Investigating ontology based query expansion using a probabilistic retrieval model

Jagdev Bhogal

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

At

City University, London

School of Information Science

June 2011

TABLE OF CONTENTS

List of Tables ... **6**

List of Illustrations ... **8**

DECLARATION ... **12**

ACKNOWLEDGEMENTS ... **13**

ABSTRACT ... **14**

1: INTRODUCTION ... **15**

1.1 Thesis Background ... 15

1.2 Motivation ... 19

1.3 Aims and Objectives ... 20

1.4 Scope ... 20

1.5 Main Findings of Thesis ... 20

1.6 Structure of the Thesis ... 21

1.7 Statement about Publication ... 21

2. REVIEW OF CONTEXT-BASED INFORMATION RETRIEVAL ... **22**

2.1. Introduction ... 22

2.2 Personalisation ... 22

2.3 Link Analysis ... 24

2.4 User Interface ... 25

2.5 Language Models ... 28

2.6 Ubiquitous Computing ... 29

2.7 Query Expansion ... 32

2.7.1 Query Expansion Using Relevance Feedback ... 32

2.7.1.1 Selection Of New Terms ... 33

2.7.1.2 Weighting Of New Terms ... 34

2.7.1.3 Sample Size Of Relevant Documents ... 35

2.7.2 Query Expansion Using Term Co-Occurrence ... 35

2.7.3 Query Expansion Using Lexical Networks ... 37

2.7.4 Query Expansion Using Ontologies ... 40

2.7.4.1 Ontology Definition ... 40

2.7.4.2 Ontologies for Query expansion ... 42

2.7.4.3 Case Studies Of Query Expansion Using Domain-Independent Ontologies ... 42

2.7.4.4 Case Studies Of Query Expansion Using Domain-Specific Ontologies ... 47

2.7.4.5 Success factors in ontology based query expansion ... 49

2.8 Review of News Applications ... 52

2.9 Summary ... 56

3. METHODOLOGY ... **58**

3.1 Introduction ... 58

3.2 News Domain ... 58

3.3 Collections ... 59

3.3.1 News Ontologies and Test Collections ... 59

3.3.1.1 World News Ontology (WNO) 60
3.3.1.2 News Engine Web Services Ontology (NEWS) 63
3.3.2 Justification for chosen Ontology 65
3.3.3 Justification for chosen Test collection 65
3.4 Experiment Method 67
3.4.1 Design 67
3.4.2 Metrics 70
3.4.2.1 Recall 70
3.4.2.2 Precision 71
3.4.2.3 Mean Average Precision 71
3.4.2.4 Binary Preference (BPREF) 72
3.4.2.5 Document level averages 72
3.4.2.6 Precision-Recall Graphs 73
3.4.2.7 T-test metric for measuring statistical significance 74
3.5 Summary 75
4. SYSTEM 77
4.1 Introduction 77
4.2 Indexing the Collection 78
4.3 Probabilistic Retrieval Model 78
4.4 Search Method 80
4.5 Term Selection Method 80
4.6 Term Weighting Method 81
4.7 Algorithms 82
4.7.1 Building Parent-Child database 83
4.7.2 Algorithm for Search routine 84
4.7.3 Algorithm for Pseudo Relevance Routine 85
4.7.4 Algorithm for Rfb routine 86
4.7.5 Top 3 terms 87
4.8 Summary 88
5. EXPERIMENT RESULTS FOR RELEVANCE FEEDBACK RUNS: STANDARD AND TOP 3 TERMS 89
5.1 Introduction 89
5.2 Experiments Using Standard 20 Documents And 20 Terms and All Terms in Expanded Query 90
5.2.1 Document Level Averages Graphs 90
5.2.2 Precision-Recall Graphs 96
5.3 Experiments Using Standard 20 Documents And 20 Terms And Top 3 Terms In Expanded Query 98
5.3.1 Document Level Averages Graphs 98
5.3.2 Precision-Recall Graphs 103
5.4 Summary 107
6. EXPERIMENT RESULTS FOR VARYING NUMBER OF TERMS RELEVANCE FEEDBACK PARAMETER 108
6.1 Introduction 108

6.2 Experiments Using 20 Documents And 5 Terms And All Terms In Expanded Query 109
6.2.1 Document Level Averages Graphs 109
6.2.2 Precision-Recall Graphs 114
6.3 Experiments Using Standard 20 Documents And 10 Terms And All Terms In Expanded Query . 116
6.3.1 Document Level Averages Graphs 116
6.3.2 Precision-Recall Graphs 120
6.4 Experiments Using Standard 20 Documents And 15 Terms And All Terms In Expanded Query . 121
6.4.1 Document Level Averages Graphs 121
6.4.2 Results – Precision-Recall Graphs 123
6.5 Experiments Using Standard 20 Documents And 100 Terms And All Terms In Expanded Query 124
6.5.1 Document Level Averages Graphs 124
6.5.2 Results – Precision-Recall Graphs 128
6.6 Experiments Using Standard 20 Documents And 200 Terms And All Terms In Expanded Query 129
6.6.1 Document Level Averages Graphs 129
6.6.2 Precision-Recall Graphs 134
6.7 Summary 135
7. EXPERIMENT RESULTS FOR VARYING NUMBER OF DOCUMENTS RELEVANCE FEEDBACK PARAMETER 136
7.1 Introduction 136
7.2 Experiments Using 5 Documents And Standard 20 Terms And All Terms In Expanded Query ... 137
7.2.1 Document Level Averages Graphs 137
7.2.2 Precision-Recall Graphs 140
7.3 Experiments Using 10 Documents And Standard 20 Terms And All Terms In Expanded Query . 141
7.3.1 Document Level Averages Graphs 141
7.3.2 Precision-Recall Graphs 145
7.4 Experiments Using 15 Documents And Standard 20 Terms And All Terms In Expanded Query . 146
7.4.1 Document Level Averages Graphs 146
7.4.2 Precision-Recall Graphs 149
7.5 Experiments Using 100 Documents And Standard 20 Terms And All Terms In Expanded Query 151
7.5.1 Document Level Averages Graphs 151
7.5.2 Precision-Recall Graphs 154
7.6 Experiments Using 200 Documents And Standard 20 Terms And All Terms In Expanded Query 156
7.6.1 Document Level Averages Graphs 156
7.6.2 Precision-Recall Graphs 159
7.7 Summary 161
8. DISCUSSION OF SINGLE VALUE RESULTS 162
8.1 Introduction 162
8.2 Recall results 162
8.3 MAP Results 168
8.4 BPref results 174
8.5 Summary 180
9. DISCUSSION 182

9.1 Introduction 182
9.2 Analysis of results 182
9.2.1 Document Level Averages 182
9.2.2 Precision-Recall 182
9.2.3 Recall 183
9.2.4 MAP 183
9.2.5 BPref 183
9.3 Explanation of results 183
9.4 Other influential Factors 187
9.4.1 Quality of the Ontology and the Document collection 187
9.4.2 Stemming 187
9.4.3 Topic Hardness 188
9.4.4 Data Quality 190
9.5 Recommended Guidelines for Ontology Based Query Expansion 191
9.6 Summary 192
10. CONCLUSION 194
REFERENCES 199
APPENDIX A – Recall Precision Graphs 213
APPENDIX B – Parent list for Standard run (Topic 51-100) 235
APPENDIX C – Parent list for Top 3 terms run (Topic 51-100) 241
APPENDIX D – Child list for Standard run (Topic 51-100) 247
APPENDIX E – Child list for Top 3 terms run (Topic 51-100) 253
APPENDIX F – TREC output Listings 262
APPENDIX G – Single value results 310

List of Tables

Table 3.1 Basic Document Collection statistics **65**

Table 3.2 Summary of Experiments **69**

Table 3.3 Difference in the number of documents retrieved for pf and rfb standard run and top 3 terms run **69**

Table 4.1: Topic Length statistics (including stop words). **80**

Table 5.1 Number of relevant documents returned for each topic set **90**

Table 5.2: T-Test results (Document Level Averages) **97**

Table 5.3: T-Test results (Precision-Recall) **97**

Table 5.4: T-Test results (Document Level Averages) **106**

Table 5.5: T-Test results (Precision-Recall) **106**

Table 6.1 Difference in the number of documents retrieved for pf and rfb runs (varying number of terms parameter) **108**

Table 6.2: T-Test results (Document Level Averages) **116**

Table 6.3: T-Test results (Precision-Recall) **116**

Table 6.4: T-Test results (Document Level Averages) **121**

Table 6.5: T-Test results (Precision-Recall) **121**

Table 6.6: T-Test results (Document Level Averages) **123**

Table 6.7: T-Test results (Precision-Recall) **124**

Table 6.8: T-Test results (Document Level Averages) **129**

Table 6.9: T-Test results (Precision-Recall) **129**

Table 6.10: T-Test results (Document Level Averages) **134**

Table 6.11: T-Test results (Precision-Recall) **135**

Table 7.1 Difference in number of relevant documents retrieved for pf and rfb (varying number of documents parameter) **136**

Table 7.2: T-Test results (Document Level Averages) **141**

Table 7.3: T-Test results (Precision-Recall) **141**

Table 7.4: T-Test results (Document Level Averages) **145**

Table 7.5: T-Test results (Precision-Recall) **146**

Table 7.6: T-Test results (Document Level Averages) **150**

Table 7.7: T-Test results (Precision-Recall) **150**

Table 7.8: T-Test results (Document Level Averages) **155**

Table 7.9: T-Test results (Precision-Recall) **155**

Table 7.10: T-Test results (Document Level Averages) **160**

Table 7.11: T-Test results (Precision-Recall) **160**

Table 8.1 Ttest (Recall) **165**

Table 8.2 Ttest (Recall) **168**

Table 8.3 Ttest (MAP) **171**

Table 8.4 Ttest (MAP) **174**

Table 8.5 Ttest (BPref) **177**

Table 8.6 Ttest (BPref) **180**

Table 9.1 Number of terms in topics251-300 with at least one parent/child term **186**

Table 9.2 Summary of statistics given in Table 9.1 186
Table 9.3 Difference in the number of documents retrieved for pf and rfb standard run............... 187
Table 9.4 Overall Results summary... 189

List of Illustrations

Figure 2.1 Recommendation System example 24
Figure 2.2: Links for "Bush" 25
Figure 2.3 – Example of category Inline search interface 26
Figure 2.4 Example of Faceted Search. 27
Figure 2.5 Point Features displayed over a backdrop map 31
Figure 2.6 top 30 concepts retrieved for TREC4 query 214 37
Figure 2.7 Fragment of concept hierarchy from TREC topic 230 39
Figure 2.8 Schema Ontology and instantiation using data values 41
Figure 2.9 Relations defined for the six senses of the noun "swing" in WordNet 44
Figure 2.10 Example navigation using a concept lattice 46
Figure 2.11 Yahoo Directory structure 48
Figure 2.12 Screenshot of YourNews news filtering system. 53
Figure 2.13 News@hand architecture 55
Figure 2.14 Overview of Papyrus 56
Figure 3.1 NewsCodes for Crime,Law and justice concept and Crime concept 59
Figure 3.2 Hierarchical structure of Newscodes ontology 60
Figure 3.3 Sample extract taken from the WNO 61
Figure 3.4 : Diagrammatic form of XML tree format shown in Figure 3.3 62
Figure 3.5 Sample extract from NEWS ontology 64
Figure 3.6 Hierarchical Structure for Sample extract shown in Figure 3.5 64
Figure 3.7 Sample contents of the .bib file 66
Figure 3.8 Sample extract from the topic file 66
Figure 3.9 Diagram of Search Strategy 68
Figure 3.10 Precision-Recall Graph for example system 73
Figure 4.1 : System Overview 77
Figure 4.2 Algorithm for Building the Parent-Child database 83
Figure 4.3 Search Routine Algorithm 84
Figure 4.4 Algorithm for Pf Routine 86
Figure 4.5 Algorithm for Rfb routine 87
Figure 4.6 Algorithm for Top 3 terms routine 88
Figure 5.1 Topic 51-100 Retrieval Results – Document Level Averages 90
Figure 5.2 Topic 101-150 Retrieval Results – Document Level Averages 92
Figure 5.3 Topic 151-200 Retrieval Results – Document Level Averages 94
Figure 5.4 Topic 201-250 Retrieval Results – Document Level Averages 95
Figure 5.5 Topic 251-300 Retrieval Results – Document Level Averages 96
Figure 5.6 Topic 51-100 Retrieval Results – Precision-Recall 96
Figure 5.7 Topic 51-100 Results – Document Level Averages 98
Figure 5.8 Topic 101-150 Results – Document Level Averages 100
Figure 5.9 Topic 151-200 Results – Document Level Averages 101
Figure 5.10 Topic 201-250 Results – Document Level Averages 102
Figure 5.11 Topic 251-300 Results – Document Level Averages 103

Figure 5.12 Topic 51-100 Results – Precision-Recall **104**
Figure 5.13 Topic 101-150 Results – Precision-Recall **104**
Figure 5.14 Topic 151-200 Results – Precision-Recall **105**
Figure 5.15 Topic 201-250 Results – Precision-Recall **105**
Figure 5.16 Topic 251-300 Results – Precision-Recall **106**
Figure 6.1 Topic 51-100 Results – Document Level Averages **109**
Figure 6.2 Topic 101-150 Results – Document Level Averages **112**
Figure 6.3 Topic 151-200 Results - Document Level Averages **112**
Figure 6.4 Topic 201-250 Results – Document Level Averages **113**
Figure 6.5 Topic 251-300 Results – Document Level Averages **114**
Figure 6.6 Topic 51-100 Results – Precision-Recall **114**
Figure 6.7 Topic 101-150 Results – Precision-Recall **115**
Figure 6.8 Topic 151-200 Results – Precision-Recall **115**
Figure 6.9 Topic 51-100 Results – Document Level Averages **117**
Figure 6.10 Topic 101-150 Results – Document Level Averages **118**
Figure 6.11 Topic 151-200 Results – Document Level Averages **118**
Figure 6.12 Topic 201-250 Results – Document Level Averages **119**
Figure 6.13 Topic 251-300 Results – Document Level Averages **120**
Figure 6.14 Topic 251-300 Results – Precision-Recall **120**
Figure 6.15 Topic 51-100 Results – Document Level Averages **122**
Figure 6.16 Topic 101-150 Results – Document Level Averages **122**
Figure 6.17 Topic 251-300 Results – Precision-Recall **123**
Figure 6.18 Topic 51-100 Results – Document Level Averages **124**
Figure 6.19 Topic 101-150 Results – Document Level Averages **125**
Figure 6.20 Topic 151-200 Results – Document Level Averages **126**
Figure 6.21 Topic 201-250 Results – Document Level Averages **127**
Figure 6.22 Topic 251-300 Results – Document Level Averages **127**
Figure 6.23 Topic 251-300 Results – Precision-Recall **128**
Figure 6.24 Topic 51-100 Results – Document Level Averages **130**
Figure 6.25 Topic 101-150 Results – Document Level Averages **130**
Figure 6.26 Topic 151-200 Results – Document Level Averages **132**
Figure 6.27 Topic 201-250 Results – Document Level Averages **132**
Figure 6.28 Topic 251-300 Results – Document Level Averages **133**
Figure 6.29 Topic 51-100 Results – Precision-Recall **134**
Figure 7.1 Topic 51-100 Results – Document Level Averages **137**
Figure 7.2 Topic 101-150 Results – Document Level Averages **138**
Figure 7.3 Topic 151-200 Results – Document Level Averages **138**
Figure 7.4 Topic 201-250 Results – Document Level Averages **139**
Figure 7.5 Topic 251-300 Results – Document Level Averages **140**
Figure 7.6 Topic 51-100 Results – Precision-Recall **140**
Figure 7.7 Topic 51-100 Results – Document Level Averages **142**
Figure 7.8 Topic 101-150 Results – Document Level Averages **142**

Figure 7.9 Topic 151-200 Results – Document Level Averages **143**
Figure 7.10 Topic 201-250 Results – Document Level Averages **144**
Figure 7.11 Topic 251-300 Results – Document Level Averages **144**
Figure 7.12 Topic 51-100 Results – Precision-Recall **145**
Figure 7.13 Topic 51-100 Results – Document Level Averages **146**
Figure 7.14 Topic 101-150 Results – Document Level Averages **147**
Figure 7.15 Topic 151-200 Results – Document Level Averages **148**
Figure 5.61 Topic 201-250 Results – Document Level Averages **148**
Figure 7.17 Topic 251-300 Results – Document Level Averages **149**
Figure 7.18 Topic 51-100 Results – Precision-Recall **150**
Figure 7.19 Topic 51-100 Results – Document Level Averages **151**
Figure 7.20 Topic 101-150 Results – Document Level Averages **152**
Figure 7.21 Topic 151-200 Results – Document Level Averages **152**
Figure 7.22 Topic 201-250 Results – Document Level Averages **153**
Figure 7.23 Topic 251-300 Results – Document Level Averages **154**
Figure 7.24 Topic 51-100 Results – Precision-Recall **155**
Figure 7.25 Topic 51-100 Results – Document Level Averages **156**
Figure 7.26 Topic 101-150 Results – Document Level Averages **157**
Figure 7.27 Topic 151-200 Results – Document Level Averages **157**
Figure 7.28 Topic 201-250 Results – Document Level Averages **158**
Figure 7.29 Topic 251-300 Results – Document Level Averages **159**
Figure 7.30 Topic 51-100 Results – Precision-Recall **160**
Figure 8.1 Recall results for Topic 51-100 .... **162**
Figure 8.2 Recall results for Topic 101-150 .... **163**
Figure 8.3 Recall results for Topic 151-200 .... **163**
Figure 8.4 Recall results for Topic 201-250 .... **164**
Figure 8.5 Recall results for Topic 251-300 .... **164**
Figure 8.6 Recall results for Topic 51-100 .... **165**
Figure 8.7 Recall results for Topic 101-150 .... **166**
Figure 8.8 Recall results for Topic 151-200 .... **166**
Figure 8.9 Recall results for Topic 201-250 .... **167**
Figure 8.10 Recall results for Topic 251-300 .... **167**
Figure 8.11 MAP results for Topic 51-100 .... **168**
Figure 8.12 MAP results for Topic 101-150 .... **169**
Figure 8.13 MAP results for Topic 151-200 .... **169**
Figure 8.14 MAP results for Topic 201-250 .... **170**
Figure 8.15 MAP results for Topic 251-300 .... **170**
Figure 8.16 MAP results for Topic 51-100 .... **171**
Figure 8.17 MAP results for Topic 101-150 .... **172**
Figure 8.18 MAP results for Topic 151-200 .... **172**
Figure 8.19 MAP results for Topic 201-250 .... **173**
Figure 8.20 MAP results for Topic 251-300 .... **173**

Figure 8.21 BPref results for Topic 51-100 174
Figure 8.22 BPref results for Topic 101-150 175
Figure 8.23 BPref results for Topic 151-200 175
Figure 8.24 BPref results for Topic 201-250 176
Figure 8.25 BPref results for Topic 251-300 176
Figure 8.26 BPref results for Topic 51-100 177
Figure 8.27 BPref results for Topic 101-150 178
Figure 8.28 BPref results for Topic 151-200 178
Figure 8.29 BPref results for Topic 201-250 179
Figure 8.30 BPref results for Topic 251-300 179

DECLARATION

I grant powers of discretion to the University Librarian to allow this thesis to be copied in whole or in part without further reference to me. This permission only covers single copies made for study purposes subject to normal conditions of acknowledgement.

ACKNOWLEDGEMENTS

Sincere thanks to my supervisors Dr. A. MacFarlane and Dr. P. Smith for their patience, knowledge and excellent supervision. In addition, thanks to Dr. A. MacFarlane and City University for allowing me to use their information retrieval software.

I would like to thank Prof. S. Robertson for his invaluable advice.

My heartfelt thanks to my husband Kalwant and beautiful children Kaldev, Dilraj and Sharandeep for their love and continuous support throughout my studies.

Thank you to Leonidas Kallipolitis, Vassilis Karpis and Isambo Karali (University of Athens) for allowing me to use their news ontology.

Finally thanks to Birmingham City University for their financial support.

ABSTRACT

This research briefly outlines the problems of traditional information retrieval systems and discusses the different approaches to inferring context in document retrieval. By context we mean word disambiguation which is achieved by exploring the generalisation-specialisation hierarchies within a given ontology. Specifically, we examine the use of ontology based query expansion for defining query context. Query expansion can be done in many ways and in this work we consider the use of relevance feedback and pseudo-relevance feedback for query expansion. We examine relevance feedback and pseudo-relevance to ascertain the existence of performance differences between relevance feedback and pseudo-relevance feedback. The information retrieval system used is based on the probabilistic retrieval model and the query expansion method is extended using information from a news domain ontology. The aim of this project is to assess the impact of the use of the ontology on the query expansion results. Our results show that ontology based query expansion has resulted in a higher number of relevant documents being retrieved compared to the standard relevance feedback process. Overall, ontology based query expansion improves recall but does not produce any significant improvements for the precision results. Pseudo-relevance feedback has achieved better results than relevance feedback. We also found that reducing or increasing the relevance feedback parameters (number of terms or number of documents) does not correlate with the results. When comparing the effect of varying the number of terms parameter with the number of documents parameter, the former benefits the pseudo-relevance feedback results but the latter has an additional effect on the relevance feedback results. There are many factors which influence the success of ontology based query expansion. The thesis discusses these factors and gives some guidelines on using ontologies for the purpose of query expansion.
.

1: INTRODUCTION

1.1 Thesis Background

Information Retrieval (IR) is the process of translating a set of information needs into queries and determining which set of documents satisfy the information needs. In this chapter we will examine the different approaches to IR, the strengths and weaknesses of the techniques used to improve query formulation and the advantages and disadvantages of using ontologies to infer query context.

There are four approaches to IR, namely Boolean Retrieval, Vector Space, Latent Semantic Indexing and Probabilistic Retrieval. We will start with the classic Boolean Retrieval model which uses a binary approach where documents are considered either relevant or non-relevant. Relevant documents are considered to be those which include the term contained in the query. The Boolean model uses "exact match" retrieval where a document is considered to be either relevant or non-relevant. It does not take into account degree of relevance. There are two methods for matching query terms with the document terms namely 'exact match' and 'partial matching'. The first method is the "exact match" principle which results in all query terms having equal importance because there is no mechanism for assigning a degree of importance to query terms. The returned results set is in list format which is not ranked in terms of relevancy. So for "OR" searches, documents containing at least one of the query terms are as relevant as documents containing all query terms.

The second approach to IR is the vector space model (Salton 1983) which enables "partial matching" and ranks the query results using weighted vectors. The Term frequency-Inverse document frequency (tf-idf) weighting scheme is normally used to weight the vectors. Tf is a count of how many times a term occurs within a document. Idf is a count of how many times a term occurs across all of the relevant documents. The best terms usually have a high within-document frequency but low frequency across the entire document collection. A study (Paris and Tibbo 1998) found that although the exact match searches had better results more often than the partial match searches, neither mechanism demonstrated superior performance for every query. These results do not in any way prove the superiority of partial match techniques or exact match techniques, but they do suggest that different queries demand different techniques.

With the increase in the size of the document collection, automatic classification can take place by calculating a similarity measure between documents. A representative document *group vector* is then chosen as the cluster, and a search request is initially checked against all the cluster vectors only. Thereafter, the request is checked against only those individual documents where the cluster vectors show a high score with the request. The Cluster Hypothesis (Van Rijsbergen 1979) states "Closely associated documents tend to be relevant to the same requests. Document clustering is carried out using document-document similarity measure and should result in more effective and efficient retrieval." Salton believes that although document clustering saves time, it necessarily reduces the effectiveness of a retrieval system. Jardine and Van Rijsbergen (1971) disagree and states that document clustering has *potential* for improving the effectiveness.

A disadvantage of conventional IR is that with the inverted index, the information pertaining to a document is scattered among many inverted term lists. Data about related items should appear close together. To improve the computational efficiency of information retrieval, various techniques were developed. The Rocchio Classification is based on the assumption that most users have a general conception of which documents should be denoted as relevant or non-relevant. The search query is revised to include terms from the relevant documents in order to increase recall and precision. However, the Rocchio algorithm often fails to classify multimodal classes whereby a class belongs to more than one cluster. For example, a news article about A company merger in the United Kingdom could be labelled as 'business_mergers' and 'UK'. The label 'business_mergers' label is based on a specific topic but the label 'UK' is more general and just refers to location. A disadvantage of clustering is the conceptual problems in clustering large files and maintaining clustered organizations in dynamic file environment. Testing of document clustering requires large quantities of data and computer time.

There are several limitations with these retrieval models. Both retrieval models assume that the query terms are independent and they also use a literal search whereby only documents which contain the query terms are retrieved in the final results set. Synonymy and polysemy are ignored. Polysemy occurs when a word has different meanings. For example, there is a problem if a query term has multiple meanings eg bed – place to sleep or plant flowers. Synonymy occurs when different words are used to refer to the same concept. Synonyms are also problematic. For example the query 'heart attack' would only retrieve documents that contain the terms "heart" or "attack", even though documents containing "cardiac arrest" are relevant, they would not appear in the result because the terms do not appear in the query. Therefore literal searches often return irrelevant results (false-positive) and miss information that is relevant (false-negative). Also long documents are poorly represented because they have poor similarity values.

The third approach to IR is the Latent Semantic Indexing (LSI) which is an indexing and retrieval method that handles synonymy and polysemy using a mathematical technique called Singular Value Decomposition (SVD) to identify patterns in the relationships between the terms and concepts contained in an unstructured collection of text (Deerwester et al, 1988). LSI assumes that words that are used in the same contexts tend to have similar meanings. LSI aims to extract the conceptual content of a document by establishing associations between those terms that occur in similar contexts. Queries, or concept searches, against a set of documents that have undergone LSI will return results that are conceptually similar in meaning to the search criteria even if the results don't share a specific word or words with the search criteria.

All of these previous models have used inverse document frequency to judge the usefulness of a term for for indexing purposes. The important distinction between terms in relevant and non-relevant documents is ignored. The fourth approach to IR is the probabilistic retrieval model which is a highly effective retrieval model that makes explicit distinctions between occurrences of terms in relevant and non-relevant documents (Sparck-Jones et al, 2000). It calculates the probability of a document being relevant if it contains certain terms.

As we have seen, in classical information retrieval systems, documents and queries are usually represented by sets of weighted terms. To evaluate to what extent a document is relevant to a query, a retrieval status value (rsv) is computed by aggregating the above weights for the terms present in the query. Then documents are ranked in descending rsv order. In addition to tf-idf, other factors such as document length and use of constants also need to be taken into account. The chances of matching terms in longer documents is higher so unless document length is taken into account, longer documents will appear higher up in the results set. A relevant document should be treated with equal importance regardless of its length so a normalisation factor is used in the term weighting formula to equalise the length of the document vectors.

An ideal query cannot be generated without knowing a great deal about the document collection. It is customary to conduct searches iteratively. The first query is usually tentative then query formulation is improved using subsequent searches based on evaluations of previous retrievals. One method for improved query formulation is Relevance feedback. Relevance feedback is a form of query expansion using local analysis because it requires users to give input on the relevance of each document in the results set. Pseudo-relevance feedback (also known as blind feedback) takes terms from the top N highest ranking documents in the result set and uses these terms for query expansion. Relevance feedback extracts terms, from documents, which users have judged to be relevant to given query topics and uses these terms for query expansion.

In relevance feedback, users give additional input on documents by marking documents in the results set as relevant or not and this input is used to reweight the terms in the query for documents. In query expansion, users give additional input on query words or phrases, possibly suggesting additional query terms (Manning et al, (2008). Query expansion allows the user to carry out searches on morphological variations of the original term(s) and/or include any other terms which result from word sense disambiguation. Various approaches exist for conducting query expansion. Query expansion can be done by global analysis or local analysis. Global analysis uses a thesaurus as a source for query expansion terms. This has the advantage of not requiring any user input. Local analysis analyses each document in the result set.

As we have just mentioned, relevance feedback relies on additional user input. However since the users might be reluctant to provide feedback, researchers started focusing on contextual IR. Contextual IR integrates the user context into the retrieval process. Context can be inferred in many different ways (see chapter 2). Since the 1980s much work has been done in the area of text categorisation/classification. Text categorization can be used for word sense disambiguation and query expansion. With text categorisation the documents are assigned to categories. Whether this process is done manually (using domain experts) or automatically (using machine learning algorithms), there is a need to understand the context of the document content so that a suitable category can be assigned. With the number of digital documents on the rise, there is a need to automate the process of text categorization. Sebastiani (2002) discusses the main machine learning approaches used for text categorization. Machine learning algorithms achieve a level of accuracy that is comparable to that achieved by human experts. More recently, ontologies have been used in an interactive manner for supporting faceted search queries. The faceted search facility has become a standard component of web site design, especially for retail web sites. Faceted searches are better suited for ontologies which have wide breadth and shallow depth. Users can interact with the faceted search facility and enable or disable facets according to their search needs. Ontologies can also be used to infer context for ambiguous queries. The concepts in the ontology can be used for word sense disambiguation and subsequent query expansion. An ontological model can effectively disambiguate meanings of words from free text sentences (Buckland 2003). This will be a form of global analysis but instead of using a thesaurus, we will be using an ontology. An ontology is a collective body of knowledge which is usually created and shared by users who are experts in that domain. An ontology that is a collective view of the domain is more likely to be an accurate and comprehensive representation of that domain. A domain ontology models a specific domain. It represents the particular meanings of terms as they apply to that domain. An upper ontology is a model of common objects that apply across a wide range of domains ontologies. The ontology can be collection dependent or collection independent. The advantage of using collection dependent ontologies is that the relevant documents are indexed with the ontology terms. However, these types of ontologies are harder to maintain for two reasons. Firstly, the ontology will have to be updated with new incoming documents. Secondly, if new concepts are added to the ontology then changes have to be reapplied to the documents that have already been categorized within the ontology. Collection independent ontologies do not suffer from this “update” problem as much. This does not mean the ontology is not updated at all but ontology concepts evolve slowly over a longer period so updates are less frequent.

Text categorization is a much widely researched topic and according to Cleverdon (1984)the effectiveness levels of automated text categorization are growing at a steady pace and even if they reach a plateau, this plateau will be higher than the effectiveness levels of manual text categorization. The machine learning approach requires good quality training data and documents must be a representative sample of unseen incoming documents, otherwise the approach loses effectiveness. Faceted search will not be used for this research project because the topics in the chosen query sets are predefined and do not require interactive search. If the design of an interactive user interface was within the scope of this project then most probably a faceted search facility would have been incorporated. For these reasons we have decided not to use text categorization or faceted search and opted to use ontology based query expansion instead. We will be using a collection independent ontology for our experiments. Another reason for using ontologies is that most ontologies are designed for the semantic web and are written in a portable language.

Broder (2002) define 3 types of information retrieval tasks – navigation tasks (such as finding a website), information query tasks and transaction tasks. Ontology based query expansion is more suited for information tasks. We are interested in the news domain because this is a domain which is information intensive. Another reason for examining the news domain is that in future it will not be free and will be available supscription pay per view basis, so there is a definite requirement for the service to be accurate otherwise online news will not survive. The online Times news website is an example of this, where mandatory registration is required by users in order to generate digital revenue (Halliday 2010). News is the communication of information on current events which is presented by print, broadcast, internet or word of mouth to a third party or mass audience. A news ontology is usually created and shared by a group of specialists in the news field such as journalists, editors and Press standards organisations. Domain specific ontologies are used to model specialised vocabulary from that field such as medical terms. The news domain doesn't have a specific vocabulary as such it just uses plain English language in an accepted journalistic style. However what is important within this domain is the structure of news items. News writing attempts to answer event based questions. The important features of a news item usually includes who, what, when, where and why. The structure of a news item includes: Headline, subheading, event description. News ontologies can be used to assist in different tasks such as news categorisation/classification, reasoning; searching; news annotation; updating, news summarization and news alerts. We will be using the news ontology for the searching task.

1.2 Motivation

Query expansion seems to be more successful only on relevant documents (Ogawa and Mano 2001, Billerbeck and Zobel 2004). There is still scope to improve the techniques, interfaces or algorithms used to infer context more accurately in order to improve the results even further. The most recent query expansion technique involves the use of ontologies to infer context for ambiguous queries. The concepts in the ontology can be used for word sense disambiguation and subsequent query expansion. Gonzalo et al, 1998 state that "Ontologies have been used to aid query expansion since the early nineties with mixed success". According to Manning et al, (2008), query expansion is less successful than relevance feedback but it may be as good as pseudo-relevance feedback. A detailed investigation into query expansion using ontologies is needed to study the reasons for their success/failure. Therefore, our motivation for this research is to carry out such an investigation and address questions such as whether the use of query expansion increases recall, precision or both and secondly how ontology based query expansion compares with relevance feedback/pseudo-relevance feedback techniques.

This research attempts to combine both approaches of relevance feedback query expansion and ontology based query expansion. The purpose of this research project is to carry out a detailed investigation into the area of query expansion using a news ontology in a probabilistic retrieval environment. Since we are interested in the news domain, an appropriate document collection and domain-specific ontology will be selected. The ontology is written in eXtensible Mark-up language (XML) to make it portable. We need to analyse the ontology and transfer the knowledge in an appropriate format so it is accessible to the Okapi software. The Okapi model already uses pseudo relevance and relevance feedback techniques (Robertson et al, 1997) and the relevance feedback information can be based on pre-stored relevance judgements which indicate for each document whether it is relevant to the topic query or not. The techniques have proved to be successful to a certain extent so we do not want to discard them. The two main parameters of relevance feedback are: selection of terms and the sample size of relevant documents. In the Okapi system traditionally these have been 20 terms and 20 documents. Billerbeck and Zobel (2004) state that the choice of query expansion parameters used can affect the retrieval performance. As part of this research we will experiment in varying these relevance feedback parameters and analyse the impact on the results. Another question that will be addressed is whether to use all expanded terms or select the top 3 query expansion terms. Experimental results will be evaluated using retrieval effectiveness metrics. The initial study recommended conditions where ontology based query expansion is likely to be successful. The news ontology will be analysed to see which of these conditions or success factors were applicable if any. All of the findings of this research will be discussed.

1.3 Aims and Objectives

Research experiments are needed to examine the effects of using an ontology for query expansion in the newswire domain. The aim of this research is to investigate the effectiveness of ontology based query expansion techniques. The central hypothesis is that the use of a news domain simple ontology for query expansion in a probabilistic retrieval model will improve retrieval effectiveness. This aim can be broken down into a number of objectives.

- Firstly, a document collection will be selected together with an appropriate ontology. The document collection will be indexed in the standard manner.
- The second objective will focus on building a separate database containing semantic information such as parent-child relationships between ontology nodes. This information will be used to supply additional terms for expanding the original query terms.
- There is an objective for designing and conducting laboratory experiments in order to compare and contrast the performance of the standard retrieval model with the revised retrieval model.
- Finally the findings of this research will be summarized and an overall conclusion given with recommended areas for future work.

1.4 Scope

Although ontologies have been used in news retrieval systems, they have mostly been used for news classification, news annotation and building user profiles. We would like to explore the use of ontologies to for query expansion. This thesis focuses on the use of ontologies for query expansion in the news domain. General domains and ontologies from other types of domains will not be discussed. The probabilistic retrieval model will be used, other types of retrieval models will not be examined. We are not including broadcast news, our research uses news in text/print format. We are only looking at the searching feature of news retrieval systems and do not include any of the other features. At present the use of personalized news stories for mobile devices falls outside the scope of our research.

1.5 Main Findings of Thesis

Based on the aims and objectives declared above, our results show the following main findings :

- Use of the news domain simple ontology for query expansion has resulted in a higher number of relevant documents being retrieved compared to the standard relevance feedback process.
- Query expansion on our chosen ontology improves recall but does not produce any significant improvements for the precision results.

- Using the news domain simple ontology, pseudo-relevance feedback has gained more from ontology based query expansion. The pseudo-relevance feedback results have obtained better precision at rank results higher up the rank, whereas the relevance feedback results start improving from the lower end of the ranked set of documents. Even for the single value results, the ontology has more of a positive impact on e pseudo-relevance feedback compared to relevance feedback.
- The experiment results show that reducing or increasing the relevance feedback parameters (number of terms or number of documents) does not have any correlation with the results.
- The results also show that the number of terms parameter benefits the pseudo-relevance feedback results but the number of documents parameter has an additional effect on the relevance feedback results.
- There are many factors which influence the success of ontology based query expansion such as the quality of the ontology in terms of coverage of the domain; use of fewer general terms and lastly the similarity match between the ontology and the document collection is important to increase the number of relevant documents retrieved for the ontology based query expansion terms.

1.6 Structure of the Thesis

The thesis consists of 10 chapters with additional appendices. In the next chapter, a detailed literature review is given of the different approaches to Context-Based Information Retrieval. Chapter 3 describes the methodology used to achieve the research objectives. It includes sections on experiment design and metrics. A description of the design and implementation of the system is given in Chapter 4. Chapters 5-8 detail the range of experiments conducted and deploy the metrics discussed in Chapter 3 to measure retrieval performance. The discussion in Chapter 9 describes the results and also discusses the issues surrounding ontology based query expansion. Finally in Chapter 10, the findings of this research will be summarized and an overall conclusion given with recommended areas for future work.

1.7 Statement about Publication

Much of the literature review given in chapter 2 on approaches to Context Based Information Retrieval has been published earlier (Bhogal et al, 2007).

2. REVIEW OF CONTEXT-BASED INFORMATION RETRIEVAL

2.1. Introduction

Traditionally information retrieval systems used fairly small static document collections. The early search engines used standard information retrieval techniques (Salton 1989) to return documents whose index terms matched the terms in the user query. These search engines aimed for high precision and high recall. The search process was iterative. Relevance feedback information was taken from the user so the retrieval process could be repeated using the additional relevance information. Over the past decade or so, search engines have been used to carry out information retrieval on the web. Information space on the web is comparatively larger and combined with the ambiguity of the English language, a long list of results is returned, much of which is not always relevant to the user's information needs. The main difference in online retrieval systems and traditional information retrieval systems is that the former are usually web-based and as a result the document collection is more dynamic or fluid. For traditional information retrieval systems and web based retrieval systems, the inadequacies of standard search engines means that the user community suffers from information overload. To increase the number of relevant documents retrieved queries need to be disambiguated by looking at their context. In this section different approaches to context are described including query expansion using ontologies which forms the focus for this research project.

Third generation search engines known as meta-search engines attempt to determine the context of the user query and allow the user to obtain more meaningful results. In other words these search engines are focusing more on achieving high precision. The next section examines the different approaches to handling context in information retrieval.

Context does not have a standard definition (Finkelstein et al, 2002; Goker et al, 2009). With Domain Specific search engines the context is implied by the chosen database which contains information specific to that domain (Lawrence 2000). The problem with this is that ambiguity might still exist within a domain. Another approach to defining context is to distinguish between novice and expert users (Torrison 1998). The following sections summarise the various IR approaches to context some of which overlap with the disciplines of artificial intelligence and statistics.

2.2 Personalisation

The Personalisation approach to context is based on an individual user's search profile. The system records the history of queries and documents viewed and reuses this information in future searches. Personalization involves learning user interest and preferences over time. Learning techniques can be used to help the information retrieval system adapt to user's specific needs. Goker (1999) argues that the system can learn from user searches to improve subsequent searches. The main assumption is that user queries remain within a particular context over several online sessions. The user models provide context for queries and other interactions with the information system. So two users with the same query could end up with different results if they have different profiles.

Many personalisation systems exist (Bharat, 2000, Bauer and Leake 2001, Rhodes and Maes 2000, Goker 1999). Another application area for Personalisation is in Web Usage mining (Srivastava et al, 2000) whereby user access patterns are automatically discovered from one or more Web servers. This is especially useful in helping to target advertisements to specific groups of users.

Recently personalisation has moved towards user community based information examples of which are collaborative filtering and collaborative querying. Collaborative filtering is a process which looks for votes from users with the same voting patterns as a given user or active user and then attempts to automatically predict votes for the active user. Recommendation systems use the votes predicted by collaborative filtering to make automatic recommendations to the active user. The assumption is that those who agreed in the past tend to agree again in the future. For example a music collaborative filtering or recommendation system can recommend music that other users with similar taste liked. Collaborative querying reduces the time taken by users to formulate queries by recommending similar queries carried out by other users (Fu et al, 2005b). Another interesting dimension in Recommender Systems is that of trust (O'Donovan and Smyth 2005). It may not be enough to just look at users with similar preferences but also look at the previous ratings history of each of these users and highlight those users that have been reliable in their ratings who thereby gain higher trust and thus higher weighting in the recommendation process. The collaborative approach can also be used for web page ranking and classification. Lifantsev (1998) (as cited by Huang 2000) describe the OpenGrid project which proposes to extend the HTML standard and allow users to submit their opinions about the pages they browse. These opinions are collected regularly and are used to influence page ranking and classification. The more recent approaches such as Collaborative filtering and Collaborative querying both aim to get the best out of sharing information between the user community and extracting the relevant parts to make specific recommendations to a user. Figure 2.1 shows products being recommended on the basis of previous purchases by the user. Also the products are ranked in descending order of customer review rating

Figure 2.1 Recommendation System example

. Personalised systems have their limitations such as they are based on the local context of a single user profile and thus may be seen to be intrusive; some require a high level of input from the user; some use context information at run-time to re-rank documents thus creating a performance overhead. Also the user goal is based on past use of the system not initial description of intent which may change from one session to another. Knowledge models such as user profiles and ontologies both require storage space and need to be updated but the user profile probably needs to be updated more frequently than an ontology because the needs of an individual user change more rapidly than a community view of a subject area.

2.3 Link Analysis

Link analysis interprets context as information space. Information space could be a web page and its different types of hyperlinks such as links to companies and homepages The content and structure of the information space surrounding documents can be used to provide the context of a web document (Lowe 2000). This contextual information can then be used to improve the accuracy of relevance rankings, assigned to documents. Pages which contain the context term(s) or pages that have links that contain context term(s) and have small semantic distance are ranked higher. Some improvements in the results are shown but the methods used to select context terms need to be less manual. El-Beltagy et al(2001) amend web pages by adding multi-destination links to them. The user interests define which links are to be exported for that user and a document's content defines which of these links are to be rendered in that document. In link analysis, the user is presented with related links, whereas with query expansion, a list of related terms is given.

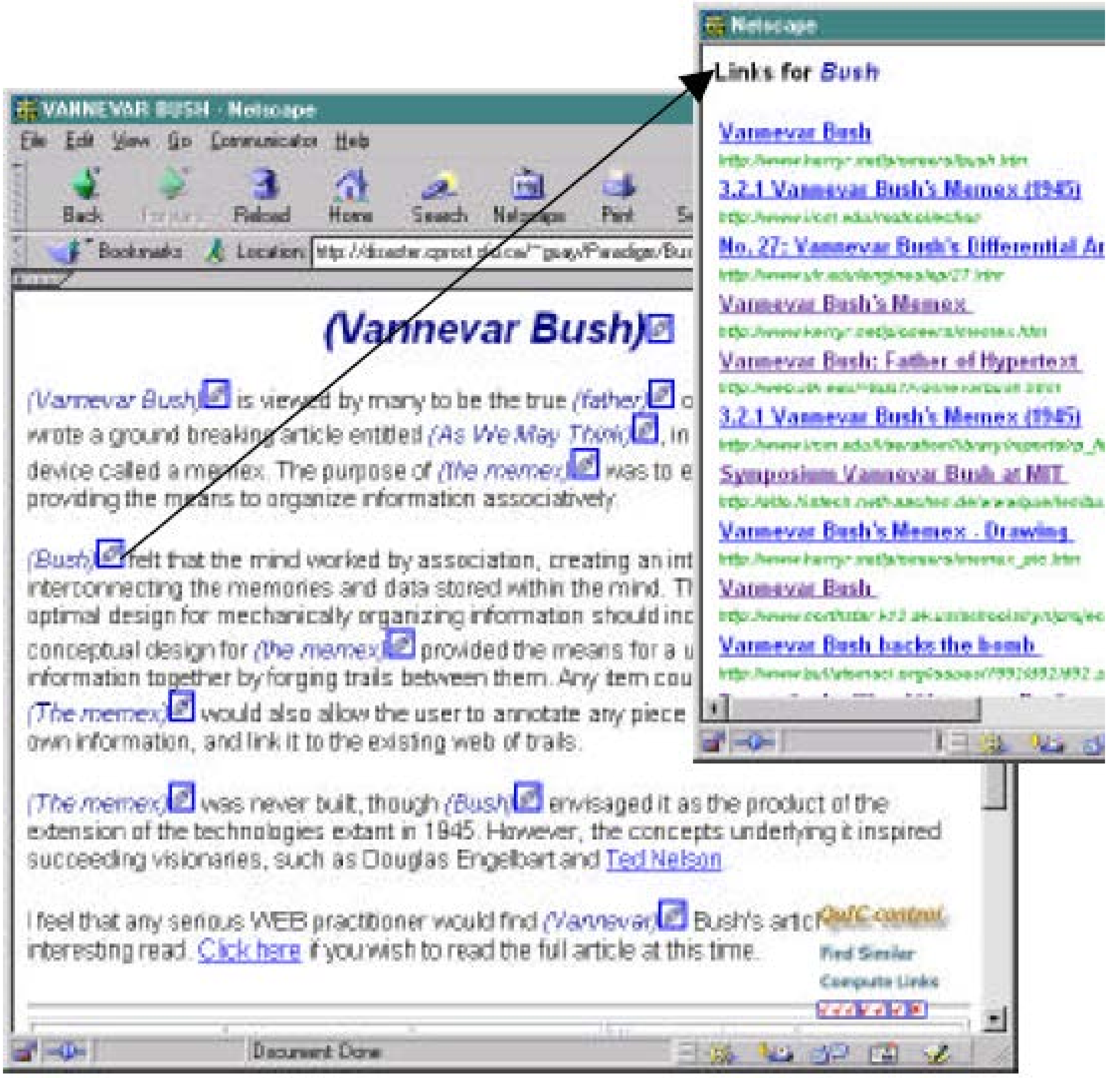


Figure 2.2: Links for "Bush"

Linkages among documents may indicate the importance of documents on the basis that important documents are pointed to by many documents (Yu et al, 2001). This is useful for deciding which databases are appropriate to search thus preventing the need to search all of the databases.

One flaw of this approach is that it relies on a correct and complete set of links being put in place otherwise missing/incorrect links affect the quality of the results.

2.4 User Interface

The assumption that for domain-specific search engines the context is implied by the chosen database which contains domain specific information might be an over generalisation because firstly the problem with this is that ambiguity might still exist within a domain (Krovetz and Croft 1992) and secondly, even within a particular domain, users will differ in their information needs and information seeking behaviour. Differences in user levels can be addressed through the user interface design. Novice users need more support from the user-interface for navigating and querying than expert users (Torrissen 1998).

Another interpretation of context relates to the appropriate way to display information rather than using traditional list interfaces. Dumais et al, (2001) developed 7 interfaces for integrating semantic category information with Web search results. The interfaces ranged from normal list interfaces to category interfaces formatted in different ways. The interface that performed the best was Category Inline. This is where the resulting documents are presented under different category headings and an inline summary of each document is given under the document title. Dumais et al, (2001) state the reason for this is that the category headings help the user to disambiguate ambiguous queries. For example, a query on Jaguar produced documents under the following categories: Computers and Internet, Automotive, Entertainment and Media, Travel, Finance to name but a few (see Figure 2.3)

Figure 2.3 – Example of category Inline search interface

If the user is interested in Jaguar cars, then only the documents under the Automotive category need to be followed up thus speeding up the search process for documents relevant to the user's information need. In some cases there may be the need to introduce sub-categories. Another problem might be where a document is related across multiple categories, these have to be searched separately. Hearst (1994) developed a task related specialised search interface which shows a list of categories at the top and the top three categories are shown as a Venn diagram. One drawback of the Venn diagram approach is that it does not cater for subtopics and is limited to displaying three categories.

The limitations of the above user interface designs is that they are more suitable for documents that only belong to one category. if a document belongs to multiple categories then all of the categories have to be searched separately.

The design of the user interface is increasingly being used for faceted search whereby the user interface is designed to enhance the user experience in navigating and querying. Faceted search has become the de-facto standard for e-commerce and product related websites. Users can deconstruct large set of results into bite-size pieces and the user can select which section to navigate based on its importance (Lemieux 2009). A facet is a specific perspective on content that is clearly bounded and mutually exclusive. Values in facets can be a flat list (eg shoe size) or hierarchical drill-down multiple levels (eg women->coat->black). Faceted search passes the control to the users and allows them to create custom navigation by combining various facets rather than forcing the navigation through a specific path. This is advantageous because users have such disparate information needs. Also users can make judgements on the relevance of the chosen facets and so the search path is more likely to meet user needs and results can be achieved more efficiently than a search path chosen by the retrieval system.

Care must be taken when designing the interface to not have too many facets, thus avoiding information overload. Use cases and user access patterns must be examined for Facet selection. Faceted search interfaces are ideal for broad but shallow taxonomies because the user has control over displaying only the relevant/important facets and hiding the irrelevant facets. The information from collaborative feedback is being used to create facets based on social tags such as pros and cons of a product.

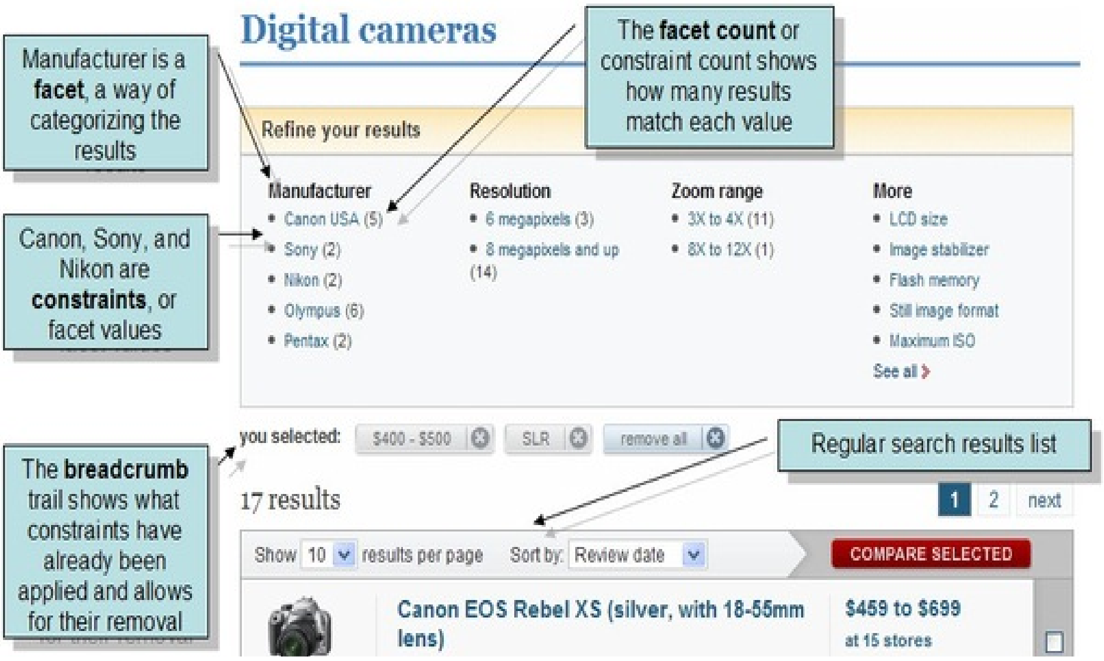


Figure 2.4 Example of Faceted Search.

Figure 2.4 shows an example of faceted search for a digital camera. (http://www.lucidimagination.com/devzone/technical-articles/faceted-search-solr). At any point in time, the user can add/remove filters as necessary.

2.5 Language Models

Language models can be used to represent context and support context-based techniques such as relevance feedback and query disambiguation (Croft et al, 2001). Language models are based on statistical language modelling (SLM) which has been used in related fields such as speech recognition, machine translation and summarization. Each document is represented using a document language model. The language model is a probability distribution that captures statistical regularities of natural language use. In other words, how likely the ith word in a sequence would occur given the identities of the preceding i-1 words. So in a collection of computing documents the probability of 'computer' following 'apple' is more probable than 'pie'. In other words 'apple computer' matches the context of the computing documents whereas 'apple pie' is unrelated. The query is treated as sample of text from a language model. The query is assessed for ambiguity and the level of ambiguity is quantified. A clarity value (non-negative number) is assigned to the query based on how different its associated language model is from the corpus language model. If the query is ambiguous then probable contexts are identified. Sentences that are representative of each context (ie the sentences have high probabilities of generation in those contexts) are shown to the user for clarification. Documents are ranked according to the probability that the document language model could generate the query text.

Xu and Croft (1999) propose a new approach to distributed retrieval by combining language modeling and document clustering. Liu and Croft (2002) state research on language modelling can also be effectively used for passages. Passage level evidence brings benefits for documents that are long or span several subjects. Documents are segmented into different passage types including structural (Hearst and Plaunt 1993), semantic (Hearst 1993), window-based (fixed no of bytes or words) and arbitrary (passage can start at any word in the document, can be fixed/variable length and defined at run-time). A language model is built for each passage. Passages are ranked according to the probability that the query could have been generated by each of them. Documents are ranked based on the score of their best passage. Experiments conclude that passage retrieval is better for long documents and provides more reliable performance than retrieval based on full documents.

Performance might be a problem with language models because they require a large collection of training data for translation probabilities and so it is important that the documents are ranked efficiently. Current language models have several weaknesses (Rosenfeld 2000). Firstly they are sensitive to changes in the style, topic or genre of the text on which they are trained. Secondly language modeling techniques assume some form of independence among different portions of the same document. Such independence assumptions are false and usually lead to overly sharp distributions.

2.6 Ubiquitous Computing

In an increasingly mobile computing environment, the scope of context can be extended to include physical context. Information retrieval in ubiquitous computing is proactive. It tries to predict which information the user will need next and retrieves it at the right time. With the approaches described, information retrieval is making use of physical contextual information to provide enriched results. Weiser (1991) introduced the area of ubiquitous computing whereby information and services could be provided when and where desired as a result of augmenting people and environments with computational resources. In mobile computing environments the physical environment provides important information in establishing the context for the user's information needs. Continuous change in the physical environment implies that the context of user information needs are also constantly changing. Sensory equipment is required to record users physical contextual information such as location, presence, identity and activity (Dey and Abowd 2000). Technical constraints such as low bandwidth means precision is more important than recall. Due to these limitations, Jones and Brown (2003) use the ideas behind link analysis to establish document authority whereby the document with the highest authority is sent to the user instead of sending all documents to the user, even though they may all be relevant.

Personalisation is used to tailor the information delivery to match the interests of a particular user. Eg for mobile tourist application, if the tourist is interested in avoiding traffic jam information then only information which relates to traffic jams specifically on the tourists journey is retrieved from a range of sources such as satellites and the internet.

Hattori et al, (2006) propose two novel methods for query modification based on real world contexts such as geographic location and the objects surrounding him/her, aiming to enhance location-awareness, and moreover, context-awareness, to the existing location-free information retrieval systems. They combine user profile information to come up with 6 relevance statistics: relevance between place name and query, relevance between object name and query, place name and object name ; user and query; user and place name; user and object name.

A problem with context management systems is that they are environment specific and this can cause interoperability problems between different context management systems when the user moves to different environments. Attempts have been made to address this problem using bridging systems which resolve functional differences between context management systems rather than focusing on resolving data model differences (Hesselman et al, 2008, Bunningen et al, 2006).

Recent research shows a rise in the use of ontologies in ubiquitous computing. Hilera and Ruiz (2006) offer a taxonomy for classifying ontologies used in Ubiquitous Computing. Several ontologies exist in this area of computing. One of them is SOUPA (Standard Ontology for Ubiquitous and Pervasive Applications) which offers developers combined vocabularies from different consensus ontologies (Chen et al, 2004). Soupa includes concepts such as Agent (to represent human users), Action, time and device or location. Another ontology is CONON (CONtext Ontology) which models these aspects and provides flexible extensibility to add specific concepts in different application domains (Wang et al, 2004).

In ubiquitous computing, location is an important aspect of context information. Spatial context is when location is modelled. Location models are used to define spatial relations between locations. Choosing a suitable location model for the spatial structure of a context model is important because the possible spatial queries depend on the location model. Space takes into account the position of entities (eg Where is printer HP13?) and the spatial relation to other entities within a distinct area or range (eg what devices are on floor 3 of engineering building?). Nearest neighbour queries such as "where is the nearest printer?" are also possible. Within context-aware mobile applications, location information usually plays a major role for information selection and adaptivity. Ahlers and Boll (2009) explore the geospatial dimension between simple position-aware and fully context-aware information systems by examining in-depth the features of spatial context beyond mere position. Features of spatial context can also be viewport (map visualization to define the extent of the current view), speed (average and current), heading (determine direction of travel or gaze), current time and date (temporal aspects for current location and trip), past track of previous locations (range estimations or coarse prediction of future locations), elapsed duration of a trip, location of the departure point of a trip and spatial environment features such as road networks which help in understanding user movements in a given environment. They describe how these features can be used to create spatial queries in a mobile information retrieval system and further discuss the influence of spatial context to select and adapt the query results and its relation to mobile user's information needs.

There are several approaches to defining spatial context in ubiquitous computing. Schilit et al, (1994) proposed a taxonomy which names *spatial context* (where you are), *social context* (who you are with), and *computing context* (what resources are nearby). Some consider this taxonomy to be incomplete (Chen and Kotz 2000). Giaglis (2003) put forward a classification framework / taxonomy of mobile location services. Dix et al. (2000) set up an algebraic specification for the type *space*, consisting of location, nearness, and regions, and for the type *world*, consisting of spaces and bodies. With these elements they set up a kind of top-level ontology of spatial context in the environment in the mind of a mapping system. Becker and Nicklas (2004) argue the case for a combined approach providing the efficiency of context management through spatial context models combined with the semantic expressiveness of ontologies. Efficient processing depends on the underlying spatial structure and the involved co-ordinates of the positioning system. Instead of providing a new taxonomy. Freksa, Klippel and Winters (2005) offer an alternative formal approach to defining spatial context by using cognitive approaches. They provide an operational definition of context by examining how contexts are created and used. Context information is implicit in the environment, from which cognitive agent (user) creates mental map in his mind and a physical map. A map enables a cognitive agent to solve spatial problems that cannot be solved by inspecting the environment nor by inspecting its mental representation.

Temporal context relates to modelling time as a way of determining context. Even though, temporal context might still be appropriate for desktop based applications such as news retrieval, we have included it within ubiquitous computing because the vast majority of applications which require temporal context are mobile computing applications. According to Mozetic and Bojadzijev (2006), temporal information is implicit context information. They use a formal algebraic model to record temporal information and make it more explicit. The algebraic model takes into account two aspects of time namely, linearity and granularity. Granularity refers to the level of time detail required. So for example, high granularity would be days and low granularity could be months. Their system will use the time-stamp part of the metadata about news items and temporal models of the events reported, to distinguish related news items from unrelated ones.

Mizzaro et al, (2008) present a benchmark approach for the evaluation of applications retrieval in a context-aware environment. The results show that Description domain has the highest Mean Average Precision (MAP) because it contains a textual description. The fields with lowest MAP are time and location because of the difficulty carrying out exact matches on time which can be represented by a label (eg morning, afternoon) or numerical time value. Mountain and Macfarlane (2007) identified 4 geographical post query filters: spatial proximity, temporal proximity, speed heading prediction services, visibility (locations that can be seen). Spatial proximity assumes a closer information source to be more relevant. Figure 2 shows an image of point features displayed over a backdrop map.

Figure 2.5 Point Features displayed over a backdrop map

Temporal proximity assumes that regions reached in shorter time period to be more relevant. Users found information sorted by spatial proximity more relevant than prediction surface of likely future locations. So spatial context is very important in mobile applications. They point out another useful relevance measure could be Cognitive relevance whereby users score the returned results in terms of informativeness, novelty and quality.

Ubiquitous computing still involves several challenges such as having the ability to deal with information from multiple sources in different formats and delivering precise information in a timely manner. Other hurdles relate to the cost of the sensory equipment; user familiarisation and acceptance of the equipment, and finally what backup mechanisms exist in the event of equipment failure.

2.7 Query Expansion

Query expansion is needed due to the ambiguity of natural language and also the difficulty in using a single term to represent an information concept. Krovetz and Croft (1992) observed that most benefit is achieved with high-recall searches that depend on matches of single concepts.

With query expansion, the user is guided to formulate queries which enable useful results to be obtained. The main aim of query expansion (also known as query augmentation) is to add new terms (automatically derived from the context) to the initial query. This process of adding terms can either be manual, automatic or user-assisted. Manual query expansion relies on user expertise to make decisions on which terms to include in the new query. In the case of automatic query expansion, weightings are calculated for all terms and the terms which have the highest weighting are added to the initial query. Different weighting functions produce different results, therefore retrieval performance depends on how the weightings have been calculated. With user-assisted query expansion, the system generates possible query expansion terms and the user selects which of these to include.

The new terms resulting from the chosen term selection method should provide contextual information for the initial query with a view to improving the retrieval results. The contextual information for resolving ambiguities and expanding queries can be acquired from relevance feedback, term co-occurrence and more recently it has been derived from knowledge models such as ontologies. The following sections explain how each approach works.

2.7.1 Query Expansion Using Relevance Feedback

Relevance Feedback is obtained by modelling the user's interests in a single search session. It involves the modification of the initial query using words from top-ranked or identified relevant documents. Relevance feedback is a fairly established technique for modification of the initial query using words from top-ranked or identified relevant documents (Salton and McGill 1983). Based on user query and document corpus, possible contexts for the query are inferred and used to suggest additional terms. It is an easier way of improving the retrieved document set as opposed to the user having to construct a new query.

The relevance feedback loop requires the user to enter an initial query which results in a display of ranked documents (usually titles/abstracts). From this display, the user makes relevance judgements and selects the relevant documents. The relevant terms from these documents are added to the initial query. An alternative to this is pf where the top ranked n documents are assumed to be relevant. Terms from these documents are selected and used for expanding the query. Whether pseudo-relevance feedback is used or traditional relevance feedback, the term selection method is a key factor in the performance of expanded queries. It needs to consider how to weight the new terms, whether to exclude the original query terms, whether to include all of the new terms or just some of them and if so how many new terms to include. These issues are described below.

2.7.1.1 Selection Of New Terms

It is possible that the query expansion process generates such a large number of candidate terms that it might not be practical to use all of these terms. Some research has been carried out on the optimum number of terms to include and there are differing viewpoints ranging from a one-third of the terms as suggested by Robertson and Willett (1993), 20 terms (Harman 1992) to massive query expansion (Buckley et al, 1995). In the latter, 300-530 terms were added to the original query. The terms came from known relevant documents / top retrieved documents and seemed to improve effectiveness from 7%-25%. Buckley concluded that massive query expansion is effective for routing ie moving the query vector towards the centroid of relevant documents and away from non-relevant documents. A contrasting view provided by Sihvonen and Vakkari (2004) is that the number of terms used for query expansion is less important than the type and quality of the terms chosen. Vakkari et al, (2003) compared interactive query expansion with automatic query expansion based on relevance feedback. Interactive query expansion is better if all retrieved relevant items were counted otherwise it made no difference if only those items recognised relevant by users were observed. Billerbeck and Zobel (2004) systematically studied the effect of the number of query expansion terms on performance. They concluded that one size does not fit all because the optimal number varies from query to query. Ruthven and Lalmas (2003) state that domain specific collections perform better with relevance feedback than domain independent collections because it is easier to select good expansion terms from this type of a collection or because the ambiguity of search terms is less significant. So a consensus view does not exist on the optimum number of selected terms to be used for query expansion.

Collaborative learning is a more recent term selection approach whereby term concepts learned by other queries can be used for query expansion (Klink et al, 2002). The relevant document set is an accumulation of the relevant document sets for all queries which contain a term from the original query. Terms which are most similar to the concept of individual query terms are selected as opposed to selecting terms that are similar to the entire query. Results of this approach look positive.

2.7.1.2 Weighting Of New Terms

In answer to the question whether all terms in the expanded query should have equal weighting or whether the new terms should have a higher/lower weighting, Voorhees (1994) found that assigning lower weights to added concepts enhances retrieval accuracy. She used a factor between 0 and 1 for weighting added terms. The original f4.5 formula (Robertson and Sparck-Jones 1976) was modified and renamed f4modified (Robertson 1986). The f4modified formula took into account the addition of new terms to the original query and achieved the desired effect of bringing terms with low frequency to the top of the ranked list. In 1990, Robertson developed a new ranking algorithm which was based on the Probability Ranking principle (Robertson 1990). This principle requires for each document an initial relevance judgement from which the probability of a document being relevant to a query can be calculated. According to this principle, the retrieved documents are presented in decreasing probability of relevance to the user who submitted the query.

The inclusion of term t in the search formulation with weight w_t will increase the effectiveness of retrieval by

$$WPQ_t = w_t(p_t - q_t) \qquad (2.1)$$

(where, w_t is a weighting function, which in this case is the F4.5 formula; p_t is the probability of term t occurring in a relevant document; and q_t is the probability of a term t occurring in a non-relevant document.

$$WPQ_t = \log \frac{(r + .5)(N - n - R + r + .5)}{(n - r + .5)(R - r + .5)} \cdot \left(\frac{r}{R} - \frac{n - r}{N - R} \right) \qquad (2.2)$$

where N is the total number of documents in the collection; R is the sample of relevant documents as defined by the user's feedback; n is the number of documents indexed by term t; r is the number of relevant documents (from the sample R) assigned to term t.

Rocchio's method (1971) used an information retrieval system based on the vector-space model whereby documents are represented as vectors in the information space and so are queries. The vectors are weighted, with higher weightings being given to relevant vectors. After relevance feedback the weights are adjusted and only those new terms that occur in the initial query or occur in at least half of the relevant documents are added to the original query. This produced positive results.

An extension of this approach is based on predictive algorithms for collaborative filtering (Hoashi et al, 2001). The query-document similarity is used for calculating term weights during query expansion. The user database of votes is viewed as a set of vectors where each vector expresses a user and the elements of each vector express the votes of the active user. The main aim of the collaborative filtering process is to predict the values of empty elements of the active user's vector. Such predictive algorithms are showing signs of effectiveness.

2.7.1.3 Sample Size Of Relevant Documents

Efthimiadis (1996) describes the various approaches adopted by different researchers with respect to sample size. Okapi uses a sample size of 3 documents, some recommend a sample size of 5 documents and others use a larger sample size. There is no clear effect of the sample size of relevant documents on the search performance and the area of selecting the optimal sample size is still an open area of IR research. In the absence of relevance judgements, a common alternative is to have a cut-off point in the form of 'top X documents' which are all treated as relevant. The sample size of relevant documents can be selected from the documents retrieved prior to the cut-off point. In many relevance feedback experiments the sample is defined at a cutoff level of the 10 or 20 top-ranked documents.

Relevance feedback is not very successful in search engines with the closest feature being "find more like this" (Croft et al, 2001). The main problem is trying to get users to provide relevance information. Simply indicating 'Relevant/Not relevant' does not give users enough incentive. Instead user feedback can be of different types (Spink 1997). Users can give feedback on the content of the retrieved documents; state which of the relevance feedback terms are relevant; give feedback on the magnitude of the retrieved set as to whether it is too large, too small or just right; and give feedback on terminology by judging the relevance of the terms in the inverted file. Robertson (1990) also found that when a user is asked to make a relevance judgement this is slightly ambiguous. For a given relevance judgement, it is important to distinguish between whether it is relevant to the topic and/or relevant to the user. Some research has been done on passive measures such as time spent browsing a page or number of links followed from a page. If the user can indicate relevant sections or even phrases in a document, relevance feedback is more accurate. This implies we need more feedback from users not less. Successful application of relevance feedback involves sophisticated interface design and good algorithms for inferring context.

In conclusion, effectiveness of query expansion using relevance feedback can vary depending on many factors such as choice of parameters in the term weighting process, number of relevant documents in the document collection, facilities provided for users to give good quality relevance feedback with ease and finally whether the collection is domain specific or domain independent.

2.7.2 Query Expansion Using Term Co-Occurrence

Previous research concentrated on obtaining context from the document collection using techniques such as stemming and clustering. Stemming is a process whereby all variations of a term are generated by adding/removing prefixes and suffixes as appropriate. If the stemmer over-stems the words this can have an a negative impact on overall performance (Ogawa et al, 2000)

During the late 60s and early 70s, query expansion using term clustering was investigated by many researchers. Latent Semantic Indexing (LSI) uses term clustering and was one of the early techniques for global query expansion (Deerwester et al, 1990). Similar documents are placed in a cluster. It is assumed that similar documents are relevant to the same requests. On this basis, if the query terms mapped onto one or more clusters the terms from the cluster(s) would be used for expanding the query. The usefulness of this technique was found to be marginal due to poor clusters resulting from small document collections (Lesk 1969) or insufficient differences in vocabulary between relevant and non-relevant documents (Sparck-Jones 1973). Another problem with clustering is that it works on the assumption that a term can only belong to one cluster. For ambiguous terms this may not be the case (Deerwester et al, 1990).

Another area of investigation to assist in inferring context is that of Term co-occurrence which refers to two or more terms that are situated next/near to each other in the source document. Smeaton and Van Rijsbergen (1983) ran experiments which used new terms generated from sources such as maximum spanning trees and found very little improvement. One explanation of this is that similar terms have comparable frequencies. Query terms have high collection frequencies and as a result so will the candidate terms (Peat and Willett 1991). Since high frequency terms do not discriminate between relevant and non-relevant documents the addition of these terms for query expansion is ineffective. Expansion of phrasal terms results in improvement when used with longer queries (Ogawa et al, 2000).

Other work looks at using two complementary term suggestors. Schatz et al, (1996) used subject thesauri and co-occurrence lists. The thesauri are generated by human indexers who based on their subject knowledge decide where in a subject hierarchy a term should be placed. The co-occurrence lists are computer generated and terms are placed in frequency order of co-occurrence. The overall conclusion was that multiple views results in a better quality search. This is supported by Mandala et al, (1999) who also suggest that it is better to use a combination of query expansion techniques than to use a single technique. Similar findings were made by Huang et al, 2005 as a result of conducting experiments which combined conceptual indexing and keyword indexing.

Phrases together with pf work better than phrase expansion alone (Eguchi 2005). This is supported by Vechtomova et al, (2003) who uses long-span collocates where words significantly co-occur in topic-size windows with query terms. Global collocation analysis where collocates of query terms are extracted from the entire collection. Secondly, local collocation analysis is carried out to extract terms from a subset of retrieved documents. The experiments showed that global collocation analysis performed worse than unexpanded queries. This may be due to the fact that terms extracted from the global collection are too general and they need to have a more specific context. The local collocation experiments produced better results. However if the top retrieved documents are not rich enough they may exclude documents which are still relevant to the query.

Xu and Croft (2000) show that global analysis techniques such as word context and phrase structure on the local set of documents produce results that are both more effective and more predictable than simple local feedback. Figure 2.6 shows the top 30 concepts retrieved by Phrasefinder for the TREC4 query 214 “What are the different techniques used to create self induced hypnosis”.

hypnosis	meditation	practitioners
dentists	antibodies	disorders
psychiatry	immunodeficiency-virus	anesthesia
susceptibility	therapists	dearth
atoms	van-dyke	self
confession	stare	proteins
katie	johns-hopkins-university	growing-acceptance
reflexes	voltage	ad-hoc
correlation	conde-nast	dynamics
ike	illnesses	hoffman

Figure 2.6 top 30 concepts retrieved for TREC4 query 214

Another property which affects the complexity of the query expansion algorithm is whether the phrase is compositional or non-compositional. With compositional phrases each term in the phrase can be expanded using substitute terms and the final expanded phrase will retain its meaning whereas non-compositional phrases (or idiomatic phrases) phrases which take on meanings that go beyond the meanings of their parts. Lin (1999) defines a non-compositional phrase as a phrase where 'its mutual information differs significantly from the mutual information of phrases obtained by substituting one of the word in the phrase with a similar word'. For example, the phrase 'red tape' can be used to refer to bureaucracy. If we replace the colour red with another colour eg 'yellow tape', this does not achieve the same meaning as the original phrase. An example of a compositional phrase is 'search engine'. Cui et al, (2003) identify compositional phrases using n-grams from query logs. They filter out those phrases that do not appear in the documents. The general findings show that short phrases are a more accurate representation of information requirements.

Also phrases have a higher inverse-document frequency because terms are more common in document collections than phrases. If phrases offer more selectivity due to higher inverse document frequency then, the area of phrase based query expansion has further scope for investigation.

2.7.3 Query Expansion Using Lexical Networks

Lexical networks are another important source for deriving context. Lexical networks containing domain-specific vocabularies and relationships are automatically extracted from the collection and play an important role in this navigation process. Lexical relationships between terms are utilised to suggest additional terms. Text analysis tools can be used to develop the vocabulary for the lexical network. Some systems indicate the strength of the relationship between two terms. The general tendency is to disambiguate terms during the search process and not store the disambiguated terms prior to the search.

In the early sixties, word sense was frozen into the lexicon. In other words the word sense could not be updated afterwards. However Pustejovsky's work stems from the realization that full lexical knowledge comes from the texts themselves (Pustejovsky

1995). He proposed the use of a generative lexicon to disambiguate word sense. A generative lexicon uses machine-readable dictionaries and large text corpora to construct a core lexical engine which acquires new lexical entries and refines existing ones through statistically based corpus acquisition methods. Coates-Stephens (1991) established an algorithm for obtaining the meaning of proper nouns from the text. Callan et al, (1992) describe a retrieval system (INQUERY) that provides support for complex query formulation. The system is based on a type of probabilistic retrieval model called the inference net. The inference net has two component networks, one for the document collection and one for each query. Probability can be represented by weights on arcs. Query nodes are linked to concept nodes and these in turn are linked to the document nodes. The concept-document node link can be true or false depending on whether the concept is observed in the document. Therefore relevant documents can be retrieved by following the arcs from the concept nodes which have a true value to the associated document nodes.

Barzilay and Elhadad (1997) claim that WordNet was not created as a tool for automatic text processing because it lacks necessary information such as relations between different parts of speech and conceptual relations such as situation-participant, situations-domain or object-property (Climent et al, 1996). Instead they use WordNet relations to construct lexical chains. A lexical chain is a chain of words in which the criterion for inclusion of a word is some kind of cohesive relationship to a word that is already on the chain (Morris and Hirst 1991).

Finkelstein et al, (2002) use a semantic network to derive the context for ambiguous queries. Keywords are extracted from the surrounding text to augment the query. This involves semantic keyword extraction and clustering to generate new queries. The results show that using context to guide a user's search process offers definite improvements.

Instead of string based searches, concept based searches are carried out on concept hierarchies which generally produce more useful results than the former (Jarvelin et al, 2001). Sanderson (2004) introduced a query expansion technique that makes use of concept hierarchies which have been generated from the document collection.

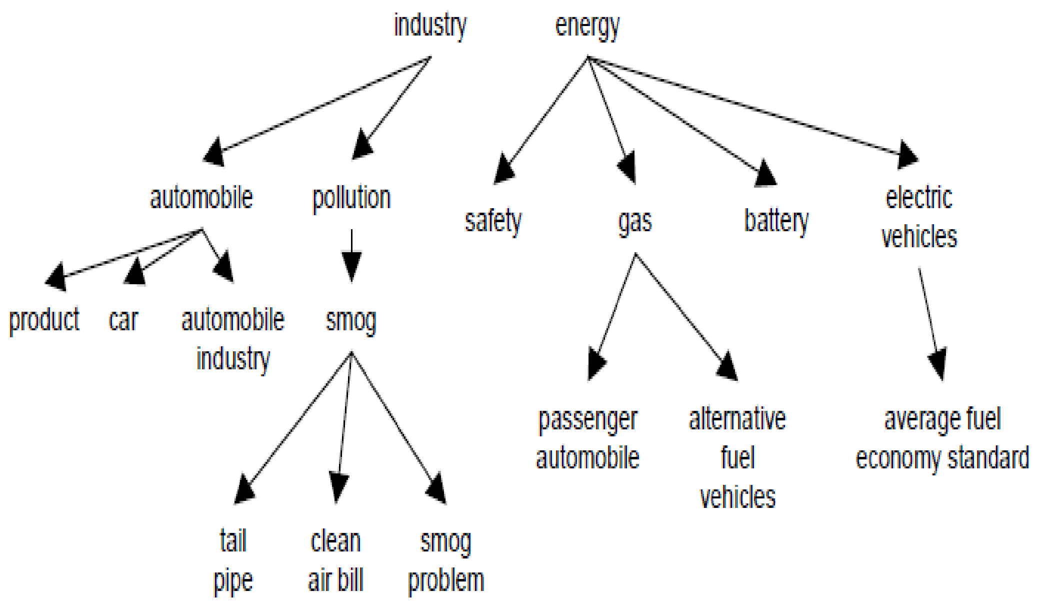


Figure 2.7 Fragment of concept hierarchy from TREC topic 230

Chu et al, (2002) also map the query terms to general concept terms which have been mined from the corpus and for the query expansion process substitute general concept terms by a set of specific concept terms. Since the expanded query matches with the document index terms much better, experimental results reveal that such query expansion produces better retrieval effectiveness than the unexpanded ones. The downside of this approach is that specific terms reduce recall. This approach is only suited to situations when precision is of more importance than recall.

For static collections this is a reasonable approach to use however if the document collection is not static then the concept hierarchies have to be generated repeatedly, which is a time consuming activity. Another issue to be resolved in relation to this is whether to archive previous versions of concept hierarchies and their document collections in case the user wants to return to older queries.

