token. Consequently, plural as singular concepts are annotated. A XML tag different was assigned to each sentence of each discourse element type and for each domain concept and referenced authors. Below is a small extract of a scientific document tagged by the obtained GATE pipeline:

```xml
<?xml version="1.0" encoding="UTF-8"?>
<GateAnnotation>
  <findingsSentence rule="findingsSentence">
    As a result of these different processes,
    <concept propertyValue="men" class="http://../gendStud#men" URI="http://../gendStud#men">men</concept>
  </findingsSentence>
</GateAnnotation>
```

\(^{17}\)Ontology OWLIM2, OntoRoot Gazetteer

\(^{18}\)For the description of the ontologies, readers can refer to the chapter 3
achieve an ego structure that defines itself as more separate and distinct, with a greater sense of rigid ego boundaries.

Third, we implemented a JAVA application (see figure 4.4) using the OWL API to transform the GATE’s XML files into an RDF representation of the text. Each XML tag corresponding to concept or object properties in the ontology were transformed. The sentences that did not contain one of the four discourse elements (definition, hypothesis, finding or methodology) were annotated with the tag `<Non-DefinedDE>`, allowing for the annotation of each sentence of the entire document, even those not assigned to discourse elements. For the creation of the referenced document, each time that a tag `authorCite` was found, the program created a new document with the author’s name and the year of publication. The author’s name was created if it did not already exist. This process was the same as that applied for the author’s name of the annotated document.

Figure 4.4: Annotation algorithm model

For the following sentence, five documents were created:

"<authorCite>Cherlin and Walters (1981)</authorCite> and <authorCite>Ransford and Miller (1983)</authorCite>, for example, examined attitudes about the suitability and appropriateness of women’s presence in the public worlds of work and politics, while <authorCite>Porter Gump (1975)</authorCite>, <authorCite>Randon Hershey (1978)</authorCite>, and <authorCite>Hermons (1980)</authorCite> measured women’s identification with the values of "femininity" and the private roles of wife and mother". [Dugger 1988, p.429]
4.4. Automatic annotation

Because some properties of the bibliographic data were more difficult to automatically discriminate than others, it was not possible to have a completely automatic process. Some properties, such as journal name or the year of publication, were manually assigned. For the other properties, some regular expressions were used to find the authors, the title, and, when possible, the year of publication. These difficulties explain why we annotated 1,400 documents.

The different RDF representations created by the JAVA application were loaded into an RDF triple store. We chose Allegrograph\textsuperscript{19} because it supports RDFS++ reasoning in addition to SPARQL query execution.

Even after applying a cleaning process during the transformation from PDF to text, some documents still contained noise. By counting the number of paragraphs (fragments) in a document, we deleted from the triple store all documents that contained more than 450 paragraphs.

From Allegrograph, we created new triples, transforming the rules presented in chapter 3 section 3.3.2.5 into SPARQL queries. For the "Lineage of Ideas" (see Figure 4.5), relations between discourse elements of the same type and documents that cite the same author/discourse element but do not cite each other and describe the same concepts, we inferred 1 relationship between 2 definitions, 38 relationships between 74 findings, 3 relationships between 6 hypotheses, and 3 relationships between methodologies (see figure 4.6).

Figure 4.5: Inference for the "Lineage of Idea"

\[\text{http://franz.com/agraph/allegrograph/}\]
Figure 4.6: An example for the inference "lineage of Idea"
4.4. Automatic annotation

4.4.3 Corpus statistics

In the following section, certain statistical results are presented. The possibilities for analysing the corpus are endless; thus, we present only a few. The table 4.16 presents the statistics concerning the number of fragments, sentences and the coverage by journals. The table 4.17 presents the distribution of the discourse elements by journal. We can observe that the number of referenced documents is greater than the number of related works. This result is observed because authors generally refer to several documents simultaneously rather than to a single one.

Table 4.16: Annotated corpus statistics

<table>
<thead>
<tr>
<th>Journal Name</th>
<th>No fragments by journal</th>
<th>No sentences</th>
<th>No discourse elements</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender and Society</td>
<td>17638</td>
<td>63832</td>
<td>6368</td>
<td>9.98%</td>
</tr>
<tr>
<td>Feminist Studies</td>
<td>31627</td>
<td>135989</td>
<td>12394</td>
<td>9.11%</td>
</tr>
<tr>
<td>Gender issues</td>
<td>6042</td>
<td>20412</td>
<td>2397</td>
<td>11.74%</td>
</tr>
<tr>
<td>Signs</td>
<td>14310</td>
<td>43218</td>
<td>4258</td>
<td>9.85%</td>
</tr>
<tr>
<td>American Historical Review</td>
<td>1435</td>
<td>6284</td>
<td>571</td>
<td>9.09%</td>
</tr>
<tr>
<td>American Journal Of Sociology</td>
<td>40511</td>
<td>167268</td>
<td>22372</td>
<td>13.37%</td>
</tr>
<tr>
<td>Feminist economist</td>
<td>30610</td>
<td>101129</td>
<td>14159</td>
<td>14.00%</td>
</tr>
<tr>
<td>Total</td>
<td>142173</td>
<td>538132</td>
<td>62519</td>
<td>11.62%</td>
</tr>
</tbody>
</table>

Table 4.17: Annotated corpus statistics by Discourse elements

<table>
<thead>
<tr>
<th>Journal Name</th>
<th>Def.</th>
<th>Find.</th>
<th>Hypo.</th>
<th>Meth.</th>
<th>Related Work</th>
<th>Referenced documents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender and Society</td>
<td>745</td>
<td>2945</td>
<td>1021</td>
<td>1742</td>
<td>986</td>
<td>4855</td>
</tr>
<tr>
<td>Feminist Studies</td>
<td>2201</td>
<td>4091</td>
<td>2545</td>
<td>3600</td>
<td>177</td>
<td>5377</td>
</tr>
<tr>
<td>Gender Issues</td>
<td>280</td>
<td>1126</td>
<td>414</td>
<td>611</td>
<td>267</td>
<td>1566</td>
</tr>
<tr>
<td>Signs</td>
<td>789</td>
<td>1566</td>
<td>712</td>
<td>1221</td>
<td>516</td>
<td>3129</td>
</tr>
<tr>
<td>American Historical Review</td>
<td>97</td>
<td>219</td>
<td>87</td>
<td>170</td>
<td>15</td>
<td>440</td>
</tr>
<tr>
<td>American Journal Of Sociology</td>
<td>1776</td>
<td>10160</td>
<td>4316</td>
<td>6742</td>
<td>2907</td>
<td>13323</td>
</tr>
<tr>
<td>Feminist economist</td>
<td>1381</td>
<td>6940</td>
<td>2025</td>
<td>4169</td>
<td>2288</td>
<td>9600</td>
</tr>
<tr>
<td>Total</td>
<td>7269</td>
<td>27047</td>
<td>11120</td>
<td>18315</td>
<td>7156</td>
<td>38290</td>
</tr>
</tbody>
</table>
Some discourse elements appear more frequently than others (see Figure 4.7), as is the case in manual annotation. In addition, the distribution changes depending on the journal. It is interesting to note that the percentage of methodology elements remains stable.

Figure 4.7: Percentage discourse elements by journals (1= Gender and Society; 2= Feminist Studies; 3=Gender Issues; 4=Signs; 5=American Historical Review; 6= American Journal Of Sociology; 7=Feminist economist)

Figure 4.8 reports the distribution of concepts in the corpus and the "long tail". The average number of concepts by sentence is four. The distribution follow Zip's law, as we would expect. In the Figure 4.9, we extracted the number of times the concepts "sex", "gender" and "discrimination" were evoked by year through the corpus (normalised by the total number of concepts by year). These terms seem to have been more popular in some years than in others. The term "gender" seems to be used more than "sex", which was used more often previously. Of course, these observations are not generalisable because of the low representativeness of the scientific documents compared to the entire lot of scientific documents in gender studies.
4.4. Automatic annotation

Figure 4.8: Distribution of the concepts in the corpus

Figure 4.9: Number of times the concepts "gender", "sex" or "discrimination" are evoked by year (normalised by the number of concepts by year)
4.4.4 Annotation evaluation

To test the quality of patterns, we uploaded 555 manually annotated sentences that constitute our gold standard into GATE and processed the same sentences used to test the JAPE rules. These sentences were extracted from the manually annotated corpus. We did not use any of the sentences analysed to create the JAPE rules to construct the gold standard to avoid bias. We performed measurements of precision and recall on these sentences (see Table 4.18). The results indicated good precision but low recall.

Table 4.18: Precision/recall values

<table>
<thead>
<tr>
<th>Discourse element type</th>
<th>No of sentences</th>
<th>Precision</th>
<th>Recall</th>
<th>F1.0s</th>
</tr>
</thead>
<tbody>
<tr>
<td>Findings</td>
<td>168</td>
<td>0.82</td>
<td>0.39</td>
<td>0.53</td>
</tr>
<tr>
<td>Hypothesis</td>
<td>104</td>
<td>0.62</td>
<td>0.29</td>
<td>0.39</td>
</tr>
<tr>
<td>Definitions</td>
<td>111</td>
<td>0.80</td>
<td>0.32</td>
<td>0.46</td>
</tr>
<tr>
<td>Methodology</td>
<td>172</td>
<td>0.83</td>
<td>0.46</td>
<td>0.59</td>
</tr>
</tbody>
</table>

To increase recall, we added heuristics such as the following: if a fragment (paragraph) contains unrecognised elements (sentences) and at least three elements with the same type $T$, then assign type $T$ to the unrecognised elements. With these rules, we created 341 findings, 130 methodologies, 29 hypothesis and 6 additional definitions. Nevertheless, the coverage was significantly lower than that encountered in the manual annotation due to a very conservative automatic annotation.

4.4.5 Discussion

In this section, we have presented the automated annotation process. The SciAnnotDoc model operates with an automated annotation process; however, it was necessary to simplify the original SciAnnotDoc model to some extent to make it compatible with an automatic annotation system. It is only possible (at the moment) to link from the discourse elements to the documents, not from the discourse elements to other discourse elements, that is, within document relations. In addition, it is not possible to distinguish between the different types of CiTO reference relationships; therefore, we used the generic cites. To improve the detection of the type of citations between documents, sentiment analysis approaches can be used. To improve the granularity of the annotation of cited documents, a first step is to have the different documents that cite or are cited by one of the annotated documents to detect the paragraph or the sentence to which the sentence of the citing document refers and to analyse the cited document using the same processes. Using API such as Microsoft Academic Search or CiteSeer to gather more precise information on bibliographic data for the annotated documents and the referenced documents could help to improve this detection.
4.5 Conclusion

For definitions, we were not able to distinguish the definiendum from the definition. Currently, only the definitions and the concepts that are described in the definitions are annotated. In other words, it is only possible to query a definition with the term that uses the definition, not the definiendum directly. By using more advanced Natural Language Processing such as the detection of the subject in the sentence, the detection of the defined term should be improved.

For the findings, no distinction was made between the types of findings, and all the findings were annotated with the generic *findings* concept. To improve this detection, one approach would be to define more JAPE rules. The structural part of the document was also not annotated.

To improve the recall index, the first solution was to use heuristics to annotate non-defined sentences as one of the discourse element types, as we propose above. Another solution could be to create more JAPE rules. However, introducing a larger number of rules might also increase the risk of adding noise to the annotation. Another solution could be to create a training corpus and use a machine-learning system instead of a rule-based system.

4.5 Conclusion

In this chapter, we have discussed the advantages and the disadvantages of manual, crowdsourcing and automatic annotation processes. The manual annotation process allows the user to have the most expressive model. The fine subtleties of the semantics in each discourse element can be annotated, but the process is very time consuming. This type of annotation also requires users to be very familiar with a field. Thus, to attain a high level of inter-judge agreement, annotators may need some training, even if they are "experts" in the research domain.

Crowdsourcing is a very powerful way to harness the intelligence of a mass of people, but this intelligence has some limitations. As observed, although the level of inter-judge agreement was high between the workers of Amazon Mechanical Turk, the number of errors committed was significant. We can also note that the batch was rarely finished by any annotator; to have a completed project, it would be necessary to either provide a greater incentive or propose a smaller batch of sentences.

Automated annotation allows the user to have a large number of documents annotated quickly, and all the documents can be annotated in a similar manner. However, the automation also presents certain limitations, as we have discussed above, particularly in terms of precision and recall as well as in terms of the level of detail for the annotation. To be able to automate the processes, we had to simplify the annotation model. However, as mentioned above, the tool can be improved using different approaches.
Chapter 5

Query model and Faceted search interface

Contents

5.1 Introduction .......................................................... 131
5.2 Query model ......................................................... 132
  5.2.1 Simple queries .................................................. 135
  5.2.2 Queries with multiple elements ............................. 136
  5.2.3 Examples of queries in SPARQL extracted from interviews. 138
5.3 Faceted search on the SciAnnotDoc (FSAD) .................. 140
  5.3.1 Meta-data facet or dimension ................................. 142
  5.3.2 Conceptual facet/dimension ................................. 142
  5.3.3 Discursive facet/dimension .................................. 143
  5.3.4 Relationships facet/dimension .............................. 144
  5.3.5 Result’s visualisation ......................................... 145
5.4 Conclusion ............................................................ 145

5.1 Introduction

In this chapter, we present the query model and the user interface we have created to query the graph that represents the annotated documents uploaded in the Allegrograph triple store (1410 scientific documents). The construction of the query model as well as the user interface were based on 1) the SciAnnotDoc model and its different components and 2) the precise questions indicated by the scientists (see Chapter 2). This set of "precise questions"\(^1\) served as the basis for determining how the scientists wanted to find their information in the model. The implementation of the query model and the user interface were intrinsically linked.

In existing scientific information retrieval systems (SIRSs), the query model is generally based on keywords, which are sometimes mixed with facets such as author

\(^1\)In this thesis, we define the term "precise question" as questions that could resolve complex problems or conduct deep analysis for a given problem.
name, article title, journal name or year of publication. In SIRSs with more advanced search schemes (PubMed\(^2\), CiteSeer\(^3\)), it is possible to search for keywords in certain structural parts, such as the abstract. In the present thesis, the aim is to search for information contextualised in a discourse element of documents. In [Salton et al. 1993], the authors argue that users do not want to see entire documents when they are searching for information; rather, they want to concentrate on the most immediately relevant passages of text. Current studies in the field of focused retrieval have the same aim as that of Salton et al.: they focus their set of answers on the passages that are the most relevant to the user. This field of research encompasses a variety of studies ranging from structured document retrieval to passage retrieval, structured text retrieval, querying semi-structured data, and XML retrieval [Trotman et al. 2007, Kamps et al. 2008, Lalmas 2009].

Faceted-search interfaces represent an interaction paradigm that organises the information provided by search engines according to a faceted classification or faceted dimensions. A review of the literature concerning this interaction paradigm is beyond the scope of this thesis. If the reader is interested in learning more about faceted-search interfaces, a more complete review of such studies can be found in [Yee et al. 2003, Hearst 2006, Broughton 2006, Ben-Yitzhak et al. 2008, Tunkelang 2009, Wei et al. 2013]. We chose to use an adaptive-faceted search interface in the present thesis because this technique allows users to navigate intuitively through an important set of information [Kules et al. 2009, Bennett et al. 2014], and it is very well adapted to scientific documents and to RDF navigating data [Oren et al. 2006]. In addition, the facets can be easily described using the SciAnnotDoc model.

The query model of the present thesis as well as the user interface search were built on the principle that scientists want to focus initially only on a particular part of documents, especially when several documents are displayed. Only in a second phase, and if a document seems interesting, may scientists want to read entire documents.

5.2 Query model

The query model that we defined (see Figure 5.1) is based on the discourse elements of the SciAnnotModel. Each basic query may be expressed in terms of discourse elements and other search criteria. The query model is based on the idea that it is not the entire document that interests users but only smalls chunks of text—in this case, discourse elements (definition, finding, hypothesis, methodology and related work). Following the conclusions of our study on how scientists search for information, the set of responses provided will be discourse elements (sentences) and not the entire document. The discourse element is the core of the SciAnnotDoc query model.

The structure of a query is formally defined in the form of a UML class dia-

\(^2\)http://www.ncbi.nlm.nih.gov/pubmed/advanced
\(^3\)http://citeseerx.ist.psu.edu/advanced_search
gram in figure 5.1. In this query model, a query specifies a number of conditions that a discourse element must satisfy. A condition may specify the type(s) of the element (TypeSpecification), the citation(s) it must contain (CitationSpecification), the scientific object concepts, the methodological concepts and the domain concepts that refer to (ScientificObjectConceptSpecification, MethodologyConceptSpecification, DomainConceptSpecification). When a condition specification has more than one component (e.g. several element types), they are connected by a logical operator (AND or OR). In addition, a query may specify metadata conditions that must be satisfied by the document that contains the discourse element.

To express the semantics of the query model, we will show how queries are transformed into SPARQL queries on the RDF graph of annotated documents.

We can distinguish between two families of queries. The first family, simple queries, only asks for one discourse element. A simple query represents, for example, all the findings describing the term X. The second family of queries is complex queries. A complex query contains a minimum of two distinct elements. An example of a complex query is the findings describing the term X and a hypothesis describing the term Y, where X and Y are not in the same element (sentence) and the two elements are in the same document.
Figure 5.1: Query model
5.2. Query model

5.2.1 Simple queries

A simple query contains only one discourse element specification. Figure 5.2 shows an instance of such a query. The SPARQL representation of such a query is a straightforward transformation of the query graph into a SPARQL query graph, replacing the UML associations with the appropriate RDF properties of the SciDocAnnot schema. When a logical operator is a disjunction, the SPARQL query is slightly more complex because it must use the UNION construct. Figures 5.3 and 5.4 show the SPARQL representation of the query of figure 5.2 for the conjunctive ($\theta = AND$) and disjunctive ($\theta = OR$) cases.

The answer set is thus the set of elements that satisfy the specified conditions and that are in a document that satisfies the metadata conditions.

![Figure 5.2: A instance of the query model (simple query)](image)

![Figure 5.3: Transformation of a simple query in SPARQL with an AND between the discourse elements](image)

The figure 5.4 is a SPARQL representation of a simple query that describes a disjunction (the diamond) between two types of discourse elements.
5.2.2 Queries with multiple elements

A query may have one main discourse element specification E and several other discourse element specifications F1, F2, ..., Fn (see figure 5.5). In this case, the intention is to select all the elements that satisfy E and that are in a document containing elements satisfying F1, F2, ..., Fn. A query with multiple elements is necessary, for instance, to express the information need "Retrieve the findings that refer to concept X and that are in a document containing a hypothesis about concept Y".

The SPARQL representation of this type of query is obtained by taking the SPARQL representation of each discourse element specification and connecting them to a node representing the common document to which they belong, as shown in figure 5.6.
5.2. Query model

Figure 5.6: Transformation of a complex query in SPARQL with an AND between the discourse elements

\[
\begin{align*}
?doc & \quad \text{in} \quad ?frag \\
?doc & \quad \text{in} \quad ?frag_n
\end{align*}
\]

\[
\begin{align*}
?e & \quad \text{type} \quad t_1 \\
?e_n & \quad \text{type} \quad t_2
\end{align*}
\]

5.2.3 Examples of queries in SPARQL extracted from interviews

To test the query model, each specific question from the interviews was transformed into one query (or several queries when necessary). In addition, the questions that did not pertain to gender studies because the researchers were specialists in other fields were generalised before the transformation into queries. We present a few examples of these transformations, where Q1 and Q2 correspond to a simple query and Q3 corresponds to a complex query. We tested the different queries in SPARQL on the annotated corpus uploaded in the triple store (Allegrograph).

**Precise question** Gender, historicity of the concept, its various definitions and how it evolves and anchors culturally. Highlight the main controversy

- Q1 Find all the definitions of the term gender
- Q2 Find all the definitions of the term that cites author Y

**Precise question** Intra-individual variability in terms of the brain and behaviour. The method of analysis, which included the indices of variability, was generalised and transformed as follows:

- Q3 Obtaining all the findings that describe X and are in a document using methodology Y with statistical test Z

The representation of Q1 in SPARQL.

Figure 5.7: Representation of Q1 in SPARQL.
SELECT DISTINCT *
{
?concept gendStud:term "gender".
}

The representation of Q2 in SPARQL.

Figure 5.8: The representation of Q2 in SPARQL

SELECT DISTINCT *
{
?concept gendStud:term "gender".
?e1 cito:cites ?citesX.
?authorX gendStud:name "butler"
}

The representation of Q3 in SPARQL.

SELECT DISTINCT *
{
{
?concept gendStud:term "academic".
}
5.3 Faceted search on the SciAnnotDoc (FSAD)

In this section, we will describe a faceted search in SciAnnotDoc (FSAD), a faceted search interface we developed for scientists to query the annotated scientific documents uploaded in Allegrograph. Because scientists are not all experts in description logic or other formal query languages (such as SPARQL), we propose a faceted search interface based on SciAnnotDoc. A major goal of the interface is to hide the complexity of semantic search from scientists and to make it easy to use and effective.

The use of a faceted-search interface to mask the complexity of SPARQL has the disadvantage of being less expressive than other query languages [Ferré et al. 2011]. However, it is still possible to have sufficient expressiveness to allow end users to find their desired information. We represent the different dimensions of the SciAnnotDoc model (see Chapter 3 section 3.3) as facets. Thus, the resulting interface is a faceted-search interface in the sense that the SciAnnotDoc dimensions structure the information that users can explore. The interface is also an adaptive search
interface in the sense that users should be able to add as many discourse elements
as they desire to build "complex" queries. In the implemented interface, users are
only able to query two disjoint discourse elements due to some implementation prob-
lems. The interface was built using JAVA and SWING. The query model classes
were implemented using the OWL API and the Allegrograph API.

FSAD was built in several iterations. During the different cycles of iteration,
scientists were asked twice to evaluate the usability interface through heuristic eval-
uations [Nielsen & Molich 1990]. We defined three tasks that scientists had to per-
form using the interface. We provided a small tutorial to explain how to perform a
query using the interface. We will detail only the latest version of the interface:
Figure 5.10: Adaptative-faceted search interface (with the four dimensions /facets)

- **Meta-Data dimension**
- **Discursive dimension**
- **Conceptual dimension**
- **Relationships dimension**
- **Result’s Visualisation**
5.3.1 Meta-data facet or dimension

The meta-data facet allows users to query the meta-data of the SciAnnotDoc model. Although the aim of this system is not to find a document, the meta-data can reduce the space of the set of results by, for example, specifying only part of a title or a word or by specifying a range of years, a single author’s name or part of an author’s name.

An example (extract only for meta-data) of a SPARQL query was generated automatically by the system. The semantics of the query involve a document that has a title that contains "sex", was published between 1980 and 2001 and was written by "cotter".

```
BIND (xsd:int(?year) as ?l)
FILTER (regex(str(?title), "sex", "i")
  && (?l=1980&&?l<=2001)
  &&regex(str(?name), "cotter", "i"))
```

5.3.2 Conceptual facet/dimension

The conceptual facet allows users to specify what term or terms (keywords) discourse elements describe. Due to a technical implementation problem, only the conjunction between terms is functional; however, the disjunction should also be functional.

To determine whether a user is using a conjunction, the user must enter "AND" between terms. The reason "AND" must be included is that in scientific vocabulary, it is often the case that a term is composed of more than one word; thus, it is essential that the system transforms the query of a user depending on the composition of the words used. For example, if users want to search for an element that contains "gender gap", they do not mean an element that contains "gender" or "gap". The semantics of the query are not the same; thus, the construction will not be similar.

First, the system built a SPARQL query by querying the gendStud ontology. A filter was used to allow the user to commit some mistakes in entering the term, such as typing a capital letter. If the term was not found, the "magical property" free text of Allegrograph was used, allowing the system to search in the index created by Allegrograph from the text contained in the comments of the discourse element. For scientific objects or uses of methodology, the interface directly displayed the different concepts extracted from the two ontologies (SciObj and SciMeth (see Chapter 3 section ?? and ?? ).

---

4 A magic property is a predicate in a SPARQL query that produces bindings using an approach other than simple subgraph matching. These extensions provide a much richer query environment at the cost of non-portability. AllegroGraph has long supported a magic property to enable free text queries and to interface to Solr and MongoDB. [http://franz.com/agraph/support/documentation/current/magic-properties.html](http://franz.com/agraph/support/documentation/current/magic-properties.html).
5.3. Faceted search on the SciAnnotDoc (FSAD)

An example (extracted only for the conceptual dimensions) of a SPARQL query was generated automatically by the system. The semantics of the query involve a single discourse element that describes the terms "academic" and "income".

```sparql
?concept0 gendStud:term ?con0.
FILTER (regex(str(?con0),"academic", "i"))
FILTER (regex(str(?con1),"income", "i"))
```

An example (extracted only for the discursive dimensions) of a SPARQL query was generated automatically by the system if the set of answers was empty. The semantics of the query involved a single discourse element that describes the terms "academic" and "income".

```sparql
?p fti:matchExpression '(and "academic" "income")'.
```

5.3.3 Discursive facet/dimension

The discursive facet allows users to specify in which discourse element they want to apply the other search criteria. As previously indicated, users can search for a discourse element by specifying one or more types describing the same discourse element (simple query). If users want to search only for a definition, they tick only one check-box for the first discourse element. If they want to search for a sentence that belongs to two discourse elements, such as a definition that is also a finding, they tick two check-boxes of the first discourse element.

An example (extracted only for the discursive dimensions) of a SPARQL query was generated automatically by the system when the user specified one check-box of the first box. The semantics of the SPARQL query involve a discourse element that is a "findings".

```sparql
```

An example (extracted only for the discursive dimensions) of a SPARQL query was generated for the case in which the user specified two check-boxes of the first discourse element. The system allows for a conjunction or a disjunction of the two discourse elements (the example shows a conjunction). The semantics of the SPARQL query involve a discourse element that is a definition and a finding.

```sparql
```
Users may want to search for two discourse elements disjointedly in the same text (complex query), such as findings describing X based on a hypothesis describing Y. Users must tick the "taking into account second element" and one (or more) checkboxes of the first discourse element box and another (or more) of the second discourse element box. For a simple query, the user can specify the concepts/keywords for each discourse element box. The same algorithm applies for the search of terms in the discourse elements; if no discourse element is found, the "free text" is used.