Tombros et al, (2001) found that good quality expansion terms can only be generated if the original document collection contains a large number of relevant documents. If the set of relevant documents is small then the choice of query expansion terms can be poor. An additional factor which affects the quality of expansion terms produced is that the results will vary depending on which summarization technique is used. An ideal solution should be able to produce good results and not be dependent on the document collection or the choice of summarization technique. Global techniques and local techniques can be used for automatic query expansion. A comparison of the two techniques was carried out by Xu and Croft (2000). The problem with global techniques is that of performance because of the time taken to analyze the whole collection to discover word relationships. The local techniques focus on the most frequently occurring terms in the top ranked documents. Xu and Croft found that local techniques only work if all of the top-ranked documents are relevant. The performance improvement by local context analysis over local feedback is due to a better metric for selecting expansion terms. Corpus dependent query expansion is more suitable for static document collections. For web collections, the knowledge models would have to be constantly updated/regenerated because the collections on the web are more dynamic/fluid in nature (Huang 2000).

A commonly experienced problem of using a lexical network for navigation and display purposes is that depending on the degree of complexity, users might find it overwhelming to navigate. A common problem is the size of the lexical network frequently too big to fit onto one screen. If leveling is used to overcome this problem, it is quite easy for the user to 'lose' their way in the vast information space. Added to this is that if the lexical network has a large number of nodes then users are faced with choosing the right node to satisfy their information need. If the correct node is not selected then the users might spend longer traversing the network until the desired results are reached.

2.7.4 Query Expansion Using Ontologies

Ontologies have been used for information retrieval in various fields such as medical domain (Leroy et al, 2000; Goble et al, 1998; Abasolo and Gomez 2000); the legal field (Lame 2003; Sais 2002); image retrieval (Khan and Wang 2002; Hyvonen 2003); and cross language information retrieval (Pavel et al, 2003; Leger et al, 2001; Grover et al, 2003).

Ontologies have been used for a range of information retrieval tasks such as thematic summarization (Loukachevitch and Dobrov 2000); word sense disambiguation (Gonzalo et al, 1998); query formulation (Suomela and Kekalaninen 2005); indexing (Stairmand 1997); text classification (Scott and Matwin 1998); browsing (Davies and Weeks 2004); image retrieval (Khan and Wang 2002), (Hyvonen 2003); and for query expansion.

2.7.4.1 Ontology Definition

An ontology is a model of reality, it is not reality itself. In a natural language, a word may have multiple meanings depending on the applicable context. The purpose of an ontology is to provide a context for the vocabulary it contains. In a computer system, context may be represented and constrained by an ontology. Therefore, an ontological model can effectively disambiguate meanings of words from free text sentences. Ontologies provide consistent vocabularies and world representations necessary for clear communication within knowledge domains (Leroy et al, 2000).

An ontology is more accurately described as "a classification, thesaurus or a set of concept clusters" ((Bates, 2002; Soergel, 1999). Another definition of ontologies is 'classifications, lists of indexing terms, or concept term clusters' (Communications of the ACM, 2002). Ontologies improve the accuracy in fuzzy information search and facilitate mono- and multi-lingual human-computer dialogues by paraphrasing the query of the user through context identification and disambiguation (Leger et al, 2001). (Gruber1993) defines an ontology to be a 'specification of a conceptualisation'. Gruber explains that ontologies were first used in philosophy then Artificial Intelligence.

Ontologies range from general to domain-specific. WordNet, EuroWordNet and Cyc are examples of a general ontology. Many domain-specific ontologies exist for example in the medical and legal domains. (Buckland 2003) distinguishes between three different kinds of ontology. Axiomatic ontologies contain abstract concepts that facilitate reasoning. Terminological Ontologies composed of structures of lexicalized concepts and Domain ontologies which represent the knowledge organization systems used for documentary purposes such as Dewey Decimal Classification system which uses a numerical notation to denote an unlimited number of topics and has an English language index to the classification numbers. Similarly, Davies et al (2009) define different ontology categories, namely schema-ontologies, topic-ontologies and lexical ontologies which are all based on the different types of semantics being modelled. Figure 2.8 shows the schema ontology for person and instantiations of the schema ontology properties.

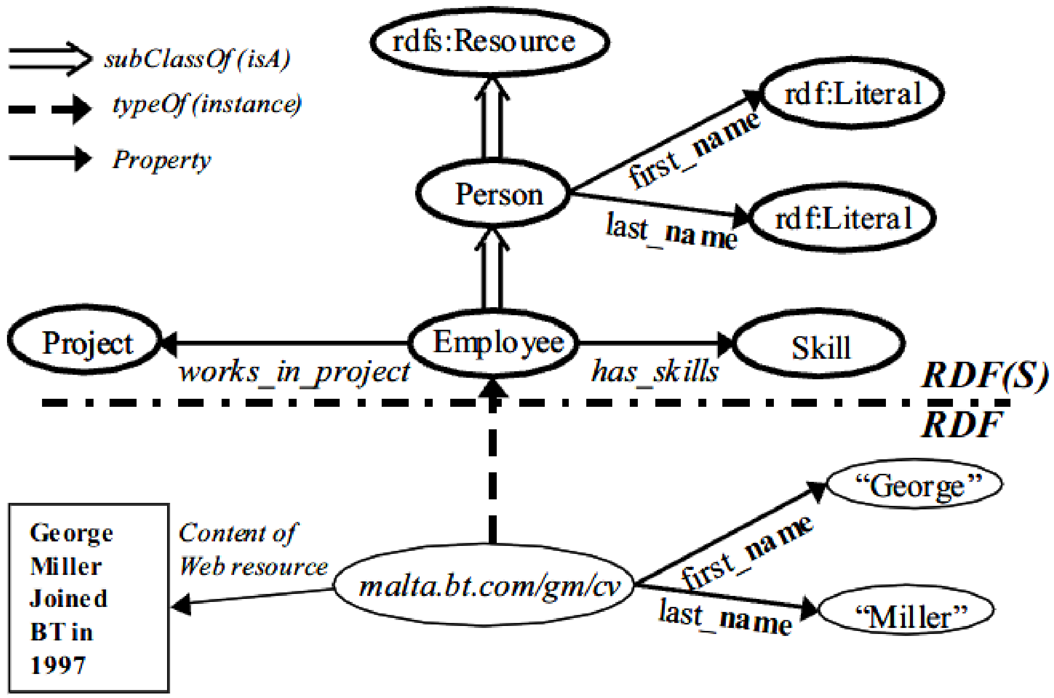


Figure 2.8 Schema Ontology and instantiation using data values

Ontologies are not without their problems. The first issue is related to vocabulary mismatch between the query terms and the concepts in the ontology. A mapping process needs to take place to overcome this problem. Secondly, if an ontology for a particular domain does not exist then a lot of effort is required to construct ontologies from scratch not just from a technical point of view but more importantly the process of knowledge extraction from domain experts and arriving at a consensus view. The design and construction of domain ontologies is labour intensive, time consuming and difficult (Kashyap 2001). Kashyap proposes a methodology for creating an ontology with minimal involvement of the domain expert by reusing readily available information such as schemas, queries, data dictionaries and thesauri. Hwang (1999) proposed one method for automatic generation of ontology started from the seed-words suggested by domain experts. This system collected relevant documents from the Web, extracted phrases containing seed-words, generated corresponding concept terms and located them in the 'right' place of the ontology. Several kinds of relations are extracted: is-a, part-of, manufactured-by or owned-by etc. It also collects "context lines" for each concept generated, showing how the concept was used in the text, as well as frequency and co-occurrence statistics for word association discovery and data mining. The drawback is that it fully depends on the seed-words provided by the domain experts. (Lame 2003) presents a method to identify ontologies' components. The method relies on text analysis to extract concepts and relations among these concepts. (Saias 2002) states that for certain areas such as legal domains, web legal information retrieval systems need the capability to reason with the knowledge modelled by legal ontologies.

Advantages offered by ontologies are that they are readily available (Bateman 2005); they routinely include proper nouns: personal names and place names; many software tools exist to automate the creation and support the evolution of ontologies; finally most ontologies have been defined in a portable language such as XML (eXtensible Markup Language). Also the extensible features of XML can be used to capture changes in the evolving ontologies.

2.7.4.2 Ontologies for Query expansion

As previously mentioned, ontologies have been used for a range of information retrieval tasks. In our research we focus on the use of ontologies for query expansion.

The problem with traditional relevance feedback techniques and corpus dependent query expansion is that they are content driven. The corpus content is analysed to extract candidate terms for query expansion. This can only work if there are sufficient relevant documents to work with and also that these documents contain a reasonable set of terms that represent the subject area for the query. Corpus independent knowledge models do not suffer from this drawback.

Corpus independent knowledge models can be in the form of a thesaurus or an ontology although according to some, the distinction between the two is blurred.

Before discussing the benefits of ontologies, we will examine case studies of query expansion using general ontologies followed by query expansion using domain specific ontologies. The last section reviews some existing news applications which support a range of different tasks such as news annotation, news categorization/classification, news summarization, browsing/searching and personalized news. The review shows various news applications making use of news ontologies in different ways.

2.7.4.3 Case Studies Of Query Expansion Using Domain-Independent Ontologies

Several domain independent ontologies or Upper Level ontologies exist such as Cyc and WordNet. Much effort has been invested in developing Cyc. Cyc is a domain-independent ontology which attempts to capture and encode large amounts of common sense knowledge about the real world. It contains more than 4.5 million assertions (facts and rules) describing more than 250,000 terms, including 15,000 predicates. It contains temporal knowledge, spatial knowledge, event information, geography information and othe general information such as emotional information. A version of the Cyc ontology, called ResearchCyc, has been released for the scientific community. ResearchCyc contains both intensional information (entity types, relationship types, integrity constraint), extensional information (representation of individuals and their relationship to space, time and human perception) and linguistic information. In particular, the ResearchCyc lexicon contains entities for over 20,000 single-word noun, verb, adjective and adverb forms, 40,000 multi-word phrases, and over 100,000 proper names. Unlike dictionary-based taxonomies such as WordNet, in which every node is identified with a word sense, the Cyc ontology is not an ontology of word senses.

Conesa et al, (2008) state that linguistic repositories do not capture semantic relationships or integrity constraints between concepts, and semantic repositories do not represent linguistic relationships of the concepts. They combine semantic knowledge about different application domains from Research-Cyc as well as linguistic knowledge from WordNet to support query expansion. They present a query classification scheme that explains why in some cases the query expansion may not be done successfully. They distinguish between intensional queries and extensional queries. Intensional queries directed by the need for the user to increase his knowledge. Ontologies are suitable for this type of query. Extensional queries such as “Where can I buy Nike shoes in Georgia?” is where the user is interested in learning specific knowledge to solve a practical information need. The user is not interested in general facts about Georgia, Nike or USA. Knowledge repositories are poor at handling extensional queries. Curtis et al, (2006) believe query expansion needs to deal with domain related knowledge, common sense inferences and semantic relationships of the query concepts. They use Cyc to create a semantic method for word sense disambiguation. The benefit of this method against the ones that use more linguistic ontologies, such as Wordnet, is the use of a much richer set of relations among word senses, enabling the development of more robust mechanisms for determining semantic closeness. The Cyc taxonomic knowledge has been used in other approaches to facilitate word sense disambiguation with human intervention. Curtis et al, (2005) also use Cyc to carry out deductive Question Answering and produce results superior to what each question answering technique could produce alone.

WordNet has been a popular general ontology used in the area of query expansion as the following works show. (Gonzalo et al, 1998) use a manually disambiguated test collection of queries and documents derived from the SEMCOR semantic concordance. Their experiment covers three types of index spaces: original terms; word senses derived from manual disambiguation and finally WordNet synsets. The authors observe that if queries are not disambiguated, indexing by synsets performs only as good as standard word indexing. According to Gonzalo, indexing with word sense improves information retrieval by more than 29%.

Voorhees (1993) carried out experiments to exploit the semantics contained within WordNet to improve retrieval effectiveness by indexing with word senses instead of word stems. The results showed that the effectiveness of the vectors produced by this disambiguation technique was worse than word stem vectors for all five collections. The findings indicate that short query statements can be difficult to disambiguate because the IS-A hierarchy is not sufficient to reliably select the correct sense of the noun.

Incorrect sense resolution or query expansion using automatically generated synsets resulted in missing correct matches which in turn have a deteriorating effect on retrieval performance than using spurious matches. Voorhees (1994) used WordNet to conduct experiments on small single domain TREC collections. The results suggest that query expansion can improve problems of mismatched vocabularies especially in cases where the terms that are expanded are lexically related to the query terms. Figure 2.9 shows the six WordNet senses for "swing". The most relevant synset is added in order to expand the query. In contrast, query expansion makes little difference in retrieval effectiveness for long queries because they usually contain a full description of the information request. Voorhees found that some long queries could benefit from other techniques such as relevance feedback.

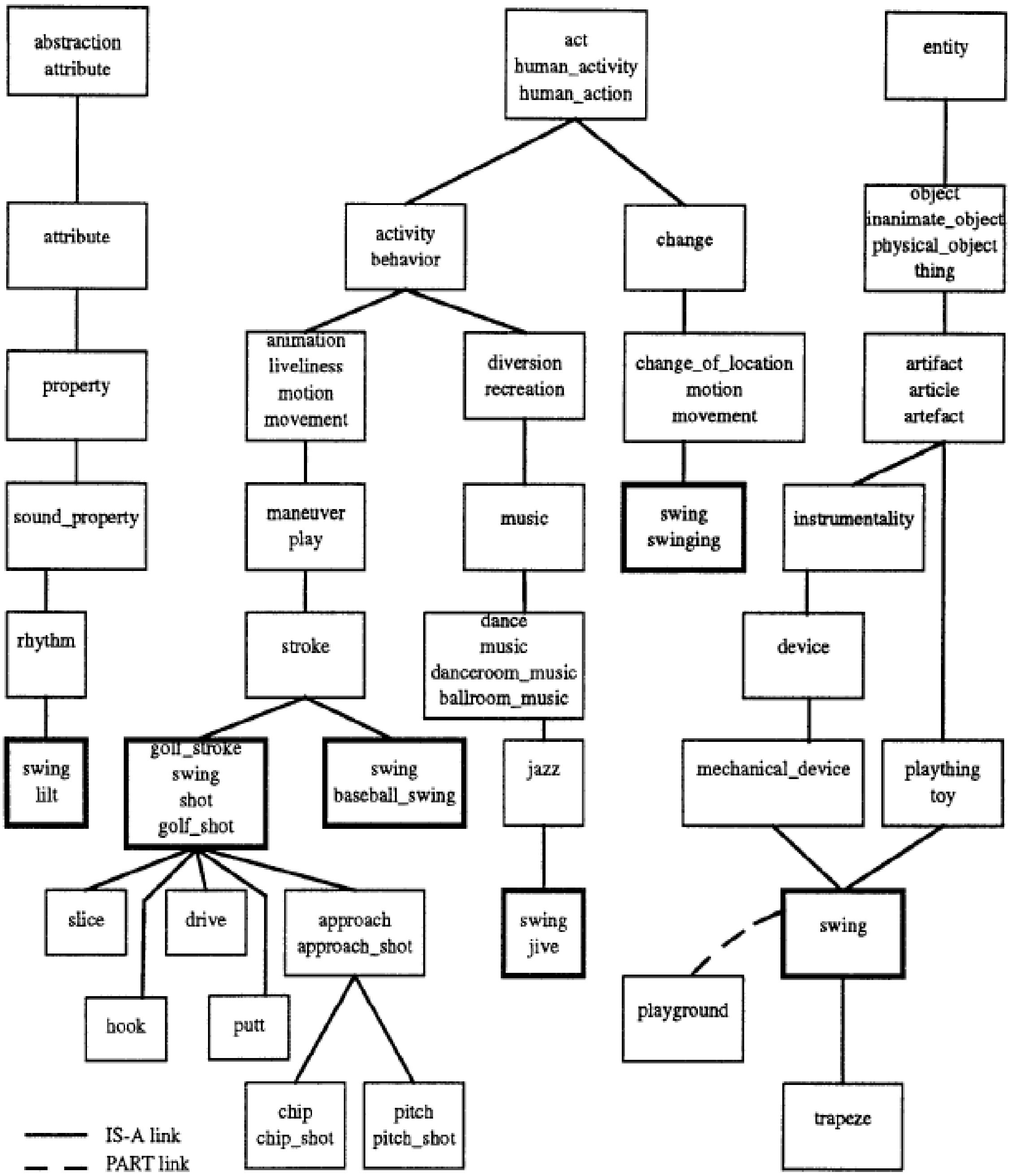


Figure 2.9 Relations defined for the six senses of the noun "swing" in WordNet.

Hearst (1992) describes a pattern-matching method for the automatic acquisition of the hyponymy lexical relation from unrestricted text. For example "Bruises, wounds, broken bones or other injuries …" the method infers that each of the items on the list are hyponyms of the term injury. The hyponymy relation is also used to identify the general meaning of unfamiliar noun phrases eg hyponym ("broken bone", "injury"). The unfamiliar noun phrase "broken bone" is unlikely to appear in a dictionary but due to the hyponym relation, "broken bone" can be understood at some level as an injury without determining the correct senses of the component words. This approach avoids the use of pre-coded knowledge and can be applied across a wide range of text. According to Hearst, list items are usually similar so they are used to indicate synonyms and can be used for synonym expansion. Instances of the hyponymy relation ("is-a") that are found using this approach are compared with the relations in existing lexicons such as WordNet's noun hierarchy. The results show a high level of similarity between Hearst's approach and WordNet's noun hierarchy.

Finkelstein et al, (2002) describe Intellizap which is a context based search system. Query ambiguity is eliminated by deriving the context from the text surrounding the marked query in a given document. Keywords are extracted from the surrounding text to augment the query. This involves semantic keyword extraction and clustering to generate new queries. The expanded query is then submitted to various search engines. The results are then re-ranked. The system uses a semantic network for measuring the distances between words. It uses the vector space model to represent 27 domains. Linguistic information (such as hypernym and meronym) is obtained from the WordNet dictionary. The system combines the WordNet metric and the correlation metric to statistically analyse the relationship between words. The results show that using context to guide a user's search process offers definite improvements.

Navigli and Velardi (2003) use sense information and ontologies for query expansion. They argue that expanding with synonyms and hyperonyms has a limited effect on web information retrieval performance. They suggest that other types of semantic information derivable from an ontology is more effective such as gloss words and common nodes. This is because words in the same semantic domain and same level of generality are best candidates for expansion. The ontology is used to extract the semantic domain of a word and then the query is expanded further using co-occurring words. Effectiveness of using ontologies to improve retrieved results depends on the type of task (ie subject finding vs. site retrieval) and query length. They concur with the view that query expansion is suitable for short queries. Their experiments used TREC 2001 web track, WordNet 1.6 for the ontology and Google. A semantic network is created for each word sense. The relevant semantic networks are then intersected pair-wise and a score is assigned to the configuration based on the number of common nodes where common nodes are those nodes in the semantic network which can be reached by both semantic network centres through directed paths. The experiment results showed a systematic improvement over the unexpanded query.

Baziz et al, (2005) state ontology based information retrieval is promising in increasing the quality of responses since document semantics are captured. They used a small document collection and document content is represented using an optimum semantic network called document semantic core. WordNet concepts (which include words and phrases) are extracted and then globally disambiguated with reference to document terms to produce the optimum semantic network. Similarity measures between connected nodes weight the links. Four similarity measures are used, of which the first three are 'is-a' based and the fourth one is based on gloss overlaps. Scored concepts are used for document conceptual indexing. In automated word sense disambiguation, the challenge is to keep relevant concepts and discard irrelevant concepts.

Jones et al, (1995) analysed the INSPEC thesaurus and used 8 relational database tables to store thesaurus information such as terms, equivalence, hierarchical, associative, components, classes, facets and words. Their experiments found no correspondence between the number of terms chosen and the query performance. Also the other main finding was that the number of terms selected is dependent on the distance from the original node. In other words there was a tendency for nodes with a short distance from the original node to be selected. Jones et al, (1995) state that one of the main success factors in using a thesaurus for query expansion is to have a good match in the thesaurus to start with. For this reason terms should be expanded before carrying out the search of the document database to overcome the fact that there might not be an exact match of query terms with a WordNet node (Jones 1993). Contrary to Efthimiadis (1996) who suggests that query expansion using knowledge models may be done without term re-weighting, Jones et al, (1995) suggests that term weighting is a good idea. A reference is made to Shoval who bases the weightings on criteria such as number of connections; type of relationship; number of co-occurrences and path length (Shoval 1985). A further reference is made to Chen who suggests that users should be able to adjust the term weightings (Chen et al, 1993). Jones found that quality of the thesaurus is paramount. A thesaurus which has greater coverage, depth and accuracy has improved chances of producing better results.

Grootjen and van der Weide (2006) generate a local thesaurus by projecting a global thesaurus onto the top ranked documents resulting from an initial short two word query. A concept lattice is produced for the local thesaurus, however to solve the problem of finding a good starting concept, the user has to navigate the lattice. Figure2.10 illustrates the navigation process for query 59. Navigation starts at the top node (empty expansion). The user selects flood as an expansion candidate. By doing so the mean average precision rises form 0.0231 to 0.0725. The process continues until the end concept is reached.

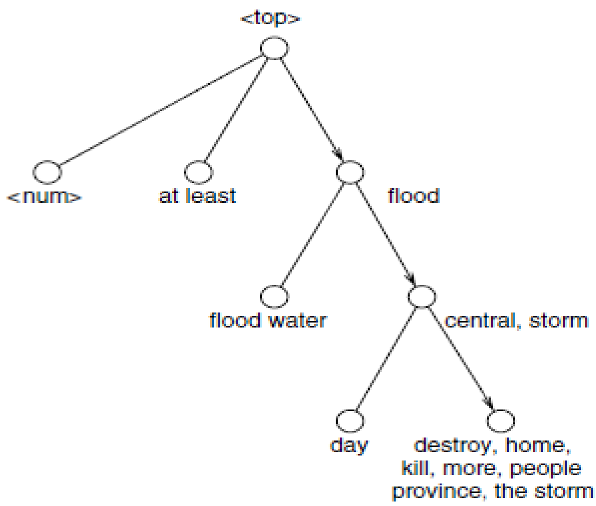


Figure 2.10 Example navigation using a concept lattice

Even if the top ranked documents are not relevant they still cover topics related to the query. According to Mandala et al, (1998), WordNet has two major weaknesses, namely the inadequate proper nouns and secondly the inability to form relations between terms that belong to different parts of speech because words in WordNet are organized on the basis of part-of speech. In other words we cannot form relationships between an adjective term and a noun term even though such a relationship might exist in reality. Also the meronym relationship is too narrow in its interpretation. To overcome this, Grootjen and van der Weider use a hybrid approach consisting of a global thesaurus and a dynamically created local thesaurus.

2.7.4.4 Case Studies Of Query Expansion Using Domain-Specific Ontologies

The problem with domain-independent ontologies such as WordNet is that because they have a broad coverage, ambiguous terms within the ontology can be problematic. For narrower search tasks, domain-specific ontologies are the preferred choice. A domain-specific ontology models terms and concepts which are specifically used in a given domain. Domain-specific ontologies have been constructed in many different application areas such as law, medicine, archaeology, agriculture, geography, multimedia, business, economics, history, and even the news domain to name but a few.

Fu et al, (2005a) present query expansion techniques based on both a domain and a geographical ontology. In their work, a query is expanded by derivation of its geographical footprint. Spatial terms such as place names are modeled in the geographical ontology and non-spatial terms such as 'near' are encoded in a tourism domain ontology. The experiments showed that this method results in improved searches.

In the TREC Genomics Track, Hersh et al, (2003) ran one experiment using phrases based on gene name synonyms and another experiment assessed query expansion using external knowledge resources. The results for the first experiment were better than the results for the query expansion experiment and they conclude that the query expansion results could improve if the query is for a specific task.

Nilsson et al, (2005) use a domain specific ontology based on Stockholm University Information System (SUiS) to carry out query expansion. SUis differs from other question answering systems because it does not allow free-form questions. The question types are restricted to who, what, when and where. Instead of expanding queries with all semantic relationships provided by an ontology such as WordNet, only synonyms and hyponyms are used to increase precision. The experiments have shown an improvement in results.

Huang (2000) describes hierarchical directories as ontologies of the web. Each node on the hierarchy acts as a portal the contents of which are related to a particular category. The user does not have to search the entire web, the hierarchical directory allows the user to search the portal which is relevant to that category. A popular web directory is Yahoo!(see Figure 2.11) The other advantage of such hierarchies is that they can be searched to provide a list of topic paths first and if the user is satisfied with the topic path only then the documents for that topic path are retrieved. This means that a user does not have to search through pages of retrieved results to find relevant documents, any ambiguities can be resolved prior to retrieving a document results set. However, such hierarchies need to be kept uptodate and an automatic updating process is preferred due to the continuous growth of the Web compared to updates being carried out manually. Agrawal et al, (1998) (as cited by Huang 2000) describe TAPER as an example of an authomatic classification system.

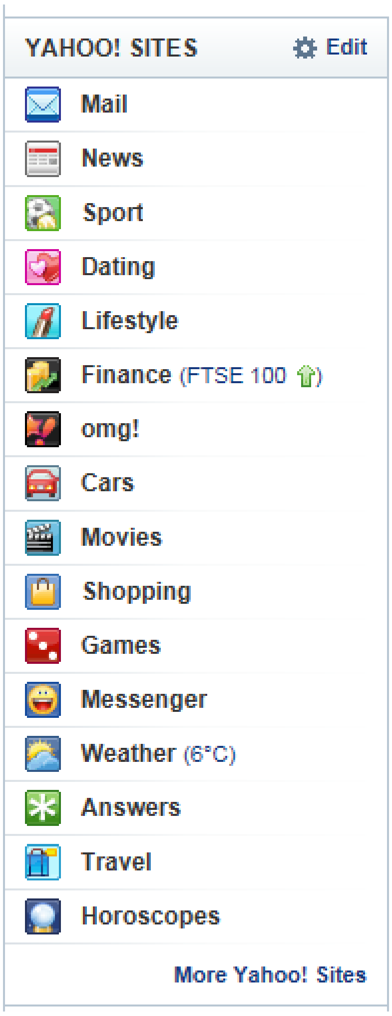


Figure 2.11 Yahoo Directory structure

The use of general/domain-specific ontologies, have advantages and disadvantages. Firstly, ontologies are becoming readily available in an increasing number of subject areas and the predominant use of the semantic web for information retrieval means that the ontologies are usually are encoded in a portable language. Where an ontology is not available for a given subject area or the ontology lacks suitability for a specific application then the process of creating an ontology from scratch is labour intensive and time consuming even though software tools for ontology construction do exist. Also some ontologies might not provide reasoning and inference for applications which would benefit from such features.

2.7.4.5 Success factors in ontology based query expansion

To conclude this section, ontologies have been used for a wide range of information retrieval tasks. Ingwersen (2000) states that tasks can be categorised as either work-tasks or social tasks. Domain specific ontologies are more suitable for work-tasks. The terminology in these ontologies is subject specific and less ambiguous therefore short queries can be expanded with a higher chance of accuracy. General ontologies cover a large area and can be applied for information type broad queries in any domain. The general ontologies tend to include linguistic information and/or semantic information but ambiguous words in a general ontology can have a large number of meanings so the process of identifying the correct sense of ambiguous words is a difficult one and the query expansion process may need some guidance/interaction from the user.

The trend now is to use multiple ontologies to satisfy a user search query (Mena et al, 2000). Magnini and Speranza (2002) state that "Linguistic ontologies encompass ontological and lexical information so partly overcome other limitations such as insufficient broad coverage and the need to be constantly updated". They merge a global ontology with a specialized linguistic ontology such as an economics ontology which includes specific terms and common terms. O'Sullivan et al, (1995) merged WordNet with a domain specific ontology created for word processing software. Bao et al, (2004) integrate a domain-specific ontology and a domain-independent ontology for colonoscopy video database annotation. The domain specific ontology contains information on colonoscopy and the domain-independent ontology contains general information on properties of videos.So hybrid ontologies combine the use of a general ontology and a domain specific ontology to benefit from both approaches. Hybrid approaches can also refer to the use of a bag of words search of the schematic ontology to obtain an initial set of relevant documents which can be refined further by conducting a structured search on an instantiated ontology using proper nouns.

Ontologies seem to be a promising way forward in query expansion. They improve the accuracy in fuzzy information search by paraphrasing the query of the user through context identification and disambiguation (Leger et al, 2001). The success of using an ontology for query expansion depend on various factors. These are described below.

According to (Cheng and Pan 2004), "the quality of the interpretation of free text is strongly dependent on the quality of the model. Coherence, stability, and resistance to inconsistency and ambiguity are desirable ontological model characteristics". This is supported by Jones (1993) who states that the quality of a knowledge model or thesaurus is of paramount importance. The model must be accurate, stable, comprehensive and up-to-date. If a data model does not cover the subject area in a comprehensive way then queries which are relevant to a subject area will not get any results because the model is suffering from some omissions. Voorhees (1993) found the IS-A hierarchy insufficient for selecting the correct sense of the noun when disambiguating short queries. Incorrect sense resolution or query expansion using automatically generated synsets resulted in missing correct matches which in turn have a deteriorating effect on retrieval performance. According to Mandala et al, (1998), WordNet has two major weaknesses, namely the inadequate proper nouns and secondly the inability to form relations between terms that belong to different parts of speech because words in WordNet are organized on the basis of part-of speech. In other words we cannot form relationships between an adjective term and a noun term even though such a relationship might exist in reality. Also the meronym relationship is too narrow in its interpretation.

Suoemela and Kekalaninen (2005) argue that the search process has a higher chance of success if the user is familiar with the knowledge model and is confident at navigating the ontology. Also the initial query formulation starts within the ontology, so its possible for the user to lose their sense of direction or be distracted by the different number of paths during the navigation process and the time taken to traverse those paths. This viewpoint is shared by Sihvonen and Vakkari (2004) who state that query expansion using a thesaurus, is only beneficial if the searcher is familiar with the search topic.

If a user can navigate a knowledge model with ease, this increases its effectiveness. Some ontologies are hundreds of megabytes in size so suitable mechanisms should be used to allow large ontologies to fit onto one screen otherwise users may 'lose' their way in the vast information space and have difficulty in navigating large knowledge models. To overcome the difficulties users have in navigating ontologies, a mixed approach might be better whereby the system automatically searches the ontology for expansion terms which are suggested to the user who will then interact with the system by selecting the relevant terms. Term suggestion is used to enhance recall while user interactions enables precision to be maintained. This is supported by Efthimiadis (1992) who carried out experiments to study the behaviour of ranking algorithms for query expansion and also end-users during the process of query expansion especially how they select terms. The results provided evidence for effectiveness of interactive query expansion.

The above three success factors are related to the properties of the ontology. In addition to these, other factors however also influence the effectiveness of the ontology for conducting query expansion. Firstly query terms need to be mapped onto corresponding ontology concepts. If an exact match is not found then the mapping process must find the 'next best' match. The entry point into the ontology forms the basis of any subsequent expansion so it is crucial to get this process right. Jones et al, (1995) state that one of the main success factors in using a thesaurus for query expansion is to have a good match in the thesaurus to start with. For this reason terms should be expanded before carrying out the search of the document database to overcome the fact that there might not be an exact match of query terms with a WordNet node (Jones 1993).

Secondly query length determines whether there will be any resulting benefit from conducting query expansion. Voorhees (1994) used wordnet to conduct experiments on small single domain TREC collections. The results suggest that query expansion can improve problems of mismatched vocabularies. In contrast, query expansion makes little difference in retrieval effectiveness for long queries because they usually contain a full description of the information request. Voorhees found that some long queries could benefit from other techniques such as relevance feedback. The majority of queries have less than 3 terms (Walker and Beaulieu, 1991; Jansen, 1998). This is insufficient to describe user's need. So it is argued that shorter queries are ideal candidates for query expansion because they tend to be more ambiguous (Navigli and Velardi 2003). However the query expansion method used must expand the query accurately otherwise a degradation in performance will be evident (Voorhees 1993). Accurate query expansion is when a higher percentage of the suggested terms are relevant to the query and thus produce more relevant documents in the output.

Thirdly, as argued by Broder (2002), general, broader information queries benefit more from query expansion than navigational or transactional queries. The reason for this might be that the information query is less specific than the other query types. Navigational queries typically are used to find homepages. Information queries are used to find information relevant to a given topic. Transactional queries are just that, they allow a user to locate a website offering a service such as shopping and enable the user to carry out a purchase transaction. Ontologies have been used for a wide range of information retrieval tasks. Tasks can be categorised as either work-tasks or social tasks. Domain specific ontologies are more suitable for work-tasks. The terminology in these ontologies is less ambiguous therefore short queries can be expanded with a higher chance of accuracy. General ontologies would be suitable for information type broad queries however the query expansion process may need some guidance or interaction from the user. Hersh et al, (2003) conclude that the query expansion results could improve if the query is for a specific task. The problem with domain-independent ontologies such as WordNet is that because they have a broad coverage, ambiguous terms within the ontology can be problematic. For narrower search tasks, domain-specific ontologies are the preferred choice. A domain-specific ontology models terms and concepts which are specifically used in a given domain. Domain-specific ontologies have been constructed in many different application areas such as law, medicine, archaeology, agriculture, geography, multimedia, business, economics, history, and even the news domain to name but a few. The trend now is to combine the use of domain specific ontologies with general ontologies to satisfy a user search query (Mena et al, 2000; Magnini and Speranza, 2002; Mandala et al, 1999; Fu et al, 2005a; Bao et al, 2004). It is suggested that using combined query expansion techniques with an ontology produces better results than using a single technique (Schatz et al, 1996).

A fourth area of discussion is the weighting mechanism used for ontology terms. Jones et al, suggests that term weighting is a good idea. A reference is made to Shoval who bases the weightings on criteria such as number of connections; type of relationship; number of co-occurrences and path length (Shoval 1985). A further reference is made to Chen who suggests that users should be able to adjust the term weightings (Chen et al, 1993). Contrary to Efthimiadis (1996) who suggests that query expansion using knowledge models may be done without term re-weighting.

Fifthly, the number of ontology terms selected for query expansion also needs to be considered. Jones et al, (1995) analysed the INSPEC thesaurus and used 8 relational database tables to store thesaurus information such as terms, equivalence, hierarchical, associative, components, classes, facets and words. Their experiments found no correspondence between the number of terms chosen and the query performance. Also the other main finding was that the number of terms selected is dependent on the distance from the original node. In other words there was a tendency for nodes with a short distance from the original node to be selected.

Finally, knowledge resources are assumed to be context-free however Ruthven (2004) points out that whilst the knowledge resources are useful in deriving context, the knowledge resources themselves have a certain context attached based on how they were created and why they were created. To use these knowledge resources effectively, it is useful to understand the "built-in" context otherwise we will not be using the resources to their full potential. For example a mathematics knowledge resource that has been created by a school teacher will probably include information relevant to teaching mathematics at school level and is unlikely to include information that would be required by post-graduate students.

In the previous sections we have examined the different approaches to inferring context. In the next section we will discuss how some of these approaches have been applied in the news domain.

2.8 Review of News Applications

News is the communication of information on current events which is presented by print, broadcast or internet to a mass audience. News systems can support a range of different tasks namely news annotation, news categorization/classification, news summarization, browsing/searching and personalized news alerts. In the late nineties broadcast news systems were of interest to information retrieval researchers. spoken document retrieval systems were researched into. Broadcast news systems includes spoken or visual news. Recent work has focussed on providing personalized news services for mobile devices based on user profiles. Ontologies have been used for news annotation and news classification. Concept nodes within ontologies have been used to represent user profiles. Then the move was to group profiles of users with related interests and conduct collaborative filtering. Finally news ontologies can be used in conjunction with other domain ontologies to satisfy searches for other domain applications. Each of these areas will be discussed in more detail in the following paragraphs.

Johnson et al, (1999) describe a spoken document retrieval (SDR) system and assess its performance using transcriptions of about 50 hours of broadcast news data. They use error rates as a more relevant measure of transcription accuracy for SDR purposes. The implication is that documents with least errors are more relevant than documents with a higher rate of transctiption errors. Billsus and Pazzani (1999) use synthesized speech to read news stories to a user. The user gives voice feedback and the system automatically adapts to the user's preferences. Timecoded feedback (assumes a story is relevant if a user spends more time listening to it) is also used to automatically induce accurate profile interests and help classify news stories. When users track different threads of ongoing recent events, these are known as short term interests. Short term interests are represented using Nearest neighbour. Long term interests reflect general user news interests and will retrieve news stories that may not be related to a recent event but still meet the user's interests. Long term interests are represented using a Bayesian classifier. Instead of giving feedback on relevance of retrieved items, users give feedback on the concepts used for the classification of a news item. Abberley et al, (1998) develop a realtime broadcast news vocabulary speech recognizer and integrate it into a spoken document retrieval system. Two advances were made for this task: automatic segmentation and statistical query expansion using a secondary corpus. Ariki and Sugiyama (1997) describe a system which automatically classifies TV news articles using a keyword spotting technique and can answer queries from users interactively. The keyword spotting technique extracts a keyword sequence from the newsreader's off-line script with their probabilities and the article is annotated with these keywords for classification and retrieval purposes. Liu and Sha (2006) incorporate automatic speech recognition based on video transcripts and visual information in the search mechanisms.

Over the past 5 years, much work has focussed on providing personalized news services for mobile devices. EPaper is a personalized newspaper service for mobile devices (Tenenboim et al, 2008). It uses language modelling techniques to classify incoming news articles against concepts in the ontology. It combines the use of the ontology and the user profile to deliver personalised news stories to mobile devices. Huhn et al, (2005) developed a news alert system called P-Alert and found that ontologies improve the quality of the result set if the expanded query terms have hits in the ontology. The adaptation process in Adaptive news delivery systems has become more complicated and less transparent to users (Ahn et al, 2007). YourNews is a system which uses open user models whereby users can view and edit their interest profiles (see Figure 2.12).Users prefer transparency but open user models may harm the system's performance and has to be used with caution because the user actions of adding/removing terms from the query might make an already good profile worse.

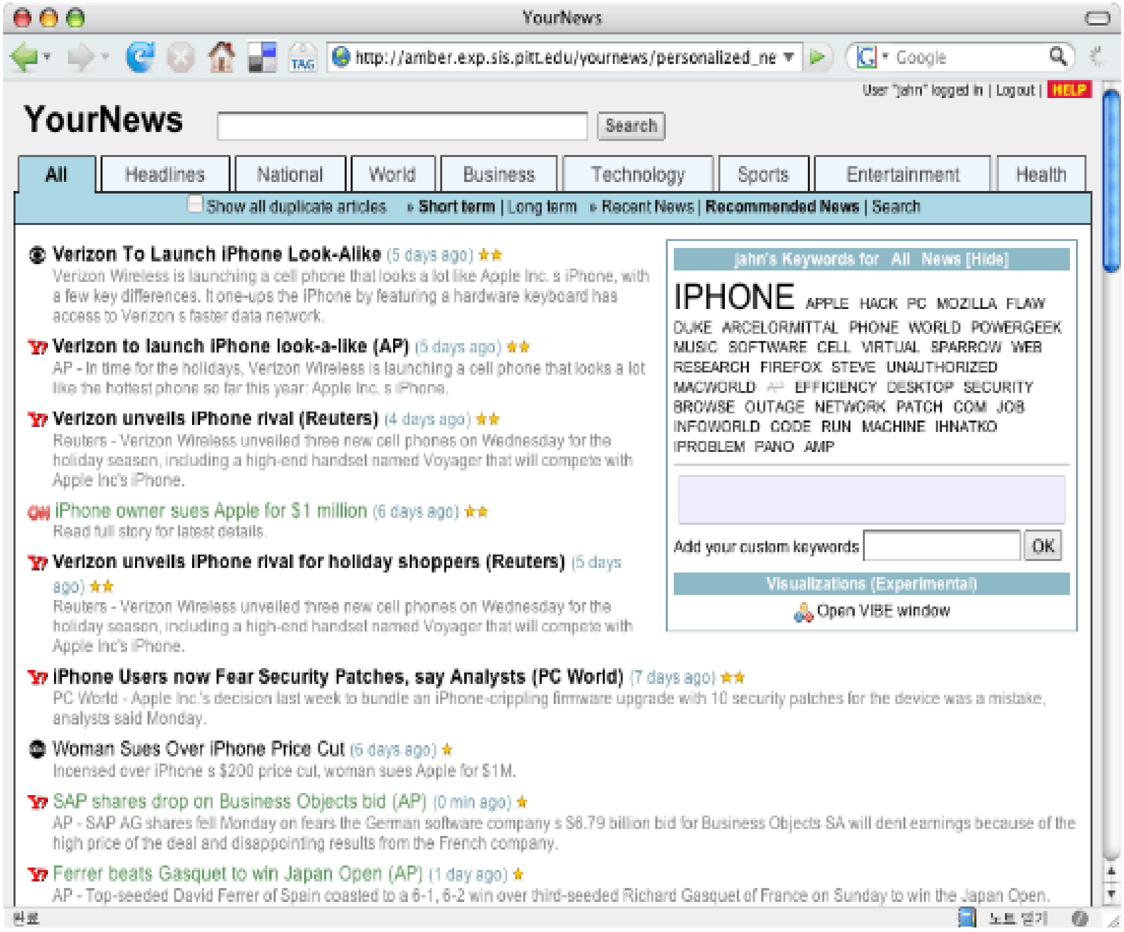


Figure 2.12 Screenshot of YourNews news filtering system.

Lang, (1995) shows that a learning algorithm based on the Minimum Description Length was able to raise the percentage of interesting articles to be shown to users from 14% to 52% on average. The learning algorithm outperformed tf-idf technique by 21%.

Personalized content retrieval aims at improving the retrieval process by taking into account the particular interests of individual users. Vallet et al, (2007) state that not all user preferences are relevant in all situations and should be understood in context with the user goals and tasks at hand. The user profile is built using explicit feedback or implicitly by the user actions. They build a dynamic representation of the semantic context of ongoing retrieval tasks which is used to activate different subsets of user interests at runtime. The runtime context is represented using a set of weighted concepts from a domain ontology, user preferences are also considered to be concepts in the same domain. A similarity measure based on the number and length of path distances is calculated between the runtime context and the user preferences. A high similarity measure indicates that the user preferences are relevant to the runtime context. The ontology-based representation of user interests is richer, more precise, less ambiguous than a keyword-based or item-based model. It provides an adequate grounding for the representation of coarse to fine-grained user interests. An ontology provides further formal, computer-processable meaning on the concepts and makes it available for the personalization system to take advantage of.

Recent research has looked at the use of ontologies for obtaining and recording user profile information. Vallet et al, (2007) use an ontology-based scheme for the semiautomatic annotation of documents, and use an annotation weighting algorithm and a ranking algorithm to produce a modified version of the vector-space retrieval model. The search system takes advantage of both detailed instance-level knowledge available in the KB, and topic taxonomies for classification. Zhang et al, (2009) use compact concept ontology to determine whether to do query expansion for a given query. Users have the option of exploring different semantic levels by using different compact concept ontologies. They integrate the methods into a textbased video search system. PlanetOnto is a knowledge-based news server supporting ontology driven story enrichment and knowledge retrieval. Journalists send news stories to an automated agent. The agent annotates the stories and informs users that new stories have been added. News@hand (Cantador et al, 2008) applies semantic web technologies to describe and relate news contents and user preferences in order to produce enhanced recommendations (see figure 2.13).

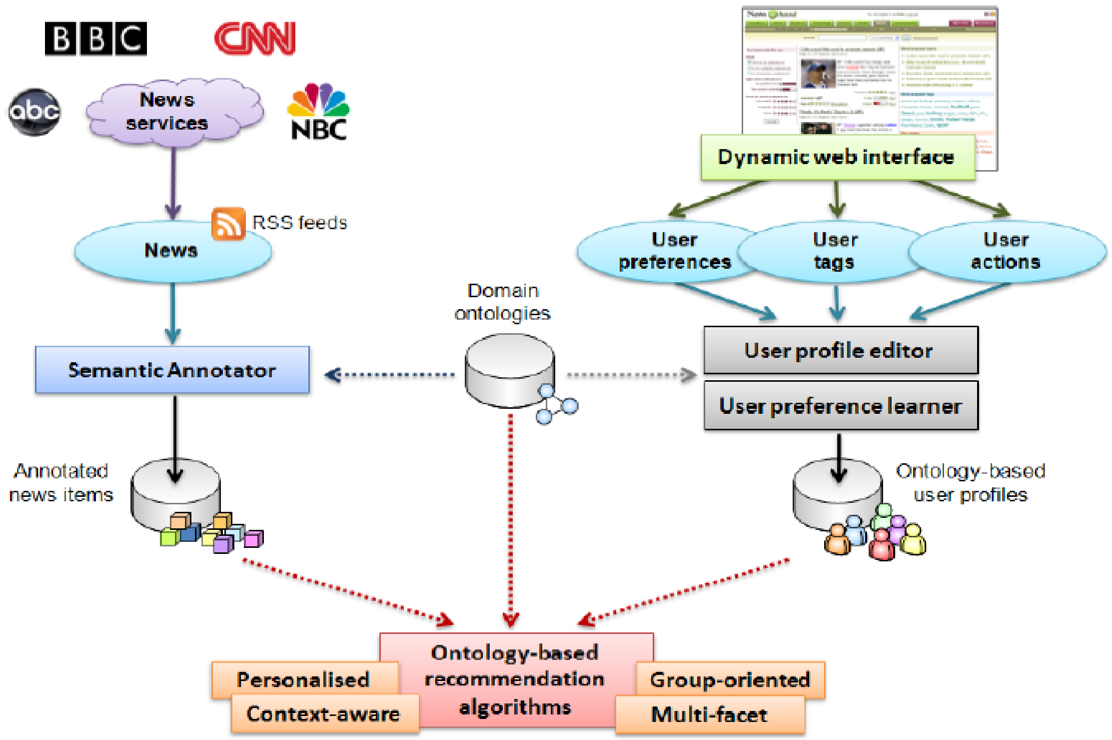


Figure 2.13 News@hand architecture

They exploit meta-information in the form of ontologies that describe items and user profiles in general, portable way along with the capability of inferring knowledge from the semantic relations defined in the ontologies. News items and user profiles are represented in terms of concepts appearing in 17 domain ontologies (adaptions of the IPTC ontology) and semantic relations between these concepts are exploited to enrich the above representation of news items and user profiles and enhance recommendations. The vector based preference description facilitates combination of multiple profiles to generate shared profile for group of users. The News@hand system uses natural language programming to annotate news titles and summaries. A major problem is that if the ontology hierarchy changes annotation must be done again.

Collaborative filtering (CF) applications adapt to user communities who interact with the system, where the users have common interests. Claypool et al, (1999) apply a combined collaborative and content based filtering approach to an online newspaper. Das et al, (2007) state recommender systems can be categorized into 2 types: Content based and collaborative filtering. The content based filtering approaches calculate similarity of items defined in terms of their content to other items that have been highly rated. Cantador et al, (2008) describe a clustering strategy that automatically identifies Communities of Interest (CoI) from the tastes and preferences expressed by users in personal ontology-based profiles, and evaluate how these CoI can be applied to recommend annotated items combining several content-based collaborative recommendation techniques. Clustering enables common topics to be obtained and with these topics preferences are partitioned into different layers.

Papyrus is based on a history ontology and a news ontology (Kiyavitskaya et al, 2010). Ideally, users should be able to start from the history domain, where they could identify relevant concepts and related information (see Figure 2.14).

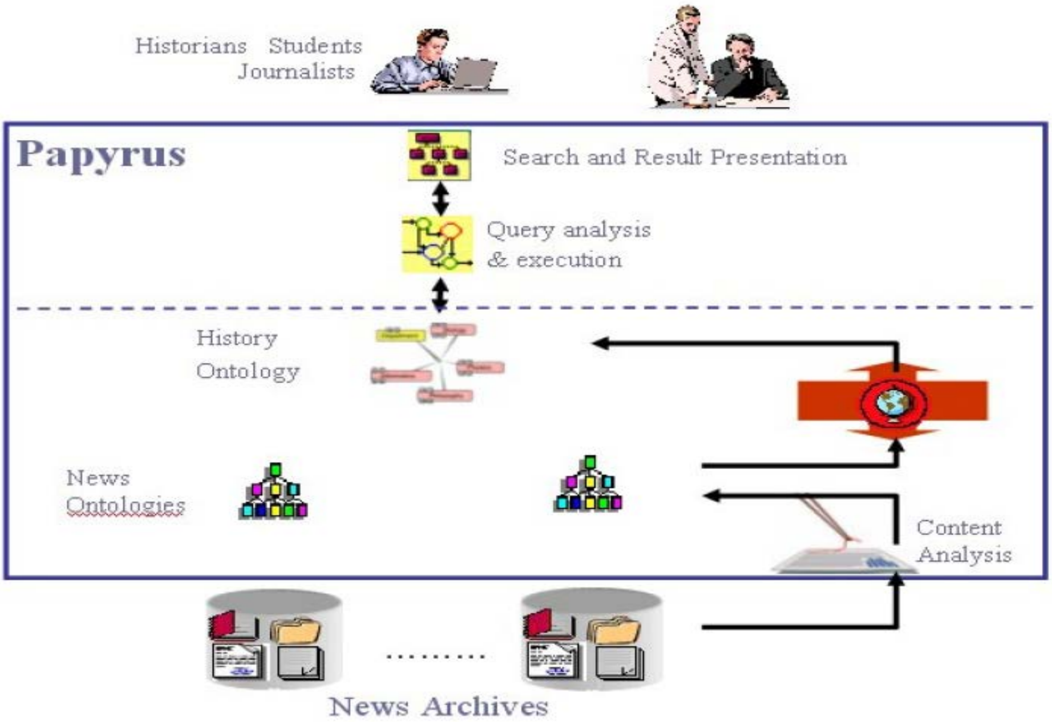


Figure 2.14 Overview of Papyrus

Through them they should be able to move to the news domain where they could review related vocabulary and concepts, and finally, reach the relevant multimedia content, the news items. Although both news and history are about events, a news item conveys journalistic views (when it happened, how it happened; who was involved) and history records why it happened, what the consequences were and how it could have been avoided. Papyrus models both domains and maps the correspondences.

We are not including broadcast news, our research uses news in text/print format. We are only looking at the searching feature of news retrieval systems and do not include any of the other features. At present the use of personalized news stories for mobile devices falls outside the scope of our research. PlanetOnto is closely dependent on its own information retrieval system so we would not be able to use it in our research project because we are using Okapi. Although ontologies have been used in news retrieval systems, they have mostly been used for news classification, news annotation and building user profiles. We would like to explore the use of ontologies to for query expansion.

2.9 Summary

The problem of context with respect to information retrieval has been described. The various approaches to handling context have been analysed. Early research work adopted a single approach for handling context. Recent context aware applications have shown an increase in combining several approaches eg Ubiquitous computing and Personalisation. Ubiquitous computing and Ontologies. The chapter analyses the use of relevance feedback and term co-occurrence as ways of handling context within query expansion. A section on the use of ontologies for a range of information retrieval tasks is presented and in particular their use in the area of query expansion. Researchers have reported mixed success in ontology based query expansion. This section has tried to identify the properties of an ontology which assist in successful query expansion. In addition to these inherent properties, there are some external factors which also influence the effective use of an ontology during query expansion.

This research seeks to investigate possible ways of using context information in the ontology for query expansion using a probabilistic retrieval model. So we will be combining the use of relevance feedback techniques for query expansion with expanding terms using the ontology. The challenge is to keep relevant concepts and discard irrelevant concepts so that query expansion results in improved retrieval effectiveness. To overcome the difficulties users have in navigating ontologies, this research will enable the system to automatically search the ontology and identify expansion terms. Our view of context is similar to collaborative filtering in the sense both use a community of users. An ontology is a conceptual view which is shared by a set of users whereas collaborative filtering is a view of searches shared by a user community. As suggested by Jones (1993), the queries will be expanded first to ensure a good match in the ontology. Experiments will be carried out on ontology term selection using different depths and varying the relevance feedback parameters. A report of the findings will be given.

3. METHODOLOGY

3.1 Introduction

In order to give the background for the news domain, section 3.2 describes the structure of a news story and the professional body involved in setting the eXtensible Mark-up Language (XML) standard for news. The first objective of this research "Select a document collection with an appropriate ontology and index the document collection in the standard manner" is covered by section 3.3 COLLECTIONS. Section 3.3 compares and contrasts two news ontologies and gives justification for the choice of ontology. We also need to select a document collection and justify our choice. The chosen document collection will be indexed in the standard manner. Further details on indexing the document collection can be found in chapter 4. Section 3.4 EXPERIMENT METHOD covers objective 3 "Design and conduct laboratory experiments in order to compare and contrast the performance of the standard retrieval model with the revised retrieval model". In section 3.4 we will describe the range of experiments that need to be carried out to test the research hypothesis and an explanation will be given of the metrics that have been chosen to evaluate the results.

3.2 News Domain

Domain –specific ontologies such as those in the medical domain use subject related specialist terminology. A news ontology overlaps the two definitions of domain-specific and domain-independent. It is domain-independent in the sense it doesn't have subject specific terminology, however it is domain-specific in terms of its structure; every news item has a set structure which includes date, heading, author and story content.

Many of the news retrieval systems make use of the ontology defined by the International Press Telecommunications Council (IPTC) which is an organisation which was established in 1965. It is a consortium of 70 world's major news agencies and news industry vendors. It develops and maintain standards for improved news exchange that are used by virtually every major news organisation in the world. It is involved in setting XML based business-to-business standards for sharing news, and development of advanced metadata to describe and classify news text, images and other media. News Industry text format (NITF) is an XML specification published by the IPTC that is designed to standardize the content and structure of text-based news articles. NewsML is an XML standard developed by IPTC for multimedia news. IPTC has creates and maintains NewsCodes which are controlled vocabularies - or taxonomies - to be used to categorize news content. The Newscodes are grouped to cover aspects such as administration, description and transmission (IPTC 2011). The advantage of using numeric codes is that they are unique in value and there is no risk of the codes being ambiguous or duplicated.

Concept Id: subj:02000000	created: 2000-10-30T12:00:00+00:00	
Type (Qcode) = cpnat:abstract	Name in en-GB	Abstract concept
Name in en-GB is:	crime, law and justice	
Definition in en-GB is:	Establishment and/or statement of the rules of behaviour in society, the enforcement of these rules, breaches of the rules and the punishment of offenders. Organizations and bodies involved in these activities.	

Concept Id: subj:02001000	created: 2000-10-30T12:00:00+00:00	
Type (Qcode) = cpnat:abstract	Name in en-GB	Abstract concept
Name in en-GB is:	crime	
Definition in en-GB is:	Violation of established laws by individuals, companies or organizations	

Broader concept:	http://cv.iptc.org/newscodes/subjectcode/02000000	of type cpnat:abstract

Figure 3.1 NewsCodes for Crime,Law and justice concept and Crime concept

In figure 3.1 we can see that crime, law and justice is a broader concept for the crime concept. The IPTC ontology which contains concepts of multiple domains such as education, culture, politics, religion, science, technology, business, health, entertainments, sports etc.

3.3 Collections

3.3.1 News Ontologies and Test Collections

Since XML is used to describe document structure, it is ideal for representing the structure of news articles. We will briefly describe XML technologies before describing the news ontologies and test collections, The News domain is of interest because of the need to manage huge amount of dynamic content. Since the test collection is based on news articles, it is sensible to examine some news ontologies. This section describes a number of ontologies in the news domain with a focus on details such as whether the ontology was indexed manually or using some automation technique and the language used to implement the ontology.

The inherent structure of an ontology can be represented using XML which is a widely used standard for encoding structured documents. XML documents have a ordered, labelled tree structure which can be simple or complex where attributes/nodes are nested. XML itself is a metalanguage to design markup languages, i.e. text language where semantic and structure are added to the content using extra "markup" information enclosed between angle brackets. HTML is the most well-known markup language. Each node of the tree is an XML element and is written with an opening and closing tag. An element can have one or more XML attributes. In Figure 3.2 the <SUBJECT> element has child elements, TITLE and SUBJECTMATTER. Figure 3.2 shows figure 3.1 as a tree. The leaf nodes <TITLE> consist of text. The internal nodes <SUBJECTMATTER> and <SUBJECTDETAIL> encode the structure of the document or metadata. With a Document Object Model (DOM) API we can process an XML document by starting at the root element and then descending down the tree from parents to children. XPath is a standard for searching paths in an XML document collection. The XPath expression node selects all nodes of that name. Successive elements of a path are separated by slashes, so act/scene selects all scene elements whose parent is an act element. Double slashes indicate an arbitrary number of elements can intervene on a path. An XML schema is required to put constraints on the structure of allowable XML documents. Two standards for XML documents are XML DTD (document type definition) and XML Schema.

3.3.1.1 World News Ontology (WNO)

This system performs semantic search on the World News domain (Kallipolitis et al, 2007) It is based on metadata files created for every single world news HTML web page.The World News Ontology was developed using logic programming as the basic way of data representation and it was implemented using XML. World News Finder is a Java based information retrieval system which uses XML implementations for the metadata representation and the ontology.

The metadata representation is based on NewsML (newsml.org). NewsML is an XML-based standard developed by International Press Telecommunications Council (IPTC) to represent and manage news throughout its lifecycle including production, interchange and consumer use. NewsML provides a set terms for the news domain. This set of terms also known as Newscodes includes a hierarchy of terms and concepts that can be used to describe news in any field of interest. This hierarchical structure or taxonomy shown in Figure 3.2 consists of three levels:

Subject: topics at this level provide a description of the editorial content of news at a high level
Subjectmatter: a Subjectmatter provides a more precise description
Subjectdetail: provides the most specific description compared to the higher levels.

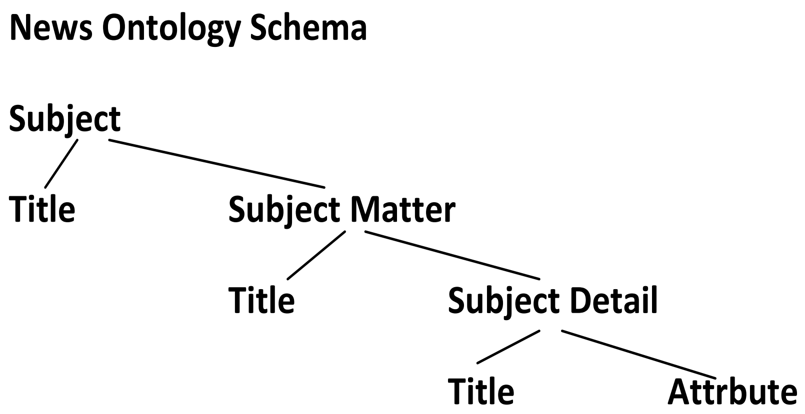


Figure 3.2 Hierarchical structure of Newscodes ontology

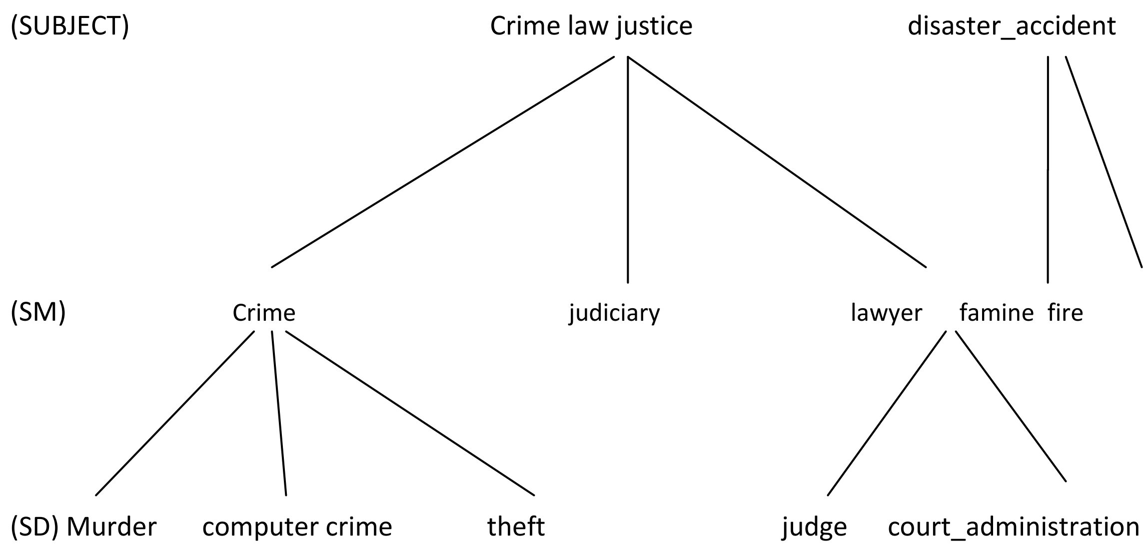
Figure 3.3 shows a sample extract of the xml definitions for the three levels of the ontology.

```
<ONTOLOGY>

 <SUBJECT>
  <TITLE>crime_law_justice</TITLE>
    <SUBJECTMATTER>
      <TITLE>crime</TITLE>
        <SUBJECTDETAIL><TITLE>murder</TITLE></SUBJECTDETAIL>
        <SUBJECTDETAIL><TITLE>computer_crime</TITLE></SUBJECTDETAIL>
        <SUBJECTDETAIL><TITLE>theft</TITLE></SUBJECTDETAIL>
        .
.
.
    </SUBJECTMATTER>
      <SUBJECTMATTER>
      <TITLE>judiciary</TITLE>
    </SUBJECTMATTER>
      <SUBJECTMATTER>
      <TITLE>lawyer</TITLE>
      <ATTRIBUTE>name</ATTRIBUTE>
        <SUBJECTDETAIL>
          <TITLE>judge</TITLE>
          <ATTRIBUTE>name</ATTRIBUTE>
        </SUBJECTDETAIL>

<SUBJECTDETAIL><TITLE>court_administration</TITLE></SUBJECTDETAIL>
    </SUBJECTMATTER>
</SUBJECT>
<SUBJECT>
  <TITLE>disaster_accident</TITLE>
    <SUBJECTMATTER>
      <TITLE>famine</TITLE>
    </SUBJECTMATTER>
    <SUBJECTMATTER>
      <TITLE>fire</TITLE>
    </SUBJECTMATTER>
  </SUBJECT>
.
</ONTOLOGY>
```

Figure 3.3 Sample extract taken from the WNO



KEY: SM = Subject Matter
SD = Subject Detail

Figure 3.4 : Diagrammatic form of XML tree format shown in Figure 3.3

Note: At present the <ATTRIBUTE> child element is not used, we represent documents as trees using just the <TITLE> child element nodes.

The authors of this ontology studied a large number of international news articles from news agency websites and as a result based the ontology on 11 subjects which they felt were sufficiently representative in the domain of world news. Logic Programming (LP) is used to express the ontology and the metadata, which is then transformed into XML format. With LP, new rules can be added allowing processing and reasoning by a LP language.

Metadata for a page consist of a set of topics which can have attribute values. Each topic is assigned a weight to indicate the page's event relevance to this topic. All this information for each page is put in a metadata wrapper called hat. Each topic is associated with ontology subject(s). The subject of each page is deduced from the topics that appear in its hat and their corresponding weights. Each topic contributes to the subject it belongs to by a degree based on its weight. The subject that ends up with the highest degree is considered to be the subject of that page. This might be more than one subject. The hats were created and imported in the database manually. Users can create new hats and record them in the system. The intention is to create hats automatically by parsing large number of articles from the websites of various news agencies daily.

The ontology is simplistic because it only shows an IS-A hierarchy between nodes, but the relations between concepts do not have any other semantics. The ontology could be improved if it contained more semantic information such as homonyms, hypernyms, holonyms and meronyms. Homonym relations are equivalent to the child nodes in a parent-child hierarchy. Hypernym relations are equivalent to the parent nodes in a parent-child hierarchy. Whole-part relations are represented using holonyms (whole) and meronyms (part). The query topics we have used for our experiments are varied in the breadth /depth of information required for the results. For example a topic which requires a wide set of results would in addition to exploring the parent link of the search term from a hypernym relation, also base the expansion on sibling nodes, synonym concepts and holonym relations. For queries requiring more specific results then in addition to exploring the child nodes of a search term, also search nodes from the hyponym and meronym semantic relations.

The entire WNO ontology was used for our experiments because it is fairly small in size and it felt unnecessary to tailor/prune it. However, only results for level1 of the ontology are discussed because the level2 results did not offer any significant improvements. For larger ontologies there may be a case of pruning/tailoring the ontology based on some criteria such as the initial term and the domain frequency of any nodes connected to the initial term node. Nodes with a higher domain frequency would be favoured and the remaining nodes would not be considered. The set of queries would need to be examined and sections of the ontology which are considered to be irrelevant to the query set could be removed/excluded from the ontology. The benefit of doing this would be to save storage space, increase search efficiency and reduce information overload for the user.

3.3.1.2 News Engine Web Services Ontology (NEWS)

NEWS is a lightweight Resource Description framework schema (RDFS) ontology for Spanish EFE News Agency (Sanchez-Fernandez et al, 2005). RDFS is a general-purpose language for representing information in the Web (W3C 2004). The main commercial databases used are EFEdata and Fototeca. The first stores more than 9 million text news documents from 1988 and the second contains and pictures from 1998. News items are categorized manually using a categorization system internal to the news agency. If it was done automatically then allow usage of more complex categorization systems like the one provided by the IPTC: the Subject Code NewsCodes (IPTC 2011). News items are represented in XML based formats like News Industry Text Format (NITF) and NewsML which is a solution for exchanging multimedia news (IPTC 2011).

Content annotation is done for management purposes (author, creation time, date, keywords and creation location). The contents of news items are not annotated, so the basic entities (organizations, persons, places, etc) which are mentioned inside the news item are not tagged. Such content annotations are useful in fine grained news item selection, especially in advanced push services where users are more likely to be interested in specific news items such as news about a certain person or organization. The sample extract from the NEWS ontology in Figure 3.5 corresponds to the hierarchical structure shown in figure 3.6.

```
<rdfs:Class rdf:about="&Content;Park"
       rdfs:label="Content:Park">
      <rdfs:comment>Documentation on Park(MILO): A LandArea which is
intended to be used for recreation and/or exercise.Note that a Park
can be either publicly or privately owned.</rdfs:comment>
      <rdfs:subClassOf rdf:resource="&Content;LandArea"/>
</rdfs:Class>
<rdfs:Class rdf:about="&Content;LandArea"
       rdfs:label="Content:LandArea">
      <rdfs:comment>Documentation on LandArea(SUMO):  An area which is
predominantly solid ground, a mountain, a desert, etc. Note that a
LandArea may contain some relatively small WaterAreas. For example,
Australia is a LandArea even though it contains various rivers and
lakes.</rdfs:comment>
      <rdfs:subClassOf rdf:resource="&Content;GeographicArea"/>
</rdfs:Class>
<rdfs:Class rdf:about="&Content;GeographicArea"
       rdfs:label="Content:GeographicArea">
      <rdfs:comment>Documentation on GeographicArea(SUMO):  A
geographic location, generally having definite boundaries. Note that
this differs from its immediate superclass Region in that a
GeographicArea is a three-dimensional Region of the Earth.
Accordingly, all astronomical objects other than Earth and all one-
dimensional and two-dimensional Regions are not classed under
GeographicArea.</rdfs:comment>
      <rdfs:subClassOf rdf:resource="&Content;Region"/>
</rdfs:Class>
<rdfs:Class rdf:about="&Content;Region"
       rdfs:label="Content:Region">
      <rdfs:comment>Documentation on Region(SUMO):  A topographic
location. Regions encompass surfaces of Objects, imaginary places, and
GeographicAreas. Note that a Region is the only kind of Object which
can be located at itself. Note too that some regions, e.g.
archipelagos, have parts which are not connected with one another.
Regions and locations are very interesting entities for the journalism
world as reflects the fact that IPTC NewsCodes include several Topic
Types related with this issue eg: Location or City.</rdfs:comment>
      <rdfs:subClassOf rdf:resource="&Content;Object"/>
</rdfs:Class>
```

Figure 3.5 Sample extract from NEWS ontology

Region

GeographicArea

LandArea

Park

Figure 3.6 Hierarchical Structure for Sample extract shown in Figure 3.5

3.3.2 Justification for chosen Ontology

The choice was made between WNO and NEWS ontologies. The PlanetOnto and EPaper ontologies have a different scope and purpose so were not included in the selection process. The ontology has been implemented in XML form because of the wide availability of XML-processing software and in order to achieve portability. The NEWS ontology has been implemented in rdf which means it will contain richer information but at the same time is also more difficult to process. Also the WNO adheres to a tree structure but the NEWS ontology adopts a lattice or network hierarchy which allows child nodes to be linked to more than one parent node. We can see that the ontology hierarchy in WNO will always have a maximum depth of 2 levels whereas the NEWS ontology can have many levels for a given class. In the example given, Region has a depth of 3. Previous research has shown that terms from level1 produce the most beneficial results so having an increased levels in the ontology would mean extra processing for no real gain. Scalability is important when selecting representation language used for the ontologies. Both ontologies are easily scalable, so scalability was not a distinguishing factor. WNO is a much smaller ontology at 29.Kb whereas NEWS is 47.2Kb in size. The WNO ontology is of a manageable size which makes it easier to navigate and process programmatically. Class information in WNO is given in a top-down fashion whereas the NEWS ontology gives information in a bottom-up fashion in the sense that for each class it states which class it is a subclass of. This would mean that we would have to process the ontology in a bottom up fashion. Taking everything into consideration the WNO ontology was selected for this research.

3.3.3 Justification for chosen Test collection

In the past, TREC conferences have centred around two main tasks based on traditional information retrieval modes: a routing task and an adhoc task. In the routing task it is assumed that the same questions are always being asked but that new data is being searched. In the adhoc task, it is assumed that new questions are being asked against a static set of data. For this research we will be using the adhoc task. The known documents were on TIPSTER disk 2 and topics 51-300 used. Disk 2 contains over 231,000 documents from WSJ (Wall Street Journal -1990, 1991, 1992), AP (AP Newswire -1988), Ziff (Articles from Computer Select disks) and FR (Federal Register). Table 3.1 (Voorhees and Harman 1997) shows some basic document collection statistics.

	Size (megabytes)	# Docs	Mean # of words
Wall Street Journal 1990 – 1992 (WSJ)	242	74,520	508.4
Associated Press newswire 1988 (AP)	237	79,919	468.7
Computer Selects article, Ziff-Davis (ZIFF)	175	56,920	451.9
Federal Register, 1988 (FR)	209	19,860	1378.1

Table 3.1 Basic Document Collection statistics

The collection sizes are roughly equivalent and document length ranges from very short to very long. As can be seen from Table 3.1, the document collection contains news articles (WSJ and AP) and non-news based articles (ZIFF and FR government documents). Disk2 is a smaller collection size in comparison to other document collections but the advantage of Disk 2 is that it contains a wider range of topics. This ontology and document collection have not been used in conjunction with each other before so this work will set a baseline for future research experiments by other authors. News based articles were not separated out and the entire collection was used because the aim was to use the as many relevance assessments in the document collection as possible. The non-news articles in the collection introduced “noise” to discover whether the news ontology ranked news articles higher than non-news articles. Upon examination of the relevance assessments and trec ids of the documents in the ranked results, there is no strong evidence to suggest that the news ontology favours news articles (WSJ & AP) over non-news articles (ZIFF & FR). The possible reason for this is that we are not putting any emphasis on the structure of the news articles. Only key terms are being used for the search thus we are treating all articles news or non-news in the same manner. If any structural feature of news articles is incorporated in the search process then it is likely that news articles would appear higher up in the ranked set of results. The documents are formatted using SGML tags as shown in Figure 3.7 below.

```
<DOC>
<DOCNO> AP880212-0001 </DOCNO>
<FILEID>AP-NR-02-12-88 2344EST</FILEID>
<FIRST>u i AM-Vietnam-Amnesty     02-12 0398</FIRST>
<SECOND>AM-Vietnam-Amnesty,0411</SECOND>
<HEAD>Reports Former Saigon Officials Released from Re-education Camp</HEAD>
<DATELINE>BANGKOK, Thailand (AP) </DATELINE>
<TEXT>
   More than 150 former officers of the
overthrown South Vietnamese government have been released from a
re-education camp after 13 years of detention, the official Vietnam
News Agency reported Saturday.
.
.
</TEXT>
</DOC>
```

Figure 3.7 Sample contents of the .bib file

The topic files contain tag definitions for each topic. Figure 3.8 shows the definition for topic 251.

```
<top>

<num> Number: 251
<title> Exportation of Industry

<desc> Description:
Documents will report the exportation of some part
of U.S. Industry to another country.

<narr> Narrative:
Relevant documents will identify the type of industry
being exported, the country to which it is exported; and as well
will reveal the number of jobs lost as a result of that exportation.

</top>
```

Figure 3.8 Sample extract from the topic file

3.4 Experiment Method

3.4.1 Design

The problem with traditional relevance feedback techniques and corpus dependent query expansion is that they are content driven. In other words, the corpus content is analysed to extract candidate terms for query expansion. This can only work if there are sufficient relevant documents to work with and also that these documents contain a reasonable set of terms that represent the subject area for the query. Corpus independent knowledge models do not suffer from this drawback. Therefore, we will be using an ontology to assist in the query expansion process.

The central hypothesis of this thesis "The use of ontology based query expansion in a probabilistic retrieval model will improve retrieval effectiveness" will be tested by carrying out the steps given in this methodology. A general newswire document collection will be used (Trec disk2 topic 51-100, topic 101-150, topic 151-200, topic 201-250, topic 251-300) together with a News domain specific ontology. The topics are based on domains such as International Economics, Science, Technology, International relations , Law, Government and Politics. The document collection will be indexed in Okapi which uses the probabilistic retrieval model (Sparck-Jones et al, 2000). Additionally, the News domain specific ontology will be searched and hierarchical node relationship information for that term will be included into the parent-child database. The original Okapi system just uses relevance feedback (rfb) and pseudo-relevance feedback (pf) for query expansion. The new system will employ rfb and pf techniques but in addition it will expand the query further by making use of the parent-child information obtained from the ontology. So for each experiment there is a pf run and a rfb run. The parent node(s) of a query term will broaden the query and the child node(s) of a query term will make a query more specific. The search strategy used is depicted in Figure 3.9.

The search strategy used is simplistic in nature in the sense that we do not analyse the query to determine whether a broad / narrow search is required. In future work it is recommended that some type of query analysis is conducted prior to conducting the ontology search. For example Topic 185 "Reform of the U.S. Welfare System" would probably search for relevant documents at three levels of government – local, state and federal. Depending on the entry point into the ontology, this type of search would require broad and narrow searches. Topic 68 "Health Hazards from Fine-Diameter Fibers" is an example of a query which would benefit from a narrow search because the scope of the query is not from health hazards in general but those health hazards that result from fine-diameter fibers. This type of search strategy is pointed out in chapters 9 and 10 in which we discuss our recommended guidelines on using an ontology for query expansion.

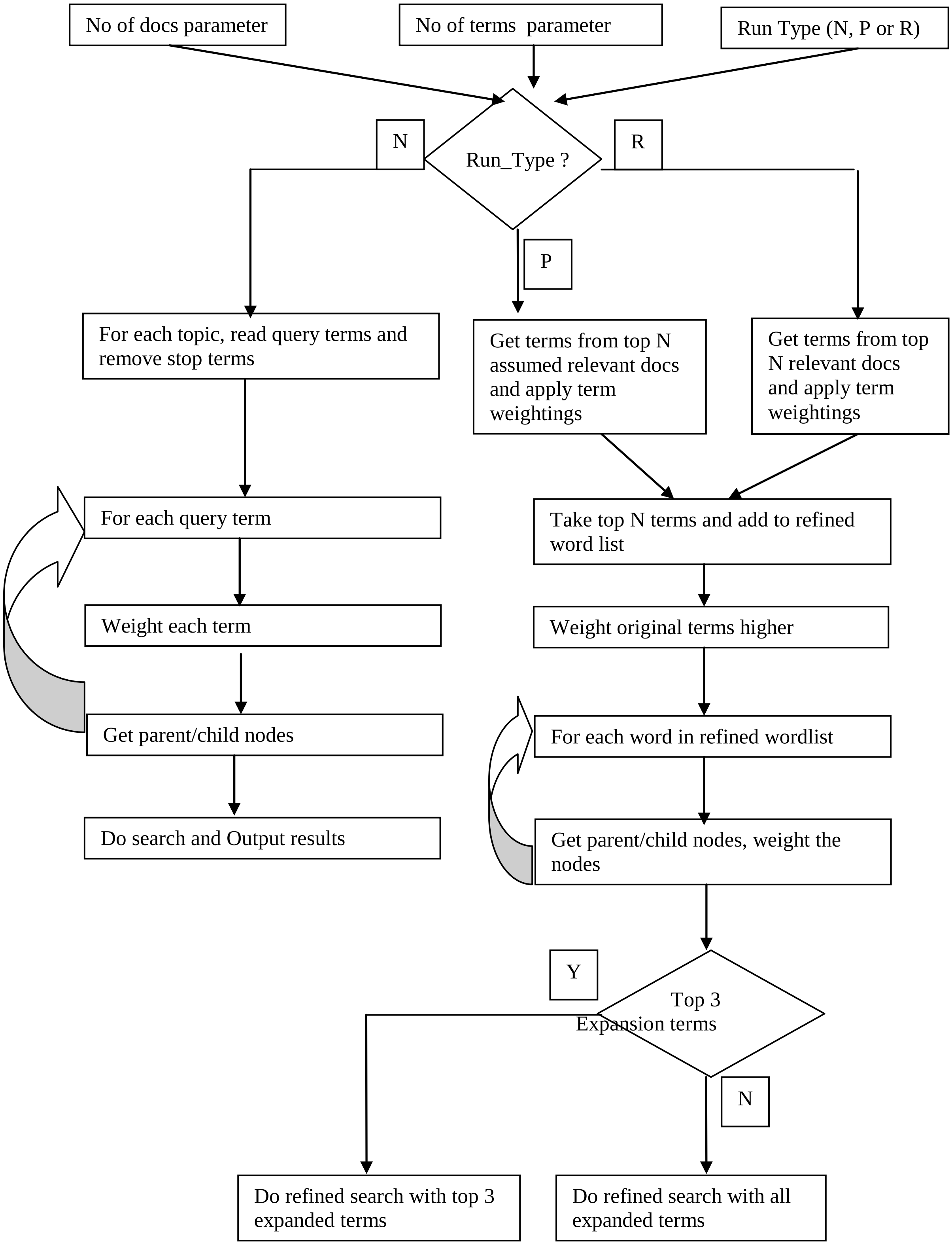


Figure 3.9 Diagram of Search Strategy

There is no real consensus on the optimum number of documents to use for Query expansion. Sparck-Jones (1979) used 20, Robertson et al, (1995) used 1000 (too much effort for very little return). We wanted to test this statement using experiments where the number of documents parameter was varied. Laboratory experiments will be used instead of user based experiments because of the repetitive nature of the experiments and the need to scale the experiments with ease. Table 3.2 describes the purpose of the experiments which are designed to meet objective 3.

Purpose of experiment	Experiment Number
Test ontology based query expansion compared to original system in order to show the baseline results	Experiment 1 uses standard relevance feedback parameter values of 20 documents and 20 terms
Test the effect of varying the number of terms relevance feedback parameter	Experiments 2, 3, 4, 5, 6 use term relevance feedback parameters of 5, 10, 15, 100 and 200 respectively
Test the effect of varying the number of documents relevance feedback parameter	Experiments 7, 8, 9, 10, 11 use document relevance feedback parameters of 5, 10, 15, 100 and 200 respectively
Test the effect of selecting a subset of the expanded terms	Experiment 12 uses standard relevance feedback parameter values of 20 documents and 20 terms but only selects the top 3 expansion terms

Table 3.2 Summary of Experiments

In testing out the hypothesis, it is important to examine what is meant by improved retrieval effectiveness. Improved retrieval effectiveness in its basic sense could just mean an increase in the number of relevant document returned as a result of ontology based query expansion. Table 3.3 shows for each experiment the increase/decrease in the number of relevant documents retrieved for pf and rfb runs. The table shows, with the exception of 1 case, the remaining 20 cases, the use of the ontology has resulted in an increase in the number of relevant documents retrieved. The approach used for relevance feedback where the top 3 query expansion terms are used is showing some success.

Run Type	Topics	+/- #docs retrieved (Pf)	+/-#docs retrieved (Rfb)
Standard	51-100	161	393
	101-150	103	252
	151-200	354	376
	201-250	249	323
	251-300	91	116
Top 3 terms	51-100	553	494
	101-150	539	440
	151-200	584	556
	201-250	337	330
	251-300	179	154

Table 3.3 Difference in the number of documents retrieved for pf and rfb standard run and top 3 terms run

However, improved retrieval effectiveness also refers to improved recall and precision. We also need to know whether the retrieved relevant documents appear high up in the ranked set of results. Clearly it is desirable to have as many relevant documents higher up in the ranked list of retrieved documents. Therefore the results will be analysed using the 5 metrics outlined in the next section.

3.4.2 Metrics

For the evaluation, the expanded queries should produce relevant documents higher up in the ranked list and there should be an increase in the number of relevant documents and a decrease in the number of non-relevant documents. This would indicate an improvement in retrieval results. The trec_eval program (Buckley, 2009) will be used to evaluate the resulting ranked list of documents.

Several standard metrics have been developed to measure retrieval effectiveness. One of the criticisms of early IR systems were they lacked robust and consistent testbeds and benchmarks. In early 1990s, Text Retrieval Evaluation Conference (TREC) was set up (Harman 1993) which is dedicated to experimentation with a large test collection. TREC uses summary table statistics, Precision-Recall averages and document level averages. Single value measures regarding the set of all queries can be stored in a table known as Summary Table Statistics. Summary Table Statistics record information such as the number of queries, total number of documents retrieved by all queries, total number of relevant documents which were effectively retrieved when all queries are considered, total number of relevant documents which could have been retrieved by all queries.

Our system will use the TREC measures such as Recall, Precision, document level averages, MAP (Mean Average Precision, sometimes referred to as AVEP) and Precision-Recall graphs. We will also be using t-tests on all of the results to measure their statistical significance.These measures are commonly used by other information retrieval systems thus making it easier to compare our results against those of other systems. The four measures are discussed below.

3.4.2.1 Recall

Recall is a measure of the ability of a system to present all relevant items.

$$\text{Recall} = \frac{\text{number of relevant items retrieved}}{\text{Number of relevant items in collection}} \quad (3.1)$$

All TRECs have used the pooling method to obtain relevance assessments. In the pooling method, human assessors make judgements on the top 100 documents. This method assumes that most relevant documents are in the pool and documents not in the pool are not relevant. (Baeza- Yates and Ribeiro-Neto 1999). It is important to measure recall for circumstances where the searcher wants as much information on the topic as possible and therefore is interested in retrieving as many relevant results as possible. Recall on its own is not very useful, we need to compare it with the number of non-relevant documents by calculating precision.

3.4.2.2 Precision

Precision is a measure of the ability of a system to present only relevant items. It takes all retrieved documents into account.

$$\text{Precision} = \frac{\text{number of relevant items retrieved}}{\text{Total number of items retrieved}} \qquad (3.2)$$

IR systems aim to have high precision because this means that the majority of documents retrieved are relevant to the user needs. So if System A retrieves only 2 relevant documents out of a total of 100 documents, the precision value will be .02 whereas if System B retrieves 50 relevant documents out of a total of 100 documents, the precision value will be .50. So clearly System B performs better because it has higher precision. Searchers who are interested in retrieving some but not necessarily all relevant documents require high precision results for their query. Generally, Recall and Precision clearly trade off against one another. Precision usually decreases as the number of documents retrieved is increased. The ideal would be to achieve high recall and high precision.

3.4.2.3 Mean Average Precision

MAP is a single-figure measure of quality across recall levels. MAP is used to rank relevant documents higher up in the resulting document set. The measure is not an average of the precision at standard recall levels. Rather, it is the average of the precision value obtained after each relevant document is retrieved. MAP has been shown to have especially good discrimination and stability. When a relevant document is not retrieved at all, its precision is assumed to be 0.

Precisions are computed at the point of each of the relevant documents in the ranked sequence and an average precision is produced. It is calculated as follows:

$$\mathrm{AveP} = \frac{\sum_{r=1}^{N} (P(r) \times \mathrm{rel}(r))}{\text{number of relevant documents}} \qquad (3.3)$$

where *r* is the rank, *N* the number retrieved, *rel()* a binary function on the relevance of a given rank, and *P(r)* precision at a given cut-off rank:

$$\mathrm{P(r)} = \frac{|\{\text{relevant retrieved documents of rank r or less}\}|}{r} \qquad (3.4)$$

As an example, consider a query that has four relevant documents which are retrieved at ranks 1, 2, 4, and 7. The actual precision obtained when each relevant document is retrieved is 1, 1, 0.75, and 0.57, respectively, the mean of which is 0.83. Thus, the average precision over all relevant documents for this query is 0.83. In contrast another query has 4 relevant documents which are retrieved at ranks 1, 2, 9 and 10. The actual precision obtained when each relevant document is retrieved is 1, 1 and 0.33, 0.40 respectively, the mean of which is 0.68. We can say query 1 performs better than query

2 because it places the relevant documents at a higher ranking and thus has a higher MAP.

3.4.2.4 Binary Preference (BPREF)

MAP is only calculated for relevant documents. It does not make any distinction in pooled collections between documents that are explicitly judged as non-relevant and documents that are assumed to be non-relevant because they are not judged, so valuable information might be ignored or discarded. Bpref and MAP scores aren't comparable directly, as they measure different things. Bpref is a measure which is meant to be robust in the face of missing relevance judgements, assuming that those judgements are missing uniformly.

Bpref allows us to view how known relevant and non-relevant documents are ranked rather than just focusing on all the relevant documents in the collection. For document collections where relevance judgement information is incomplete, we can use the bpref measure (Buckley and Voorhees 1994). Bpref uses binary relevance judgements to compute a preference of whether judged relevant documents are retrieved ahead of judged non-relevant documents. It is inversely related to the fraction of judged non-relevant documents that are retrieved before relevant documents. Queries that have a higher bref are better than queries with low bpref. For a topic with R relevant documents where r is a relevant document and n is a member of the first R judged non-relevant documents as retrieved by the system:

$$\text{bpref} = \frac{1}{R}\sum_{r} 1 - \frac{|n \text{ ranked higher than } r|}{R} \qquad (3.5)$$

The trec_eval program uses a slight variation on the above formula:

$$bpref = \frac{1}{R}\sum_{r}\left(1 - \frac{|n \text{ ranked higher than } r|}{min(R,N)}\right) \qquad (3.6)$$

R: number of judged relevant documents
N: number of judged non-relevant documents
r: relevant document retrieved
n: member of the first R judged non-relevant documents retrieved

MAP is a stable and sensitive measure but has been criticized as favouring the first few retrieved relevant documents. With bpref, each relevant document's score is independent of all other relevant documents' scores.

3.4.2.5 Document level averages

Average precision (over all topics) is computed at specified document cutoff values (instead of standard recall levels). The precision computed after a given number of documents have been retrieved reflects the actual measured system performance as a user might see it. Each document precision average is computed by summing the precisions at the specified document cut-off value and dividing by the number of topics

(50). For instance, the average precision might be computed when 5, 10, 15, 20, 30, 100, 200, 500 and 1000 relevant documents have been seen. This is similar to precision-recall graphs but plots precision against document cutoff values instead of recall.

3.4.2.6 Precision-Recall Graphs

To evaluate ranked lists, precision can be plotted against recall. For example a given query q has 10 relevant documents

R_q={d3,d5,d9,d25,d39,d44,d56, d71,d89,d123}

The retrieval systems retrieves only 5 of the relevant documents d123, d56, d9, d25, d3) with rankings 1, 3, 6, 10 and 15 respectively:

Precision for each relevant document retrieved is calculated as:

$$P(r) = \frac{|\{\text{relevant retrieved documents of rank r or less}\}|}{r} \qquad (3.7)$$

So at recall .10, precision for d123 is 1/1 (=1); at recall .20, precision for d56 is 2/3 (=0.66) ; at recall .30, precision for d9 is 3/6 (=0.50); at recall .40, precision for d25 is 4/10 (=0.40); at recall .50, precision for d3 is 5/15 (=0.30)

The results can be plotted on a precision-recall graph as shown in figure 3.10. Typically these graphs slope downward from left to right, enforcing the notion that as more relevant documents are retrieved (recall increases), the more non-relevant documents are retrieved (precision decreases). This graph is the most commonly used method for comparing systems. The plots of different runs can be superimposed on the same graph to determine which run is superior. Curves closest to the upper right-hand corner of the graph (where recall and precision are maximized) indicate the best performance.

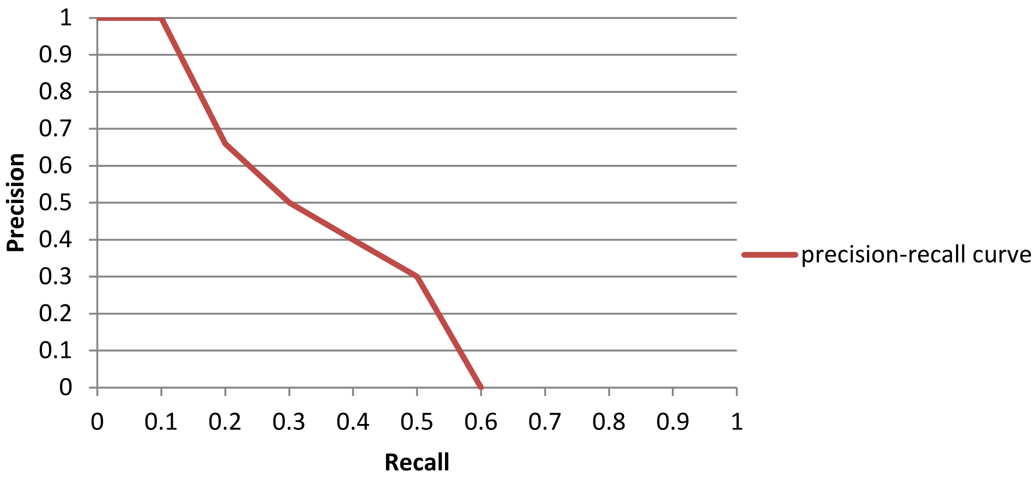


Figure 3.10 Precision-Recall Graph for example system

Precision-Recall is usually based on 11 standard recall levels: 0%, 10%, ..., 100%

However, to compute the average performance over a set of topics each with a different number of relevant documents, individual topic precision values are interpolated to a set of standard recall levels (0 to 1 in increments of .1). Interpolation is when average precision is calculated for a number of queries at each given recall level. For our purposes, Precision-Recall graphs will be used to compare the retrieval performance of distinct retrieval algorithms. They allow us to quantitatively evaluate the quality of the overall answer set and the breadth of the retrieval algorithm.

3.4.2.7 T-test metric for measuring statistical significance

The standard metrics used for evaluation in information retrieval have certain pitfalls. Single metrics such as precision and recall, provide limited information, whereas decision curves such as precision-recall graphs illustrate which model is best for a specific region and provide better visualisation. The evaluation methods assume that the relevance of one document is treated as independent of the relevance of other documents in the collection. Relevance of a document to an information need is treated as an absolute, either the document is relevant or it is non-relevant. Judgements of relevance are subjective, varying across people.

We also have to assume that users' information needs do not change as they start looking at retrieval results. Any results based on one collection are heavily skewed by the choice of collection, queries, and the relevance judgment set: the results may not translate from one domain to another or to a different user population. T-tests are carried out to check the reliability (statistical significance) of results and that the difference in results is not just due to chance/error. T-tests were carried out for all results in the PF runs and the rfb runs separately. For the t-tests we have chosen to use the standard confidence level p=0.05, and 1-tailed tests have been used for testing whether the ontology based query expansion algorithm produces an improvement in the retrieval results compared to the baseline retrieval performance. The null hypothesis assumes that the means for ontology based query expansion algorithm and the original algorithm are the same. The alternate hypothesis states that the mean for the the new algorithm is greater than the mean produced by the original algorithm. The null hypothesis is rejected for t values of <0.05 and the alternative hypothesis is accepted.

T-tests will be carried out on Document level averages metric, Precision-Recall graphs, MAP, and Average Recall to measure the statistical significance of these results. For example if the MAP for topic 51-100 shows improvement in the precision and the t-test result for MAP is <0.05 then we can say that it is likely that the ontology has improved retrieval effectiveness in terms of precision because the t-test results are statistically significant.

The t-test and is a parametric test because it assumes certain conditions about the parameters of the population from which the samples are drawn. One of the problems with T-test is that it assumes normal distribution. An alternative significance measure is the chi-squared test. Chi-square statistics use nominal (categorical) or ordinal level data, thus instead of using means and variances, this test uses frequencies.

The Chi-square is a statistical test commonly used to compare observed data with data we would expect to obtain according to a specific hypothesis. It is non-parametric because it does not require the sample data to be normally distributed. The chi-square test is always testing what scientists call the **null hypothesis,** which states that there is no significant difference between the expected and observed result.

The formula for calculating chi-square (χ^2) is:

$$\chi^2 = (o-e)^2/e \quad (3.8)$$

That is, chi-square is the sum of the squared difference between observed (*o*) and the expected (*e*) data (or the deviation, *d*), divided by the expected data in all possible categories. This statistic is then compared to a chi-square distribution table with known degrees of freedom in order to arrive at the p-value. The null hypothesis is that the observed values are close to the predicted values. We use the p-value to decide whether or not we can reject the null hypothesis. If the p-value is less than .05, then we can reject the null hypothesis. There are two types of chi-square test.

- *The Chi-square test for goodness of fit* which compares the expected and observed values to determine how well an experimenter's predictions fit the data.
- *The Chi-square test for independence* which compares two sets of categories to determine whether the two groups are distributed differently among the categories. In this context independence means that the two factors are not related. It is important to keep in mind that the chi-square test for independence only tests whether two variables are independent or not, it cannot address questions of which is greater or less. (McGibbon, 2006)

The main weakness of nonparametric tests is that they are less powerful than parametric tests. They are less likely to reject the null hypothesis when it is false. When the assumptions of parametric tests can be met, parametric tests should be used because they are the most powerful tests available (Key 1997).

In addition, Sanderson and Zobel (2005) found that past work on significance testing overestimated the error of t-tests and identified artificially high error rates for significance measures. They found that the t-test is more reliable than the Wilcoxon signed rank test and the latter produced a higer rate of Type-I errors. Type-I errors are known as false-positives (retrieved document is not relevant) which means they reject the null hypothesis when in fact it is true. They found that as topic set size increased, the error rates reduced with clear exponential trends. T-test results based on false-positive rate of 5% and MAP difference of > 10% are reliable. The t-test is also more relevant than showing large percentage difference in effectiveness measures between IR systems.

3.5 Summary

The methodology shown here describes the steps taken to select a suitable document collection and an appropriate news ontology; enable the two to interface with each other; modify the relevance feedback routines to include ontology information in the query expansion process and finally measure the effects of using ontology based query expansion.

TREC is a good document collection to use because it is large and has readily available relevance judgements and queries. The chosen ontology is one that has been derived from news articles so it is appropriate to use it for the TREC document collection. This step has ensured we have achieved most of objective 1 “Select a document collection with an appropriate ontology and index the document collection in the standard manner”. The remaining part of objective 1 which constitutes the indexing of the document collection will be discussed in chapter 4. The experiment design section shows how we have achieved objective 3 “Design and conduct laboratory experiments in order to compare and contrast the performance of the standard retrieval model with the revised retrieval model. The experiment design section outlines the techniques used to enable the ontology to interface with the document collection. The standard retrieval model uses rfb and PF techniques. The revised retrieval model will build on the existing retrieval model and incorporate the use of the ontology information into the query expansion process. Different types of evaluation metrics are required to evaluate the performance of each retrieval model and conduct a comparison. Single-value metrics and Ranked metrics have been discussed. Recall and Precision are single-value metrics which evaluate the quality of an unordered set of documents returned by the system. For systems that return a ranked sequence of documents, it is desirable to also consider the order in which the returned documents are presented. There are three metrics used for this purpose, namely mean average precision, bpref and precision-recall curves. A discussion on the use of t-tests to indicate the statistical significance of the results has been included.

This is the first time this particular TREC document collection and news ontology are being used in conjunction with each other so the results produced can provide useful baseline statistics for other researchers who want to carry out retrieval experiments using this particular combination of document collection and ontology.

4. SYSTEM

4.1 Introduction

Okapi is an experimental IR system, written to examine various aspects of interactive IR research, including such tasks as bibliographic search and full-text search. (Macfarlane et al, 2010). The Okapi software comprises a low level basic search system (BSS), together with data conversion and inversion utilities. There were also various scripts and programs for generating query terms, term weighting, merging sets of terms and performing the evaluation. The main code is written in C. The evaluation program is from Chris Buckley at Cornell University.

A single processor Sun SS10 with 64MB of core and about 12GB of disk was used as the main development machine and file server.The document collection in the disk2 database was used. The system uses the Probabilistic Retrieval Model and BM25 weighting functions are used to rank the documents (Sparck-Jones et al, 2000). BM25 is a best match operator which retrieves ore relevant documents higher up the rank (MacFarlane and Tuson 2009). Figure 4.1 shows an overview of all the components which make up the final system.

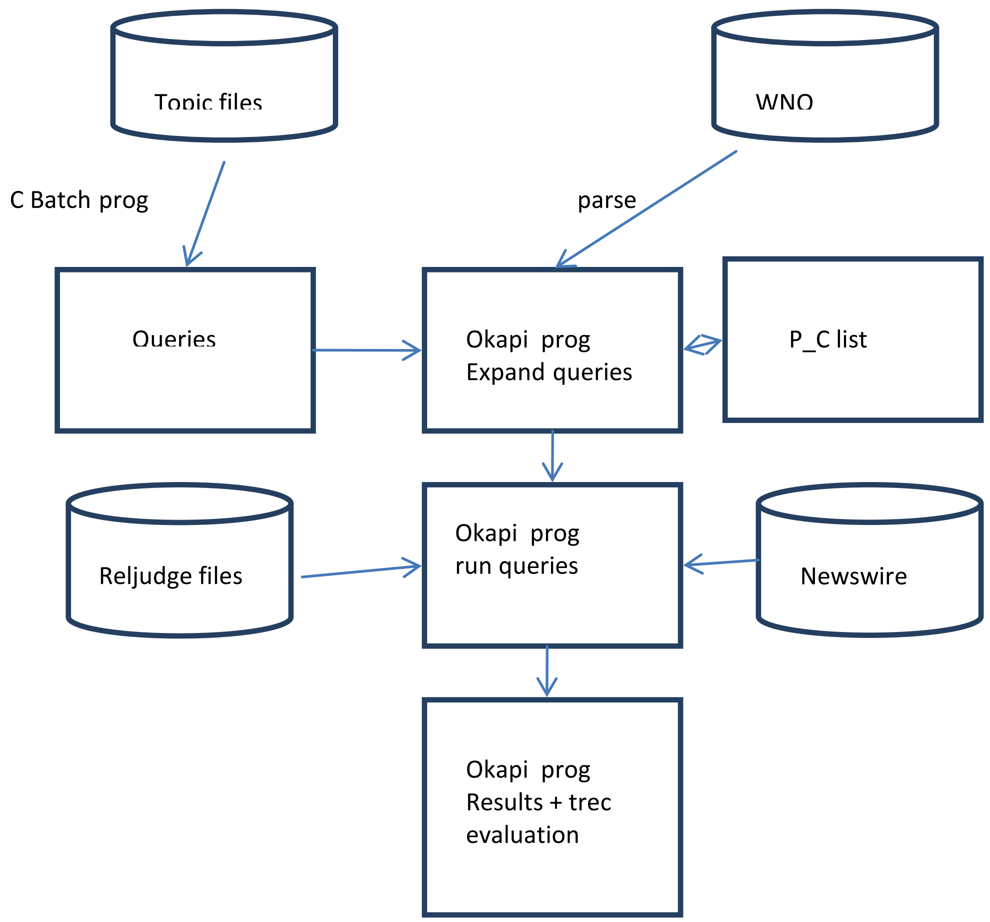


Figure 4.1 : System Overview

4.2 Indexing the Collection

This section shows the achievement of the indexing part of objective 1: A document collection will be selected together with an appropriate ontology. The document collection will be indexed in the standard manner.

Document pre-processing has to take place prior to indexing the collection. Document pre-processing consists of 3 stages: Lexical analysis, stop-word elimination and stemming. Lexical analysis is when the stream of characters are converted into stream of words by eliminating hyphens, punctuation marks, and any casing of letters. Then stop-words are eliminated because they have a low discrimination value for retrieval purposes and the elimination process helps to reduce the size of the index. Finally, the remaining words are stemmed with the objective of removing affixes and allowing retrieval of documents containing syntactic variations of query terms. We use the Porter algorithm (1980) to carry out word stemming. Sometimes noun groups are used to select which words/stems are going to be used for indexing.

The indexing information is contained in two separate files – database parameter file and attribute parameter file. The main database parameter specifies which fields are to be used as the source of keys, the number of indexes (attributes), the index type of each and where its files are to be placed. The attribute parameter file <db name>.search_groups contains information on which fields are to be used as source of keys, the indexing regime used to parse the source and extract keys, the stemming function used, name of the language database used if any, index type and the attribute name to be assigned (Okapi-Pack 2001).

For document collections which are in sgml/xml format, the convert_runtime process is not required. When constructing an index, the indexing program reads documents in internal record number sequence, extracting terms and positional information for the required index from the specified fields. Separate files exist for each of the document collections. These files were merged into a single file in sgml format using zcat. Several alterations had to be carried out on the sgml file before it could be indexed. Firstly, the xml headerline had to be inserted at the beginning of the sgml file and <DISK2DATA> document root tags were added. Also ENTITY declarations using DOCTYPE to include references to all dtd files were added to the disk2.xml.bib file. For the parameter file, all instances of tagnum= had to be surrounded with quotes and the tags in the parameter files had to be written in uppercase. The sgml file was then indexed. The resulting indexed sgml file had to be edited to remove some spurious characters or format the tags properly.

4.3 Probabilistic Retrieval Model

The Okapi system is based on the probabilistic retrieval (Sparck-Jones et al, 2000) model which uses Bayes Theorem to determine the probability that a specific document will be judged relevant to a specific query. This model is based on the assumption that the terms are distributed differently in relevant and non-relevant documents. Matching documents are ranked according to their relevance to a given search query using a "best match" ranking function called BM25 (sometimes referred to as OkapiBM25 because the Okapi information retrieval system was the first to implement the BM25 function). BM25 is based on the probabilistic retrieval framework developed in the 1970s and 1980s (Robertson and Sparck-Jones 1976). The BM25 is a progression of the F4 ranking function. The 'F4' formula, point-5 version is

$$w = \log \frac{(r+0.5)\,(N-R-n+r+0.5)}{(R-r+0.5)\,(n-r+0.5)} \tag{4.1}$$

where N = collection size
n = number of postings of term
R = total known relevant documents
r = number of these posted to the term

BM25 is a bag-of-words retrieval function that ranks a set of documents based on the query terms appearing in each document, regardless of the inter-relationship between the query terms within a document (e.g., their relative proximity). It is not a single function, but actually a whole family of scoring functions, with slightly different components and parameters. One of the most prominent instantiations of the function is as follows.

Given a query Q, containing keywords $q_1,\ldots,q_n$, the BM25 score of a document D is:

$$\text{score}(D,Q) = \sum_{i=1}^{n} \text{IDF}(q_i) \cdot \frac{f(q_i,D)\cdot(k_1+1)}{f(q_i,D) + k_1 \cdot \left(1 - b + b \cdot \frac{|D|}{\text{avgdl}}\right)}, \tag{4.2}$$

where $f(q_i,D)$ is q_i's term frequency in the document D, $|D|$ is the length of the document D in words, and $avgdl$ is the average document length in the text collection from which documents are drawn. k_1 and b are free parameters, usually chosen as $k_1 = 2.0$ and $b = 0.75$. $\text{IDF}(q_i)$ is the IDF (inverse document frequency) weight of the query term q_i. It is usually computed as:

$$\text{IDF}(q_i) = \log \frac{N - n(q_i) + 0.5}{n(q_i) + 0.5}, \tag{4.3}$$

where N is the total number of documents in the collection, and $n(q_i)$ is the number of documents containing q_i.

In the original BM25 derivation, the IDF component is derived from the Binary Independence Model where relevant documents have a score of 1 and non-relevant documents have a score of 0. Rather than just providing a term weighting method for terms in a user's query, rfb can also involve augmenting the query (automatically or with manual review) with some (say 20) of the top terms in the known-relevant documents the above formula can be used with such an augmented query vector.

4.4 Search Method

We will be using the search method used in the probabilistic retrieval model whereby query terms are matched with terms in the index and terms are weighted to produce a ranked list of documents. The first step is to load the original keyword set and the refined keyword set. Then we weight the sets and merge the sets into a single set. The terms are weighted using the BM25 formula (Sparck-Jones et al, 2000) and the documents are ranked in descending order of probability of them being relevant to the user. With the revised system we will search each term in the refined keyword set and extend the set further by adding the parent node and child node information for each term. The same weighting mechanism will be applied to the new terms obtained from the ontology. The refined query (consisting of original query terms, terms obtained from relevance feedback and terms derived from the ontology) will be used to do a further search.

4.5 Term Selection Method

Query files can be based on title, title and description or title, description and narrative. Table 3.1 shows that the topic length based on title is much shorter in comparison to the topic length for Description and Narrative (Robertson et al, 1995. As highlighted in chapter 2, short queries are better candidates for query expansion because they have insufficient terms to describe the information need and tend to be more ambiguous (Navigli and Velardi 2003). Therefore the query files will be based on the topic titles only because they form shorter queries compared to queries based on the topic description.

	Min	Max	Mean
TREC-1(51-100)	44	250	107.4
Title	1	11	3.8
Description	5	41	17.9
Narrative	23	209	64.5
TREC-2 (101-150)	54	231	130.8
Title	2	9	4.9
Description	6	41	18.7
Narrative	27	165	78.8
TREC-3(151-200)	49	180	103.5
Title	2	20	6.5
Description	9	42	22.3
Narrative	26	146	74.6
TREC-4(201-250)	8	33	16.3
Description	8	33	16.3
TREC-5(251-300)	29	213	82.7
Title	2	10	3.8
Description	6	40	15.7
Narrative	19	168	63.2

Table 4.1: Topic Length statistics (including stop words).

One of the definitions of an ontology states “Ontologies provide consistent vocabularies and world representations necessary for clear communication within knowledge domains” (Leroy et al, 2000). Therefore it made sense to use an ontology within the news knowledge domain for query expansion. The overall purpose of our experiments was to assess the usefulness of ontology based query expansion in providing contextual information to resolve ambiguous queries and improve the search results. The Okapi system already provides query expansion using relevance feedback features but we wanted to extend the query expansion mechanism by combining it with the use of an ontology. The scope of the experiments is limited because we have used a simple ontology which does not include rich semantic information and it is also relatively small in size. In addition to this our experiments are limited because we are only using the news domain and not looking at any other type of domain. However, despite these limitations we think it is a good starting point and will provide a benchmark for future research in this area. Apart from the detailed investigation into ontology based query expansion, we also wanted to test the effect of varying the relevance feedback parameters for pf and rfb to see if a trend could be detected or an optimum combination could be found.

For the experimental runs, all documents in the disk 2 training collection were used. Experiments were done using a total of 250 queries in the Ad-hoc TREC task. These queries were based on topics 51-100 (TREC-1), 101-150 (TREC-2) ,151-200 (TREC-3), 201-250 (TREC-4) and 251-300 (TREC-5) Like most traditional retrieval collections, there are three distinct parts to the collection – the documents, the questions or topics and the relevance judgements. The newswire document collection has associated topics/queries and a set of relevance judgements. Therefore it is ideal to use as a test collection for information retrieval evaluation.

The query terms for each topic will be constructed automatically using an Okapi query generator utility program. An important question to address is whether to expand all queries or just the ambiguous queries. For the purposes of our system, all queries are expanded. We have not yet designed a mechanism to identify ambiguous queries. With regards to term selection, all query terms are used for query expansion. Each query term in the Okapi Index will be searched to provide new query terms, however in addition to this, the parent-child database will be searched to provide ontology based query expansion terms.

4.6 Term Weighting Method

The Okapi system makes use of the Best Match weighting function BM25. In query expansion after relevance feedback in Okapi, terms from the relevant items are ranked according to some selection value which is intended to measure how useful they would be if added to the query. The formula used for this purpose is the Retrieval Selection Value (RSV). RSV is calculated as follows:

$$w(p-q) \tag{4.4}$$

where w is the weight to be assigned to ther term, p is the probability of the term occurring in a relevant document and q is the probability that it occurs in a non-relevant document (Robertson et al, 1995).

The laboratory style experiments will be designed to test issues such as whether to use a bottom-up/top-down search strategy for the ontology and search depth (ie how many levels to search in the hierarchy).

Separate experiments will be conducted for PF runs and Rfb runs. Relevance feedback works when the searcher's vocabulary matches the collection vocabulary and relevant documents have similar term frequencies (Manning et al, 2008). Pf is also known as blind relevance feedback. The process of giving feedback is automated and the user gets improved retrieval results without any extra interaction. The pf method retrieves an initial set of documents based on a normal search and then *assumes* that the top n ranked documents are relevant. The feedback terms are taken from the top n ranked documents. Rfb is where a group of users make judgements on the relevance or non-relevance of documents. A binary scoring mechanism is used whereby 1 indicates a document is relevant and 0 indicates it is not relevant. Each of the runs will use 20 documents and 20 terms. The next step will be to run the experiments and analyse results in order to assess the effectiveness of the query expansion procedure. The results will show whether improvements are made in the area of recall or that of precision.

Term weighting functions can also depend on the path length and type of relationship. Investigations can be carried out on whether terms that are derived from a shorter path distance should be assigned a higher weighting than other nodes which are further in distance from the current node. General ontologies such as WordNet contain information on homonyms, synonyms etc. In such cases term weighting can also vary depending on the type of relationship being examined. For example terms that are nouns could have a higher weighting than synonyms. There is some evidence that local analysis works better than global analysis (Manning 2008). Query Expansion is effective in increasing recall, it is less successful than rfb and may be as good as pf (Billerbeck and Zobel 2003). Our research experiments can be used to test these claims.

4.7 Algorithms

Before describing the algorithms, we need to explain the parsing process needed for all XML files. XML Parsers read the xml file into memory, it will check if it is well formed and if given a Document Type Definition (DTD) will check the document's validity. When an xml document is parsed, the parser returns a tree built during the document analysis referred to as the Document Object Model (DOM). The XML DOM defines a standard way for accessing and manipulating XML documents. The DOM presents an XML document as a tree-structure.

The value returned is an **xmlDocPtr** (i.e., a pointer to an **xmlDoc** structure). This structure contains information such as the file name, the document type, and a **children** pointer which is the root of the document (or more exactly the first child under the root which is the document). The tree is made of **xmlNode**s, chained in double-linked lists of siblings and with a children<->parent relationship. An xmlNode can also carry properties (a chain of xmlAttr structures).

Many XML Parsers exist but we chose Libxml2 (Velliard n.d) because it is written in the same programming language as the Okapi software. Libxml2 is the XML C parser and toolkit which is open source software.

4.7.1 Building Parent-Child database

This section discusses the achievement of objective 2 "building a separate database containing semantic information such as parent-child relationships between ontology nodes. This information will be used to supply additional terms for expanding the original query terms."

For this algorithm XPath is not required, instead the XML tree structure is traversed in order to obtain the parent child information for each node.

A routine was written to parse the ontology file (WNO.xml) and obtain the DOM tree structure. Starting from the root element node, each node in the tree structure was processed to obtain its child nodes and all parent nodes. This was a recursive routine to ensure all nodes in the tree structure had been processed. The algorithm for building the parent-child database is shown in Figure 4.2

Parse the file and get the DOM
Get the root element node.
Recursive routine:
    Process every node in the ontology
        get its childnode
        record parent-child information
End Recursive routine

Figure 4.2 Algorithm for Building the Parent-Child database

The information is stored in memory using a list structure which consists of:
(childnode; parentnode; original term; weighting, r; nwords; levelno).

Childnode is the descendant of the original terms, parentnode is the ancestor of the original term and original term is either a term from the query topic or from the refined wordlist resulting from the relevance feedback process. Weighting is the weighting value; r is the number of relevant documents for the term; nwords is the number of words in the list and levelno is distance from the original term to the parent/child term. So for example :

Crime_Law_Justice
|
Crime
|
Police
|
Law_Enforcement

If we searched for the term Crime then there would be the following list entries:
(Police; Crime_Law_Justice; Crime; 0.0, r; nwords;1);

(Law_Enforcement; Crime_Law_Justice; Crime; 0.0; r; nwords; 2);

The weighting values are updated during the three types of search routines which are explained below.

4.7.2 Algorithm for Search routine

The basic search routine uses word stemming to and removes stop words before conducting a search. The basic search routine was amended to include the parent child information for each query term. This ontological information is then used for conducting a subsequent refined search. Figure 4.3 shows the algorithm for the basic search routine.

While not end of topic file
- Do_exp_search
- Read query terms for given topic
- For each query tem:
  - Remove any stopterms before doing a search
  - If the term exists in the Okapi index, then determine the set and weight and append (stem, original term, 0, weight) to wordlist
  - *Search parent-child list to get the parents of this term and record into broadlist*
  - *Search parent-child list to get the child nodes of this term and record into narrowlist*
- Return bss document set number and number of postings for the best match search on all terms
- For each item in the document set, record the topic number, document number and weighting information in the results file

End while

Figure 4.3 Search Routine Algorithm

The results file sometimes contained multiple instances of document id numbers which was problematic for the trec evaluation program. In order overcome this problem. the routine was modified to add the topic number and the sequence number (using underscores in between) to the document id to make it unique. As an illustration, the results file for topic 51-100 contains the following sample lines of data:

Q0 AP880619_51_445 445 33.727001 ok-test

Q0 AP880619_51_446 446 33.689999 ok-test

The first field is the set id, the second field is the document id , the third field is the sequence number, the fourth field is the weighting and the final field is the run-id. In this example there are two instances of the same document id AP880619. If we look at the first instance, the topic number 51 has been appended to the document id followed by the sequence number (445) in order to make each document id occurrence in the results file unique. These changes to ensure unique document id numbers in the results file have also been applied to the Pf routine and the rfb routine.

Another point worth noting is that the original wordlist has a simpler structure compared to the expanded wordlist. It has 5 parameters as follows:
(stem,original term, r, weight)

Where r is the no of relevant documents containing the search term.
The new list structure which is used for broadlist and narrowlist has 7 parameters as described earlier in section 4.7.1.

4.7.3 Algorithm for Pseudo Relevance Routine

This routine and the Rfb routine also use word stemming prior to searching the ontology in order to maximise the chances of finding a match and avoiding non-match situations.

With PF it is assumed that the top N documents are relevant. As expected of PF, the top N keywords are extracted from the top N documents where N is a parameter value supplied at run-time by the user. The standard values used are 20 terms and 20 documents. The terms are then weighted. Figure 4.4 shows the weighting algorithm which uses a tuning constant ie a weighting factor of 4. The value of the tuning constant is set experimentally to find one which works the best. The original query terms are given a higher weighting so as not to distort the query results. The algorithm uses a weighting factor of 4. This is supported by Voorhees(1994) who states that the assignment of lower weights to added concepts enhances the retrieval accuracy. The weighting formulas are part of the original source code for Okapi and these have not been amended in any way.

Further query expansion takes place by searching for and adding the parent-child ontological information for each term in the refined list. Any additions to the original Okapi routine are shown in italics. The new terms that have been obtained from the ontology are weighted. A final refined search is conducted which is based on the query expansion terms obtained from PF and from the ontology.
Jones et al, (1995) state that terms should be expanded before carrying out the search of the document database to overcome the fact that there might not be an exact match of query. The algorithm used for the Pf routine is shown in Figure 4.4.

Do_pseudo

Strip terms from top N documents and weight them using Robertson Selection Value (RSV)
Take Top N keywords and add to original list to form refinedwordlist
Weight the terms in the refinedwordlist using FUNC2
Weight the terms in the original list using FUNC2
Multiply weights in the original list by 4.
Merge original list and refined wordlist
Save refined keywords (topicno, weight, refined keyword) to file on disk
Do refined search with refinedwordlist
For each word in the refined wordlist expand it further:
Search parent-child list to get the parents of this term and record the parent node details in the list
Search parent-child list to get the children of this term and record the child node details in the list
Terms are weighted using same weighting function that is used for the Top N keywords (FUNC2)
Merge the narrowlist with refined wordlist
Save refined keywords for narrowlist (topicno, weight, original term, childterm, levelno) to disk
Merge the broadlist with refined wordlist
Save refined keywords for broadlist (topicno, weight, original term, parent term, levelno) to disk
If ontology search produced childnodes/parentnodes then
Do refined search with the expanded list
endfor

Figure 4.4 Algorithm for Pf Routine

4.7.4 Algorithm for Rfb routine

With Rfb, instead of assuming the top N documents to be relevant, the top N documents that are used have actually been judged as relevant. Figure 4.5 shows the algorithm for the Rfb routine.

Do Rfb

Strip terms from N relevant documents and weight them using RSV
Take Top N keywords and add to original list to form refinedwordlist
Weight keywords in original list using FUNC2
Merge original list and refined wordlist
Save refined keywords
Do refined search with refinedwordlist
For each word in the refined wordlist expand it further:
Search parent-child list to get the parents of this term and record orig query term, current term, parentterm, levelcount into broadlist
Search parent-child list to get the children of this term and record orig query term, current term, childterm, levelcount into narrowlist
Terms are weighted using same weighting function that is used for the Top N keywords (FUNC2)
Merge the narrowlist with refined wordlist
Save refined keywords for narrowlist (topicno, weight, original term, childterm, levelno) to disk
Merge the broadlist with refined wordlist
Save refined keywords for broadlist (topicno, weight, original term, parent term, levelno) to disk
If ontology search produced childnodes/parentnodes then
Do refined search with the expanded list
endfor

Figure 4.5 Algorithm for Rfb routine

A different weighting function is used to weight the terms taken from the top N relevant documents. The remainder of the routine is similar to the Pf routine. In query expansion after rfb in Okapi, terms from the top N relevant items are weighted using the RSV formula. RSV multiples the calculated weight by r (no of relevant documents containing the term which is being weighted.) The RSV ranking is intended to measure how useful they would be if added to the query (Robertson 1990). The formula given in that reference is w(p-q) where w is the weight to be assigned to the term, p is the probability of the term occurring in a relevant document and q is the probability that it occurs in a non-relevant document. For RSV, w is interpreted as the usual Robertson-Sparck Jones relevance weight, p is estimated as r/R and q is assumed to be negligible. A good term is one which tends to occur more frequently in relevant documents than in non-relevant documents. In the past this ranking of terms has been used to select the top n terms where n is fixed between topics.

4.7.5 Top 3 terms

With the previous Pf and Rfb routines, all of the query expansion terms have been used when conducting the final refined search. An alternative is to select some of the expanded terms for the refined search. One mechanism of selecting some of the terms is to pick the top N terms from the refined word list. Algorithms 4..4 and 4.5 are modified to select the top 3 terms from the expanded list. Figure 4.6 shows the relevant section of the code which has been modified to just use the top N terms from the expanded list where we have chosen N to be 3.

For each word in the refined wordlist expand it further:
Search parent-child list to get the parents of this term and record orig query term, current term, parentterm, levelcount into broadlist
Search parent-child list to get the children of this term and record orig query term, current term, childterm, levelcount into narrowlist
Terms are weighted using same weighting function that is used for the Top N keywords (FUNC2)
Merge the narrowlist with refined wordlist
Save refined keywords for narrowlist (topicno, weight, original term, childterm, levelno) to disk
Merge the broadlist with refined wordlist
Save refined keywords for broadlist (topicno, weight, original term, parent term, levelno) to disk
If ontology search produced childnodes/parentnodes then
Pick top 3 terms from the complete list of expanded terms
Do refined search with the top 3 terms
endfor

Figure 4.6 Algorithm for Top 3 terms routine

4.8 Summary

In this chapter we have achieved objectives 1 and 2 by giving a description of our information retrieval system and explaining the indexing process used. We present a brief explanation of the probabilistic retrieval model used by Okapi and then describe the algorithms which are used to extract the parent-child information from the chosen ontology, and integrate this semantic information into the use of rfb query expansion process and the PF query expansion process.

5. EXPERIMENT RESULTS FOR RELEVANCE FEEDBACK RUNS: STANDARD AND TOP 3 TERMS

5.1 Introduction

A number of experiments were used to investigate whether query expansion benefits from the use of an ontology. These experiments were designed to test different dimensions:

- effect of varying relevance feedback parameters such as number of documents (see chapter 7) and number of terms (see chapter 6) compared to standard relevance feedback parameters;
- to use all query expansion terms or a small selection (sections 5.2-5.3);
- assess which technique out of relevance feedback or pseudorelevance feedback is more effective (this is done in all experiments) and
- effect of searching at level 1 of the ontology compared to searching at level 2 (section 5.1)

The structure of this chapter is as follows: section 5.2 shows experiment set 1 which is run using the standard relevance feedback parameters of 20 documents and 20 terms and all terms from the expanded query ie no term selection. Sections 5.3 shows section experiment 12 in which the standard relevance feedback parameters of 20 documents and 20 terms and the top 3 query expansion terms are used. Finally the chapter is summarised in section 5.4. The remaining experiments are discussed in Chapter 6 (experiments 2-6 where the number of terms parameter are set to 5, 10, 15, 100 and 200 respectively), Chapter 7 (experiments 7-11 where the number of documents used for relevance feedback are set to 5, 10, 15, 100 and 200 respectively), Chapter 8 (discussion of the results for Recall, MAP and Bpref), and Chapter 9 (detailed discussion on topic analysis which tries to establish an explanation of the results achieved).

Although experiments for ontology based query expansion to a depth of level2 were conducted, their results have not been included in this discussion because level2 ontology nodes added very little gain if any. This is illustrated in Table 5.1 which shows the number of relevant documents returned at level1 query expansion and level2 query expansion for each topic set.

The table shows that for the pf run, level1 returns more results than the baseline but level2 only produces a handful of additional documents compared to level1 for topic 51-100, topic 101-150 and topic 151-200. For the rfb run, level1 returns more results than the baseline but level2 offers no improvement on level1. Such small gains do not make the graph curves for level1 and level2 distinctive, in fact most of the time because the results are identical, the level1 and level2 curves overlap on top of each other. Therefore we took the decision not to discuss the level2 results or show the level2 results on the graphs because they did not produce any significant improvement in the results.

Topic	Num_rel	Rel returned (PF)	L1 (PF)	L2 (PF)	Rel returned (RFB)	L1 (RFB)	L2 (RFB)
51-100	16386	1036	1197	1200	570	963	963
101-150	11645	1343	1446	1450	865	1117	1117
151-200	9805	1069	1423	1429	827	1203	1203
201-250	6503	733	982	979	490	813	813
251-300	5524	390	481	471	278	394	394

Table 5.1 Number of relevant documents returned for each topic set

Also for the Precision-Recall graphs, only the graphs which show a difference in the curves are discussed. Any Precision-Recall graphs which are highly similar have been included in the appendix.

5.2 Experiments Using Standard 20 Documents And 20 Terms and All Terms in Expanded Query

5.2.1 Document Level Averages Graphs

Table 5.1 shows that both pf and rfb runs retrieve a higher number of relevant documents compared to the standard run but the pf run shows a bigger improvement higher up the ranked set of results. For topics51-100, Figure 5.1 shows an improvement in retrieval performance for the pf run with the ontology over the pf baseline especially between ranks 5 to 30. At rank 10 there is a 17% improvement and the t-tests (Table 5.1) show these results to be statistically significant. However the ontology has not benefited rfb up until rank 100 and then after that point the improvement is only slight. At rank 5, there is a 18 % deterioration in results. The rfb t-test result is not statistically significant.

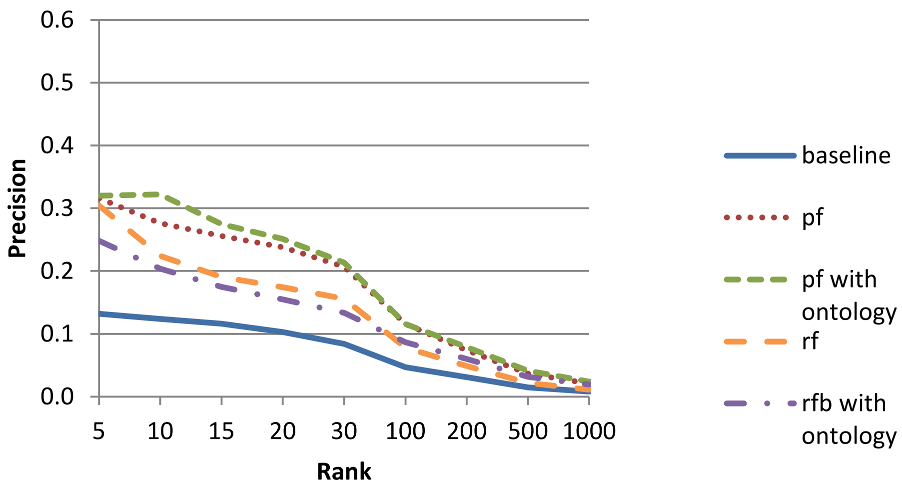


Figure 5.1 Topic 51-100 Retrieval Results – Document Level Averages

It seems that the rfb run is retrieving relevant documents but with a lower weighting than those retrieved for the pf run, thus the improvement for rfb is only showing at the lower end of the ranked list of documents. Some examples where the search for narrower terms could have a positive effect on precision is shown below:

```
Topic no = 67 (politically motivated civil disturbances)
ORIG WORD IS civil
     --> revolutions
     --> rebellions
     --> political_dissent
     --> religious_conflict
     --> social_conflict
     --> protest

Topic no = 89 (downstream investments opec member states)
ORIG WORD IS state
     --> public_finance
```

The topic description for topic 67 states that ocuments for this topic will focus on " civil disturbance in any country, involving citizens of that country protesting a political position of their own country's government". Clearly the child terms such as ***political_dissent*** and ***protest*** are directly related to topic 67 and will improve the retrieval results.

In some cases, many of the child terms are related to the search term but only a few of the child terms are relevant to the query as shown for below:

```
Topic no = 90 (data proven reserves oil natural gas producers)
ORIG WORD IS natural
     --> earthquake
     --> tsunami
     --> flood
     --> drought
     --> avalanche
     --> landslide
     --> land_resources
     --> parks
     --> forests
     --> wetlands
     --> mountains
     --> rivers
     --> oceans
     --> wildlife
     --> energy_resources
     --> geology
     --> paleontology
     --> geography
     --> physiology
     --> botany
     --> astronomy
     --> biology
```

The TIPSTER description for topic 90 covers documents that "provide totals or specific data on changes to the proven reserve figures for any oil or natural gas producer". In this case although we have many child terms, the only term that could be remotely related to the topic is ***energy_resources.***

In other cases, many of the child terms are related to the search term but not relevant to the query as shown for topic 68:

```
Topic no = 68 (health hazards finediameter fibers)
ORIG WORD IS health
     --> disease
     --> epidemic_plague
     --> health_treatment
     --> prescription_drugs
     --> medical_procedure
     --> therapy
     --> health_org
     --> medical_research
     --> medical_staff
     --> medicine
     --> preventative_medicine
     --> injury
     --> hospital
     --> clinic
     --> illness
```

The TIPSTER topic description for topic 68 refers to documents that "report studies or unsubstantiated concerns about the safety to manufacturing employees and installation workers of fine-diameter fibers used in insulation and other products". None of the child terms retrieved for topic 68 would be related to that topic.

In the case of Topics101-150, the ontology has not benefited the pf run. Figure 5.2 shows both pf curves to be virtually identical. The rfb run has not benefited from the ontology at all between ranks 5-30 after which the two curves are identical. At rank 5 there is a 30% deterioration. The t-test results for this topic set show statistical significance (Table 5.3).

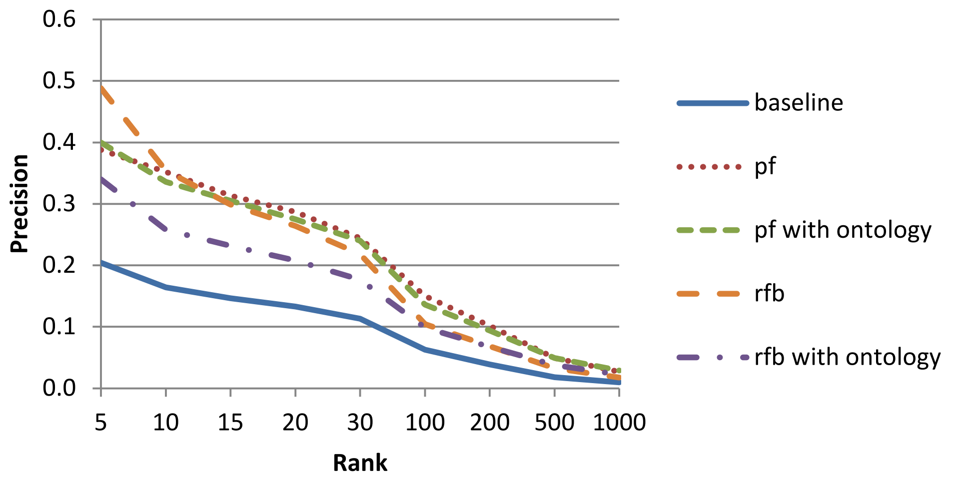


Figure 5.2 Topic 101-150 Retrieval Results – Document Level Averages

PF / RFB have not benefitted because the ontology terms chosen might have 1 or two relevant terms but the results are distorted by all of the documents retrieved for the terms that are not relevant to the query:

For example:

```
Topic no = 101 (design star wars antimissile defense)
ORIG WORD IS defenc
     --> veterans_affairs
     --> national_security
     --> security_measures
     --> troops_withdrawal
     --> armed_forces
     --> military_equipment
     --> firearms
     --> biological_chemical_weapons
     --> missile_systems
     --> nuclear_weapons
```

The TIPSTER topic description states that relevant documents will "provide information on the proposed configuration, components and technology of the U.S. star wars anti-missile defense system". Of the above 10 terms, probably only two terms (***national_security*** and ***missile systems***) appear to be related to the query. This term is picked up 3 times, so any distortion that takes place with precision is multiplied three times.

Also some of the ontology terms are general (see topic 123 below) so they will retrieve a huge number of documents containing the term ***survey*** but but it is likely that many of the documents will not be relevant to the query topic ***research control carcinogens***.

```
Topic no = 123
ORIG WORD IS research
     --> survey

TOPIC NUMBER = 115 (impact immigration law)
current word is immigration
     -->  demographics
 TOPIC NUMBER = 115 (impact immigration law)
current word is law
     -->  crime_law_justice
```

The TIPSTER topic description for topic 115 refers to documents that "report specific consequence(s) of the U.S's Immigration Reform and Control Act of 1986". Whilst these child terms are related to the search term and will retrieve many documents in these broad areas, how many are actually related to the ***impact of immigration law***?

```
TOPIC NUMBER = 126 (medical ethics modern technology)
current word is medical
     -->  health_treatment
```

The TIPSTER topic description for topic 126 refers to documents that "discuss ethical issues attendant to contemporary advances in medical technology". With the above example whilst many documents containing the word ***medical*** will be retrieved, it is difficult to guarantee that all of the documents retrieved will be related to ***medical ethics***.

The pf runs generally retrieve more ontology matches so this implies that the pf will have better results than rfb but we can't just go by number of ontology nodes retrieved alone, it depends on the relevance of the ontology parent/child node to the topic query.

In the case of Topics151-200, the ontology has not benefited the pf run. Figure 5.3 shows both pf curves to be virtually identical. The rfb run has not benefited from the ontology at all between ranks 5-30 after which the two curves are identical. At rank 5 there is a 21% deterioration.

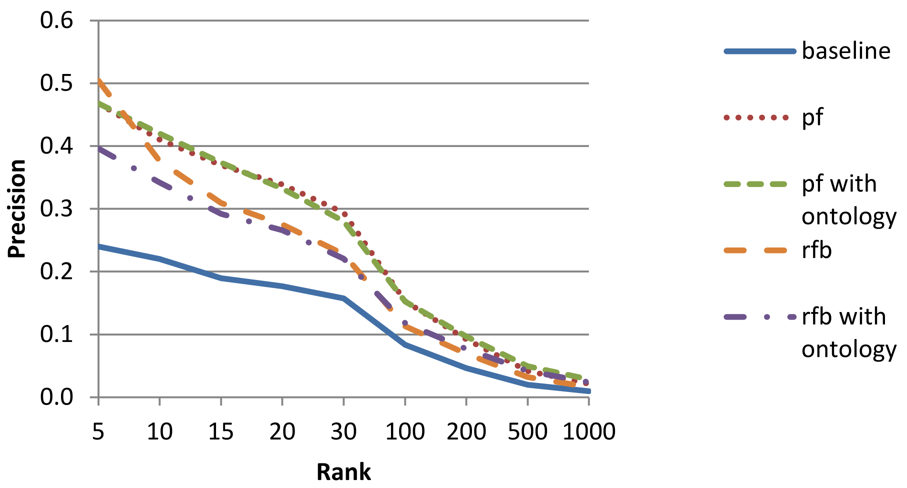


Figure 5.3 Topic 151-200 Retrieval Results – Document Level Averages

```
Topic no = 179 (restaurants foreign lands)
ORIG WORD IS foreign
     --> summit
     --> meeting
     --> diplomat
     --> embassy
     --> embassador
     --> delegation
     --> economic_sanction
```

The TIPSTER topic description for topic 179 refers to documents that "identify those countries where U.S. fast food restaurant chains are now or will be located". Whilst all of these terms are related to ***foreign***, they bear little relation to the query topic. Also topic 200 picks up on the term ***foreign*** which bears little relation to the query ***impact foreign textile imports us industry***.

With rfb there might be a few terms that help to improve precision, overall rfb also suffers from the same problem of ontology terms which have little relevance to the query topic.

In some cases we have good choice of ontology terms being retrieved

```
TOPIC NUMBER = 199 (legality medically assisted suicides)
current word is suicid
     -->  euthanasia
```

Euthanasia is related to the query topic ***legality medically assisted suicides***.

In the case of Topics201-250, the ontology has resulted in very slight improvement on the pf baseline. Figure 5.4 shows this improvement from rank 20 onwards. The rfb run has not benefited from the ontology at all between ranks 5-30 after which the two curves are identical. At rank 5 there is a 21% deterioration.

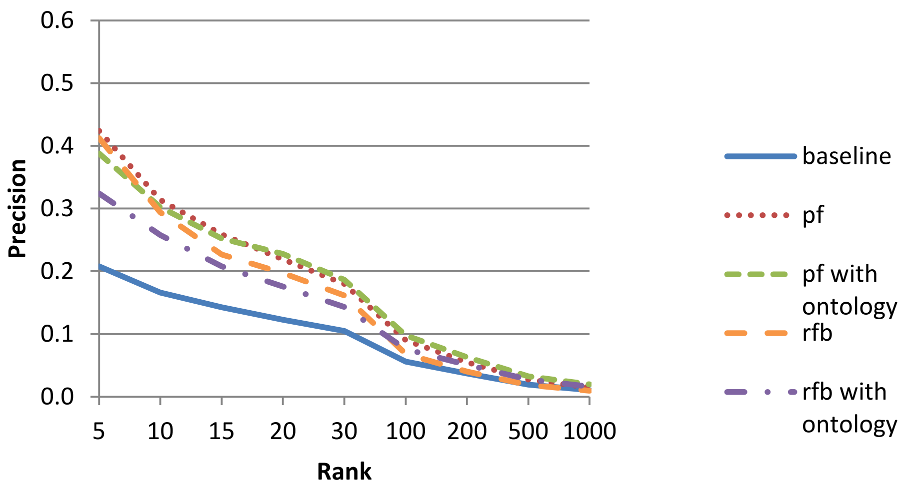


Figure 5.4 Topic 201-250 Retrieval Results – Document Level Averages

Stemming is used to maximise the chance of obtaining a hit in the ontology. However sometimes the stemming process can hinder the performance as the following example shows:

```
Topic no = 223 (responsible great emergence Microsoft computer
industry)
ORIG WORD IS emergenc
     --> explosion
```

Emergence has been stemmed to ***emergenc,*** and ontology picks up non-relevant term ***explosion*** which is more related to emergency than emergence.

In the case of Topics251-300, the ontology has improved the pf run by 7% between ranks 5-10 after which the performance deteriorates as can be seen in Figure 5.5. The rfb run has not benefited from the ontology at all between ranks 5-30 after which the two curves are identical. At rank 5 there is a 19% deterioration.

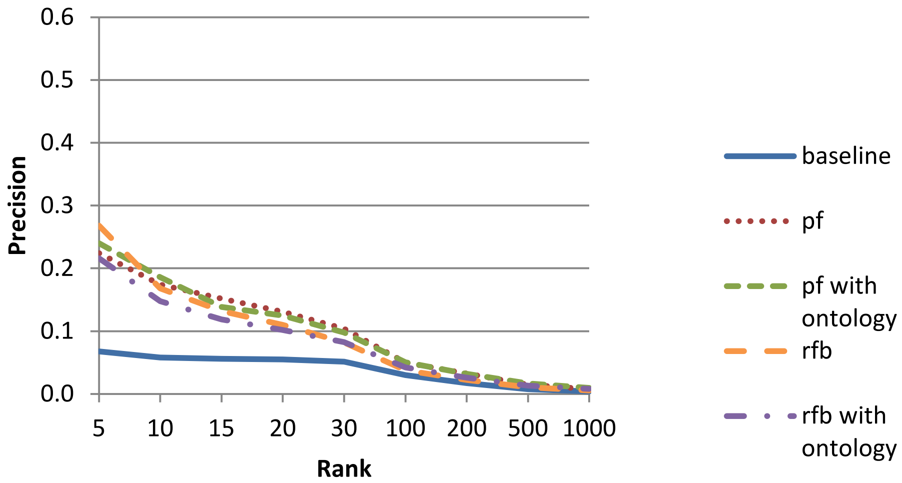


Figure 5.5 Topic 251-300 Retrieval Results – Document Level Averages

5.2.2 Precision-Recall Graphs

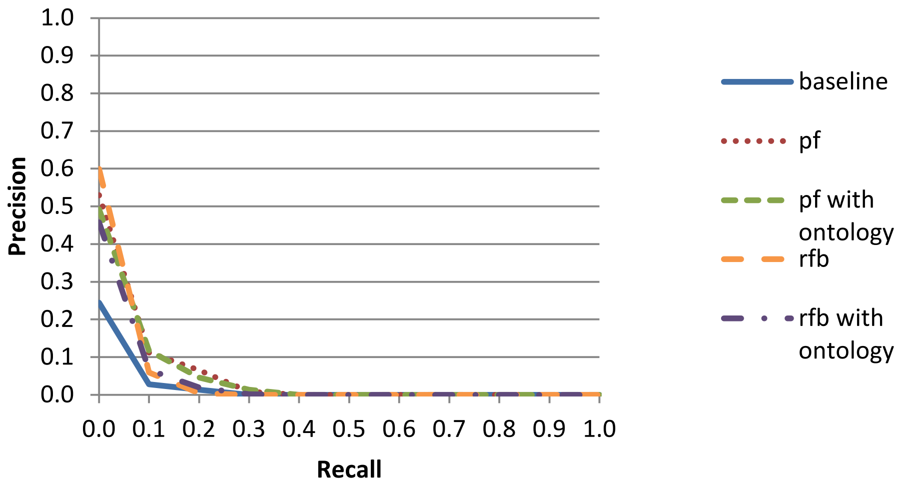


Figure 5.6 Topic 51-100 Retrieval Results – Precision-Recall

The performance of the pf and the rfb curves, with and without ontology is virtually identical for all topics. In figure 5.6, at 0.0 recall, precision is 24% less than rfb baseline whereas pf is 7% less than the pf baseline. The PF t-test results (Table 5.2) are statistically significant for all topic sets except for topic 51-100. The RFB t-test results do not show any statistical significance. So query expansion has not proved to be of any benefit for the rfb runs presented in this section. The results for the other topic sets are quite similar and so are not discussed individually and have been included in Appendix A1.

The ontology has picked up quite a few good terms to enhance recall. For example :

```
TOPIC NUMBER = 88 (crude oil price trends)
current word is oil
     -->  economy_business_finance
```

The topic query is focussing on ***price trends*** so the ontology term ***economy_business_finance*** is related to the topic query.

The success of improving precision-recall depends on the number of documents which contain these terms and the ranked weightings.

Sometimes comes up with ontology terms that are not related to the query. This is shown in the example below where the ontology has interpreted national to be related to defence whereas the topic description states it is more to do with demographics and population movements between different countries:

```
 TOPIC NUMBER = 73 (demographic shifts national boundaries)
current word is national
     -->  defence
```

Table 5.2 shows the t-test results on the document level averages for all topicsets and for both pf and rfb runs. Majority of the results are not significant except for topic 51-100 (pf) and topic 101-150 (rfb) which are significant.

Topics	PF	Rfb
51-100	0.023	0.094
101-150	0.055	0.017
151-200	0.316	0.116
201-250	0.387	0.079
251-300	0.408	0.094

Table 5.2: T-Test results (Document Level Averages)

Table 5.3 shows the t-test results on the document level averages for all topicsets and for both pf and rfb runs. In Table 5.3, for the pf run, the highlighted cells show that topic 101-150, 151-200, 2and 251-300 are significant; topic 201-250 are very significant. For the rfb run, none of the results show statistical significance.

Topics	PF	Rfb
51-100	0.145	0.217
101-150	0.037	0.140
151-200	0.016	0.062
201-250	0.008	0.085
251-300	0.050	0.092

Table 5.3: T-Test results (Precision-Recall)

5.3 Experiments Using Standard 20 Documents And 20 Terms And Top 3 Terms In Expanded Query

5.3.1 Document Level Averages Graphs

For topic 51-100 (Figure 5.7) from rank 5 onwards there is a 21% improvement on pf baseline at rank 5. After rank 5 the performance continues to improve on original pf baseline. For rfb at rank 5 there is a 19% deterioration on the rfb baseline and improvement takes place from rank 15 onwards.The pf t-test results are very significant but the rfb t-test results are not significant.

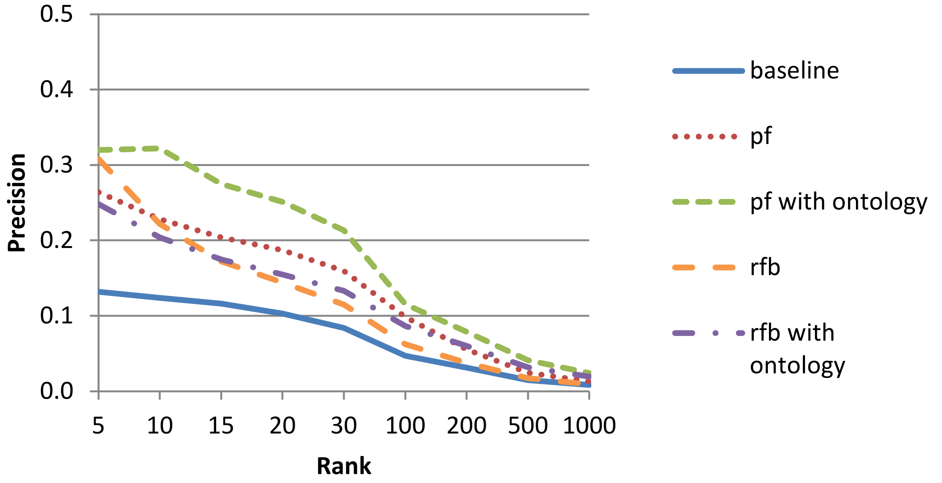


Figure 5.7 Topic 51-100 Results – Document Level Averages

Compared to the pf baseline in the standard run, the pf baseline in the top3 run has been degraded and there is no improvement on the pf curve with ontology from the standard run to the top3 run. For example, at rank 5, the pf baseline in the standard run is 0.316; the pf baseline in the top3 run is 0.264 and the pf with ontology curves for the standard run and the top 3 run are 0.320. Similarly, the rfb baseline in the top 3 run has increased only at rank 5 after which it is degraded and the rfb curve with ontology is unchanged. Looking at rank 5 again, the rfb baseline for the standard run is 0.304; the rfb baseline for the top3 run is 0.308 and the rfb with ontology curve for both runs is 0.248. A full set of the TREC output figures for the standard run and the top3 run can be seen in Appendix F (F1 and F2 respectively).

Appendix C shows that the 5 extra ontology terms picked up compared to the standard run (Appendix B). The 5 extra ontology terms are: ***weather, civil, health, conflict and scienc***. Some of the ontology child terms are relevant to the ontology term but not directly related to the topic. For example:

```
Topic no = 59 (weather related fatalities)
ORIG WORD IS weather
     --> forecast
     --> global_change
     --> report
     --> statistic
     --> warning
```

None of these ontology terms are directly related to topic 59.

In other cases, the child terms are relevant to the ontology search term and the topic statement as the following example shows:

```
Topic no = 67 (politically motivated civil disturbances)
ORIG WORD IS civil
     --> revolutions
     --> rebellions
     --> political_dissent
     --> religious_conflict
     --> social_conflict
     --> protest
```

In other cases, the child terms appear to be related but we cannot be certain because the topic statement is not specific enough as the following example shows:

```
Topic no = 74 (conflicting policy)
ORIG WORD IS conflict
     --> peacekeeping_force
```

The TIPSTER topic description just referes to documents that “cite an instance in which the U.S. government propounds two conflicting or opposing policies”. This statement is quite broad so we cannot guarantee that ***peacekeeping_force*** is directly related to the topic.

For topic 101-150 (Figure 5.8) at rank the two curves are identical and from rank 15 the performance continues to improve on original pf baseline. For rfb at rank 5 there is a 17% deterioration on the rfb baseline and improvement takes place from rank 30 onwards. The pf t-test results are very significant but the rfb t-test results are not significant.

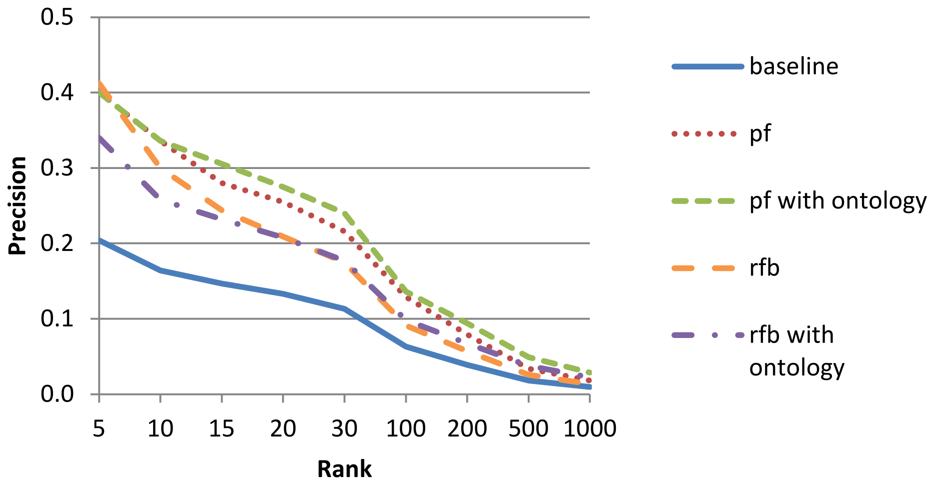


Figure 5.8 Topic 101-150 Results – Document Level Averages

Compared to the pf baseline in the standard run, the pf baseline in the top3 run (topic 101-150) has been degraded from rank 5 onwards and there is no improvement on the pf curve with ontology from the standard run to the top3 run. For example, at rank 5, the pf baseline in the standard run is 0.388; the pf baseline in the top3 run is 0.400 and the pf with ontology curves for the standard run and the top 3 run are 0.400. The rfb baseline in the top 3 run is worse and the rfb curve with ontology is unchanged. Looking at rank 5 again, the rfb baseline for the standard run is 0.488; the rfb baseline for the top3 run is 0.412 and the rfb with ontology curve for both runs is 0.340. A full set of the TREC output figures for the standard run and the top3 run can be seen in Appendix F (F1 and F2 respectively).

The top3 run (topic 101-150) did not find any extra ontology terms compared to the standard run (topic 101-150).

For topic 151-200 (Figure 5.9) from rank 5 onwards there is a 3% improvement on pf baseline at rank 5. After rank 5 the performance continues to improve on original pf baseline. For rfb at rank 5 there is a 16% deterioration on the rfb baseline and improvement takes place from rank 15 onwards. The pf t-test results are very significant but the rfb t-test results are not significant.

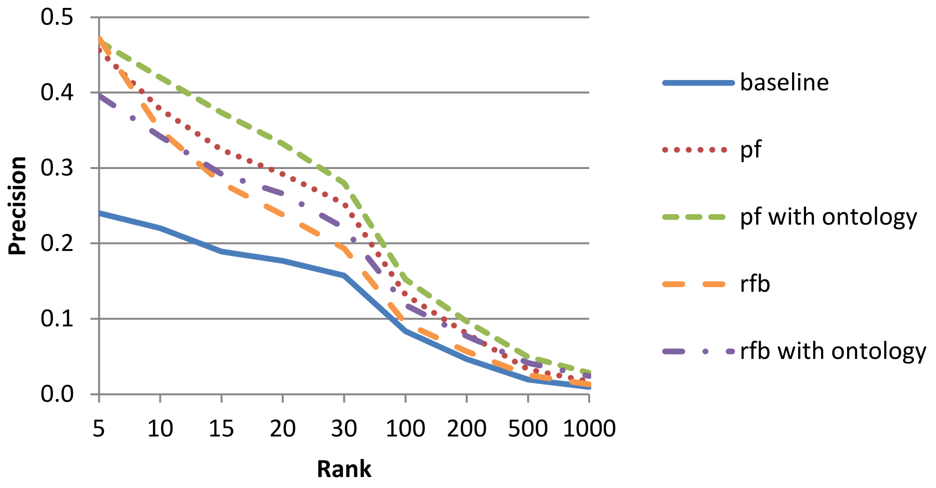


Figure 5.9 Topic 151-200 Results – Document Level Averages

Compared to the pf baseline in the standard run, the pf baseline in the top3 run (topic 151-200) has been degraded from rank 5 onwards and there is no improvement on the pf curve with ontology from the standard run to the top3 run. For example, at rank 5, the pf baseline in the standard run is 0.468; the pf baseline in the top3 run is 0.456 and the pf with ontology curves for the standard run and the top 3 run are 0.468. The rfb baseline in the top 3 run is worse and the rfb curve with ontology is unchanged. Looking at rank 5 again, the rfb baseline for the standard run is 0.504; the rfb baseline for the top3 run is 0.472 and the rfb with ontology curve for both runs is 0.396. A full set of the TREC output figures for the standard run and the top3 run can be seen in Appendix F (F1 and F2 respectively).

The top3 run (topic 151-200) only found 2 extra ontology terms compared to the standard run (topic 151-200), namely ***defenc and health***.

```
Topic no = 152 (accusations cheating contractors us defense
projects)
ORIG WORD IS defenc
     --> veterans_affairs
     --> national_security
     --> security_measures
     --> troops_withdrawal
     --> armed_forces
     --> military_equipment
     --> firearms
     --> biological_chemical_weapons
     --> missile_systems
     --> nuclear_weapons
```

```
Topic no = 153 (insurance coverage pays term care)
ORIG WORD IS health
     --> disease
     --> epidemic_plague
     --> health_treatment
     --> prescription_drugs
     --> medical_procedure
     --> therapy
     --> health_org
     --> medical_research
     --> medical_staff
     --> medicine
     --> preventative_medicine
     --> injury
     --> hospital
     --> clinic
     --> illness
```

In both of these examples, the child terms do not appear to be directly relevant to each the topic statements.

For topic 201-250 (Figure 5.10) from rank 5 onwards there is a 11% improvement on pf baseline at rank 5. After rank 5 the performance continues to improve on original pf baseline. For rfb at rank 5 there is a 5% deterioration on the rfb baseline and improvement takes place from rank 10 onwards. The pf t-test results are very significant but the rfb t-test results are significant.

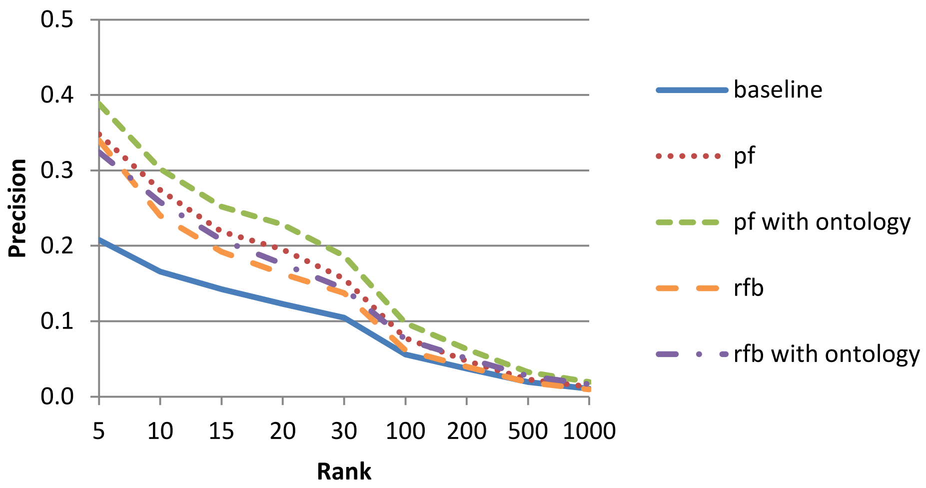


Figure 5.10 Topic 201-250 Results – Document Level Averages

Compared to the pf baseline in the standard run, the pf baseline in the top3 run (topic 201-250) has been degraded and there is no improvement on the pf curve with ontology from the standard run to the top3 run. For example, at rank 5, the pf baseline in the standard run is 0.424; the pf baseline in the top3 run is 0.348 and the pf with ontology curves for the standard run and the top 3 run are 0388. The rfb baseline in the top 3 run is worse and the rfb curve with ontology is unchanged. Looking at rank 5 again, the rfb baseline for the standard run is 0.412; the rfb baseline for the top3 run is 0.340 and the rfb with ontology curve for both runs is 0.324. A full set of the TREC output figures for the standard run and the top3 run can be seen in Appendix F (F1 and F2 respectively).