An example (extracted only for the discursive dimensions) of a SPARQL query was generated automatically by the system when the user specified two discourse elements disjointedly for a document. The system generated a sub-level query. The semantics of the SPARQL query involve a document that contains a fragment that has a discourse element that is a methodology and a different fragment that has a discourse element that is a definition.

```
{  
{  
SELECT DISTINCT ?f1 ?p1 ?doc WHERE  
{  
}  
}  
}
```

### 5.3.4 Relationships facet/dimension

The relationships facet allows users to indicate whether a discourse element cites another document by specifying the author’s name. Users can specify more than one author’s name by using "AND" between the two names. The disjunction was not implemented but should function.

An example (extracted only for the relationships dimensions) of a SPARQL query was generated automatically by the system when the user specified two cited author’s names. The semantics of the SPARQL query involve an element that cites "butler" and "cotter".

```
?f SciAnnotDoc:has_discourse_element ?p  
?p cito:cites ?cites0.  
?cites0 SciAnnotDoc:written_by ?authorC0.  
?authorC0 gendStud:name ?nameC0.  
FILTER (regex(str(?nameC0),"butler", "i"))  
```
5.3.5 Result’s visualisation

The visualisation of the set of results is a table of the discourse elements. The table is constructed such that the tuples are the results, and the columns are (in order) year of publication, the author’s names, the sentence linked to the discourse element (that is contained in the rdfs:comment) and the text of the paragraph contained in the fragment. The paragraph allows users to contextualise the information given in the discourse element.

Each column of the table can be used to sort the set of results.

For a complex query, the interface only displays the first discourse element, taking the second one as a criterion of research for the first one. In other words, if users want to search for findings describing the term X and for a methodology in the same document describing the term Y, the first element should contain the findings and the second element should be a methodology. In contrast, if users want to search for the methodology, they will invert the query.

5.4 Conclusion

In this chapter, we presented a query model and the faceted-SciAnnotDoc-search interface, including the semantics of a query, the query expression that we applied on the graph that represents the annotated corpus, and the way the system built the query depending on the different characteristics the user entered in the interface. We tested the different queries created from the precise question extracted from the interviews, and—with the exception of queries such as "Authors X that disagree with author Y"—all the queries were possible (the aforementioned query is not yet possible, not because of the query model but because of the missing characteristics of the citation type in the annotation process). Thus, although the query model does not allow for the full expressiveness that SPARQL could provide, the defined query model is sufficiently flexible to be able to query an annotation to meet scientists’ information needs.

This interface was built with the aim of allowing for evaluation based on scientists’ perceptions regarding the annotated corpus. As will be discussed in the next chapter, it was possible for scientists without any knowledge of Semantic Web technology such as SPARQL to use the interface and query the triple store. Certain search criteria of the query model were not possible to implement, and we reduced certain capabilities of the query model, such as the disjunction of the term and the author’s name. We had initially intended to produce a more interactive view of the set of results, such as allowing a user to click on a discourse element and view the corresponding document or prompt the system to redefine the SPARQL request as.
a function of the author’s name if a user clicked on the name of an author. However, time did not allow for such capabilities to be incorporated into the interface. The way in which the data extracted from the annotated corpus can be visualised is important, such as creating a visualisation that helps the reader to read various documents in parallel or visualise the lineage of an idea. To build a good visualisation data, the nature of the needed information can also be a variable that could change the type of visualisation used. The user studies made in the beginning of this thesis can be a very strong base to build the visualisation data model. These concepts will constitute the basis of a future project (Post-Doc). However, even if FSAD could be more developed, this is sufficient to be use for the evaluation test with the "real users", as we will see in the next chapter.

At the end of the implementation, we realised that the choice of JAVA as the programming language was not the most judicious. One reason is that programming the interface was more computationally expensive than expected, and the architecture of the language made changing the interface difficult. Another reason concerns the evaluation of the system. Because this is a JAVA program, it is not possible to launch a large evaluation campaign on the Web as if it was a Web page. A third reason is the increasing number of existing APIs in JavaScript or PHP for visualising RDF data.
6.1 Introduction

In this chapter, we present the evaluation of the SciAnnotDoc model using the Faceted-SciAnnotDoc-search (FSAD) interface presented in the previous chapter. The aim of this study is to evaluate the effectiveness of the information retrieval process. We propose a new search model, and the evaluation focuses on the fact that scientists find this new search model interesting and highly effective.
Evaluation in information retrieval has a long tradition, starting with the Cranfield tests [Cleverdon 1967] and culminating in modern campaigns such as TREC\(^1\) and CLEF\(^2\). Relevance\(^3\) is one of the measures tested in these campaigns. However, these types of tests normally require a test collection to calculate the relevance level of the results. Unfortunately, a test collection of this kind to test the relevance of the SciAnnotDoc system results does not yet exist. Nevertheless, some conferences and challenges have been initiated in the domain of semantic publishing; for example, the Semantic Publishing Challenge began this year\(^4\), and some test collections appear to be on the verge of being created.

In the present evaluation, the relevance of the results will be tested through the subjective impressions of scientists, which were gathered using a survey. To compare the effectiveness of the system and the relevance of the set of results, we will compare the SciAnnotDoc system with a more "conventional" system based on a keyword search.


### 6.2 Methodology

We implemented a keyword-based search interface to evaluate the Faceted search on SciAnnotDoc search system (FSAD). Because Allegrograph allows "free text" indexing, we use the same triple store, and we index the text only by the keywords contained in sentences. The system queries at the sentence level, not at the document level. In other words, if participants search using the keywords "gender" and "definition", both terms must be in the same sentence. The FSAD interface and the keyword search interface display only sentences, not the entire document.

The design of the experiment was based on a Latin square rotation of tasks (see table 6.1) to control for a possible learning effect of the interface on the participants.

The objectives of the search tasks to be performed by the participants were defined by the author of this thesis and tested with two persons before the experiment using the two interfaces:

**Task 1** Find all the definitions of the term "feminism".

**Task 2** Show all findings of studies that have addressed the issue of gender inequality in academia.

**Task 3** Show all findings of studies that have addressed the issue of gender equality in terms of salary.

---

\(^1\)http://trec.nist.gov

\(^2\)http://www.clef-initiative.eu

\(^3\)TREC uses the following working definition of relevance: "If you were writing a report on the subject of the topic and would use the information contained in the document in the report, then the document is relevant". [TREC 2000]

\(^4\)http://challenges.2014.eswc-conferences.org/index.php/SemPub
6.2. Methodology

Figure 6.1: Keyword search interface

Table 6.1: Design experiment (Latin square rotation of tasks)

<table>
<thead>
<tr>
<th>Participants</th>
<th>System</th>
<th>Tasks passation order</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>keywords</td>
<td>T1,T3,T2</td>
</tr>
<tr>
<td>P2</td>
<td>FSAD</td>
<td>T1,T2,T3</td>
</tr>
<tr>
<td>P3</td>
<td>FSAD</td>
<td>T2,T1,T3</td>
</tr>
<tr>
<td>P4</td>
<td>Keywords</td>
<td>T2,T1,T3</td>
</tr>
<tr>
<td>P5</td>
<td>Keywords</td>
<td>T1,T2,T3</td>
</tr>
<tr>
<td>P6</td>
<td>FSAD</td>
<td>T1,T3,T2</td>
</tr>
<tr>
<td>P7</td>
<td>FSAD</td>
<td>T1,T3,T2</td>
</tr>
<tr>
<td>P8</td>
<td>keywords</td>
<td>T3,T2,T1</td>
</tr>
</tbody>
</table>

The corpus was the annotated corpus (1410 documents) in gender studies. All documents were in English. A short tutorial was presented to the participants, and the participants were allowed to ask questions to the moderator during the experiment. The moderator also provided some advice on the terms of query (the level of English of some participants was low, and they sometimes had difficulty finding terms). The experiment was not performed in a closed environment; except for two occasions, the experiment was conducted on the same computer in the office of the author of this thesis. Six participants were examined using the same computer (an iMac 27") in the office of the author of this thesis, and two were examined using a MacBook Pro 13" in the participants’ offices. The average time for the experiment was 1 hour.

The end of a task was determined by the participants, who decided whether they had obtained enough information on the given subject. During the experiment, the
moderator sat next to the participants and took note of their comments. A small program was written to capture the logs of the two systems. The logs recorded the time stamps indicating when the participants clicked "search", the SPARQL query, which elements were checked (in FSAD), the different search criteria, and the time stamp indicating when a query was returned by the system. We calculated the length of time during which a participant looked at a set of results based on the interval between the time stamp of a query’s end and the next click on the "search" button. Thus, the time spent looking at a set of results was approximate.

The experiment began with the presentation of a brief tutorial (on paper) to the participants. When necessary, the moderator responded to the participants’ questions. After reading the tutorial, the participants sat in front of the computer and first responded to a socio-demographic survey composed of 11 questions (age, sex, level of education, knowledge of search information and computer, knowledge of gender studies). Then, the moderator presented the interface to the participants and asked them to perform the first task (the order was pre-determined). When the participants indicated that they had gathered enough information, they were asked to complete a new questionnaire (10 questions, see below). This procedure (the performance of the task and the completion of the questionnaire) was repeated 2 more times. At the end of the last search task, the participants completed a final questionnaire concerning the system rather than the specific task (11 questions). The evaluation was performed in French. The questionnaires were conducted on LimeSurvey.

The social-demographic questionnaire was composed of the following items:

Q1 Year of birth
Q2 Sex
Q3 Domain of research
Q4 Level of education
Q5 Average number of scientific documents that you retrieve per week
Q6 Average number of scientific documents that you read per week
Q7 Do you consider yourself an expert in scientific document information retrieval? (1 = not an expert, 5 = expert)
Q8 Do you find it difficult to find the information you need in the documents you found? (1 = very easy, 5 = very difficult)
Q9 Which search engines do you typically use to search scientific literature? (if more than one, please put them in order of preference)
Q10 What is your level of knowledge in gender studies? (1 = absent, 5 = excellent)
6.2. Methodology

Q11 Are you frustrated with the results provided by academic search systems? (Google Scholar, specialised database, etc.)? (1 = very content, 5 = very frustrated)

The sample was small; it consisted of 8 participants (see table 6.2).

Table 6.2: Sample Population (the number of the question related to the socio-demographic questionnaire)

<table>
<thead>
<tr>
<th></th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
<th>Q5</th>
<th>Q6</th>
<th>Q7</th>
<th>Q8</th>
<th>Q9</th>
<th>Q10</th>
<th>Q11</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>1976</td>
<td>F</td>
<td>Sociology, sexual health, gender studies</td>
<td>Master</td>
<td>10</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>Google Scholar, PubMed, Jstor, Elsevier</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>P2</td>
<td>1985</td>
<td>F</td>
<td>Translation and communication</td>
<td>PhD</td>
<td>30</td>
<td>15</td>
<td>3</td>
<td>4</td>
<td>Translation studies, Google Scholar, Erudit</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>P3</td>
<td>1982</td>
<td>F</td>
<td>International relations</td>
<td>Master</td>
<td>15</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>Google</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>P4</td>
<td>1983</td>
<td>H</td>
<td>Computer science</td>
<td>Master</td>
<td>6</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>Google Scholar, Google</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>P5</td>
<td>1978</td>
<td>F</td>
<td>Computer science, Information systems</td>
<td>PhD</td>
<td>30</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>Google Scholar, PubMed</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>P6</td>
<td>1957</td>
<td>H</td>
<td>Physics, history</td>
<td>PhD</td>
<td>30</td>
<td>30</td>
<td>4</td>
<td>3</td>
<td>Google</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>P7</td>
<td>1977</td>
<td>F</td>
<td>Literature</td>
<td>PhD</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>Google</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>P8</td>
<td>1983</td>
<td>F</td>
<td>Gender Studies</td>
<td>Master</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>Rero, Cairn, Google, Google Scholar</td>
<td>5</td>
<td>4</td>
</tr>
</tbody>
</table>
6.3 Results

6.3.1 Questions answered after each task

We computed the average response for the three tasks (see table 6.3). We tested the difference between the participants who had to evaluate the FSAD versus the keyword search using analysis of variance (ANOVA) tests.

6.3.1.1 Relevance or irrelevance of the set of results (scale)

**Q1** Do you think the set of results was relevant to the task? (1=not useful, 5=useful)

**Q2** Do you think the number of results was too large to be useful? (1 = totally unusable, 5 = usable)

<table>
<thead>
<tr>
<th>Question</th>
<th>Interface type</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>FSAD</td>
<td>4</td>
<td>4.0</td>
<td>.81</td>
<td>.780</td>
</tr>
<tr>
<td></td>
<td>Keywords</td>
<td>4</td>
<td>3.75</td>
<td>1.5</td>
<td></td>
</tr>
<tr>
<td>Q2</td>
<td>FSAD</td>
<td>4</td>
<td>4.5</td>
<td>.577</td>
<td>.045</td>
</tr>
<tr>
<td></td>
<td>Keywords</td>
<td>4</td>
<td>2.75</td>
<td>1.258</td>
<td></td>
</tr>
</tbody>
</table>

For the first question (relevance), no significant difference was observed between participants using the FSAD versus the keyword search interface. The two groups seemed to think that the set of results was sufficiently relevant to achieve the task. For the second question (number of results), a significant difference was observed ($p = 0.045$) between the two groups. The group that used the keyword search system seemed to find that the set of results was too large to be useful to achieve the task, whereas the participants who used the FSAD seemed to find that the number of results was useful.

6.3.1.2 Relevance or irrelevance of the set of results (in percent) (see table 6.4)

**Q3** How many elements correspond to your request? (1 = 0-5%, 2 = 6-15%, 3 = 16-30%, 4 = 31-50%, 5 = 51-75%, 6 = 76-90%, 7 = +90%)

**Q4** How many elements were completely irrelevant? (1 = 0-5%, 2 = 6-15%, 3 = 16-30%, 4 = 31-50%, 5 = 51-75%, 6 = 76-90%, 7 = +90%)

For the third question, FSAD participants found that 51%-75% of the results were useful, whereas the participants who used the keyword search interface found that 31%-50% of the results were useful; the difference was not significant.
6.3. Results

Table 6.4: Relevance of the set of results in percent

<table>
<thead>
<tr>
<th>Question</th>
<th>Type interface</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q3</td>
<td>FSAD</td>
<td>4</td>
<td>5.0</td>
<td>1.63</td>
<td>.494</td>
</tr>
<tr>
<td></td>
<td>Keywords</td>
<td>4</td>
<td>4.25</td>
<td>1.25</td>
<td></td>
</tr>
<tr>
<td>Q4</td>
<td>FSAD</td>
<td>4</td>
<td>1.5</td>
<td>.57</td>
<td>.031</td>
</tr>
<tr>
<td></td>
<td>Keywords</td>
<td>4</td>
<td>3.75</td>
<td>1.5</td>
<td></td>
</tr>
</tbody>
</table>

For the fourth question, FSAD participants found that 6%-15% of the results were useless, whereas the participants who used the keyword search interface found that 31%-50% of the results were useless. The difference was significant ($p = 0.031$).

Overall, the FSAD participants seemed to find the set of results more relevant than the keywords search participants did (see figure 6.3, 6.2).

Figure 6.2: Relevance of the set of results by interface
Figure 6.3: Irrelevance of the set of results by interface (some scores have been inverted, so that all are in the same direction (1= irrelevant or useless; 5 or 7=very useful or relevant)
6.3. Results

6.3.1.3 Time to obtain information (see table 6.5)

Q5 Do you think that these results are obtained faster in this way than when using a common scientific information search system (Google Scholar, etc.)? (1 = not at all faster, 5 = much faster)

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q5</td>
<td>FSAD</td>
<td>4</td>
<td>3.5</td>
<td>.51</td>
</tr>
<tr>
<td></td>
<td>Keywords</td>
<td>4</td>
<td>3.6</td>
<td>.14</td>
</tr>
</tbody>
</table>

Table 6.5: Time to obtain information

For the fifth question, no difference was observed between the groups. Both groups appeared to have mixed feelings about the speed of the systems. It should be noted that the systems performed poorly at times, resulting in a long time required to obtain a set of results. Thus, the participants may have been confused between the performance in terms of the time required to obtain the set of results and the time required to achieve the task.

6.3.1.4 Number of requests

Q6 Do you think the number of requests you made was adequate to achieve good results? (1 = very adequate, 5 = not at all suitable) (see table 6.6)

<table>
<thead>
<tr>
<th></th>
<th>Type of interface</th>
<th>N</th>
<th>Mean</th>
<th>sd</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q6</td>
<td>FSAD</td>
<td>4</td>
<td>2.16</td>
<td>.21</td>
<td>.223</td>
</tr>
<tr>
<td></td>
<td>Keywords</td>
<td>4</td>
<td>2.83</td>
<td>.44</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.6: Number of requests

More FSAD participants than keyword search interface participants seemed to find that the number of queries was adequate, but the difference was not significant.

6.3.1.5 Satisfaction with the set of results and the overall results

Q7 Did you obtain satisfactory results for each query you made? (1 = not at all satisfied, 5 = very satisfied) (see table 6.7)

The FSAD participants found more results to be satisfactory than did the keyword search interface participants, but the difference was not significant.

Q8 Are you satisfied with the overall results provided? (1 = not at all satisfied, 5 = completely satisfied) (see table 6.8)
Table 6.7: Satisfaction for each query

<table>
<thead>
<tr>
<th>Type of interface</th>
<th>N</th>
<th>Mean</th>
<th>sd</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FSAD</td>
<td>4</td>
<td>3.83</td>
<td>.34</td>
<td>.484</td>
</tr>
<tr>
<td>Keywords</td>
<td>4</td>
<td>3.41</td>
<td>.43</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.8: Satisfaction with the overall results

<table>
<thead>
<tr>
<th>Type of interface</th>
<th>N</th>
<th>Mean</th>
<th>sd</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FSAD</td>
<td>4</td>
<td>4.16</td>
<td>.28</td>
<td>.273</td>
</tr>
<tr>
<td>Keywords</td>
<td>4</td>
<td>3.41</td>
<td>.55</td>
<td></td>
</tr>
</tbody>
</table>

The participants who used the keyword search interface seemed to be less satisfied with the overall results (Q8) than the participants who used the FSAD, but the difference was not significant.

For the two questions regarding satisfaction, the FSAD group seemed to be more satisfied than the keyword search interface group; however, the difference was not significant.

6.3.1.6 Frustration linked to the set of results

Q9 Are you frustrated by the set(s) of results provided? (1 = totally frustrated, 5 = not at all frustrated) (see table 6.9)

Table 6.9: Frustration with the set of results

<table>
<thead>
<tr>
<th>Type of interface</th>
<th>N</th>
<th>Mean</th>
<th>sd</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FSAD</td>
<td>4</td>
<td>3.16</td>
<td>.56</td>
<td>.220</td>
</tr>
<tr>
<td>Keywords</td>
<td>4</td>
<td>4.25</td>
<td>.55</td>
<td></td>
</tr>
</tbody>
</table>

The FSAD group seemed to be less frustrated with the set of results than the keyword search interface group, but the difference was not significant. No significant correlation was found between the level of frustration indicated by the responses to this question and that indicated by the responses to question Q11 (socio-demographic questions; Are you frustrated with the results provided by academic research systems (Google Scholar, specialised database, etc.)?)

6.3.1.7 Difficulty of the task

Q10 Did you find this task difficult? (1 = very difficult, 5 = very easy) (see table 6.10)
6.3. Results

Table 6.10: Difficulty of the task

<table>
<thead>
<tr>
<th>Q10</th>
<th>N</th>
<th>Mean</th>
<th>sd</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>FSAD</td>
<td>4</td>
<td>3.50</td>
<td>.38</td>
<td>.780</td>
</tr>
<tr>
<td>Keywords</td>
<td>4</td>
<td>3.66</td>
<td>.49</td>
<td></td>
</tr>
</tbody>
</table>

No significant difference was observed between the two groups. Both groups seemed to find the task to be of average difficulty. No significant correlation was found between the level of frustration indicated by the responses to this question and that indicated by the responses to question Q8 (socio-demographic questions; Do you find it difficult to find the information you need in the documents retrieved?)

6.3.2 Questions asked after completion of the three tasks

6.3.2.1 Usefulness and precision of the system

Q1 Do you think this system is useful? (1 = useless and 5 = very useful)

Participants seemed to think that both systems were useful (FSAD and keyword search interface $M = 4.25$).

Q2 I think the indexing of these documents is not sufficiently precise for my information needs. (1 = totally disagree, 5 = totally agree)

The FSAD group found the system precise enough ($M = 2.00$), more so than the keyword search interface group ($M = 3.50$), but the difference was not significant.

Q3 Do you think the search by type of sentence is more effective than the search by keywords in conventional systems? (1 = disagree, 5 = strongly agree)

This question was only asked to the users of the FSAD, who agreed that this system is more effective than conventional systems ($M = 4.25$).

6.3.2.2 Comparison with other search engines

Q4 Do you think this system is better for finding the information you need than your usual system (Google scholar, etc.)? (1 = disagree, 5 = strongly agree)

No significant difference was found with respect to the type of interface. Both groups agreed that the system they used in the present study (either the FSAD or keyword search interface) was superior to their usual system (Google scholar, etc.) for finding the required information ($M = 4.00$ for FSAD and $M = 3.67$ for keywords).

Q5 I would use this system instead of the search engines that I use every day. (1 = disagree, 5 = totally agree)
No significant difference was found regarding the type of interface. Both groups agreed moderately with this statement ($M = 3.70$ for FSAD and $M = 3.25$ for keywords). One participant did not answer this question. Thus, the participants seemed to find the system they used slightly better than their usual system and seemed to appreciate the fact that only one sentence was displayed rather than an entire document.

P4 indicated that the possibility of focusing on only a single paragraph was very useful, making it unnecessary to read an entire document.

P5 indicated that in Google Scholar, she would have had to open the document and find the most relevant sentence or paragraph. She noted that she liked the "disarticulation" of the article into sentences.

P6 indicated that if the system had operated completely correctly (there were some bugs), it would certainly be better than Google.

P7 indicated that Google can cause the reader to adopt a different type of reading style, whereas this system is more focused. She was frustrated that sentences were not linked to the document.

P8 indicated that the proposed system allowed for finer queries than those allowed by other search engines and that the search results should be more precise.

### 6.3.2.3 Usability of the system

Questions concerning usability were inspired by the USE questionnaire\(^5\).

**Q6** I found the system unnecessarily complex. (1 = disagree, 5 = totally agree)

No significant difference between the two groups was observed for this question. Neither group found the system used to be unnecessarily complex ($M = 2.25$ for FSAD and $M 2.00$ for keyword search).

**Q7** I think this system is easy to use. (1 = disagree, 5 = totally agree)

No significant difference was observed for this question. Both groups agreed that the system was easy to use ($M = 3.75$ for FSAD and $M 3.75$ for keyword search).

**Q8** I think I need support/technical assistance to use this system. (1 = disagree, 5 = totally agree)

No significant difference was observed for this question. Both groups declared that they did not need support or technical assistance ($M = 2.25$ for FSAD and $M 2.25$ for keyword search).

**Q9** I think that users (scientists) could easily learn to use this system. (1 = disagree, 5 = totally agree)

\(^5\)http://hcibib.org/perlman/question.cgi?form=USE
6.3. Results

No significant difference was observed for this question. Both groups thought that scientists could easily learn to use this system ($M = 4.75$ for FSAD and $M = 4.50$ for keyword search). Some doubts were expressed by some participants regarding the ease of use for a "general population".

Q10 I need to learn many new things before I feel comfortable and can use this system easily. ($1 = $disagree, $5 = $totally agree)

Participants in the FSAD group indicated that less learning was necessary before they became comfortable with the system than did participants in the keyword search interface group ($M = 1.00$ for FSAD and $M = 2.25$ for keywords search), but no significant difference was observed.

Q11 I felt comfortable using this system. ($1 = $disagree, $5 = $totally agree)

No significant difference was observed for this question. Both groups seemed comfortable using each system ($M = 3.75$ for FSAD and $M = 3.75$ for keyword search).

6.3.3 Logs

Task 1 The analysis of the log for task 1 (the definition of feminism) reveals that SciAnnotDocSearch (FSAD) participants queried the system more often ($M = 5.7$) than the keyword search participants did ($M = 4.75$). To calculate the probability of significance, we used a Student’s t-test. The difference was not significant ($p = 0.266$). We also tested whether expertise in the domain of gender studies had an effect on the number of queries. Experts queried the system more often ($M = 6.75$) than non-experts did ($M = 3.75$). The difference was significant ($p = 0.023$).

The FSAD participants clicked on the definition check box. P3 initially had some difficulties in changing the strategies she normally uses in querying search engines such as Google, and she added "definition" and "define" while the definition check box remained clicked. At the end of her search, P7 clicked the definition box off and performed two queries, typing in the keyword text field "define AND feminism" and "definition AND feminism" to determine whether some definitions were missing. She observed that all the definitions considered when she was using the facet definition and only typing "feminism" in the keyword text field were present in the set of results obtained with the term "define" or the term "definition". However, she added that the "new set" of definitions obtained only by the keywords was not as complete as the one obtained with the facet definition and the term "feminism".

In general, keyword search participants typed "feminism" into the keyword field first to search for the definition. Because they received a set of results containing 3,000 answers (the upper limit in the query), they refined their query with words such as "define", "definition", or "meaning" or short sentences such as "what is feminism".
For both systems, each time participants received a set of results containing more than 100 answers, they refined their query.

The average number of answers was less important for the FSAD group (\(M = 30.28\)) than for the keyword search group (\(M = 731.63\)). The difference is significant (\(p = 0.012\)). The difference in the average number of answers between the experts (\(M = 306.38\)) and non-experts (\(M = 462.75\)) was not significant (\(p = 0.314\)).

The largest set of results obtained using the FSAD contained 148 definitions when the participants only clicked on the definition check-box and typed "feminism". For the keyword search interface, "definition" combined with "feminism" generated a set of results containing 16 answers versus 12 for "define" combined with "feminism".

We calculated the average time that participants focused on one set of results and on one answer (see table 6.11).

Table 6.11: Average time (minutes) spent on a set of results and on an answer (the results for the FSAD participants are shown in bold)

<table>
<thead>
<tr>
<th></th>
<th>set of results (min)</th>
<th>answer (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>1.16</td>
<td>0.001</td>
</tr>
<tr>
<td>P2</td>
<td>4.25</td>
<td>0.076</td>
</tr>
<tr>
<td>P3</td>
<td>1.02</td>
<td>0.021</td>
</tr>
<tr>
<td>P4</td>
<td>2.22</td>
<td>0.002</td>
</tr>
<tr>
<td>P5</td>
<td>1.17</td>
<td>0.001</td>
</tr>
<tr>
<td>P6</td>
<td>2.42</td>
<td>0.113</td>
</tr>
<tr>
<td>P7</td>
<td>1.36</td>
<td>0.044</td>
</tr>
<tr>
<td>P8</td>
<td>1.03</td>
<td>0.002</td>
</tr>
</tbody>
</table>

FSAD participants spent, on average, more time on the set of results (\(M = 2.26\)) and on an answer (\(M = 0.064\)) than keyword search participants did (\(M = 1.39\) for the set of results and \(M = 0.0019\) for an answer). The difference was not significant for the set of results (\(p = 0.329\)) but was significant for the answers (\(p = 0.05\)). It is likely that both differences would become significant with a larger sample.

Non-experts queried the system more (\(M = 15\)) than experts did (\(M = 10\)). The number of participants was already insufficient for the test to be powerful; therefore, we did not test whether the expertise effect varied with the type of interface.

6.3.3.1 Task 2

The analysis of the log for Task 2 (academia and gender gap) also showed that FSAD participants introduced more queries (\(M = 14.25\)) than the keyword search participants did (\(M = 11.25\)). The difference was not significant (\(p = 0.531\)). Non-experts queried the system more (\(M = 15\)) than experts did (\(M = 10\)). The
difference was not significant \( (p = 0.190) \). Participant P6 started to "play" with the system for this query to observe what happened depending on the characteristics of the query and tried several queries without actually reading the results.

In this task, all users used the second element as a condition for the first one. All FSAD participants clicked on "findings" for the first element of discourse. They typed in two general terms when they did not check a second discourse element and one term if a second element was checked. For example, P2 clicked on the finding element for a query typed "equality AND academic" and did not click a second element. When she added the second element, she typed "equality" in the findings and "academic" in the second finding element. The second element checked was findings, methodology or hypothesis.

The average number of answers (including the queries that did not yield answers) was less important for the FSAD \( (M = 10.06) \) than for the keyword search interface \( (M = 16.27) \). The difference was not significant \( (p = 0.131) \). The average number of answers did not differ between the experts \( (M = 13.47) \) and non-experts \( (M = 12.91) \) \( (p = 0.902) \).

In contrast with Task 1, the set of results was the largest for the FSAD interface: 206 for a query with the term "women" in methodology and "university" in the findings. For the keyword search interface, the largest set of results was 30 for the expression "equality AND academic". This difference seems plausible because the keyword search interface, like the FSAD interface, queries only keywords associated with one sentence and not an entire document. Moreover, even in documents about inequality in academia, it may be plausible that "equality" and "academia" are not in the same sentence.

We calculated the average time that participants focused on one set of results and on one answer (see table 6.12).

Table 6.12: Average time (minutes) spent on a set of results and on an answer (the results for the FSAD participants are shown in bold)

<table>
<thead>
<tr>
<th></th>
<th>set of results (min)</th>
<th>answer (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>0.32</td>
<td>0.006</td>
</tr>
<tr>
<td>P2</td>
<td>0.43</td>
<td>0.0031</td>
</tr>
<tr>
<td>P3</td>
<td>1.03</td>
<td>0.0036</td>
</tr>
<tr>
<td>P4</td>
<td>0.24</td>
<td>0.0009</td>
</tr>
<tr>
<td>P5</td>
<td>2.06</td>
<td>0.02</td>
</tr>
<tr>
<td>P6</td>
<td>1.0</td>
<td>0.003</td>
</tr>
<tr>
<td>P7</td>
<td>0.57</td>
<td>0.0035</td>
</tr>
<tr>
<td>P8</td>
<td>0.57</td>
<td>0.006</td>
</tr>
</tbody>
</table>

The two groups did not differ in time spent on a set of results or on answers. The difference was not significant between experts and non-experts (expert set of results: \( M = 0.927 \), answer: \( M = 0.007 \); non-expert set of results: \( M = 0.278 \),
answer $M = 0.004$).

6.3.3.2 Task 3

The analysis of the log for Task 3 (salary and gender gap) shows that the FSAD participants introduced more queries ($M = 12$) than the keyword search participants did ($M = 8$). The difference was not significant ($p = 0.291$). The difference between experts ($M = 9.5$) and non-experts ($M = 10.5$) was not significant ($p = 0.890$). The participant P6 started to "play" with the system for a query to observe what happened depending on the characteristics of the query and tried several queries without actually reading the results.

In this task, all users used the second element as a condition for the first one. All FSAD participants clicked on "findings" for the first element of discourse. They typed in two general terms when they did not check a second discourse element and one term if a second element was checked. The second element was findings. For example, in a query, P6 clicked on findings and added the terms "salary AND equality", and in a second query, he clicked on findings for the first element and added the term "women"; for the second element, he clicked on "findings" and added the term "wage". P3 did not click the check-box for "taking into account the second element", and the second element was not considered.

The largest set of results for the FSAD interface was 161 ("gender" in findings and "salary" in findings) versus 850 for the keyword search interface ("gender AND wage").

Table 6.13: Average time (minutes) spent on a set of results and on an answer (the results for the FSAD participants are shown in bold)

<table>
<thead>
<tr>
<th></th>
<th>set of results (min)</th>
<th>answer (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>0.29</td>
<td>0.020</td>
</tr>
<tr>
<td><strong>P2</strong></td>
<td>3.33</td>
<td>0.302</td>
</tr>
<tr>
<td>P3</td>
<td>0.49</td>
<td>0.013</td>
</tr>
<tr>
<td>P4</td>
<td>0.35</td>
<td>0.032</td>
</tr>
<tr>
<td>P5</td>
<td>22.24</td>
<td>0.002</td>
</tr>
<tr>
<td><strong>P6</strong></td>
<td>0.47</td>
<td>0.0125</td>
</tr>
<tr>
<td>P7</td>
<td>1.01</td>
<td>0.0122</td>
</tr>
<tr>
<td>P8</td>
<td>3.17</td>
<td>0.143</td>
</tr>
</tbody>
</table>

FSAD participants spent, on average, less time on the set of results ($M = 1.325$) and on an answer ($M = 0.085$) than keyword search participants did ($M = 6.512$ for the set of results and $M = 0.361$ for an answer). The difference was not significant.

Non-experts spent more time on the set of results ($M = 6.597$) and on an answer ($M = 0.093$) than experts did ($M = 1.24$ for the set of results, $M = 0.047$ for an answer). The difference was not significant. Because P5 spent a considerable
amount of time evaluating the set of results that contained 850 answers and 1) did not perform any other query and 2) was the only one to spend this amount of time, she could be considered an outlier, even if the sample were larger.

Figure 6.4 reports the total time spent on a set of queries by the two groups of participants for each task.

Figure 6.4: Total time (in minutes) spent on query by group and by task
6.3.3.3 Evaluation of the precision/recall

The estimated precision of the FSAD on task 1 (based on a sample of 80 results out of 148) was 0.61. This precision is calculated on the number of the definition of the term "feminism" and not on the precision of any definition that is higher (0.76). The estimate based on the automated annotation recall was 0.32. The F1 was 0.419. The estimated number of relevant result provided by FSAD was 90. The precision of the keyword search obtained with the combination of the term "define AND feminism" and "definition AND feminism" was 0.82 (24 relevant results out of 29). Therefore, the ratio between FSAD and the keyword search was 3.77, so the estimated recall of the keyword search was 0.084. The F1 for the keywords-search interface was 0.152 (see figure 6.5).

Figure 6.5: Estimated precision and recall for the task 1
6.4 Discussion and Conclusion

In summary, keyword search interface participants seemed to find the set of results less useful than the FSAD participants did (significant difference). This difference appears to hold true even independently of the number of results in a set. In Task 2, FSAD participants received more answers than the keyword search participants did, but they found the number of irrelevant answers less important than the keyword search participants did (see figure 6.6). This finding may support the fact that the keyword search interface participants found the set of results less useful than the FSAD participants did.

Figure 6.6: Comparison of the number of results in a set of results and the irrelevance of the set of result, task 2

A small difference was observed between the two groups with respect to the relevance of the set of results, with the FSAD participants finding the results slightly more relevant. However, the difference was not significant. There may be several reasons for this lack of significance. First, participants might have built search strategies for searching for scientific documents using a search engine, and those strategies might have yielded "strange" results when applied to the FSAD interface. This finding would correspond to an effect of the novelty of the FSAD interface.

Second, it might be easier for participants to declare that a set of results is not relevant than to find them relevant, especially for non-expert participants. Because the number of participants was small, it was not possible to test the difference between experts and non-experts with respect to each interface.

A third reason for the lack of difference might be that the sample was composed of persons who knew the moderator (i.e., the author of this thesis). It is somewhat simpler to test acquaintances for a system that still presents some "bugs". Even though the moderator indicated to the participants and specified in the description
of the task that they should not "behave nicely" with her system, some participants did not follow this part of the instructions. During the debriefing that followed, the moderator discovered that some participants simply wanted to provide results that would please her. Of course, this bias could exist for both interfaces.

The logs showed that FSAD participants introduced, on average, more queries than the keyword search participants did. In the questionnaire, the FSAD participants nevertheless indicated that the number of queries was adequate (see figures 6.7, 6.8, 6.9), except for query 3. An interpretation of this result could be that even if participants have more "bad" answers in the set of results of one query, such answers could affect the perception of the adequate number of queries. It would be interesting to test whether a larger number of queries that retrieve more "interesting" and "good" answers would alter the perception of the number of queries introduced.

Figure 6.7: Number of queries and feeling of adequacy by interface for task 1 (adequacy; 1= very adequate; 2 = not at all adequate)
6.4. Discussion and Conclusion

Figure 6.8: Number of queries and feeling of adequacy by interface for task 2 (adequacy; 1= very adequate; 2 = not at all adequate)

Figure 6.9: Number of queries and feeling of adequacy by interface for task 3 (adequacy; 1= very adequate; 2 = not at all adequate)
Another factor that might have influenced the results was the curiosity of the moderator and the participants. The experimenter was not always strict about the instructions. For example, for task 3, P5 decided to check all 850 results in her set of results to determine whether the relevance of the set was good. Because the moderator was curious about the relevance of the set of results, she was glad that a participant agreed to look at all 850 results. A last factor could be the familiarity of the participants with the language English that can have an influence on the results.

In the future, the moderator should be stricter during the experiment. The number of participants should also be considered more carefully.

The evaluation provided a good opportunity to test the precision and the recall of the annotation. For Task 1\(^6\), we extracted 80 answers from the set of results that participants obtained in the FSAD query, and we evaluated them manually. The precision of the annotation was nearly equal to that obtained in the test performed with the 555 sentences (see chapter 4, table 4.18). We also observed that a good proportion (80.33\%) of the definitions were indeed definitions of the term feminism (direct and indirect definitions and sub-definitions). Because the JAPE rules (see chapter 4) did not distinguish the definiendum in the definition, it could have been the case that many definitions containing the word "feminism" but with a different definiendum be retrieved. It was therefore surprising to find such a high proportion.

Some measure of self-efficacy should be added to the questionnaire for the participants. However, 1) the questionnaire was already long enough, and 2) this test was a pilot evaluation. In the future, some of the intervening variables (proposed by Wilson in 1986 [Wilson 1999] see figure 2.6) should be tested, particularly when subjective measures are used. The psychological features of the participants may have an effect on the results as well [Heinström 2005].

Despite some methodological problems in the evaluation with the users, we observed a preference among the participants for the FSAD interface and for a search engine that focuses on a sentence rather than an entire document. We also observed that the recall and by consequence the F1 measure are better for the FSAD for the task 1 than for the keywords-search interface. These findings are encouraging for further refinements to the system.

\(^6\)definition of the term feminism
Chapter 7

Conclusion

In each chapter of the present thesis, a conclusion was presented to the reader. Thus, in this final conclusion, we will only briefly summarise the main contributions and results reported in the other chapters and suggest some ideas for future work. The structure of the final conclusion follows the one adopted in the rest of the thesis.

7.1 User model

The first contribution of the present thesis is to propose a user model. In the two studies we conducted, we observed that scientists seeking information focus on specific parts of documents rather than linearly reading the entire document. Thus, scientists do not adopt the same search strategy depending on their objectives and needs. From these observations, we created a user model of scientists.

One part of the survey and the interview aimed to uncover the nature of the necessary information on which scientists mainly focus. We found that the most frequently sought elements of a document can be classified into findings, methodology, definition, related work and hypothesis. Although the nature of the information changes depending on scientists' needs, these five elements are the ones to which scientists devote the most attention. The observations from the interviews and survey were used to conceptualise the annotation model. However, although the survey and the interviews provide a sufficiently solid foundation for developing the annotation model, the number of participants for each study was not sufficient to generalise the results or to have obtain significant results for the differences in usage between areas of research. Future studies should consider using larger samples. In future studies, it would be interesting to observe whether a difference exists regarding the precision and the recall scientists want in their set of results. In other words, it would be interesting to observe whether for a certain task, the precision of the annotation is more necessary than the recall or vice versa. The actual observations and future results could also be used to build a better search interface as well as to improve data visualisation.

7.2 Annotation models for scientific documents

The second contribution of the present thesis is an annotation model for scientific documents (SciAnnoDoc model). The constraints to build the annotation model were as follows: 1) the model has to be useful to scientists and based on their needs; 2) the model has to be flexible and should represent the different dimensions
contained in a scientific document; and 3) the model has to be generic to represent a large spectrum of the scientific literature.

The model is based on the experiment we conducted regarding information needs and the nature of the information needed. It was developed from observations from the survey and the interviews. We retained the five most frequently used discourse elements defined in the user model (finding, definition, hypothesis, methodology and related work), which constitute the core of the SciAnnotDoc model. However, in the future, it might be interesting to refine or extend the SciAnnotDoc model to new discourse elements, such as the aim or problem statement.

The analysis we conducted on scientific documents showed that the same chunk of text can belong to different discourse elements. The analysis of the precise questions extracted from the interviews conducted with the scientists and the manual annotation by scientists and non-scientists showed that the classification varied with the perspective of the annotator. These results are in agreement with [Teufel & Moens 2002, Ibekwe-SanJuan 2006], who argued that one chunk of text can belong to several patterns. For example, a chunk of text can be both a definition and a finding. Because it is important to retain this characteristic of a chunk of text, the annotation model allows us to annotate the same chunk of text with many type it is necessary.

We also observed that scientists are not always interested in finding the original ideas of another author but rather search for information concerning a given fact. That is, they do not really care whether the author cited is the one who "discovered" or first mentioned a fact, but they need to know whether a given piece of information has already been mentioned by others. This is why the documents were annotated from a "universal" point of view and not an author-centered point of view. The consequence of this observation is that related work always belongs to another discourse element. For example, the following example is annotated as a definition, a finding and a related work.

"As an antidote to the production and reinforcement of prevailing notions of sexual identity, Butler argues that homosexuality and heterosexuality-like gender-exist as enactments of cultural and aesthetic performances; even as these identities are performed and repeated, they are (in true Foucauldian form) being contested and unraveled".[Deveaux 1994, p.239]

To be able to provide an answer to a precise question, annotating the elements of text with discourse elements only is not sufficient. The SciAnnotDoc model was built on four dimensions:

1. the metadata
2. the textual contents (conceptual dimension)
3. the discourse elements

"As an antidote to the production and reinforcement of prevailing notions of sexual identity, Butler argues that homosexuality and heterosexuality-like gender-exist as enactments of cultural and aesthetic performances; even as these identities are performed and repeated, they are (in true Foucauldian form) being contested and unraveled".[Deveaux 1994, p.239]
4. the relationships between the elements.

Using these four dimensions makes the description of a scientific document more precise. Therefore, each element was annotated along the four dimensions. We created or imported different ontologies representing the different dimensions.

SciDeo is the ontology that represents the discursive dimension (dimension 3 above). This ontology is independent of the domain of research. We tested it with scientific documents from various research domains: computer science, physics, sociology and gender studies. The last one is the domain we retained as a use case for the remainder of the present thesis.

SciMeth, SciObj and GendStud are ontologies that represent textual content (dimension 2 above). SciMeth is an ontology that represents the different methods used in sciences in general. The SciObj ontology represents the different scientific objects in various domains of research. We defined a scientific object as an object that was used in more than one study and was generalised. This ontology will continue to evolve, like most ontologies. For example, a model that previously was present in only one study (or studies of the same group) might later be used by several groups. Consequently, this model will become a scientific object, and the ontology will have to be evaluated as well. The concepts of a specific domain are represented in domain-specific ontologies. Because no gender studies ontology existed, we created the GendStud ontology within the framework of the present thesis.

The relationships dimension (dimension 4 above) was represented by the CiTO ontology [Shotton 2010], which we imported in the SciDeo ontology. In the SciAnnotDoc, the relations between documents relate two discourse elements or discourse elements/document rather than the whole document, as is often the case. This approach provides more precise information on the semantics of the relations between authors and their writing. For example, a scientist using SciAnnotDoc can find that author A has cited author B for definition X but has cited author C for finding Z.

With this level of precision, we can also infer some new information, such as the lineage of an idea, and create a citation network in which the semantics of the context of the citation are present. Currently, we infer that if A cites B and B cites C or C cites B describing the same domain concept and the same discourse element, A and C have a relation. In the future, we will use the properties of the CiTO ontology to create a matrix of relations depending on the nature of the relations between the elements of discourse. For example, if A agrees with B and B disagrees with C on the discourse element level, we would like to define the nature of the relation between A and C.

The three ontologies SciMeth, SciObj and GendStud were developed during the annotation process. Consequently, they are not complete. In the future, they could be improved by adding missing terms, especially SciMeth and SciObj, which focused on the gender studies domain of research. If the aim in future studies is to annotate different areas of research, specific terms extracted from the new area of research should be added, or new ontology created whether none exist in the specific
domain. Some methods exist to populate ontologies from documents, and it would be interesting to evaluate some of these methodologies to automatically complete these ontologies.

7.3 Corpus annotation

To test and evaluate the SciAnnotDoc model with scientists, we needed a sufficiently large corpus of annotated scientific documents in a given domain of research. In the present thesis, we selected gender studies. As explained, we chose this domain because it was relatively unexplored and challenging because the scientific documents are representative of social and human sciences and are less structured than in other research domains (such as life sciences or "hard" sciences). We tried different approaches and techniques. In particular, we used both a manual and an automatic annotation technique.

On the one hand, manual annotation presents some advantages. For instance, manual annotation allows for subtleties in the four dimensions of the model, such as the nature of the citation between two discourse elements, the discovery of the definiendum of a definition, the type of findings or the link between different discourse elements in the same documents. However, manual annotation also has disadvantages; in particular, it is very time consuming, and the probability of obtaining high inter-judge agreement is not very high. We attempted to use crowd-sourcing to annotate the scientific documents, but we observed that even if the level of inter-judge agreement was high among the workers of Amazon Mechanical Turk, the number of mistakes in classifying the type of discourse elements was too large. We also observed that when using this kind of crowd-sourcing, the task has to be short; otherwise, the batch is rarely finished. By using scientists for the manual annotation, we observed that the number of mistakes in classifying the type of discourse elements was less important, but the time spent by scientists to classify a sentence is long. It could be possible to reduce the time of this task using experts in the domain of area. However, even if they were faster, the time they would take to classify the discourse element into one or several categories would certainly still be too long.

On the other hand, an automatic annotation has the advantage of making it possible to annotate a large number of documents in a short amount of time. However, some subtleties are not possible. We simplified the annotation model to be able to use automatic annotation. We used GATE and created several JAPE rules based on cue phrases built on syntactic patterns and concepts from the domain ontologies that can be used as heuristics to classify the sentences in the different types of discourse element. We obtained good precision (\textit{findings} = 0.82, \textit{hypothesis} = 0.62, \textit{definition} = 0.8, \textit{methodology} = 0.83), but the recall was low (\textit{findings} = 0.39, \textit{hypothesis} = 0.29, \textit{definition} = 0.32, \textit{methodology} = 0.46). The aim of the annotation process was to have an annotated corpus large enough to evaluate some cases of use with scientists; 1410 documents were annotated. This corpus is sufficient to evaluate the SciAnnotDoc model with scientists, and it makes
it possible to compute some interesting meta-statistics on the corpus itself or on the
domain of research, such as the number of times a concept appears by year. This
type of statistic opens novel paths for empirical research on the scientific documents
themselves and on the domain of research. The SciAnnotDoc not only allows users
to introduce more precise queries that correspond more closely to the user’s needs,
but it also enables meta-analysis of the domain of research, such as tracing the
"trend" of a term.

The consequences of the simplification of the model imply that the annotations
of the scientific document are less precise. For example, it was not possible to
annotate the definiendum of a definition, citations were made for the discourse
element/document and not the discourse element/discourse element and did not
describe the nature of the relation (the upper level "cites" property was used), or
we did not detect a relation between the discourse element in the same document.
To correct this lack of precision, other JAPE rules could be created. For the nature
of the relation between discourse elements, sentiment analysis approaches could be
tested.

To improve the automatic annotation process, particularly recall, one solution
could be to create more JAPE rules. Another solution could be to test whether some
hybrid approaches that combine a rules-based approach and a machine-learning ap-
proach may improve precision and recall. Another solution would be to ask experts
not to classify the sentences into categories but to confirm the types of categories
into which a sentence is already classified. By using this type of methodology, we
can improve and enlarge a training corpus that could be used to improve the preci-
sion and recall of the actual annotation process, and that could be used in hybrid

7.4 Query model and FSAD

The query model was built according to the needs expressed by scientists in their
specific questions. The semantics of queries were expressed through a mapping to
SPARQL queries on the SciAnnotDoc model. The query model does not allow for
the full expressiveness that SPARQL could provide, but it is sufficiently flexible to
be able to query documents to meet scientists’ information needs.

We implemented in JAVA a faceted search interface based on the SciAnnotDoc
model (Faceted search for Scientific Annotated Document, FSAD). The first aim of
this interface was to evaluate scientists’ perceptions regarding the annotated corpus.
The visualisation of the set of results was reduced in a simple way (possibly too
simple). In the future, we intend to produce a more interactive view of the set of
results. Furthermore, some simple improvements, such as the ability to click on
the discourse element to navigate through the whole document, could be made as
well as more sophisticated solutions, such as a different visualisation depending on
the task scientists need to achieve, and the nature of the needed information. For
example, scientists that seek all the definitions of a term and the contradictions
between the different definitions may want a different visualisation of the set of results than scientists who are seeking findings that observe that population X is better at calculus than population Y. The variation in the visualisation of the set of data depending on the task, and the nature of the needed information is an area of research that I intend to examine further in my Post-Doc.

7.5 User evaluation

Considering the other scientific annotation models that exist in the literature, to our knowledge, our approach is the only one that has been evaluated with real end-users (scientists).

Despite the small number of participants, this evaluation demonstrated that the SciAnnotDoc model seems to outperform a keyword-search system in terms of answering a precise question for users. The relevance of the results seems higher, as shown by the statements expressed by the scientists. The estimation of the recall and the F1 score show that the FSAD has better recall and a better F1 than the keyword search, although the precision is slightly lower for the FSAD.

Scientists seem to find that the SciAnnotDoc delivers less "bad" answers than a keyword-search system, and they appear to be less frustrated and more satisfied by the set of results.

It is intriguing to observe that even when the number of retrieved results is more important in the FSAD than in the keyword-search system, users nevertheless seem to find that the relevance is higher for the FSAD than for the keyword search. The paradigm of a "best" document, on which most of the conventional scientific search engines are based, is not necessarily the best one for searching scientific documents.

In conclusion, scientists want and need better tools to find information. The different types of discourse elements we selected from the interviews and surveys seem to be sufficient to represent a large panel of precise questions expressed by scientists.

In the future, we will evaluate the SciAnnotDoc with more participants, including participants that do not personally know the author of this thesis, to prevent bias. It would also be interesting to compare the precision and recall for queries other than the definition of the term "feminism". The number of documents that a scientist needs to read to be satisfied with the answer is a research question that requires further exploration. The first issue will be to analyse whether the number of documents needed varies depending on the task and, consequently, whether the precision and the recall of the annotation should vary depending on the information scientists want to find.

The SciAnnotDoc model is a realistic model that can be used in an automatic annotation process. The results provided on the basis of the annotation process seem to be better in terms of precision (and even recall) to fulfill the information needs
of scientists, as the evaluation shows. Participants seem to find that the relevance of the set of results is better for the FSAD and that the irrelevance of the set of results is more important for the keyword search. Although these results are not completely conclusive, they show that even if the process of annotating scientific documents by their discourse element, the relation between the different documents and the terms (keywords) is costly in terms of time for the indexing process and that the query time for users is more important, the process of annotating the scientific documents with the fourth defined facets (metadata, conceptual, discourse element and relationship dimensions) seems to be worthwhile to provide better results for a precise question.

The potential impact of this thesis is important, as are the various possibilities for studies and applications resulting from this research. In the short term, the findings of this study could be integrated into systems such as the open archives of a university or research centre to conduct testing with "real" users in a "real" environment and to help scientists in their information-seeking behaviour. In the longer term, it should be possible, as Bush [Bush 1945] proposed in 1945, to make a real use of the ocean of publications that exist today.


[Ou & Kho 2008] Shiyan Ou and Christopher S. G. Kho. Aggregating search results for social science by extracting and organizing research concepts and


[Sendhilkumar et al. 2013] S Sendhilkumar, E Elakkiya and GS Mahalakshmi. *Citation Semantic Based Approaches To Identify Article Quality.* In ICC-SEA, SPPR, CSIA, WimoA, SCAI, 2013. (Cited on page 70.)