The top3 run (topic 201-250) did not find any extra ontology terms compared to the standard run (topic 201-250).

For topic 251-300 (Figure 5.11) from rank 5 onwards there is a 18% improvement on pf baseline at rank 5. After rank 5 the performance continues to improve on original pf baseline. For rfb at rank 5 there is a 4% improvement on the rfb baseline. After rank 5 the performance continues to improve on original pf baseline. The pf t-test results are significant but the rfb t-test results are very significant.

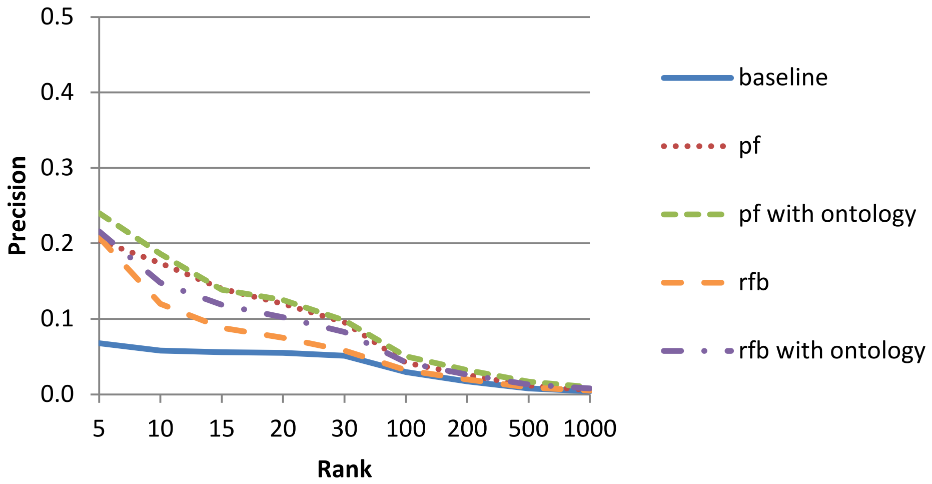


Figure 5.11 Topic 251-300 Results – Document Level Averages

Compared to the pf baseline in the standard run, the pf baseline in the top3 run (topic 251-300) has been degraded and there is no improvement on the pf curve with ontology from the standard run to the top3 run. For example, at rank 5, the pf baseline in the standard run is 0.224; the pf baseline in the top3 run is 0.204 and the pf with ontology curves for the standard run and the top 3 run are 0240. The rfb baseline in the top 3 run is worse and the rfb curve with ontology is unchanged. Looking at rank 5 again, the rfb baseline for the standard run is 0.268; the rfb baseline for the top3 run is 0.208 and the rfb with ontology curve for both runs is 0.216. A full set of the TREC output figures for the standard run and the top3 run can be seen in Appendix F (F1 and F2) respectively.

The top3 run (topic 251-300) did not find any extra ontology terms compared to the standard run (topic 251-300).

5.3.2 Precision-Recall Graphs

At recall 0.0 (figure 5.12), the two pf curves are identical and a 23% deterioration on the rfb baseline. Between recall 0.1 and 0.2 there is an improvement after which the two curves are identical. The pf t-test results for are significant. There is an improvement for the rfb curve from recall 0.1 onwards and from recall 0.3 the two curves are the same. The rfb t-test score for this topic set is not significant. So query expansion has not proved to be of any benefit for the pf or the rfb runs presented in this section. The Precision-Recall graphs for the other topic sets are all very similar except for topic 251-300 where the pf with ontology is better than the pf baseline. These graphs can be found in Appendix G.

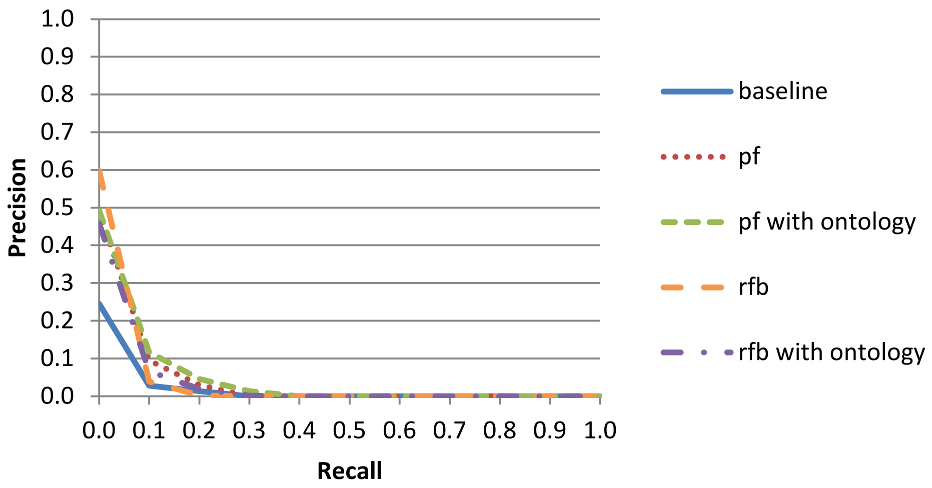


Figure 5.12 Topic 51-100 Results – Precision-Recall

At recall 0.0 (figure 5.13), the two pf curves are identical and a 17% deterioration on the rfb baseline. Between recall 0.1 and 0.2 there is an improvement after which the two curves are identical. The pf t-test results for are not significant. There is an improvement for the rfb curve from recall 0.1 onwards and from recall 0.3 the two curves are the same. The rfb t-test score for this topic set is not significant.

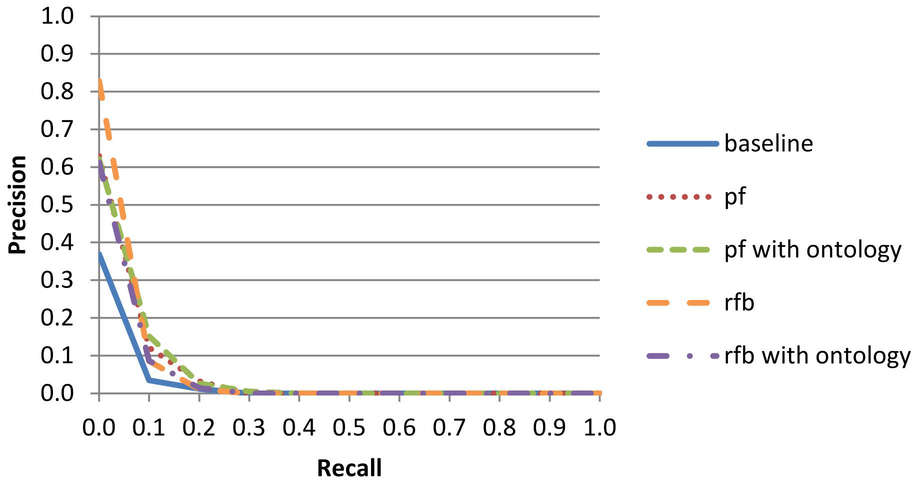


Figure 5.13 Topic 101-150 Results – Precision-Recall

At recall 0.0 (figure 5.14), there is a 4% deterioration on the pf baseline and a 24% deterioration on the rfb baseline. Between recall 0.1 and 0.2 there is an improvement after which the performance declines. From recall 0.8 onwards, the two curves are identical. The pf t-test results for are not significant. There is an improvement for the rfb curve at recall 0.1, at recall .2 a drop in performance, at 0.3 to 0.5 an improvement then from 0.6 – 0.8 a drop, and then from 0.9 onwards the two curves are the same. The rfb t-test score for this topic set is not significant.

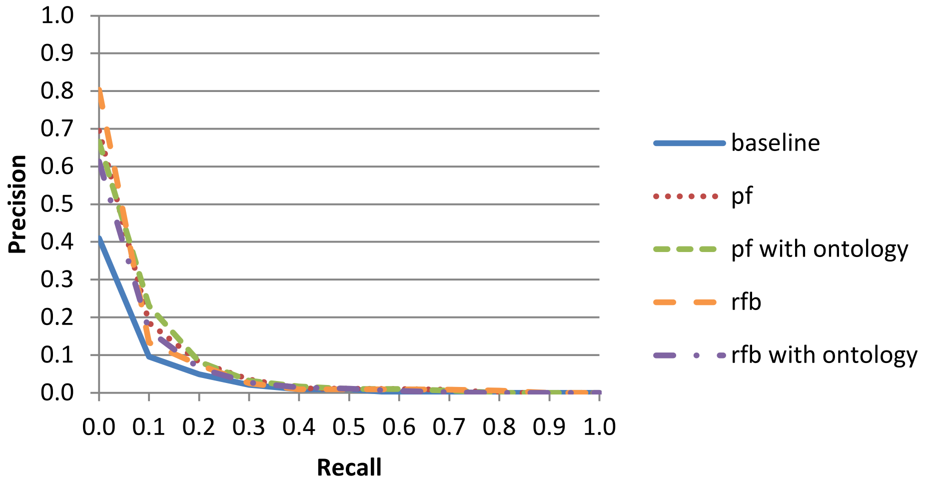


Figure 5.14 Topic 151-200 Results – Precision-Recall

At recall 0.0 (figure 5.15), there is a 6% improvement on the pf baseline and a 17% deterioration on the rfb baseline. The improvement continues until recall 0.5 after which the two curves are identical. The pf t-test results for are significant. There is an improvement for the rfb curve at recall 0.1 – 0.2, then recall 0.3-0.5 a drop in performance, and then from 0.6 onwards the two curves are the same. The rfb t-test score for this topic set is not significant.

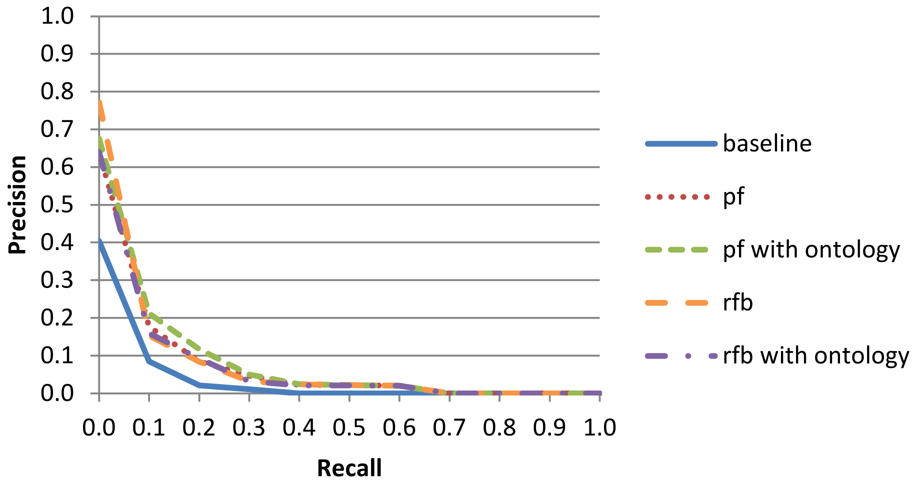


Figure 5.15 Topic 201-250 Results – Precision-Recall

At recall 0.0 (figure 5.16), there is a 3% improvement on the pf baseline and a 18% deterioration on the rfb baseline. The biggest improvements start from recall 0.6 onwards. The pf t-test results are not significant. There is an improvement for the rfb curve at recall 0.1 – 0.3, and then from 0.4 onwards the two curves are the same. The rfb t-test score for this topic set is not significant.

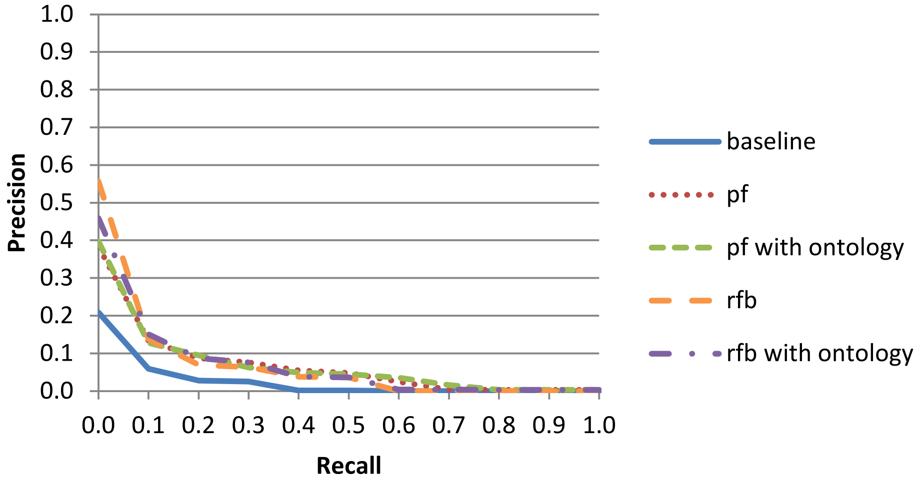


Figure 5.16 Topic 251-300 Results – Precision-Recall

In Table 5.4, for the pf run, the highlighted cells show that topic 251-300 is significant and all remaining topics are very significant. For the rfb run, topic 201-250 is significant and topic 251-300 is very significant.

Topics	PF	Rfb
51-100	0.001	0.383
101-150	0.002	0.172
151-200	0.000	0.293
201-250	0.000	0.017
251-300	0.022	0.002

Table 5.4: T-Test results (Document Level Averages)

In Table 5.5, for the pf run, the highlighted cells show that topic 51-100 and topic 201-250 are significant. For the rfb run, none of the topics are significant.

Topics	PF	Rfb
51-100	0.045	0.271
101-150	0.269	0.176
151-200	0.388	0.199
201-250	0.032	0.190
251-300	0.299	0.374

Table 5.5: T-Test results (Precision-Recall)

5.4 Summary

In this chapter we have presented the results for the standard run which uses 20 documents and 20 terms for relevance feedback. We have also presented the results for the run where only the top 3 query expansion terms have been used.

Firstly we found that for all of the topic sets, the top 3 run retrieves a higher number of relevant documents compared to the standard run. (see Chapter 3 - Table 3.3).

When looking at the precision at rank results, only topic 51-100 found 5 extra ontology hits but this has not improved the pf results instead the rfb results have improved at rank 5 after which there are no further improvements. The remaining topic sets did not find any extra ontology terms compared to the standard run. Apart from a slight improvement on the pf baseline at rank 5 (topic 101-150) the remaining pf baselines and rfb baselines are worse than the standard run. The pf with ontology and the rfb with ontology curves have not improved, they are identical to the standard run. All topic sets have precision in the range 0.3 to 0.5 except for topic 251-300 which has lower pre.cision at 02.

For the precision-recall results, the top3 runs were worse than the standard run.

To conclude, any improvements/degradations have occurred on the pf/rfb baselines. The ontology has resulted in an improvement for pf with ontology and rfb with ontology curves in the standard run, but the use of the ontology for the top3 run has not resulted in any further improvements. This supports Jones et al, (1995) who found no correspondence between the number of terms chosen and the query performance.

6. EXPERIMENT RESULTS FOR VARYING NUMBER OF TERMS RELEVANCE FEEDBACK PARAMETER

6.1 Introduction

In this chapter we present the results for experiments 2-6 where the number of terms parameter is set to 5, 10, 15, 100 and 200 respectively. We can see from table 6.1 that ontology based query expansion has for most cases resulted in a large increase in the number of relevant documents retrieved.

#terms parameter	Topics	+/- #docs retrieved (Pf)	+/-#docs retrieved (Rfb)
5	51-100	407	460
5	101-150	406	348
5	151-200	452	529
5	201-250	257	345
5	251-300	130	138
10	51-100	298	426
10	101-150	193	270
10	151-200	475	439
10	201-250	195	335
10	251-300	162	140
15	51-100	271	403
15	101-150	104	288
15	151-200	391	414
15	201-250	163	345
15	251-300	129	123
standard	51-100	161	393
standard	101-150	103	252
standard	151-200	354	376
standard	201-250	249	323
standard	251-300	91	116
100	51-100	81	268
100	101-150	-37	198
100	151-200	287	339
100	201-250	185	247
100	251-300	176	33
200	51-100	723	355
200	101-150	847	485
200	151-200	870	556
200	201-250	415	200
200	251-300	267	104

Table 6.1 Difference in the number of documents retrieved for pf and rfb runs (varying number of terms parameter)

Sections 6.2 – 6.6 will examine the effects of varying the number of terms parameter used for relevance feedback**.**

6.2 Experiments Using 20 Documents And 5 Terms And All Terms In Expanded Query

6.2.1 Document Level Averages Graphs

For topic 51-100 (Figure 6.1) the use of the ontology has resulted in a substantial improvement on the pf baseline. At rank 10, there is a 26% improvement and an overall average improvement of 27%. The rfb curve with the ontology improves on the rfb baseline for rank 30, 100, 200, 500 and 1000 by 2%, 23%, 43%, 67% and 91% respectively. Compared to the other graphs in this run, the pf with ontology curve performs much better than the pf curve. An explanation for this is that this is not considered to be a hard topic. The topic hardness measure is oriented towards high-recall performance (Buckley 1996). For example topic 75 ***automation*** is a prime candidate for increasing recall because it is such a broad term. Table 5.3 shows the pf results to be very significant but the rfb results for this topicset are not significant.

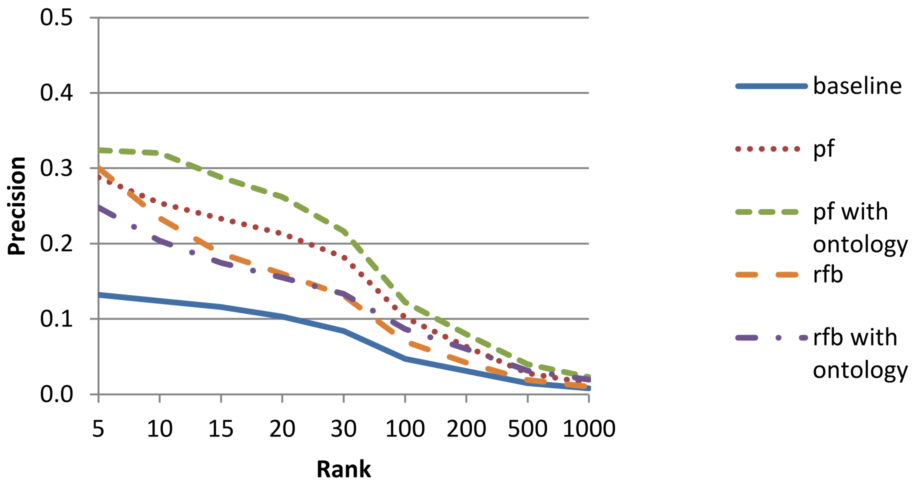


Figure 6.1 Topic 51-100 Results – Document Level Averages

Compared to the pf baseline in the standard run, the pf baseline in the 5 term run (topic 51-100) has been degraded however there is an improvement on the pf curve with ontology from the standard run to the 5 terms run. For example, at rank 5, the pf baseline in the standard run is 0.316; the pf baseline in the 5 term run is 0.288 and the pf with ontology curves for the standard run and the 5 term run are 0.320 and 0.324 respectively. The rfb baseline in the 5 term run is worse and the rfb curve with ontology is unchanged. Looking at rank 5 again, the rfb baseline for the standard run is 0.304; the rfb baseline for the 5 term run is 0.300 and the rfb with ontology curve for both runs is 0.248. A full set of the TREC output figures for the standard run and the 5 term run can be seen in Appendix F (F1 and F3 respectively).

The 5 term run (topic 51-100) did not find all of the terms from the standard run. The standard run had ontology hits ***for state, natural, employment, crime and un***. Whereas the 5 term run had ontology hits for ***state, natural, crime, weather, civil, health and conflict***. Section 5.2.1 has already explained the relevance of terms ***state, natural and civil***; the non-relevant terms have been explained in section 5.2.1 (***health***) and

5.3.1(***weather***). The following example is related to the topic and result in improving the results for precision and recall:

```
Topic no = 94 (computeraided crime)
ORIG WORD IS crime
     --> crime
     --> murder
     --> computer_crime
     --> theft
     --> drug_trafficking
     --> sexual_assault
     --> assault
     --> torture
     --> kidnapping
     --> arson
     --> gang_activity
     --> criminal
     --> murderer
     --> offender
     --> accused
     --> crime_victim
     --> stolen
     --> judiciary
     --> lawyer
     --> police
     --> investigation
     --> punishment
     --> prison
     --> laws
     --> justice_rights
     --> trials
     --> organized_crime
     --> international_law
     --> corporate_crime
     --> war_crime
```

Computer_crime is strongly related to the topic ***computeraided crime***.

In the case of un and employment, these terms are not related to the topic statements so the ontology terms retrieved will also not be related as shown below:

```
Topic no = 95 (computeraided crime detection)
ORIG WORD IS un
     --> oposing_group
     --> truce
     --> armed_conflict
     --> civil_unrest
     --> coup_detat
     --> terrorism
     --> massacre
     --> riots
     --> demonstration
     --> turf_war
     --> war
     --> conflict
     --> crisis
     --> weaponry
     --> bombings
     --> invasion
     --> war_victim
```

Un is not related to topic 95 so it would have a degrading effect on the results.

```
Topic no = 92 (international military equipment sales)
ORIG WORD IS employment
     --> labor_market
     --> job_layoffs
     --> child_labour
     --> occupations
```

The TIPSTER topic statement for topic 92 refers to documents that “identify a proposed or recently concluded sale of military equipment in the international arms market”. It is clear that neither the selected ontology term or its child terms are related to the topic. So this would have a degrading effect on precision and recall.

For topics101-150 (figure 6.2), the ontology does not produce any improvements for pf. At rank 5 there is a 7% deterioration for the pf curves and a 24% deterioration for the rfb curves. For topics 101-150 the ontology seems to benefit the pf and rfb from rank 100 onwards. This is the only graph where the rfb curve performs higher than the other curves between ranks 5-10. This shows that the ranking algorithm is working because the top 10 documents contain the most relevant terms. Only the results for rfb are significant (Table 5.3).

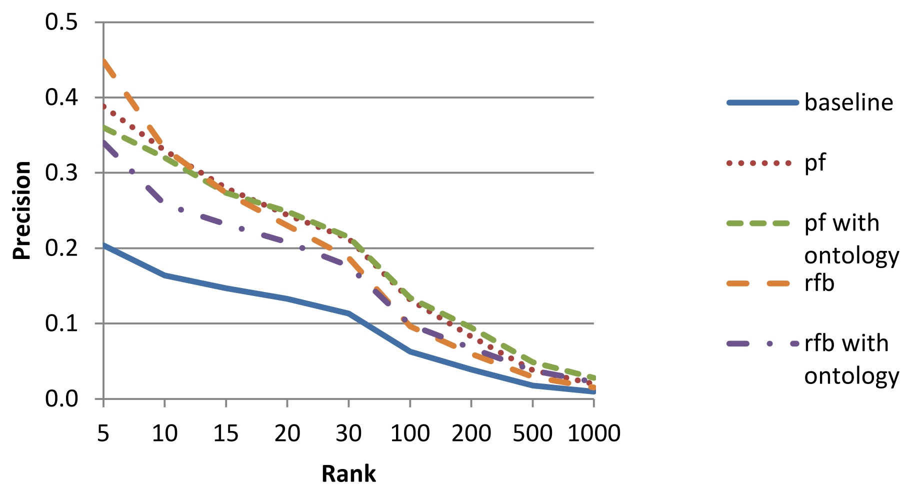


Figure 6.2 Topic 101-150 Results – Document Level Averages

Compared to the pf baseline in the standard run, the pf baseline and the pf with ontology curve in the 5 term run (topic 101-150) have been degraded The rfb baseline in the 5 term run is worse and the rfb curve with ontology is unchanged. A full set of the TREC output figures for the standard run and the 5 term run can be seen in Appendix F (F1 and F3 respectively). The 5 term run (topic 101-150) has fewer ontology hits compared to the standard run.

For topics151-200, figure 6.3, the ontology does not produce any improvements for pf . At rank 5 there is a 20% deterioration for the pf and the rfb curves. For topics 151-200 the ontology seems to benefit the pfb from rank 100 onwards however the improvement for rfb occurs higher up the rankings from rank 10 onwards. Only the results for pfb are significant.

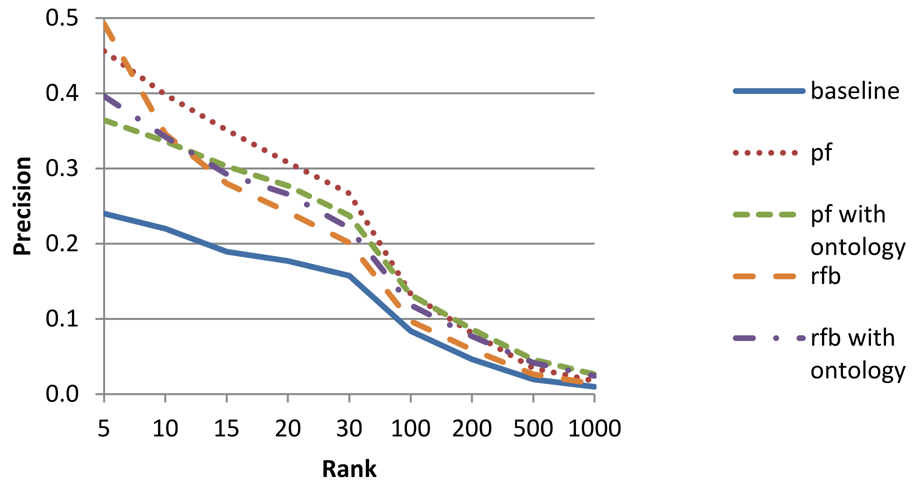


Figure 6.3 Topic 151-200 Results - Document Level Averages

Compared to the pf baseline in the standard run, the pf baseline and the pf with ontology curve in the 5 term run (topic 151-200) have been degraded The rfb baseline in the 5 term run is worse and the rfb curve with ontology is unchanged. A full set of the TREC output figures for the standard run and the 5 term run can be seen in Appendix F (F1 and F3 respectively). The 5 term run (topic 151-200) has fewer ontology hits compared to the standard run.

For topics201-250 (figure 6.4), the ontology does not produce any improvements for pf. At rank 5 there is a 12% deterioration and for rfb at rank 5 there is a 7% deterioration. The ontology has not benefited the pfb however the improvement for rfb occurs much higher up from rank 10 onwards. Only the pfb results show significance.

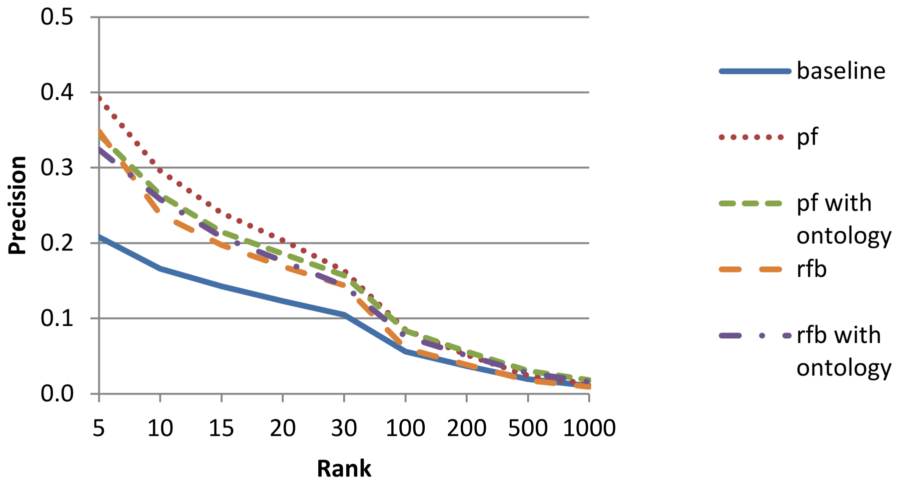


Figure 6.4 Topic 201-250 Results – Document Level Averages

Compared to the pf baseline in the standard run, the pf baseline and the pf with ontology curve in the 5 term run (topic 201-250) have been degraded The rfb baseline in the 5 term run is worse and the rfb curve with ontology is unchanged. A full set of the TREC output figures for the standard run and the 5 term run can be seen in Appendix F (F1 and F3 respectively). The 5 term run (topic 201-250) has fewer ontology hits compared to the standard run.

For topics251-300 (figure 6.5), the ontology does produce slight improvements for pf. At rank 5 there is a 9% improvement on the pf baseline but for rfb at rank 5 there is a 10% deterioration. The pfb results are very significant. The improvement for rfb occurs much higher up from rank 10 onwards and table 5.3 shows the rfb results to be significant.

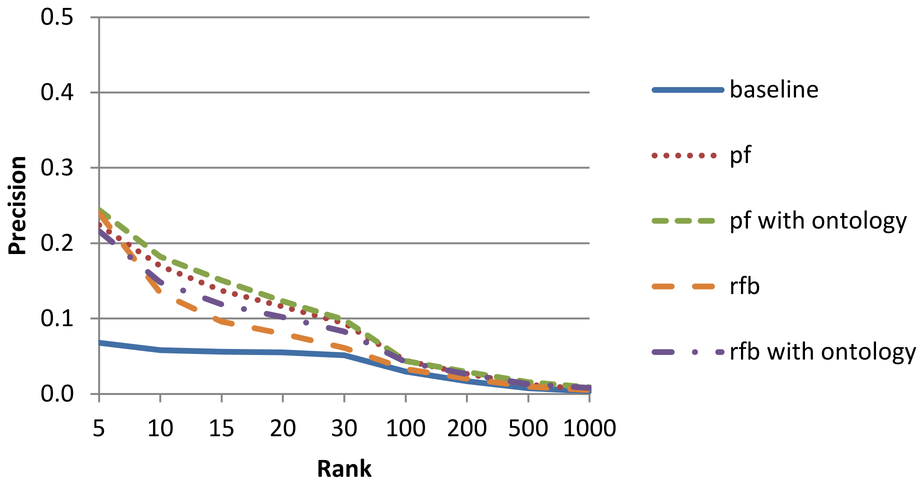


Figure 6.5 Topic 251-300 Results – Document Level Averages

Compared to the pf baseline in the standard run, the pf baseline has been degraded but the pf with ontology curve in the 5 term run (topic 251-300) improves at ranks 5 and 15. The rfb baseline in the 5 term run is worse and the rfb curve with ontology is unchanged. A full set of the TREC output figures for the standard run and the 5 term run can be seen in Appendix F (F1 and F3 respectively). The 5 term run (topic 251-300) has fewer ontology hits compared to the standard run.

6.2.2 Precision-Recall Graphs

The performance of the pf and the rfb curves, with and without ontology is virtually identical for all topics. At 0.0 recall (figure .6), precision is 23% less than rfb baseline whereas pf is 4% less than the pf baseline. So query expansion has not proved to be of any benefit for the rfb runs presented in this section. The results for the other topic sets are quite similar and so are not discussed individually and have been included in Appendix A2.

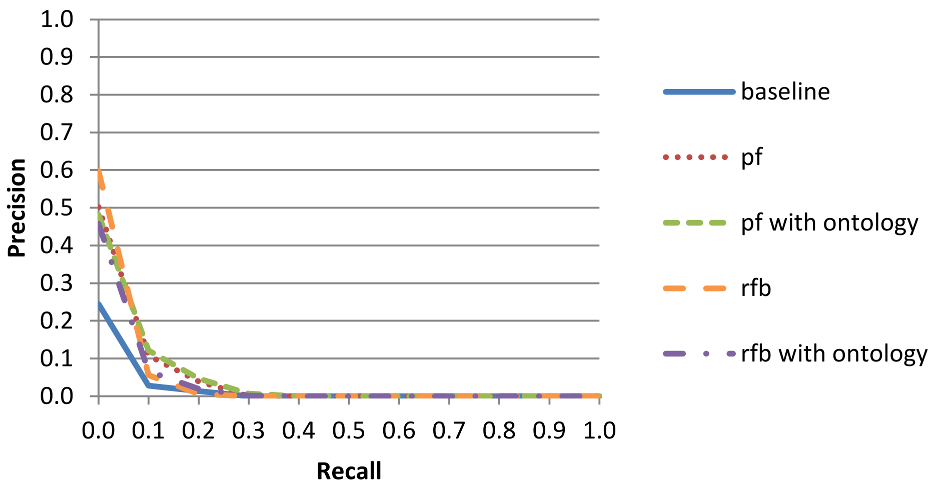


Figure 6.6 Topic 51-100 Results – Precision-Recall

At 0.0 recall (figure 6.7), precision is 27% less than rfb baseline whereas pf is 15% less than the pf baseline. At rank 0.1, pf improves by 29% and rfb improves by 7% after which the performance deteriorates again. As can be seen from Table 5.4, none of the results are of statistical significance.

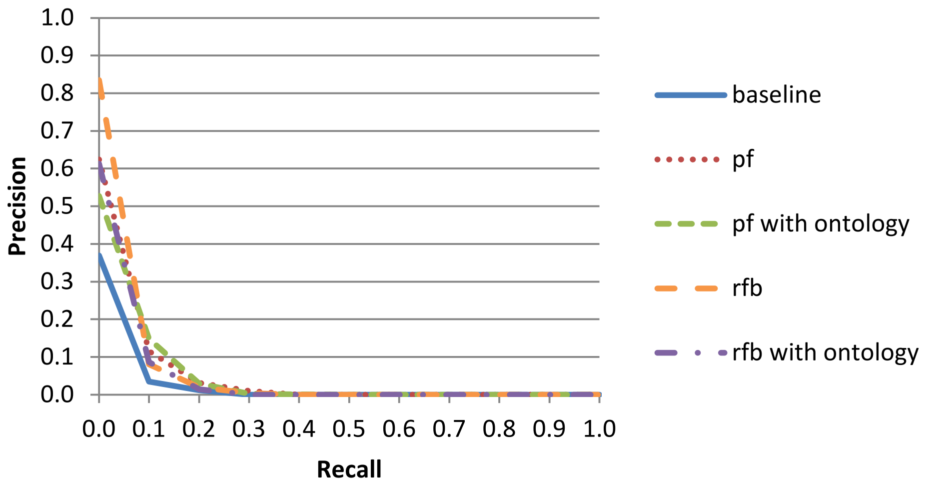


Figure 6.7 Topic 101-150 Results – Precision-Recall

At 0.0 recall (figure 6.8), precision is 25% less than rfb baseline whereas pf is 17% less than the pf baseline. So query expansion has not proved to be of any benefit for either pf or rfb runs.

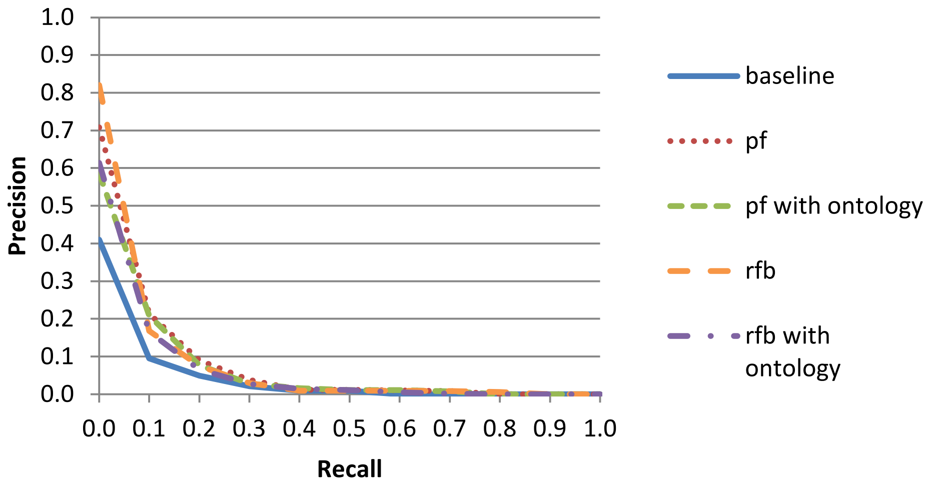


Figure 6.8 Topic 151-200 Results – Precision-Recall

The results for the other topic sets are quite similar and so are not discussed individually and have been included in Appendix A3.

Table 6.2 shows the t-test results on the document level averages for all topicsets and for both pf and rfb runs. The highlighted cells show that topic 51-100 and topic 251-300 are very significant; topic 151-200 and topic 201-250 are significant. Only topic 101-150 is not significant. For the rfb run, only topic 101-150 is significant, the remaining topics are not significant.

Topics	PF	Rfb
51-100	0.001	0.290
101-150	0.449	0.050
151-200	0.029	0.430
201-250	0.042	0.085
251-300	0.005	0.057

Table 6.2: T-Test results (Document Level Averages)

Table 6.3 shows the Precision-Recall t-test results for all topicsets and for both pf and rfb runs. In Table 5.4, none of the results show statistically significance.

Topics	PF	Rfb
51-100	0.394	0.223
101-150	0.254	0.167
151-200	0.123	0.143
201-250	0.057	0.262
251-300	0.261	0.194

Table 6.3: T-Test results (Precision-Recall)

6.3 Experiments Using Standard 20 Documents And 10 Terms And All Terms In Expanded Query

6.3.1 Document Level Averages Graphs

For topic 51-100 (Figure 6.9) the use of the ontology has resulted in a very slight improvement on the pf baseline. Compared to other graphs in this run, the pf and pf with ontology curve perform much better. This is a similar situation to that of the experiment using 5 terms for relevance feedback (see 6.2.1). If other retrieval systems found it difficult to retrieve the relevant documents on this topic then it would be considered to be hard (Buckley 1996). However, other systems did not find it difficult to achieve high recall performance for this topic set. Another example from this topic set for enhancing peformance is topic 85 ***official corruption*** which is a broad topic and also is a topical one in the news industry. The rfb curve with the ontology improves on the rfb baseline from rank 100 onwards. At rank 5 there is a 2% deterioration on the pf baseline but for rfb at rank 5 there is a 16% deterioration. The t-test result for this topicset did not show statistical significance.

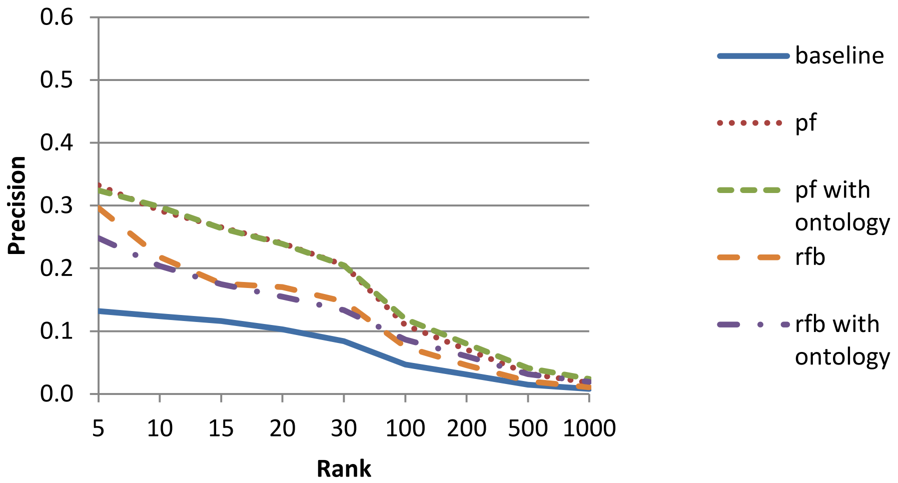


Figure 6.9 Topic 51-100 Results – Document Level Averages

Compared to the pf baseline in the standard run, the pf baseline has improved but the pf with ontology curve in the 10 term run (topic 51-100) improves at rank 5 after which the performance deteriorates. The rfb baseline in the 10 term run is worse and the rfb curve with ontology is unchanged. A full set of the TREC output figures for the standard run and the 10 term run can be seen in Appendix F (F1 and F4 respectively). The 10 term run (topic 51-100) has the same ontology terms as the standard run except for employment. Instead the 10 term run the same terms as the 5 term run and in addition has the term ***scienc***.

```
Topic no = 74 (conflicting policy)
ORIG WORD IS scienc
     --> applied_science
     --> engineering
     --> natural_science
     --> research
     --> scientific_exploration
     --> space_programme
     --> standards
     --> mathematics
     --> biotechnology
     --> agricultural_research_technology
     --> nanotechnology
     --> IT_computer_science
     --> scientific_institutions
```

The term ***scienc*** is not related to the topic statement so any child nodes retrieved will not be relevant unless the policy is research-based or educational.

For topic 101-150 (Figure 6.10) the use of the ontology has not resulted in any significant improvements on the pf or rfb baseline. At rank 5 there is a 13% deterioration on the pf baseline but for rfb at rank 5 there is a 30% deterioration. In Table 5.5 we can see that the PF run results were of statistical significance for topics101-150. The rfb were very significant.

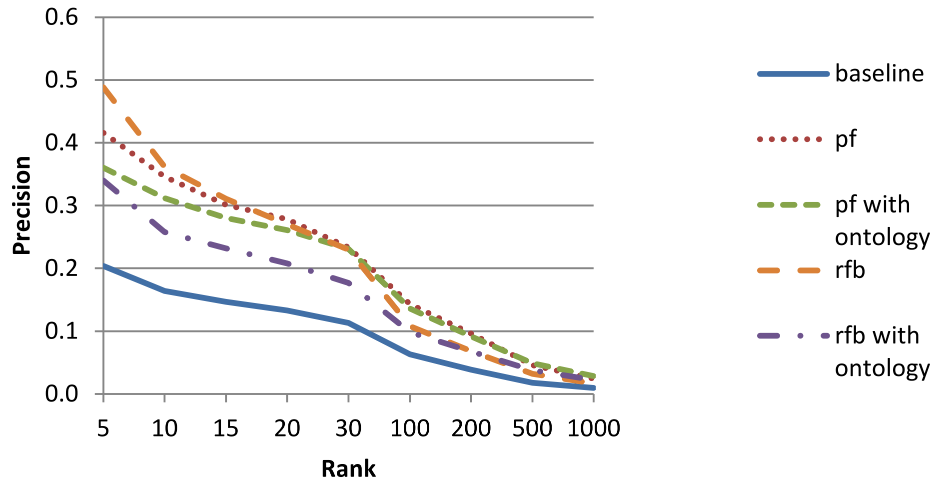


Figure 6.10 Topic 101-150 Results – Document Level Averages

Compared to the pf baseline in the standard run, the pf baseline has improved but the pf with ontology curve in the 10 term run (topic 101-150) is degraded. The rfb baseline in the 10 term run is improved but the rfb curve with ontology is unchanged. A full set of the TREC output figures for the standard run and the 10 term run can be seen in Appendix F (F1 and F4 respectively). The 10 term run (topic 101-150) has the same ontology terms as the standard run except for employment. Instead the 10 term run the same terms as the 5 term run and in addition has the terms natural, punishment and diseas..

For topic 151-200 (Figure 6.11) the use of the ontology has resulted in a slight improvement on the pf and the rfb baseline from rank 10 and rank 30 onwards . At rank 5 there is a 8% deterioration on the pf baseline but for rfb at rank 5 there is a 22% deterioration. The pf and rfb t-test results for this topic set did not show statistical significance.

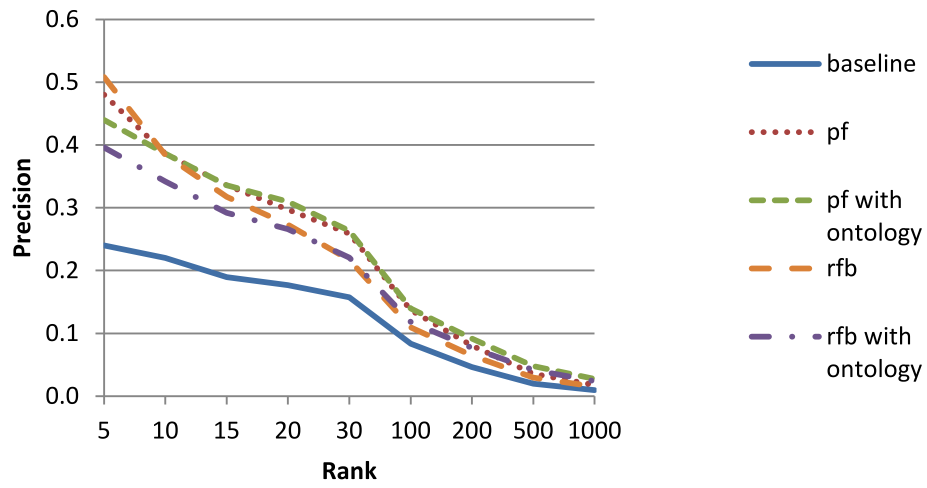


Figure 6.11 Topic 151-200 Results – Document Level Averages

Compared to the pf baseline in the standard run, the pf baseline has improved at rank 5 only, but the pf with ontology curve in the 10 term run (topic 151-200) is degraded. The rfb baseline in the 10 term run is improved but the rfb curve with ontology is unchanged. A full set of the TREC output figures for the standard run and the 10 term run can be seen in Appendix F (F1 and F4 respectively). The 10 term run (topic 151-200) has some terms the same as the standard run (ap, foreign, un, trial) and some terms the same as 5 term run (defenc and corporat).

For topic 201-250 (Figure 6.12) the use of the ontology has resulted in a slight improvement on the pf and the rfb baseline from rank 100 onwards . At rank 5 there is a 9% deterioration on the pf baseline and for rfb at rank 5 there is a 16% deterioration. The pf and rfb t-test results for this topic set did not show statistical significance.

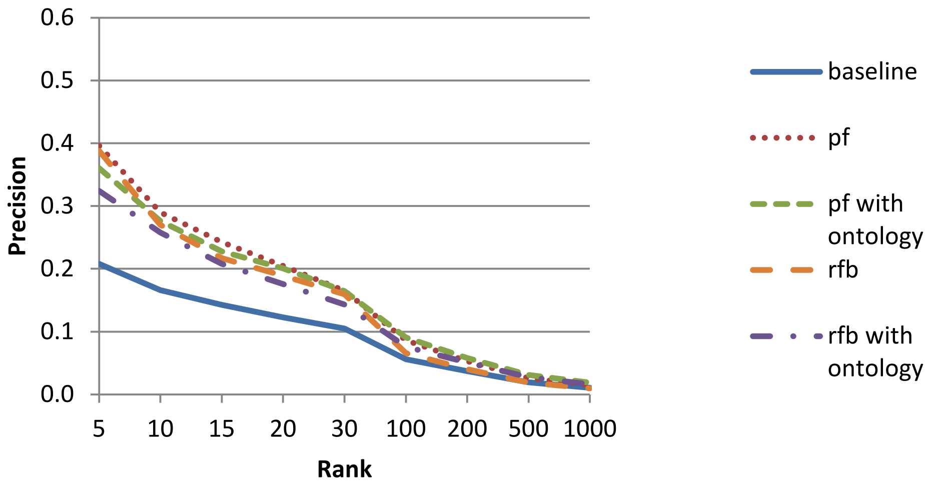


Figure 6.12 Topic 201-250 Results – Document Level Averages

Compared to the pf baseline in the standard run, the pf baseline and the pf with ontology curve in the 10 term run (topic 201-250) is degraded. The rfb baseline has also deteriorated and the rfb curve with ontology is unchanged. A full set of the TREC output figures for the standard run and the 10 term run can be seen in Appendix F (F1 and F4 respectively). The 10 term run (topic 201-250) has fewer terms than the standard run, eg it does not have election, lab, ap and applied.

For topic 251-300 (Figure 6.13) the use of the ontology has resulted in a better improvement on the pf in comparison to the other topic sets. The improvement on the rfb baseline is from rank 15 onwards . At rank 5 there is a 18% improvement on the pf baseline and for rfb at rank 5 there is a 13% deterioration. The pf t-test results are very significant for this topic set. The rfb t-test results for this topic set did not show statistical significance.

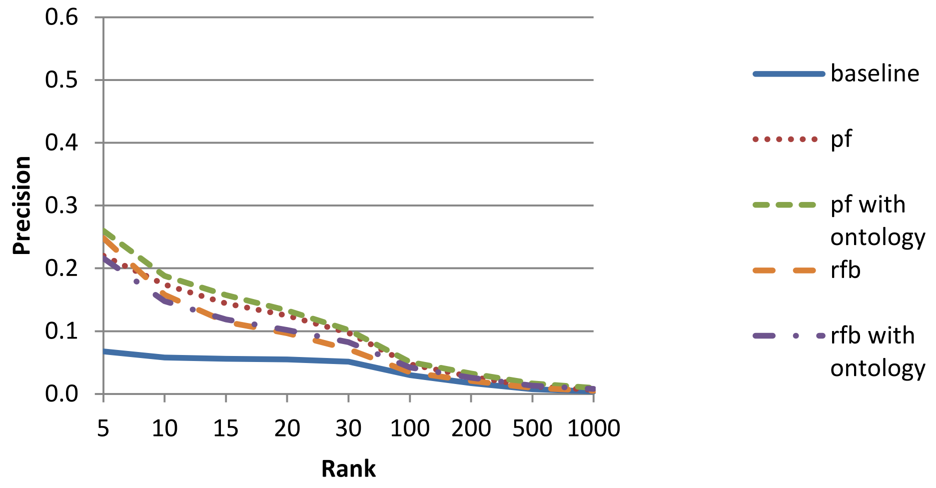


Figure 6.13 Topic 251-300 Results – Document Level Averages

Compared to the pf baseline in the standard run, the pf baseline and the pf with ontology curve in the 10 term run (topic 251-300) is degraded. The rfb baseline has also deteriorated and the rfb curve with ontology is unchanged. A full set of the TREC output figures for the standard run and the 10 term run can be seen in Appendix F (F1 and F4 respectively). The 10 term run (topic 251-300) has fewer terms than the standard run, eg it does not have election, lab, ap and applied.

6.3.2 Precision-Recall Graphs

The performance of the pf and the rfb curves, with and without ontology is virtually identical for all topics. At 0.0 recall (figure 6.14), precision is 22% less than rfb baseline whereas pf is 13% improvement on the pf baseline. There is a slight improvement on the pf baseline from 0.3 recall onwards. Table 5.6 shows topic 151-200 results to be significant and the t-test results for topic 201-250 and topic 251-300 are very significant. For the rfb run, none of the results were of statistical significance. The remaining Precision-Recall graphs have been included in Appendix A4.

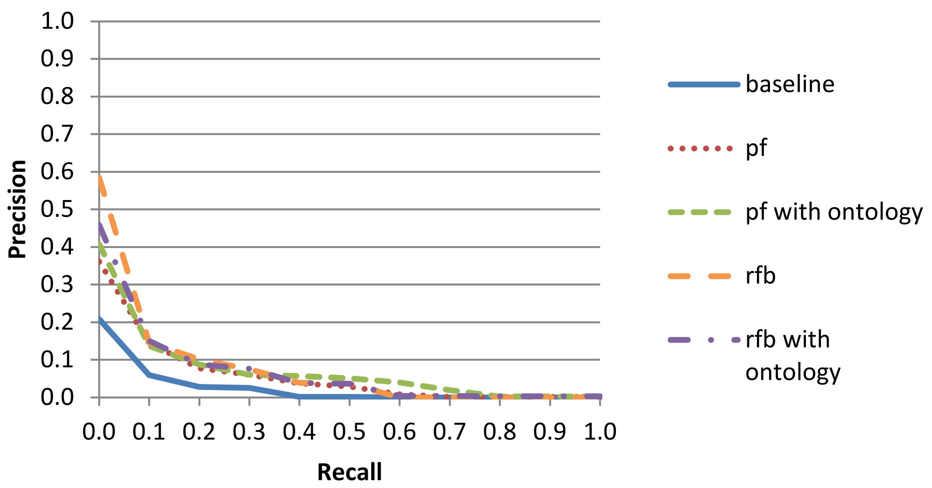


Figure 6.14 Topic 251-300 Results – Precision-Recall

Table 6.4 shows the t-test results on the document level averages for all topicsets and for both pf and rfb runs. The highlighted cells show that topic 251-300 is very significant and topic 101-150 is significant. For the rfb run, only topic 101-150 is very significant, the remaining topics are not significant.

Topics	PF	Rfb
51-100	0.079	0.229
101-150	0.025	0.013
151-200	0.394	0.141
201-250	0.126	0.151
251-300	0.013	0.481

Table 6.4: T-Test results (Document Level Averages)

Table 6.5 shows the Precision-Recall t-test results for all topicsets and for both pf and rfb runs. For the pf run, the highlighted cells show that topic 151-200 is significant; topic 201-250 and topic 251-300 are very significant. In Table 5.6, none of the rfb results show statistical significance.

Topics	PF	Rfb
51-100	0.262	0.217
101-150	0.143	0.150
151-200	0.023	0.070
201-250	0.010	0.101
251-300	0.012	0.193

Table 6.5: T-Test results (Precision-Recall)

6.4 Experiments Using Standard 20 Documents And 15 Terms And All Terms In Expanded Query

6.4.1 Document Level Averages Graphs

For topic 51-100 (Figure 6.15) the use of the ontology has resulted in the best improvement on the pf in comparison to the other topic sets in this run. In other graphs for this run, the rfb curves perform better but in this graph the performance of the pf curves is superior to that of the rfb curves. This is a similar situation to that of the experiment using 5 terms for relevance feedback (see 6.2.1). An example from this topic set which is suitable for improving performance is topic 94 ***computeraided crime*** because it is relevant to the International Economics and Technology areas covered by the newswire document collection. The improvement on the rfb baseline is from rank 100 onwards . At rank 5 there is a 12% improvement on the pf baseline and for rfb at rank 5 there is a 16% deterioration. The pf t-test results showed statistical significance for this topic set. The rfb t-test results for this topic set did not show statistical significance.

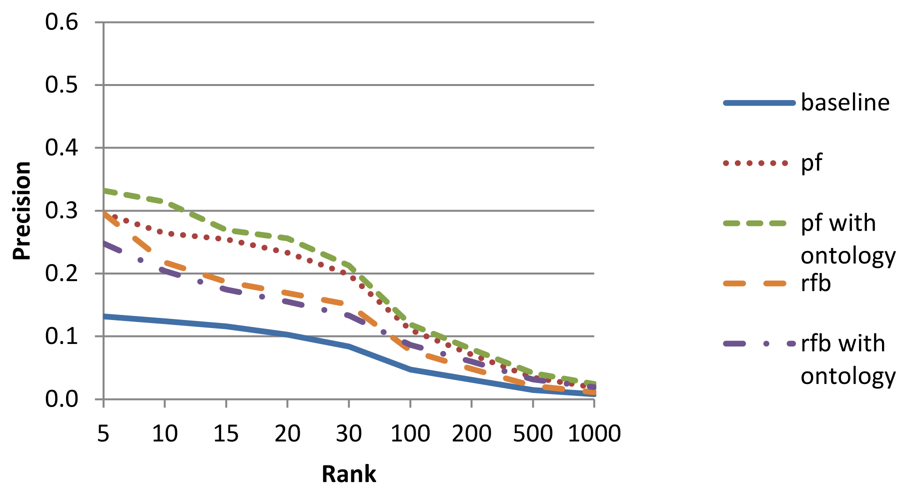


Figure 6.15 Topic 51-100 Results – Document Level Averages

For the 15 term run (topic 51-100), the pf baseline, pf with ontology curve, and the rfb baseline have all degraded compared to the standard run. The rfb curve with ontology is unchanged. A full set of the TREC output figures for the standard run and the 15 term run can be seen in Appendix F (F1 and F5 respectively). The 15 term run (topic 51-100) has some terms the same as the standard run, 5 term run and 10 term run. No additional ontology terms have been found.

For topic 101-150 (Figure 6.16) the use of the ontology has not resulted in any significant improvement on the pf. The very slight improvement on the rfb baseline is from rank 500 onwards . At rank 5 there is a 9% deterioration on the pf baseline and for rfb at rank 5 there is a 30% deterioration. The pf t-test results showed statistical significance for this topic set. The rfb t-test results for this topic set were very significant.

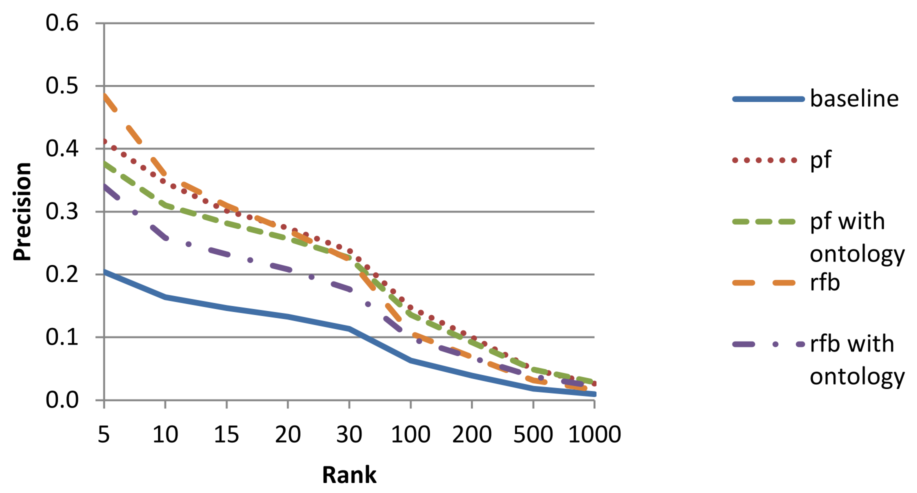


Figure 6.16 Topic 101-150 Results – Document Level Averages

For the 15 term run (topic 101-150), the pf baseline has improved at rank5 only but the , pf with ontology curve has deteriorated. The rfb baseline has improved at ranks 10, 15, 20, 30, 100 and 200 but the the rfb curve with ontology is unchanged. A full set of the TREC output figures for the standard run and the 15 term run can be seen in Appendix F (F1 and F5 respectively). The 15 term run (topic 101-150) has some extra terms such as lab, parliament, reli and lawyer. The precision at rank graphs for the remaining topics have been included in Appendix A5 because their results are very similar.

6.4.2 Results – Precision-Recall Graphs

The performance is virtually identical for all topics except for topic 251-300 where the Precision-Recall starts to perform better at recall 0.2 onwards (figure 6.17). The significance of this result is supported by the t-test scores for this topic set. There is no improvement for the rfb curve. So query expansion has not proved to be of any benefit for the rfb runs presented in this section. The Precision-Recall graphs for the other topic sets are all very similar and can be found in Appendix D.

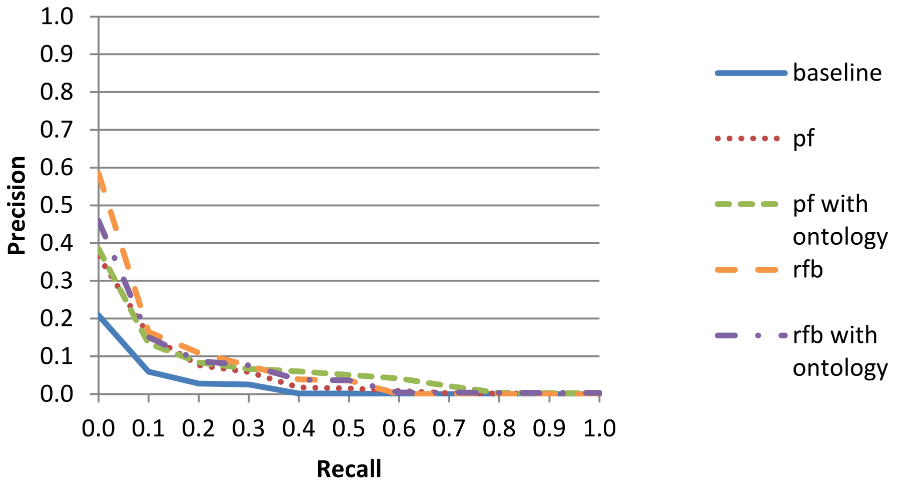


Figure 6.17 Topic 251-300 Results – Precision-Recall

In Table 6.6 we can see from the highlighted cells that the pf run results were of statistical significance for all topics except for topics151-200. The only results of statistical significance for the rfb run were for topics101-150.

Topics	PF	Rfb
51-100	0.003	0.146
101-150	0.005	0.014
151-200	0.312	0.123
201-250	0.044	0.151
251-300	0.032	0.085

Table 6.6: T-Test results (Document Level Averages)

In Table 6.7, for the pf run, the highlighted cells show that topic 151-200 is very significant; topic 201-250 and topic 251-300 are significant. None of the results for the rfb run are statistically significant.

Topics	PF	Rfb
51-100	0.376	0.207
101-150	0.058	0.146
151-200	0.009	0.063
201-250	0.046	0.124
251-300	0.019	0.146

Table 6.7: T-Test results (Precision-Recall)

6.5 Experiments Using Standard 20 Documents And 100 Terms And All Terms In Expanded Query

6.5.1 Document Level Averages Graphs

For topic 51-100 (Figure 6.18) the use of the ontology has resulted in a slight improvement on the pf. At rank 5 there is a 1% improvement on the pf baseline but this increases to 9% for rank 10. For rfb at rank 5 there is a 22% deterioration. This is the only graph for this run set where the rfb curve is lower than the pf curves for ranks 5-10. For this particular topic, the use of 100 terms does not benefit the rfb run possibly due to query drift. In other words the additional relevance feedback terms have extracted terms from the ontology which have resulted in fewer relevant documents being retrieved. The very slight improvement on the rfb baseline is from rank 200 onwards . The pf t-test results showed statistical significance for this topic set. The rfb t-test results for this topic set were also significant.

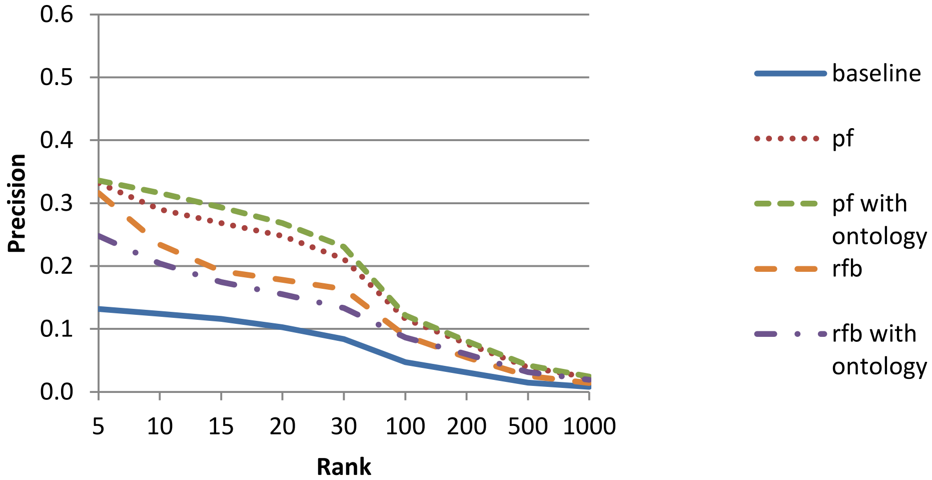


Figure 6.18 Topic 51-100 Results – Document Level Averages

For the 100 term run (topic 51-100), the pf baseline, pf with ontology curve and the rfb baseline have all improved compared to the standard run. Its only the the rfb curve with ontology which remains unchanged. A full set of the TREC output figures for the standard run and the 100 term run can be seen in Appendix F (F1 and F6 respectively). The 100 term run (topic 51-100) has quite a few extra terms such as ***defenc, law,***

lawyer, election, war, minor, di, polic, de, la, employment and ***rep*** which are causing the improvement in the pf with ontology curve. For example, ***polic*** is related to the topic ***greenpeace*** because during the protests, ***law enforcement*** takes place and sometimes ***arrests*** are made.

```
Topic no = 78 (greenpeace)
ORIG WORD IS polic
      --> law_enforcement
      --> operation
      --> arrest
```

For topic 101-150 (Figure 6.19) the use of the ontology has resulted in a slight improvement on the pf from rank 30 onwards. At rank 5 there is a 13% deterioration on the pf baseline. For rfb at rank 5 there is a 32% deterioration. The very slight improvement on the rfb baseline is from rank 500 onwards . The pf and rfb t-test results did not show statistical significance.

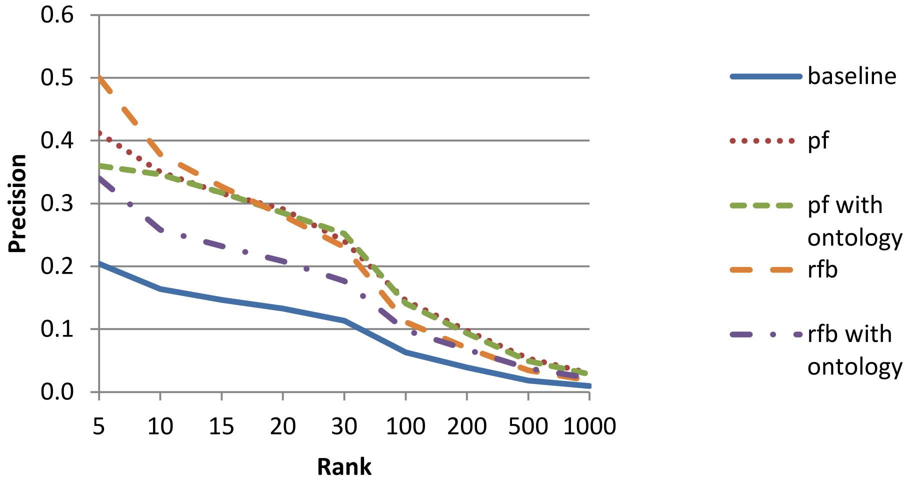


Figure 6.19 Topic 101-150 Results – Document Level Averages

For the 100 term run (topic 101-150), the pf baseline has improved at rank 5 only, and the rfb baseline has improved compared to the standard run. The pf with ontology curve has improved (except at rank5) but the rfb curve with ontology the same for the standard run and the 100 term run. A full set of the TREC output figures for the standard run and the 100 term run can be seen in Appendix F (F1 and F6 respectively). The 100 term run (topic 101-150) has quite a few extra terms such as ***justice, transport, trial, unrest, global, relief and engineer.*** These additional terms are causing the improvement in the pf with ontology curve as the following example shows:

```
Topic no = 140 (political impact Islamic fundamentalism)
ORIG WORD IS unrest
      --> oposing_group
      --> truce
      --> armed_conflict
      --> civil_unrest
      --> coup_detat
```

```
--> terrorism
--> massacre
--> riots
--> demonstration
--> turf_war
--> war
--> conflict
--> crisis
--> weaponry
--> bombings
--> invasion
--> war_victim
```

The child terms for topic 140 would help to enhance recall results the term ***unrest*** and its child terms are related to the topic.

For topic 151-200 (Figure 6.20) the use of the ontology has resulted in a slight improvement on the pf from rank 100 onwards. At rank 5 there is a 2% improvement on the pf baseline. For rfb at rank 5 there is a 23% deterioration. The very slight improvement on the rfb baseline is from rank 100 onwards. The pf results did not show statistical significance but the rfb t-test scores were statistically significant.

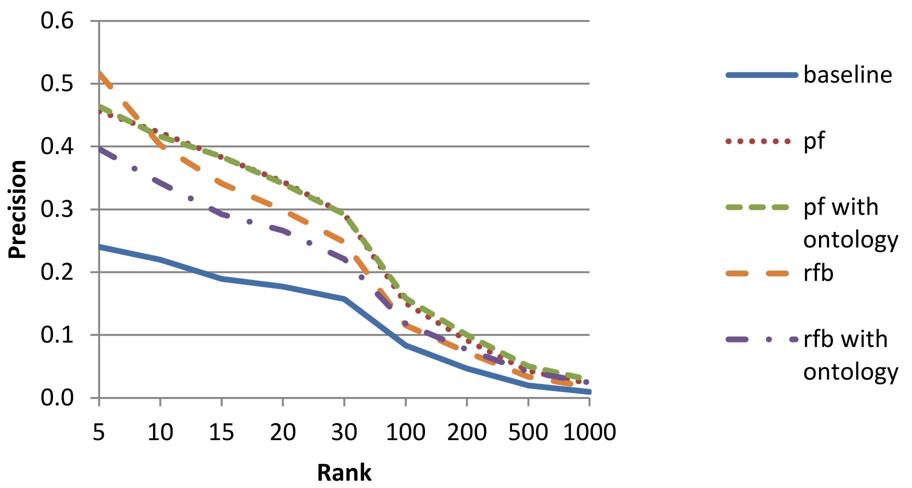


Figure 6.20 Topic 151-200 Results – Document Level Averages

For the 100 term run (topic 151-200), the pf baseline has improved (except for rank5) , and the rfb baseline has also improved compared to the standard run. The pf with ontology curve has improved except at ranks 5 and 10, but the rfb curve with ontology the same for the standard run and the 100 term run. A full set of the TREC output figures for the standard run and the 100 term run can be seen in Appendix F (F1 and F6 respectively). The 100 term run (topic 151-200) has quite a few additional terms such as justice, prevent elect, politic, diseas, health, law, employment, civil and movement.

For topic 201-250 (Figure 6.21) the use of the ontology has resulted in a consistent improvement on the pf from rank 5 onwards. At rank 5 there is a 6% improvement on the pf baseline. For rfb at rank 5 there is a 28% deterioration. The very slight improvement on the rfb baseline is from rank 100 onwards. The pf t-test results were very significant and the rfb t-test scores were significant.

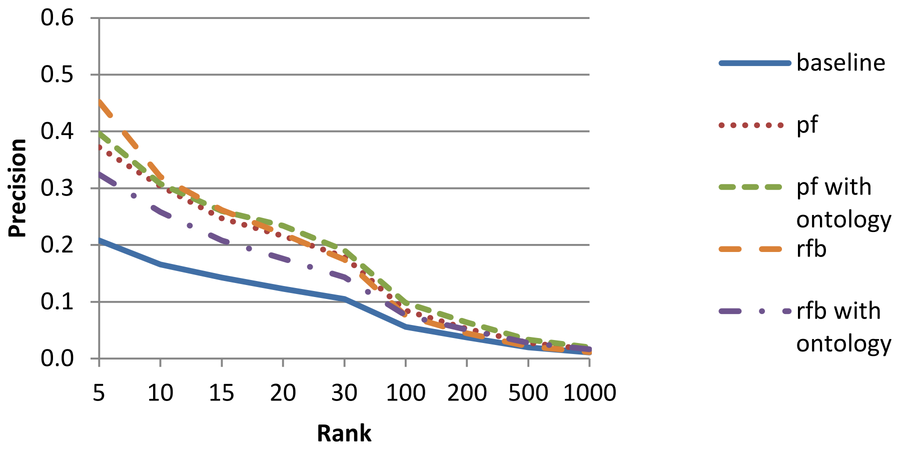


Figure 6.21 Topic 201-250 Results – Document Level Averages

For the 100 term run (topic 201-250), the pf baseline has deteriorated but the rfb baseline has improved compared to the standard run. The pf with ontology curve has improved, but the rfb curve with ontology is the same for the standard run and the 100 term run. A full set of the TREC output figures for the standard run and the 100 term run can be seen in Appendix F (F1 and F6 respectively). The 100 term run (topic 201-250) has quite a few additional terms such as health, mass, war, politic, trial, un, engineer, treat, civil, global, and justic.

For topic 251-300 (Figure 6.22) the use of the ontology has resulted in a consistent improvement on the pf from rank 5 onwards. At rank 5 there is a 31% improvement on the pf baseline. For rfb at rank 5 there is a 25% deterioration. The pf t-test results were very significant and the rfb t-test scores were significant.

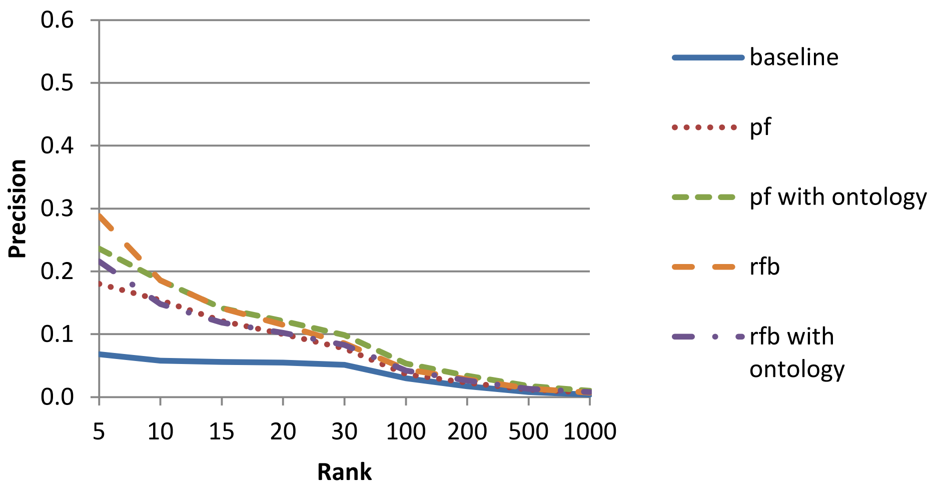


Figure 6.22 Topic 251-300 Results – Document Level Averages

For the 100 term run (topic 251-300), the pf baseline is worse, but the rfb baseline has improved compared to the standard run. The pf with ontology curve has improved except at ranks 5 and 20, but the rfb curve with ontology the same for the standard run and the 100 term run. A full set of the TREC output figures for the standard run and the 100 term run can be seen in Appendix F (F1 and F6 respectively). The 100 term run (topic 251-300) has quite a few additional terms such as engineer, civil, ap, mass, di, justice, interior, election, global, rep, movement, treat and law.

6.5.2 Results – Precision-Recall Graphs

The performance is virtually identical for all topics except for topic 251-300 where the Precision-Recall is consistently better than the pf baseline. The significance of this result is supported by the t-test scores for this topic set with the exception of topic 201-250 which does not show statistical significance. There is a very slight improvement for the rfb curve at recall 0.6 onwards (figure 6.23). So query expansion has not proved to be of any benefit for the rfb runs presented in this section. The Precision-Recall graphs for the other topic sets are all very similar and can be found in Appendix A6.

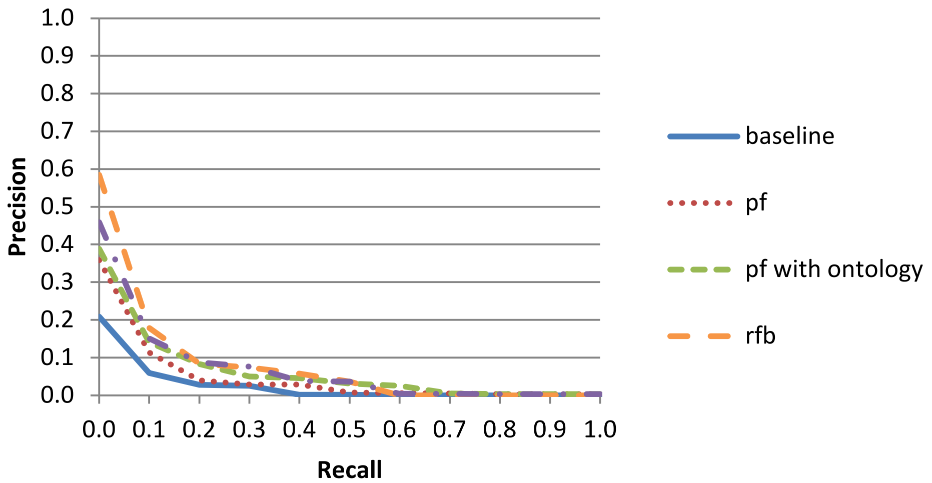


Figure 6.23 Topic 251-300 Results – Precision-Recall

In Table 6.8 for the pf run, the highlighted cells show that topic 51-100, topic 201-250 and topic 251-300 are very significant. For the rfb run, all topics are significant except for topic 101-150 which is very significant.

Topics	PF	Rfb
51-100	0.004	0.034
101-150	0.133	0.011
151-200	0.057	0.035
201-250	0.000	0.028
251-300	0.002	0.036

Table 6.8: T-Test results (Document Level Averages)

In Table 6.9, the highlighted cells show that all topics produced statistically significant results for the pf run. In the rfb only topic 151-200 and topic 201-250 are of statistical significance.

Topics	PF	Rfb
51-100	0.045	0.203
101-150	0.016	0.131
151-200	0.032	0.054
201-250	0.055	0.044
251-300	0.002	0.136

Table 6.9: T-Test results (Precision-Recall)

6.6 Experiments Using Standard 20 Documents And 200 Terms And All Terms In Expanded Query

6.6.1 Document Level Averages Graphs

By increasing the number of feedback terms to 200, this has had quite a positive impact on the pf curves. The ontology has produced significant improvements. For the rfb curves there is a slight improvement for topics 101-150, 151-200 from rank 100 onwards.

For topic 51-100 (Figure 6.24) the use of the ontology has resulted in a consistent improvement on the pf baseline. The biggest improvement on the pf baseline is at rank 20. Most graphs in this run set do show the pf with ontology curve to be higher than the other curves but this graph shows the largest gap between the pf with ontology curve and the other curves. An explanation for this is is that topic 51-100 is not considered to be a hard topic. This is a similar situation to that of the experiment using 5 terms for relevance feedback (see 6.2.1). For example topic 53 ***leveraged buyouts*** is commonly taking place in the field of Economics and the newswire document collection contains documents related to International Economics, For rfb at rank 5 there is a 22% deterioration. The pf t-test results were very significant and the rfb t-test scores were significant.