[Shotton 2009a] David Shotton. *CiTO, the Citation Typing Ontology, and its use for annotation of reference lists and visualization of citation networks.* Bio-Ontologies 2009 Special Interest Group meeting at ISMB, 2009. (Cited on pages xi and 71.)


**bouncing** "Bouncing describes a form of behaviour in which users view only a few web pages from the vast numbers available and generally do not return to the same website very often, if at all." [Nicholas *et al.* 2007, p.1086]. 21

**conceptual indexing** "The conceptual indexing consists in representing documents (queries) by concepts instead of words. Thus, during the retrieval process, the matching between a query and a document is done based on a non-ambiguous vocabulary (concepts)." [Radhouani *et al.* 2008, p.3]. 54

**definition** In philosophy, the specification of the meaning of an expression relative to a language. Definitions may be classified as lexical, ostensive, and stipulative. Lexical definition specifies the meaning of an expression by stating it in terms of other expressions whose meaning is assumed to be known (e.g., a ewe is a female sheep). Ostensive definition specifies the meaning of an expression by pointing to examples of things to which the expression applies (e.g., green is the color of grass, limes, lily pads, and emeralds). Stipulative definition assigns a new meaning to an expression (or a meaning to a new expression); the expression defined (definiendum) may either be a new expression that is being introduced into the language for the first time, or an expression that is already current. [Encyclopaedia Britannica Online 2014a] 79

**information behaviour** "Information behaviour is meant those activities a person may engage in when identifying his or her own needs of information, searching for such information in any way, and using or transferring that information." [Wilson 1999, p.249]. 4, 12

**information need** "Signifies a consciously identified gap in the knowledge available to an actor. Information need may lead to information seeking and formation of requests of information. Information need can be different for scientist that it is for "normal situation", in the sense that scientist can have the knowledge, but in some circumstance need to reference or confirm this knowledge with somebody else knowledge (e.g. scientist can know that hearth is round, but they have to prove it, or to reference other work that prove this fact)." [Ingwersen & Järvelin 2006, p. 20]. 15

**information-seeking behaviour** "Human information behaviour dealing with searching or seeking information by means of informations sources and (interactive) information retrieval systems; also called IS&R behaviour." [Ingwersen & Järvelin 2006, p. 386]. 3, 4, 8, 12, 13, 15, 20, 21, 24
inter-individual variations "Inter-individual variations: differences between individual scientists (traditionally referred to as "individual differences")"
[Garvey et al. 1974, p.115]. 8

intra-individual variations "Intra-individual variations: changes which occur within individual scientists as their scientific work progresses"
[Garvey et al. 1974, p.115]. 8

ontology An ontology is an explicit specification of a conceptualization. The term is borrowed from philosophy, where an Ontology is a systematic account of Existence. For AI systems, what "exists" is that which can be represented. When the knowledge of a domain is represented in a declarative formalism, the set of objects that can be represented is called the universe of discourse. This set of objects, and the describable relationships among them, are reflected in the representational vocabulary with which a knowledge-based program represents knowledge. Thus, in the context of AI, we can describe the ontology of a program by defining a set of representational terms. In such an ontology, definitions associate the names of entities in the universe of discourse (e.g., classes, relations, functions, or other objects) with humanreadable text describing what the names mean, and formal axioms that constrain the interpretation and well-formed use of these terms.

Formally, an ontology is the statement of a logical theory. Ontologies are often equated with taxonomic hierarchies of classes, but class definitions, and the subsumption relation, but ontologies need not be limited to these forms. Ontologies are also not limited to conservative definitions, that is, definitions in the traditional logic sense that only introduce terminology and do not add any knowledge about the world (Enderton, 1972) . To specify a conceptualization one needs to state axioms that do constrain the possible interpretations for the defined terms."[Gruber 1995, p.2] 54

self-efficacy "Perceived self-efficacy is defined as people’s beliefs about their capabilities to produce designated levels of performance that exercise influence over events that affect their lives."[Bandura 1994, p.71]. 13

semantic Publishing "In the present context, I define ?semantic publishing? as anything that enhances the meaning of a published journal article, facilitates its automated discovery, enables its linking to semantically related articles, provides access to data within the article in actionable form, or facilitates integration of data between papers. Among other things, it involves enriching the article with appropriate metadata that are amenable to automated processing and analysis, allowing enhanced verifiability of published information and providing the capacity for automated discovery and summarization. These semantic enhancements increase the intrinsic value of journal articles, by increasing the ease by which information, understanding and knowledge can be
extracted. They also enable the development of secondary services that can integrate information between such enhanced articles, providing additional business opportunities for the publishers involved. Equally importantly, readers benefit from more rapid, more convenient and more complete access to reliable information.\cite{Shotton 2009b, p.86} 3

**Strategic reading** "strategic reading relies on human recognition of scientific terms from plain text, perhaps assisted with simple string searching and mental calculation of ontological relationships, combined with burdensome tactics such as bookmarking, note-taking, and window arrangement." \cite{Reenar 2009, p.2}. 22
Question 1: Do you work with /read scientific papers like articles, abstracts, books etc.?

Question 2: What is your research domain? (Arts, Economics, Medicine, Engineering, Law, Sciences, Social Science, Economics science, Theology, Psychology and Educational sciences, Translation and interpretation, Other)

Question 3: Year of birth?

Question 4: You are? (Female, Male)

Question 5a: What is/are the reason(s) why you are reading scientific papers/books: to stay informed. (Yes, No)

Question 5b: What is/are the reason(s) why you are reading scientific papers/books: to find all/some other studies in your research domain. (Yes, No)

Question 5c: What is/are the reason(s) why you are reading scientific papers/books: to analyses how studies are conducted in your field. (Yes, No)

Question 5d: What is/are the reason(s) why you are reading scientific papers/books: to compare studies. (Yes, No)

Question 5e: What is/are the reason(s) why you are reading scientific papers/books: to find references for writing your own papers. (Yes, No)

Question 5f: What is/are the reason(s) why you are reading scientific papers/books: to find ideas for your own research. (Yes, No)

Question 6a: Do you read scientific paper to respond to a precise need?. (Yes, No)


Question 6c: Where do you generally find the answers: Definition? (Always, Very Frequently, Occasionally, Rarely, Very Rarely, Never).


Question 6g: Where do you generally find the answers: Conclusion? (Always, Very Frequently, Occasionally, Rarely, Very Rarely, Never).

Question 7: Do you often do a review of the literature.? (Yes, No)

Question 8a: Do you easily find papers/books you are looking for, without too much noise (e.g. a paper which is out of the field you are looking for)? (Strongly agree, Agree, Neither agree or disagree, Disagree, Strongly disagree)

Question 8b: Do you use more often google scholar than a specialised database (in your domain)? (Strongly agree, Agree, Neither agree or disagree, Disagree, Strongly disagree)
Appendix B

Semi-direct interview question

Démographie :
1. Année de naissance
2. Sexe
3. Discipline de recherche
4. Interdisciplinaire : oui ou non
5. Localisation de votre université ou centre de recherche
6. Plus haut degré d’études
7. Doctorat : oui ou non
8. Si oui, quelle année l’avez-vous obtenu ?
9. Nombre d’articles écrit en 1er auteur ?
10. Nombre d’articles auquel vous avez participé (1er et autre auteur) ?
11. Nombre d’année de pratique de la recherche ?
12. Nombre d’articles lu environ par mois en moyenne ?
13. Pourquoi recherchez-vous un article ?
14. Recherchez-vous souvent des articles pour les raisons suivantes ?
   • Rechercher une définition
   • Trouver un autre point de vue
   • Rester informé
   • Voir les nouveautés
   • Se faire un bagage de connaissance
   • Conseillé par quelqu’un
   • Par curiosité
   • Rechercher une méthodologie
   • Comparer les articles avec vos idées
   • Comparer les articles avec vos articles à écrire
202 Appendix B. Semi-direct interview question

- Résoudre un problème
- Construire un modèle
- Apprendre / enrichir des connaissances
- Chercher une base théorique à vos recherches

15. Comment recherchez-vous un article ? (mots-clés, composer une hypothèse pour répondre à une question, bibliographie, meta-data, email liste, référence des conférence, browsing liste, suivre les auteurs, etc)

16. Quelles sont les bases de données ou les moteurs de recherche que vous utilisez?

17. Recherchez-vous des articles pour répondre à une question précise ?

18. Pour chaque tâche dites-moi comment vous recherchez des informations ?

19. Est-ce que suivant la tâche pour laquelle vous cherchez, vous cherchez dans des BDD ou systèmes différents ?
   - Se tenir au courant
   - Découvrir un nouveau domaine
   - Ecrire un article
   - Trouver de nouvelles idées
   - Trouver une méthode
   - Trouver des "findings" (résultats, conclusion, données, ...)
   - Trouver une/des définition s
   - Trouver les Hypothèses
   - Pour le plaisir
   - Tache applicative
   - Conseiller un article
   - Apprendre / enrichir des connaissances
   - Reviewer un article
   - Ecrire un SOA
   - Ecrire une proposition de projet
   - Dissémination, poster site web de votre projet ?

20. La connaissance du domaine vous fait-elle changer votre façon de chercher votre information ? Comparer les manières de chercher de l’information suivant le domaine connu ou moins connu ?

21. Pourquoi lisez-vous un article ?

22. Lisez-vous des passages spécifiques ? Lisez-vous des passages en priorité ?
23. Est-ce que vous vous concentrez sur les différents passages en fonction de la tâche que vous avez à effectuer ?

24. Lisez-vous un article différemment suivant la connaissance que vous avez du domaine ?

25. Lisez-vous des passages différents suivant votre connaissance du domaine ?

26. Lisez-vous des fragments/ passages différents suivant la tâches que vous avez à effectuer ?

- Future work
- Exemple
- Findings
- Definition
- Central problem
- Aim
- Background
- Hypothesis
- Methodology
- New idea
- Positioning

- Se tenir au courant
- Découvrir un nouveau domaine
- Ecrire un article
- Trouver de nouvelles idées
- Trouver une méthodologie
• Trouver des résultats
• Trouver une définition
• Conclusion
• Hypothèses
• Pour le plaisir
• Tache applicative
• Conseiller un article
• Apprendre / enrichir des connaissances
• Reviewer un article
• Ecrire un SOA
• Ecrire une proposition de projet
• Dissémination, poster site web de votre projet ?
• Comprendre ou on en est sur une question ? Se mettre à niveau sur une question précise ou large ? Ex : algorithme d’animation sur l’ordinateur, etc ?
SciDeo ontology

Appendix C

Classes

**Analysed_findings**
Analysed_findings ⊑ Findings

**Definiendum**
Definiendum ⊑ ∀ defined Definiens
Definiendum ⊑ Thing
Definiendum ⊑ ∃ part_of Definition
Definiendum ⊑ ¬ Document
Definiendum ⊑ ¬ Fragment
Definiendum ⊑ ¬ DiscourseElement
Definiendum ⊑ ¬ Person
Definiendum ⊑ ¬ Definiens
Definiendum ⊑ ¬ DomainConcept
Definiendum ⊑ ¬ LogicalStructure

**Definiens**
Definiens ⊑ ∀ define Definiendum
Definiens ⊑ Thing
Definiens ⊑ ∃ part_of Definition
Definiens ⊑ ¬ DomainConcept
Definiens ⊑ ¬ DiscourseElement
Definiens ⊑ ¬ Document
Definiens ⊑ ¬ Person
Definiens ⊑ ¬ LogicalStructure
Definiens ⊑ ¬ Fragment

**Definition**
Definition ⊑ ∃ contains Definiens
Definition ⊑ ¬ DiscourseElement
Definition ⊑ ¬ Document

**DiscourseElement**
DiscourseElement ⊑ ∃ belong_to Fragment
DiscourseElement ⊑ ∃ refers DiscourseElement
DiscourseElement ⊑ ∃ cites DiscourseElement
DiscourseElement ⊑ ∃ describe DomainConcept
DiscourseElement ⊑ ¬ Person
DiscourseElement ⊑ ¬ Definiens
DiscourseElement ⊑ ¬ Definiendum
DiscourseElement ⊑ ¬ DomainConcept
DiscourseElement ⊑ ¬ LogicalStructure
DiscourseElement ⊑ ¬ Document

**Document**
Document ⊑ Thing
Document ⊑ ¬ Source
Document ⊑ ∀ written_by Person
Document ⊑ ∀ contains Fragment
Document ⊑ ¬ Definiendum
Document ⊑ ¬ LogicalStructure
Document ⊑ ¬ Fragment
Document ⊑ ¬ Definiens
Document ⊑ ¬ DomainConcept
Document ⊑ ¬ Person
Document ⊑ ¬ DiscourseElement

**DomainConcept**
DomainConcept ⊑ ∀ is_describe_by DiscourseElement
DomainConcept ⊑ Thing
DomainConcept ⊑ ¬ Definiens
DomainConcept ⊑ ¬ Fragment
DomainConcept ⊑ ¬ Document
DomainConcept ⊑ ¬ DomainConcept
DomainConcept ⊑ ¬ LogicalStructure
DomainConcept ⊑ ¬ Person

**Findings**
Findings ⊑ DiscourseElement

**Fragment**
Fragment ⊑ ∃ is_in_logical_structure LogicalStructure
Fragment ⊑ ∀ part_of Document
Fragment ⊑ Thing
Fragment ⊑ ¬ Document
Fragment ⊑ ¬ Person
Fragment ⊑ ¬ DomainConcept
Fragment ⊑ ¬ LogicalStructure
Fragment ⊑ ¬ Definiens

**Hypothesis**
Hypothesis ⊑ ¬ DiscourseElement

**LogicalStructure**
LogicalStructure ⊑ ∃ contains_fragment Fragment
LogicalStructure \notin \text{Person}
LogicalStructure \notin \text{Document}
LogicalStructure \notin \text{Definiendum}
LogicalStructure \notin \text{Definiens}
LogicalStructure \notin \text{DomainConcept}
LogicalStructure \notin \text{Fragment}

Methodology
Methodology \notin \text{DiscourseElement}

Person
Person = \text{Name}
Person \forall \text{write Document}
Person \notin \text{Thing}
Person \notin \text{LogicalStructure}
Person \notin \text{Document}

Raw_findings
Raw_findings \notin \text{Findings}

RelatedWork
RelatedWork \notin \text{DiscourseElement}

Thing

Object properties

\text{belong_to} \notin \text{topObjectProperty}
<http://cui.unige.ch/~deribauh/sciDeo>
#\text{belong_to} \equiv <http://cui.unige.ch/~deribauh/
/sciDeo#has_discourse_element>-\exists \text{belong_to Thing} \notin \text{DiscourseElement}
T \forall \text{belong_to Fragment}

cites
contains
<http://cui.unige.ch/~deribauh/sciDeo>
#\text{part_of} \equiv <http://cui.unige.ch/~
deribauh/sciDeo#contains>-contains\_fragment
<http://cui.unige.ch/~deribauh/sciDeo>
#\text{is_inLogicalStructure} \equiv <http://cui.unige.ch/
/sciDeo#contains\_fragment>-\exists \text{contains\_fragment Thing} \notin \text{LogicalStructure}
T \forall \text{contains\_fragment Fragment}

define
<http://cui.unige.ch/~deribauh/sciDeo>
#\text{define} \equiv <http://cui.unige.ch/
/~deribauh/sciDeo#defined>-\exists \text{define Thing} \notin \text{Definiens}

\top \forall \text{define Definiendum}

defined
<http://cui.unige.ch/~deribauh/sciDeo>
#\text{define} \equiv <http://cui.unige.ch/~deribauh/
/sciDeo#defined>-\exists \text{define Thing} \notin \text{Definiendum}
T \forall \text{define Definiens}

describe
<http://cui.unige.ch/~deribauh/sciDeo>
#\text{describe} \equiv <http://cui.unige.ch/~deribauh/
/sciDeo#has\_discourse\_element>-\exists \text{has\_discourse\_element Thing} \notin \text{Fragment}
T \forall \text{has\_discourse\_element DiscourseElement}

\text{infered\_relation}
\text{infered\_relation\_fragment}

\text{is\_describe\_by}
\text{is\_in\_logical\_structure}

\text{is\_referred\_by}
\text{part\_of}
\text{refers}

\text{topObjectProperty}

\text{write}

<http://cui.unige.ch/~deribauh/sciDeo>
#\text{write} \equiv <http://cui.unige.ch/~deribauh/
/sciDeo#written\_by>-\exists \text{written\_by Thing} \notin \text{Definiens}

T \forall \text{write Definiendum}

defined
Data properties

<table>
<thead>
<tr>
<th>Property</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>JournalName</td>
<td></td>
</tr>
<tr>
<td>LogicalStructures</td>
<td></td>
</tr>
<tr>
<td>Name</td>
<td></td>
</tr>
<tr>
<td>Source</td>
<td></td>
</tr>
<tr>
<td>Title</td>
<td></td>
</tr>
<tr>
<td>Year</td>
<td></td>
</tr>
</tbody>
</table>

Individuals

<table>
<thead>
<tr>
<th>Property</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>abstract</td>
<td>LogicalStructure</td>
</tr>
<tr>
<td>background</td>
<td>LogicalStructure</td>
</tr>
</tbody>
</table>

Datatypes

<table>
<thead>
<tr>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENTITIES</td>
</tr>
<tr>
<td>Literal</td>
</tr>
<tr>
<td>PlainLiteral</td>
</tr>
<tr>
<td>string</td>
</tr>
</tbody>
</table>

SWRL rules

Rule 1: DiscourseElement(? w), Document(? x), Fragment(? z), Person(? y), belong_to(? w, ? z), contains(? x, ? z), written_by(? x, ? y) → write(? y, ? w), write(? y, ? z)

Rule 2: Definiendum(? x), Definiens(? y), Definition(? z), part_of(? x, ? z), part_of(? y, ? z) → define(? y, ? x)

Rule 3: DiscourseElement(? y), Document(? x), cites(? y, ? x) → RelatedWork(? y)

Rule 4: DiscourseElement(? x), DiscourseElement(? y), DiscourseElement(? z), cites(? x, ? y), cites(? y, ? z) → inferred_relation(? x, ? z)

Rule 5: DiscourseElement(? x), DiscourseElement(? y), Fragment(? F x), Fragment(? F y), inferred_relation(? x, ? y), belong_to(? x, ? F x), belong_to(? y, ? F y) → inferred_relation_fragment(? F x, ? F y)
Appendix D

Methodology ontology

Classes

2 independent sample t-test
2_independent_sample_t-test ⊆ Statistical_tests
Accidental sampling
Accidental_sampling ⊆ Nonprobability_sampling
Ad hoc quotas
Ad_hoc_quotas ⊆ Nonprobability_sampling
Analysis of Covariance
Analysis_of_Covariance ⊆ Statistical_tests
Analysis of variance
Analysis_of_variance ⊆ Statistical_tests
Bernoulli sampling
Bernoulli_sampling ⊆ Poisson_sampling
Bernoulli trial
Bernoulli_trial ⊆ Trial
Binomial test
Binomial_test ⊆ Statistical_tests
Canonical correlation
Canonical_correlation ⊆ Statistical_tests
Case study
Case_study ⊆ Nonprobability_sampling
Chi-squared test
Chi_squared_test ⊆ Statistical_tests
Chi Square goodness-of-fit
Chi_Square_goodness-of-fit ⊆ Statistical_tests
Chi square test
Chi_square_test ⊆ Statistical_tests
Cluster sampling
Cluster_sampling ⊆ Random_sampling
Content analysis
Content_analysis ⊆ Methods_techniques
Correlation
Correlation ⊆ Statistical_tests
Cronbach's alpha
Cronbach’s_alpha ⊆ Statistical_tests
Demon algorithm
Demon_algorithm ⊆ Random_sampling
Dependent variable
Dependent_variable ⊆ Variable
Deviant Case
Deviant_Case ⊆ Nonprobability_sampling
Direct interview
Direct_interview ⊆ Interview
Directif interview
Directif_interview ⊆ Interview
Discriminant analysis
Discriminant_analysis ⊆ Statistical_tests
Distance sampling
Distance_sampling ⊆ Random_sampling
Existing data
Existing_data ⊆ Methods_techniques
Experiment
Experiment ≡ Trial
Experiment ⊆ MethodsOntology
Experimental design
Experimental_design ⊆ Methods_techniques
Extraneous variable
Extraneous_variable ⊆ Variable
Factor analysis
Factor_analysis ⊆ Statistical_tests
Factorial ANOVA
Factorial_ANOVA ⊆ Statistical_tests
Fisher’s Exact test
Fisher’s_Exact_test ⊆ Statistical_tests
Focus group
Focus_group ⊆ Methods_techniques
Friedman test
Friedman_test ⊆ Statistical_tests
General Linear Models
General_Linear_Models ⊆ Statistical_tests
Gradsect
Gradsect ⊆ Random_sampling
Historical comparison
Historical_comparison ⊆ Methods_techniques
IRM
IRM ⊆ Methodology_Tools
Implementation
Implementation ⊆ Methods_techniques
Appendix D. Methodology ontology

Independent variable
Independent_variable ⊑ Variable

Interview
Interview ⊑ Methods_techniques

Judgmental sampling
Judgmental_sampling ≡ Purposive_sampling
Judgmental_sampling ⊑ Nonprobability_sampling

Kish grid
Kish_grid ⊑ Random_sampling

Kruskal Wallis
Kruskal_Wallis ⊑ Statistical_tests

Large scale data
Large_scale_data ⊑ Methods_techniques

Latin hypercube sampling
Latin_hypercube_sampling ⊑ Random_sampling

Line intercept sampling
Line-intercept_sampling ⊑ Random_sampling

Log-linear, logistic regression
Log-linear,_logistic_regression ⊑ Statistical_tests

Logistic regression
Logistic_regression ⊑ Statistical_tests

MANOVA
MANOVA ⊑ Statistical_tests

Mann Whitney
Mann_Whitney ⊑ Statistical_tests

Mann Whitney U
Mann_Whitney_U ⊑ Statistical_tests

McNemar
McNemar ⊑ Statistical_tests

Mean
Mean ⊑ Statistical_tests

Mean square weighted deviation
Mean_square_weighted_deviation ⊑ Statistical_tests

Median
Median ⊑ Statistical_tests

Methodology Tools
Methodology_Tools ⊑ MethodsOntology

Methods
Methods ⊑ MethodsOntology

Methods ⊑ ∃ Measure Variable

Methods ⊑ ∃ Use_tools Methodology_Tools

Methods ⊑ ∃ Uses Sampling_techniques

Methods ⊑ MethodsOntology

Methods ⊑ ∃ Uses Methods_techniques

Methods techniques
Methods_techniques ⊑ MethodsOntology

Multiple linear regression
Multiple_linear_regression ⊑ Statistical_tests

Multistage sampling
Multistage_sampling ⊑ Cluster_sampling

Multivariate multiple linear regression
Multivariate_multiple_linear_regression ⊑ Statistical_tests

Non-parametric correlation
Non-parametric_correlation ⊑ Statistical_tests

Non directif interview
Non_directif_interview ⊑ Interview

Nonprobability sampling
Nonprobability_sampling ⊑ Sampling_techniques

Observation
Observation ⊑ Methods_techniques

On line survey
On_line_survey ⊑ Methodology_Tools

One-sample median
One-sample_median ⊑ Statistical_tests

One-sample t-test
One-sample_t-test ⊑ Statistical_tests

One-way ANOVA
One-way_ANOVA ⊑ Statistical_tests

Paired t-test
Paired_t-test ⊑ Statistical_tests

Paper questionnaire
Paper_questionnaire ⊑ Methodology_Tools

Participant observatory
Participant_observatory ⊑ Methods_techniques

Pearson product-moment correlation coefficient
Pearson_product-moment_correlation_coe ⊑ Statistical_tests

Poisson sampling
Poisson_sampling ⊑ Random_sampling

Population
Population ⊑ MethodsOntology

Proportion
Proportion ⊑ Statistical_tests

Purposive sampling
Purposive_sampling ≡ Judgmental_sampling
Purposive_sampling □ Nonprobability_sampling
Qualitative_methods
Qualitative_methods □ Methods
Quantitative_methods
Quantitative_methods □ Methods
Questionnaire
Questionnaire □ Methodology_Tools
Quota_sampling
Quota_sampling □ Nonprobability_sampling
Random_sampling
Random_sampling □ Sampling_techniques
Regression_analysis
Regression_analysis □ Statistical_tests
Repeated_measures_ANOVA
Repeated_measures_ANOVA □ Statistical_tests
Sampling_techniques
Sampling_techniques □ Sample_as_Population
Sampling_techniques □ MethodsOntology
Sampling_techniques □ □ has_category Types_of_variables
Secondary_data
Secondary_data □ Methods_techniques
Semi_directif_interview
Semi_directif_interview □ Interview
Simple_linear_regression
Simple_linear_regression □ Statistical_tests
Simple_random_sample
Simple_random_sample □ Random_sampling
Snowball_sampling
Snowball_sampling □ Nonprobability_sampling
Spearman’s_rank_correlation_coefficient
Spearman’s_rank_correlation_coefficient □ Statistical_tests
Square_root_biased_sampling
Square_root_biased_sampling □ Random_sampling
Statistical_analysis
Statistical_analysis □ Methods_techniques
Statistical_tests
Statistical_tests □ MethodsOntology
Story_life
Story_life □ Methods_techniques
Student’s_t-test
Student’s_t-test □ Statistical_tests
Survey
Survey □ Methods_techniques
Systematic_sampling
Systematic_sampling □ Random_sampling
Time_series_analysis
Time_series_analysis □ Statistical_tests
Trial
Trial □ Experiment
Trial □ MethodsOntology
Types_of_variables
Types_of_variables □ MethodsOntology
Types_of_variables □ □ has_category Variable
Variable
Variable □ □ has_category Types_of_variables
Variable □ □ is_used_in Methods
Variable □ MethodsOntology
Wilcoxon_rank_sum_test
Wilcoxon_rank_sum_test □ Statistical_tests
Wilcoxon_signed_ranks_test
Wilcoxon_signed_ranks_test □ Statistical_tests

Object properties
Create_sample
<http://cui.unige.ch/~deribauh/Methods.owl#Create_sample>

Is_measured_by
<http://cui.unige.ch/~deribauh/Methods.owl#Is_measured_by>

Is_sampled_by
<http://cui.unige.ch/~deribauh/Methods.owl#Is_sampled_by>

Is_subjected_to
<http://cui.unige.ch/~deribauh/Methods.owl#Is_subjected_to>

Measure
<http://cui.unige.ch/~deribauh/Methods.owl#Measure>
Appendix D. Methodology ontology

#Sample_as>
≡ <http://cui.unige.ch/~deribauh/
Methods.owl#Is_sampled_by>

TopObjectProperty

Use tools
∀ Use_tools Methodology_Tools

Used by
∃ Used_by Thing ⊑ Methodology_Tools
Appendix E

Scientific object ontology

**Classes**

**Academic_achievement**

Academic_achievement ⊆ Measured_variables

Academic_achievement ⊆ 3 is_measured_by Kaufman_Tests_of_Educational_Achievement

Academic_achievement ⊆ 3 is_measured_by Wechsler_Individual_Achievement_Test

**Achievement_tests**

Achievement_tests ⊆ 3 measure Academic_achievement

Achievement_tests ⊆ Test

**Afrobarometer**

Afrobarometer ⊆ Test

Afrobarometer ⊆ 3 measure Behaviour_patterns ∩ 3 measure People_Attitudes

**Algorithm**

Algorithm ⊆ Scientific_Object

**Attention**

Attention ⊆ Measured_variables

Attention ⊆ 3 is_measured_by Rey-Osterrieth_Complex_Figure

**Auditory_Processing**

Auditory_Processing ⊆ Measured_variables

**Behaviour_patterns**

Behaviour_patterns ⊆ Measured_variables

Behaviour_patterns ⊆ 3 is_measured_by Afrobarometer ∩ 3 is_measured_by European_Social_Survey

**Comprehension-knowledge**

Comprehension-knowledge ⊆ Measured_variables

**Conceptual_model**

Conceptual_model ⊆ Model

**Contemplative_theory**

Contemplative_theory ⊆ Theory

**Critical_thinking_skills**

Critical_thinking_skills ⊆ Measured_variables

Critical_thinking_skills ⊆ 3 is_measured_by Graduate_record_examination

**Data**

Data ⊆ Scientific_Object

**Dementia**

Dementia ⊆ 3 is_measured_by Mini-Mental_State_Examination

Dementia ⊆ Measured_variables

**Demographic_characteristics**

Demographic_characteristics ⊆ Measured_variables

Demographic_characteristics ⊆ 3 is_measured_by European_Social_Survey

**Differential_Ability_Scales**

Differential_Ability_Scales ⊆ IQ_test

**Economic_model**

Economic_model ⊆ Model

**Equation**

Equation ⊆ Scientific_Object

**European_Social_Survey**

European_Social_Survey ⊆ Test

European_Social_Survey ⊆ 3 measure Behaviour_patterns ∩ 3 measure Demographic_characteristics ∩ 3 measure People_Attitudes ∩ 3 measure People_Beliefs

**Figure**

Figure ⊆ Scientific_Object

**Fluid_Reasoning**

Fluid_Reasoning ⊆ Measured_variables

Fluid_Reasoning ⊆ 3 is_measured_by Woodcock-Johnson_Tests_of_Cognitive Abilities

**Four_Picture_Test**

Four_Picture_Test ⊆ 3 measure People_Attitudes

**General_social_Survey**

General_social_Survey
General_social_Survey $\sqsubseteq$ $\exists$ measure Behaviour_patterns $\sqcap$ $\exists$ measure Demographic_characteristics $\sqcap$ $\exists$ measure People_Beliefs

Graduate_record_examination
Graduate_record_examination $\sqsubseteq$ $\exists$ measure Mathematical_capacities $\sqcap$ $\exists$ measure Quantitative_reasoning $\sqcap$ $\exists$ measure Critical_thinking_skills $\sqcap$ $\exists$ measure Verbal_capacities $\sqcap$ $\exists$ measure Writing_Skills

Graduate_record_examination $\sqsubseteq$ Test

Household_survey
Household_survey $\sqsubseteq$ Scientific_Object

IQ_test
IQ_test $\sqsubseteq$ Test
IQ_test $\sqsubseteq$ $\exists$ measure Intelligence

ISCO-08
ISCO-08 $\sqsubseteq$ Tools_for_deriving_occupational_status

ISCO-68
ISCO-68 $\sqsubseteq$ Tools_for_deriving_occupational_status

ISCO-88
ISCO-88 $\sqsubseteq$ Tools_for_deriving_occupational_status

ISCO-88
ISCO-88 $\sqsubseteq$ Tools_for_deriving_occupational_status

ISKO-88

ISKO-88 $\sqsubseteq$ Tools_for_deriving_occupational_status

Information_model
Information_model $\sqsubseteq$ Model

Intelligence
Intelligence $\sqsubseteq$ Measured_variables
Intelligence $\sqsubseteq$ $\exists$ is_measured_by IQ_test

International_Socio-Economic_Index_of_Occupational_Status
International_Socio-Economic_Index_of_Occupational_Status $\sqsubseteq$ Tools_for_deriving_occupational_status

Kaufman_Test_of_Educational_Achievement
Kaufman_Test_of_Educational_Achievement $\sqsubseteq$ Achievement_tests

Literacy_skills
Literacy_skills $\sqsubseteq$ $\exists$ is_measured_by SAT_reasoning_test
Literacy_skills $\sqsubseteq$ Measured_variables

Logical_model
Logical_model $\sqsubseteq$ Model

Long-term_retrieval
Long-term_retrieval $\sqsubseteq$ $\exists$ is_measured_by Woodcock-Johnson_Tests_of_Cognitive_Abilities

Mathematical_capacities
Mathematical_capacities $\sqsubseteq$ Measured_variables
Mathematical_capacities $\sqsubseteq$ $\exists$ is_measured_by Graduate_record_examination $\sqcap$ $\exists$ is_measured_by Preliminary_SAT_National_Merit_Scholarship_Qualifying_Test

Mathematical_model
Mathematical_model $\sqsubseteq$ Model

Measured_variables
Measured_variables $\sqsubseteq$ Scientific_Object
Measured_variables $\sqsubseteq$ $\exists$ is_measured_by Test

Memory
Memory $\sqsubseteq$ $\exists$ is_measured_by Rey-Osterrieth_Complex_Figure
Memory $\sqsubseteq$ Measured_variables

Mini-Mental_State_Examination
Mini-Mental_State_Examination $\sqsubseteq$ $\exists$ measure Dementia

Mini-Mental_State_Examination $\sqsubseteq$ Test

Model
Model $\sqsubseteq$ Scientific_Object

Otis-Lennon_School_Ability_Test
Otis-Lennon_School_Ability_Test $\sqsubseteq$ IQ_test

Panel_Study_of_Income_Dynamics
Panel_Study_of_Income_Dynamics $\sqsubseteq$ Household_survey

People_Attributes
People_Attributes $\sqsubseteq$ Measured_variables
People_Attributes $\sqsubseteq$ $\exists$ is_measured_by Afrobarometer $\sqcap$ $\exists$ is_measured_by European_Social_Survey $\sqcap$ $\exists$ is_measured_by Four_Picture_Test

People_Beliefs
People_Beliefs $\sqsubseteq$ Measured_variables
People_Beliefs $\sqsubseteq$ $\exists$ is_measured_by European_Social_Survey $\sqcap$ $\exists$ is_measured_by World_Values_Survey

People_values
People_values $\sqsubseteq$ $\exists$ is_measured_by World_Values_Survey
People_values $\sqsubseteq$ Measured_variables

Planning
Planning $\sqsubseteq$ $\exists$ is_measured_by Rey-Osterrieth_Complex_Figure
Planning $\sqsubseteq$ Measured_variables

Preliminary_SAT_National_Merit_Scholarship_Qualifying_Test
Preliminary_SAT_National_Merit
Scholarship_Qualifying_Test □ △ measure Mathematical_capacities ∩ △ measure Verbal_capacities ∩ △ measure Writing_Skills
Preliminary_SAT_National_Merit_Scholarship_Qualifying_Test □ Test
Processing_Speed
Processing_Speed □ △ is_measured_by Woodcock-Johnson_Tests_of_Cognitive_Abilities
Processing_Speed □ Measured_variables
Quantitative_reasoning
Quantitative_reasoning □ △ is_measured_by Graduate_record_examination
Quantitative_reasoning □ Measured_variables
Rational_theory
Rational_theory □ Theory
Rey-Osterrieth_Complex_Figure
Rey-Osterrieth_Complex_Figure □ Test
Rey-Osterrieth_Complex_Figure □ ∃ measure Attention ∩ △ measure Memory ∩ △ measure Planning ∩ △ measure Visuospatial_abilities ∩ △ measure Working_memory
SAT_reasoning_test
SAT_reasoning_test □ Test
SAT_reasoning_test □ △ measure Literacy_skills ∩ △ measure Writing_Skills
Scientific_Object
Short-Term_Memory
Short-Term_Memory □ △ is_measured_by Woodcock-Johnson_Tests_of_Cognitive_Abilities
Short-Term_Memory □ Measured_variables
Social_mode
Social_mode □ Model
Standard_International_Occupational_Scale
Standard_International_Occupational_Scale □ Tools_for_deriving_occupational_status
Stanford-Binet_Intelligence_Scales
Stanford-Binet_Intelligence_Scales □ IQ_test
Statistical_model
Statistical_model □ Model
Table
Table □ Scientific_Object
Test
Test □ ≥ 1 term
Test □ △ measure Measured_variables
Test □ Scientific_Object
Theory
Theory □ Scientific_Object
Tools_for_deriving_occupational_status
Tools_for_deriving_occupational_status □ Scientific_Object
Verbal_capacities
Verbal_capacities □ △ is_measured_by Graduate_record_examination ∪ △ is_measured_by Preliminary_SAT_National_Merit_Scholarship_Qualifying_Test
Verbal_capacities □ Measured_variables
Visuospatial_abilities
Visuospatial_abilities □ △ is_measured_by Rey-Osterrieth_Complex_Figure
Visuospatial_abilities □ Measured_variables
Wechsler_Adult_Intelligence_Scale
Wechsler_Adult_Intelligence_Scale □ IQ_test
Wechsler_Individual_Achievement_Test
Wechsler_Individual_Achievement_Test □ Achievement_tests
Wechsler_Intelligence_Scale_for_Children
Wechsler_Intelligence_Scale_for_Children □ IQ_test
Wechsler_Preschool_and_Primary_Scale_of_Intelligence
Wechsler_Preschool_and_Primary_Scale_of_Intelligence □ IQ_test
Woodcock-Johnson_Tests_of_Cognitive_Abilities
Woodcock-Johnson_Tests_of_Cognitive_Abilities □ IQ_test
Woodcock-Johnson_Tests_of_Cognitive_Abilities □ △ measure Auditory_Processing ∩ △ measure Comprehension-knowledge ∩ △ measure Fluid_Reasoning ∩ △ measure Long-term_retrieval ∩ △ measure Processing_Speed △ △ measure Short-Term_Memory
Working_memory
Working_memory □ Measured_variables
Working_memory □ △ is_measured_by Rey-Osterrieth_Complex_Figure
World_Values_Survey
World_Values_Survey □ Test
World_Values_Survey □ △ measure People_Beliefs ∩ △ measure People_values
Writing_Skills
Writing_Skills □ △ is_measured_by Graduate_record_examination △ △ is_measured_by Preliminary_SAT_National_Merit_Scholarship_Qualifying_Test □ △ is_measured_by
SAT_reasoning_test
Writing_Skills ⊆ Measured_variables

Object properties

is_measured_by
<http://cui.unige.ch/~deribauh/SciObj#measure> ≡ <http://cui.unige.ch/~deribauh/SciObj#is_measured_by>

Data properties

term
Gender studies ontology

Classes

A_little_ketchup_with_my_steak
A_little_ketchup_with_my_steak ⊑ Menstrual_euphemisms
A_snatch_box_decorated_with_red_roses
A_snatch_box_decorated_with_red_roses ⊑ Menstrual_euphemisms
Ability
Ability ⊑ Variable_analyzed
Absent_parent
Absent_parent ⊑ Variable_analyzed
Academia
Academia ⊑ Variable_analyzed
Access_to_power
Access_to_power ⊑ Variable_analyzed
Active
Active ⊑ Variable_analyzed
Adulthood
Adulthood ⊑ Variable_analyzed
African_american
African_american ⊑ Variable_analyzed
Age
Age ⊑ Variable_analyzed
Alcohol
Alcohol ⊑ Variable_analyzed
All_anal_sex_week
All_anal_sex_week ⊑ Menstrual_euphemisms
All_over
All_over ⊑ Menstrual_euphemisms
Alternative_feminism
Alternative_feminism ⊑ Feminism
Amazon
Amazon ⊑ Variable_analyzed
American
American ⊑ Variable_analyzed
Analytical
Analytical ⊑ Variable_analyzed
Anarchist
Anarchist ⊑ Variable_analyzed
Androcentrism
Androcentrism ⊑ Variable_analyzed
Androgyny
Androgyny ⊑ Variable_analyzed
Android_morphology
Android_morphology ⊑ Variable_analyzed
Anglo-american
Anglo-american ⊑ Variable_analyzed
Animal
Animal ⊑ Variable_analyzed
Annual_earnings
Annual_earnings ⊑ Earnings
Anthropology
Anthropology ⊑ Field_of_research
Anthropology
Anthropology ⊑ Variable_analyzed
Anti-discrimination_law
Anti-discrimination_law ⊑ Variable_analyzed
Anti_realism
Anti_realism ⊑ Variable_analyzed
Anticipation_of_retirement
Anticipation_of_retirement ⊑ Variable_analyzed
Antifeminism
Antifeminism ⊑ Variable_analyzed
Anxiety
Anxiety ⊑ Variable_analyzed
Archaeology
Archaeology ⊑ Variable_analyzed
Architecture
Architecture ⊑ Variable_analyzed
Arousal
Arousal ⊑ Variable_analyzed
Art
Art ⊑ Variable_analyzed
Art_history
Art_history ⊑ Field_of_research
Arts_and_crafts_week_at_Panty_Camp
Arts_and_crafts_week_at_Panty_Camp ⊑ Menstrual_euphemisms
Asian
Asian ⊑ Variable_analyzed
Aspiration
Aspiration ⊑ Variable_analyzed
Appendix F. Gender studies ontology

At_high_tide
At_high_tide ⊂ Menstrual_euphemisms

Atheist
Atheist ⊂ Variable_analyzed

Attracting_the_lesbian_vampires
Attracting_the_lesbian_vampires ⊂ Menstrual_euphemisms

Aunt_Aggie
Aunt_Aggie ⊂ Menstrual_euphemisms

Aunt_Dot
Aunt_Dot ⊂ Menstrual_euphemisms

Aunt_Fanny
Aunt_Fanny ⊂ Menstrual_euphemisms

Aunt_Flo’s_in_town
Aunt_Flo’s_in_town ⊂ Menstrual_euphemisms

Aunt_Flo_sent_someone_else_in_her_place
Aunt_Flo_sent_someone_else_in_her_place ⊂ Menstrual_euphemisms

Aunt_Rose_visiting
Aunt_Rose_visiting ⊂ Menstrual_euphemisms

Aunt_Ruby
Aunt_Ruby ⊂ Menstrual_euphemisms

Aunt_from_Reading_was_in_town
Aunt_from_Reading_was_in_town ⊂ Menstrual_euphemisms

Average_annual_earnings
Average_annual_earnings ⊂ Earnings

Average_earning
Average_earning ⊂ Earnings

Back_in_the_saddle_again
Back_in_the_saddle_again ⊂ Menstrual_euphemisms

Barbequed_chicken
Barbequed_chicken ⊂ Menstrual_euphemisms

Battered_Women
Battered_Women ⊂ Variable_analyzed

Behavior
Behavior ⊂ Variable_analyzed

Belief
Belief ⊂ Variable_analyzed

Biais
Biais ⊂ Variable_analyzed

Biological_determinism
Biological_determinism ⊂ Variable_analyzed

Biological_sex
Biological_sex ⊂ Variable_analyzed

Biology
Biology ⊂ Variable_analyzed

Bitchy_witchy_week
Bitchy_witchy_week ⊂ Menstrual_euphemisms

Black
Black ⊂ Variable_analyzed

Black_men
Black_men ⊂ Variable_analyzed

Black_women
Black_women ⊂ Variable_analyzed

Bleedies
Bleedies ⊂ Menstrual_euphemisms

Bleeding_out_my_vagina
Bleeding_out_my_vagina ⊂ Menstrual_euphemisms

Blobbing
Blobbing ⊂ Menstrual_euphemisms

Blobs
Blobs ⊂ Menstrual_euphemisms

Bloody_Mary
Bloody_Mary ⊂ Menstrual_euphemisms

Blow_job_week
Blow_job_week ⊂ Menstrual_euphemisms

Bodily_functions
Bodily_functions ⊂ Menstrual_euphemisms

Boys
Boys ⊂ Variable_analyzed

Business
Business ⊂ Variable_analyzed

Business_environments
Business_environments ⊂ Variable_analyzed

Business_owner
Business_owner ⊂ Variable_analyzed

CEDAW
CEDAW ⊂ Variable_analyzed

Calendar_days
Calendar_days ⊂ Menstrual_euphemisms

Calling_vision_for_my_people
Calling_vision_for_my_people ⊂ Menstrual_euphemisms

Canada
Canada ⊂ Country

Cardiovascular_trials
Cardiovascular_trials ⊂ Variable_analyzed

Career_choice
Career_choice ⊂ Variable_analyzed

Career_path
Career_path ⊂ Variable_analyzed

Career_progression
Career_progression ⊂ Variable_analyzed

Career_relevant
Career_relevant ⊂ Variable_analyzed

Charlie_popped_out_of_the_bush_and_I_need_allies
Charlie_popped_out_of_the_bush_and_I_need_allies Menstrual_euphemisms
Checking_into_the_Red_Roof_Inn Checking_into_the_Red_Roof_Inn Menstrual_euphemisms
Chicana Chicana Variable_analyzed
Child Child Period_of_age Variable_analyzed
Child_bird Child_bird Variable_analyzed
Childcare Childcare Variable_analyzed
Childless Childless Variable_analyzed
Cholesterol Cholesterol Variable_analyzed
Christian Christian Variable_analyzed
Church Church Variable_analyzed
Cinema Cinema Field_of_research
Civil_liberties Civil_liberties Variable_analyzed
Clean-up_on_aisle_one Clean-up_on_aisle_one Menstrual_euphemisms
Clogging_molly Clogging_molly Menstrual_euphemisms
Closed_for_business Closed_for_business Menstrual_euphemisms
Closed_for_maintenance Closed_for_maintenance Menstrual_euphemisms
Closed_shop Closed_shop Menstrual_euphemisms
Code_red Code_red Menstrual_euphemisms
Cohort Cohort Variable_analyzed
Cohort_control Cohort_control Cohort
Cohort_rempacement Cohort_rempacement Cohort
Collective_power Collective_power Variable_analyzed
Colonialism Colonialism Variable_analyzed
Communists_in_the_Summer_House Communists_in_the_Summer_House Menstrual_euphemisms
Compensation_strategie Compensation_strategie Variable_analyzed
Competance Competance Variable_analyzed
Complementarianism Complementarianism Variable_analyzed
Composition_studies Composition_studies Variable_analyzed
Computer_expert Computer_expert Computer_skills
Computer_novice Computer_novice Computer_skills
Computer_skills Computer_skills Variable_analyzed
Condensation Condensation Menstrual_euphemisms
Confidence Confidence Variable_analyzed
Conflict Conflict Variable_analyzed
Conservative Conservative Variable_analyzed
Conservative Conservative Variable_analyzed
Conservative_politic Conservative_politic Variable_analyzed
Constrast_sensitivity_ability Constrast_sensitivity_ability Variable_analyzed
Continental_european_countries Continental_european_countries Country
Control Control Variable_analyzed
Convention_on_the_Elimination_of_All_Forms_of_Discrimination_Against_Women Convention_on_the_Elimination_of_All_Forms_of_Discrimination_Against_Women Variable_analyzed
Corporate_club Corporate_club Variable_analyzed
Country Country Variable_analyzed
Cousin_Claudia Cousin_Claudia Menstrual_euphemisms
Cousin_Tom Cousin_Tom Menstrual_euphemisms
Criminology Criminology Variable_analyzed
Crimson_curse Crimson_curse Menstrual_euphemisms
Crimson_tide Crimson_tide Menstrual_euphemisms
Appendix F. Gender studies ontology

Crimson_wave
Crimson_wave ⊑ Menstrual_euphemisms

Criteria
Criteria ⊑ Variable_analyzed

Cross-disciplinary
Cross-disciplinary ⊑ Variable_analyzed

Cultural
Cultural ⊑ Variable_analyzed

Cultural_belief
Cultural_belief ⊑ Variable_analyzed

Cultural_change
Cultural_change ⊑ Variable_analyzed

Cultural_milieu
Cultural_milieu ⊑ Variable_analyzed

Cultural_schema
Cultural_schema ⊑ Variable_analyzed

Culture
Culture ⊑ Variable_analyzed

Cumulative_average_earning
Cumulative_average_earning ⊑ Earnings

Cumulative_earning
Cumulative_earning ⊑ Earnings

Cyber
Cyber ⊑ Variable_analyzed

Dam_has_burst
Dam_has_burst ⊑ Menstrual_euphemisms

Danish_Act_of_Succession_referendum
Danish_Act_of_Succession_referendum ⊑ Variable_analyzed

Dark_week
Dark_week ⊑ Menstrual_euphemisms

Dead_week
Dead_week ⊑ Menstrual_euphemisms

Death
Death ⊑ Variable_analyzed

Demand_side
Demand_side ⊑ Variable_analyzed

Demographic
Demographic ⊑ Variable_analyzed

Depreciation
Depreciation ⊑ Variable_analyzed

Developed_family
Developed_family ⊑ Variable_analyzed

Deviance
Deviance ⊑ Variable_analyzed

Diabetic
Diabetic ⊑ Variable_analyzed

Difference
Difference ⊑ Variable_analyzed

Disability
Disability ⊑ Variable_analyzed

Discrimination
Discrimination ⊑ Variable_analyzed

Dishonorable_discharge_from
Dishonorable_discharge_from ⊑ the_Uterine_Navy

the_Uterine_Navy
Dishonorable_discharge_from ⊑ the_Uterine_Navy

Dominance
Dominance ⊑ Variable_analyzed

Dracula's_tea_bag
Dracula's_tea_bag ⊑ Menstrual_euphemisms

Dropping_clots
Dropping_clots ⊑ Menstrual_euphemisms

Earning_inequality
Earning_inequality ⊑ Earnings

Earnings
Earnings ⊑ Paid

Earnings ⊑ Variable_analyzed

East
East ⊑ Variable_analyzed

Eco
Eco ⊑ Variable_analyzed

Economic
Economic ⊑ Variable_analyzed

Economics
Economics ⊑ Variable_analyzed

Education
Education ⊑ Variable_analyzed

Effects_on_society
Effects_on_society ⊑ Variable_analyzed

Egalitarianism
Egalitarianism ⊑ Variable_analyzed

Egality
Egality ⊑ Variable_analyzed

Egalitarianism_essentialism
Egalitarianism_essentialism ⊑ Variable_analyzed

Embarrassment
Embarrassment ⊑ Variable_analyzed

Emergent_feminism
Emergent_feminism ⊑ Variable_analyzed

Employer
Employer ⊑ Variable_analyzed

Employment
Employment ⊑ Variable_analyzed

End_of_sentence
End_of_sentence ⊑ Menstrual_euphemisms

Engineering
Engineering ⊑ Variable_analyzed

Entrepreneurs
Entrepreneurs ⊑ Variable_analyzed
Entreprise
Entreprise $\subseteq$ Variable_analyzed

Epicene
Epicene $\subseteq$ Variable_analyzed

Epistemology
Epistemology $\subseteq$ Variable_analyzed

Equal_Pay_Act
Equal_Pay_Act $\subseteq$ Variable_analyzed

Equal_power
Equal_power $\subseteq$ Variable_analyzed

Equality
Equality $\subseteq$ Variable_analyzed

Ethics
Ethics $\subseteq$ Variable_analyzed

Ethnicity
Ethnicity $\subseteq$ Variable_analyzed

European_charter_for_equality_of_women_and_men_in_local_life
European_charter_for_equality_of_women_and_men_in_local_life $\subseteq$ Variable_analyzed

European_countries
European_countries $\subseteq$ Country

European_employment
European_employment $\subseteq$ Variable_analyzed

Evaluation
Evaluation $\subseteq$ Variable_analyzed

Executive_week
Executive_week $\subseteq$ Menstrual_euphemisms

Exercise
Exercise $\subseteq$ Variable_analyzed

Exercise_training
Exercise_training $\subseteq$ Variable_analyzed

Existentialism
Existentialism $\subseteq$ Variable_analyzed

Expenditure
Expenditure $\subseteq$ Variable_analyzed

Exploitation
Exploitation $\subseteq$ Variable_analyzed

Falling_to_the_communists
Falling_to_the_communists $\subseteq$ Menstrual_euphemisms