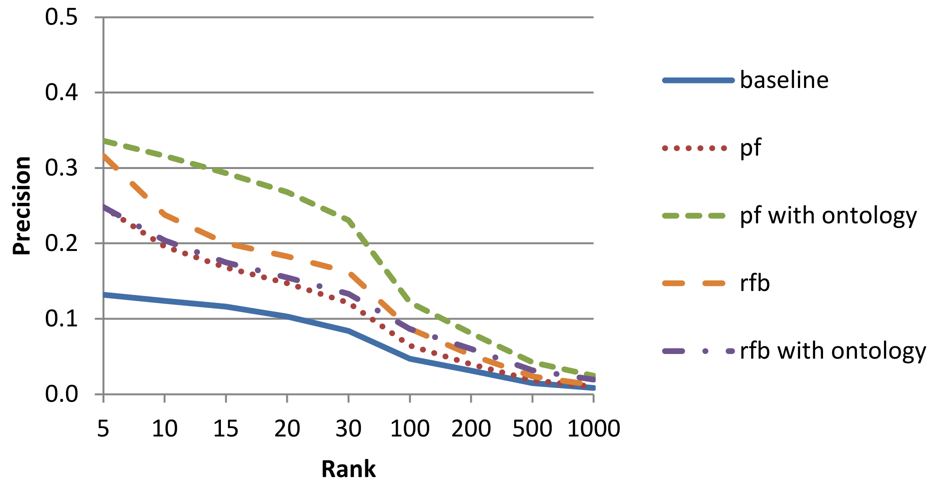


Figure 6.24 Topic 51-100 Results – Document Level Averages

For the 200 term run (topic 51-100), the pf baseline is worse, but the rfb baseline has improved compared to the standard run. The pf with ontology curve has improved except at rank 10, but the rfb curve with ontology the same for the standard run and the 200 term run. A full set of the TREC output figures for the standard run and the 200 term run can be seen in Appendix F (F1 and F7 respectively). The 200 term run (topic 51-100) has the same ontology terms as the 100 term run (topic 51-100).

For topic 101-151 (Figure 6.25) the use of the ontology has resulted in a consistent improvement on the pf baseline from rank 10 onwards. There is only 1% deterioration on pf baseline at rank 5. For rfb at rank 5 there is a 23% deterioration. However this is the only graph in this run set where the rfb curve is higher than the pf with ontology curve for ranks 5-10. The increased number of relevance feedback terms have extracted ontology terms which have produced an increase in the number of relevant documents. The improvement on the rfb baseline takes place from rank 100 onwards. The pf t-test results were very significant and the rfb t-test scores were significant.

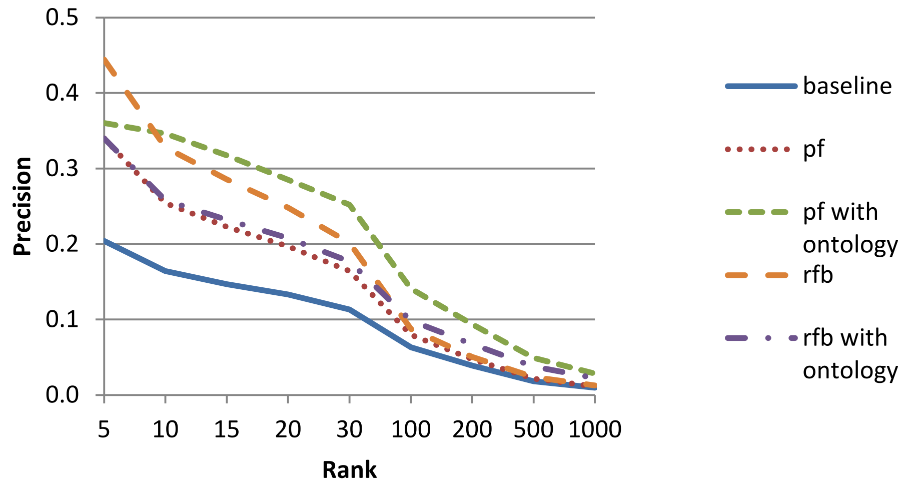


Figure 6.25 Topic 101-150 Results – Document Level Averages

For the 200 term run (topic 101-150), the pf baseline, the rfb baseline and the pf with ontology curve are worse compared to the standard run. The rfb curve with ontology is the same for the standard run and the 200 term run. A full set of the TREC output figures for the standard run and the 200 term run can be seen in Appendix F (F1 and F7 respectively). The 200 term run (topic 101-150) has one more ontology term than the 100 term run (topic 101-150), namely ***nature***.

```
Topic no = 123 (research into & control of carcinogens)
ORIG WORD IS natural
     --> earthquake
     --> tsunami
     --> flood
     --> drought
     --> avalanche
     --> landslide
     --> land_resources
     --> parks
     --> forests
     --> wetlands
     --> mountains
     --> rivers
     --> oceans
     --> wildlife
     --> energy_resources
     --> geology
     --> paleontology
     --> geography
     --> physiology
     --> botany
     --> astronomy
     --> biology
```

In this case ***natural*** might be very loosely related to the topic, especially the child terms ***physiology*** and ***biology***. This would improve recall but not precision.

For topic 151-200 (Figure 6.26) the use of the ontology has resulted in a consistent improvement on the pf baseline from rank 5 onwards. There is a 17% improvement on pf baseline at rank 5. For rfb at rank 5 there is a 7% deterioration which is much less than other topic sets. The improvement on the rfb baseline takes place from rank 100 onwards. The pf t-test results were very significant however the rfb t-test scores for this topic set were not significant.

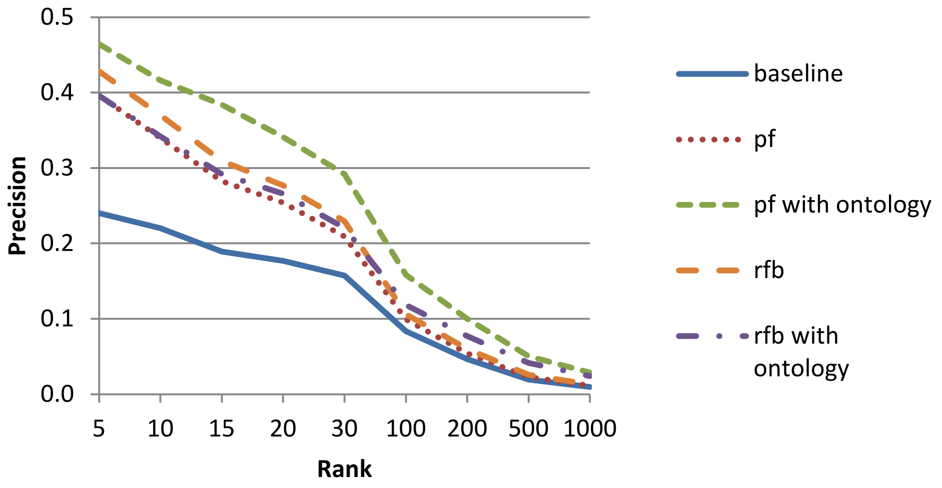


Figure 6.26 Topic 151-200 Results – Document Level Averages

For the 200 term run (topic 151-200), the pf baseline is lower than the standard run and the rfb baseline is worse except at ranks 20 and 30. The pf with ontology curve has improved except at ranks 5 and 10. The rfb curve with ontology is the same for the standard run and the 200 term run. A full set of the TREC output figures for the standard run and the 200 term run can be seen in Appendix F (F1 and F7 respectively). The 200 term run (topic 151-200) does not have any additional ontology terms compared to the standard run and the 100 term run.

For topic 201-250 (Figure 6.27) the use of the ontology has resulted in a consistent improvement on the pf baseline from rank 5 onwards. There is a 19% improvement on pf baseline at rank 5. For rfb at rank 5 there is a 9% deterioration on the rfb baseline. The improvement on the rfb baseline takes place from rank 100 onwards. The pf t-test results were very significant and the rfb t-test scores for this topic set were significant.

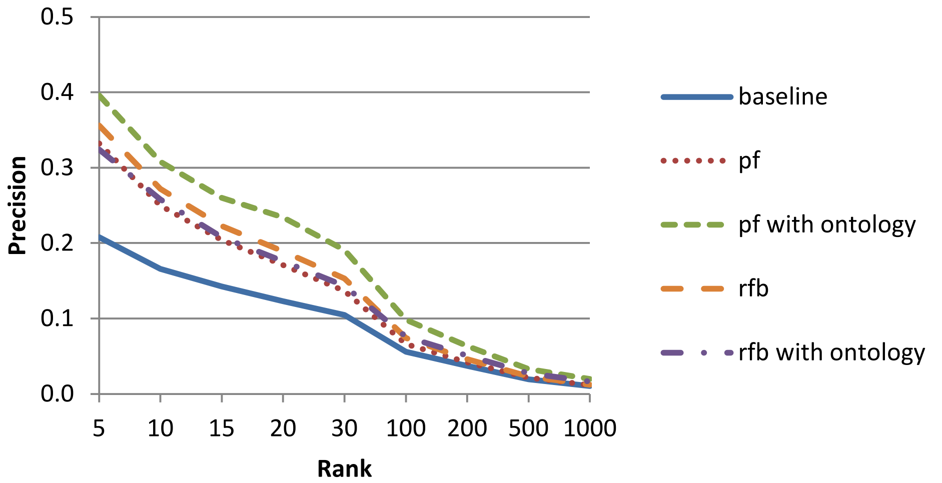


Figure 6.27 Topic 201-250 Results – Document Level Averages

For the 200 term run (topic 201-250), the pf baseline and the rfb baseline are lower than the standard run. The pf with ontology curve has improved but the rfb curve with ontology is the same for the standard run and the 200 term run. A full set of the TREC output figures for the standard run and the 200 term run can be seen in Appendix F (F1 and F7 respectively). The 200 term run (topic 201-250) has all of the ontology terms up to the 100 term run but also has an additional term relief.

```
Topic no = 225 (main function federal emergency management
agency fema funding level meet emergencies resources available
people equipment facilities)

ORIG WORD IS relief
     --> relief_aid_organisation
```

It looks like the ontology word ***relief*** and its child node would be related to the topic statement.

For topic 251-300 (Figure 6.28) the use of the ontology has resulted in a slightly smaller but consistent improvement on the pf baseline from rank 10 onwards. There is a 9% improvement on pf baseline at rank 5. For rfb at rank 5 there is a 13% deterioration on the rfb baseline. The improvement on the rfb baseline takes place from rank 200 onwards. The pf t-test results were very significant and the rfb t-test scores for this topic set were significant.

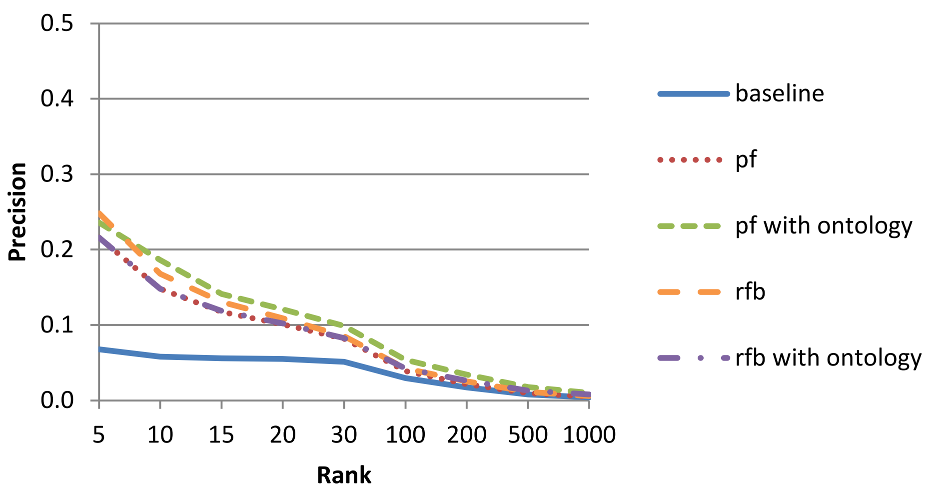


Figure 6.28 Topic 251-300 Results – Document Level Averages

For the 200 term run (topic 251-300), the pf baseline is lower than the standard run and the rfb baseline is worse except at ranks 30,100 and 200. The pf with ontology curve has improved except at rank 5. The rfb curve with ontology is the same for the standard run and the 200 term run. A full set of the TREC output figures for the standard run and the 200 term run can be seen in Appendix F (F1 and F7 respectively). The 200 term run (topic 251-300) has all of the ontology terms up to the 100 term run and no additional terms.

6.6.2 Precision-Recall Graphs

At recall 0.0 (figure 6.29), there is an 8% improvement on the pf baseline but a 17% deterioration on the rfb baseline. There is a consistent improvement on the pf curve. The significance of this result is supported by the t-test scores for this topic set. There is a very slight improvement for the rfb curve at recall 0.2 onwards. The rfb t-test score for this topic set is not significant. So query expansion has not proved to be of any benefit for the rfb runs presented in this section. The Precision-Recall graphs for the other topic sets are all very similar except for topic 251-300 where the Precision-Recall is slightly worse than the pf baseline. These graphs can be found in Appendix A7.

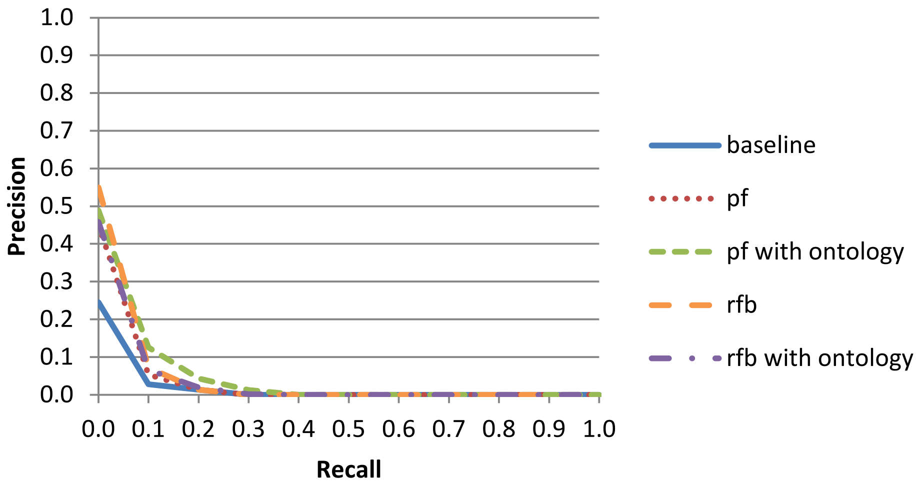


Figure 6.29 Topic 51-100 Results – Precision-Recall

In Table 6.10 for the pf run, the highlighted cells show that all topic sets are very significant. For the rfb run, all topics are significant except for topic 151-200. With the exception of topics101-150 all other topics produced statistically significant results for the rfb run.

Topics	PF	Rfb
51-100	0.000	0.035
101-150	0.000	0.052
151-200	0.000	0.247
201-250	0.000	0.053
251-300	0.000	0.044

Table 6.10: T-Test results (Document Level Averages)

In Table 6.11, for the pf run, topic 51-100 and topic 151-200 are significant and topic 201-250 is very significant. For the rfb run, only topic 201-250 is very significant.

Topics	PF	Rfb
51-100	0.041	0.177
101-150	0.135	0.083
151-200	0.044	0.189
201-250	0.011	0.008
251-300	0.184	0.232

Table 6.11: T-Test results (Precision-Recall)

6.7 Summary

In this chapter we have presented the results for experiments where the number of terms relevance feedback parameter has been varied.

Firstly, we note that the all of our experimental runs (with the exception of the run where 100 terms are used for relevance feedback) retrieved a higher number of relevant documents compared to the standard run.

Secondly, there doesn't seem to be any clear trend or correlation between the number of terms parameter and the increase in the number of relevant documents retrieved.

For the precision at rank results, the use of ontology based query expansion has improved the pf with ontology and rfb with ontology curves in the standard run. However for all of the other runs where the number of terms relevance feedback parameter has been varied, it is the pf with ontology curve which is affected by the ontology, more so for topic 51-100 which is considered a non-hard topic (see chapter 9.4.3 for discussion on Hard Topics).The rfb with ontology curve has not been affected by ontology based query expansion and has identical performance to the standard run.

Finally the tables in appendix F (F1, F3-F7) for the precision-recall results, the pf with ontology curve has slightly better performance for10 terms (topic 251-300), 100 terms (topic 151-200) and 200 terms (topic 151-200). Tables F8-F12 show slight improvements in the pf figures for the 10 document run (topic 51-100). The rfb figures are better for the 100 document run (all topic sets) and the 200 document run (all topic sets). As the precision-recall graphs in this chapter show, the improvement is only slight ie 1%.

7. EXPERIMENT RESULTS FOR VARYING NUMBER OF DOCUMENTS RELEVANCE FEEDBACK PARAMETER

7.1 Introduction

In this chapter we present the results for experiments 7-11 where the number of documents parameter is set to 5, 10, 15, 100 and 200 respectively. We can see from table 7.1 that ontology based query expansion has for most cases resulted in a large increase in the number of relevant documents retrieved.

#docs parameter	Topics	+/- #docs retrieved (Pf)	+/-#docs retrieved (Rfb)
5	51-100	222	413
5	101-150	161	450
5	151-200	418	607
5	201-250	97	296
5	251-300	78	125
10	51-100	402	750
10	101-150	185	250
10	151-200	-46	471
10	201-250	101	261
10	251-300	40	151
15	51-100	146	429
15	101-150	112	255
15	151-200	362	443
15	201-250	222	312
15	251-300	101	110
standard	51-100	161	393
standard	101-150	103	252
standard	151-200	354	376
standard	201-250	249	323
standard	251-300	91	116
100	51-100	446	103
100	101-150	365	-35
100	151-200	420	144
100	201-250	119	86
100	251-300	154	30
200	51-100	471	-101
200	101-150	561	-122
200	151-200	453	20
200	201-250	247	-52
200	251-300	202	36

Table 7.1 Difference in number of relevant documents retrieved for pf and rfb (varying number of documents parameter)

Sections 7.2-7.6 will examine the effects of varying the number of terms parameter used for relevance feedback**.**

7.2 Experiments Using 5 Documents And Standard 20 Terms And All Terms In Expanded Query

7.2.1 Document Level Averages Graphs

For topic 51-100 (Figure 7.1) from rank 5 onwards there is a 10% improvement on pf baseline at rank 5. After rank 5 the performance deteriorates and matches the original pf run from rank 100 onwards. For rfb at rank 5 there is a 29% deterioration on the rfb baseline. The improvement on the rfb baseline takes place from rank 100 onwards. The pf t-test results did not show statistical significance however the rfb t-test scores for this topic set were significant.

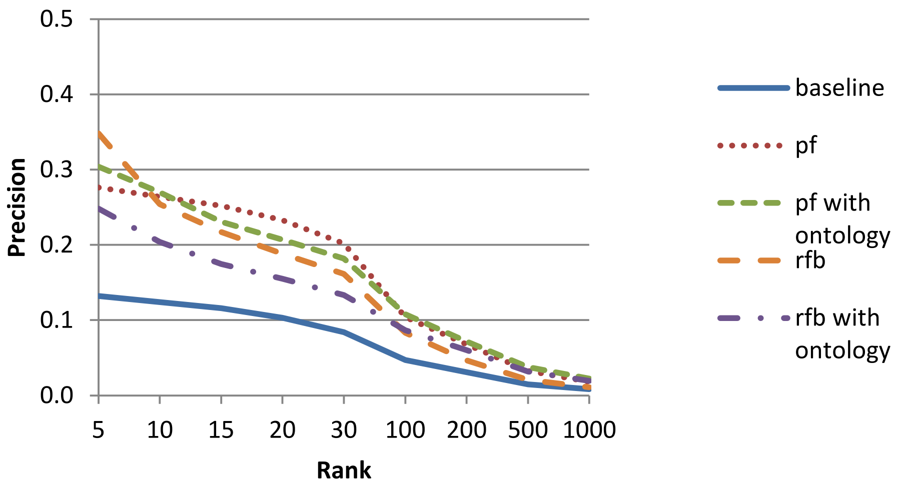


Figure 7.1 Topic 51-100 Results – Document Level Averages

For the 5 document run (topic 51-100), the pf baseline and the pf with ontology curve is lower than the standard run. The rfb baseline is improved but the rfb curve with ontology is the same for the standard run and the 5 document run. A full set of the TREC output figures for the standard run and the 5 document run can be seen in Appendix F (F1 and F8 respectively). The 5 document run (topic 51-100) has some additional terms such as weather, civil, health, conflict, lawyer, report and research.

For topic 101-150 (Figure 7.2) from rank 5 onwards there is a 12% improvement on pf baseline at rank 5. After rank 5 the performance deteriorates and matches the original pf run from rank 500 onwards. For rfb at rank 5 there is a 14% deterioration on the rfb baseline. The improvement on the rfb baseline takes place from rank 100 onwards. The pf and rfb t-test results did not show statistical significance

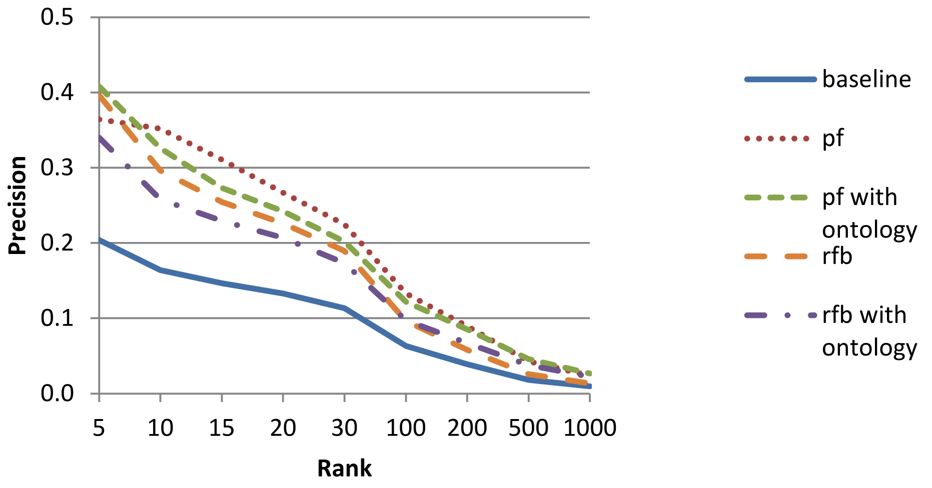


Figure 7.2 Topic 101-150 Results – Document Level Averages

For the 5 document run (topic 101-150), the pf baseline, the pf with ontology curve (except at rank5) and the rfb baseline are lower than the standard run. The rfb curve with ontology is the same for the standard run and the 5 document run. A full set of the TREC output figures for the standard run and the 5 document run can be seen in Appendix F (F1 and F8 respectively). The 5 document run (topic 101-150) has some additional terms such as di, global and justice.

For topic 151-200 (Figure 7.3) from rank 5 onwards there is a 3% deterioration on pf baseline at rank 5. After rank 5 the performance deteriorates and matches the original pf run from rank 100 onwards. For rfb at rank 5 there is a 11% deterioration on the rfb baseline. The improvement on the rfb baseline takes place from rank 20 onwards. The pf and rfb t-test results did not show statistical significance

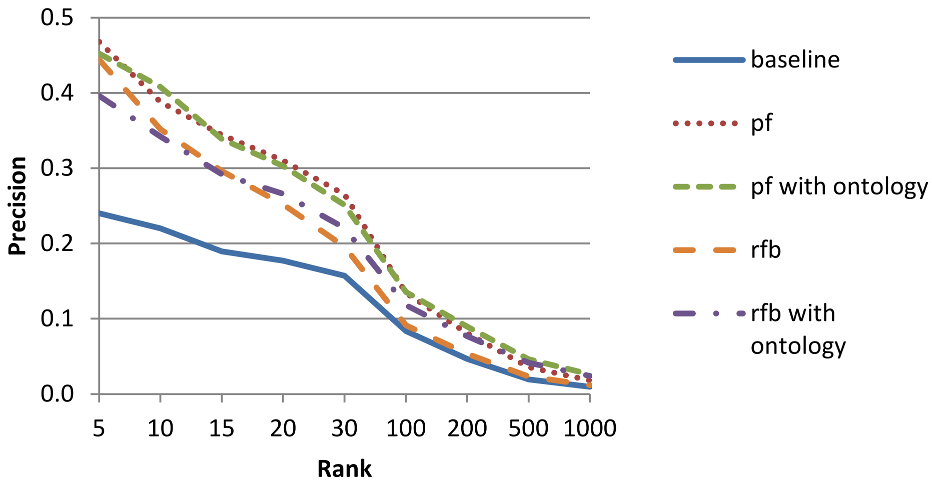


Figure 7.3 Topic 151-200 Results – Document Level Averages

For the 5 document run (topic 151-200), the pf baseline (except at rank 5), the pf with ontology, and the rfb baseline are lower than the standard run. The rfb curve with ontology is the same for the standard run and the 5 document run. A full set of the TREC output figures for the standard run and the 5 document run can be seen in

Appendix F (F1 and F8 respectively). The 5 document run (topic 151-200) has some additional terms such as defenc, corporat and elect.

For topic 201-250 (Figure 7.4) the two pf curves are identical at rank 5. After rank 5 the performance deteriorates and matches the original pf run from rank 100 onwards. For rfb at rank 5 there is a 10% deterioration on the rfb baseline. The improvement on the rfb baseline takes place from rank 100 onwards. The pf and rfb t-test results did not show statistical significance

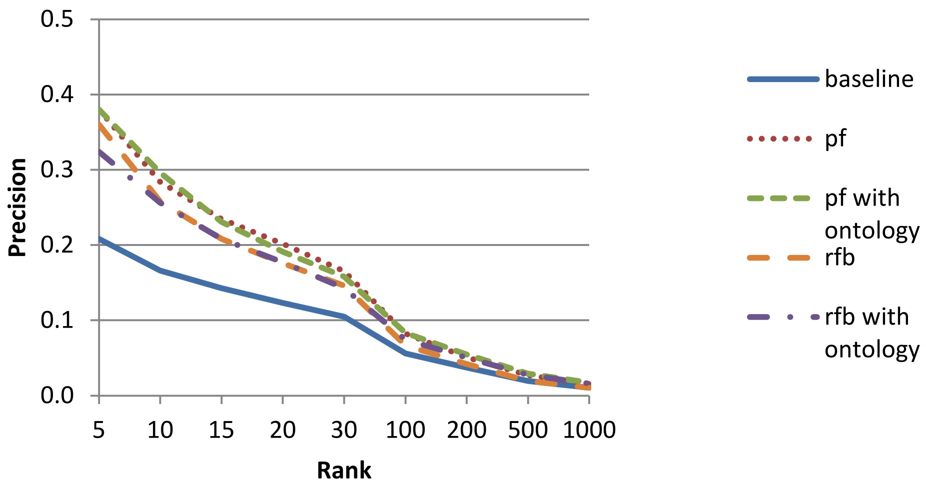


Figure 7.4 Topic 201-250 Results – Document Level Averages

For the 5 document run (topic 201-250), the pf baseline, the pf with ontology and the rfb baseline are lower than the standard run. The rfb curve with ontology is the same for the standard run and the 5 document run. A full set of the TREC output figures for the standard run and the 5 document run can be seen in Appendix F (F1 and F8 respectively). The 5 document run (topic 201-250) has some additional terms such as health, politic, trial, engineer, treat, global and relief.

For topic 251-300 (Figure 7.5) from rank 5 onwards there is a 23% improvement on pf baseline at rank 5. After rank 5 the performance is still an improvement on the baseline, until rank 200 when the two curves are the same. For rfb at rank 5 there is a 7% deterioration on the rfb baseline. The improvement on the rfb baseline takes place from rank 100 onwards. The pf and rfb t-test results both show statistical significance

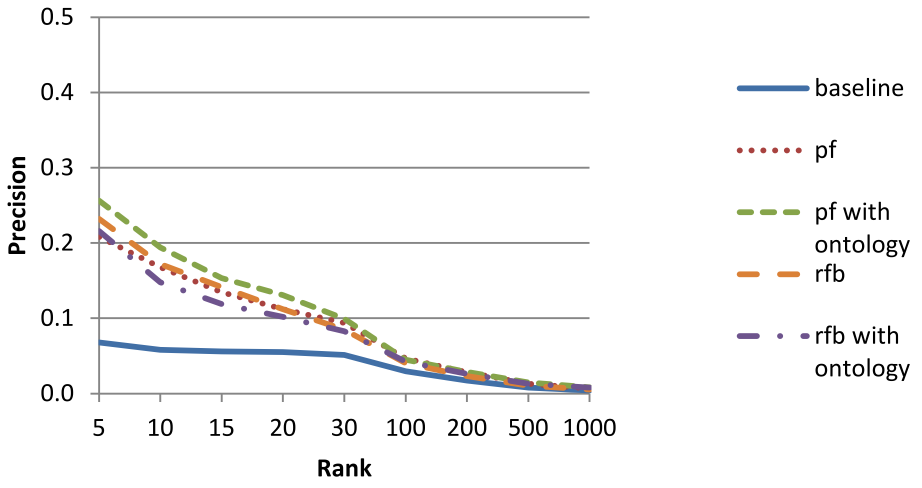


Figure 7.5 Topic 251-300 Results – Document Level Averages

For the 5 document run (topic 251-300), the pf baseline is lower than the standard run. However, the pf with ontology and the rfb baseline (except at rank5) are higher than the standard run. The rfb curve with ontology is the same for the standard run and the 5 document run. A full set of the TREC output figures for the standard run and the 5 document run can be seen in Appendix F (F1 and F8 respectively). The 5 document run (topic 251-300) has some additional terms such as ap, global, treat, law, state and report.

7.2.2 Precision-Recall Graphs

At recall 0.0 (figure 7.6), there is an 9% deterioration on the pf baseline and a 21% deterioration on the rfb baseline. There is a consistent deterioration on the pf curve. The significance of this result is supported by the t-test scores for this topic set. There is a very slight improvement for the rfb curve at recall 0.2 after which the two curves are the same. The rfb t-test score for this topic set is also significant. So query expansion has not proved to be of any benefit for the pf or the rfb runs presented in this section. The Precision-Recall graphs for the other topic sets are all very similar except for topic 251-300 where the Precision-Recall is slightly better than the pf baseline. These graphs can be found in Appendix A8.

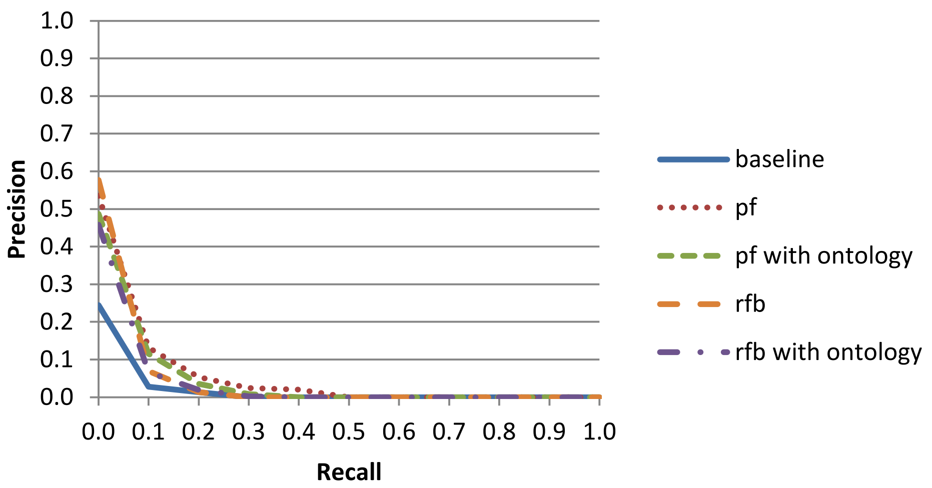


Figure 7.6 Topic 51-100 Results – Precision-Recall

In Table 7.2, for the pf run, the highlighted cell shows only topic 251-300 is significant. For the rfb run, only topic 51-100 and topic 2351-300 are significant.

Topics	PF	Rfb
51-100	0.382	0.044
101-150	0.163	0.058
151-200	0.425	0.222
201-250	0.469	0.403
251-300	0.021	0.046

Table 7.2: T-Test results (Document Level Averages)

In Table 7.3, for the pf run, topic 51-100 and topic 151-200 are significant. For the rfb run, only topic 251-300 is significant.

Topics	PF	Rfb
51-100	0.020	0.182
101-150	0.358	0.176
151-200	0.026	0.256
201-250	0.068	0.229
251-300	0.056	0.042

Table 7.3: T-Test results (Precision-Recall)

7.3 Experiments Using 10 Documents And Standard 20 Terms And All Terms In Expanded Query

7.3.1 Document Level Averages Graphs

For topic 51-100 (Figure 7.7) from rank 5 onwards there is a 4% improvement on pf baseline at rank 5. After rank 5 the performance deteriorates and matches the original pf run from rank 100 onwards. For rfb at rank 5 there is a 23% deterioration on the rfb baseline. The improvement on the rfb baseline takes place from rank 100 onwards. The pf t-test and rfb results did not show statistical significance.

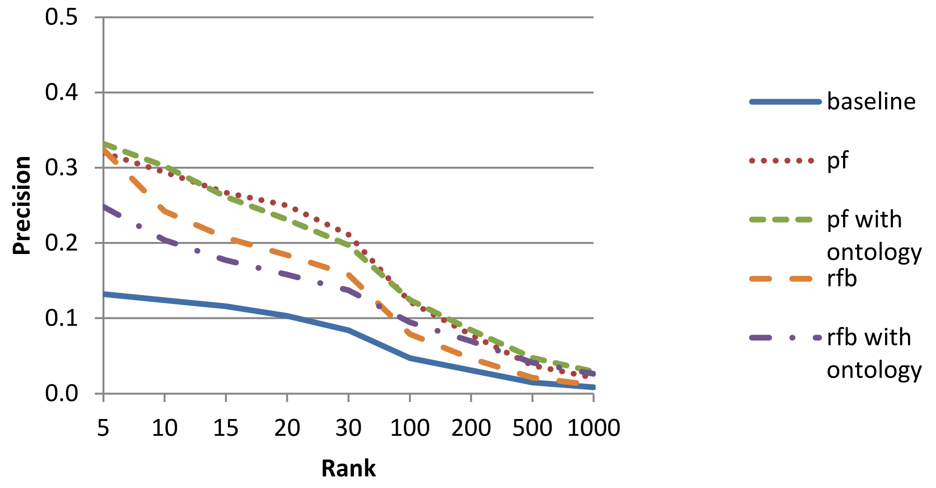


Figure 7.7 Topic 51-100 Results – Document Level Averages

For the 10 document run (topic 51-100), the pf baseline, pf with ontology curve (except ranks 10-30) and the rfb baseline (except ranks 200, 500 and 1000) are higher than the standard run. In addition, the rfb with ontology curve (except ranks 5 and 10) is higher than the standard run. A full set of the TREC output figures for the standard run and the 10 document run can be seen in Appendix F (F1 and F9 respectively). The 10 document run (topic 51-100) has the same terms as the 5 document run and some additional terms such as election, di and research.

For topic 101-150 (Figure 7.8) from rank 5 onwards there is a 23% improvement on pf baseline at rank 5. There is a improvement on the pf baseline curve up until rank 100. After rank 100 the performance deteriorates and matches the original pf run from rank 500 onwards. For rfb at rank 5 there is a 19% deterioration on the rfb baseline. The improvement on the rfb baseline takes place from rank 500 onwards. The pf t-test and did not show statistical significance. However the rfb results were very significant.

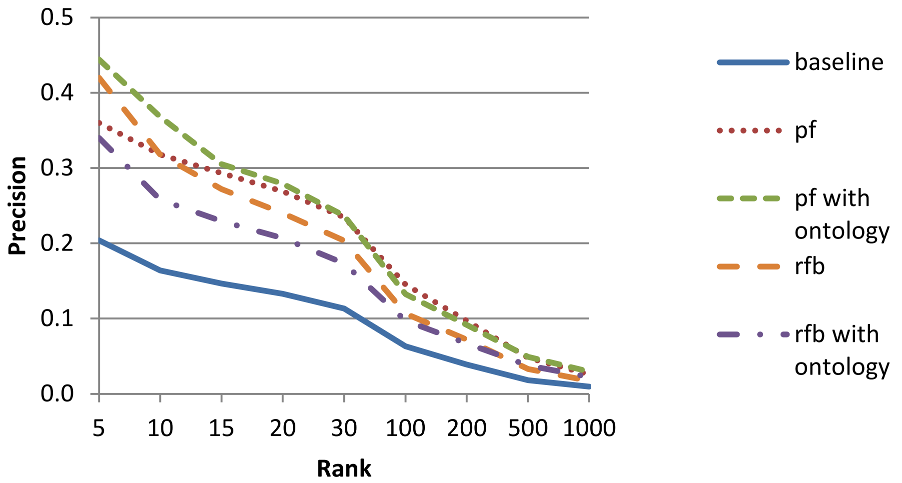


Figure 7.8 Topic 101-150 Results – Document Level Averages

For the 10 document run (topic 101-150), the pf baseline, rfb baseline, and the rfb with ontology curve (except at ranks 5, 10, 500 and 1000) is lower than the standard run. However, the pf with ontology curve (except ranks 30, 100 and 200) is higher than the standard run. A full set of the TREC output figures for the standard run and the 10 document run can be seen in Appendix F (F1 and F9 respectively). The 10 document run (topic 101-150) has the same terms as the 5 document run and some additional terms such as parliament, trial and global.

For topic 151-200 (Figure 7.9) there is a 54% deterioration on pf baseline at rank 5. After which there is a consistent deterioration in performance. This is the only graph for this run set where the pf with onto curve performs worse than the other curves and is virtually identical to the baseline. An explanation for this is that the terms in these 10 documents have caused query drift. In pseudo-relevance feedback, problems arise when terms or phrases taken from assumed-to-be relevant documents that are actually non-relevant are added to the query causing a drift in the focus of the query (Song et al, 2007). An example of this is topic 57 'mci' which could be abbreviation for different organisations or concepts so some of the documents retrieved for this topic might not be relevant. For rfb at rank 5 there is a 19% deterioration on the rfb baseline. The improvement on the rfb baseline takes place from rank 20 onwards. The pf t-test were very significant. However the rfb results did not show any statistical significance.

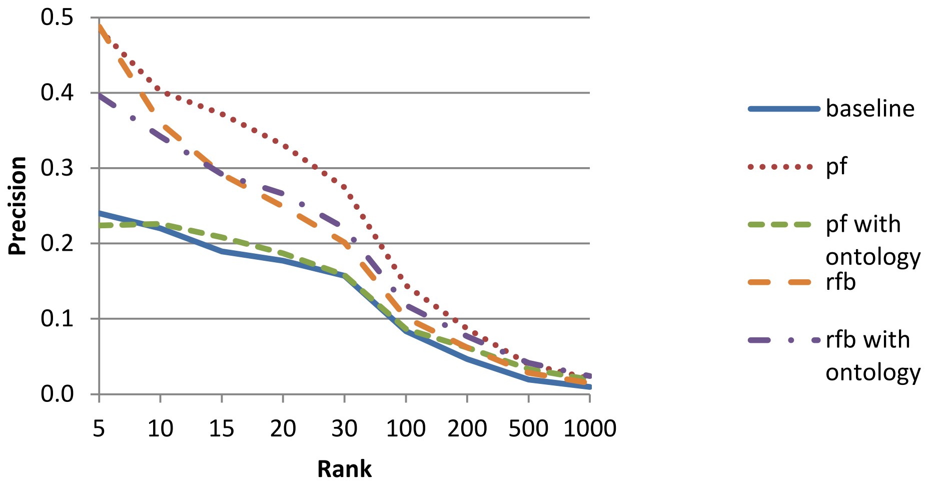


Figure 7.9 Topic 151-200 Results – Document Level Averages

For the 10 document run (topic 151-200), the pf baseline (except at ranks 5 and 15), rfb baseline, and the rfb with ontology curve are lower than the standard run. However, the rfb with ontology curve is the same as the standard run. A full set of the TREC output figures for the standard run and the 10 document run can be seen in Appendix F (F1 and F9 respectively). The 10 document run (topic 151-200) has the same terms as the 5 document run and some additional terms such as movement and prevent.

For topic 201-250 (Figure 7.10) there is a 10% deterioration on pf baseline at rank 5. After which there is a consistent deterioration in performance until rank 200 when the performance starts to improve on the baseline. For rfb at rank 5 there is a 20% deterioration on the rfb baseline. The improvement on the rfb baseline takes place from rank 200 onwards. The pf t-test were very significant. However the rfb results did not show any statistical significance.

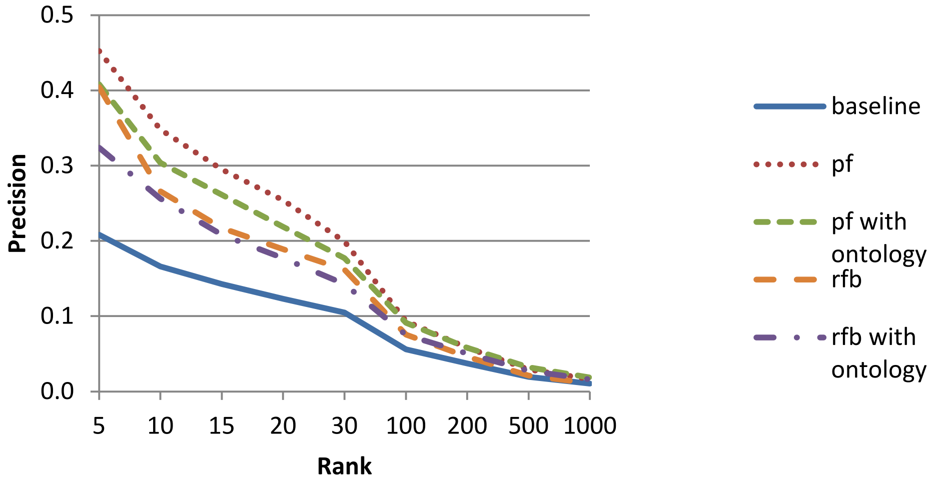


Figure 7.10 Topic 201-250 Results – Document Level Averages

For the 10 document run (topic 201-250), the pf baseline and the pf with ontology curve are higher than the standard run but the rfb baseline is lower than the standard run. The rfb with ontology curve (except at ranks 5, 10, 500 and 1000) is the same as the standard run except at ranks 10, 20, 100 and 200. A full set of the TREC output figures for the standard run and the 10 document run can be seen in Appendix F (F1 and F9 respectively). The 10 document run (topic 201-250) has fewer terms than the 5 document run.

For topic 251-300 (Figure 7.11) there is a 17% improvement on pf baseline at rank 5. After which there is a consistent deterioration in performance until rank 500 when the performance starts to improve on the baseline. For rfb at rank 5 there is a 10% deterioration on the rfb baseline. The improvement on the rfb baseline takes place from rank 100 onwards. The pf and rfb t-test results were not significant.

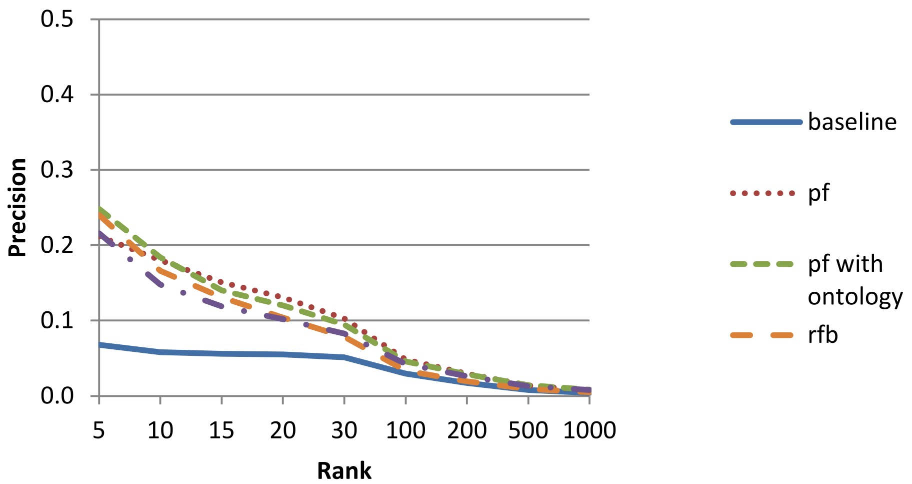


Figure 7.11 Topic 251-300 Results – Document Level Averages

For the 10 document run (topic 251-300), the pf baseline (except at ranks10), the pf with ontology curve (except at rank 5 and 15) and the rfb baseline are lower than the standard run but the rfb with ontology curve is the same as the standard run. The rfb with ontology curve (except at ranks 5, 10, 500 and 1000) is the same as the standard run except at ranks 10, 20, 100 and 200. A full set of the TREC output figures for the standard run and the 10 document run can be seen in Appendix F (F1 and F9 respectively). The 10 document run (topic 251-300) does not have any additional ontology hits.

At recall 0.0 (7.12), there is a 4% deterioration on the pf baseline and a 23% deterioration on the rfb baseline. There is a consistent deterioration on the pf curve. The t-test results are very significant. There is an improvement for the rfb curve at recall 0.1 to 0.2 after which the two curves are the same. The rfb t-test score for this topic set is not significant. So query expansion has not proved to be of any benefit for the pf or the rfb runs presented in this section. The Precision-Recall graphs for the other topic sets are all very similar except for topic 151-200 where the Precision-Recall is substantially worse than the pf baseline ompared to the other topic-sets. These graphs can be found in Appendix A9.

7.3.2 Precision-Recall Graphs

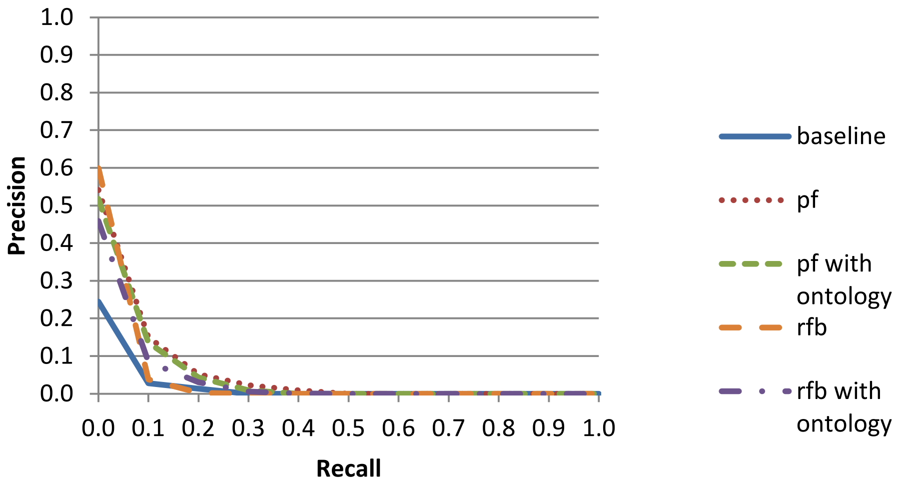


Figure 7.12 Topic 51-100 Results – Precision-Recall

In Table 7.4, for the pf run, the highlighted cells show that topic 151-200 and topic 201-250 are very significant. For the rfb run, only topic 101-150 is very significant.

Topics	PF	Rfb
51-100	0.402	0.142
101-150	0.079	0.011
151-200	0.004	0.432
201-250	0.010	0.094
251-300	0.427	0.214

Table 7.4: T-Test results (Document Level Averages)

In Table 7.5, for the pf run, the highlighted celles show that topic 51-100 is very significant; topic 151-200 and topic 251-300 are significant. For the rfb run, only topic 251-300 produced a statistically significant result.

Topics	PF	Rfb
51-100	0.014	0.355
101-150	0.383	0.129
151-200	0.036	0.170
201-250	0.248	0.088
251-300	0.043	0.047

Table 7.5: T-Test results (Precision-Recall)

7.4 Experiments Using 15 Documents And Standard 20 Terms And All Terms In Expanded Query

7.4.1 Document Level Averages Graphs

For topic 51-100 (Figure 7.13) from rank 5 onwards there is a 2% deterioration on pf baseline at rank 5. After rank 5 the performance deteriorates and matches the original pf run from rank 200 onwards. For rfb at rank 5 there is a 15% deterioration on the rfb baseline. The improvement on the rfb baseline takes place from rank 100 onwards. The pf t-test and rfb results did not show statistical significance.

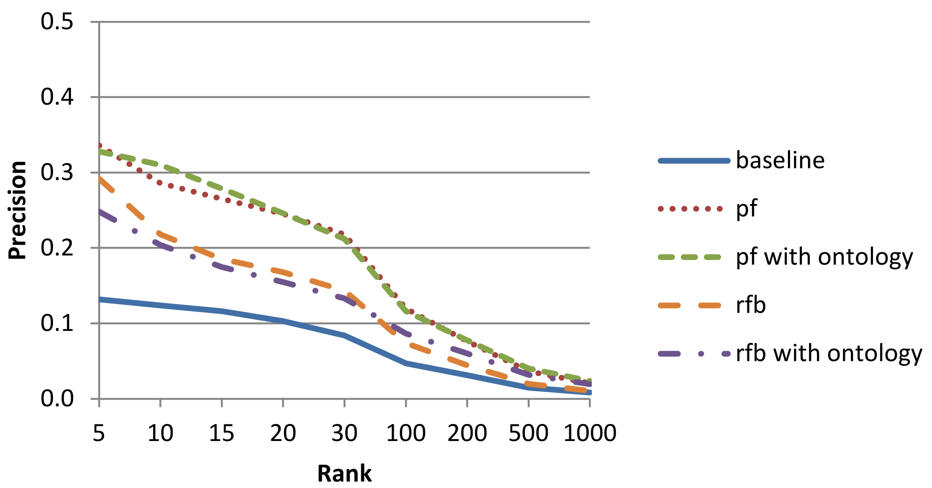


Figure 7.13 Topic 51-100 Results – Document Level Averages

For the 15 document run (topic 51-100), the pf baseline and the pf with ontology curve (except at rank 20, 30, 200 and 500) are lower than the standard run but the rfb baseline is lower than the standard run and the rfb with ontology curve is the same as the standard run. A full set of the TREC output figures for the standard run and the 15 document run can be seen in Appendix F (F1 and F10 respectively). The 15 document run (topic 51-100) has the same terms as the 10 document run with the addition of scienc.

For topic 101-150 (Figure 7.14) from rank 5 onwards there is a 2% deterioration on pf baseline at rank 5. There is an improvement on the pf baseline between ranks 10 and 100 after which the curves are virtually identical. For rfb at rank 5 there is a 25% deterioration on the rfb baseline. The improvement on the rfb baseline takes place from rank 500 onwards. The pf t-test results did not show statistical significance however the rfb results are very significant.

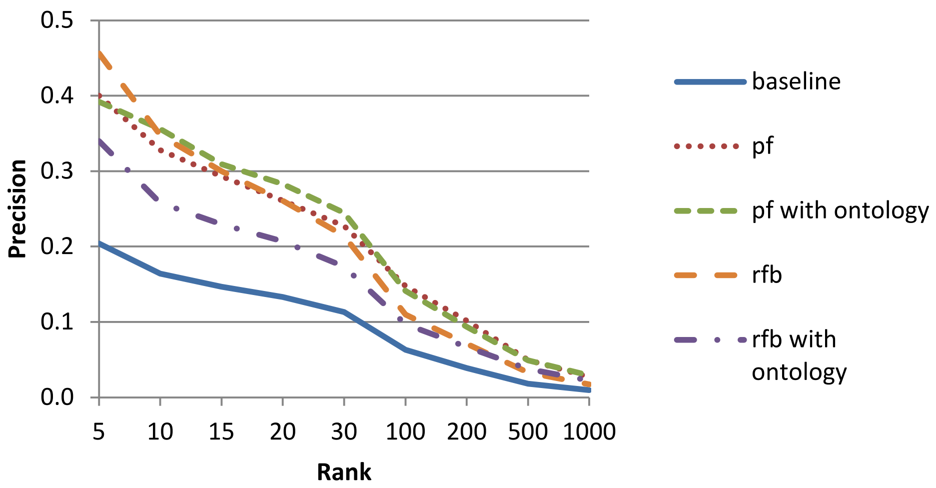


Figure 7.14 Topic 101-150 Results – Document Level Averages

For the 15 document run (topic 101-150), the pf baseline (except at rank5) is lower than the standard run and the pf with ontology curve (except at rank 5) is higher than the standard run. but the rfb baseline (except at rank 15 and 100) is lower than the standard run and the rfb with ontology curve is the same as the standard run except at ranks 15, 20, 30, 100 and 200 where it is only 0.001 lower. A full set of the TREC output figures for the standard run and the 15 document run can be seen in Appendix F (F1 and F10 respectively). The 15 document run (topic 101-150) has only 1 additional term which is unrest.

For topic 151-200 (Figure 7.15) from rank 5 onwards there is a 10% deterioration on pf baseline at rank 5. After rank 5 the performance deteriorates and matches the original pf run from rank 100 onwards. For rfb at rank 5 there is a 18% deterioration on the rfb baseline. The improvement on the rfb baseline takes place from rank 15 onwards. The pf and rfb t-test results did not show statistical significance.

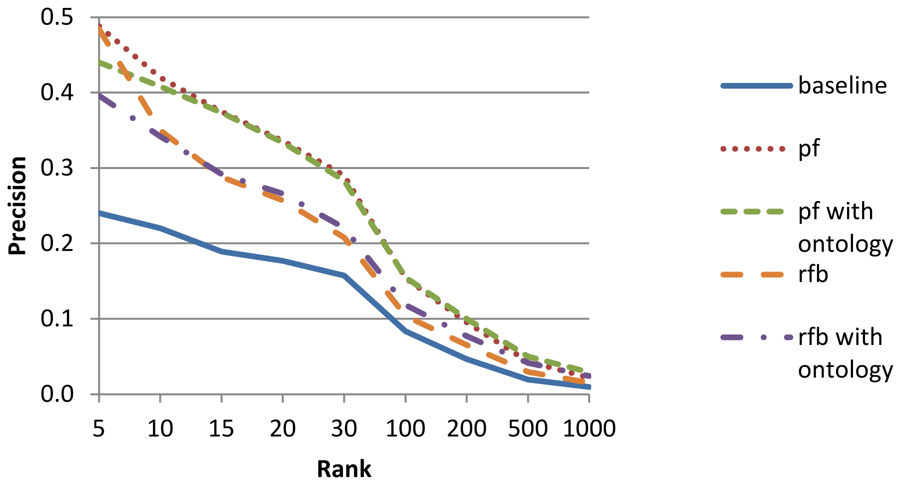


Figure 7.15 Topic 151-200 Results – Document Level Averages

For the 15 document run (topic 151-200), the pf baseline is higher than the standard run but the pf with ontology curve and the rfb baseline are lower than the standard run. The rfb with ontology curve is the same as the standard run and the rfb with ontology curve is the same as the standard run except at ranks 15, 20, 30, 100 and 200 where it is only 0.001 lower. A full set of the TREC output figures for the standard run and the 15 document run can be seen in Appendix F (F1 and F10 respectively). The 15 document run (topic 151-200) has only 2 additional terms namely health and lawyer.

For topic 201-250 (Figure 7.16) from rank 5 onwards there is a 8% deterioration on pf baseline at rank 5. After rank 5 the performance deteriorates and matches the original pf run from rank 30 onwards. For rfb at rank 5 there is a 16% deterioration on the rfb baseline. The improvement on the rfb baseline takes place from rank 100 onwards. The pf and rfb t-test results did not show statistical significance.

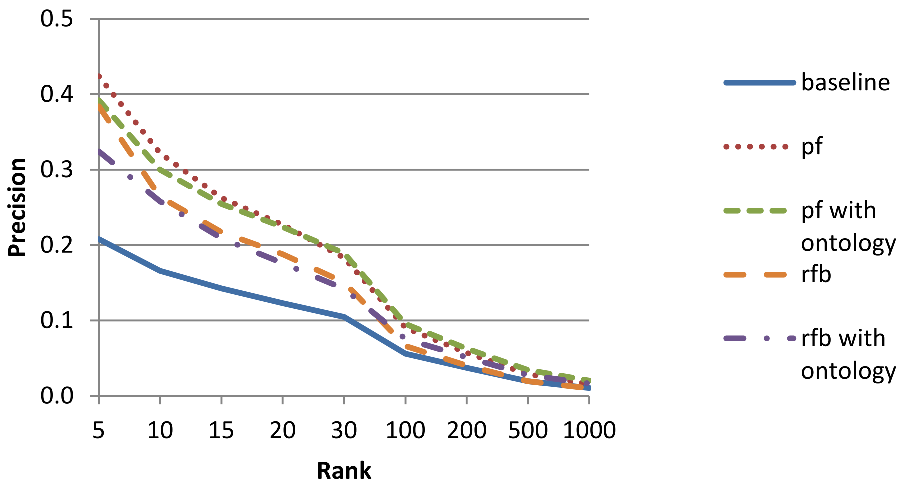


Figure 5.61 Topic 201-250 Results – Document Level Averages

For the 15 document run (topic 201-250), the pf baseline is higher and the pf with ontology curve (except ranks 10, 20, 30 and 100) are higher than the standard run. The rfb baseline is lower than the standard run. The rfb with ontology curve is the same as the standard run. A full set of the TREC output figures for the standard run and the 15 document run can be seen in Appendix F (F1 and F10 respectively). The 15 document run (topic 201-250) does not have any additional terms.

For topic 251-300 (Figure 7.17) from rank 5 onwards there is a 13% improvement on pf baseline at rank 5. After rank 15 the performance deteriorates and matches the original pf run from rank 200 onwards. For rfb at rank 5 there is a 7% deterioration on the rfb baseline. The improvement on the rfb baseline takes place from rank 10 onwards. The pf and rfb t-test results did not show statistical significance.

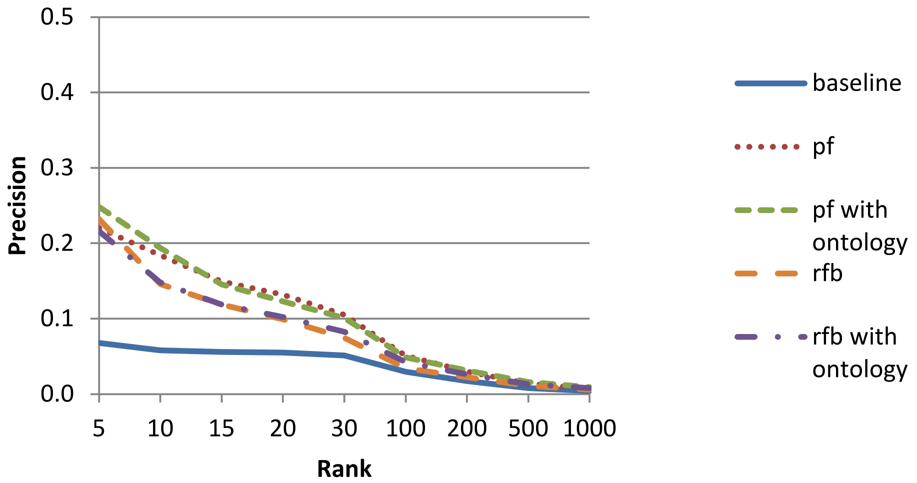


Figure 7.17 Topic 251-300 Results – Document Level Averages

For the 15 document run (topic 251-300), the pf baseline (except ranks 10, 20, 30 and 100), the pf with ontology curve (except ranks 5, 10 and 15) and the rfb baseline are lower than the standard run. The rfb with ontology curve is the same as the standard run. A full set of the TREC output figures for the standard run and the 15 document run can be seen in Appendix F (F1 and F10 respectively). The 15 document run (topic 251-300) only has 1 extra term which is punish.

7.4.2 Precision-Recall Graphs

At recall 0.0 (figure 7.18), there is a 8% deterioration on the pf baseline and a 23% deterioration on the rfb baseline. There is a consistent deterioration on the pf curve until recall 0.5 after which the two curves are identical. The pf t-test results for all topics are significant except for topic 251-300 where the results are very significant. There is an improvement for the rfb curve at recall 0.1 to 0.2 after which the two curves are the same. The rfb t-test score for this topic set is not significant. So query expansion has not proved to be of any benefit for the pf or the rfb runs presented in this section. The Precision-Recall graphs for the other topic sets are all very similar except for topic 251-300 where the Precision-Recall is slightly better than the pf baseline compared to the other topicsets. These graphs can be found in Appendix A10.

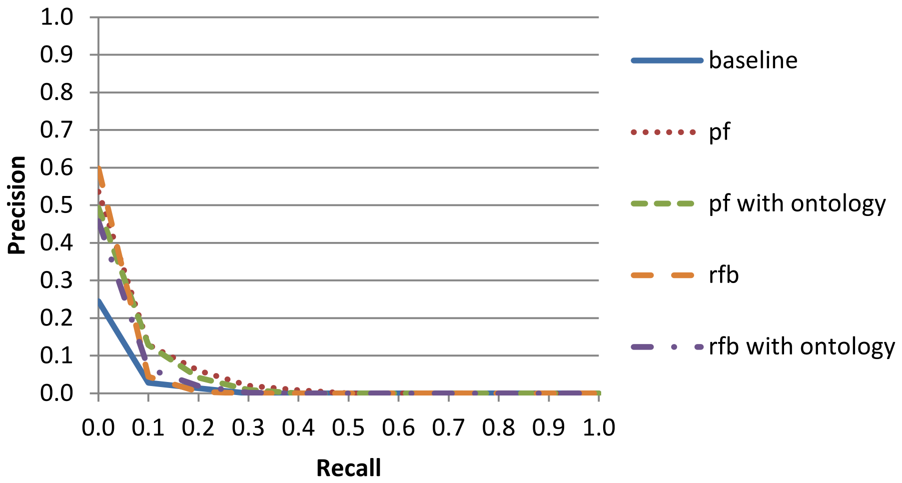


Figure 7.18 Topic 51-100 Results – Precision-Recall

In Table 7.6, for the pf run, none of the topics are significant. For the rfb run, only topic 101-150 is very significant.

Topics	PF	Rfb
51-100	0.174	0.237
101-150	0.084	0.010
151-200	0.177	0.409
201-250	0.189	0.193
251-300	0.247	0.271

Table 7.6: T-Test results (Document Level Averages)

In Table 7.7, for the pf run, the highlighted cells show that topic 251-300 is very significant and the remaining topics are significant. For the rfb run, none of the topics are significant.

Topics	PF	Rfb
51-100	0.046	0.252
101-150	0.018	0.131
151-200	0.033	0.086
201-250	0.052	0.197
251-300	0.007	0.351

Table 7.7: T-Test results (Precision-Recall)

7.5 Experiments Using 100 Documents And Standard 20 Terms And All Terms In Expanded Query

7.5.1 Document Level Averages Graphs

For topic 51-100 (Figure 7.19) from rank 5 onwards there is a 3% deterioration on pf baseline at rank 5. After rank 5 the performance improves on original pf baseline. For rfb at rank 5 there is a 40% deterioration on the rfb baseline. The improvement on the rfb baseline takes place from rank 500 onwards. The pf t-test and rfb results are very significant.

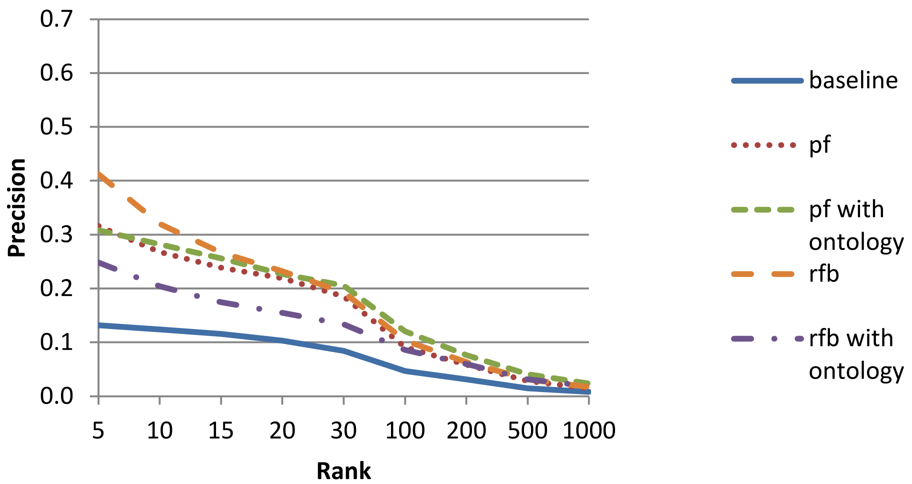


Figure 7.19 Topic 51-100 Results – Document Level Averages

For the 100 document run (topic 51-100), the pf baseline and the pf with ontology curve (except rank 100) is lower higher than the standard run. The rfb baseline is higher than the standard run but the rfb with ontology curve is the same as the standard run. A full set of the TREC output figures for the standard run and the 100 document run can be seen in Appendix F (F1 and F11 respectively). The 100 document run (topic 51-100) has the same terms as the 15 document runs and does not have any additional terms.

For topic 101-150 (Figure 7.20) from rank 5 onwards there is a 24% deterioration on pf baseline at rank 5. After rank 5 the performance deteriorates and matches the original pf run from rank 100 onwards.. For rfb at rank 5 there is a 40% deterioration on the rfb baseline. The improvement on the rfb baseline takes place from rank 500 onwards. The pf t-test are significant and the rfb results are very significant.

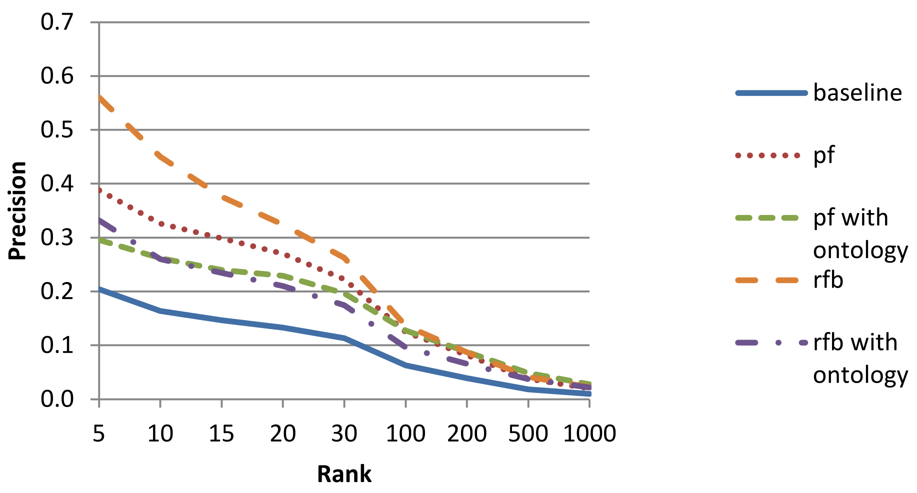


Figure 7.20 Topic 101-150 Results – Document Level Averages

For the 100 document run (topic 101-150), the pf baseline and the pf with ontology curve are lower than the standard run. The rfb baseline and the rfb with ontology curve (except ranks 5, 30, 100, 200 and 500) are higher than the standard run. A full set of the TREC output figures for the standard run and the 100 document run can be seen in Appendix F (F1 and F11 respectively). The 100 document run (topic 101-150) has the same terms as the 15 document runs and a few additional terms such as terror and organ.

For topic 151-200 (Figure 7.21) from rank 5 onwards there is a 16% deterioration on pf baseline at rank 5. After rank 5 the performance deteriorates and matches the original pf run from rank 15 onwards.. For rfb at rank 5 there is a 40% deterioration on the rfb baseline. The improvement on the rfb baseline takes place from rank 500 onwards. The pf t-test are not significant but the rfb results are significant.

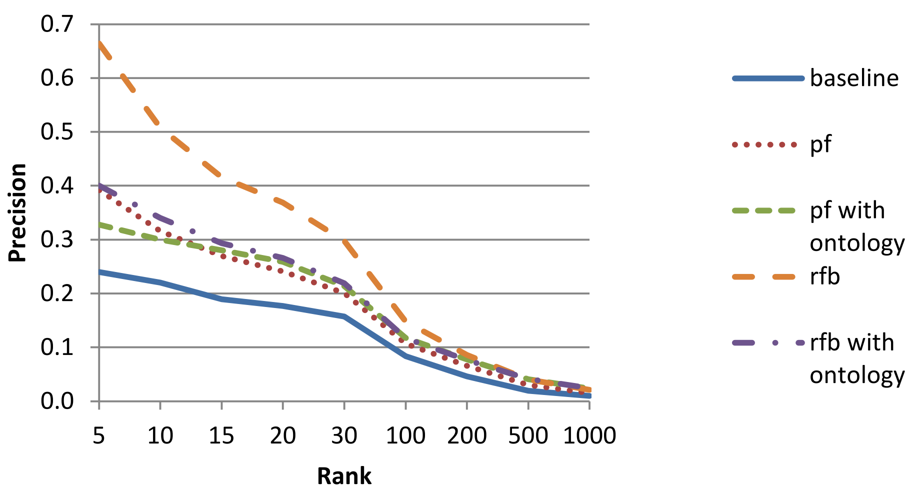


Figure 7.21 Topic 151-200 Results – Document Level Averages

For the 100 document run (topic 151-200), the pf baseline and the pf with ontology curve are lower than the standard run. The rfb baseline and the rfb with ontology curve (except ranks 10, 30 and 100) are higher than the standard run. A full set of the TREC output figures for the standard run and the 100 document run can be seen in Appendix F (F1 and F11 respectively). The 100 document run (topic 151-200) has the same terms as the 15 document runs and a few additional terms such as punishment and engineer.

For topic 201-250 (Figure 7.22) from rank 5 onwards there is a 4% deterioration on pf baseline at rank 5. After rank 5 the performance deteriorates and matches the original pf run from rank 200 onwards.. For rfb at rank 5 there is a 34% deterioration on the rfb baseline. The improvement on the rfb baseline takes place from rank 1000 onwards. The pf t-test are not significant but the rfb results are very significant.

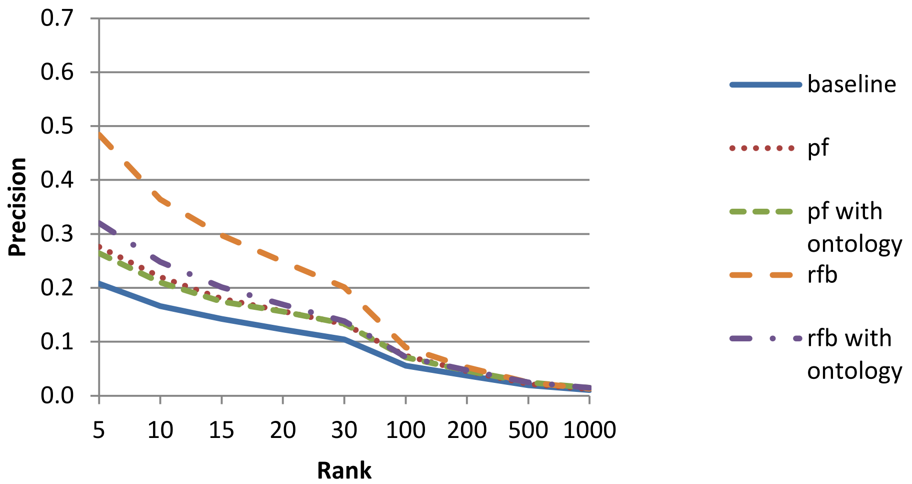


Figure 7.22 Topic 201-250 Results – Document Level Averages

For the 100 document run (topic 201-250), the pf baseline, pf with ontology curve and the rfb with ontology curve are lower than the standard run. However, the rfb baseline is higher than the standard run. A full set of the TREC output figures for the standard run and the 100 document run can be seen in Appendix F (F1 and F11 respectively). The 100 document run (topic 201-250) has the same terms as the 15 document runs and an additional term, namely punishment.

For topic 251-300 (Figure 7.23) from rank 5 onwards there is a 4% deterioration on pf baseline at rank 5. After rank 5 the performance deteriorates and matches the original pf run from rank 200 onwards.. For rfb at rank 5 there is a 34% deterioration on the rfb baseline. The improvement on the rfb baseline takes place from rank 1000 onwards. The pf and rfb t-test results are very significant.

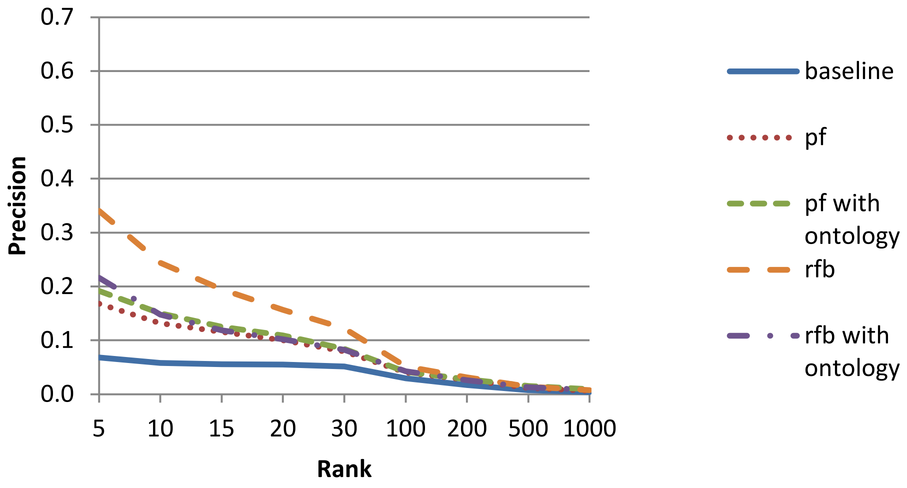


Figure 7.23 Topic 251-300 Results – Document Level Averages

For the 100 document run (topic 251-300), the pf baseline and the pf with ontology curve are lower than the standard run. However, the rfb baseline is higher than the standard run. The rfb with ontology curve is the same as the standard run. A full set of the TREC output figures for the standard run and the 100 document run can be seen in Appendix F (F1 and F11 respectively). The 100 document run (topic 251-300) has the fewer terms than the 5, 10 and 15 document runs.

7.5.2 Precision-Recall Graphs

At recall 0.0 (figure 7.24), there is a 8% deterioration on the pf baseline and a 40% deterioration on the rfb baseline. Between recall 0.1 and 0.2 there is an improvement and between recall .3 and .4 there is a deterioration. After recall 0.5 the two curves are identical. The pf t-test results for all topics are not significant except for topic 201-250 where the results are significant. There is an improvement for the rfb curve at recall 0.2 after which the two curves are the same. The rfb t-test score for this topic set is not significant. So query expansion has not proved to be of any benefit for the pf or the rfb runs presented in this section. The Precision-Recall graphs for the other topic sets are all very similar and can be found in Appendix A11.

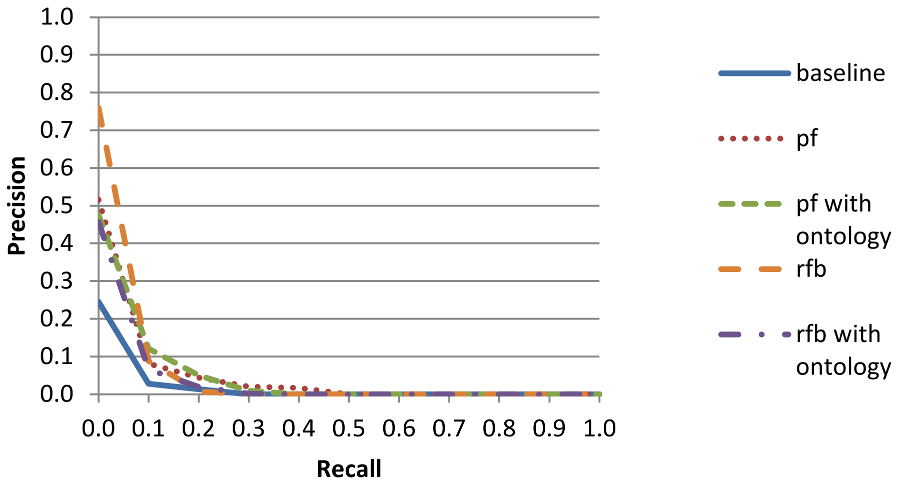


Figure 7.24 Topic 51-100 Results – Precision-Recall

In Table 7.8, for the pf run, topic 51-100 and topic 251-300 are very significant and topic 101-150 is significant. For the rfb run, all cells are highlighted showing that all topics are very significant.

Topics	PF	Rfb
51-100	0.002	0.009
101-150	0.027	0.005
151-200	0.480	0.011
201-250	0.078	0.008
251-300	0.008	0.009

Table 7.8: T-Test results (Document Level Averages)

In Table 7.9, for the pf run, only topic 201-250 is significant. For the rfb run, onlyttopic 151-200 is significant.

Topics	PF	Rfb
51-100	0.361	0.164
101-150	0.244	0.110
151-200	0.120	0.046
201-250	0.024	0.096
251-300	0.214	0.156

Table 7.9: T-Test results (Precision-Recall)

7.6 Experiments Using 200 Documents And Standard 20 Terms And All Terms In Expanded Query

7.6.1 Document Level Averages Graphs

For topic 51-100 (Figure 7.25) from rank 5 onwards there is a 13% deterioration on pf baseline at rank 5. After rank 5 the performance improves on original pf baseline. For rfb at rank 5 there is a 46% deterioration on the rfb baseline and there is no improvement at any of the ranks. The pf t-test and rfb results are very significant.