Family
Family $\subseteq$ Variable_analyzed

Family_attitude
Family_attitude $\subseteq$ Type_of_attitude

Family_income
Family_income $\subseteq$ Variable_analyzed

Family_status
Family_status $\subseteq$ Variable_analyzed

Family_stress
Family_stress $\subseteq$ Variable_analyzed

Family_structure
Family_structure $\subseteq$ Variable_analyzed

Fat
Fat $\subseteq$ Variable_analyzed

Father
Father $\subseteq$ Variable_analyzed

Female
Female $\subseteq$ Variable_analyzed

Female_advantaging
Female_advantaging $\subseteq$ Variable_analyzed

Female_earnings
Female_earnings $\subseteq$ Earnings

Female_economic_activity
Female_economic_activity $\subseteq$ Variable_analyzed

Female_education
Female_education $\subseteq$ Variable_analyzed

Female_occupation
Female_occupation $\subseteq$ Variable_analyzed

Femicide
Femicide $\subseteq$ Variable_analyzed

Feminazi
Feminazi $\subseteq$ Variable_analyzed

Femininity
Femininity $\subseteq$ Variable_analyzed

Feminism
Feminism $\subseteq$ Variable_analyzed

Feminism_in_culture
Feminism_in_culture $\subseteq$ Variable_analyzed

Feminist
Feminist $\subseteq$ Person

Fertility
Fertility $\subseteq$ Variable_analyzed

Field_of_research
Field_of_research $\subseteq$ Person

Film_theory
Film_theory $\subseteq$ Variable_analyzed

Financial
Financial $\subseteq$ Variable_analyzed

Flying_the_red_flag
Flying_the_red_flag $\subseteq$ Menstrual_euphemisms

Focus_group
Focus_group $\subseteq$ Variable_analyzed

Force
Force $\subseteq$ Variable_analyzed

France
France $\subseteq$ Country

French
French $\subseteq$ Variable_analyzed

French_structuralist
French_structuralist $\subseteq$ Variable_analyzed

Friend_from_the_Red_Sea
Appendix F. Gender studies ontology

Friend_from_the_Red_Sea ⊑ Menstrual_euphemisms

Fruit
Fruit ⊑ Variable_analyzed

Full_time
Full_time ⊑ Variable_analyzed

Full_time_worker
Full_time_worker ⊑ Variable_analyzed

Game_Day_for_the_crimson_tide
Game_Day_for_the_crimson_tide ⊑ Menstrual_euphemisms

Gaze
Gaze ⊑ Variable_analyzed

Gender
Gender ⊑ Variable_analyzed

Gender_Equality_Duty_in_Scotland
Gender_Equality_Duty_in_Scotland ⊑ Gender

Gender_Equity_Education_Act
Gender_Equity_Education_Act ⊑ Gender

Gender_Parity
Gender_Parity ⊑ Gender

Gender_apartheid
Gender_apartheid ⊑ Gender

Gender_attitude
Gender_attitude ⊑ Gender

Gender_based
Gender_based ⊑ Gender

Gender_belief
Gender_belief ⊑ Gender

Gender_bias
Gender_bias ⊑ Gender

Gender_difference
Gender_difference ⊑ Gender

Gender_dissociated
Gender_dissociated ⊑ Gender

Gender_egalitarian
Gender_egalitarian ⊑ Gender

Gender_egalitarianism
Gender_egalitarianism ⊑ Gender

Gender_equality
Gender_equality ⊑ Gender

Gender_gap
Gender_gap ⊑ Gender

Gender_history
Gender_history ⊑ Gender

Gender_identity
Gender_identity ⊑ Gender

Gender_inequality
Gender_inequality ⊑ Gender

Gender_issue
Gender_issue ⊑ Gender

Gender_mainstreaming
Gender_mainstreaming ⊑ Gender

Gender_queer
Gender_queer ⊑ Gender

Gender_related
Gender_related ⊑ Gender

Gender_relation
Gender_relation ⊑ Gender

Gender_role
Gender_role ⊑ Gender

Gender_role_attitude
Gender_role_attitude ⊑ Gender

Gender_roles
Gender_roles ⊑ Gender

Gender_scale_trend
Gender_scale_trend ⊑ Gender

Gender_segregation
Gender_segregation ⊑ Gender

Gender_sensitization
Gender_sensitization ⊑ Gender

Gender_studies
Gender_studies ⊑ 3 analyses Race

Gender_studies ⊑ 3 analyses Sexuality

Gender_studies ⊑ 3 analyses Ethnicity

Gender_studies ⊑ 3 analyses Nationality

Gender_studies ⊑ Field_of_research

Gender_studies ⊑ 3 analyses Disability

Gender_variance
Gender_variance ⊑ Gender

Geography
Geography ⊑ Variable_analyzed

Germany
Germany ⊑ Country

Gift-giving
Gift-giving ⊑ Variable_analyzed

Girl_power
Girl_power ⊑ Variable_analyzed

Girl_time
Girl_time ⊑ Menstrual_euphemisms

Girls
Girls ⊑ Variable_analyzed

Global
Global ⊑ Variable_analyzed

Global_Gender_Gap_Report
Global_Gender_Gap_Report ⊑ Variable_analyzed

Got_the_things
Got_the_things ⊑ Menstrual_euphemisms

Governing
Governing ⊑ Variable_analyzed

Great_britain
Great_britain ⊑ Uk

Great_britain ⊑ Country

Group_difference
Group_difference ⊑ Variable_analyzed
<table>
<thead>
<tr>
<th>Term</th>
<th>Analyzed Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gynaecoid_morphology</td>
<td>Variable analyzed</td>
</tr>
<tr>
<td>Gynocentrism</td>
<td></td>
</tr>
<tr>
<td>Habitus</td>
<td>Variable analyzed</td>
</tr>
<tr>
<td>Hammock_season</td>
<td>Menstrual euphemisms</td>
</tr>
<tr>
<td>Happy_Fun_Time</td>
<td>Menstrual euphemisms</td>
</tr>
<tr>
<td>Having_the_painters_in</td>
<td>Menstrual euphemisms</td>
</tr>
<tr>
<td>Headscarf</td>
<td>Variable analyzed</td>
</tr>
<tr>
<td>High_aspiration</td>
<td></td>
</tr>
<tr>
<td>High_educated</td>
<td>Level_of_education</td>
</tr>
<tr>
<td>High_education</td>
<td>Level_of_education</td>
</tr>
<tr>
<td>High_school_girl</td>
<td>Variable analyzed</td>
</tr>
<tr>
<td>High_self-esteem</td>
<td>Self-esteem</td>
</tr>
<tr>
<td>High_status_conditions</td>
<td>Variable analyzed</td>
</tr>
<tr>
<td>Highly_educated</td>
<td>Level_of_education</td>
</tr>
<tr>
<td>History</td>
<td>Field_of_research</td>
</tr>
<tr>
<td>Home</td>
<td>Variable analyzed</td>
</tr>
<tr>
<td>Home_care</td>
<td>Variable analyzed</td>
</tr>
<tr>
<td>Home_time</td>
<td>Variable analyzed</td>
</tr>
<tr>
<td>Homemaking</td>
<td>Variable analyzed</td>
</tr>
<tr>
<td>Hours_of_work</td>
<td>Variable analyzed</td>
</tr>
<tr>
<td>Household</td>
<td>Variable analyzed</td>
</tr>
<tr>
<td>Human_capital</td>
<td>Variable analyzed</td>
</tr>
<tr>
<td>Human_capital_perspective</td>
<td>Variable analyzed</td>
</tr>
<tr>
<td>Human_capital_theory</td>
<td>Variable analyzed</td>
</tr>
<tr>
<td>Human_development</td>
<td>Field_of_research</td>
</tr>
<tr>
<td>Hummer_week</td>
<td>Menstrual euphemisms</td>
</tr>
<tr>
<td>Hunt_for_Red_October</td>
<td>Menstrual euphemisms</td>
</tr>
<tr>
<td>Identity</td>
<td>Variable analyzed</td>
</tr>
<tr>
<td>Identity_formation</td>
<td>Identity</td>
</tr>
<tr>
<td>Ideological_variable</td>
<td>Variable analyzed</td>
</tr>
<tr>
<td>Ideology</td>
<td>Variable analyzed</td>
</tr>
<tr>
<td>Illustrators_for_Gender_Equality</td>
<td>Variable analyzed</td>
</tr>
<tr>
<td>In_season</td>
<td>Menstrual euphemisms</td>
</tr>
<tr>
<td>Income</td>
<td>Variable analyzed</td>
</tr>
<tr>
<td>Independent_women</td>
<td>Variable analyzed</td>
</tr>
<tr>
<td>Individual</td>
<td>Variable analyzed</td>
</tr>
<tr>
<td>Individual_behavior</td>
<td>Variable analyzed</td>
</tr>
<tr>
<td>Individualalist</td>
<td>Variable analyzed</td>
</tr>
<tr>
<td>Institution</td>
<td>Variable analyzed</td>
</tr>
<tr>
<td>Intellectual_ability</td>
<td>Variable analyzed</td>
</tr>
<tr>
<td>Intelligence_quotient</td>
<td>Variable analyzed</td>
</tr>
<tr>
<td>Interdiscipline</td>
<td>Field_of_research</td>
</tr>
<tr>
<td>Internalization</td>
<td>Variable analyzed</td>
</tr>
<tr>
<td>International_Men’s_Day</td>
<td>Variable analyzed</td>
</tr>
<tr>
<td>International_relations</td>
<td>Variable analyzed</td>
</tr>
<tr>
<td>Invisible</td>
<td>Variable analyzed</td>
</tr>
<tr>
<td>Islam</td>
<td>Religion</td>
</tr>
<tr>
<td>Islamic</td>
<td>Variable analyzed</td>
</tr>
</tbody>
</table>
Appendix F. Gender studies ontology

Islamic □ Religion
Jewish
Jewish □ Variable_analyzed
Job
Job □ Occupation
Job □ Variable_analyzed
Kitty has a nosebleed
Kitty has a nosebleed □ Menstrual_euphemisms
Knowledge
Knowledge □ Variable_analyzed
Kyriarchy
Kyriarchy □ Variable_analyzed
LGBT studies
LGBT_studies □ Gender_studies
Labor
Labor □ Variable_analyzed
Labor_economics
Labor_economics □ Variable_analyzed
Labor_force
Labor_force □ Variable_analyzed
Labor_market
Labor_market □ Variable_analyzed
Lady_days
Lady_days □ Menstrual_euphemisms
Lady_in_the_red_dress
Lady_in_the_red_dress □ Menstrual_euphemisms
Language
Language □ Field_of_research
Later_ages
Later_ages □ Variable_analyzed
Law
Law □ Field_of_research
Leak_week
Leak_week □ Menstrual_euphemisms
Leaky_basement
Leaky_basement □ Menstrual_euphemisms
Legal
Legal □ Variable_analyzed
Legal_theory
Legal_theory □ Variable_analyzed
Lesbian
Lesbian □ Variable_analyzed
Lesbianism
Lesbianism □ Variable_analyzed
Less_educated
Less_educated □ Level_of_education
Level_of_competence
Level_of_competence □ Variable_analyzed
Level_of_education
Level_of_education □ Variable_analyzed
Liberal
Liberal □ Variable_analyzed
Liberal_attitude
Liberal_attitude □ Type_of_attitude
Liberal_cohort
Liberal_cohort □ Cohort
Liberal_gender
Liberal_gender □ Variable_analyzed
Liberalism_feminism
Liberalism_feminism □ Variable_analyzed
Liberalizing_trends
Liberalizing_trends □ Variable_analyzed
Lifecycle_pattern
Lifecycle_pattern □ Variable_analyzed
Lifetime_earning
Lifetime_earning □ Variable_analyzed
Lifetime_earning □ Earnings
Lilly_Ledbetter_Fair_Pay_Act
Lilly_Ledbetter_Fair_Pay_Act □ Variable_analyzed
Lining_the_drawers
Lining_the_drawers □ Menstrual_euphemisms
Lipstick
Lipstick □ Variable_analyzed
List_of_gender_equality_lawsuits
List_of_gender_equality_lawsuits □ Variable_analyzed
Literacy_theory
Literacy_theory □ Field_of_research
Literary_criticism
Literary_criticism □ Variable_analyzed
Literature
Literature □ Field_of_research
Little_Red_Riding_Hood
Little_Red_Riding_Hood □ Menstrual_euphemisms
Little_enemy
Little_enemy □ Menstrual_euphemisms
Little_ketchup_with_my_steak,_A
Little_ketchup_with_my_steak,_A □ Menstrual_euphemisms
Location
Location □ Variable_analyzed
Low_education
Low_education □ Level_of_education
Low_self-esteem
Low_self-esteem □ Self-esteem
Low_status_conditions
Low_status_conditions □ Variable_analyzed
Macro_level
Macro_level □ Variable_analyzed
Male
Male □ Variable_analyzed
Male advantage
Male advantage □ Variable_analyzed
Male advantaging
Male advantaging □ Variable_analyzed
Male domination
Male domination □ Variable_analyzed
Male occupation
Male occupation □ Variable_analyzed
Male trends
Male trends □ Variable_analyzed
Man studies
Man studies □ Gender studies
Marginalisation
Marginalisation □ Variable_analyzed
Marital status
Marital status □ Variable_analyzed
Marital stress
Marital stress □ Variable_analyzed
Market economy
Market economy □ Variable_analyzed
Married
Married □ Marital status
Married man
Married man □ Marital status
Married women
Married women ≡ Wife
Married women □ Marital status
Marxist
Marxist □ Variable_analyzed
Masculinity
Masculinity □ Variable_analyzed
Masculism
Masculism □ Variable_analyzed
Massacre at the Y
Massacre at the Y □ Menstrual euphemisms
Massacre at the beef curtain
Massacre at the beef curtain □ Menstrual euphemisms
Material
Material □ Variable_analyzed
Mathematical ability
Mathematical ability ≡ Mathematical skills
Mathematical ability □ Variable_analyzed
Mathematical performance
Mathematical performance □ Variable_analyzed
Mathematical skills
Mathematical skills ≡ Mathematical ability
Mathematical skills □ Variable_analyzed
Mathematics
Mathematics □ Variable_analyzed
Matriarchy
Matriarchy □ Variable_analyzed
Matriarchy Women’s studies
Matriarchy Women’s studies □ Variable_analyzed
Matrimonial
Matrimonial □ Variable_analyzed
Matriname
Matriname □ Variable_analyzed
Mechanical tasks
Mechanical tasks □ Variable_analyzed
Medecine
Medecine □ Field of research
Media studies
Media studies □ Field of research
Medical
Medical □ Variable_analyzed
Mediterranean diet
Mediterranean diet □ Variable_analyzed
Medium education
Medium education □ Level of education
Men
Men □ Variable_analyzed
Men’s rights
Men’s rights □ Variable_analyzed
Men’s studies
Men’s studies □ Variable_analyzed
Menial position
Menial position □ Variable_analyzed
Menopause
Menopause □ Variable_analyzed
Menopause
Menopause □ Menstrual euphemisms
Menstrual
Menstrual □ Variable_analyzed
Menstrual euphemisms
Menstrual euphemisms □ Variable_analyzed
Menstrual taboo
Menstrual taboo □ Variable_analyzed
Menstruation
Menstruation ≡ Period
Menstruation □ Variable_analyzed
Mental ability
Mental ability □ Variable_analyzed
Middle class
Middle class □ Social class
Migrant
Migrant □ Variable_analyzed
Misandry
Misandry □ Variable_analyzed
Misogyny
Misogyny □ Variable_analyzed
Miss Scarlett’s come home to Tara
Miss_Scarlett’s_come_home_to_Tara ⊆ Menstrual_euphemisms
Mixed-sex_education
Mixed-sex_education ⊆ Variable_analyzed
Mobility
Mobility ⊆ Variable_analyzed
Moderate_daily
Moderate_daily ⊆ Variable_analyzed
Moderately_educated
Moderately_educated ⊆ Level_of_education
Moderately_educated
Moderately_educated ⊆ Level_of_education
Monkey_has_a_nosebleed
Monkey_has_a_nosebleed ⊆ Menstrual_euphemisms
Monthlies
Monthlies ⊆ Menstrual_euphemisms
Monthlies
Monthlies ⊆ Menstrual_euphemisms
Monthly_bill
Monthly_bill ⊆ Menstrual_euphemisms
Monthly_bill
Monthly_bill ⊆ Menstrual_euphemisms
Monthly_curse
Monthly_curse ⊆ Menstrual_euphemisms
Monthly_evacuations
Monthly_evacuations ⊆ Menstrual_euphemisms
Monthly_time
Monthly_time ⊆ Menstrual_euphemisms
Monthly_trouble
Monthly_trouble ⊆ Menstrual_euphemisms
Monthly_visitor
Monthly_visitor ⊆ Menstrual_euphemisms
Moodiness
Moodiness ⊆ Variable_analyzed
Moody_monthly
Moody_monthly ⊆ Menstrual_euphemisms
Moon_time
Moon_time ⊆ Menstrual_euphemisms
Mortality
Mortality ⊆ Variable_analyzed
Moses_parting_the_Red_Sea
Moses_parting_the_Red_Sea ⊆ Menstrual_euphemisms
Mother
Mother ⊆ Variable_analyzed
Mother_Nature’s_gift
Mother_Nature’s_gift ⊆ Menstrual_euphemisms
Mother_nature’s_curse
Mother_nature’s_curse ⊆ Menstrual_euphemisms
Motherhood_gaps
Motherhood_gaps ⊆ Variable_analyzed
Mucking_out
Mucking_out ⊆ Menstrual_euphemisms
Mudslide_in_Crotch_Canyon
Mudslide_in_Crotch_Canyon ⊆ Menstrual_euphemisms
Mutual_concession
Mutual_concession ⊆ Variable_analyzed
My_Friends
My_Friends ⊆ Menstrual_euphemisms
My_cup_of_joy_is_overflowing
My_cup_of_joy_is_overflowing ⊆ Menstrual_euphemisms
My_little_friend_came_to_visit
My_little_friend_came_to_visit ⊆ Menstrual_euphemisms
My_red-headed_cousin_from_Virginia
My_red-headed_cousin_from_Virginia ⊆ Menstrual_euphemisms
My_red-headed_cousin_from_down_south
My_red-headed_cousin_from_down_south ⊆ Menstrual_euphemisms
My_steak_is_raw
My_steak_is_raw ⊆ Menstrual_euphemisms
Myocardical_infarction
Myocardical_infarction ⊆ Variable_analyzed
Mythology
Mythology ⊆ Variable_analyzed
National_currency
National_currency ⊆ Variable_analyzed
National_identity
National_identity ⊆ Identity
Nationality
Nationality ⊆ Variable_analyzed
Negative_feedback
Negative_feedback ⊆ Variable_analyzed
Negative_stereotype
Negative_stereotype ⊆ Variable_analyzed
Neo
Neo ⊆ Variable_analyzed
Netherlands
Netherlands ⊆ Country
Networked_feminism
Networked_feminism ⊆ Variable_analyzed
New_Postcolonial
New_Postcolonial \(\subseteq\) Variable_analyzed
Non-mothers
Non-mothers \(\subseteq\) Variable_analyzed
Nordic_countries
Nordic_countries \(\subseteq\) Country
Nosebleed_in_Australia
Nosebleed_in_Australia \(\subseteq\) Menstrual_euphemisms
Number_of_children
Number_of_children \(\subseteq\) Variable_analyzed
Number_of_hours
Number_of_hours \(\subseteq\) Variable_analyzed
Obstacles
Obstacles \(\subseteq\) Variable_analyzed
Occupation
Occupation \(\equiv\) Job
Occupation \(\subseteq\) Variable_analyzed
Occupational_prestige
Occupational_prestige \(\subseteq\) Variable_analyzed
Occupational_segregation
Occupational_segregation \(\subseteq\) Variable_analyzed
Occupational_status
Occupational_status \(\subseteq\) Variable_analyzed
Oil_change
Oil_change \(\subseteq\) Menstrual_euphemisms
Older_women
Older_women \(\subseteq\) Women_age
On_the_blob
On_the_blob \(\subseteq\) Menstrual_euphemisms
On_the_rag
On_the_rag \(\subseteq\) Menstrual_euphemisms
Ordering_l'Omelette_Rouge
Ordering_l'Omelette_Rouge \(\subseteq\) Menstrual_euphemisms
Out_of_commission
Out_of_commission \(\subseteq\) Menstrual_euphemisms
Out_of_order
Out_of_order \(\subseteq\) Menstrual_euphemisms
Outcome
Outcome \(\equiv\) Paid
Outcome \(\subseteq\) Variable_analyzed
Outside_work
Outside_work \(\subseteq\) Variable_analyzed
Pad_straddling
Pad_straddling \(\subseteq\) Menstrual_euphemisms
Paid
Paid \(\equiv\) Earnings
Paid \(\equiv\) Outcome
Paid \(\equiv\) Wage
Paid \(\subseteq\) Variable_analyzed
Painters_are_in
Painters_are_in \(\subseteq\) Menstrual_euphemisms
Panty_line
Panty_line \(\subseteq\) Menstrual_euphemisms
Panty_painting
Panty_painting \(\subseteq\) Menstrual_euphemisms
Panty_shields_up_Captain!
Panty_shields_up_Captain! \(\subseteq\) Menstrual_euphemisms
Parent
Parent \(\subseteq\) Variable_analyzed
Part_time
Part_time \(\subseteq\) Variable_analyzed
Participation_rates
Participation_rates \(\subseteq\) Variable_analyzed
Particular事业
Particular事业 \(\subseteq\) Variable_analyzed
Patriarchy
Patriarchy \(\subseteq\) Variable_analyzed
Pattern
Pattern \(\subseteq\) Variable_analyzed
Paycheck_Fairness_Act
Paycheck_Fairness_Act \(\subseteq\) Variable_analyzed
Performance
Performance \(\subseteq\) Variable_analyzed
Performance_expectation
Performance_expectation \(\subseteq\) Performance
Period
Period \(\equiv\) Menstruation
Period \(\subseteq\) Variable_analyzed
Period_of_age
Period_of_age \(\subseteq\) Variable_analyzed
Periodical
Periodical \(\subseteq\) Menstrual_euphemisms
Person
Person \(\subseteq\) ~ Variable_analyzed
Person \(\subseteq\) School_of_thought
Person \(\subseteq\) Field_of_research
Philosophy
Philosophy \(\subseteq\) Variable_analyzed
Philosophy
Philosophy \(\subseteq\) Field_of_research
Physical_danger
Physical_danger \(\subseteq\) Variable_analyzed
Physical_inactivits
Physical_inactivits \(\subseteq\) Variable_analyzed
Pink_Light
Pink_Light \(\subseteq\) Menstrual_euphemisms
Planting_cotton
Planting_cotton \(\subseteq\) Menstrual_euphemisms
Playground’s_muddy
Playground’s_muddy \(\subseteq\) Menstrual_euphemisms
Playing_banjo_in_Sgt.
Zygote’s Ragtime Band
Playing banjo in Sgt. Zygote’s Ragtime Band
Menstrual euphemisms
Policy
Policy in Variable_analyzed
Policy
Political
Political in Variable_analyzed
Political activism
Political activism in Variable_analyzed
Political conservatism
Political conservatism in Variable_analyzed
Political ecology
Political ecology in Variable_analyzed
Political identity
Political identity in Variable_analyzed
Political liberalism
Political liberalism in Variable_analyzed
Political party identification
Political party identification in Variable_analyzed
Political science
Political science in Field_of_research
Political theory
Political theory in Variable_analyzed
Poor diet
Poor diet in Variable_analyzed
Poorly educated
Poorly educated in Level_of_education
Popular culture
Popular culture in Variable_analyzed
Population
Population in Variable_analyzed
Positional power
Positional power in Variable_analyzed
Positive emotion
Positive emotion in Variable_analyzed
Positive stereotype
Positive stereotype in Variable_analyzed
Postfeminism
Postfeminism in Variable_analyzed
Postmodern
Postmodern in Variable_analyzed
Postpone
Postpone in Variable_analyzed
Poststructural
Poststructural in Variable_analyzed
Potty parity
Potty parity in Variable_analyzed
Power
Power in Variable_analyzed
Power relation
Power relation in Variable_analyzed
Power resource
Power resource in Variable_analyzed
Powerful position
Powerful position in Variable_analyzed
Preference
Preference in Variable_analyzed
Preschool
Preschool in Variable_analyzed
Preteen
Preteen in Period_of_age
Private sector
Private sector in Variable_analyzed
Pro-feminism
Pro-feminism in Variable_analyzed
Pro-life
Pro-life in Variable_analyzed
Procreation
Procreation in Variable_analyzed
Professional managerial status
Professional managerial status in Variable_analyzed
Professional position
Professional position in Variable_analyzed
Promotion
Promotion in Variable_analyzed
Protestant
Protestant in Variable_analyzed
Proto
Proto in Variable_analyzed
Psychanalysis
Psychanalysis in Field_of_research
Psychoanalysis
Psychoanalysis in Variable_analyzed
Psychology
Psychology in Variable_analyzed
Psychology
Psychology in Field_of_research
Psychosocial stress
Psychosocial stress in Variable_analyzed
Quaker Testimony of Equality
Quaker Testimony of Equality in Variable_analyzed
Quiff cry
Quiff cry in Menstrual euphemisms
Race
Race in Variable_analyzed
Race sex
Race sex in Variable_analyzed
Radical
Radical in Variable_analyzed
Rag
Menstrual euphemisms