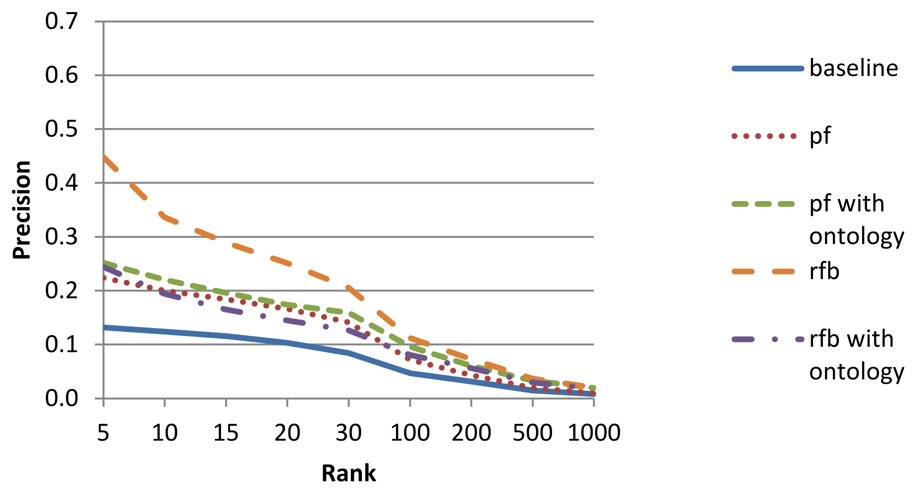


Figure 7.25 Topic 51-100 Results – Document Level Averages

For the 200 document run (topic 51-100), the pf baseline, pfb with ontology curve and the rfb with ontology curve are lower than the standard run. However, the rfb baseline is higher than the standard run. A full set of the TREC output figures for the standard run and the 200 document run can be seen in Appendix F (F1 and F12 respectively). The 200 document run (topic 51-100) has the same terms as the 100 document runs with an additional term intern.

For topic 101-150 (Figure 7.26) from rank 5 onwards there is a 14% deterioration on pf baseline at rank 5. After rank 5 the performance improves on original pf baseline. For rfb at rank 5 there is a 44% deterioration on the rfb baseline and there is no improvement at any of the ranks. The pf t-test results are significant and rfb results are very significant.

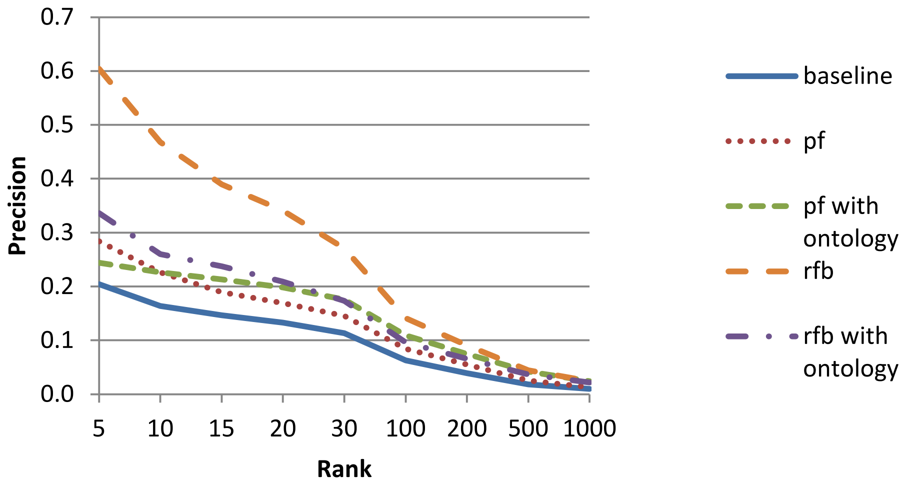


Figure 7.26 Topic 101-150 Results – Document Level Averages

For the 200 document run (topic 101-150), the pf baseline, pfb with ontology curve and the rfb with ontology curve (except at rank 10, 15 and 20) are lower than the standard run. However, the rfb baseline is higher than the standard run. A full set of the TREC output figures for the standard run and the 200 document run can be seen in Appendix F (F1 and F12 respectively). The 200 document run (topic 101-150) has the same terms as the 100 document runs.

For topic 151-200 (Figure 7.27) from rank 5 onwards there is a 13% deterioration on pf baseline at rank 5. After rank 5 the performance improves on original pf baseline. For rfb at rank 5 there is a 39% deterioration on the rfb baseline and there is no improvement at any of the ranks except for rank 1000. The pf t-test results are significant and rfb results are very significant.

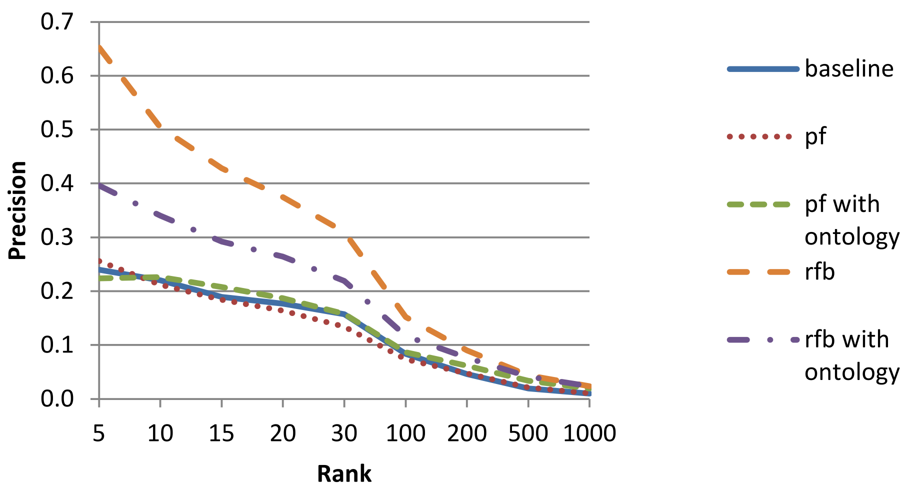


Figure 7.27 Topic 151-200 Results – Document Level Averages

For the 200 document run (topic 151-200), the pf baseline, pfb with ontology curve and the rfb with ontology curve are lower than the standard run. However, the rfb baseline is higher than the standard run. A full set of the TREC output figures for the standard run and the 200 document run can be seen in Appendix F (F1 and F12 respectively). The 200 document run (topic 151-200) has the same terms as the standard run with the addition of euthanasia.

For topic 201-250 (Figure 7.28) from rank 5 onwards there is a 5% deterioration on pf baseline at rank 5. After rank 5 the performance improves on original pf baseline. For rfb at rank 5 there is a 46% deterioration on the rfb baseline and there is no improvement at any of the ranks. The pf and rfb t-test results are very significant.

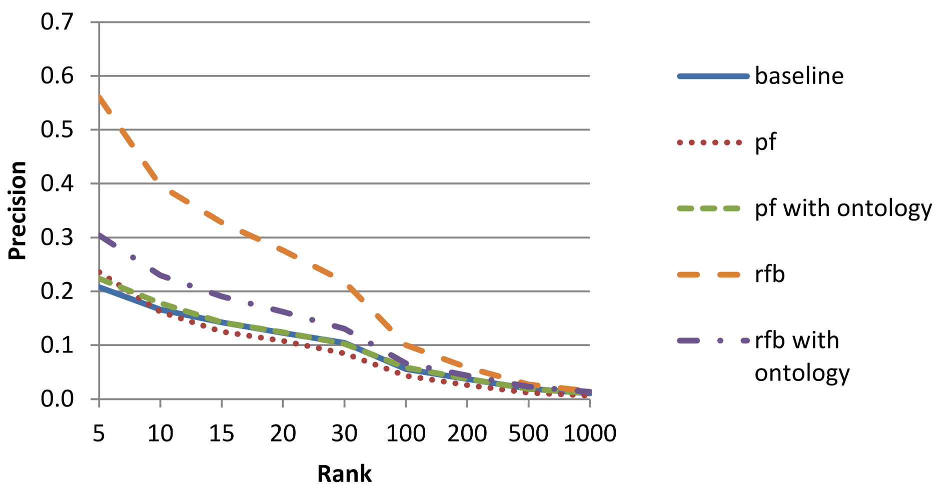


Figure 7.28 Topic 201-250 Results – Document Level Averages

For the 200 document run (topic 201-250), the pf baseline, pfb with ontology curve and the rfb with ontology curve are lower than the standard run. However, the rfb baseline is higher than the standard run. A full set of the TREC output figures for the standard run and the 200 document run can be seen in Appendix F (F1 and F12 respectively). The 200 document run (topic 201-250) has the same terms as the 100 document run.

For topic 251-300 (Figure 7.29) at rank 5 the two pf curves are identical. After rank 5 the performance improves on original pf baseline. For rfb at rank 5 there is a 51% deterioration on the rfb baseline and there is no improvement at any of the ranks until rank 1000. The pf and rfb t-test results are very significant.

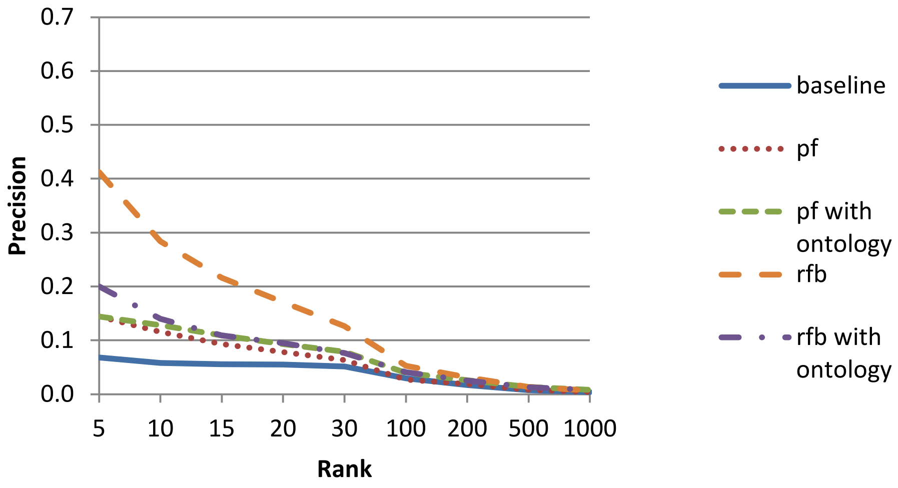


Figure 7.29 Topic 251-300 Results – Document Level Averages

For the 200 document run (topic 251-300), the pf baseline, pfb with ontology curve and the rfb with ontology curve are lower than the standard run. However, the rfb baseline is higher than the standard run. A full set of the TREC output figures for the standard run and the 200 document run can be seen in Appendix F (F1 and F12 respectively). The 200 document run (topic 251-300) has some terms the same as the standard runs and some additional terms dea, weather and organ.

7.6.2 Precision-Recall Graphs

At recall 0.0 (figure 7.30), there is a 7% deterioration on the pf baseline and a 34% deterioration on the rfb baseline. Between recall 0.1 and 0.2 there is an improvement and between recall .3 and .4 there is a deterioration. After recall 0.5 the two curves are identical. The pf t-test results for all topics are not significant except for topic 251-300 where the results are significant. There is an improvement for the rfb curve at recall 0.2, then the performance drops at recall 0.3 and from rank 0.4 onwards the two curves are the same. The rfb t-test score for this topic set is not significant. So query expansion has not proved to be of any benefit for the pf or the rfb runs presented in this section. The Precision-Recall graphs for the other topic sets are all very similar except for topic 251-300 where the pf with ontology is better than the pf baseline. These graphs can be found in Appendix A12.

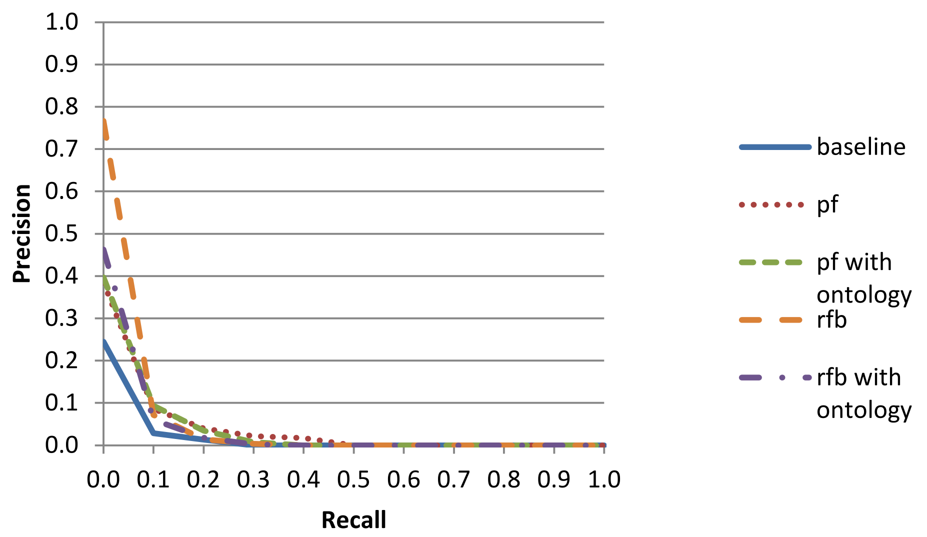


Figure 7.30 Topic 51-100 Results – Precision-Recall

In Table 7.10, for the pf run, topic 151-200 is significant; topic 51-100, topic 201-250 and topic 251-300 are very significant. For the rfb run, all cells have been highlighted to show that all topics are very significant.

Topics	PF	Rfb
51-100	0.000	0.005
101-150	0.058	0.005
151-200	0.042	0.007
201-250	0.006	0.007
251-300	0.000	0.013

Table 7.10: T-Test results (Document Level Averages)

In Table 7.11, for the pf run, only topic 251-300 is significant. For the rfb run, only topic 151-200 is significant.

Topics	PF	Rfb
51-100	0.284	0.163
101-150	0.454	0.118
151-200	0.060	0.040
201-250	0.350	0.056
251-300	0.031	0.144

Table 7.11: T-Test results (Precision-Recall)

7.7 Summary

In this chapter we have presented the results for the experiments in which the number of documents relevance feedback parameter has been varied.

Our findings show that only the rfb runs for experiments which use number of documents 5, 10 and 15 retrieve a higher number of relevant documents compared to the standard run. However, when the number of documents parameter is set to 100 and 200 then it is the pf runs which retrieve more relevant documents instead of the rfb runs. So there doesn't seem to be any clear trend or correlation between the number of documents parameter and the increase in the number of relevant documents retrieved.

The results do not give a clear picture of which particular topic-set benefits the most from ontology based query expansion.

For the precision at rank results, the use of ontology based query expansion has improved the pf with ontology and rfb with ontology curves in the standard run. However, varying the number of documents parameter has a more complex effect. The pf and pf with ontology curves are more sensitive to varying the number of terms relevance feedback parameter and the rfb with ontology curve does not fluctuate. Whereas varying the number of documents parameter also affects the rfb baseline and the rfb with ontology curve resulting in degradation / improvement across the different runs.

Finally for the precision-recall results, the ontology produces a slight improvement on the pf curve where the number of documents relevance feedback parameter has been set to 5, 10 and 15 (see Appendix F – Tables F8, F9 and F10. However when this parameter is set to 100 and 200 documents, it is the rfb curve which has a slight improvement on the rfb curve in the standard run (see Appendix F – Tables F11 and F12).

Although our experiments have shown that use of our chosen ontology for query expansion has resulted in some improvements, there may be some issues which might be preventing the widespread use of ontologies in query expansion. The success rate depends on the level of similarity between the ontology and the document collection. A common problem is that of query drift which is caused by ontology terms which retrieve non-relevant documents. Usually ontologies which are derived from a document collection result in improved retrieval. However as the document collection changes over a period of time, constant effort is required to maintain the ontology in order to reflect these changes. Success also depends on the domain coverage in terms of breadth and depth. Lastly its possible that the ontologies for a particular domain might be virtually non-existent in which case alot of time and effort will need to be invested into creating a new ontology from scratch.

8. DISCUSSION OF SINGLE VALUE RESULTS

8.1 Introduction

In this chapter, the single value results (recall, map and bpref) for pf , pf with ontology, rfb and rfb with ontology can be compared across all topic sets for different relevance feedback parameters.

8.2 Recall results

Appendix G1 shows the recall results by topic set for all the variations of the number of terms parameter. We will analyse each topic in turn.

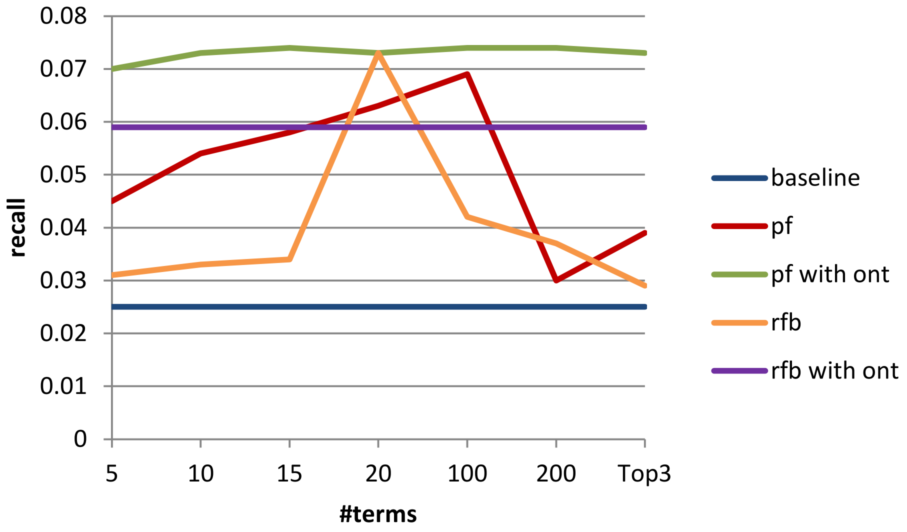


Figure 8.1 Recall results for Topic 51-100

For topic 51-100, the pf with ontology curve has the highest recall across all the different variations of the number of terms parameter (figure 8.1) whereas the rfb with ontology curve has lower recall. For the pf curve, 100 terms seems to produce the optimum recall and for the rfb curve, 20 terms produces the optimum results. However the pf curve dips at 200 terms.

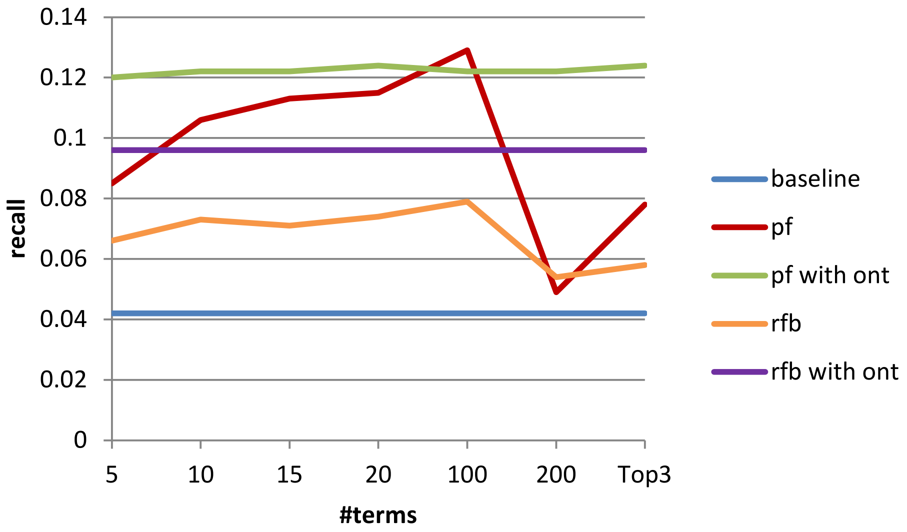


Figure 8.2 Recall results for Topic 101-150

For topic 101-150, the pf with ontology curve has the highest recall across all the different variations of the number of terms parameter (figure 8.2) whereas the rfb with ontology curve has lower recall. For the pf and rfb curves, 100 terms produces the optimum results. However the pf and rfb curves dip at 200 terms.

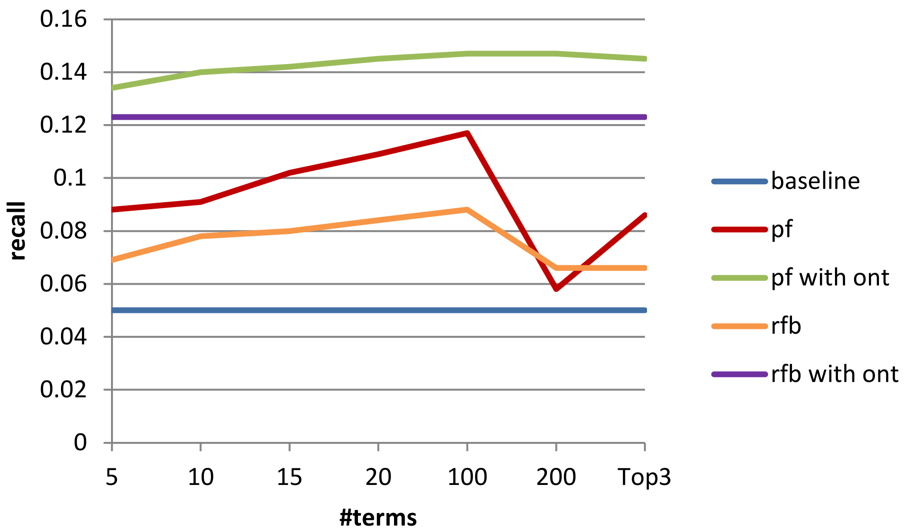


Figure 8.3 Recall results for Topic 151-200

For topic 151-200, the pf with ontology curve has the highest recall across all the different variations of the number of terms parameter (figure 8.3) whereas the rfb with ontology curve has lower recall. For the pf and rfb curves, 100 terms produces the optimum results. However the pf and rfb curves dip at 200 terms.

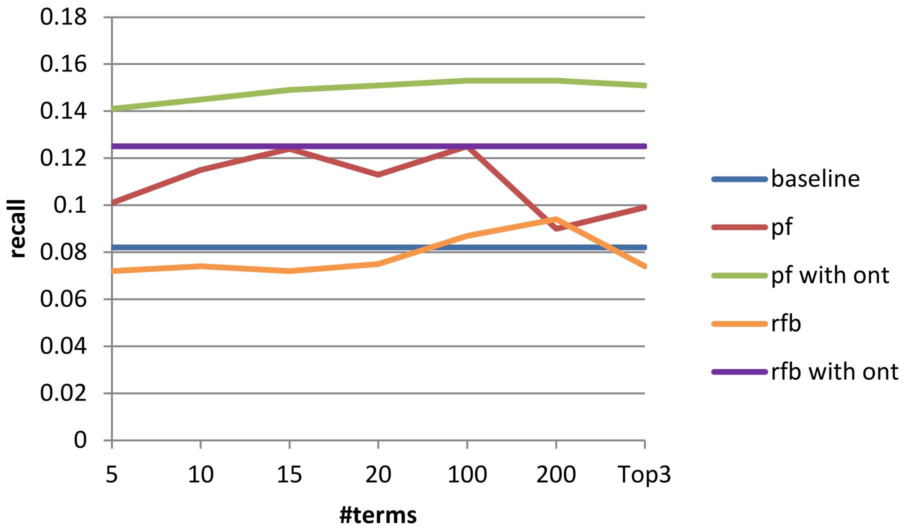


Figure 8.4 Recall results for Topic 201-250

For topic 201-250, the pf with ontology curve has the highest recall across all the different variations of the number of terms parameter (figure 8.4) whereas the rfb with ontology curve has lower recall. For the pf curve 100 terms produces the optimum results. Up to rank 20 the rfb curve performs worse than the baseline. After 20 terms, the rfb curve improves and peaks at 200 terms. However the pf curve dips at 200 terms.

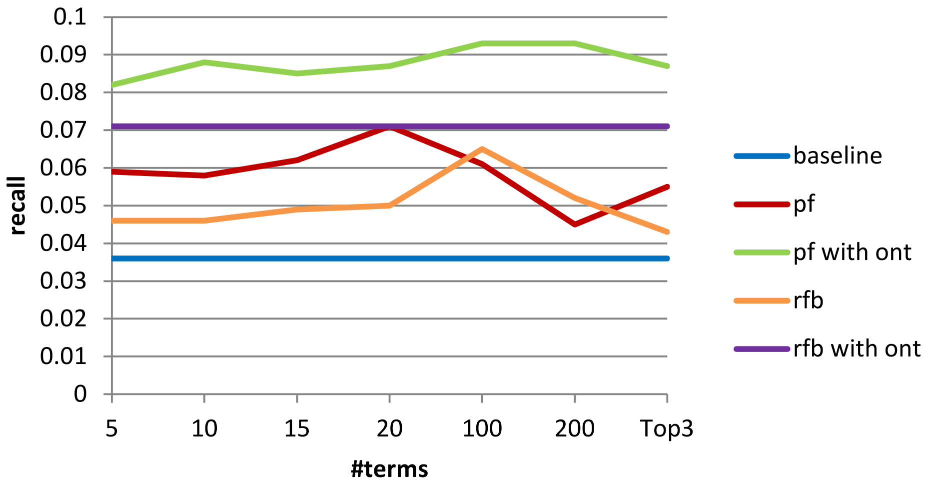


Figure 8.5 Recall results for Topic 251-300

For topic 251-300, the pf with ontology curve has the highest recall across all the different variations of the number of terms parameter (figure 8.5) whereas the rfb with ontology curve has lower recall. For the pf curve, 20 terms produces the optimum recall results and for the rfb curve, 100 terms produces the optimum results. However the pf curve dips at 200 terms.

Table 8.1 shows the t-test results for recall figures. In the pf run, the highlighted cells show that the t-test results for all the runs are very significant except for 100 terms which are significant. In the rfb run, the t-test results for all the runs are very significant except for 20 terms which are significant.

#terms	Pf	Rfb
5	0.001	0.002
10	0.004	0.002
15	0.006	0.003
20	0.013	0.049
100	0.044	0.010
200	0.001	0.004
Top3	0.000	0.001

Table 8.1 Ttest (Recall)

Appendix G2 shows the recall results by topic-set and by varying the number of documents parameter. In the next few sections we will analyse each topic in turn.

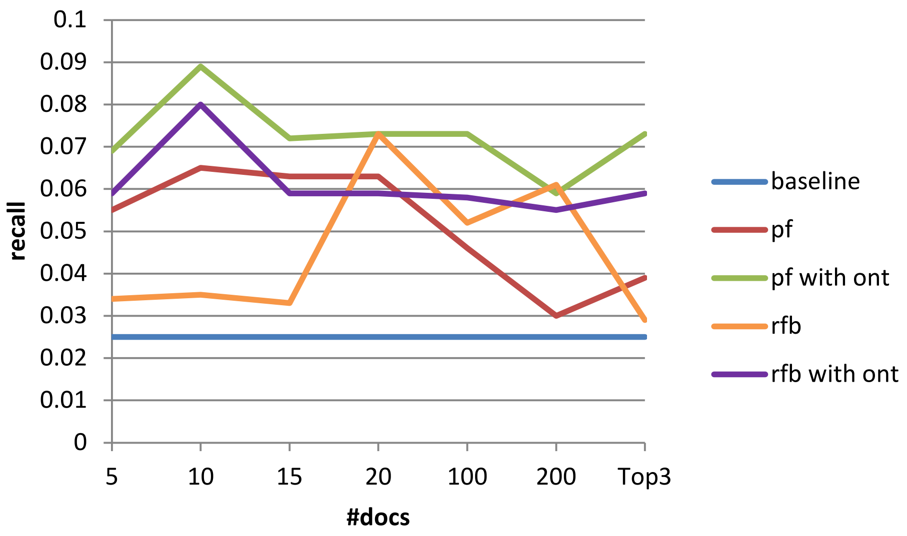


Figure 8.6 Recall results for Topic 51-100

For topic 51-100, the pf with ontology curve has the highest recall across all the different variations of the number of terms parameter (figure 8.6) whereas the rfb with ontology curve has lower recall. Unlike the graphs for recall where the number of terms parameter has been varied, the pf and rfb ontology curves have more peaks and troughs. For the pf curve, 10 documents produces the optimum recall results and for the rfb curve, 20 documents produces the optimum results. However the pf and pf with ontology curves both dip at 200 documents.

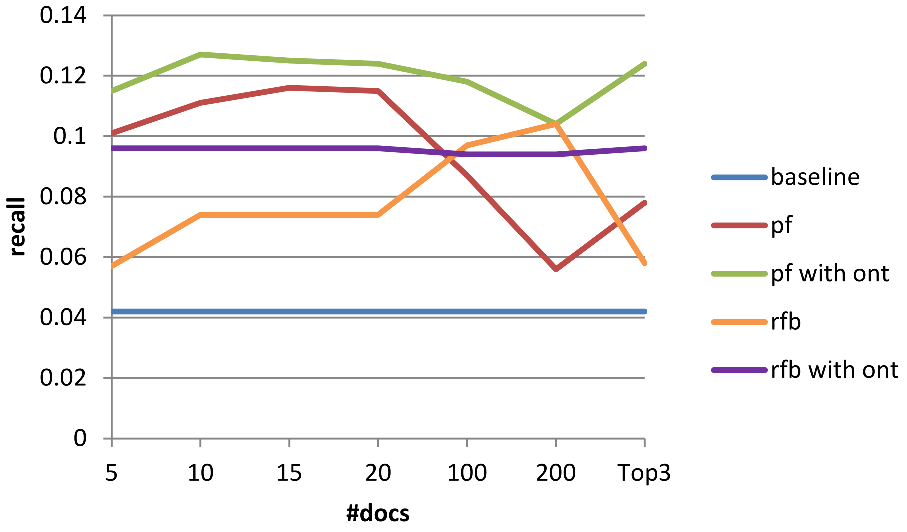


Figure 8.7 Recall results for Topic 101-150

For topic 101-150, the pf with ontology curve has the highest recall across all the different variations of the number of documents parameter (figure 8.7) whereas the rfb with ontology curve has lower recall and is more constant. For the pf curve, 20 documents produces the optimum recall results and for the rfb curve, 200 documents produces the optimum results. However the pf and pf with ontology curves both dip at 200 documents.

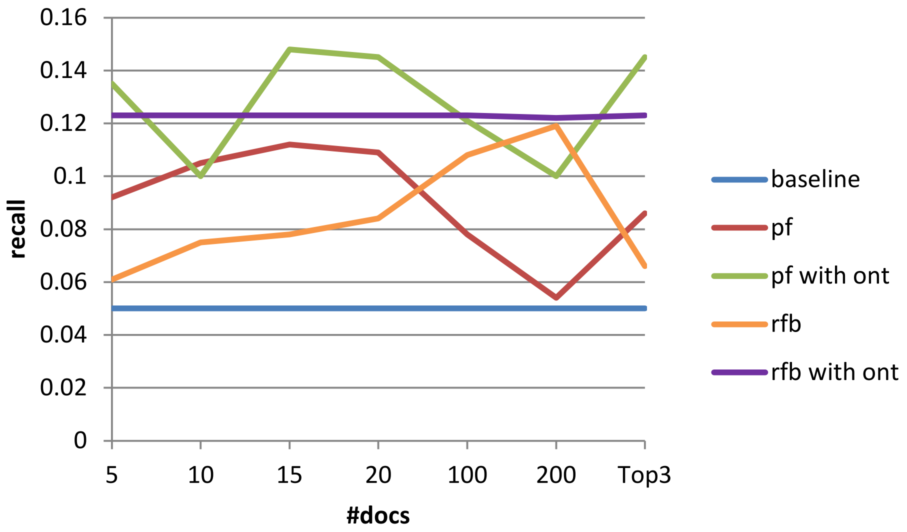


Figure 8.8 Recall results for Topic 151-200

For topic 151-200, the pf with ontology curve has the highest recall across all the different variations of the number of documents parameter (figure 8.8) and peaks at 15 documents and top3 terms. Whereas the rfb with ontology curve has lower recall. For the pf curve, 15 documents produces the optimum recall results and for the rfb curve, 200 documents produces the optimum results. However the pf and pf with ontology curves both dip at 200 documents. Unlike the other graphs where the rfb with ontology curve is flat, this is the first time when the rfb with ontology curve also dips at 200 documents. This is similar to the findings of MacFarlane et al (2010) whereby increasing the relevance feedback parameters can result in performance deterioration.

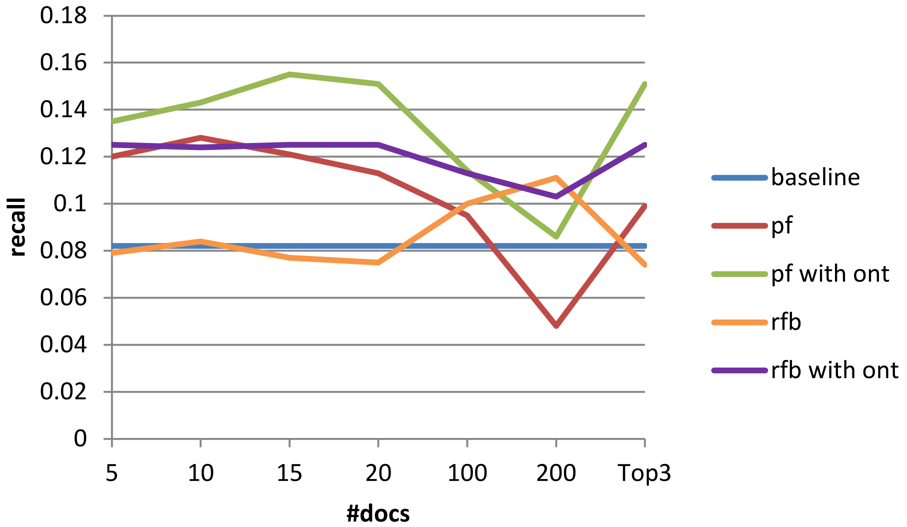


Figure 8.9 Recall results for Topic 201-250

For topic 201-250, the pf with ontology curve has the highest recall across all the different variations of the number of documents parameter (figure 8.9) and peaks at 15 documents and top3 terms. Whereas the rfb with ontology curve has lower recall. For the pf curve, 15 documents produces the optimum recall results and for the rfb curve, 200 documents produces the optimum results. Both the pf curve and the pf with ontology curve dip at 200 documents.

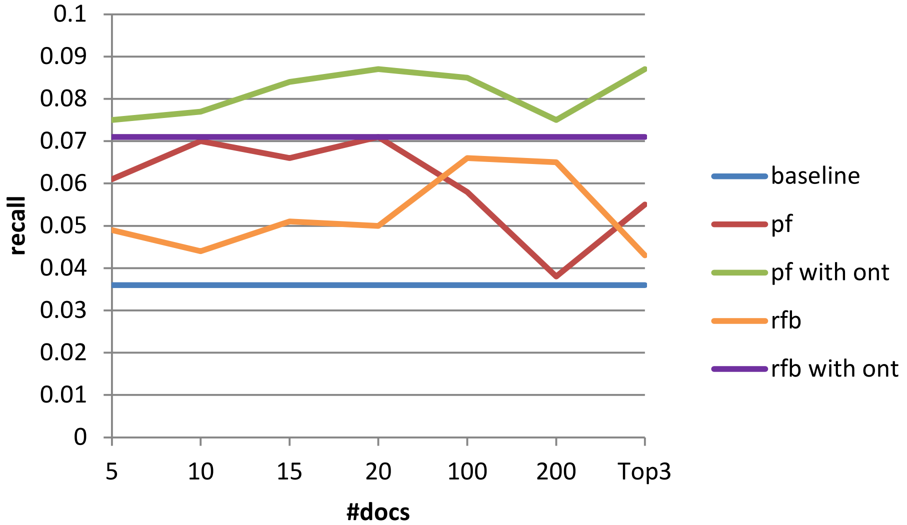


Figure 8.10 Recall results for Topic 251-300

For topic 251-300, the pf with ontology curve has the highest recall across all the different variations of the number of documents parameter (figure 8.10) and peaks at 20 documents and top3 documents. Whereas the rfb with ontology curve has lower recall.

For the pf curve, 10 and 20 documents produces the optimum recall results and for the rfb curve, 100 documents produces the optimum results

Table 8.2 shows the ttest results for for recall figures (no. of documents). In the pf run, the highlighted cells show that the t-test results for all the runs are very significant except for 10 documents which are significant. In the rfb run, the t-test results for 5, 10, 15, and top3 runs are very significant. The t-test results for 100 documents is significant. The t-test result for 200 documents is not significant.

#docs	Pf	rfb
5	.013	.003
10	.040	.001
15	.011	.003
20	.013	.049
100	.001	.044
200	.000	.198
Top3	.000	.001

Table 8.2 Ttest (Recall)

8.3 MAP Results

Appendix G3 shows the MAP results by topic-set and by varying the number of terms parameter. We will examine each topic in turn.

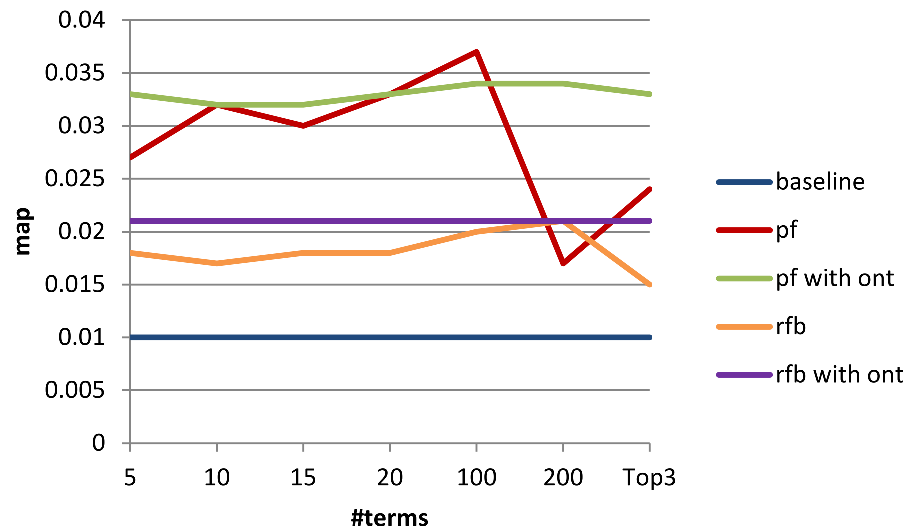


Figure 8.11 MAP results for Topic 51-100

For topic 51-100, the pf curve has the highest map across all the different variations of the number of terms parameter (figure 8.11) and peaks at 100 terms. Whereas the pf with ont curve has the next highest map and is more constant. Compared to other graphs, where the pf curve has better performance than the pf with ontology curve, this is the only graph where the pf curve with ontology curve in some places has better performance than the pf curve for number of terms 5 and 15. Chapter 6 shows that the pf with ontology curve has highest precision for topic 51-100. It follows that averaging these scores is going to result in a similarly high MAP. For the rfb curve, 200 terms produces the optimum recall results and for the rfb with ontology curve, the performance is constant across all different variations of the number of terms parameter at map of 0.02.

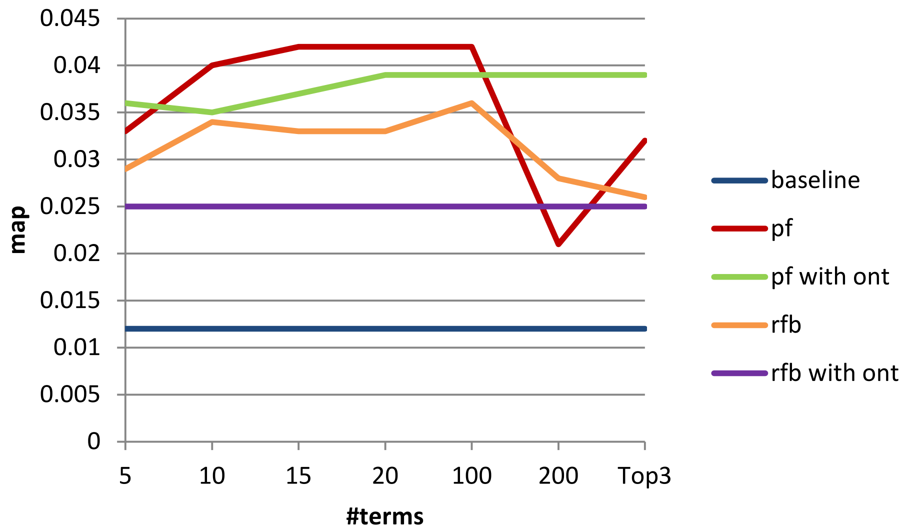


Figure 8.12 MAP results for Topic 101-150

For topic 101-150, the pf curve has the highest map across all the different variations of the number of terms parameter (figure 8.12) and 15, 20 and 100 terms produce the same peak. Whereas the pf with ont curve has the next highest map and is more constant from 20 terms onwards. For the rfb curve, 100 terms produces the optimum recall results and for the rfb with ontology curve, the performance is constant across all different variations of the number of terms parameter at map of 0.025.

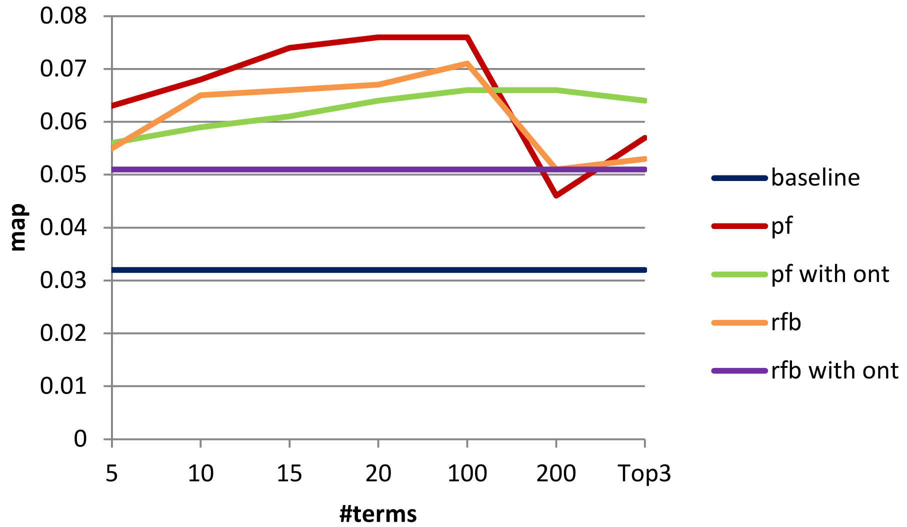


Figure 8.13 MAP results for Topic 151-200

For topic 151-200, the pf curve has the highest map across all the different variations of the number of terms parameter (figure 8.13) and peaks at 200 terms. Unlike the previous graphs, the rfb curve has the next highest map at 200 terms.This is followed by the pf with ontology curve which is more stable. For the rfb with ont curve, the performance is constant across all different variations of the number of terms parameter at map of 0.05.

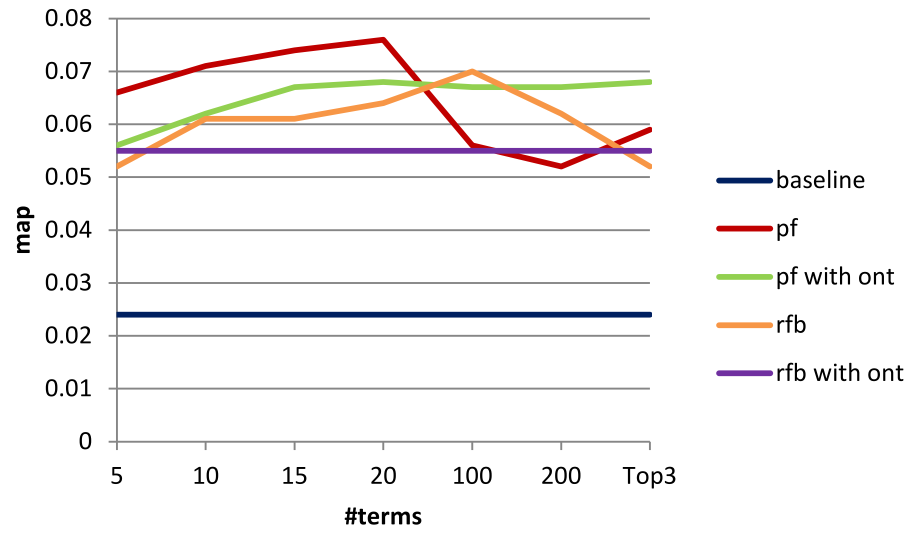


Figure 8.14 MAP results for Topic 201-250

For topic 201-250, the pf curve has the highest map across all the different variations of the number of terms parameter (figure 8.14) and peaks at 20 terms. The pf with ontology curve is lower but more stable. The rfb curve has the next highest map at 200 terms. For the rfb with ont curve, the performance is constant across all different variations of the number of terms parameter at map of 0.055.

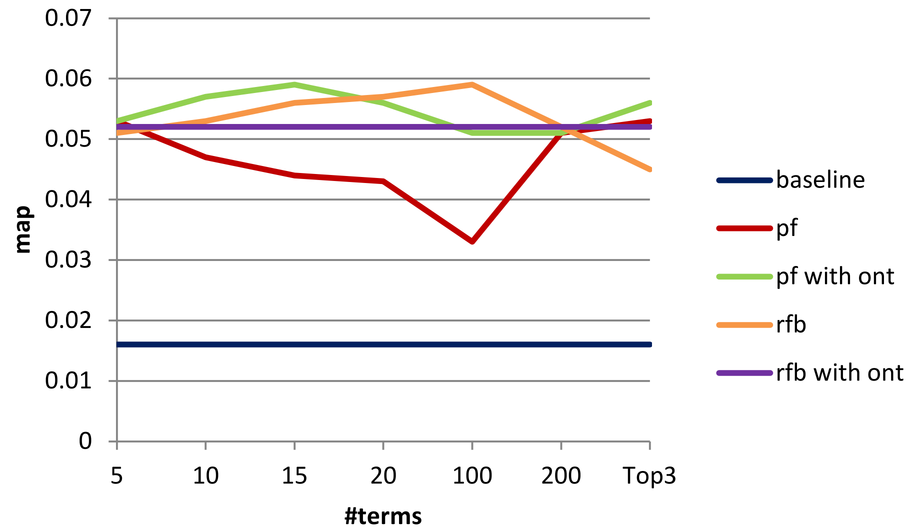


Figure 8.15 MAP results for Topic 251-300

For topic 251-300, the rfb curve peaks at 200 terms. The pf with ontology overlaps with the rfb curve. The rfb with ont curve has the constant map of 0.05 across all terms and the pf curve has the lowest map. In the pf run, the t-test results 20 terms, 200 terms and top 3 are very significant all remaining runs are not significant. In the rfb run, the highlighted cells show that the t-test results for 20 and 100 terms are significant, top3 is very significant, all remaining runs are not significant (see Table 8.3).

#terms	Pf	Rfb
5	0.309	0.389
10	0.264	0.076
15	0.374	0.053
20	0.013	0.049
100	0.333	0.022
200	0.008	0.123
Top3	0.000	0.001

Table 8.3 Ttest (MAP)

Appendix G4 shows the MAP results by topic-set and by varying the number of documents parameter. We will plot the results for each topic on a graph and analyse each topic.

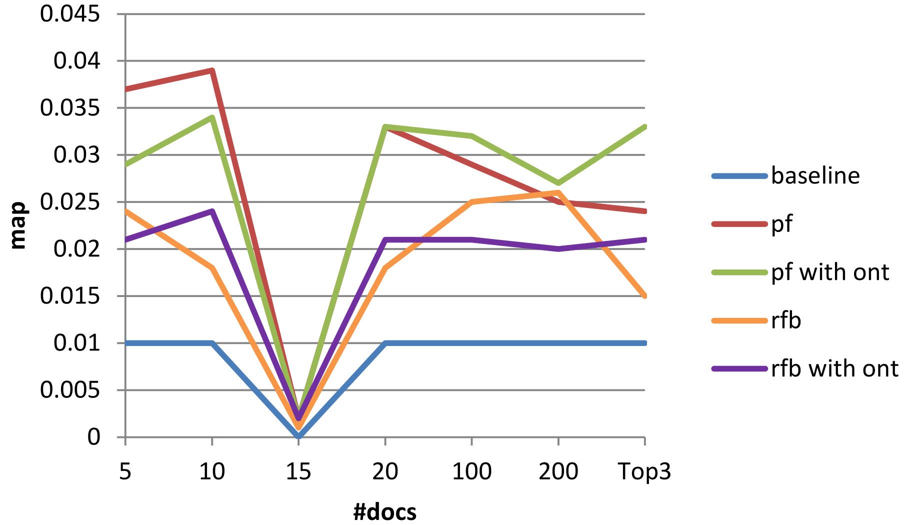


Figure 8.16 MAP results for Topic 51-100

This graph is very unusual, the map for all curves drops sharply for the 10-20 doc runs. It is difficult to say what has caused this but the process of judging the relevance of a document is subjective in nature. We would expect the person who formulates the query topic and the person who makes the relevance judgement to be the same. However this is not always the case and Voorhees (2000) found that assessors who formulated the query have a tendency to judge more documents relevant than cases where the assessor did not formulate the query topic. Also Ruthven et al (2003) point out the questions which arise during relevance feedback. For example what percentage of the document needs to be relevant before it can be marked as relevant? Is relevance to the assessor more important than relevance to the topic? So this could be one of those instances where the results have been obscured. After 20 terms, pf with ontology has the highest performance and the rfb curve is better than the rfb with ontology curve.

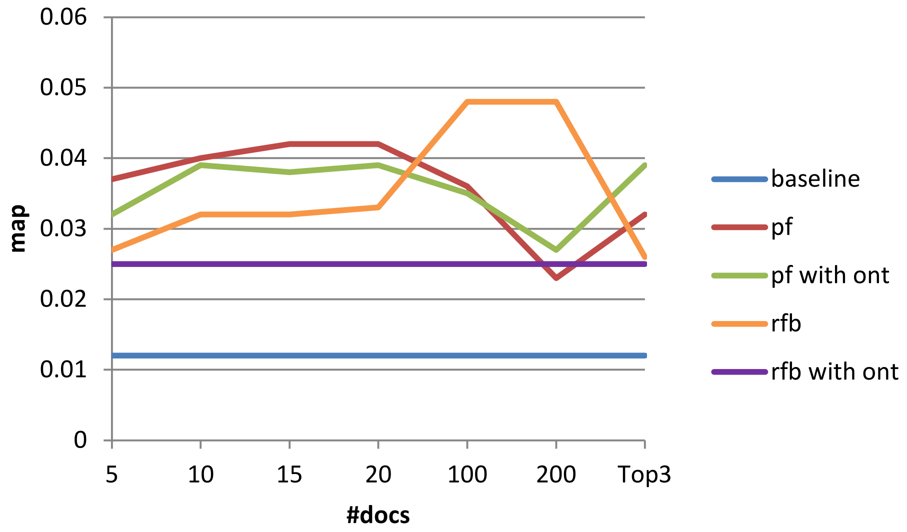


Figure 8.17 MAP results for Topic 101-150

For topic 101-150, the rfb curve has the highest map and peaks at 100 documents and 200 documents. This is the first time when the rfb curve has performed better than the pf with ontology curve. However, the pf and pf with ontology curves both dip to a low at 200 documentss. For the rfb with ont curve, the performance is constant across all different variations of the number of documents parameter at map of 0.025.

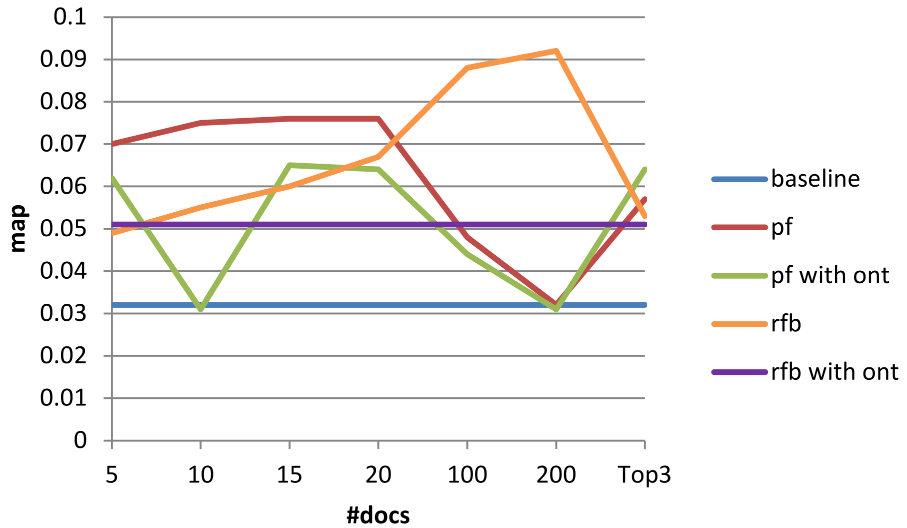


Figure 8.18 MAP results for Topic 151-200

For topic 151-200, the rfb curve has the highest map across all the different variations of the number of documents parameter (figure 8.11) and peaks at 200 documents. However, the pf the pf and pf with ontology curves both dip to a low of 0.03 at 200 documents. For the rfb with ont curve, the performance is constant across all different variations of the number of documents parameter at map of 0.05.

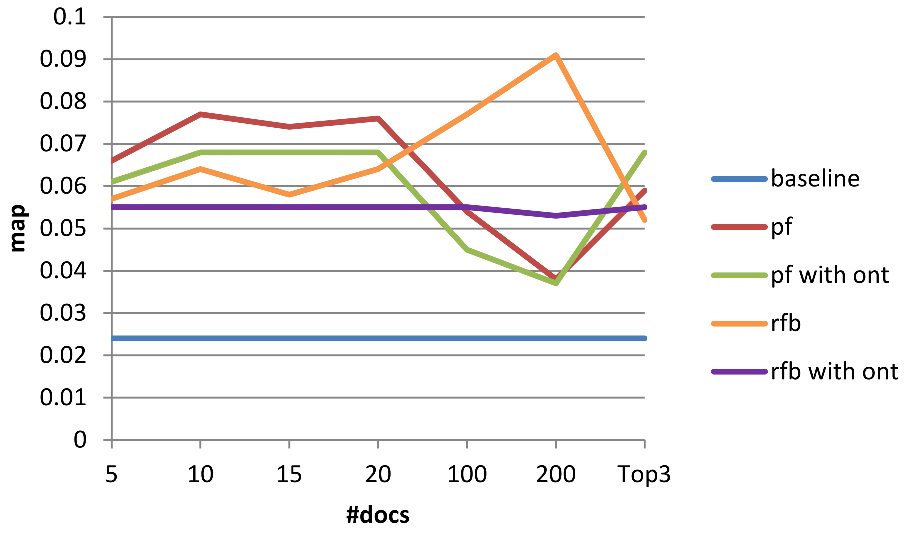


Figure 8.19 MAP results for Topic 201-250

For topic 201-250, the rfb curve has the highest map and peaks at 200 documents. However, the pf the pf and pf with ontology curves both dip to a low of 0.03 at 200 documents. For the rfb with ont curve, the performance is constant across all different variations of the number of documents parameter at map of 0.055.

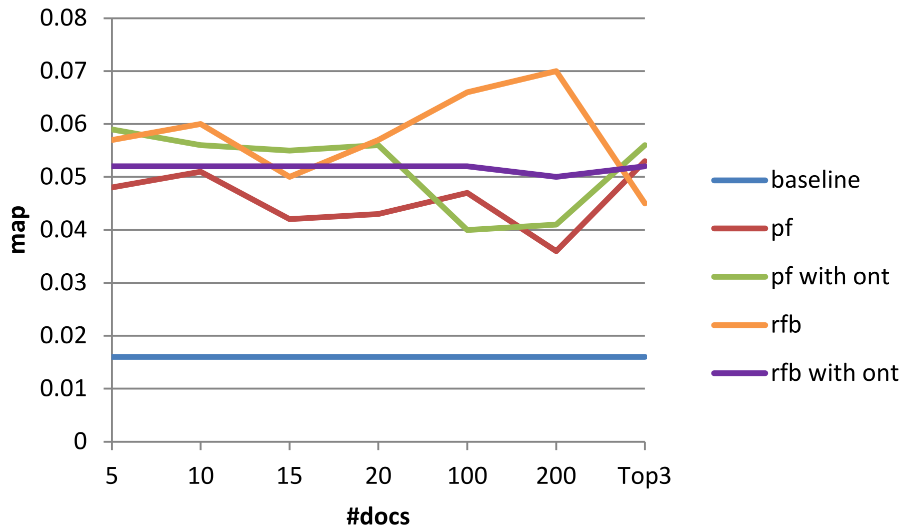


Figure 8.20 MAP results for Topic 251-300

For topic 251-300, the rfb curve has the highest map across all the different variations of the number of documents parameter (figure 8.11) and peaks at 200 documents. However, the pf dips to a low at 200 documents. The pf with ontology curve performs better than the pf curve and the rfb with ontology performs at a constant level between the pf and pf with ontology curve.

Table 8.4 shows the t-test results for for MAP figures (no. of documents). In the pf run, only the results for Top3 run are very significant. In the rfb run, the highlighted cells show that the t-test results for runs with 20 documents are significant, and the runs for 100 and 200 documents are very significant. The remaining 4 runs are not significant.

#docs	Pf	Rfb
5	.205	.078
10	.131	.095
15	.357	.089
20	.276	.039
100	.079	.011
200	.132	.008
Top3	.001	.113

Table 8.4 Ttest (MAP)

The t-test results for MAP (Table 8.3 and Table 8.4) contain fewer significant / very significant results compared to the t-test results for Recall (table 8.1 and table 8.2)

8.4 BPref results

Appendix G5 shows the BPref results by topic-set and by varying the number of terms parameter. We will examine each topic in turn.

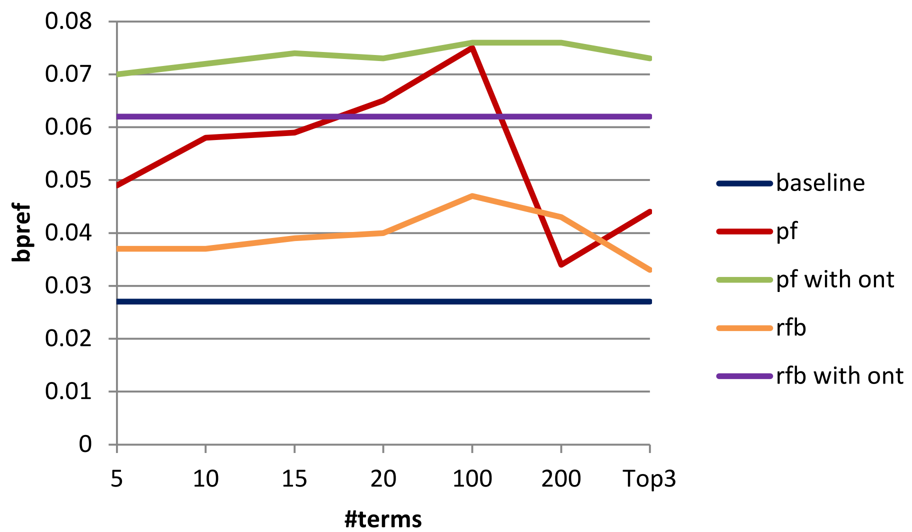


Figure 8.21 BPref results for Topic 51-100

For topic 51-100, the pf with ontology curve has the highest performance and is stable. The rfb with ontology curve has the next best performance Followed by the pf curve which peaks at 100 terms and dips at 200 terms. The rfb curve has the lowest performance and peaks at 100 terms.

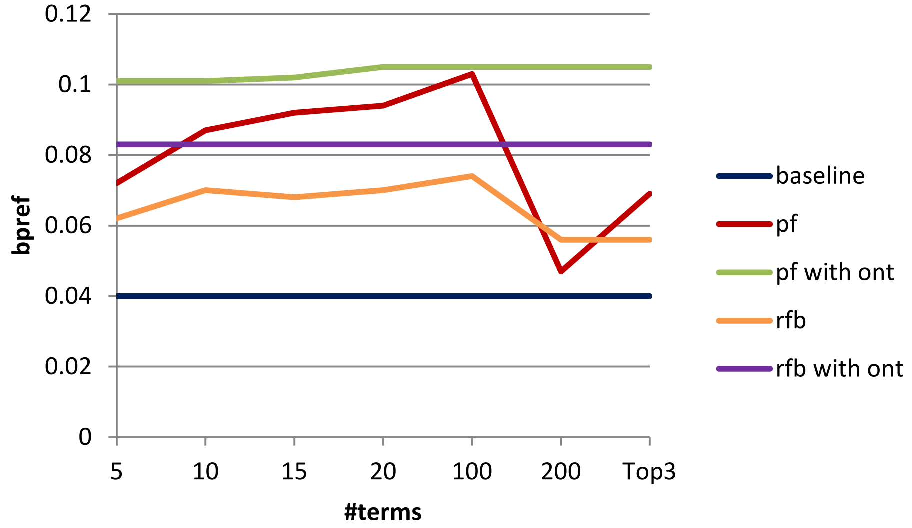


Figure 8.22 BPref results for Topic 101-150

For topic 101-150, the pf with ontology curve has the highest performance and is stable. The rfb with ontology curve has the next best performance Followed by the pf which peaks at 100 terms. The rfb curve has the lowest performance and peaks at 100 terms. The pf and rfb curve both dip at 200 terms.

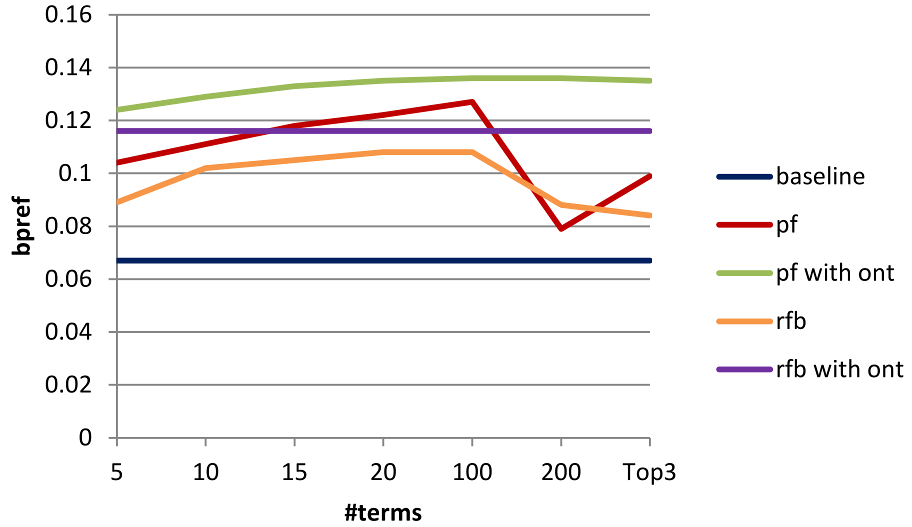


Figure 8.23 BPref results for Topic 151-200

For topic 151-200, the pf with ontology curve has the highest performance and is stable. The rfb with ontology curve has the next best performance followed by the pf which peaks at 100 terms. The rfb curve has the lowest performance and peaks at 100 terms. The pf and rfb curve both dip at 200 terms.

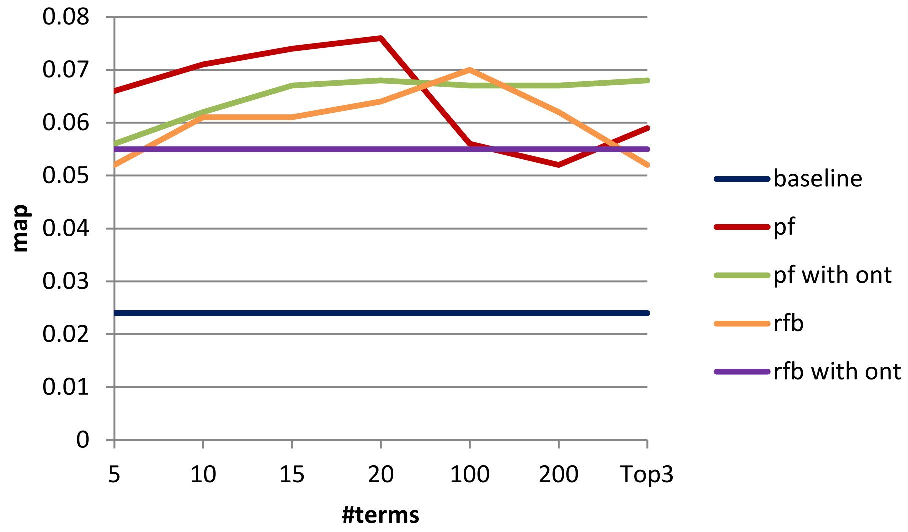


Figure 8.24 BPref results for Topic 201-250

For topic 201-250, the pf curve has the highest performance. The pf with ontology curve has the next best performance followed by the rfb curve which peaks at 100 terms. The rfb with ont curve is stable at 0.055.

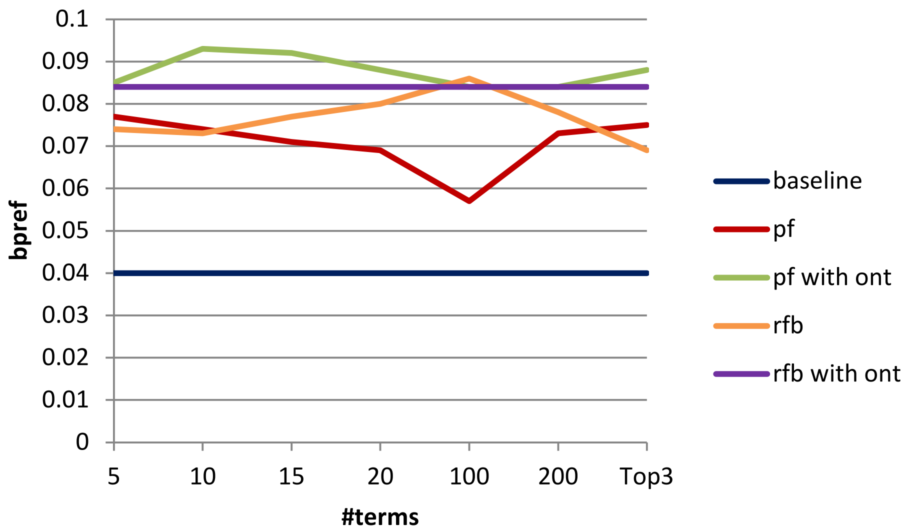


Figure 8.25 BPref results for Topic 251-300

For topic 251-300, the pf with ontology curve has the highest performance. The rfb with ontology curve has the next best performance followed by the rfb curve which peaks at 100 terms. The pf curve has the lowest performance and dips at 100 terms.

Table 8.5 shows the t-test results for for MAP figures (no. of terms). In the pf run, all runs are very significant except for 100 terms run which is significant. In the rfb run, the t-test results for all runs are very significant except for the runs with 20 and 100 documents which are significant. The highlighted cells show significant and very significant results.

#terms	Pf	Rfb
5	0.006	0.001
10	0.000	0.002
15	0.001	0.003
20	0.013	0.049
100	0.045	0.048
200	0.004	0.013
Top3	0.000	0.001

Table 8.5 Ttest (BPref)

Finally Appendix G shows the BPref results by topic-set and by varying the number of documents parameter. The results have been graphically plotted and analysed below.

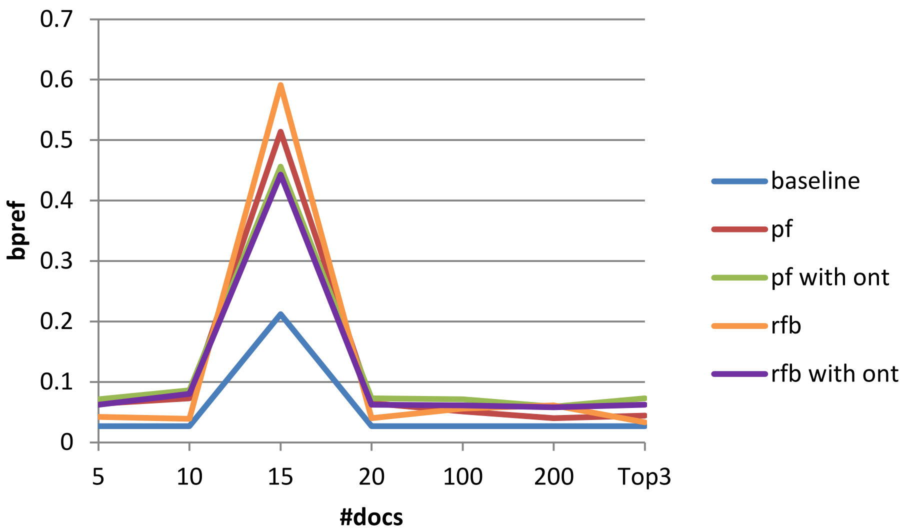


Figure 8.26 BPref results for Topic 51-100

For topic 51-100, the rfb has the highest performance followed by the pf, pf with ontology and rfb with ontology curve in this order. All of these curves, increase sharply from 10 docs , peak at 15 docs and decrease sharply at 20 docs. This could be due to reasons given for Figure 8.16.

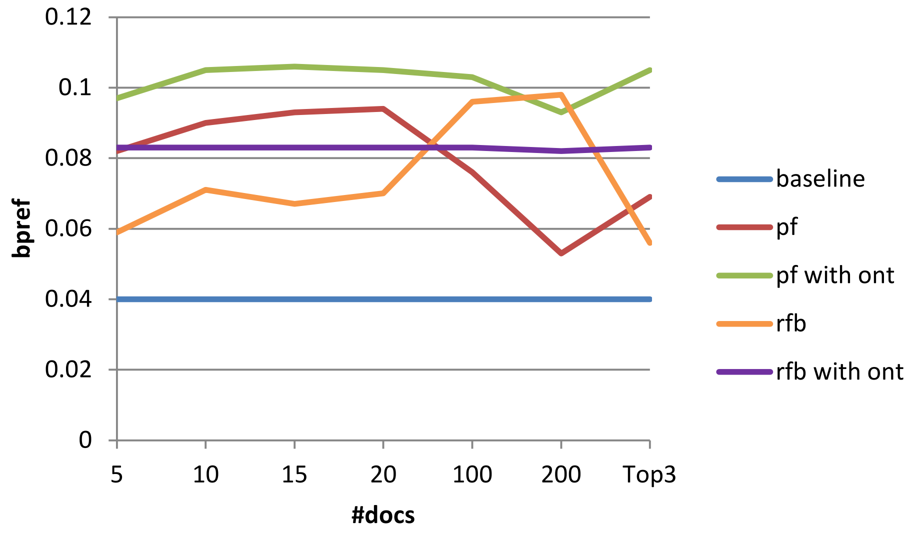


Figure 8.27 BPref results for Topic 101-150

For topic 101-150,the pf with ontology curve has the highest performance and is fairly stable and even though it dips slightly at 200 documents, the performance is same as that for pf with ontology curve. For other graphs the dip at 200 documents for the pf with ontology curve is much steeper. For this particular topic, increasing the number of documents to 200 has increased the BPREF score because further relevant documents have been found which were onriginally judged to be non-relevant. This is followed by the pf curve which peaks at 20 documents and dips at 200 documents. The next best performance is from the rfb with ontology curve which is stable at just over 0.08. The rfb curve peaks at 200 documents.

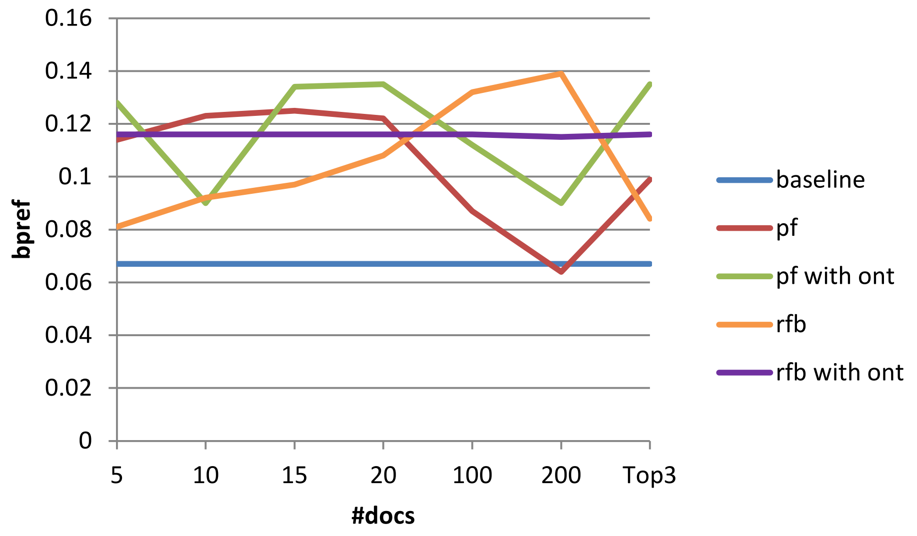


Figure 8.28 BPref results for Topic 151-200

For topic 151-200, the rfb curve has the highest performance and it peaks at 200 documents. However the pf with ontology curve and the pf curve both dip at 200 documents. The rfb with ontology curve is stable at just under 0.12.

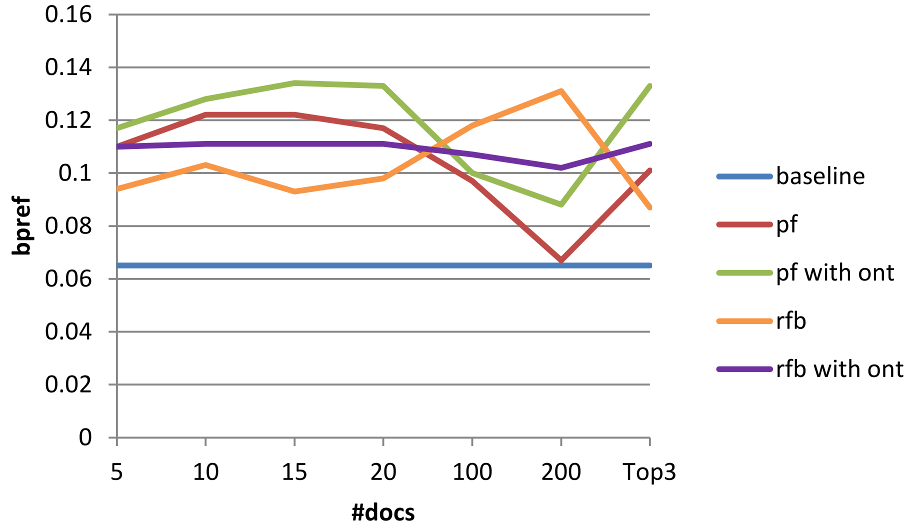


Figure 8.29 BPref results for Topic 201-250

For topic 201-250, the pf with ontology curve has the highest performance and it peaks at 20 documents. However the rfb curve peaks at 200 documents. However the pf and pf with ontology curves both dip at 200 documents. The rfb with ontology curve is stable at just under 0.12 but it too dips at 200 documents.

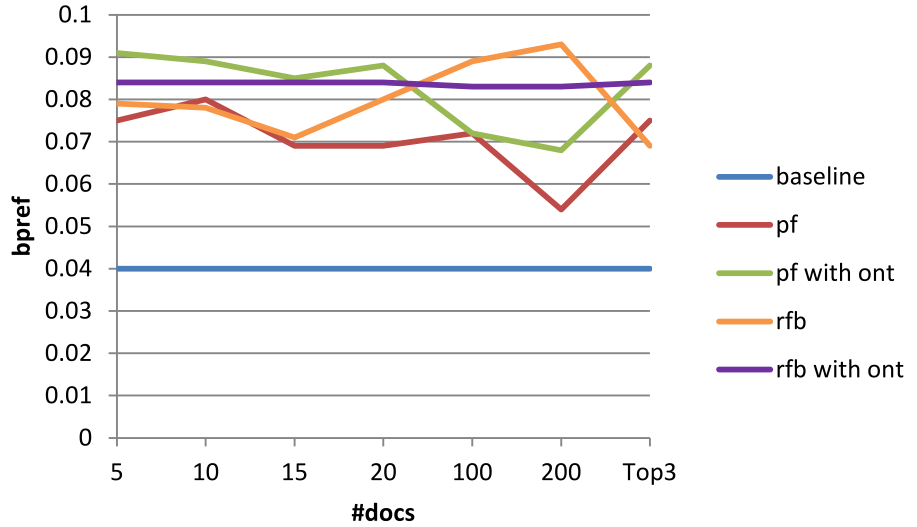


Figure 8.30 BPref results for Topic 251-300

For topic 251-300, the rfb curve has the highest performance and it peaks at 200 documents. However the pf and pf with ontology curves both dip at 200 documents. The rfb with ontology curve is stable at just over 0.08.

Table 8.6 shows the t-test results for for BPref figures (no. of documents). The pf runs with 5, 20, 200 and Top3 documents are very significant and the 100 document run is significant. The remaining two runs are not significant. In the rfb run, the highlighted cells show that the t-test results for runs with 5, 20, 200 and top3 documents are very significant and the runs with 10 and 100 documents are significant. Only the run with 15 documents is not significant.

#docs	Pf	Rfb
5	.003	.008
10	.414	.027
15	.458	.321
20	.001	.009
100	.029	.051
200	.003	.011
Top3	.001	.001

Table 8.6 Ttest (BPref)

The t-test results for BPref contain a higher number of significant/very significant results compared to the t-test results for MAP.

8.5 Summary

In this chapter we have presented the results for Recall, MAP and BPref. From our analysis, we can see that ontology based query expansion has achieved high recall results and the BPref results are satisfactory. However, ontology based query expansion has not improved MAP results. We have achieved higher recall compared to precision. Additionally the Recall and BPref t-test results are significant/very significant compared to the MAP t-test results.