Rag

Ragtime

Raining down South

Raising the Japanese flag

Rebooting the Ovarian Operating System

Red Beard's toothpaste

Red Sox are in town

Red dollar days

Red flag

Red flag is up

Red letter day

Red light district

Red moon rising

Red ribbon week

Red rider's revenge

Red river blues

Red scare

Red snow

Red special

Red thread in the sowing machine

Red wings

Red witch

Relative competence

Religion

Religion affiliation

Religion attendance

Religious

Repeated agreements

Reproduction of inequality

Researcher

Resource

Revisionist

Riding the cotton bicycle

Riding the cotton cowboy

Riding the cotton pony

Riding the crimson tide

Riding the crimson wave

Riding the red ball special

Right to equal protection

Risk factor

Role

Saddling Old Rusty
Saddling_Old_Rusty ⊆ Menstrual_euphemisms
Sanitary_napkin
Sanitary_napkin ⊆ Variable_analyzed
Santa’s bring the presents
Santa’s bring the presents ⊆ Menstrual_euphemisms
Scarlet_letter
Scarlet_letter ⊆ Menstrual_euphemisms
School_of_thought
School_of_thought ⊆ ¬ Person
Science
Science ⊆ Variable_analyzed
Science_fiction
Science_fiction ⊆ Variable_analyzed
Scientist
Scientist ⊆ Variable_analyzed
Secure_relationship
Secure_relationship ⊆ Variable_analyzed
Seeing_red
Seeing_red ⊆ Menstrual_euphemisms
Self-employment
Self-employment ⊆ Employment
Self-esteem
Self-esteem ⊆ Variable_analyzed
Self_assessment
Self_assessment ⊆ Variable_analyzed
Self_evaluate_threat
Self_evaluate_threat ⊆ Variable_analyzed
Self_perception
Self_perception ⊆ Variable_analyzed
Senior_position
Senior_position ⊆ Variable_analyzed
Separatist
Separatist ⊆ Variable_analyzed
Serving_up_the_womb_steak_medium_rare
Serving_up_the_womb_steak_medium_rare ⊆ Menstrual_euphemisms
Sex
Sex ⊆ Variable_analyzed
Sex_based_social_structures
Sex_based_social_structures ⊆ Variable_analyzed
Sex_positive
Sex_positive ⊆ Variable_analyzed
Sex_and_gender_distinction
Sex_and_gender_distinction ⊆ Variable_analyzed
Sex_categorisation
Sex_categorisation ⊆ Variable_analyzed
Sex_composition
Sex_composition ⊆ Variable_analyzed
Sex_criteria
Sex_criteria ⊆ Variable_analyzed
Sex_difference
Sex_difference ⊆ Variable_analyzed
Sex_hormones
Sex_hormones ⊆ Variable_analyzed
Sex_segregation
Sex_segregation ⊆ Variable_analyzed
Sexism
Sexism ⊆ Variable_analyzed
Sexology
Sexology ⊆ Variable_analyzed
Sexual_attractiveness
Sexual_attractiveness ⊆ Variable_analyzed
Sexual_harassment
Sexual_harassment ⊆ Variable_analyzed
Sexual_identity
Sexual_identity ⊆ Identity
Sexual_identity
Sexual_identity ⊆ Variable_analyzed
Sexual_reproduction
Sexual_reproduction ⊆ Variable_analyzed
Sexual_tolerance
Sexual_tolerance ⊆ Variable_analyzed
Sexuality
Sexuality ⊆ Variable_analyzed
Shakers
Shakers ⊆ Variable_analyzed
Shared_Earning
Shared_Earning ⊆ Variable_analyzed
Shared_Parenting_Marriage
Shared_Parenting_Marriage ⊆ Variable_analyzed
Shark_bait
Shark_bait ⊆ Menstrual_euphemisms
Shift
Shift ⊆ Variable_analyzed
Shifting
Shifting ⊆ Variable_analyzed
Sitting_on_a_sock
Sitting_on_a_sock ⊆ Menstrual_euphemisms
Small_buisness
Small_buisness ⊆ Business
Smoking
Smoking ⊆ Variable_analyzed
Snatchbox_decorated_with_red_roses
Snatchbox_decorated_with_red_roses ⊆ Menstrual_euphemisms
Social
Social ⊆ Variable_analyzed
Social_attitude
Social_attitude ⊆ Variable_analyzed
Social_capital
Social_capital ⊆ Variable_analyzed
Appendix F. Gender studies ontology

Teacher
Teacher ⊑ Variable_analyzed
Technical_difficulties
Technical_difficulties ⊑ Menstrual_euphemisms
Technoscience
Technoscience ⊑ Variable_analyzed
Teenage
Teenage ⊑ Period_of_age
Test_of_mental_ability
Test_of_mental_ability ⊑ Variable_analyzed
That_time_of_the_month
That_time_of_the_month ⊑ Menstrual_euphemisms
Thealogy
Thealogy ⊑ Variable_analyzed
Theology
Theology ⊑ Variable_analyzed
Theorical_argument
Theorical_argument ⊑ Variable_analyzed
Theory
Third_world
Third_world ⊑ Variable_analyzed
Threat
Threat ⊑ Variable_analyzed
Threat_of_social_devaluation
Threat_of_social_devaluation ⊑ Variable_analyzed
Time.Of.The.Month
Time.Of.The.Month ⊑ Menstrual_euphemisms
Time.to.change.the.filter
Time.to.change.the.filter ⊑ Menstrual_euphemisms
Tin.roof's.rusted
Tin.roof's.rusted ⊑ Menstrual_euphemisms
Title.IX_of_the_Education.Amendments
Title.IX_of_the_Education.Amendments ⊑ Variable_analyzed
Tomato.boat.has.come.in
Tomato.boat.has.come.in ⊑ Menstrual_euphemisms
Too.wet.to.plow
Too.wet.to.plow ⊑ Menstrual_euphemisms
Train.wreck
Train.wreck ⊑ Menstrual_euphemisms
Trans
Trans ⊑ Variable_analyzed
Transgender
Transgender ⊑ Variable_analyzed
Transition
Transition ⊑ Variable_analyzed
Transnational
Transnational ⊑ Variable_analyzed
Trolling_for_vampires
Trolling_for_vampires ⊑ Menstrual_euphemisms
Type_of_attitude
Type_of_attitude ⊑ Variable_analyzed
U.s.society
U.s.society ⊑ Variable_analyzed
Uk
Uk ≡ Great_britain
Uk ⊑ Country
Uniform_civil_code
Uniform_civil_code ⊑ Variable_analyzed
United.Nations_Security_Council_Resolution
United.Nations_Security_Council_Resolution ⊑ Variable_analyzed
United_states
United_states ⊑ Country
Up_on_blocks
Up_on_blocks ⊑ Menstrual_euphemisms
Vampire's_bag_lunch
Vampire's_bag_lunch ⊑ Menstrual_euphemisms
Variable_analyzed
Variable_analyzed ⊑ Person
Vegetable
Vegetable ⊑ Variable_analyzed
Verbal
Verbal ⊑ Variable_analyzed
Verbal_ability
Verbal_ability ≡ Verbal_skills
Verbal_ability ⊑ Variable_analyzed
Verbal_skills
Verbal_skills ≡ Verbal_ability
Verbal_skills ⊑ Variable_analyzed
Vigorous.exercise
Vigorous.exercise ⊑ Variable_analyzed
Visit_from.Cap'n.Bloodsnatch
Visit_from.Cap'n.Bloodsnatch ⊑ Menstrual_euphemisms
Visit_from_the/Cardinal
Visit_from_the/Cardinal ⊑ Menstrual_euphemisms
Visitation
Visitation ⊑ Menstrual_euphemisms
Vulnerability
Vulnerability ⊑ Variable_analyzed
Wage
Wage ≡ Paid
Wage ⊑ Variable_analyzed
War in Virginia
War in Virginia ⊑ Menstrual euphemisms

War on Women
War on Women ⊑ Variable analyzed

Warm relationship
Warm relationship ⊑ Variable analyzed

Wealthy
Wealthy ⊑ Variable analyzed

Weeping womb
Weeping womb ⊑ Menstrual euphemisms

Weight loss
Weight loss ⊑ Variable analyzed

West
West ⊑ Variable analyzed

White
White ⊑ Variable analyzed

White men
White men ⊑ Variable analyzed

White women
White women ⊑ Variable analyzed

Wife
Wife ≡ Married women
Wife ⊑ Variable analyzed

Womanism
Womanism ⊑ Variable analyzed

Women
Women ⊑ Variable analyzed

Women’s Equality Day
Women’s Equality Day ⊑ Variable analyzed

Women’s Petition to the National Assembly
Women’s Petition to the National Assembly ⊑ Variable analyzed

Women’s health
Women’s health ⊑ Variable analyzed

Women’s rights
Women’s rights ⊑ Variable analyzed

Women age
Women age ⊑ Variable analyzed

Women in Islam
Women in Islam ⊑ Variable analyzed

Women movement
Women movement ⊑ Variable analyzed

Women studies
Women studies ⊑ Gender studies

Work
Work ⊑ Variable analyzed

Work family conflict
Work family conflict ⊑ Variable analyzed

Work relationship
Work relationship ⊑ Variable analyzed

Work time
Work time ⊑ Variable analyzed

Working full time
Working full time ⊑ Work time

Working hours
Working hours ⊑ Variable analyzed

Working mother
Working mother ⊑ Variable analyzed

Working part time
Working part time ⊑ Work time

Working women
Working women ⊑ Variable analyzed

Wound of Eve that never heals
Wound of Eve that never heals ⊑ Menstrual euphemisms

Wounded clam
Wounded clam ⊑ Menstrual euphemisms

Year of experience
Year of experience ⊑ Variable analyzed

Younger women
Younger women ⊑ Women age

female informants
female informants ⊑ Menstrual euphemisms

female visitor
female visitor ⊑ Menstrual euphemisms

logrolling strategy
logrolling strategy ⊑ Variable analyzed

physiology
physiology ⊑ Variable analyzed

psychological gender
psychological gender ⊑ Variable analyzed

self identifies
self identifies ⊑ Variable analyzed

stereotypic
stereotypic ⊑ Variable analyzed

tampon
tampon ⊑ Menstrual euphemisms

the rag
the rag ⊑ Menstrual euphemisms

third gender
third gender ⊑ Variable analyzed

transsexualism
transsexualism ⊑ Variable analyzed

Object properties

analyses
<http://cui.unige.ch/~deribauh/genderStudies.owl#analyses> ⊑ <http://cui.unige.ch/~deribauh/genderStudies.owl#is_analyzed_by>

belong to school of thought
Appendix F. Gender studies ontology

```xml
<http://cui.unige.ch/>
  "deribauh/genderStudies.owl#contains_person" ⊑ <http://cui.unige.ch/>
  "deribauh/genderStudies.owl#belong_to_school_of_thought>-
contains_person
<http://cui.unige.ch/>
  "deribauh/genderStudies.owl#contains_person" ⊑ <http://cui.unige.ch/>
  "deribauh/genderStudies.owl#belong_to_school_of_thought>-
has_field_of_research
<http://cui.unige.ch/>
  "deribauh/genderStudies.owl#has_field_of_research>-
is_analyzed_by
<http://cui.unige.ch/>
  "deribauh/genderStudies.owl#is_analyzed_by>-
is_studied_by
<http://cui.unige.ch/>
  "deribauh/genderStudies.owl#is_studied_by>-
make_research_in
<http://cui.unige.ch/>
  "deribauh/genderStudies.owl#make_research_in>-
```

Data properties

- last_name
- name
  ⊑ topDataProperty
- term
  ⊑ topDataProperty
G.1 Macro common to all Jape rules

Phase: findings
Input: Token authorNNP
Options: control = appelt
Macro: NOUN
(
{Token.category == "NN"} | 
{Token.category == "NNP"} | 
{Token.category == "NNPS"} | 
{Token.category == "NNS"} | 
{Token.category == "NP"} | 
{Token.category == "NPS"} |
)

Macro: ADVERB
(
{Token.category == "RB"} | 
{Token.category == "RBR"} | 
{Token.category == "RBS"} |
)

Macro: DETERMINER
(
{Token.category == "CD"} | 
{Token.category == "DT"} | 
{Token.category == "PDT"} | 
{Token.category == "CC"} |
)

Macro: ADJECTIVE
(
{Token.category == "JJ"} | 
{Token.category == "JJR"} | 
{Token.category == "JJS"} | 
{Token.category == "JJSS"} |
)

Macro: PRONOUN
(
{Token.category == "PP"} | 
{Token.category == "PRP"} | 
{Token.category == "PRP$"} | 
{Token.category == "PRP$"} |
)

Macro: PUNCT
(
{Token.category == ","} |
G.2 Jape rules Findings

Phase: findings
Input: Token authorN
Options: control = appelt
Macro: RESULT

{Token.string=="result"} | {Token.string=="results"}
{Token.string=="Result"} | {Token.string=="Results"}
G.2. Jape rules Findings

Macro: EFFECT
( {Token.string="effect"} | {Token.string="effects"} )

Macro: STUDY
( {Token.string="study"} | {Token.string="studies"} )

Macro: RESEARCH
( {Token.string="research"} | {Token.string="researches"} )

Macro: WORK
( {Token.string="work"} | {Token.string="works"} )

Macro: CONFIRM
( {Token.string="Confirming"} | {Token.string="confirming"} | {Token.string="confirm"} | {Token.string="confirms"} | {Token.string="confirmed"} )

Macro: CONSISTENT
( {Token.string="Consistent"} | {Token.string="consistent"} )

Macro: ARGUE
( {Token.string="argues"} | {Token.string="argue"} | {Token.string="argued"} )

Macro: CONCLUDE
( {Token.string="concludes"} | {Token.string="conclude"} | {Token.string="concluded"} )

Macro: DEMONSTRATE
( {Token.string="demonstrate"} |
{Token.string=="demonstrates"} | {Token.string=="demonstrates"}

Macro: SHOW
{
{Token.string=="show"} | {Token.string=="shown"} | {Token.string=="shows"}
}

Macro: FOUND
{
{Token.string=="found"} | {Token.string=="find"} | {Token.string=="finds"}
}

Macro: SIGN
{
{Token.string=="significantly"} | {Token.string=="significant"} | {Token.string=="insignificant"}
}

Macro: ARTICLE
{
{Token.string=="articles"} | {Token.string=="article"} | {Token.string=="paper"} | {Token.string=="papers"} | {Token.string=="document"} | {Token.string=="documents"}
}

Macro: CORRELATION
{
{Token.string=="correlation"} | {Token.string=="correlated"}
}

Macro: FINDINGS
{
{Token.string=="findings"} | {Token.string=="finding"
}

Macro: DISCUSS
{
{Token.string=="discuss"} | {Token.string=="discussed"
}

Macro: INSUMCONCLUSION
{
((Token.string=="in") | (Token.string=="In"){Token.string=="conclusion"}) | 
((Token.string=="in") | (Token.string=="In"){Token.string=="sum"}) | 
((Token.string=="to") | (Token.string=="To"){Token.string=="conclude"}) |
G.2. Jape rules Findings

(((Token.string=="in") | {Token.string=="In"}) {Token.string=="summary"}) | 
(((Token.string=="to") | {Token.string=="To"}) {Token.string=="sum"} {Token.string=="up"}) | 
(((Token.string=="As") | {Token.string=="as"}) {Token.string=="a"}) {Token.string=="result"}) | 
(((Token.string=="The") | {Token.string=="the"}) {Token.string=="most"} {Token.string=="likely"} {Token.string=="explanation"}) | 
(((Token.string=="The") | {Token.string=="the"}) {Token.string=="lack"} {Token.string=="of"} {Token.string=="any"} {Token.string=="statistically"} {Token.string=="significant"} {Token.string=="effect"} )

Macro: APPROACH

{Token.string=="approach"} | {Token.string=="approaches"}

Rule: assumeFindings
Priority:20
( {NOUN | PRONOUN | {Token.kind==word,Token.category==RB}}?
((Token.string=="assumes") | {Token.string=="assume"} | {Token.string=="assumed"})
((Token.string=="The") | {Token.string=="the"}) {Token.string=="likely"} {Token.string=="explanation"}) | 
((Token.string=="The") | {Token.string=="the"}) {Token.string=="lack"} {Token.string=="of"} {Token.string=="any"} {Token.string=="statistically"} {Token.string=="significant"} {Token.string=="effect"} )

Rule: resultFindings
Priority:20
( {DETERMINER} | {ADJECTIVE} | {NOUN})?
(RESULT)
(PUNCT)?
((!Token.string=="section") | {DETERMINER} | {Token.kind==word,Token.category=="IN"} | {Token.string=="likely"} | {Token.string=="effect"})

Rule: studyFindings
Priority:20
( {DETERMINER} | {ADJECTIVE} | {Token.string=="likely"} | {Token.string=="effect"})

Rule: confirmFindings
Priority: 20
{
(CONFIRM)
(DETERMINER | {Token.kind==word,Token.category==IN} | ADVERB)
(NOUN | ADJECTIVE | ADVERB | DETERMINER | PRONOUN)
}:if
--> :if.confirmFind={rule="confirmFindings"}

Rule: consistentFindings
Priority: 20
{
(CONSISTENT)
{Token.kind==word,Token.category==IN}
(NOUN | ADJECTIVE | ADVERB | DETERMINER | PRONOUN)
}:if
--> :if.consistentFind={rule="consistentFindings"}

Rule: argueFindings
Priority: 20
{
(ARGUE)
({Token.kind==word,Token.category==IN} | DETERMINER)
(NOUN | ADJECTIVE | ADVERB | DETERMINER | PRONOUN)
}:if
--> :if.argueFind={rule="argueFindings"}

Rule: concludeFindings
Priority: 20
{
(CONCLUDE)
({Token.kind==word,Token.category==IN} | DETERMINER)
(NOUN | ADJECTIVE | ADVERB | DETERMINER | PRONOUN)
}:if
-->
G.2. Jape rules Findings

:if.concludeFind={rule="concludeFindings"}

Rule: showFindings
Priority:20
{
(NOUN | {Token.kind==word,Token.category==VBZ} | PUNCT | DETERMINER)
({Token})?
(SHOW)
({Token.kind==word,Token.category==IN} | DETERMINER | {Token.kind==word,Token.category==TO} | PUNCT)
}:

:if.showFind={rule="showFindings"}

Rule: foundFindings
Priority:20
{
(PRONOUN | NOUN | PUNCT | {Token.kind==word,Token.category==CC} | {Token.kind==word,Token.category==VBZ} | {Token.kind==word,Token.category==VBD})
(FOUND)
({Token.kind==word,Token.category==IN} | DETERMINER | PUNCT | {Token.kind==word,Token.category==WRB})
}:if

:if.foundFind={rule="foundFindings"}

Rule: significantFindings
Priority:20
{
(NOUN | DETERMINER)
({Token.string="made"} | {Token.kind==word,Token.category==VBZ} | {Token.kind==word,Token.category==VBP})?
(SIGN)
({Token.kind==word,Token.category==IN} | DETERMINER | PUNCT | {Token.kind==word,Token.category==WRB} | NOUN | ADJECTIVE)
}:if

:if.significantFind={rule="significantFindings"}

Rule: meanFindings
Priority:20
{
(NOUN | DETERMINER | ADJECTIVE | PUNCT)
\{\text{Token.string}=="mean"\}\text{ (NOUN} \mid \text{ADJECTIVE} \mid \{\text{Token.kind==number,Token.category==CC}\} \mid \{\text{Token.kind==word,Token.category==IN}\}\}\text{):if} \\
\text{if.meanFind} = \{\text{rule}=="meanFindings"\}

\text{Rule: correlatedFindings} \\
\text{Priority:20} \\
\{\{\text{Token.kind==word,Token.category==VBZ}\} \mid \{\text{Token.kind==word,Token.category==VBP}\}\}\text{? (CORRELATION)}\text{):if} \\
\text{if.correlatedFind} = \{\text{rule}=="correlatedFindings"\}

\text{Rule: findingsFindings} \\
\text{Priority:20} \\
\{\{\text{DETERMINER}\} \mid \{\text{Token.kind==word,Token.category==VBZ}\} \mid \text{PUNCT}\} \mid \{\text{Token.kind==word,Token.category==JJ}\}\text{ FINDINGS}\{\{\text{Token.kind==word,Token.category==VBD}\} \mid \{\text{Token.kind==word,Token.category==VBP}\} \mid \text{PUNCT}\} \mid \{\text{Token.kind==word,Token.category==IN}\}\text{):if} \\
\text{if.findingFind} = \{\text{rule}=="findingsFindings"\}

\text{Rule: literatureFindings} \\
\text{Priority:20} \\
\{\{\text{DETERMINER}\} \{\text{Token.string}=="literature"\}\text{? (Token.string}=="suggest"\} \mid \{\text{Token.string}=="on"\}\text{):if} \\
\text{if.literatureFind} = \{\text{rule}=="literatureFindings"\}

\text{Rule: theoryFindings} \\
\text{Priority:20} \\
\{\{\text{Token.kind==word,Token.category==NNP}\}\text{? (Token.string}=="theory"\} \mid \{\text{Token.string}=="theories"\}\} \\
\{\{\text{Token.string}=="brings"\} \mid \{\text{Token.string}=="provides"\} \mid \{\text{Token.string}=="argues"\}\text{):if} \\
\text{if.theoryFind} = \{\text{rule}=="theoryFindings"\}
Rule: discussFindings
Priority: 20
(
(NOUN | PRONOUN)
({Token})?
(DISCUSS)
):if
--> :if.discussFind={rule="discussFindings"}

Rule: AuthorFindings
Priority: 20
(
{authorNNP}
({Token})[0,3]
({Token.string=="recognize"} | {Token.string=="analyze"} | {Token.string=="findings"} | {Token.string=="analyzes"} | {Token.string=="provides"} | {Token.string=="noted"})
(DETERMINER | {Token.kind==word,Token.category==WRB})
):if
--> :if.authorFindings={rule="AuthorFindings"}

Rule: resultFindings
Priority: 20
(
(NOUN | DETERMINER)
(RESULT)
({Token.kind==word,Token.category==VBP} | {Token.kind==word,Token.category==VBZ})
):if
--> :if.resultFind={rule="resultFindings"}

Rule: inSumConFindings
Priority: 20
( INSUMCONCLUSION )
:if
--> :if.inSumConFind={rule="inSumConFindings"}

Rule: DemonstrateFindings
Priority: 20
(
(NOUN | PRONOUN)
(DEMONSTRATE)
):if
--> :if.DemonstrateFind={rule="DemonstrateFindings"}

Rule: SuggestFindings
Priority: 20

(DETERMINER)

(Token)?

{{Token.string=="paper"} | {Token.string=="research"} | {Token.string=="work"} | {Token.string=="study"})

(Token)?

{{Token.string=="suggests"} | {Token.string=="reported"})

{{Token.kind==word,Token.category==IN}

):if

--> :if.suggestFind={rule="SuggestFindings"}

Rule: FindShownEtc
Priority: 20

{{Token.string=="They"} | {Token.string=="He"} | {Token.string=="She"} | {Token.string=="they"} | {Token.string=="he"} | {Token.string=="she"})

{{Token.string=="have"} | {Token.string=="has"})

{{FINDINGS | SHOW | CONCLUDE | ARGUE | DEMONSTRATE | CONFIRM})

{{Token.string=="that"}

):if

--> :if.FindShownEtcR={rule="FindShownEtc"}

Rule: InAnotherStudy
Priority: 20

{{Token.string=="In"} | {Token.string=="in"})

{{Token.string=="another"} | {Token.string=="one"})

{{STUDY | RESEARCH | WORK})

{{PUNCT})

{{Token.string=="they"} | {Token.string=="he"} | {Token.string=="she"})

):if

--> :if.InAnotherStudyR={rule="InAnotherStudy"}

Rule: Approach
Priority: 20

{{Token.kind==word,Token.category==DT})

{{Token.kind==word,Token.category==JJ})

{APPROACH)

{{Token}[0,7]

{{Token.kind==word,Token.category==VBZ}

{{Token.kind==word,Token.category==VBN}

):if

--> :if.ApproachR={rule="Approach"}

Rule: StudyResearch
Priority: 20

{
G.3. Jape rules Definition

Phase: definition
Input: Token Split Lookup
Options: control = appelt
Debug=true

Macro: VERBE
( {Token.string=="is"} )

Macro: DEFINE
( {Token.string=="define"} | {Token.string=="defines"} | {Token.string=="define"} | {Token.string=="defined"} | {Token.string=="defining"} | {Token.string=="redefine"} | {Token.string=="redefines"} | {Token.string=="redefined"} )

Macro: DESCRIBE
( {Token.string=="describe"} | {Token.string=="describes"} | {Token.string=="described"} )

Macro: CONCEPTUALISATION(
{Token.string=="conceptualized"} | {Token.string=="conceptualised"} )
Macro: INTRODUCE
{
{
(\text{Token.string}=="introduce"}
| (\text{Token.string}=="introduces"}
| (\text{Token.string}=="introduced"}
)

Macro: REFER
{
{
(\text{Token.string}=="refer"}
| (\text{Token.string}=="refers"}
| (\text{Token.string}=="referred"}
)

Macro: KNOWN
{
{
(\text{Token.string}=="known"}
{\text{Token.string}=="as"}
| (\text{Token.string}=="knew"}
{\text{Token.string}=="as"}
)

Macro: TERM
{
{
(\text{Token.string}=="term"}
| (\text{Token.string}=="terms"}
| (\text{Token.string}=="word"}
| (\text{Token.string}=="words"}
)

Macro: DEFINITION
{
(\text{Token.string}=="definition"}
| (\text{Token.string}=="definitions"}
| (\text{Token.string}=="Definition"}
| (\text{Token.string}=="Definitions"}
)

Rule: NounIsNounThat
Priority: 20
\(\{
(\text{NOUN}):\text{conceptDef}
(\text{NOUN})
(\text{VERBE})
(\text{NOUN})
((\text{Token.kind}==\text{word},\text{Token.category}==\text{"WDT"}))
\):isATThat
---
:isATThat.\text{NounIsNounThat}={\text{rule}==\text{"NounIsNounThat"}}

Rule: DefinitionDetection12
Priority: 20
\(\{
(\text{DETERMINER})?
((\text{Token.string}==\text{"definition"})
((\text{Token.string}==\text{"of"}))
\):isA
---
:isA.isASentence={\text{rule}==\text{"DefinitionDetection12"}}

Rule: rule2ExcelDefinition
Priority: 20
({Token.kind==word,Token.category == "DT"}
{Token.string=="concept"})
{Token.kind==word,Token.category == "WDT"}
{Token.kind==word,Token.category == "RB"}
{Token.kind==word,Token.category == "VBZ"}
{Token.kind==word,Token.category == "DT"}(NOUN))
| ({Token.kind==word,Token.category == "JJ"}
{Token.string=="understanding"}
{Token.kind==word,Token.category == "IN"}(NOUN)
{Token.kind==word,Token.category == "IN"})
{Token.kind==word,Token.category == "WDT"}
{Token.kind==word,Token.category == "DT"}(NOUN))
| ((NOUN)
(VERBE)
({Token.string=="viewed"} | {Token.string=="view"}))
{Token.string=="as"}
{Token.kind==word,Token.category == "DT"}(NOUN)
{Token.kind==word,Token.category == "IN"}(NOUN))
| ((NOUN)
(NOUN))?((Token.string=="can"} | {Token.string=="could"} | {Token.string=="might"} | {Token.string=="may"})
({Token.string=="be")?((Token.string=="be")?((Token.string=="seen") | {Token.string=="see")}
{Token.kind==word,Token.category == "IN"}
{Token.kind==word,Token.category == "DT"}
{Token.kind==word,Token.category == "JJ"})
):isExcelDefinition2