With each of these measures, some of the pf runs outperform the standard run (20 document, 20 term). For example, with recall, the 100 term run produces better recall for pf and rfb than the standard run (Appendix G – Table G1). The 100 and 200 document runs produce better recall than the standard run (Appendix G – Table G2). Also the 15 document run produces better recall for the pf with ontology curve for all topics except topic 51-100 and topic 251-300.

Similarly, for MAP (Table G3 and G4), the 100 term runs produce some improvements on the standard run for the pf, pf with ontology and rfb curves. With the BPref runs, varying the number of documents parameter (G4) shows improvement on the standard run for all of the curves compared to varying the number of terms parameter (G5). However, none of the rfb with ontology runs outperform the standard run (see Appendix G). In other words the ontology seems to have a greater benefit on the pf runs.

For all single value results (Recall, MAP, BPref) where the number of terms parameter has been varied, the pf and rfb curves fluctuate but other curves are stable. However where the number of documents parameter has been varied, the pf, pf with ontology and rfb curves all fluctuate except for the rfb with ontology curve.

Also for topics151-200 and 201-250 the ontology benefits the pf and rfb in most of these experiments. It has been observed that there is a general trend for the results of the pf and rfb curves to decline where the relevance feedback parameter has been set to 200 terms/documents. MacFarlane et al, (2010) also found that there was very little gain after 40 terms when they conducted experiments on comparing precision for genetic algorithms and hill climbers. Additionally, in those runs where the number of documents has been set to 200 documents, the rfb curve peaks while the pf and pf with ontology curves do the exact opposite and deteriorate.

9. DISCUSSION

9.1 Introduction

We have carried out experiments to compare the impact of ontology based query expansion with query expansion based on relevance feedback alone. Since there is no real consensus on the optimum number of documents or number of terms to use for relevance feedback, we wanted to investigate the effect of varying the number of terms/documents relevance feedback parameters. We wanted to find out if an increase in number of terms/documents used for relevance feedback necessarily causes an improvement in results. An additional factor to research is whether varying the number of terms has a better impact on the results compared to varying the number of documents. Twelve experiments were used. For experiment 1 standard relevance feedback parameters are used and the results for this experiment form the baseline. For experiments 2-11 one relevance feedback parameter is adjusted per experiment to determine what values produce the optimum results. Experiment 12 uses standard relevance feedback parameters but to address the question of whether to use all expanded terms or just some of them we have ranked the expansion terms in descending order of weighting and just used the top3 expansion terms.

9.2 Analysis of results

9.2.1 Document Level Averages

Ontology based query expansion has produced improved results for the standard run (pf and rfb). However, varying the relevance feedback parameters (number of terms/documents) has not produced any improvements for the top3 run and for the remaining runs the pf and rfb baselines have fluctuated. The results show a mixed picture and a clear trend does not emerge. Chapters 6 and 7 show that varying the number of terms affects the pf with ontology curve but varying the number of documents parameter also affects the rfb curve as well. The bulk of the t-test results for pf are significant but not for rfb. However increasing the number of terms/documents to 200 for rfb produces t-test results which are significant for both type of runs.

9.2.2 Precision-Recall

Generally the precision-recall results were very low compared to the document level averages results. Increasing or decreasing the number of terms/documents does not appear to have any vast impact on the precision-recall curves. However with the use of 200 documents for relevance feedback, the rfb curve has performed much better than the rfb with the ontology. The reason for this could be that narrower searches of the ontology are producing many child nodes for a given query term, but very few of these are relevant to the query topic. For the standard run the pf results are significant and increasing the number of terms relevance feedback parameter has a positive effect on the number of significant results for the pf run. However, in general, varying the number of documents does not have a positive effect on the ttest results.

9.2.3 Recall

This metric has had significantly better results than the other metrics. For all experiments, the Average recall t-test results are statistically significant for PF and RF runs. When the relevance feedback parameter has been set to 200 terms/documents there is a large drop in performance (as shown in most graphs in chapter 8).

9.2.4 MAP

The results for this metric are very low or sometimes even negative compared with the baseline. MAP results are not significant for experiments which have less than 20 terms as the relevance feedback parameter. Increasing the number of terms parameter to 200 has produced significant t-test results for the PR run and increasing the number of documents to 100 results in significant results for the rfb run. The run where top 3 expansion terms are used produces signicant results for pf and rfb. Varying the number of documents does not produce significant results for pf, but runs with more than 20 documents have significant results for rfb.

9.2.5 BPref

Results for this metric are fairly high. This metric takes a more balanced view because the measure doesn't just take into account those documents judged to be relevant but also those that don't have a relevance judgement attached to them but may still be relevant. Most of the Bpref t-test results are significant. Only the runs with 10 documents is not significant for pf, and the runs with 15 documents are not significant for pf and rfb.

9.3 Explanation of results

The above results have highlighted areas which have shown better improvement than others. We will address the key questions with some possible explanations.

Firstly, why do the pf runs show more improvement higher up the ranked set of results than the rfb runs?

In explanation of this we have several reasons. Firstly, for the baseline, the performance of relevance feedback is generally less effective than PF runs. This trend has continued with the ontology based expansion runs. Secondly, the results file shows that the documents retrieved by the pf run have a higher weighting than the documents retrieved by the rfb run.

```
51 Q0 AP880731-0085 0 123.150002 ok-test
51 Q0 AP880316-0292 1 115.103996 ok-test
51 Q0 AP880325-0293 2 110.718002 ok-test
51 Q0 AP880406-0267 3 110.222000 ok-test
51 Q0 AP880318-0287 4 107.478996 ok-test
51 Q0 AP881118-0209 5 105.786003 ok-test
51 Q0 AP881108-0253 6 105.489998 ok-test
51 Q0 AP880627-0045 7 102.538002 ok-test
51 Q0 AP880706-0311 8 89.964996 ok-test
51 Q0 WSJ910708-0061 9 88.129997 ok-test
```

Figure 9.1 Sample extract from pf results file: Level1ontresultsbf

```
51 Q0 AP880731-0085 0 42.898998 ok-test
51 Q0 AP880316-0292 1 33.375000 ok-test
51 Q0 AP880318-0287 2 33.055000 ok-test
51 Q0 AP880325-0293 3 32.209000 ok-test
51 Q0 AP880406-0267 4 31.868000 ok-test
51 Q0 AP880706-0311 5 31.806000 ok-test
51 Q0 WSJ910708-0061 6 30.823999 ok-test
51 Q0 AP881108-02531 7 30.006001 ok-test
51 Q0 AP880627-00451 8 29.850000 ok-test
51 Q0 WSJ900720-0157 9 29.721001 ok-test
```

Figure 9.2 Sample extract from rfb results file: Level1ontresultsrfb

Figure 9.1 shows a sample extract from the results file from the standard pf with ontology run (20 terms, 20 documents) and figure 9.2 shows a sample extract from the standard rfb with ontology run (20 terms, 20 documents). From left to right the record fields are: Topicno; literal; document id; rank; weighting; runid

We can see that the documents in figure 9.1 have a much higher ranking/weighting than the documents in figure 9.2

In our opinion the rfb is harder to improve on because the top N documents used for rfb are already judged to be relevant so rfb without the use of the ontology produces good results which are hard to improve on. For the pf runs, the top N documents are assumed to be relevant because they are ranked highly by the system. These documents might not contain as many relevant query expansion terms as the rfb documents so any relevant additional ontology based query expansion terms will result in an improvement.

Secondly a question which arises is why do the graphs for document level averages show more of an improvement compared to the Precision-Recall graphs?

The reason for this is that it is easier to achieve improvements in precision in the top 5 or top 10 documents compared to achieving improvements in precision at recall .10 especially if the document collection is large. For example if the document collection is 20,000 documents, 0.10 recall calculates to 2000 documents.

Thirdly, why are the recall results much more enhanced than the precision-recall results?

An explanation for this is that quite a large number of query topic terms are being found in the ontology and even though each of these only has one parent node associated, the use of these parent nodes is retrieving more relevant documents. Sometimes when searching for parent nodes, the ontology produces relevant terms (see "oil" example on p.88). In other cases the ontology produces non-relevant terms which have a negative effect on precision and recall as shown in the example below:

TOPIC NUMBER = 90 ("data proven reserves oil natural gas producers")
current word is oil
    --> economy_business_finance

The TIPSTER description for topic 90 refers to documents that "provide totals or specific data on changes to the proven reserve figures for any oil or natural gas producer". In this case the query topic is more to do with the reserve figures for ***oil*** and is not related to ***economy_business_finance***.

However with the "narrower" searches, fewer query topic terms are matched with the ontology terms. Where a match occurs, the ontology term tends to have many more child terms associated with it but the precision-recall depends on the number of child terms that are relevant to the query topic and the number of relevant documents that contain the child term. In some cases a larger number of relevant results are produced by the ontology which result in improved precision-recall (see "civil" example p. 76). However in other cases, just because an ontology term has lots of associated child terms, does not necessarily mean that the number of relevant documents retrieved will increase vastly. An example of this is for narrower searches, where the term is quite general, many child nodes are retrieved of which only one or two might be relevant (see "natural" example p. 77). Alternatively, the term produced is so general it does not improve the precision results at all because it retrieves a large number of documents which contain the general term and many of these documents are not relevant to the query topic (see "research" example p.79). Another example to illustrate lack of improvement in performance retrieval is where many of the child terms are related to the search term but not relevant to the query (see "Health" example p.87)

This next section illustrates the point that an improvement in results depends more on the quality of the ontology nodes as opposed to the quantity of ontology terms, Table 9.1 shows for topic 251-300, a breakdown of each topic and number of terms where at least one parent node has been found and the number of terms for which at least one child has been found.

Topic	No of Terms with parent	No of Terms with children
254	1 invasive	
258	1 computer	
260	1 human	
261	1 material	
264	1 foreign	1 foreign
267	1 firefighter	
268	1 national, defense	1 defense
269	1 foreign	1 foreign
273	1 volcanic	
275	2 natural, health	2 natural, health
277	1 mines	
278	1 human	
279	1 earth	
282	1 juvenile	1 crime
284	1 drug	
289	1 hospitals	
290	1 foreign	1 foreign
292	1 welfare	
294	2 animal, animals	
299	2 local, economies	
300	1 air	
Total	24	7

Table 9.1 Number of terms in topics251-300 with at least one parent/child term

Table 9.2 shows a summary set of statistics for all topics and the number of terms with parent(s) and the number of terms with children. From this we can see that the number of terms where at least one parent node has been found exceeds the number of terms with children. The number of terms with parent nodes results in more general results being retrieved. Less narrow matches means fewer relevant results are retrieved. This is an explanation why recall is better than precision.

Topics	No of Terms with parent	No of Terms with children
51-100	22	6
101-150	30	10
151-200	29	3
201-250	21	4
251-300	24	7

Table 9.2 Summary of statistics given in Table 9.1

We would expect topic 101-150 to have the highest number of relevant documents retrieved. However table 9.3 shows that this is not the case.

Run Type	Topic	+/- #docs retrieved (Pf)	+/-#docs retrieved (Rfb)
Standard	51-100	161	393
	101-150	103	252
	151-200	354	376
	201-250	249	323
	251-300	91	116

Table 9.3 Difference in the number of documents retrieved for pf and rfb standard run

So even though topic 101-150 has the highest count for number for terms with a parent nodes and child nodes, the topic does not produce the biggest improvement in terms of number of relevant documents retrieved.

Sometimes a given ontology term and its associated children/parent terms are retrieved several times which is fine if the term has relevant ontology terms but has a detrimental effect if the ontology terms are not relevant.

In some cases the ontology produces relevant terms but an improvement is not shown in the graphs. The reason for this could be that there might not be many documents in the collection which contain that term. Generally it was found that searches for broader ontology terms resulted in fewer but high quality nodes which then resulted in an improvement in recall. The searches for narrower terms produced many resulting nodes but of these not many were relevant to the query topic, thus resulting in lowering precision results.

9.4 Other influential Factors

9.4.1 Quality of the Ontology and the Document collection

Improved retrieval results depends on the ontology coverage of the topic in breadth and depth. Also the similarity of terms between the ontology and the document collection. Finally the document collection coverage of the ontology terms.Ontology could have a lot of terms related to the topic but these terms might not be contained in many documents so minimum impact on performance.

9.4.2 Stemming

The ontology results in improvements for some topic-sets but not for others. First of all, when searching an ontology using terms from a query topic, we need to find at least one hit in the ontology for any improvements to take place. Some topics have more ontology hits than others. The second success factor relies not on just the number of hits in the ontology but on the retrieved ontology terms being relevant to the query topic.

We use stemmed keywords when searching the ontology, so its possible that the actual ontology hits are irrelevant and/or the retrieved ontology terms are irrelevant. For example:

```
TOPIC NUMBER = 102 (“laser research applicable uss strategic
defense initiative”)
current word is missil
     -->  defence
```

In other cases, its possible that the actual ontology hits are irrelevant and/or the retrieved ontology terms are irrelevant. For example:

```
Topic no = 223 (responsible great emergence Microsoft computer
industry)
ORIG WORD IS emergenc
     --> explosion
```

Emergence has been stemmed to emergenc, and ontology picks up non-relevant term explosion which is more related to emergency than emergence.

Lets suppose we find a good set of ontology terms to expand the query with, then the next factor in improving retrieval relates to finding enough documents in the document collection that contain the ontology term and are relevant to the query topic. If the match between the ontology and the document collection is poor, then even though the ontology terms are relevant to the query topic, because there aren’t enough documents containing that term, query expansion has minimal effect on recall/precision. Alternatively, if the parent/child term obtained from the ontology is too general, then many documents will be retrieved but very few of these will be relevant to the query topic.

9.4.3 Topic Hardness

A topic hardness measure is calculated as the average over a given set of runs of precision for each topic after all relevant documents have been retrieved OR after 100 documents have been retrieved if more than 100 documents are relevant. The measure is oriented towards high-recall performance and how well systems do at finding all relevant documents. If no system does well on a query then it can be called a hard query. According to TREC hardness measure given in Buckley et al, (1996) the performance for TREC 4 (topic 201-250) and TREC 5 (251-300) drops from 0.676 to 0.672 and 0.556 respectively. These are seen to be difficult topics because they are progressively shorter in length and higher level in nature. This trend is mirrored in the SMART experiments. For example in TREC1 the precision is 0.2431 and in TREC2 the best precision has improved to 0.2594 but in TREC 4 and TREC 5 the precision has dropped to 0.1507 and 0.1038 respectively.

Table 9.4 summarises for each experiment which topic set performed best for each of the pf and rfb runs across the different metrics and also shows whether the results were statistically significant or not based on the t-tests that were carried out. T-test results that are <0.05 are significant (highlighted in yellow) and those that are <0.01 are very significant (highlighted in pink).

Experiment	Doc level Averages		Precision-Recall		Recall		MAP		BPref	
	PF	RFB	PF	RFB	PF	RFB	PF	RFB	PF	RFB
5 terms	51-100	251-300	251-300	51-100	51-100	51-100	51-100	51-100	51-100	51-100
10 terms	251-300	251-300	251-300	51-100	151-200	51-100	251-300	51-100	251-300	51-100
15 terms	51-100	51-100	251-300	51-100	151-200	201-250	251-300	51-100	251-300	51-100
100 terms	251-300	51-100	251-300	251-300	251-300	201-250	251-300	51-100	251-300	51-100
200 terms	51-100	151-200	251-300	151-200	151-200	151-200	51-100	151-200	101-150	101-150
5 docs	251-300	151-200	251-300	51-100	151-200	151-200	251-300	151-200	251-300	51-100
10 docs	101-150	251-300	251-300	51-100	51-100	51-100	251-300	51-100	51-100	51-100
15 docs	251-300	251-300	251-300	51-100	151-200	51-100	251-300	51-100	251-300	101-150
100 docs	51-100	201-250; 251-300	251-300	51-100	51-100	151-200	51-100	251-300	51-100	51-100
200 docs	251-300	151-200	251-300	251-300	251-300	251-300	101-150	251-300	101-150	51-100
20 terms/docs	51-100	51-100	251-300	51-100	201-250	201-250	251-300	51-100	251-300	51-100
Top 3 expansion terms	51-100	151-200	251-300	51-100	51-100	51-100	51-100	51-100	51-100	51-100

Table 9.4 Overall Results summary

We can see that topic-set 51-100 has the best performance for 7 experimental runs (5,10,15 terms; 10, 100 documents; 20 terms/docs and top 3 expansion terms). Topic-set 251-300 has the best performance for four of the experiment runs (100 terms; 5, 15 and 200 documents). Topic-set 151-200 has the best performance for 2 runs (200 terms and 5 documents). These topics have benefited most from the use of the ontology.

For each run, we can compare across the various metrics to see which topics occur the most. Again, topics51-100 and topics251-300 have the highest frequency across the various metrics. According to Buckley (1996), topic 251-300 is considered to be a hard topic. So the ontology seems to have improved the retrieval performance for a hard topic as well as one that is considered not to be hard.

We can analyse table 9.4 for statistical significance. For the Document level averages, twice as many pf results are significant/very significant compared to the rfb results. For the Precision-Recall metric, only the pf results are significant/very significant. Recall is the metric with the highest number of statistically significant results. For recall, all results are significant/very significant except for topic 251-300 (rfb). MAP is the metric with the lowest number of statistically significant results, of which twice as many rfb results are significant/very significant compared to the pf results. With BPref, more of the pf results are significant/very significant compared to the rfb results. So Recall is the metric with the best results. The BPref results are comparable to the Recall results but in some cases not as good. MAP is the metric which has performed the worst in terms of retrieval performance and significance of results.

So we have high recall at expense of precision. This is good for the news domain because professional searchers such as lawyers and investigative journalists prefer to obtain as much information about a given news story as possible. Lawyers need to look at all case statutes in order to produce a strong argument otherwise missed case articles will weaken their evidence. In the same way investigative journalists need to ensure they have accessed all relevant articles in order to produce a thorough report on the subject they are investigating otherwise they will be open to criticism if gaps in the research are found. Also the analysis in sections 5.1 – 5.12 shows the document level average results are better than recall –precision and the document level averages (pf runs) are benefitting from the ontology higher up the rank. Again this would indicate that the ranking algorithm is working and searchers tend to concentrate on the documents occurring higher up in a ranked set of results. The documents for PF are "assumed" to be relevant because they appear high up in the system ranking, whereas the documents for rfb are judged by human assessors as actually being relevant. It would be difficult to improve retrieval performance on the rfb relevant documents, however the pf runs have more to gain from these other factors than rfb.

Robin and Ramalho (2003) used disk2 of the TREC collection and the WordNet ontology to expand query words with some of their synonyms and hypernyms. For comparison purposes, the document collection is the same but we have used a news based ontology to obtain synonyms and hypernyms instead. The other difference is that Robin and Ramalho used the F-measure metric instead of BPref. Finally they used bounds of 10, 15, 20, 30 and 50 documents, we used 5, 10, 15, 20, 100 and 200 terms/documents. They found that all expansion strategies improve overall effectiveness by improving recall more than they worsen precision (in relative terms). Their results show that recall can be boosted up by as much as 72.4% relative to the no expansion case. They also expand to the first-level in the ontology. Their best query expansion strategy yields only a 2.51% improvement reaching 9.3% and only 11% of all relevant documents together with 77.5% irrelevant ones. For bounded precision for the top 20, 30, 40 and 50 documents, precision respectively improved by 1%, 12%, 17% and 37%.

In comparison to Robin and Ramalho's work, our results are just as good if not better for recall and precision. Even though there was in improvement in our results for some topics across different runs, unlike Robin and Ramalho we did not discover any linear trend resulting from increasing the number of terms/documents.

9.4.4 Data Quality

Effectiveness of query expansion using relevance feedback can vary depending on many factors such as choice of parameters in the term weighting process, number of relevant documents in the document collection, facilities provided for users to give good quality relevance feedback with ease and finally whether the collection is domain specific or domain independent. Other factors which may affect the results relate to the quality of the data. For example if American spelling (eg fiber) has been used then documents which use English spelling (eg fibre) might not be retrieved. Secondly in some instances the query terms have not been separated by spaces. For example in topic68, we have a word ‘finediameter’ which is meant to be ‘fine diameter’. Different spelling variations of a term is sometimes not a problem. For example in topic 76 and topic 246 United States is referred to as ‘us’ whereas, topic 264 and topic 290 use ‘u.s’. In this case, the stemmer would remove the full stop and reduce ‘u.s’ to ‘us’. Another problem which sometimes occurs is mis-spelling. For example in topic 79 the word ‘frg’ and topic 81 ‘ptl’ do not make sense.

9.5 Recommended Guidelines for Ontology Based Query Expansion

This thesis has presented a thorough discussion on the use of an ontology for query based expansion.Our approach can be generalised to other retrieval models and other ontologies if the retrieval models offer a relevance feedback mechanism and ranking of retrieved results. An illustration of this is Robin & Ramalho’s work (2003) which uses ontology based query expansion with a vector space retrieval model instead of a probabilistic retrieval model and the WordNet linguistic ontology instead of a news domain ontology. Section 1.1 states that the vector space model provides document ranking. Therefore it is possible to conduct relevance feedback using vector space models. The same is true for Latent Semantic Indexing. By drawing from the lessons learnt, we can recommend a set of general guidelines on how ontologies can be best used to maximise the impact for query expansion .

These guidelines are as follows:

- Ontology based query expansion is only good for subject finding type queries not for navigational type queries such as finding a website (section 1.1 p.19).
- Only use ontology based query expansion for short queries because long queries already contain a full description of the information requirement so they won’t benefit from ontology based query expansion (section 4.5)
- There should be a close match between the ontology and the document collection in order to produce the most effective results. If the concepts contained in the document collection are not covered by the ontology then ontology based query expansion will not produce any significant improvements in the results (section 2.7.4.5 p.50)
- Expand the original query using traditional relevance feedback techniques before searching the ontology in order to maximise the chances of finding a match in the ontology (section 2.7.4.3 p.46)
- Analyse the context of the query to determine whether a broad/narrow search is required. For narrow/specific searches, child nodes in the homonym and meronym relations can be explored to narrow the scope of the topic area. If available, an instantiated version of the ontology can be used to obtain exact matches with specific proper nouns (section 3.3.1.1 p.62).
- For broad searches the parent nodes in the hypernym, synonym and holonym relations can be explored (section 3.3.1.1 p.62).
- A search query requiring alternatives to the node in the query term should use the information in the sibling nodes (section 3.3.1.1. p62).
- Information retrieved from nodes with a closest distance to the current query node is most beneficial to improving effectiveness of the query search. For example more effort required to traverse the nodes at further distances offers

very little gain. It is recommended that only nodes that are one level away from the current search node are traversed (section 5.1).

- Finally in more recent web based applications, the ontology can be used for query formulation and query expansion. For example in faceted searches, a section of the ontology hierarchy is presented to the user and the the user can filter the search results by clicking on different facets of the ontology. By removing the filters, the search becomes broader (section 2.4 p.27).
- It is recommended that faceted search be used for broad but shallow ontologies and user interactive mode has been selected (section 1.1 p.18)

9.6 Summary

Research experiments to examine the effects of using an ontology for query expansion in the newswire domain have been designed, conducted and a discussion of the results has taken place. We have presented an analysis of overall results for each type of evaluation metric. This has helped us to identify which factors could make a difference to the retrieval results.

A summary of the eight main findings for the experiments is as follows:

- Use of the ontology has increased the number of relevant documents retrieved. Varying the number of terms parameter for relevance feedback has resulted in an average increase of 88% for pf and 20% for rfb. However varying the number of documents parameter has achieved an increase of 52% for pf and a decrease of -26% for rfb.
- The ontology improves results for topics considered to be hard and topics considered to be not hard.
- The ontology has a better effect higher up the rank for the pf runs of Document Level Averages metric and the rfb runs starts improving from the lower end of the ranked set of documents which implies that the pf runs have more to gain from varying the relevance feedback parameters and do benefit from the use of the ontology. With the rfb runs, use of the ontology based terms for query expansion is distorting the retrieval of relevant documents and is only useful at the lower end of the ranked list.
- The ontology does not produce any significant improvements for the Precision-Recall results.
- The use of ontology based query expansion has achieved high Recall results. This is possibly because query topics have a higher number of hits in the ontology for broader searches and for each hit, few ontology terms are retrieved but a higher proportion of the terms retrieved are relevant compared to ontology terms retrieved for narrower searches.
- The use of ontology based query expansion has only increased mean average precision for a few cases but overall the precision is usually identical to the baseline or sometimes even below the baseline. The reason for this is that more ontology child terms are retrieved but a smaller proportion of these are actually relevant, thus having minimum impact on precision.

- The ontology has more of a positive impact on the pf runs compared to the rfb runs.
- The number of terms parameter for relevance feedback benefits the pf and pf with ontology results but the number of documents parameter also has an affect on the rfb results.

In cases where the ontology has not resulted in improvement on the baseline, this might be due to the fact that the terms obtained from relevance feedback are relevant but the ontology based query expansion process is adding terms from the ontology that are not relevant to the query and so the results show no improvement compared to the baseline. Other factors which have a negative impact on precision might be down to the fact that even though an ontology child/parent node is relevant to the query topic, only a small number of documents contain that term. In some cases, the ontology term is relevant but quite general, so many documents containing that term are retrieved but they are not relevant to the query topic. Finally, retrieval results have improved with the use of the ontology but we don't have a clear trend that increasing the number of terms/documents results in improved retrieval.

The above results are similar to the results of other experiments using other types of domains and ontologies. For example Mandala et al. [2000] compared relevance feedback to ontology-aided query expansion, and proclaimed that the latter outperforms pseudo relevance feedback remarkably but was slightly less effective than ideal relevance feedback. Also Robin & Ramalho (2003) found that all expansion strategies improve overall effectiveness by improving recall more than they worsen precision (in relative terms. In general the use of an ontology for query expansion tends to increase recall quite significantly but have less of an impact on precision.

10. CONCLUSION

Search engines have been used to carry out information retrieval on the web. Information space on the web is comparatively larger and combined with the ambiguity of the English language, a long list of results is returned, much of which is not always relevant to the user's information needs. For traditional information retrieval systems and web based retrieval systems, the inadequacies of standard search engines means that the user community is suffering from information overload.

To increase the number of relevant documents retrieved queries need to be disambiguated by looking at their context. Third generation search engines attempt to determine the context of the user query and allow the user to obtain more meaningful results. In other words these search engines are focusing more on achieving high precision.

The most recent query expansion technique involves the use of ontologies to infer context for ambiguous queries. The concepts in the ontology can be used for word sense disambiguation and subsequent query expansion (Buckland 2003). A detailed investigation into query expansion using ontologies was needed to study the reasons for their success/failure. Therefore, our motivation for this research was to carry out such an investigation and address questions such as whether the use of query expansion increases recall, precision or both compared to pf and rfb.

The aim of this research is to investigate the effectiveness of ontology based query expansion techniques. The central hypothesis is that the use of ontology based query expansion in a probabilistic retrieval model will improve retrieval effectiveness. This aim can be broken down into a number of objectives. In the following few sections we will discuss how each research objective was achieved. After the conclusion, we will discuss recommendations for future work in this area.

The first objective was to select and index a document collection and then select an appropriate ontology. We selected the TREC newswire document collection (disk2) because of its large size and also the relevance judgements were readily available. TREC document collections are widely accepted by the information retrieval research community. For the ontology we had a choice of either developing it from scratch or to use an existing ontology which was based on the news domain. Even though the number of news ontologies in existence were not great in number, we decided to use the World News Ontology developed by Kallipolitis et al, (2007).We chose this ontology for several reasons. Firstly it was written in XML which meant that it enhanced system portability and was relatively easy to process. Secondly the ontology was based on the industry standards news codes taxonomy produced by ITPC. The indexing programs provided by the Okapi system have been updated to handle XML files. So the document collection was indexed on the TREC document id (DOCNO), heading (HEAD) and description (TEXT) fields using the Okapi indexing programs.

The second objective was to build a separate database containing the semantic information such as parent-child relationships between ontology nodes. This was required so we could transfer the ontology knowledge in an appropriate format and make it accessible to the Okapi software. The Okapi program was modified to parse the

XML ontology and each node was processed recursively to obtain parent-child nodes until the entire ontology tree had been processed. The parent-child information was stored in list format as described in section 4.7.1. The level information shows whether the ontology parent-child term is obtained at a distance of 1 node (level1) or 2 nodes (level2) from the current node. This was done to find out the optimum level of ontology processing required to improve the retrieval results.

The third objective was to design and conduct laboratory experiments in order to compare and contrast performance of the standard retrieval model with the revised retrieval model. The Okapi model already uses pf and rfb techniques (Jones et al, 1997) and the relevance feedback information can be based on pre-stored relevance judgements which indicate for each document whether it is relevant to the topic query or not. The techniques have proved to be successful to a certain extent so we do not want to discard them. The two main parameters of relevance feedback are: selection of terms and the sample size of relevant documents. In the Okapi system traditionally these have been 20 terms and 20 documents. Billerbeck and Zobel (2004) state that the choice of query expansion parameters used can affect the retrieval performance. A range of experiments were designed and conducted to test our hypothesis. The primary purpose of our experiments was to judge the effect of ontology based query expansion, but in addition we experimented in varying the relevance feedback parameters and analysed the impact of doing so on the results. Another dimension to our experiments was to conduct query expansion using different depths/levels of nodes in the ontology. For our research we looked at a maximum of two levels. Another question that was addressed was whether to use all expanded terms or select the top 3 query expansion terms. Search routines were developed which used relevance feedback for query expansion and the resulting set of expanded terms were expanded even further by using associated broader and narrower ontological terms. Experimental results were evaluated using retrieval effectiveness metrics. The initial study recommended conditions where ontology based query expansion is likely to be successful. The news ontology will be analysed to see which of these conditions or success factors were applicable if any.

Our final objective was to summarize the findings of this research and recommend areas for future work. It is important to compare our findings with those of other related research. The use of ontologies for query expansion has had mixed success (Gonzalo et al, 1998) because they are effective in increasing recall and less successful than rfb but as good as PF (Billerbeck and Zobel 2004). Our findings support these statements. Our attempts at ontology based query expansion have had mixed success. Use of the ontology has vastly increased the number of relevant documents retrieved. We can conclude that for both types of query expansion, the pf results are better than the rfb results. Our findings are similar to that of Billerbeck and Zobel (2003) in that ontology based query expansion enhances recall, and produces bigger improvements for pf compared to rfb. The ontology has a better effect higher up the rank for the pf runs of Document Level Averages metric and the rfb runs starts improving from the lower end of the ranked set of documents which implies that the pf runs have more to gain from varying the relevance feedback parameters and do benefit from the use of the ontology. Query expansion seems to be more successful only on relevant documents (Ogawa and Mano 2001, Billerbeck and Zobel 2003). In support of this statement, use of the ontology based terms for query expansion in rfb runs is distorting the retrieval of relevant documents and is only useful at the lower end of the ranked list.

Also ontology based query expansion seems to have a better effect on recall compared to precision. Using ontology based query expansion we have high achieved high recall but very little improvement on MAP. The BPref results also offer a decent improvement on the baseline. This is possibly because query topics have a higher number of hits in the ontology for broader searches and for each hit, few ontology terms are retrieved but a higher proportion of the terms retrieved are relevant compared to ontology terms retrieved for narrower searches. For the latter, more ontology terms are retrieved but a smaller proportion of these are actually relevant, thus having minimum impact on precision.

The ontology produced improvements in recall but the precision results were unchanged. In some cases precision has improved but the results do not present a set pattern, they vary between topics. The use of ontological query expansion can improve the performance in terms of average precision and recall without reducing recall, but not for all cases. The pf runs benefit from varying the number of terms parameter and the rfb runs benefit from varying the number of documents parameter. We did not find any trend that increasing the number of terms/documents used for relevance feedback necessarily resulted in improvements in recall and precision. The ontology improves results for topics considered to be hard and topics considered to be not hard.

The ontology does improve the pf with ontology and rfb with ontology curves. In cases where the ontology has not resulted in improvement on the baseline, this might be due to the fact that the terms obtained from relevance feedback are relevant but the ontology based query expansion process is adding terms from the ontology that are not relevant to the query and so the results show no improvement compared to the baseline. Other factors which have a negative impact on precision might be down to the fact that even though an ontology child/parent node is relevant to the query topic, only a small number of documents contain that term. In some cases, the ontology term is relevant but quite general, so many documents containing that term are retrieved but they are not relevant to the query topic. Finally, retrieval results have improved with the use of the ontology but we don't have a clear trend that increasing the number of terms/documents results in improved retrieval.

This research has led to valuable lessons being learnt and we can draw from these lessons to produce a set of recommended guidelines for using ontologies in query expansion. These guidelines are as follows:

- Ontology based query expansion is only good for subject finding type queries not for navigational type queries such as finding a website (section 1.1 p.19).
- Only use ontology based query expansion for short queries because long queries already contain a full description of the information requirement so they won't benefit from ontology based query expansion (section 4.5)
- There should be a close match between the ontology and the document collection in order to produce the most effective results. If the concepts contained in the document collection are not covered by the ontology then ontology based query expansion will not produce any significant improvements in the results (section 2.7.4.5 p.50)
- Expand the original query using traditional relevance feedback techniques before searching the ontology in order to maximise the chances of finding a match in the ontology (section 2.7.4.3 p.46)

- Analyse the context of the query to determine whether a broad/narrow search is required. For narrow/specific searches, child nodes in the homonym and meronym relations can be explored to narrow the scope of the topic area. If available, an instantiated version of the ontology can be used to obtain exact matches with specific proper nouns (section 3.3.1.1 p.62).
- For broad searches the parent nodes in the hypernym, synonym and holonym relations can be explored (section 3.3.1.1 p.62).
- A search query requiring alternatives to the node in the query term should use the information in the sibling nodes (section 3.3.1.1. p62).
- Information retrieved from nodes with a closest distance to the current query node is most beneficial to improving effectiveness of the query search. For example more effort required to traverse the nodes at further distances offers very little gain. It is recommended that only nodes that are one level away from the current search node are traversed (section 5.1).
- Finally in more recent web based applications, the ontology can be used for query formulation and query expansion. For example in faceted searches, a section of the ontology hierarchy is presented to the user and the the user can filter the search results by clicking on different facets of the ontology. By removing the filters, the search becomes broader (section 2.4 p.27).
- It is recommended that faceted search be used for broad but shallow ontologies and user interactive mode has been selected (section 1.1 p.18)

This thesis has resulted in three key advances of work relative to previous work on ontology based query expansion. Firstly this work is original because it is the first time Okapi has been used with this news ontology. Therefore the results can act as a benchmark for any related future work. Secondly, this research differs from other research in the sense that we have conducted exhaustive testing on different relevance feedback parameters and at different levels of the ontology. Thirdly we have produced a set of recommended guidelines for using ontologies in the query expansion task.

So we have achieved our aim which was to “investigate the effectiveness of ontology based query expansion techniques”. Our experiments prove the central hypothesis of this research “the use of ontology based query expansion in a probabilistic retrieval model will improve retrieval effectiveness”. However, more work needs to be done on improving the precision results for ontology based query expansion.

Our work can be improved by conducting further research on the number of terms to use for query expansion. Robertson used 1/3 terms, Harman recommends 20 terms and Buckely did massive query expansion using 300-530 terms. Number is not important as the type of terms. Qiu and Frei (1993), argue that selecting query expansion terms based on relatedness to the whole query is more effective. In TREC 8 (Robertson and Walker 1999), a term selection measure is used for selective expansion to measure the statistical significance of any given term's association with relevance. The paper says that the choice of level (5%, 1% or 0.1%) is largely arbitrary and recommends setting the criterion in relation to the size of the vocabulary 1/Vec (c is a constant, positive or negative).

We could also investigate the impact of different types of terms used for query expansion. It is possible to handle any ambiguity in polysemous words by only expanding those terms that have a high similarity to the entire query (Mandala et al, 1998). A common assumption is that a hierarchical relationship is the strongest and most useful (Rada et al, 1991). Intelligent Information Retrieval systems move away from simple user driven thesaurus navigation and incorporate software components which act on behalf of users to find best routes and best terms to select. To increase the intelligence, the system should recognise synonyms and utilise homography which is a spelling method that represents every sound by a character. Our system does not at present have these features. Compound words add complexity to the query expansion process however, further research is needed on the effective deployment of compound words in query expansion. Jones et al, (1995) note further work is required with compound words which are prevalent in a thesaurus. Their work only records whether a word is part of a compound or not, it does not make any further use of this.

Our document collection did not distinguish between news based articles and other articles. This is because the news based articles do not use domain-specific terminology. In order to carry out this distinction it should be possible to manually or automatically assign a subject-code from the IPTC newscodes to every newsbased article. In doing so, documents using different words to describe the contents but covering a given news category would be retrieved because they carry the same subject code. This would give higher ranking to news based articles compared to standard documents. Another advantage of using numeric codes is that they are unique in value and there is no risk of the codes being ambiguous or duplicated.

Finally we could apply our query expansion algorithms to different ontologies to see what difference each ontology makes to the query expansion process and the reasons why one ontology is inherently better than another. For example the NEWS ontology (Sanchez-Fernandez et al, 2005) is larger in size which indicates it has more coverage of the news domain. It also has a more complex lattice structure and deeper levels of nodes than the ontology we used. It would be more complex to process but could produce enriched results.

Query expansion has been successful to a certain extent but there is still scope to improve the techniques for selecting and designing algorithms for optimum parameter choice and only expanding queries which would benefit from the query expansion process.

REFERENCES

Abasolo, J., Gomez, M (2000). Melisa: An ontology based agent for IR in medicine. ECDL workshop on semantic web, Lisbon, Portugal.

Abberley, D., Renals, S., Cook, G., Robinson, T. (1998). "The THISL spoken document retrieval system." Proc Sixth Text Retrieval Conference: 747-752.

Ahlers, D., Boll, S. (2009). "Beyond Position - Spatial Context for Mobile Information Retrieval Systems." Positioning, Navigation and Communication, WPNC 6th Workshop: 129-134.

Ahn, J., Brusilovsky, P., Grady, J., He, D., Syn, S. (2007). Open user profiles for adaptive news systems: help or harm? Proceedings of the 16th international conference on World Wide Web (WWW '07), Banff, Alberta, Canada, ACM.

Ariki, S., Sugiyama, Y. (1997). "A TV news retrieval system with interactive query function." 2nd International Conference on Cooperative Information Systems: 184.
Baeza-Yates, R., Ribeiro-Neto, B (1999). Modern Information Retrieval, Addison-Wesley.

Bao, J., Cao, Y., Tavanapong, W., Honavar, V (2004). Integration of Domain-Specific and Domain-Independent Ontologies for Colonoscopy Video Database Annotation. Proceedings of the International Conference on Information and Knowledge Engineering., Las Vegas, Nevada, USA, CSREA Press.

Barzilay, R., Elhadad, M (1997). Using lexical chains for text summarization. In Proceedings of the Intelligent Scalable Text Summarization Workshop,ISTS'97, Madrid, Spain.

Bateman, J.John Bateman's ontology Portal [online] available at http://www.fb10.uni-bremen.de/anglistik/langpro/webspace/jb/info-pages/ontology/ontology-root.htm [Accessed 20th July 2005]

Bates, M.After the Dot-Bomb: getting information retrieval right this time [online] available at http://www.firstmonday.dk/issues/issue7_7/bates/index.html [Accessed 1st September 2002]

Bauer, T., Leake, D. (2001). "Real Time User Context Modelling for information Retrieval agents." CIKM 2001.

Baziz, M., Boughanem, M., Aussenac-Gilles, N. (2005). Conceptual Indexing Based on Document Content Representation. Information Context: Nature, Impact, and Role: 5th International Conference on Conceptions of Library and Information Sciences, CoLIS 2005, Glasgow, UK.

Becker, C., Nicklas, D. (2004). "Where do Spatial Context Models end and where do Ontologies start? A Proposal of a combined approach." Proceedings of the 1st International Workshop on Advanced Context Modelling, Reasoning and Management UbiComp.

Bharat , K. (2000). "SearchPad:explicit capture of search context to support web search." WWW9/Computer Networks 33((1-6)): 493-501.

Bhogal, J., MacFarlane, A., Smith, P. (2007). "A review of Ontology based query expansion." Information Processing and Management 43(4): 866-886.

Billerbeck, B., Zobel, J. (2004). Questioning Query Expansion: An Examination of Behaviour and Parameters. Proceedings of the Fifteenth Australasian Database Conference, ADC 2004, Dunedin, New Zealand, Australian Computer Society.

Billsus, D., Pazzani, M.J. (1999). A Personal News Agent That Talks, Learns and Explains. Agents 99: Proceedings of the third annual conference on Autonomous Agents, Seattle, Washington.

Broder, A. (2002). "A taxonomy of Web search." SIGIR Forum 36(2): 3-10.

Buckland, M. (2003). Translingual information management using domain ontologies. [online] available at http://metadata.sims.berkeley.edu/GrantSupported/tides.html [Accessed 10th September 2003]

Buckley, C. trec_eval [online] available at http://trec.nist.gov/trec_eval/ [Accessed 14th July 2009]

Buckley, C., Salton, G., Allan, J., Singhal, A (1995). Automatic query expansion using SMART: TREC 3. Proceedings of the Third Text Retrieval Conference, NIST Special Publication 500-225, 69-80.

Buckley, C., Voorhees, E. (1994). Retrieval Evaluation with Incomplete Information. SIGIR, Sheffield, South Yorkshire, ACM.

Bunningen, A., Feng, L., Apers, P. (2006). "A Context-Aware Preference model for database querying in an Ambient Intelligent Environment." Database and Expert Systems Applications, 17th International Conference, DEXA Poland: 33-43.

Callan, J., Croft, B., Harding, S. (1992). The INQUERY Retrieval System. DEXA, 78-83.

Cantador, I., Bellogin, A., Castells, P. (2008). "A multilayer ontology-based hybrid recommendation model." AI Communications - Recommender Systems archive 21(Issue 2-3): 203-210.

Cantador, I., Bellogin, A., Castells, P. (2008). News@hand: A Semantic Web Approach to Recommending News. Adaptive Hypermedia and Adaptive Web-Based Systems, 5th International Conference, AH 2008, Hannover, Germany, July 29 - August 1, 2008. Proceedings. Lecture Notes in Computer Science 5149 Springer 2008.

Chen, G., Kotz, D. (2000). "A Survey of Context-Aware Mobile Computing Research." Time.

Chen, H., Lynch, K., Bashu, K., Ng, T. (1993). "Generating, integrating and activating thesauri for concept-based document retrieval." IEEE Expert: 25-34.

Chen, H., Perich, F., Finin, T., Joshi, A. (2004). SOUPA: Standard Ontology for Ubiquitous and Pervasive Applications. International Conference on Mobile and Ubiquitous Systems: Networking and Services, Boston.

Cheng, C., Pan, X (2004). "Using perception in managing unstructured documents,." Crossroads archive 10(2): Pages: 5 - 5.

Chu, W., Liu, Z., Mao, W (2002). Textual Document Indexing and Retrieval via Knowledge sources and data mining. Communication of the Institute of Information and Computing Machinery(CIICM), Taiwan.

Claypool, M., Miranda, T., Gokhale, A., Mir, T., Murnikov , P., Netes , D., Sartin, M. (1999). Combining Content-Based and Collaborative Filters in an Online Newspaper. Proceedings of ACM SIGIR Workshop on Recommender Systems, Part of the 22nd International Conference on Research and Development in Information Retrieval, University of California, Berkeley.

Climent, S., Rodriguez, H., Gonzalo, J. (1996). Definitions of the links and subsets for Nouns of the EuroWordNet Project., EuroWordNet, E.C. LE2-4003 Deliverable 005. Universitat de Barcelona, Universitat Politècnica de Catalunya, Universidad Nacional de Educación a Distancia. Barcelona, Madrid.

Coates-Stephens, S. (1991). Analysis and Acquisition of Proper Names for Robust Text Understanding. Computer Science. London, City University.

Communications of the ACM (2002). "Special Issue: Ontology Applications and Design." Communications of the ACM 45(2): 39-65.

Conesa, J., Storey, V., Sugumaran, V. (2008). "Improving web-query processing through semantic knowledge." Data & Knowledge Engineering 66: 18-34.

Croft, B., Cronen-Townsend, S., Lavrenko,V. (2001). Relevance Feedback and Personalization: A Language Modelling Perspective. DELOS Workshop: Personalisation and Recommended Systems in Diglibs.

Cui, H., Wen, J., Nie, J., Ma, W (2003). "Query Expansion by Mining User Logs." IEEE Transactions on Knowledge and Data Engineering 15(4).

Curtis, J., Cabral, J., Baxter, D. (2005). On the effective use of Cyc in a question answering system. IJCAI Workshop on Knowledge and Reasoning for Answering Questions., Edinburgh, Scotland.

Curtis, J., Cabral, J., Baxter, D. (2006). On the application of the Cyc ontology to word sense disambiguation. Proceedings of the 19th International Florida Artificial Intelligence Research Society Conference.

Das, A., Datar, M., Garg, A., Rajaram, S. (2007). "Google News Personalization: Scalable Online Collaborative Filtering." Proceedings of WWW: 271-280.

Davies, J., Weeks, R (2004). QuizRDF:
Search Technology for the Semantic Web. Proceedings of the 37th Hawaii International Conference on System Sciences.

Davies, J., Duke, A., Kiryakov, A. (2009). Semantic Search. In: Information Retrieval: Searching in the 21st Century. Goker, A., Davies, J. eds Chichester, Wiley**:** Chapter 9, 179-214

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., Harshman, R. (1990). "Indexing by Latent Semantic Analysis." Journal of the Society for Information Science 41(6): 391-407.

Deerwester, S. D., S.T., Landauer, T.K., Furnas, G.W., Beck, L. (1988). " Improving Information Retrieval with Latent Semantic Indexing,." Proceedings of the 51st Annual Meeting of the American Society for Information Science: 36–40.

Dey, A. A., G. (2000). Context Toolkit: Aiding the development of Context-Aware applications. workshop of s/e for wearable and pervasive computing, Ireland June 2000.

Dix, A., Rodden, T., Davies, N., Trevor, J, Friday, A., Palfreyman, K. (2000). "Exploiting space and location as a design framework for interactive mobile systems." ACM Transactions on Computer-Human Interaction (TOCHI) 7(3): 285-321.

Dumais, S., Cuttrell, E., Chen, H., (2001). "Optimizing search by showing results in context." DBLP:conf/chi/DumaisCC01.

Efthimiadis, E. (1992). Interactive Query Expansion and Relevance Feedback for Document Retrieval Systems. City University, London.

Efthimiadis, E. (1996). Query Expansion. Annual Review of Information Systems and Technology (ARIST). M. E. Williams. v31: 121-187.

Eguchi, K. (2005). Query Expansion Experiments using Term Dependence Models. Proceedings of the Fifth NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Question Answering and Cross-Lingual Information Access, Tokyo, Japan.

El-Beltagy, S., Hall, W., DeRoure, D (2001). "Linking in Context." Journal of Digital Information 2(3).

Finkelstein, L., Gabrilovich E., Matias, Y., Rivlin, E., Solan, Z., Wolfman, G., Ruppin E. (2002). "Placing search in context: the concept revisited." TOIS 20(1): 116-131.

Freksa, C., Klippel, A., Winters, S. (2005). "A Cognitive Perspective on Spatial Context." Dagstuhl Seminar 05491.

Fu, G., et al. (2005a). Ontology-Based Spatial Query Expansion in Information Retrieval. ODBASE: OTM Confederated International Conferences.
Fu, L., et al. (2005b). Evaluating the effectiveness of a collaborative querying environment. Proceedings of the 8th International Conference on Asian Digital Libraries ICADL, Lecture notes in computer science, Bangkok, Thailand.

Giaglis, G. (2003). Towards a classification framework for mobile location services. Mobile Commerce: Technology, Theory and Applications. B. S. T. Mennecke, Idea Group Publishing: 67-85.

Goble, C., Stevens, R., Baker, P., Brass, A., Bechhofer, S., Paton, N (1998). "Transparent access to multiple bioinformatics information sources." Conference Intelligent Systems for molecular biology (ISMB), Canada: 25-34.

Goker , A. (1999). Capturing Information Need by Learning User context. In Sixteenth International Joint Conference in Artificial Intelligence: Learning About Users Workshop, Stockholm, Sweden.

Goker , A., Myrhaug, H., Bierig, R. (2009). Context and Information Retrieval. In: Information Retrieval: Searching in the 21st Century. Goker, A., Davies, J. eds. Chichester, Wiley**.** Chapter 7: 131-158.

Gonzalo, J., Verdejo, F., Chugur, I., Cigarran, J. (1998). Indexing with WordNet synsets can improve text retrieval. Coling-ACL 98, Canada.

Grootjen, F., van der Weide, T (2006). "Conceptual query expansion." Data Knowl. Eng 56(2): 174-193.

Grover, C., Hachey, B., Karkaletsis, V., Valarakos, R. , Pazienza, M., Vindigni, M., Cartier, E., Coch, J., (2003). Use of Ontologies for Cross-lingual Information Management in the Web. Proceedings of the Ontologies and Information Extraction International Workshop held as part of the EUROLAN 2003, July 28 - August 8, 2003., Bucarest, Romania.

Gruber, T. (1993). "A translation approach to portable ontologies." Knowledge Acquisition 5(2): 199-220.

Guiraude, L. (2003). Using text analysis techniques to identify
legal ontologie's components. Legont2003, ICAIL 2003 w/s Legal ontologies and web base legal information management.

Halliday, J. (2010). Times loses almost 90% of online readership. [online] available at http://www.guardian.co.uk/media/2010/jul/20/times-paywall-readership [Accessed 22nd June 2011]

Harman, D. (1993). "Overview of TREC-1." Proceedings of the workshop on Human Language Technology: pp61-65.

Harman, D. K. (1988). Towards Interactive Query Expansion. Annual ACM Conference on Research and Development in Information Retrieval archive

Proceedings of the 11th annual international ACM SIGIR conference on Research and development in information retrieval, Grenoble, France.
Harman, D. K. (1992). Relevance Feedback Revisited. ACM SIGIR 15th Conference on Research and Development in Information Retrieval, Copenhagen, Denmark, ACM Press.

Hattori, S., Tezuka, T., Tanaka, K. 2006. (2006). Query Modification Based on Real-World Contexts for Mobile and Ubiquitous Computing Environments. IEEE 7th International Conference on Mobile Data Management (MDM'06).

Hearst, M. (1994). Context and Structure in Automated full-text information access. computer science. berkeley, university of California.

Hearst, M. (1994). "Using Categories to provide context for full text retrieval results." RIAO.

Hearst, M. (1996). "Improving Full text precision on short queries using simple constraints." SDAIR.

Hearst, M. (2000). "Next Generation Web Search: Setting our sites." IEEE Data Engineering Bulleting, Special issue on Next Generation Web Search.

Hearst, M. A. (1992). "Automatic acquisition of hyponyms from large text corpora." 14th International Conference on Computational Linguistics, France.

Hearst, M. A. (1993). TextTiling, a quantitative approach to discourse segmentation. Berkeley., University of California,.

Hearst, M. A., Plaunt, C (1993). Subtopic structuring for full-length document access. Proceedings of the 16th annual international ACM-SIGIR conference on research and development in information retrieval,, Pittsburgh, PA, New York:, ACM.

Hersh, W., Bhupatiraju, R., Price, S (2003). Phrases, Boosting, and Query Expansion Using External Knowledge Resources for Genomic Information Retrieval. TREC.

Hesselman, C., Benz, H., Pawar, P., Liu F., Wegdam, M., Wibbels, M., Broens, T., Brok.J. (2008). Bridging context management systems for different types of pervasive computing environments. Proceedings of the 1st international conference on MOBILe Wireless MiddleWARE, Operating Systems, and Applications (MOBILWARE '08), ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), ICST, Brussels, Belgium.

Hilera, J., Ruiz, F. (2006). Ontologies in Ubiquitous Computing. Proceedings of the I International Conference on Ubiquitous Computing: Applications, Technology and Social Issues., Alcalá de Henares, Madrid, Spain.

Hoashi, K., Matsumoto, K., Inoue, N., Hashimoto, K (2001). "Query Expansion Based on Predictive Algorithms for Collaborative Filtering." SIGIR: 414-415.

Huang, L. (2000). A Survey on Web Information Retrieval Technologies. New York, ECSL, State University of New York.
Huang, X., Huang, Y., Wen, M. (2005). A dual index model for contextual information retrieval. Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval, Salvador, Brazil, ACM Press.

Huhn, A., Hofner, P., Kiessling, W. (2005). Towards Evaluating the Impact of Ontologies on the Quality of a Digital Library Alerting System. Ercimdl.

Hwang, C. H. (1999). Incompletely and imprecisely speaking: Using dynamic ontologies for representing and retrieving information. Proceedings of the 6th International Workshop on Knowledge Representation meets Databases (KRDB'99), Linköping, Sweden.

Hyvonen, E. (2003). "Ontology Based Image Retrieval." www2003: 199.
Ingwersen, P. (2000). Users in Context. ESSIR.

Jansen, B. J., Spink, A., Bateman, J., Saracevic, T. (1998). "Real life information retrieval: A study of user queries on the web." SIGIR Forum 32(1): 5-17.

Jardine, N., Rijsbergen, C. J. (1971). "The use of hierarchic clustering in information retrieval." Information Storage and Retrieval 7: 217-240.

Jarvelin, K., Kekalainen, J., Niemi, T. (2001). "ExpansionTool: Concept-Based Query Expansion and Construction." Information Retrieval(4): 231-255.

Johnson, S., Jourlin, P., Moore,G.L., Jones, K.S., Woodland, P.C. (1999). The Cambridge University spoken document retrieval system. Acoustics, Speech, and Signal Processing, IEEE International Conference on,, Phoenix Arizona.

Jones, S. (1993). "A thesaurus data model for an intelligent retrieval system." Journal of Information Science 19: 167-178.

Jones, S., Gatford, M., Robertson, S., Hancock-Beaulieu, M (1995). "Interactive Thesaurus Navigation: Intelligence Rules OK?" Journal of the American Society for Information Science 46(1): 52-59.

Kallipolitis, L., Karpis, V., Karali, I. (2007). World News Finder: How we Cope without the Semantic Web. Artificial Intelligence and Applications (AIA), IASTED/ACTA Press.

Kashyap, V. (2001). "Design and Creation of Ontologies for Environmental Information Retrieval." AOS Workshop, Rome, November 2001.

Khan, L., Wang, L (2002). Automatic Ontology Derivation using Clustering for Image classification. 8th Intl w/s on Multimedia Information Systems, Arizona.

Kiyavitskaya, N., Katifori, A., Paci, G., Pedrazzi, G., Turra, R. (2010). The Papyrus News Ontology – A Semantic Web Approach to Large News Archives Metadata.

proceedings of the 3rd Workshop on Very Large Digital Libraries (VLDL-2010) conference, Glasgow, UK.

Klink, S., Hust, A., Junker, M., Dengel., A (2002). " Improving Document Retrieval by Automatic Query Expansion Using Collaborative Learning of Term-Based Concepts." Document Analysis Systems: 376-387.

Krovetz, R., Croft., B. (1992). "Lexical ambiguity in information retrieval." ACM Transactions on Information Systems, 10((2)): 115–141.

Lame, G. (2003). Using text analysis techniques to identify
legal ontologie's components. Legont2003, ICAIL 2003 w/s Legal ontologies and web base legal information management.

Lang, K. (1995). NewsWeeder: Learning to Filter Netnews. Proceedings of the 12th International Machine Learning Conference (ML95)

Lawrence, S. (2000). "Context in Web Search." IEEE Data Engineering Bulletin 23(3): 25-32.

Leger, A., Lehtola , A. (2001). Villagra,V.ERCIM News MKBEEM — Developing Multilingual Knowledge-Based Marketplace [online] available at http://www.ercim.org/publication/Ercim_News/enw46/leger.html; http://mkbeem.elibel.tm.fr/ [Accessed 20th December 2004]

Leroy, G., Tolle, K., Chen, H. (2000). "Customizable and Ontology-Enhanced Medical Information Retrieval Interfaces." Methods of Info in Medicine.

Lesk, M. (1969). "Word-Word Associations in Document Retrieval Systems." American Documentation 20(1): 27-38 [Accessed 10th July 2004]

Lifantsev, M. (1998). The Open GRiD Project [online] available at http://www.ecsl.cs.sunysb.edu/~maxim/OpenGRiD/ [Accessed 1st February 2006]

Lin, D. (1999). Automatic identification of non-compositional phrases. Proceedings of ACL, 1999, University of Maryland, Colledge Park, Maryland.

Liu, X., Croft, B. (2002). "Passage Retrieval Based on Language Models." CIKM 2002.

Liu, Y., Shah, M. (2006). Automatic query expansion for news video retrieval. Multimedia and Expo, IEEE international conference on, Toronto.

Loukachevitch, N., Dobrov, B. (2000). "Thesaurus-Based Structural Thematic Summary in Multilingual Information Systems." Machine Translation Review(11): 10-20.

Lowe, D. (2000). Improving web search relevance: using navigational structures to provide search context [online] available at http://ausweb.scu.edu.au/aw2k/papers/lowe/paper.html [Accessed 3rd June 2005]

MacFarlane, A., Secker, A., May, P., Timmis, J. (2010). "An experimental comparison of a genetic algorithm and a hill-climber for term selection." Journal of Documentation 66(4): 513-531.

MacFarlane, A., Tuson, A. (2009). "Local Search: A guide for the Information Retrieval Practitioner." Information Processing and Management 45(1): 159-174.

Magnini, B., Speranza, M. (2002). Merging Global and Specialized Linguistic Ontologies. Proceedings of the Workshop Ontolex-2002 Ontologies and Lexical Knowledge Bases, LREC-2002.

Mandala, R., Takenobu, T., Hozumi, T (1998). The Use of WordNet in Information Retrieval. Use of WordNet in Natural Language Processing Systems: Proceedings of the Conference.

Mandala, R., Tokunaga, T., Tanaka, H (1999). "Combining Multiple Evidence from Different Types of Thesaurus for Query Expansion." Research and Development in Information Retrieval.

Manning, D., Raghavan, P., Schutze, H. (2008). Introduction to Information Retrieval,, Cambridge University Press.

Mena, E., Kashyap, V., Illarramendi, A., Sheth, A (2000). "Imprecise Answers In Distributed Environments: Estimation Of Information Loss For Multi-Ontology Based Query Processing." International Journal of Cooperative Information Systems.

Mizzaro, S., Nazzi, E., Vassena, L. (2008). Retrieval of context-aware applications on mobile devices: How to evaluate? Proceedings of 2nd Information Interaction in Context (IIiX), New York: ACM (United States).

Morris, J., Hirst, G. (1991). "Lexical Cohesion Computed by Thesaural Relations as an Indicator of the Structure of Text." Computational Linguistics . 17: pp. 21-43.

Morris, J., Hirst, G. (1991). "Lexical cohesion, the thesaurus, and the structure of text." Computational linguistics 17(1): 21-48.

Mountain, D., MacFarlane, A. (2007). "Geographic information retrieval in a mobile environment: evaluating the needs of mobile individuals." Journal of Information Science 33(5): 515-530.

Mozetic, I., Bojadzijev, D. (2006). Reasoning with temporal context in news analysis. Contexts and Ontologies workshop ECAI-06, Riva del Garda, Italy.

Navigli, R., Velardi, P (2003). An analysis of Ontology-based query expansion strategies. Workshop on Adaptive Text Extraction and Mining (ATEM 2003), in the 14th European Conference on Machine Learning (ECML 2003),September 22-26th,, Cavtat-Dubrovnik, Croatia.

Nilsson, K., Hjelm, H, Oxhammar, H (2005). SUiS – cross-language ontology-driven information retrieval in a restricted domain. Proceedings of the 15th NODALIDA conference,, Joensuu.

O'Donovan, J., Smyth, B (2005). Trust in recommender systems. Proceedings of the 10th international conference on Intelligent user interfaces, San Diego, California, USA.

Ogawa, Y., Mano, H. (2001). Selecting expansion terms in automatic query expansion. ACM-SIGIR Annual Int. Conf. on Research and Development in Information Retrieval.

Ogawa, Y., Mano, H., Narita, M., Honma, S (2000). Structuring and expanding queries in the probabilistic model. TREC 2000, Maryland.

O'Sullivan, D., McElligott, A, Sutcliffe, R (1995). Augmenting the Princeton WordNet with a Domain Specific Ontology. Proc. Workshop on Basic Ontological Issues in Knowledge Sharing, International Joint Conference on Artificial Intelligence (IJCAI-95), Montreal, Canada.

Paris, L., Tibbo, H. (1998). "Freestyle vs. Boolean: A comparison of partial and exact match retrieval systems." Information Processing & Management. 34(Issues 2-3): 175-190.

Pavel, S., Sinopalnikova, A., Pon´y, M. (2003). Thesauri and Ontologies for Digital Libraries. Proceedings of the 5th Russian Conference on Digital Libraries RCDL, St.-Petersburg, Russia.

Peat, H., Willett, P (1991). "The limitations of term co-occurrence data for query expansion in document retrieval systems." American Society for Information Science and Technology (JASIST) 42(5): 378-383.

Pustejovsky, J. (1995). The core lexical engine: The contextual determination of word sense [online] available at http://www.cs.tufts.edu/~jacob/isgw/Pustejovsky.html [Accessed 6th September 2002]

Rada, R., Barlow, J., Potharst, J., Zanstra, P., Bijstra, D. (1991). "Document Ranking using an enriched Thesaurus." Journal of Documentation 47(3): 240-253.

Rhodes, B., Maes, P (2000). "Just-in-time information retrieval agents." IBM Systems Journal 39(3 & 4).

Robertson, A., Willett, P (1993). "A comparison of Spelling-Correction Methods for the Identification of Word Forms in Historical Text Databases." Literary & Linguistic Computing 8(3): 143-152.

Robertson, S. (1986). "On Relevance Weight Estimation and Query Expansion." Journal of Documentation. 42((3)): 182-188.
Robertson, S. (1990). "On term selection for query expansion." Journal of Documentation 46(4): 359-364.

Robertson, S., Walker, S . (1999). Okapi/Keenbow at TREC-8. TREC.

Robertson, S., Walker, S., HancockBeaulieu, M., Gatford, M., Payne, A. (1995). Okapi at TREC-4. The Fourth Text REtrieval Conference (TREC-4).

Robertson, S. E., Sparck-Jones, K. (1976). "Relevance weighting of search terms." Journal of the American Society of Information Science. 27: p129-146.

Rocchio, R., Ed. (1971). Relevance Feedback in Information Retrieval. The SMART Retrieval System: Experiments in Automatic Document Processing. Englewood Cliffs NJ, Prentice-Hall.

Rosenfeld, R. (2000). "Two Decades of Statistical Language Modeling: Where do we go from here?" Proceedings of the IEEE 88(8).

Ruthven, I. (2004). "... and this set of words represents the user's context ..." Sigir, Information retrieval in context workshop, Sheffield, ACM.

Ruthven, I., Lalmas, M. (2003). "A survey on the use of relevance feedback for information access systems." The Knowledge Engineering Review archive 18(2): 95 - 145.

Saias, J. (2002). Semantic Enrichment of a Web Legal Information Retrieval System. Jurix 2002, Conference on Legal knowledge and IS, London, IOS press.

Salton, G. (1971). The SMART Retrieval System - Experiments in Automatic Document Processing, Prentice Hall.

Salton, G. (1989). Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer. Reading, MA, Addison-Wesley.

Salton, G., Mcgill, J (1983). Introduction to Modern Information Retrieval. New York, Mc_graw Hill.

Sanchez-Fernandez, L., Fernandez-Garcia, N., Bernardi, A., Zapf, L., Penas, A., Fuentes, M. (2005). An experience with Semantic Web technologies in the news domain,. 4th International Semantic Web Conference ISWC-2005, Workshop Semantic Web Case Studies and Best Practices for eBusiness, Galway, Ireland.

Sanderson, M. (2000). "Retrieving with good sense." Information Retrieval 2(1): 49-69.

Sanderson, M. (2004). A study of user interaction with a concept based interactive query expansion support tool (CiQuest) which is integrated into Okapi. Advances in Information Retrieval: 26th European Conference on IR Research, ECIR 2004, April 5-7, 2004., Sunderland, UK, Springer-Verlag Heidelberg.

Sanderson, M., Zobel, J. (2005). Information Retrieval System Evaluation: Effort, Sensitivity, and Reliability. SIGIR 05, Salvador, Brazil, ACM.

Schatz, B., Johnson, E., Cochrane, P., Chen, H. (1996). Interactive Term Suggestion for Users of Digital Libraries. ACM Digital Library Conference.

Schilit, B., Adams, N., Want, R. (1994). Context-aware computing applications. IEEE Workshop on Mobile Computing Systems and Applications (WMCSA'94), Santa Cruz, CA, US.

Scott, S., Matwin, S. (1998). Text Classification using WordNet Hypernyms. Coling-ACL Workshop, Usage of WordNet in NLP Systems, Canada.

Sebastiani, J. (2002). Machine learning in automated text categorization. ACM Computing Surveys. 34 (1). 1-47.

Shoval, P. (1985). "Principles, procedures and rules in an expert system for information retrieval." Information Processing and Management 21: 475-487.

Sihvonen, A., Vakkari, P. (2004). "Subject knowledge improves interactive query expansion assisted by a thesaurus." Journal of Documentation 60(6): 673-690.

Smeaton, A. (1982). The Retrieval Effects of Query Expansion on a feedback document retrieval system. Dept of Computer Science. Dublin, University college.

Smeaton, A., Van Rijsbergen, C.J (1983). "The retrieval effects of query expansion on a feedback document retrieval system." Computer Journal 26(3): 239-246.

Soergel, D. (1999). "The rise of ontologies or the reinvention of classification." Journal of the American Society for Information Science 50(12): 1119-1120.

Song, M., Song, I., Hu, X., Allen, R. (2007).”Integration of association rules and ontologies for semantic query expansion”. Data & Knowledge Engineering, 63(1). 63-75.

Sparck -Jones, K. (1973). "An evaluation of Query Expansion by addition of clustered terms for a document retrieval system." Information Storage and Retrieval 9(6): 339.

Sparck-Jones, K. (1979). "Experiments in Relevance Weighting of Search Terms." Information Processing and Management" 15(3): 133-144.

Sparck-Jones, K., Walker, S., Robertson, S (2000). "A probabilistic model of information retrieval: development and status: Part 1 and Part 2." Information Processing and Management" 36(6): 779-840.

Spink, A. (1997). "Study of Interactive Feedback During Mediated Information Retrieval." JASIS 48(5): 382-394.

Srivastava, J., Cooley, R., Deshpande, M., Tan, P (2000). "Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data." SIGKDD Explorations 1(2): 12-23.

Stairmand, M. (1996). A Computational Analysis of Lexical Cohesion with Applications In Informatton Retrieval. Center for Computational Linguistics. Manchester, UMIST.

Stairmand, M. (1997). "Textual Context Analysis for Information Retrieval." SIGIR: 140-147.

Suomela, S., Kekalainen, J (2005). Ontology as a search-tool: A study of real users' query formulation with and without conceptual support. ECIR 2005, Santiago de Compostella Spain, March 2005, Springer-Verlag Berlin Heidelberg 2005.

Tenenboim, L., Shapira, B., Shoval, P. (2008). Ontology-Based Classification Of News In An Electronic Newspaper, I. International Conference "Intelligent Information and Engineering Systems" INFOS 2008, Varna, Bulgaria.

Tombros, A., Ruthven, I., Jose, J (2001). "A study on the use of summaries and summary based query expansion for a question answering task." 23rd BCS European Annual Colloquium on IR Research (ECIR).

Torrissen, B. C. (1998). Dewey goes surfing: Agent-based information retrieval and classfication support, norwegian university of science and technology.

Vakkari, P., Jones, S., Mcfarlane, A., Sormunen, E (2003). "Query Exhaustivity, relevance feedback and search success in automatic and interactive query expansion." American Society for Information Science and Technology (JASIST) 60(2).

Vallet , D., Castells , P., Fernández , M., Mylonas, P., Avrithis, Y. (2008). "Personalized content retrieval in context using ontological knowledge." The Knowledge Engineering Review archive 23(1): 73-100.

Van Rijsbergen, C. J. (1979). Information Retrieval, Butterworth-Heinemann.

Vechtomova, O., Robertson, S., Jones, S (2003). Query expansion with long-span collocates. Information Retrieval, Kluwer academic publishers.

Velliard, D.The XML C parser and toolkit of Gnome - libxml [online] available at http://www.xmlsoft.org/ [Accessed 11th October 2008]

Voorhees, E. (1993). "Using wordnet to disambiguate word senses for text retrieval." ACM SIGIR: 171-180.

Voorhees, E. (1994). Query expansion using Lexical-semantic relations. Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval table of contents, Dublin, Ireland.

Voorhees, E., Harman, D. (1997). Overview of the Sixth Text REtrieval Conference (TREC-6). In proceedings of the Sixth Text REtrieval Conference (TREC-6), Gaithersburg, Maryland, USA.

Voorhees, E. (2000). Variations in relevance judgements and the measurement of retrieval effectiveness. Information Processing and Management. 36. 697-716.

Walker, S., Hancock-Beaulieu, M. (1991). Okapi at City: an evaluation facility for interactive IR., British Library Research.

Wang, X., Gu, T., Zhang, D., Pung, H. (2004). Ontology Based Context Modelling and Reasoning using OWL. Proc 2nd IEEE Conf. Pervasive Computing and Communications, Workshop on Context Modelling and Reasoning.

Weiser, M. (1991). "The computer for the 21st century." Scientific American Special Issue on Communications, Computers, and Networks.

Xu, J., Croft, B. (1999). "Cluster-based Language Models for Distributed Retrieval." ACM SIGIR.

Xu, J., Croft, B. (2000). "Improving Effectiveness of Information Retrieval with Local Context Analysis." ACM Transactions on Information Systems 18(1): 79-112.

Yu, C., Meng, W., Liu, K (2001). "Efficient and effective metasearch for text databases incorporating linkages among documents." SIGMOD.

Zhang, C., Cui, B., Cong, G., Wang, Y. (2009). A Revisit of Query Expansion with different semantic levels. Lecture Notes in Computer Science, Database systems for Advanced Applications., Springer Berlin/Heidelberg.