:isExcelDefinition2.isExcelDefinition2={rule="rule2ExcelDefinition"}

Rule: rule3ExcelDefinition
Priority: 20

(VERBE)
{Token.kind==word,Token.category == "VBN"}
{Token.kind==word,Token.category == "IN"}(NOUN)
Rule: rule6ExcelDefinition
Priority: 20
{
  {Token.kind==word,Token.category == "JJ"} (NOUN)
  {Token.kind==word,Token.category == "WDT"} (NOUN)
  {Token.kind==word,Token.category == "DT"} (NOUN)
  {Token.kind==word,Token.category == "IN"} (NOUN)
  {Token.kind==word,Token.category == "JJ"}
  }:isExcelDefinition6
  }::isExcelDefinition6.isExcelDefinition6={rule="rule6ExcelDefinition"},

Rule: rule7ExcelDefinition
Priority: 20
{
  {Token.kind==word,Token.category == "JJ"} (NOUN):conceptDef
  {Token.kind==word,Token.category == "WDT"}
  {Token.kind==word,Token.category == "MD"}
  {Token.kind==word,Token.category == "VB"}
  {Token.kind==word,Token.category == "JJ"}
  {Token.kind==word,Token.category == "IN"}
  {Token.kind==word,Token.category == "JJ"}
  (NOUN)
  }::isExcelDefinition7
  }::isExcelDefinition7.isExcelDefinition7={rule="rule7ExcelDefinition"},

Rule: rule9ExcelDefinition
Priority: 20
{
  {Token.string=="deals"}
  {Token.kind==word,Token.category == "IN"}
  {Token.kind==word,Token.category == "DT"}
  (NOUN)
  {Token.kind==word,Token.category == "IN"}
  {Token.kind==word,Token.category == "WDT"}
  {Token.kind==word,Token.category == "DT"}
  (NOUN))
G.3. Jape rules Definition

| (NOUN):ConceptDef(NOUN)? (VERBE) {Token.kind==word,Token.category == "RB"} {Token.string="conceptualized"}{ Token.kind==word,Token.category == "IN"} {Token.kind==word,Token.category == "DT"} {Token.kind==word,Token.category == "JJ"} (NOUN) | (NOUN):conceptDef {Token.kind==word,Token.category == "VBZ"} {Token.kind==word,Token.category == "DT"} {Token.string="meaning"} {Token.kind==word,Token.category == "IN"} {Token.kind==word,Token.category == "DT"} (NOUN) | (NOUN):conceptDef (VERBE) {Token.kind==word,Token.category == "CC"} (VERBE) {Token.kind==word,Token.category == "RB"} {Token.kind==word,Token.category == "RB"} ({Token.string="used"} | {Token.string="use"}) {Token.string="as"} {Token.kind==word,Token.category == "DT"}) | (NOUN):conceptDef (VERBE) {Token.kind==word,Token.category == "VBN"} {Token.kind==word,Token.category == "IN"} {Token.kind==word,Token.category == "JJ"} (NOUN) {Token.kind==word,Token.category == "WDT"} {Token.string="related"} {Token.string="to"}) ) :isExcelDefinition9 --> :isExcelDefinition9.isExcelDefinition9={rule="rule9ExcelDefinition"}

Rule:rule2DefinitionDefine
Priority: 20
( (DEFINE) (PRONOUN)? {Token.string="as"} (PUNCT)? (DETERMINER)? (NOUN) ) :isDefine --> :isDefine.isDefine2Ann={rule="rule2DefinitionDefine"}
Rule: rule4DefinitionDefine
Priority: 20
( (PRONOUN)
 (DEFINE)
 (PRONOUN)?
 (DETERMINER)?
 {{Token.kind==word,Token.category == IN}}?
 (PUNCT)?
 (NOUN)
 ) :isDefine
 -->
 :isDefine.isDefine4Ann={rule="rule4DefinitionDefine"}

Rule: rule6DefinitionDefine
Priority: 20
( (NOUN)
 {{Token.kind==word,Token.category == MD}}
 {{Token.kind==word,Token.category == VB}}
 {{Token.kind==word,Token.category == RB}}?
 (DEFINE)
 {{Token.kind==word,Token.category == IN}}
 (DETERMINER)?
 (NOUN)
 ) :isDefine
 -->
 :isDefine.isDefine6Ann={rule="rule6DefinitionDefine"}

Rule: rule1DefinitionDefine
Priority: 20
( (NOUN)
 (VERB)
 (ADVERB)?
 (DEFINE)
 ) :isDefine
 -->
 :isDefine.isDefine1Ann={rule="rule1DefinitionDefine"}

Rule: rule7DefinitionDefine
Priority: 20
( {{Token.string=="defined"}
 {{Token.string=="as"} | {{Token.string=="as"} | {Token.string=="like"} | {Token.string=="it"})
 ) :isExcelDefinition13
 -->
 :isExcelDefinition13.isDefine7Ann={rule="rule7DefinitionDefine"}

Rule: rule10DefinitionDefine
Priority: 20
G.3. Jape rules Definition

```
( {{Token.kind==word,Token.category == JJ}}?
(NOUN)
(DEFINE)
(PRONOUN)?
(NOUN)
):isDefine

Rule:rule15DefinitionDefine
Priority: 20
( (NOUN)
{Token.kind==word,Token.category == WDT}
(DEFINE)
('{{Token.kind==word,Token.category == JJ}}?
(NOUN)
):isDefine

Rule:rule1conceptDefinition
Priority: 20
( (PRONOUN)
({Token.kind==word,Token.category == VBZ} | {Token.kind==word,Token.category == VBP})
(DETERMINER)
({Token.string=='concept'} | {Token.string=='concepts'})
{Token.kind==word,Token.category == IN}
(NOUN)
):isExcelDefinition11

Rule:rule2conceptDefinition
Priority: 20
( (NOUN)
(PUNCT)?
(DETERMINER)?
{{Token.string=='concept'} | {Token.string=='concepts'}}
{{Token.kind==word,Token.category == VBZ} | {Token.kind==word,Token.category == VBP} | {Token.kind==word,Token.category == VBN}}
):isExcelDefinition11

Rule:rule3conceptDefinition
Priority: 20
( (DETERMINER)
{{Token.string=='concept'}} |}
```
\begin{verbatim}
{Token.string=="concepts"})
{Token.kind==word,Token.category == WDT}
({{Token.kind==word,Token.category == RB}})
({{Token.kind==word,Token.category == VBZ} | 
({Token.kind==word,Token.category == VBP})
):isExcelDefinition11
-->
:isExcelDefinition11.isConcept3DefAnn={rule="rule3conceptDefinition"}

Rule:rule1DefinitionIntroduces
Priority: 20
( (PRONOUN)
(INTRODUCE)
(DETERMINER)?
(NOUN)
{Token.kind==word,Token.category==IN}
):isIntroduce
-->
:isIntroduce.isIntroduce1Ann={rule="rule1DefinitionIntroduces"}

Rule:rule1LabelDefinition
Priority: 20
( (PRONOUN)
(label)
{Token.kind==word,Token.category == DT}
({{Token.kind==word,Token.category == JJ}})
(NOUN):ConceptDef):isExcelDefinition11
-->
:isExcelDefinition11.isLabel1DefAnn={rule="rule1LabelDefinition"}

Rule:rule2LabelDefinition
Priority: 20
( (PRONOUN)
(labeled)
{Token.string=="label"}
{Token.kind==word,Token.category == DT}
({{Token.kind==word,Token.category == JJ}})
(NOUN):ConceptDef):isExcelDefinition11
-->
:isExcelDefinition11.rule2LabelDefAnn={rule="rule2LabelDefinition"}

Rule:rule3LabelDefinition
Priority: 20
( (Token KIND==word,Token.category == VB})
{Token.string=="labeled"}
(PUNCT)?
({Lookup.classURI=="http://cui.unige.ch/~deribauh/genderStudies.owl#Variable_analyzed"})
):isExcelDefinition11
-->
:isExcelDefinition11.rule3LabelDefAnn={rule="rule3LabelDefinition"}
\end{verbatim}
Rule:rule1DefinitionDescribe
Priority: 20
( (NOUN) (VERBE) (ADVERB)? (DESCRIBE)
  ({Token.kind==word,Token.category==PRP})?
  {Token.string=="as"} )
):isCall
  -->
  :isCall.isDescribe1Ann={rule="rule1DefinitionDescribe"}

Rule:rule1DefinitionRefer
Priority: 20
( (NOUN) (VERBE) (ADVERB)? (REFER) (PRONOUN)?
  {Token.kind==word,Token.category==TO} )
):isRefer
  -->
  :isRefer.isRefer1Ann={rule="rule1DefinitionRefer"}

Rule:rule2DefinitionRefer
Priority: 20
( (NOUN) (VERBE) (RB)
  {Token.kind==word,Token.category==VBN}
  ({Token})?
  {Token.kind==word,Token.category==TO} (REFER)
  {Token.kind==word,Token.category==TO} )
):isRefer
  -->
  :isRefer.isRefer2Ann={rule="rule2DefinitionRefer"}

Rule:rule3DefinitionRefer
Priority: 20
( {Token.kind==word,Token.category==JJ}
  (NOUN)
  {Token.kind==word,Token.category==WDT}
  (REFER)
  {Token.kind==word,Token.category==TO}
  (({Token.kind==word,Token.category==DT}))?
  (NOUN)
Rule: rule4DefinitionRefer
Priority: 20
(NOUN)
(REFER):

Rule: rule5DefinitionRefer
Priority: 20
({Token.kind==word,Token.category==TO})
(REFER){Token.kind==word,Token.category==TO}

Rule: DefineKnowAs
Priority: 20
(ADVERB)
(KNOWN)
(DETERMINER)
(NOUN)

Rule: DefinitionTerm1
Priority: 20
({Token.kind==word,Token.category==VBZ} | DETERMINER | {Token.kind==word,Token.category==IN} | ADJECTIVE | {Token.kind==word,Token.category==VBP})
({Token.kind==word,Token.category==DT} | (ADJECTIVE) | {Token.kind==word,Token.category==IN})
(TERM)
({Token.kind==word,Token.category==IN} | (ADJECTIVE))
(PUNCTNoPoint)

Rule: DefinitionTerm2
Priority: 20
(TERM)
(PUNCT)?
G.3. Jape rules Definition

{(NOUN)
 (PUNCT)?
 (VERBE)
 )
 (DETERMINER)?
 }:DefinitionType3
 -->
 :DefinitionType3.Term2DefAnn={rule="DefinitionTerm2"}

Rule: DefinitionTerm3
Priority: 20
(
 (PRONOUN){Token.kind==word,Token.category==VBP}
 (DETERMINER)
 (ADJECTIVE)?
 (TERM)
 ):DefinitionType3
 -->
 :DefinitionType3.Term3DefAnn={rule="DefinitionTerm3"}

Rule: DefinitionTerm4
Priority: 20
(
 ({Token.kind==word,Token.category==RB})?
 ({Token.kind==word,Token.category==VBN} | {Token.kind==word,Token.category==VBP})
 {Token.kind==word,Token.category==IN}
 (DETERMINER)
 {Token.kind==word,Token.category==VBG}
 (TERM)
 {Token.kind==word,Token.category==TO}
 ):DefinitionType3
 -->
 :DefinitionType3.Term4DefAnn={rule="DefinitionTerm4"}

Rule: DefinitionTerm5
Priority: 20
(
 (DETERMINER)
 (TERM)
 (PUNCT)?
 (NOUN)
 ):DefinitionType3
 -->
 :DefinitionType3.Term5DefAnn={rule="DefinitionTerm5"}

Rule: DefinitionTermOnto
Priority: 20
(
 ({Lookup.classURI=="http://cui.unige.ch/~deribauh/genderStudies.owl#Variable_analyzed"})
 (PUNCT)?
 (VERBE)
 )
 ( (KNOWN) | (REFER) | (DEFINE) | (DESCRIBE))
(DEFINE)
(NOUN)
):DefinitionType3
-->
:DefinitionType3.DefinedWithParaR={rule="DefinedWithPara"}

Rule: Definition
Priority: 20
(({Token.kind==word,Token.category==JJ} | {Token.kind==word,Token.category==CD})
(DEFINITION)
{Token.kind==word,Token.category==IN}
(Lookup.classURI=="http://cui.unige.ch/~deribauh/genderStudies.owl#Variable_analyzed")
):DefinitionType3
-->
:DefinitionType3.DefinitionR={rule="Definition"}

Rule: ConstruedRR
Priority: 20
(({Token.kind==word,Token.category==DT} | {Token.kind==word,Token.category==PRP} |
{Token.kind==word,Token.category==RB} | {Token.kind==word,Token.category==RBS})
{Token.string=="construed"}
(Lookup.classURI=="http://cui.unige.ch/~deribauh/genderStudies.owl#Variable_analyzed")
({Token.kind==word,Token.category==RB} | {Token.kind==word,Token.category==IN} |
{Token.kind==word,Token.category==CC})
):DefinitionType3
-->
:DefinitionType3.ConstruedR={rule="ConstruedRR"}

Rule: ConstruedRR2
Priority: 20
(({Lookup.classURI=="http://cui.unige.ch/~deribauh/genderStudies.owl#Variable_analyzed"})
({Token.kind==word,Token.category==MD})?
({Token.kind==word,Token.category==VB} | {Token.kind==word,Token.category==VBZ})
{Token.string=="construed"}
Token.kind==word,Token.category==IN} | {Token.kind==word,Token.category==CC})
):DefinitionType3
-->
:DefinitionType3.ConstruedR2={rule="ConstruedRR2"}

Rule: WouldUse
Priority: 20
( {Token.string=="she"} | {Token.string=="he"} | 
{Token.string=="they"} | {Token.string=="She"} | 
{Token.string=="He"} | {Token.string=="They"})
({Token.kind==word,Token.category==MD})?
({Token.string=="use"} | {Token.string=="uses"})
(Lookup.classURI=="http://cui.unige.ch/~deribauh/genderStudies.owl#Variable_analyzed")
):DefinitionType3
-->
:DefinitionType3.WouldUseR={rule="WouldUse"}
G.4 Jape rules Methodology

Phase: findings
Input: Token
Options: control = appelt
Macro: PUNCTNoPar

( {Token.category == ","} | {Token.category == ","} | {Token.category == ":\:"} | {Token.category == ":\:"} | {Token.category == ":\:"} | {Token.category == ":\:"} | {Token.category == ":\:"} | {Token.category == ":\:"} | {Token.category == ":\:"} | {Token.category == ":\:"} | {Token.category == ":\:"} | {Token.category == ":\:"} | {Token.category == ":\:"} | {Token.category == ":\:"} )

Macro: SUBJECT

( {Token.string=="subject"} | {Token.string=="subjects"} | {Token.string=="Subject"} | {Token.string=="Subjects"} | {Token.string=="participant"} | {Token.string=="participants"} | {Token.string=="participant"} | {Token.string=="participants"} | {Token.string=="user"} | {Token.string=="users"} )

Macro: EXPERIMENT

( {Token.string=="experiment"} | {Token.string=="experiments"} | {Token.string=="Experiment"} | {Token.string=="Experiments"} )

Macro: METHOD

( {Token.string=="method"} | {Token.string=="methods"} | {Token.string=="Method"} | {Token.string=="Methods"} )

Macro: RECORD

( {Token.string=="recorded"} | {Token.string=="records"} | {Token.string=="recording"} )

Macro: ANSWER

( {Token.string=="answered"} | {Token.string=="answers"} | {Token.string=="answer"} | {Token.string=="answering"} )

Macro: ASKED

(}
G.4. Jape rules Methodology

Macro: MEASURE
( {Token.string=="measure"} | {Token.string=="measures"} | {Token.string=="measured"} )

Macro: COMPARE
( {Token.string=="compare"} | {Token.string=="compared"} )

Macro: CONDUCTED
( {Token.string=="conducted"} | {Token.string=="conduct"} | {Token.string=="conducts"} )

Macro: RUN
( {Token.string=="runs"} | {Token.string=="run"} | {Token.string=="ran"} )

Macro: OBTAIN
( {Token.string=="obtain"} | {Token.string=="obtains"} | {Token.string=="obtained"} )

Macro: USE
( {Token.string=="use"} | {Token.string=="uses"} | {Token.string=="used"} )

Macro: REPRESENT
( {Token.string=="represent"} | {Token.string=="represents"} | {Token.string=="represented"} )

Macro: EVALUATE
( {Token.string=="evaluate"} | {Token.string=="evaluated"} )

Macro: PROCEDURE
( {Token.string=="procedure"} | {Token.string=="procedures"} | {Token.string=="Procedure"} | {Token.string=="Procedures"} )

Macro: INDEXOF
( {Token.string=="the"} | {Token.string=="The"} )
Macro: RESEARCHMETHODOLOGY
{
{Token.string="research"} {Token.string="methodology"} | 
{Token.string="researches"} {Token.string="methodologies"} | 
{Token.string="Research"} {Token.string="methodology"} | 
{Token.string="Researches"} {Token.string="methodologies"}
}

Macro: METHODOLOGY
{
{Token.string="methodology"} | 
{Token.string="methodologies"}
}

Macro: PROVIDE
{
{Token.string="provides"} | {Token.string="provide"} | {Token.string="provided"}
}

Macro: VARIABLE
{
{Token.string="composite"} {Token.string="variable"} | 
{Token.string="dependent"} {Token.string="variable"} | 
{Token.string="dependents"} {Token.string="variables"} | 
{Token.string="dependent"} {Token.string="variables"} | 
{Token.string="independent"} {Token.string="variable"} | 
{Token.string="independents"} {Token.string="variables"} | 
{Token.string="independent"} {Token.string="variables"}
}

Rule: methodSubject
{
(DETERMINER)
(NOUN)
{Token.category="IN"}
{Token.category="CD"}
(SUBJECT)
):method
-->
:method.methodMeth={rule="methodSubject"}

Rule: methodSubject2
{
({{Token.category="VBD"}})?
(SUBJECT)
{{Token.category="TO"} | ADVERB | 
{Token.category="VBD"} | {Token.category="VBZ"} | 
{Token.category="VBP"} | {Token.category="IN"} | ADJECTIVE)
{{Token.kind="punctuation"}}
):method
G.4. Jape rules Methodology

```plaintext
--> :method.methodMeth2={rule="methodSubject2"}

Rule: methodParticipant
()
SUBJECT
({Token.category=="VBD"} | {Token.category=="VBZ"})
DETERMINER
):method
--> :method.methodParticipant={rule="methodParticipant"}

Rule: methodtest
()
EXPERIMENT
{Token.category=="TO"}
({Token.category=="VBZ"} | {Token.category=="VB"})
):method
--> :method.methodTest={rule="methodtest"}

Rule: methodMethod
()
(DETERMINER | ADVERB)
METHOD
(ADVERB | {Token.category=="VBZ"} | {Token.category=="VB"} | {Token.category=="VBD"})
):method
--> :method.methodM={rule="methodMethod"}

Rule: methodProcedure
()
(DETERMINER | ADVERB)
PROCEDURE
({Token.category=="VBZ"} | {Token.category=="VB"} | {Token.category=="VBD"})
):method
--> :method.methodP={rule="methodProcedure"}

Rule: methodResearchM
()
RESEARCHMETHODOLOGY
):method
--> :method.researchMethodology={rule="methodResearchM"}

Rule: methodMethodology
()
METHODOLOGY
):method
--> :method.methodMethodology={rule="methodMethodology"}
```
Rule: methodVariable
(
{Token.category=="VBP"}
VARIABLE
):method
--> :method.methodVariable={rule="methodVariable"}

Rule: methodRecord
(
{Token.category=="VBD"}
ADVERB
RECORD
):method
--> :method.methodRecord={rule="methodRecord"}

Rule: methodRecord1
(
RECORD
(DETERMINER)?
NOUN
):method
--> :method.methodRecord1={rule="methodRecord1"}

Rule: methodCompare
(
{PRONOUN | NOUN}
COMPARE
({{Token.category=="VB"} | NOUN | ADJECTIVE}
):method
--> :method.methodCompare={rule="methodCompare"}

Rule: methodEvaluate
(
({{Token.category=="TO"})
EVALUATE
(DETERMINER | NOUN | ADJECTIVE)
):method
--> :method.methodEvaluate={rule="methodEvaluate"}

Rule: methodAnswer
(
{PRONOUN | NOUN | {Token.category=="VBD"}}
{{Token}}?
(ANSWER | ASKED | CONDUCTED | RUN | USE)
(DETERMINER | NOUN | ADJECTIVE | {Token.category=="IN"})
):method
Rule: methodIncluded
()

{Token.string=="items"}
{Token.string=="were"}
{Token.string=="included"}
)
|

{Token.string=="items"}
{Token.string=="were"}
({Token})?
{Token.string=="recoded"}
)
):method

--> :method.methodIncluded={rule="methodIncluded"}

Rule: indexOfMetho
()

(INDEXOF)

({Token.category=="VBZ"} | {Token.category=="TO"} | {Token.category=="VBN"})
):method

--> :method.indexOfMet={rule="indexOfMetho"}

Rule: representMetho
()

({Token.category=="WRB"} | {Token.category=="CD"})
(NOUN)
(REPRESENT)
(NOUN)?
):method

--> :method.representMet={rule="representMetho"}

Rule: measureMetho
()

((NOUN | DETERMINER | {Token.category=="VB"} | PRONOUN)
({Token})?
(MEASURE | OBTAIN)
):method

--> :method.measureMet={rule="measureMetho"}

Rule: provideMetho
()

(NOUN)
({Token})?
(PROVIDE)
({Token})?


(NOUN)  
):method  
-->  
:method.provideMet={rule="provideMetho"}

G.5  Jape rules Hypothesis

Phase: hypothesisDetection  
Input: Token  
Options: control = appel  
Rule: HypothesisCD  
{
  \{\{Token.string == "hypothesis"\} | \{Token.string == "Hypothesis"\} | \{Token.string == "HYPOTHESIS"\}\}
  \{Token.kind==number\}
  \{\{Token.string == ":"\} | \{Token.string == ",\"\}\}
  (DETERMINER)?
}:Hypo -->  
:Hypo.HypoCD= {rule=HypothesisCD}

Rule: Hypothesized  
{
  \{\{Token.string == "hypothesized"\} | \{Token.string == "hypothesize"\}\}
  \{\{Token.kind==word,Token.category==IN\} | \{Token.kind==word,Token.category==TO\}\}
}:Hypo -->  
:Hypo.Hypothesized= {rule=Hypothesized}

Rule: JRulePossibleReason  
{
  \{Token.string =~ \[Pp\]ossible\}({Token})?\{Token.string == "reason"\} | \{Token.string =~ \[Pp\]ossibles\}({Token})?\{Token.string == "reasons"\}
}:mainHypo -->  
:mainHypo.possibleHypo= {rule=JRuleMainHypothesis}

Rule: JRuleMainHypothesis  
\{(DETERMINER)(ADJECTIVE)?\{Token.string == "hypothesis"\}\}
:mainHypo -->  
:mainHypo.mainHypo= {rule=JRuleMainHypothesis}

Rule: suggestHypothesis  
{
  (NOUN | DETERMINER | PRONOUN)
  \{\{Token.string == "suggests"\} | \{Token.string == "suggested"\} | \{Token.string == "suggest"\}\}
  \{Token.kind==word,Token.category==IN\}
}:mainHypo -->  
:mainHypo.suggestHyp= {rule=suggestHypothesis}

Rule: ifHypothesis  
Priority:20  
{
  \{\{Token.string=="if"\} |
G.5. Jape rules Hypothesis

If (Token.string=="If")
(Adverb | Determiner | AUX)
({Token.kind==word,Token.category==VBP} | Noun | {Token.kind==word,Token.category==VBN} | PunctNoPoint | {Token.kind==word,Token.category==VBZ} | Noun)
):if
-->
:if.ifHypo={rule="ifHypothesis"}

Rule: ifHypothesis2
Priority:20
({Token.string=="if"} | {Token.string=="If"})
(Determiner | PunctNoPoint | {Token.kind==word,Token.category==IN} | {Token.kind==word,Token.category==VBZ} | {Token.kind==word,Token.category==VBP} | Noun)
):if
-->
:if.ifHypo2={rule="ifHypothesis2"}

Rule: ifHypothesis3
Priority:20
({Token.string=="if"} | {Token.string=="If"})
(Adverb | Determiner | AUX | Adjective)
({Token.kind==word,Token.category==VBP} | Noun | {Token.kind==word,Token.category==VBN} | PunctNoPoint)
):if
-->
:if.ifHypo3={rule="ifHypothesis3"}

Rule: impliesHypothesis
Priority:20
({Noun} {Token.string=="implies"} {Token.kind==word,Token.category==IN})
):if
-->
:if.impliesHyp={rule="impliesHypothesis"}

Rule: ModalAuxHyp3
Priority:20
({Token.string=="are"} | AUX | {Token.string=="is"} |
Rule: allowedTestHyp
Priority: 20
{
{Token.string=="allowed"} | ADVERB | {Token.string=="allow"}
{Token.kind==word,Token.category==TO}
{Token.string=="test"}
}:if
-->::if.allowedTest={rule="allowedTestHyp"}
When scientists are looking for information in document collections, or on the web, they generally have a precise objective in mind. Instead of looking for documents "about a topic T ", they rather try to answer specific needs such as finding the definition of a concept, finding results for a particular problem, checking whether an idea has already been tested, or comparing the scientific conclusions of several documents. To build better information system model it is important to understand the needs of scientists in terms of information and their seeking information behaviour. The contributions of this thesis are to a) offer a user model built on the basis of a survey and interviews with scientists in order to understand their specific information-seeking behaviour needs; b) propose a generic annotation model (SciAnnotDoc) for scientific documents based on the user model; c) demonstrate that the annotation model is realistic enough to be used by both manual and automatic (1400 document in gender studies) methods of annotation; d) build a faceted search interface (FSAD) based on the SciAnnotDoc and e) evaluate the model by obtaining the feedback of end users (scientists) and show that the FSAD interface is more effective than a keywords-search, notably in terms of user preference, recall and F1 measure.