APPENDIX A – Recall Precision Graphs

A1 Experiment 5.1 (20 documents and 20 Terms)

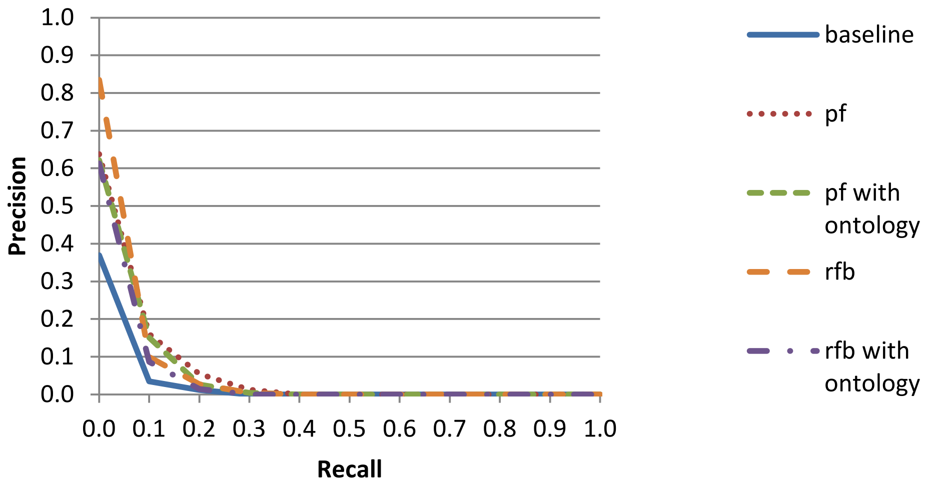


Figure A1.1 Topic 101-150 Precision-Recall Graphs

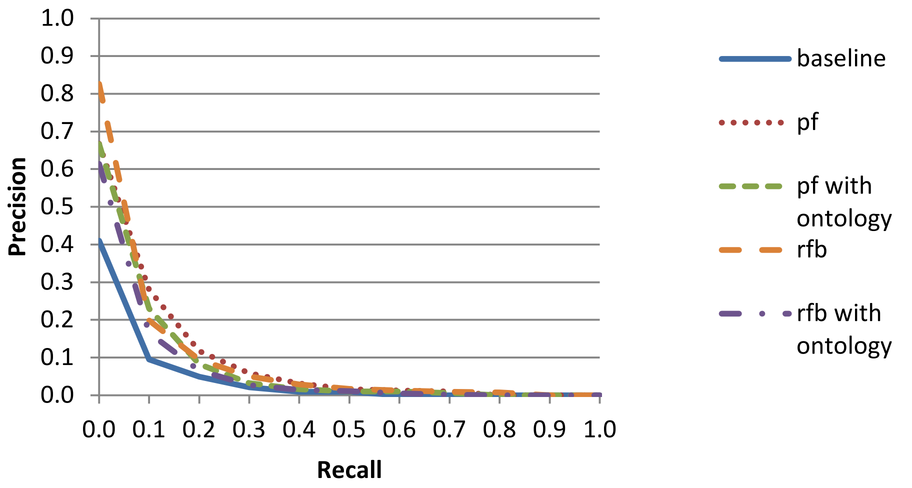


Figure A1.2 Topic 151-200 Precision-Recall Graphs

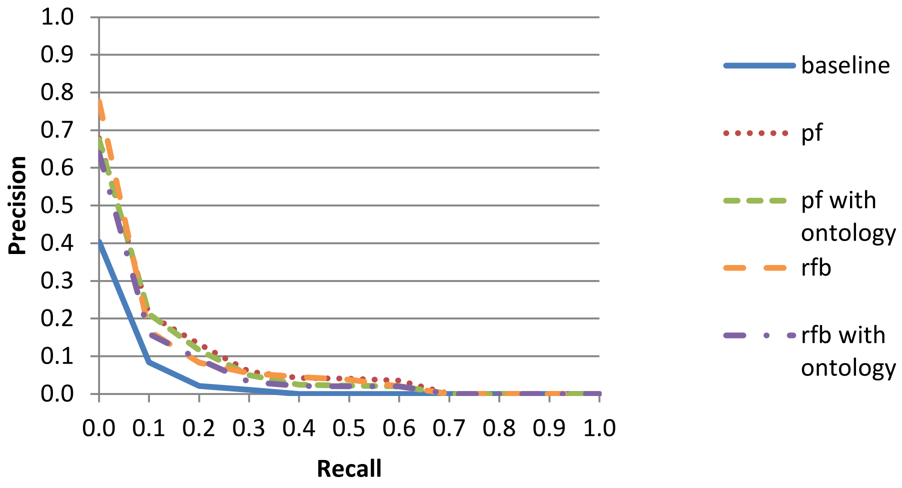


Figure A1.3 Topic 201-250 Precision-Recall Graphs

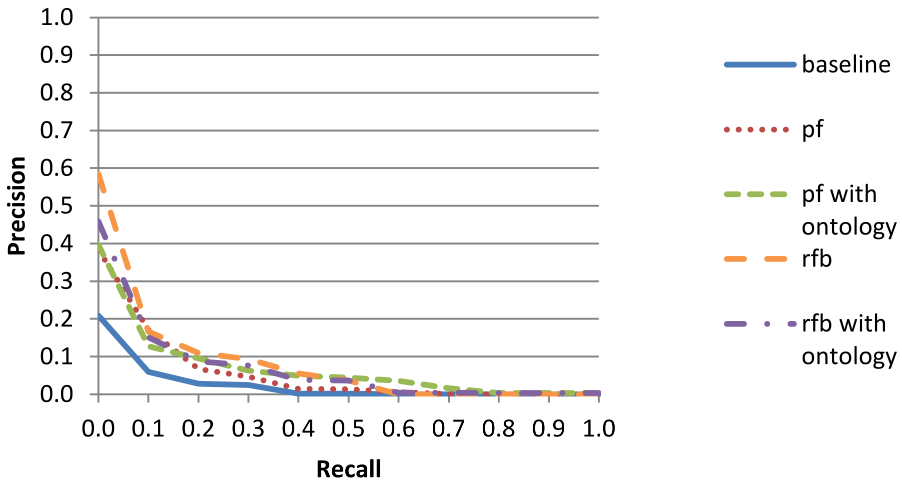


Figure A1.4 Topic 251-300 Precision-Recall Graphs

A2 Precision-Recall Graphs for Experiment 5.2 (20 documents and 5 Terms)

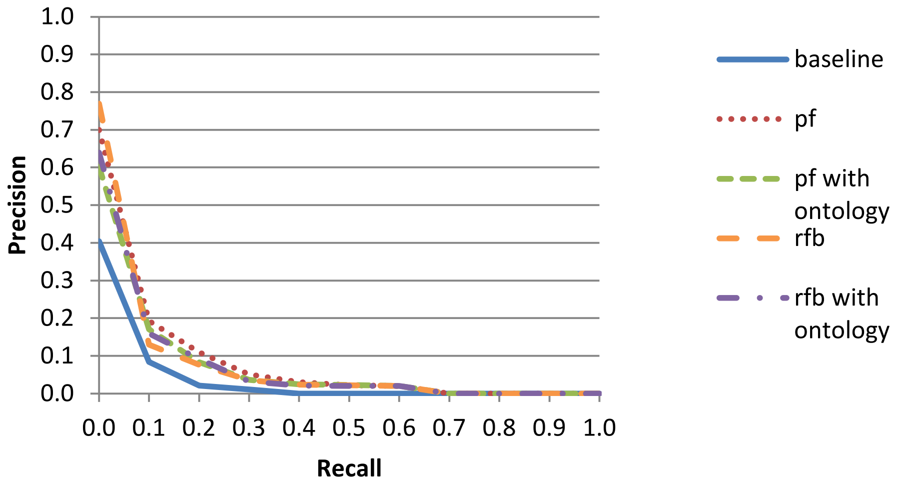


Figure A2.1 Topic 201-250 PF Results – Precision-Recall Graphs

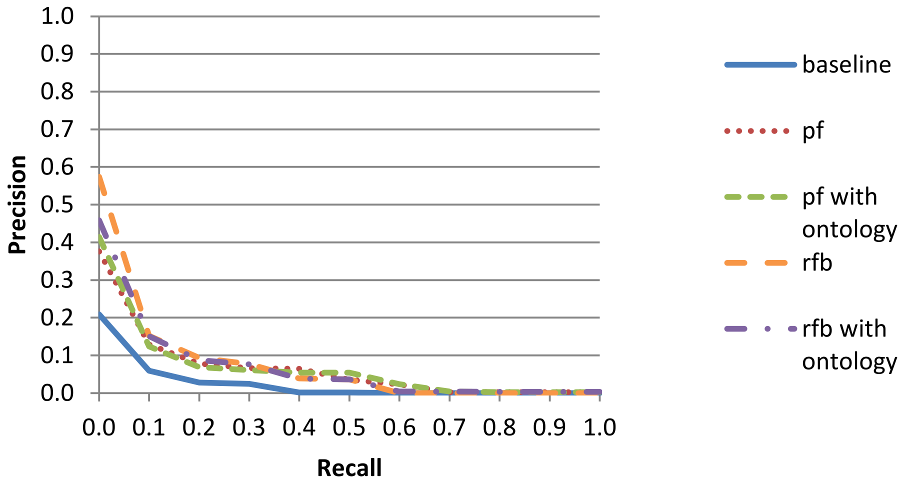


Figure A2.2 Topic 251-300 PF Results – Precision-Recall Graphs

A3 Precision-Recall Graphs for Experiment 5.3 (20 documents and 10 Terms)

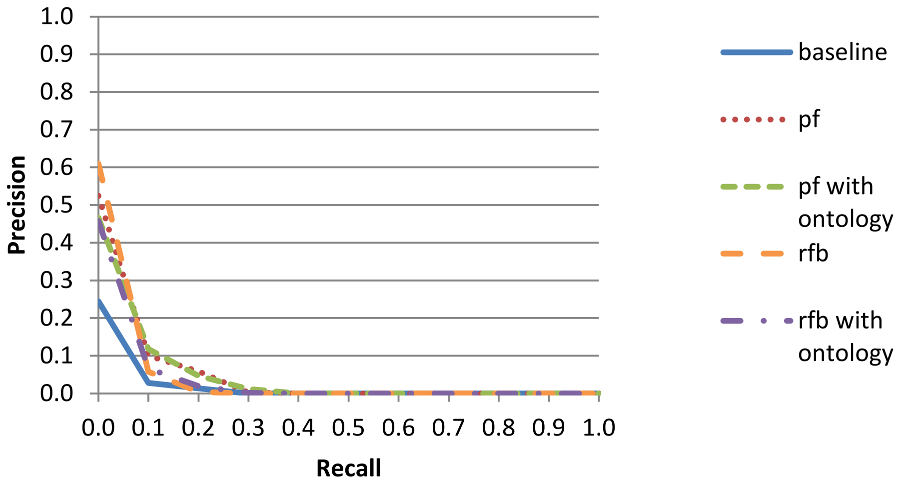


Figure A3.1 Topic 51-100 PF Results – Precision-Recall Graphs

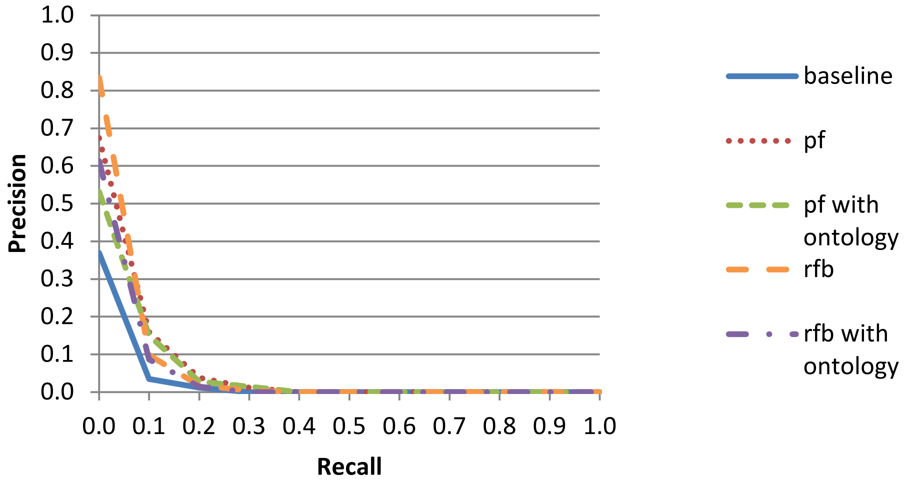


Figure A3.2 Topic 101-150 PF Results – Precision-Recall Graphs

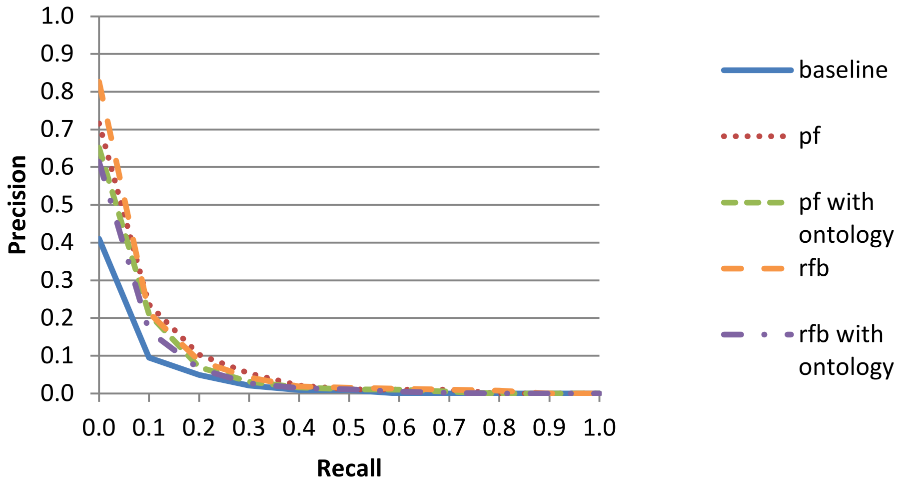


Figure A3.3 Topic 151-200 Results – Precision-Recall Graphs

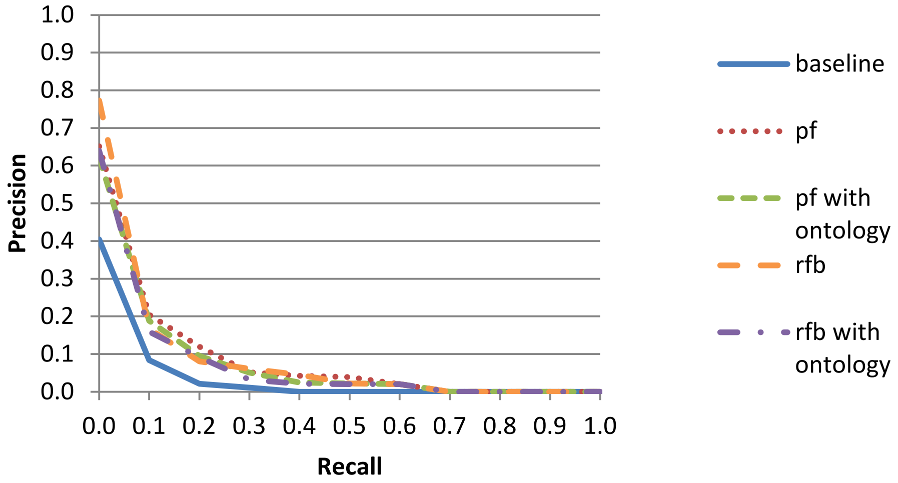


Figure A3.4 Topic 201-250 PF Results – Precision-Recall Graphs

A4 Precision-Recall Graphs for Experiment 5.4 (20 documents and 15 Terms)

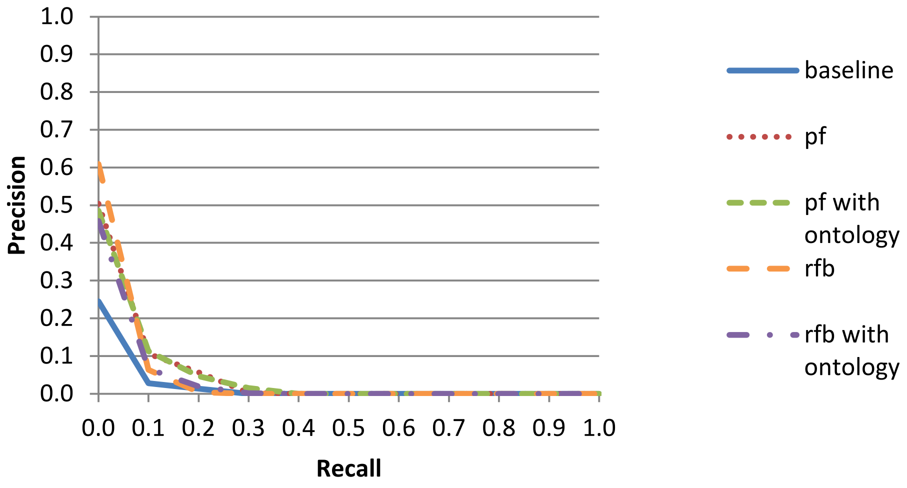


Figure A4.1 Topic 51-100 PF Results – Precision-Recall Graphs

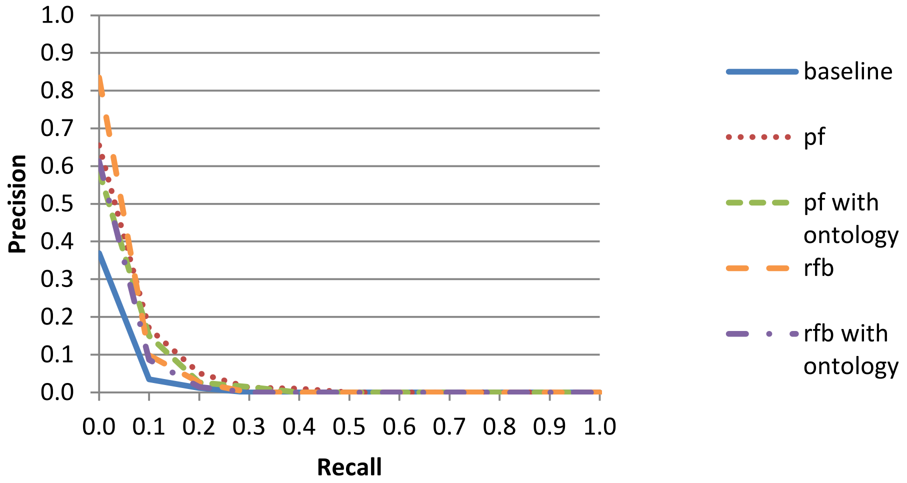


Figure A4.2 Topic 101-150 PF Results – Precision-Recall Graphs

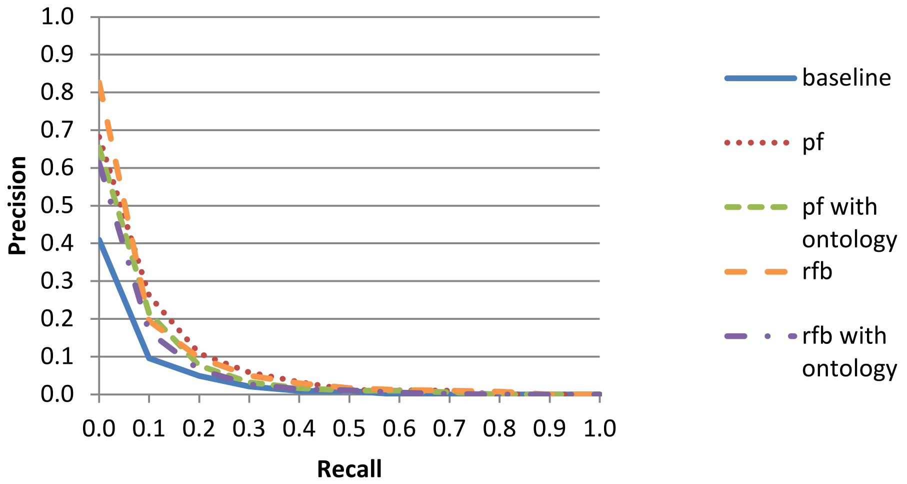


Figure A4.3 Topic 151-200 PF Results – Precision-Recall Graphs

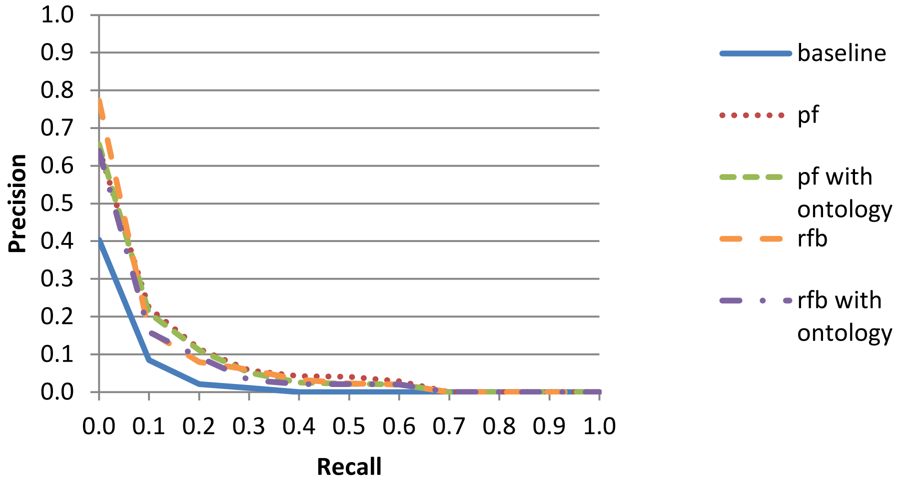


Figure A4.4 Topic 201-250 PF Results – Precision-Recall Graphs

A5 **Precision at Rank Graphs for Experiment 5.4 (20 documents and 15 Terms)**

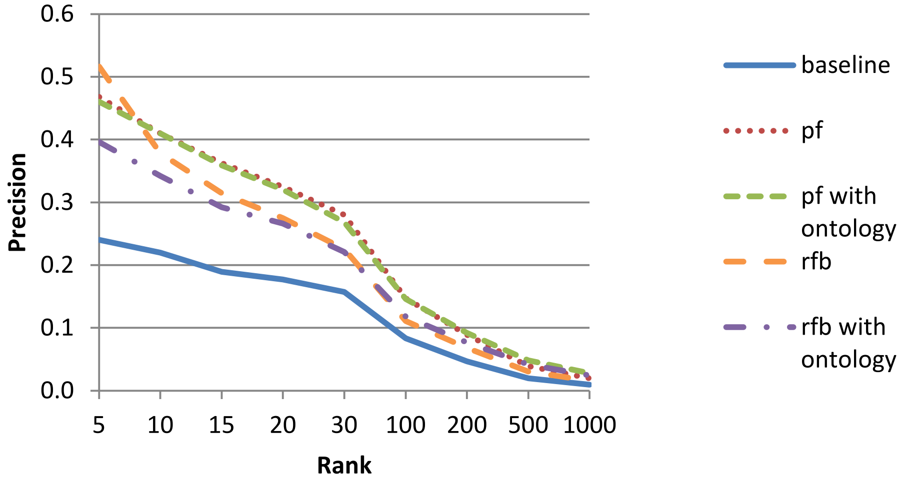


Figure A5.1 Topic 151-200 Results – Document Level Averages

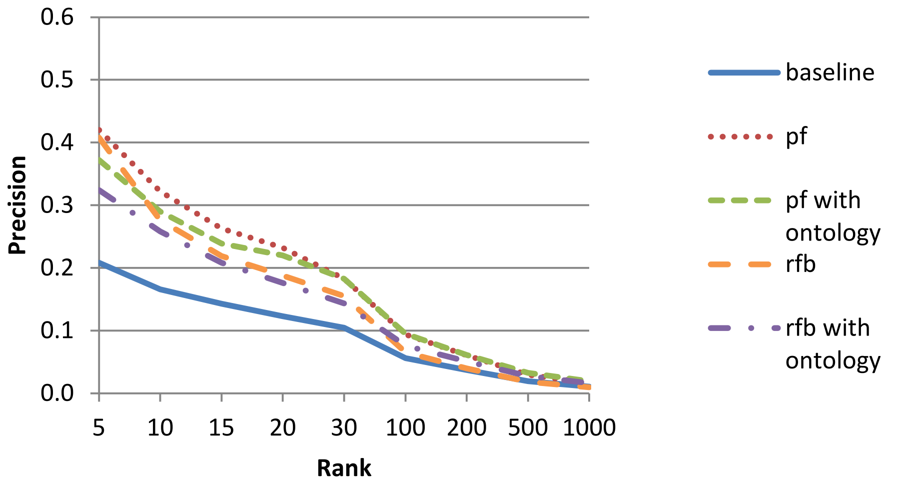


Figure A5.2 Topic 201-250 PF Results – Document Level Averages

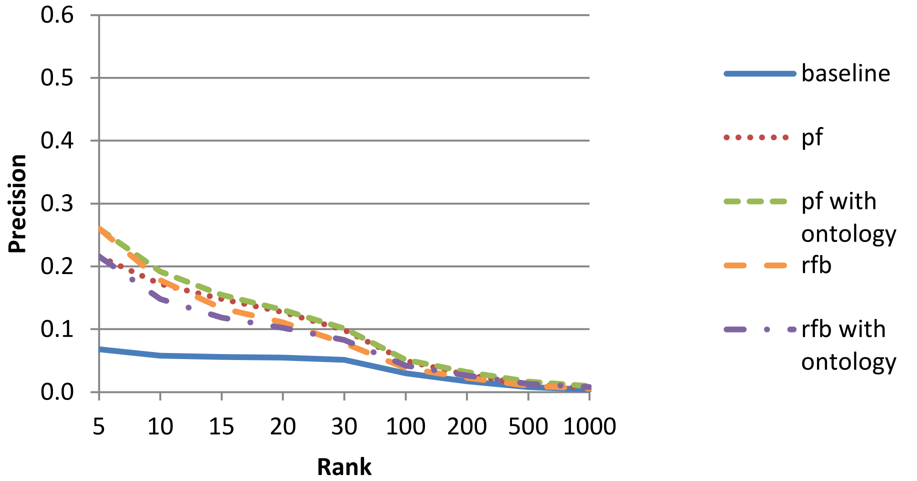


Figure A5.3 Topic 251-300 PF Results – Document Level Averages

A6 Precision-Recall Graphs for Experiment 5.5 (20 documents and 100 Terms)

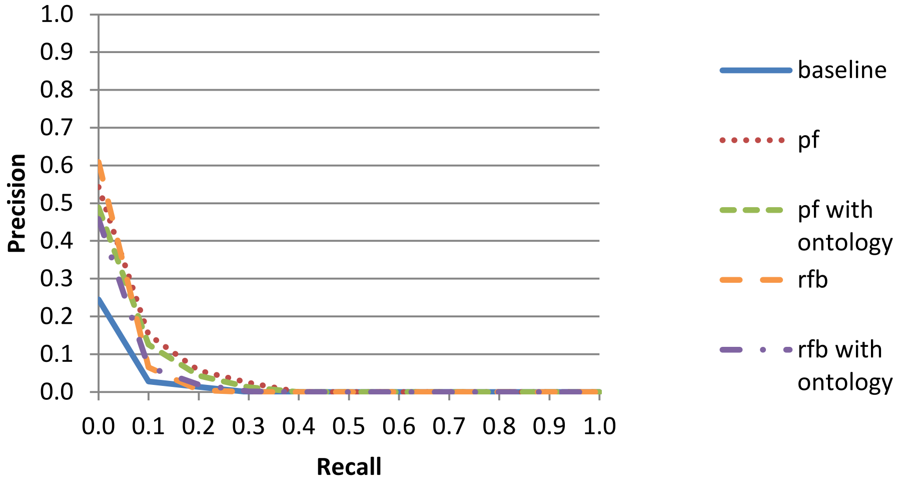


Figure A6.1 Topic 51-100 PF Results – Precision-Recall Graphs

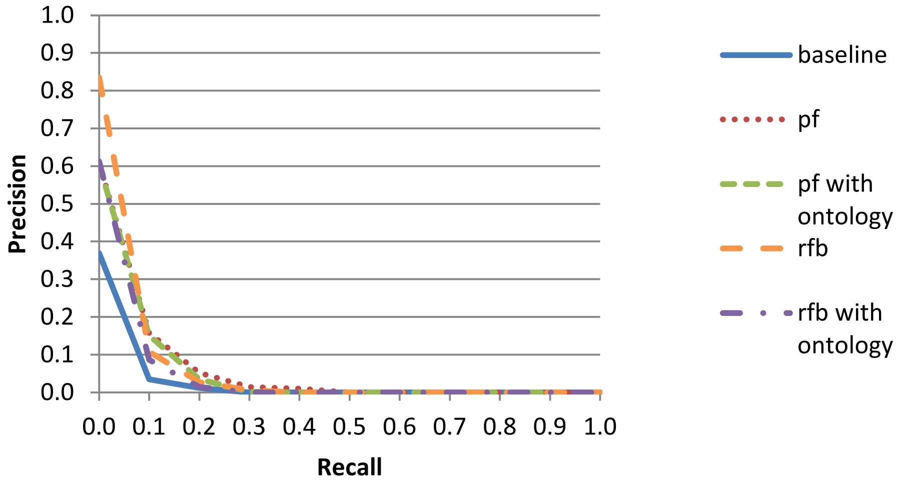


Figure A6.2 Topic 101-150 PF Results – Precision-Recall Graphs

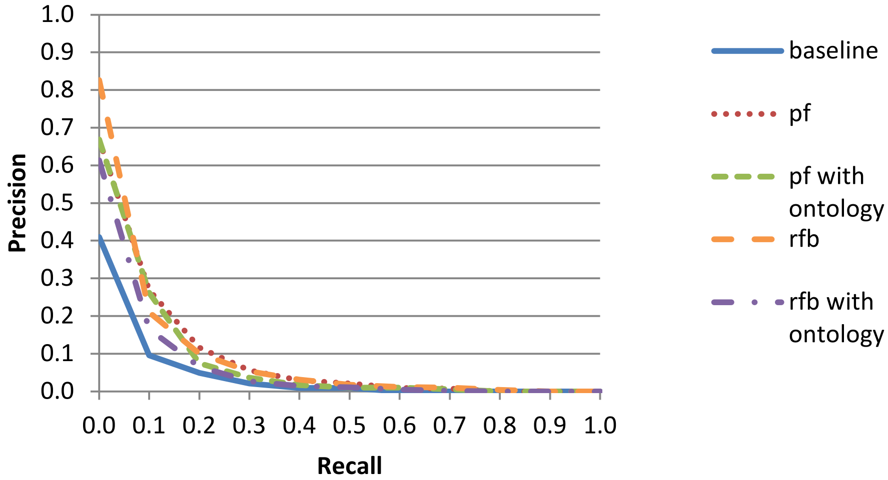


Figure A6.3 Topic 151-200 Results – Precision-Recall Graphs

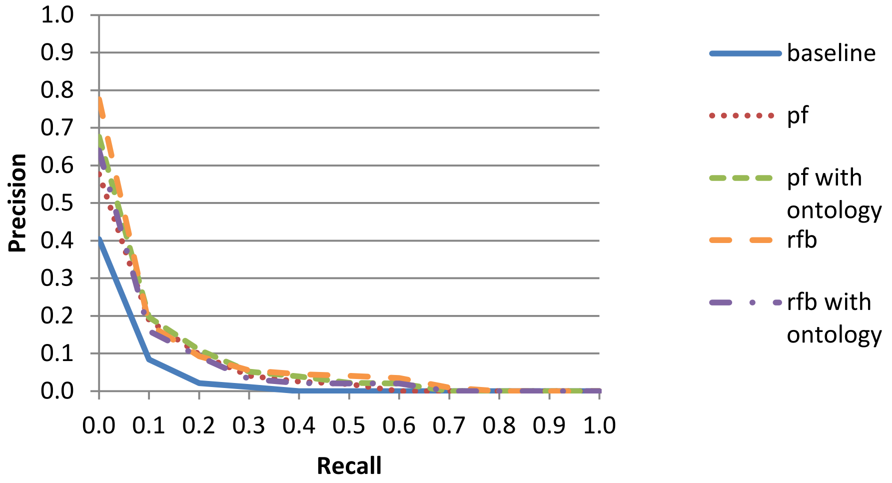


Figure A6.4 Topic 201-250 PF Results – Precision-Recall Graphs

A7 **Precision-Recall Graphs for Experiment 5.6 (20 documents and 200 Terms)**

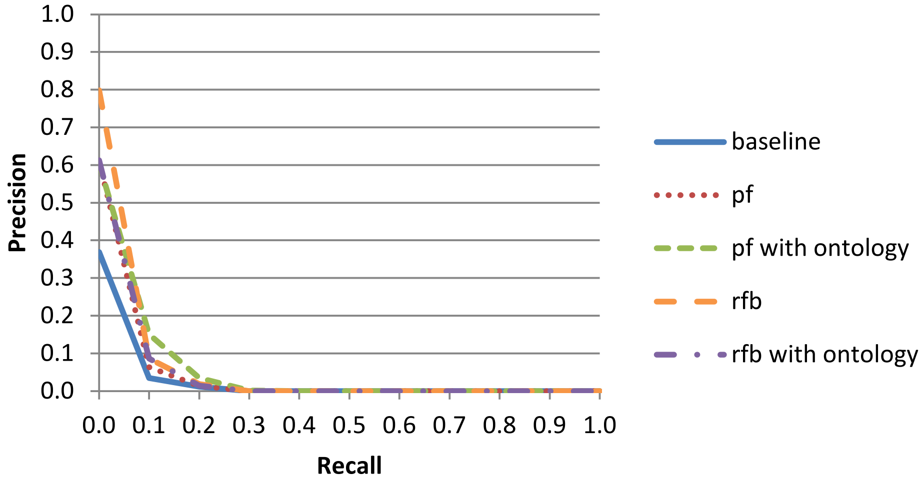


Figure A7.1 Topic 101-150 PF Results – Precision-Recall Graphs

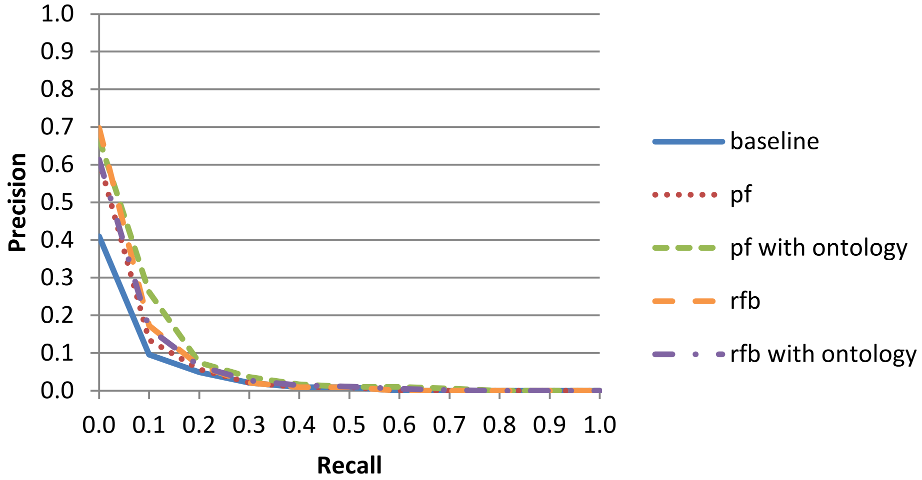


Figure A7.2 Topic 151-200 Results – Precision-Recall Graphs

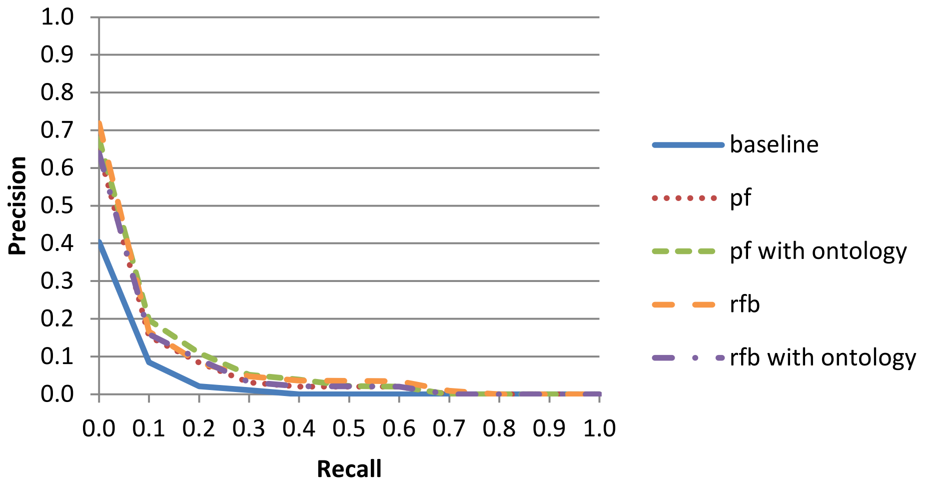


Figure A7.3 Topic 201-250 PF Results – Precision-Recall Graphs

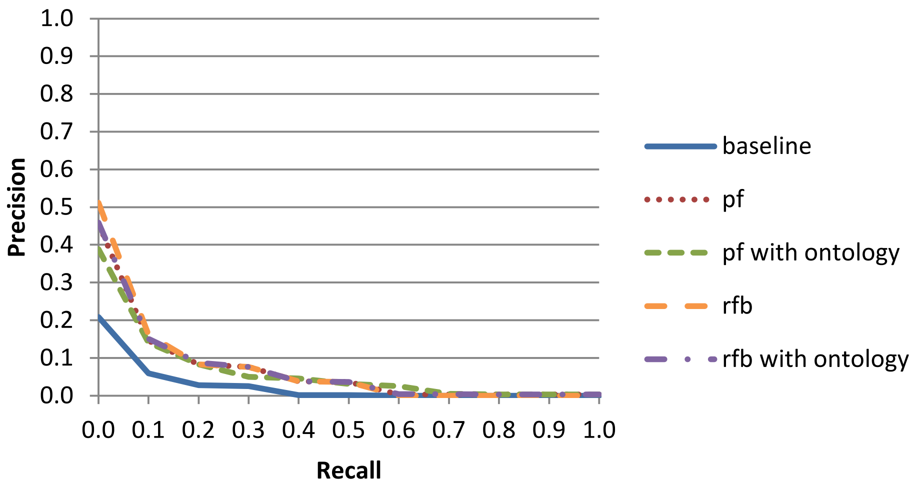


Figure A7.4 Topic 251-300 PF Results – Precision-Recall Graphs

A8 Precision-Recall Graphs for Experiment 5.7 (5 documents and 20 Terms)

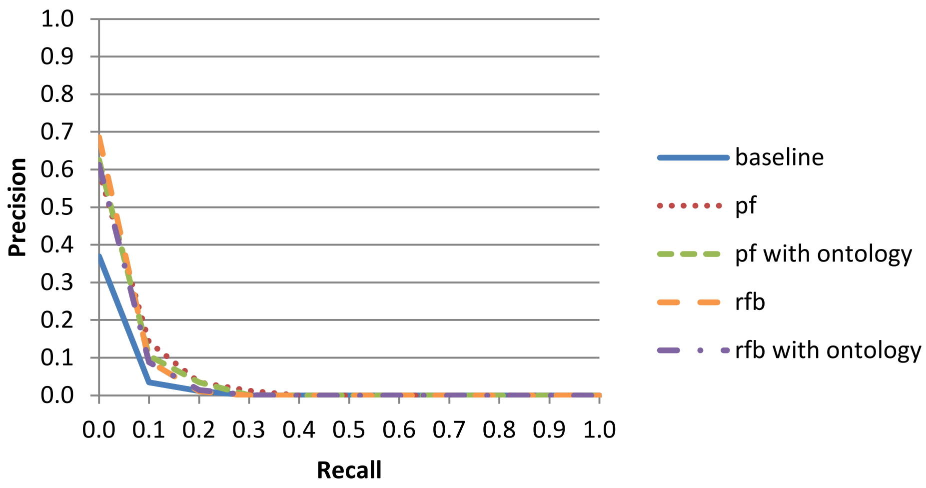


Figure A8.1 Topic 101-150 Pseudodback Results – Precision-Recall Graphs

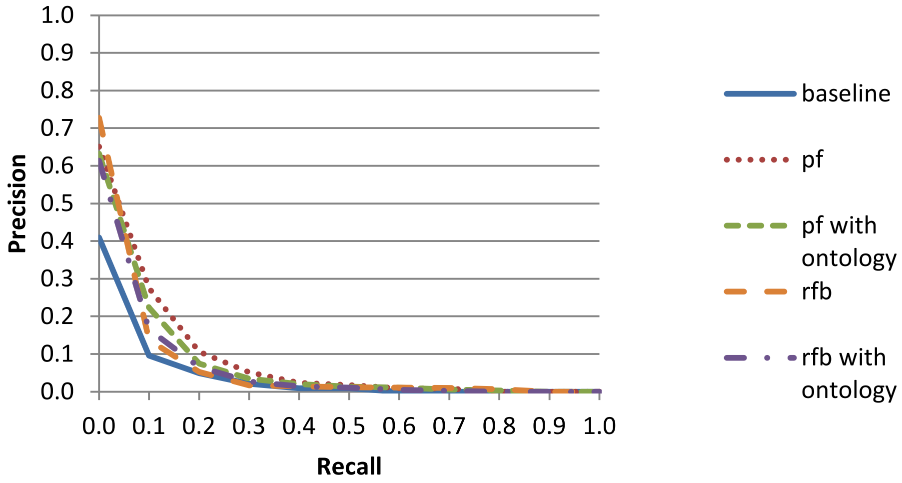


Figure A8.2 Topic 151-200 PF Results – Precision-Recall Graphs

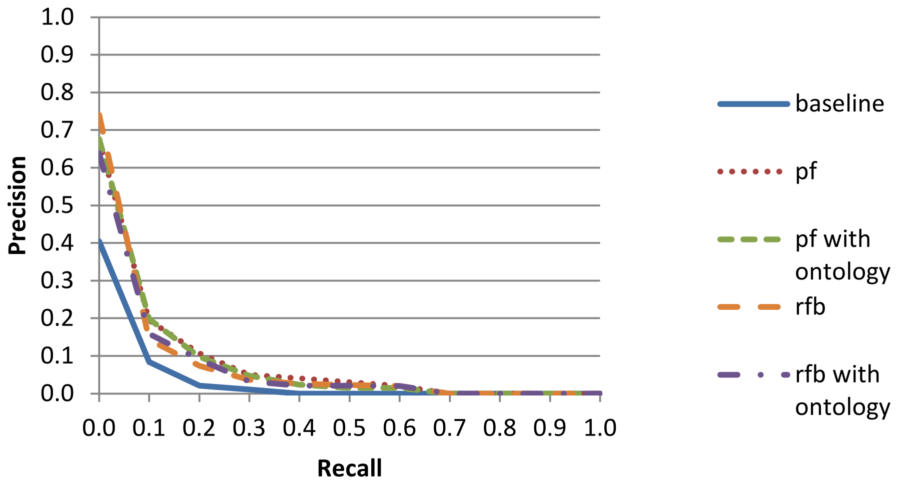


Figure A8.3 Topic 201-250 Results – Precision-Recall Graphs

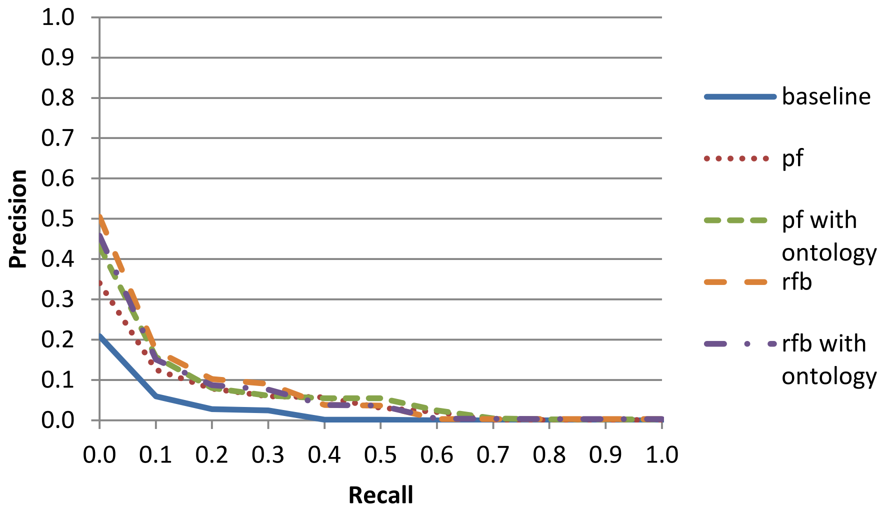


Figure A8.4 Topic 251-300 PF Results – Precision-Recall Graphs

A9 Precision-Recall Graphs for Experiment 5.8 (10 documents and 20 Terms)

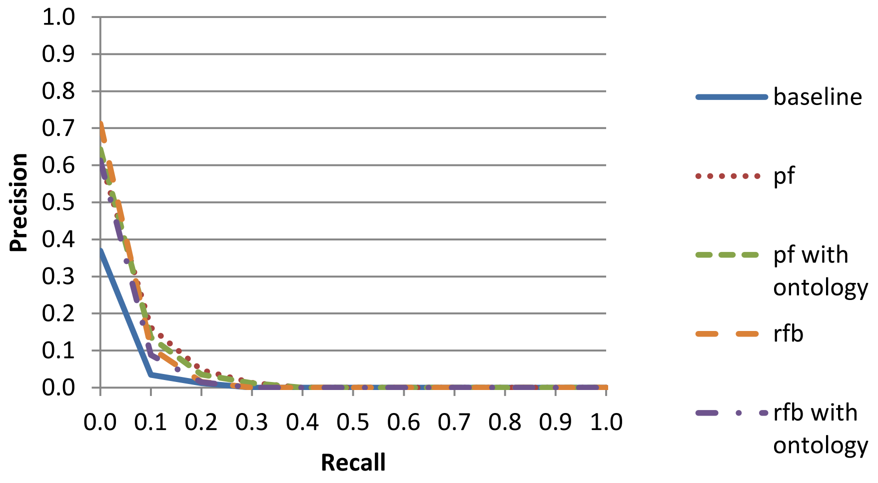


Figure A9.1 Topic 101-150 PF Results – Precision-Recall Graphs

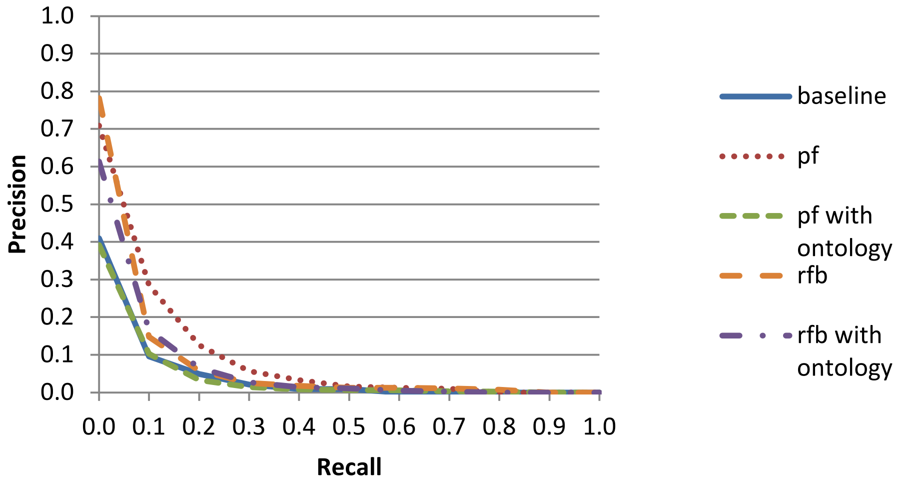


Figure A9.2 Topic 151-200 PF Results – Precision-Recall Graphs

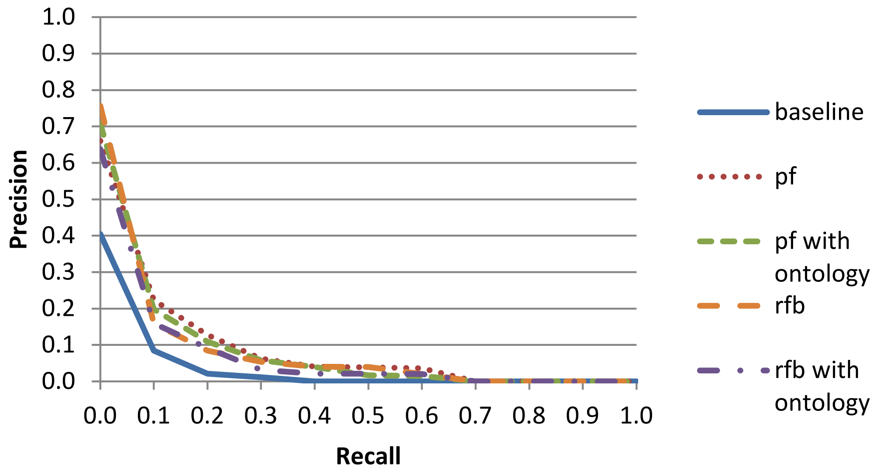


Figure A9.3 Topic 201-250 PF Results – Precision-Recall Graphs

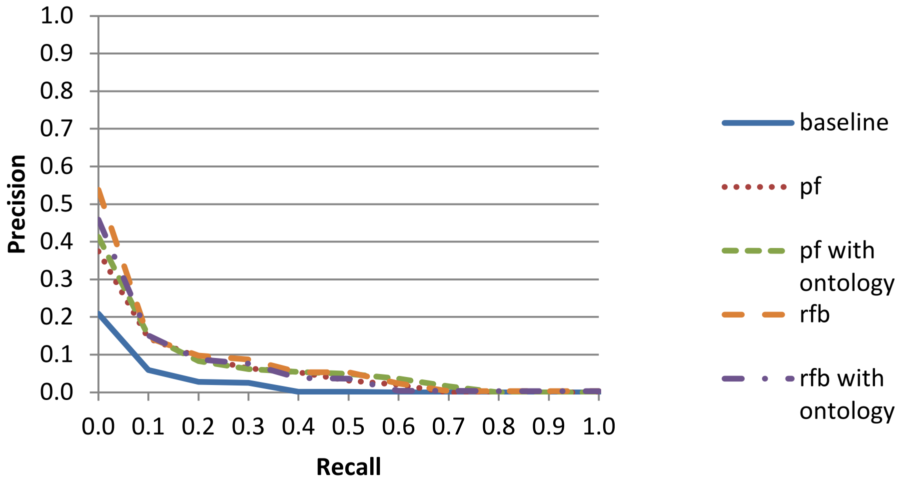


Figure A9.4 Topic 251-300 PF Results – Precision-Recall Graphs

A10 Precision-Recall Graphs for Experiment 5.9 (15 documents and 20 Terms)

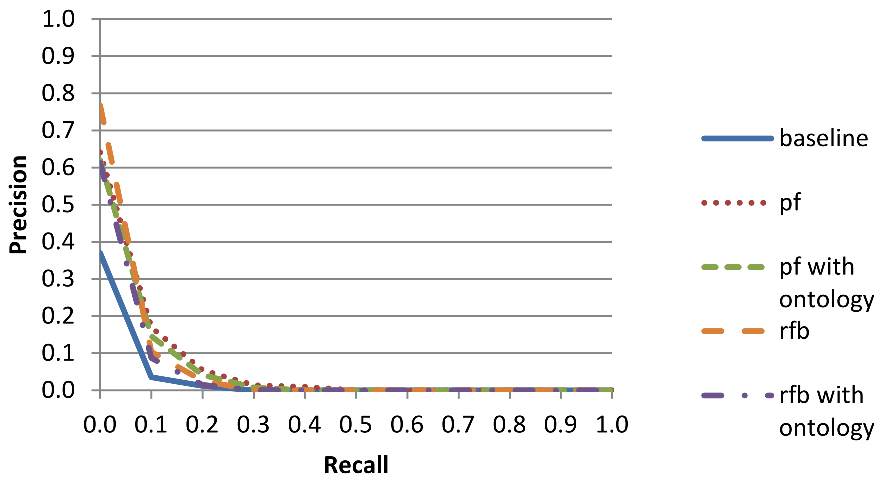


Figure A10.1 Topic 101- Results – Precision-Recall Graphs

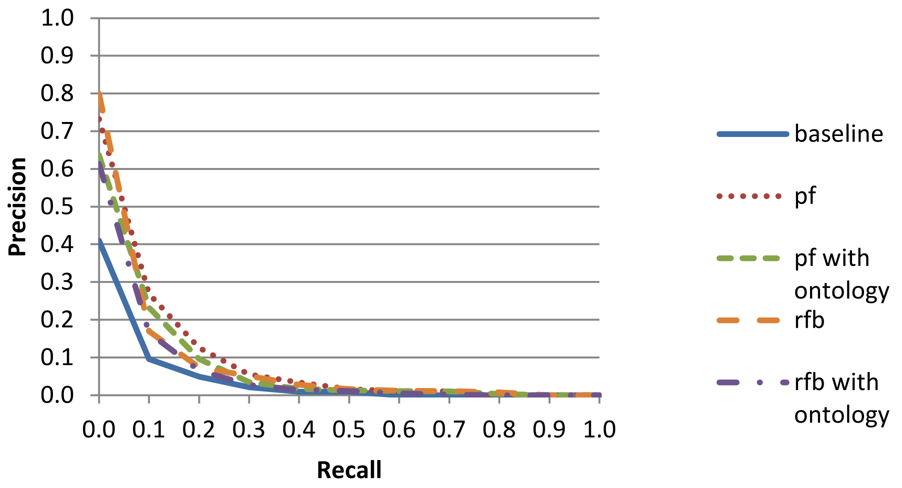


Figure A10.2 Topic 151-200 PF Results – Precision-Recall Graphs

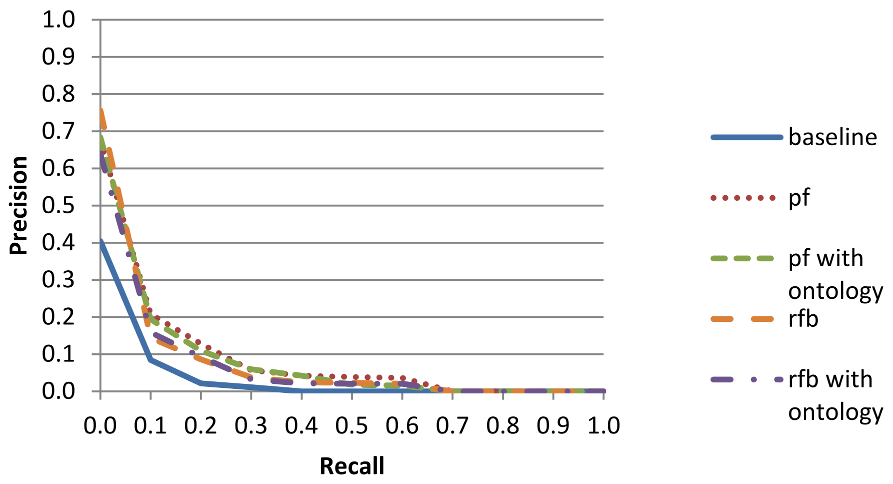


Figure A10.3 Topic 201-250 PF Results – Precision-Recall Graphs

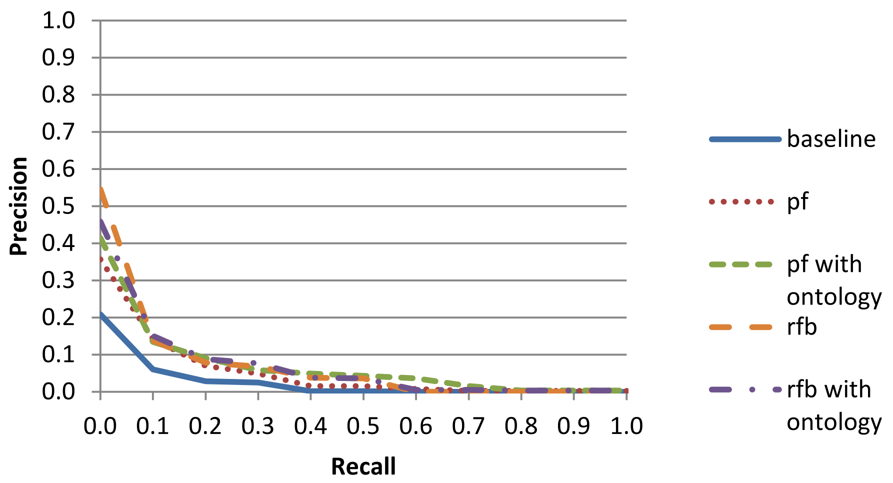


Figure A10.4 Topic 251-300 PF Results – Precision-Recall Graphs

A11 Precision-Recall Graphs for Experiment 5.10 (100 documents and 20 Terms)

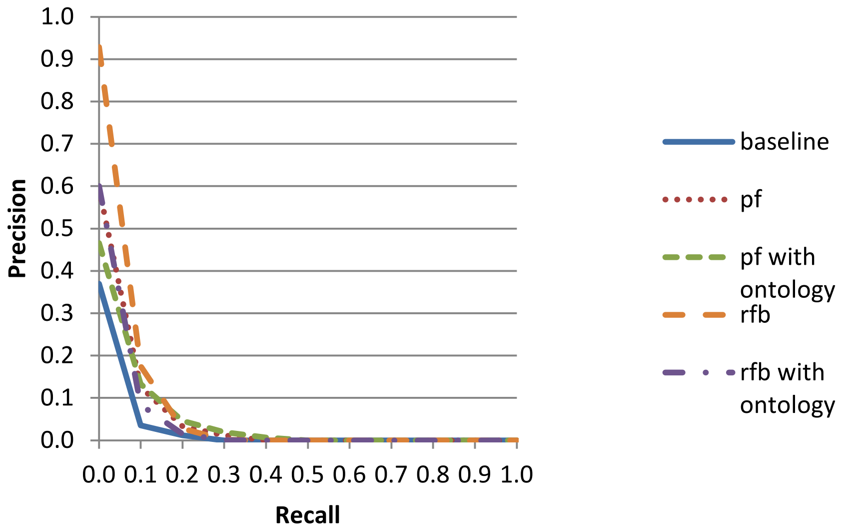


Figure A11.1 Topic 101-150 PF Results – Precision-Recall Graphs

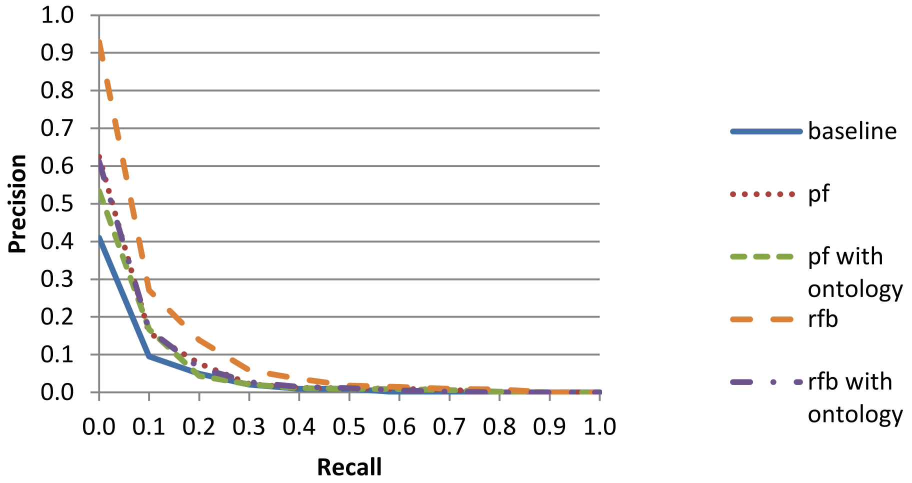


Figure A11.2 Topic 151-200 PF Results – Precision-Recall Graphs

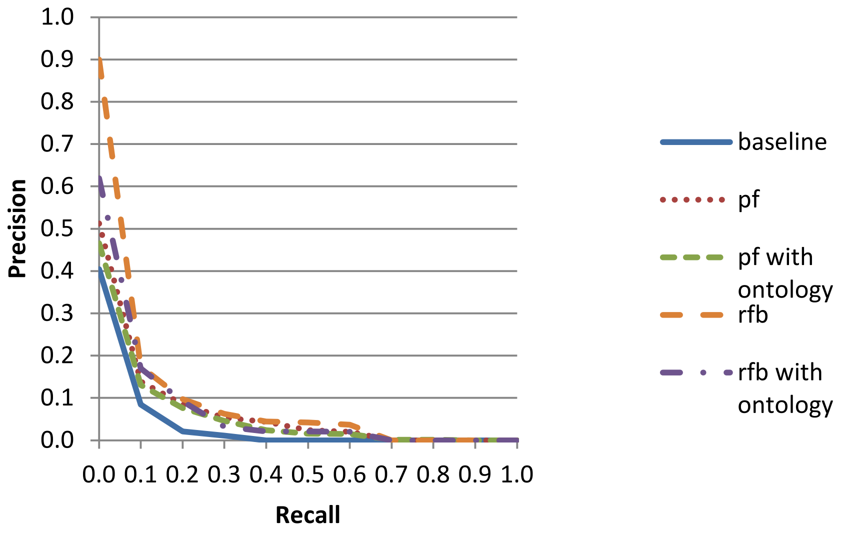


Figure A11.3 Topic 201-250 PF Results – Precision-Recall Graphs

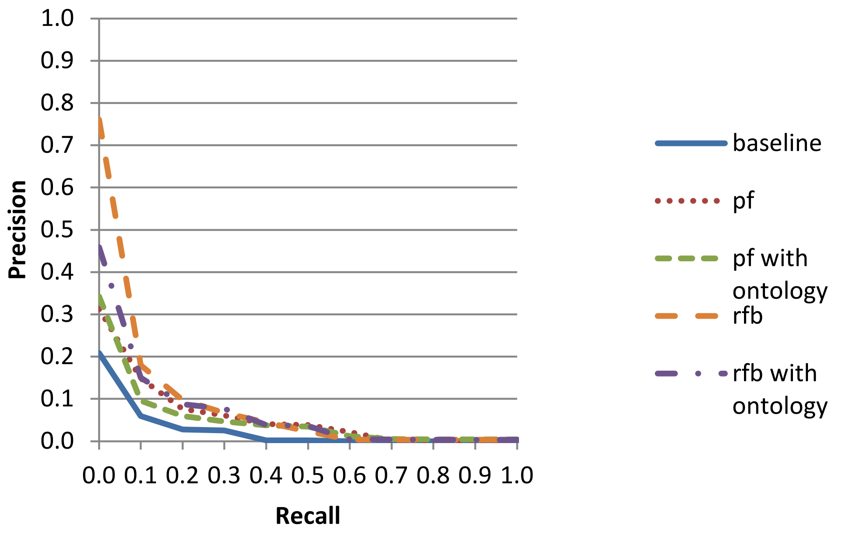


Figure A11.4 Topic 251-300 PF Results – Precision-Recall Graphs

A12 Precision-Recall Graphs for Experiment 5.11 (200 documents and 20 Terms)

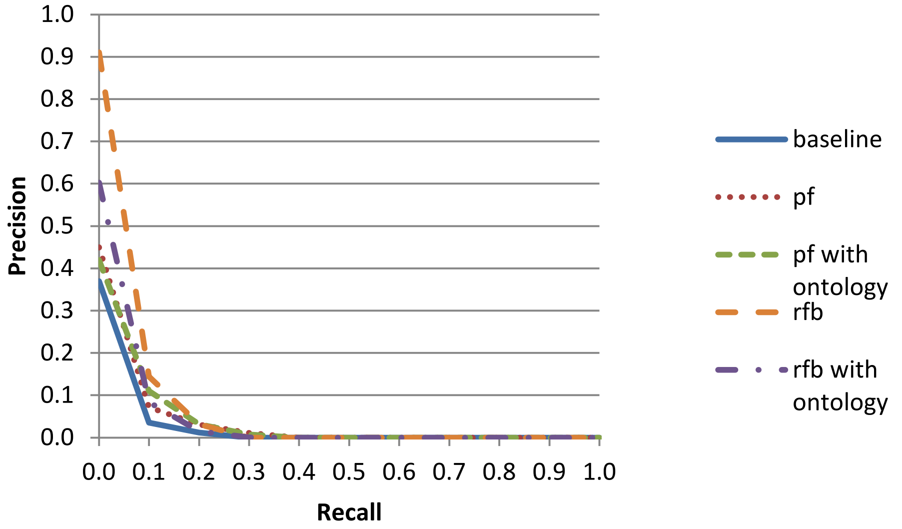


Figure A12.1 Topic 101-150 PF Results – Precision-Recall Graphs

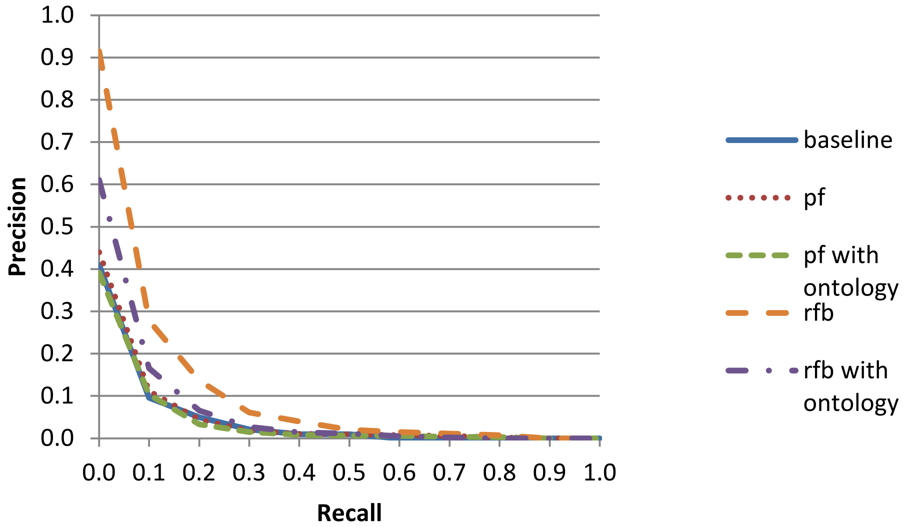


Figure A12.2 Topic 151-200 PF Results – Precision-Recall Graphs

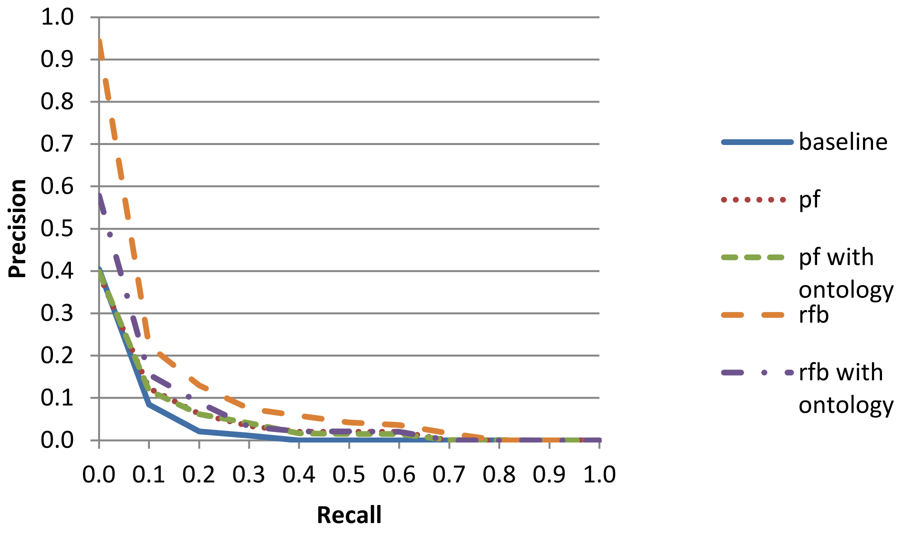


Figure A12.3 Topic 201-250 PF Results – Precision-Recall Graphs

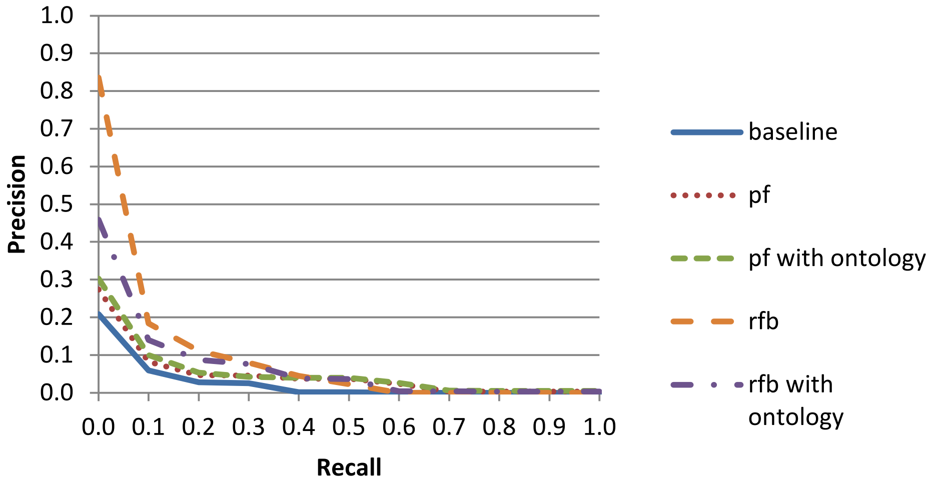


Figure A12.4 Topic 251-300 PF Results – Precision-Recall Graphs

APPENDIX B – Parent list for Standard run (Topic 51-100)

```
TOPIC NUMBER = 51
current word is conflict
     -->  unrest_conflicts_war
 TOPIC NUMBER = 52
current word is defend
     -->  trials
 TOPIC NUMBER = 53
current word is financ
     -->  economy_business_finance
 TOPIC NUMBER = 54
current word is space
     -->  science_technology
 TOPIC NUMBER = 56
current word is prime
     -->  government
 TOPIC NUMBER = 56
current word is wast
     -->  environmental_issue
 TOPIC NUMBER = 56
current word is chemical
     -->  economy_business_finance
 TOPIC NUMBER = 56
current word is prime
     -->  government
 TOPIC NUMBER = 58
current word is rail
     -->  transport_accident
 TOPIC NUMBER = 58
current word is strike
     -->  labour
 TOPIC NUMBER = 58
current word is refere
     -->  politics
 TOPIC NUMBER = 58
current word is court
     -->  lawyer
 TOPIC NUMBER = 58
current word is camp
     -->  election
 TOPIC NUMBER = 58
current word is rail
     -->  transport_accident
 TOPIC NUMBER = 58
current word is strike
     -->  labour
 TOPIC NUMBER = 59
current word is weather
 TOPIC NUMBER = 59
current word is weather
 TOPIC NUMBER = 62
current word is coup
     -->  unrest_conflicts_war
 TOPIC NUMBER = 62
current word is rebel
     -->  civil_unrest
```

```
 TOPIC NUMBER = 62
current word is troop
     -->  defence
 TOPIC NUMBER = 62
current word is coup
     -->  unrest_conflicts_war
 TOPIC NUMBER = 67
current word is civil
     -->  laws
 TOPIC NUMBER = 67
current word is prisoner
     -->  war
 TOPIC NUMBER = 67
current word is offender
     -->  crime
 TOPIC NUMBER = 67
current word is political
     -->  politics
 TOPIC NUMBER = 67
current word is suicid
     -->  euthanasia
 TOPIC NUMBER = 67
current word is civil
     -->  laws
 TOPIC NUMBER = 68
current word is health
 TOPIC NUMBER = 68
current word is hazard
     -->  environmental_issue
 TOPIC NUMBER = 68
current word is consumer
     -->  economy_business_finance
 TOPIC NUMBER = 68
current word is health
 TOPIC NUMBER = 68
current word is hazard
     -->  environmental_issue
 TOPIC NUMBER = 69
current word is summit
     -->  foreign_policy_diplomacy
 TOPIC NUMBER = 69
current word is arm
     -->  defence
 TOPIC NUMBER = 70
current word is sex
     -->  crime
 TOPIC NUMBER = 71
current word is troop
     -->  defence
 TOPIC NUMBER = 71
current word is vete
     -->  defence
 TOPIC NUMBER = 71
current word is protest
     -->  civil_unrest
 TOPIC NUMBER = 72
current word is economi
     -->  foreign_aid
```

```
 TOPIC NUMBER = 73
current word is national
     -->  defence
 TOPIC NUMBER = 73
current word is fire
     -->  disaster_accident
 TOPIC NUMBER = 73
current word is national
     -->  defence
 TOPIC NUMBER = 74
current word is conflict
     -->  unrest_conflicts_war
 TOPIC NUMBER = 74
current word is crisis
     -->  unrest_conflicts_war
 TOPIC NUMBER = 74
current word is scienc
 TOPIC NUMBER = 74
current word is investigation
     -->  crime_law_justice
 TOPIC NUMBER = 74
current word is oil
     -->  economy_business_finance
 TOPIC NUMBER = 74
current word is conflict
     -->  unrest_conflicts_war
 TOPIC NUMBER = 76
current word is constitution
     -->  politics
 TOPIC NUMBER = 76
current word is constitution
     -->  politics
 TOPIC NUMBER = 77
current word is animal
     -->  disease
 TOPIC NUMBER = 78
current word is wast
     -->  environmental_issue
 TOPIC NUMBER = 79
current word is political
     -->  politics
 TOPIC NUMBER = 79
current word is parti
     -->  applied_science
 TOPIC NUMBER = 79
current word is political
     -->  politics
 TOPIC NUMBER = 79
current word is parti
     -->  applied_science
 TOPIC NUMBER = 80
current word is candidat
     -->  election
 TOPIC NUMBER = 80
current word is candidat
     -->  election
 TOPIC NUMBER = 81
current word is renew
```

```
     -->  environmental_issue
 TOPIC NUMBER = 82
current word is biotech
     -->  science_technology
 TOPIC NUMBER = 82
current word is drug
     -->  crime
 TOPIC NUMBER = 83
current word is earth
     -->  natural_disaster
 TOPIC NUMBER = 83
current word is drought
     -->  natural_disaster
 TOPIC NUMBER = 84
current word is plant
     -->  disease
 TOPIC NUMBER = 84
current word is nuclear
     -->  disaster_accident
 TOPIC NUMBER = 84
current word is plant
     -->  disease
 TOPIC NUMBER = 85
current word is corruption
     -->  corporate_crime
 TOPIC NUMBER = 85
current word is anti
     -->  corporate_crime
 TOPIC NUMBER = 85
current word is embezzlement
     -->  corporate_crime
 TOPIC NUMBER = 85
current word is prosecutor
     -->  trials
 TOPIC NUMBER = 85
current word is corrupt
     -->  corporate_crime
 TOPIC NUMBER = 85
current word is bribe
     -->  corporate_crime
 TOPIC NUMBER = 85
current word is corruption
     -->  corporate_crime
 TOPIC NUMBER = 86
current word is regulator
     -->  politics
 TOPIC NUMBER = 87
current word is criminal
     -->  crime
 TOPIC NUMBER = 87
current word is coup
     -->  unrest_conflicts_war
 TOPIC NUMBER = 87
current word is criminal
     -->  crime
 TOPIC NUMBER = 88
current word is oil
     -->  economy_business_finance
```

```
 TOPIC NUMBER = 88
current word is crisis
     -->  unrest_conflicts_war
 TOPIC NUMBER = 88
current word is oil
     -->  economy_business_finance
 TOPIC NUMBER = 89
current word is state
     -->  politics
 TOPIC NUMBER = 89
current word is oil
     -->  economy_business_finance
 TOPIC NUMBER = 89
current word is minister
     -->  government
 TOPIC NUMBER = 89
current word is meet
     -->  foreign_policy_diplomacy
 TOPIC NUMBER = 89
current word is state
     -->  politics
 TOPIC NUMBER = 90
current word is data
     -->  interior_policy
 TOPIC NUMBER = 90
current word is oil
     -->  economy_business_finance
 TOPIC NUMBER = 90
current word is natural
     -->  disaster_accident
 TOPIC NUMBER = 90
current word is candidat
     -->  election
 TOPIC NUMBER = 90
current word is data
     -->  interior_policy
 TOPIC NUMBER = 90
current word is oil
     -->  economy_business_finance
 TOPIC NUMBER = 90
current word is natural
     -->  disaster_accident
 TOPIC NUMBER = 91
current word is weapon
     -->  unrest_conflicts_war
 TOPIC NUMBER = 91
current word is witness
     -->  trials
 TOPIC NUMBER = 91
current word is chemical
     -->  economy_business_finance
 TOPIC NUMBER = 91
current word is weapon
     -->  unrest_conflicts_war
 TOPIC NUMBER = 92
current word is illegal
     -->  demographics
 TOPIC NUMBER = 92
```

```
current word is arm
     -->  defence
 TOPIC NUMBER = 92
current word is employment
     -->  labour
 TOPIC NUMBER = 93
current word is national
     -->  defence
 TOPIC NUMBER = 93
current word is national
     -->  defence
 TOPIC NUMBER = 94
current word is crime
 TOPIC NUMBER = 94
current word is censor
     -->  politics
 TOPIC NUMBER = 94
current word is un
     -->  labour
 TOPIC NUMBER = 94
current word is terrorist
     -->  terrorism
 TOPIC NUMBER = 94
current word is anti
     -->  corporate_crime
 TOPIC NUMBER = 94
current word is drug
     -->  crime
 TOPIC NUMBER = 94
current word is crime
 TOPIC NUMBER = 95
current word is crime
 TOPIC NUMBER = 95
current word is censor
     -->  politics
 TOPIC NUMBER = 95
current word is un
     -->  labour
 TOPIC NUMBER = 95
current word is terrorist
     -->  terrorism
 TOPIC NUMBER = 95
current word is drug
     -->  crime
 TOPIC NUMBER = 95
current word is crime
 TOPIC NUMBER = 96
current word is medical
     -->  health_treatment
 TOPIC NUMBER = 96
current word is wast
     -->  environmental_issue
 TOPIC NUMBER = 96
current word is medical
     -->  health_treatment
```

APPENDIX C – Parent list for Top 3 terms run (Topic 51-100)

```
TOPIC NUMBER = 51
current word is conflict
     -->  unrest_conflicts_war
 TOPIC NUMBER = 52
current word is defend
     -->  trials
 TOPIC NUMBER = 53
current word is financ
     -->  economy_business_finance
 TOPIC NUMBER = 54
current word is space
     -->  science_technology
 TOPIC NUMBER = 56
current word is prime
     -->  government
 TOPIC NUMBER = 56
current word is wast
     -->  environmental_issue
 TOPIC NUMBER = 56
current word is chemical
     -->  economy_business_finance
 TOPIC NUMBER = 56
current word is prime
     -->  government
 TOPIC NUMBER = 58
current word is rail
     -->  transport_accident
 TOPIC NUMBER = 58
current word is strike
     -->  labour
 TOPIC NUMBER = 58
current word is refere
     -->  politics
 TOPIC NUMBER = 58
current word is court
     -->  lawyer
 TOPIC NUMBER = 58
current word is camp
     -->  election
 TOPIC NUMBER = 58
current word is rail
     -->  transport_accident
 TOPIC NUMBER = 58
current word is strike
     -->  labour
 TOPIC NUMBER = 59
current word is weather
 TOPIC NUMBER = 59
current word is weather
 TOPIC NUMBER = 62
current word is coup
     -->  unrest_conflicts_war
 TOPIC NUMBER = 62
current word is rebel
     -->  civil_unrest
```

```
 TOPIC NUMBER = 62
current word is troop
     -->  defence
 TOPIC NUMBER = 62
current word is coup
     -->  unrest_conflicts_war
 TOPIC NUMBER = 67
current word is civil
     -->  laws
 TOPIC NUMBER = 67
current word is prisoner
     -->  war
 TOPIC NUMBER = 67
current word is offender
     -->  crime
 TOPIC NUMBER = 67
current word is political
     -->  politics
 TOPIC NUMBER = 67
current word is suicid
     -->  euthanasia
 TOPIC NUMBER = 67
current word is civil
     -->  laws
 TOPIC NUMBER = 68
current word is health
 TOPIC NUMBER = 68
current word is hazard
     -->  environmental_issue
 TOPIC NUMBER = 68
current word is consumer
     -->  economy_business_finance
 TOPIC NUMBER = 68
current word is health
 TOPIC NUMBER = 68
current word is hazard
     -->  environmental_issue
 TOPIC NUMBER = 69
current word is summit
     -->  foreign_policy_diplomacy
 TOPIC NUMBER = 69
current word is arm
     -->  defence
 TOPIC NUMBER = 70
current word is sex
     -->  crime
 TOPIC NUMBER = 71
current word is troop
     -->  defence
 TOPIC NUMBER = 71
current word is vete
     -->  defence
 TOPIC NUMBER = 71
current word is protest
     -->  civil_unrest
 TOPIC NUMBER = 72
current word is economi
     -->  foreign_aid
```

```
 TOPIC NUMBER = 73
current word is national
     -->  defence
 TOPIC NUMBER = 73
current word is fire
     -->  disaster_accident
 TOPIC NUMBER = 73
current word is national
     -->  defence
 TOPIC NUMBER = 74
current word is conflict
     -->  unrest_conflicts_war
 TOPIC NUMBER = 74
current word is crisis
     -->  unrest_conflicts_war
 TOPIC NUMBER = 74
current word is scienc
 TOPIC NUMBER = 74
current word is investigation
     -->  crime_law_justice
 TOPIC NUMBER = 74
current word is oil
     -->  economy_business_finance
 TOPIC NUMBER = 74
current word is conflict
     -->  unrest_conflicts_war
 TOPIC NUMBER = 76
current word is constitution
     -->  politics
 TOPIC NUMBER = 76
current word is constitution
     -->  politics
 TOPIC NUMBER = 77
current word is animal
     -->  disease
 TOPIC NUMBER = 78
current word is wast
     -->  environmental_issue
 TOPIC NUMBER = 79
current word is political
     -->  politics
 TOPIC NUMBER = 79
current word is parti
     -->  applied_science
 TOPIC NUMBER = 79
current word is political
     -->  politics
 TOPIC NUMBER = 79
current word is parti
     -->  applied_science
 TOPIC NUMBER = 80
current word is candidat
     -->  election
 TOPIC NUMBER = 80
current word is candidat
     -->  election
 TOPIC NUMBER = 81
current word is renew
```

```
     -->  environmental_issue
 TOPIC NUMBER = 82
current word is biotech
     -->  science_technology
 TOPIC NUMBER = 82
current word is drug
     -->  crime
 TOPIC NUMBER = 83
current word is earth
     -->  natural_disaster
 TOPIC NUMBER = 83
current word is drought
     -->  natural_disaster
 TOPIC NUMBER = 84
current word is plant
     -->  disease
 TOPIC NUMBER = 84
current word is nuclear
     -->  disaster_accident
 TOPIC NUMBER = 84
current word is plant
     -->  disease
 TOPIC NUMBER = 85
current word is corruption
     -->  corporate_crime
 TOPIC NUMBER = 85
current word is anti
     -->  corporate_crime
 TOPIC NUMBER = 85
current word is embezzlement
     -->  corporate_crime
 TOPIC NUMBER = 85
current word is prosecutor
     -->  trials
 TOPIC NUMBER = 85
current word is corrupt
     -->  corporate_crime
 TOPIC NUMBER = 85
current word is bribe
     -->  corporate_crime
 TOPIC NUMBER = 85
current word is corruption
     -->  corporate_crime
 TOPIC NUMBER = 86
current word is regulator
     -->  politics
 TOPIC NUMBER = 87
current word is criminal
     -->  crime
 TOPIC NUMBER = 87
current word is coup
     -->  unrest_conflicts_war
 TOPIC NUMBER = 87
current word is criminal
     -->  crime
 TOPIC NUMBER = 88
current word is oil
     -->  economy_business_finance
```

```
 TOPIC NUMBER = 88
current word is crisis
     -->  unrest_conflicts_war
 TOPIC NUMBER = 88
current word is oil
     -->  economy_business_finance
 TOPIC NUMBER = 89
current word is state
     -->  politics
 TOPIC NUMBER = 89
current word is oil
     -->  economy_business_finance
 TOPIC NUMBER = 89
current word is minister
     -->  government
 TOPIC NUMBER = 89
current word is meet
     -->  foreign_policy_diplomacy
 TOPIC NUMBER = 89
current word is state
     -->  politics
 TOPIC NUMBER = 90
current word is data
     -->  interior_policy
 TOPIC NUMBER = 90
current word is oil
     -->  economy_business_finance
 TOPIC NUMBER = 90
current word is natural
     -->  disaster_accident
 TOPIC NUMBER = 90
current word is candidat
     -->  election
 TOPIC NUMBER = 90
current word is data
     -->  interior_policy
 TOPIC NUMBER = 90
current word is oil
     -->  economy_business_finance
 TOPIC NUMBER = 90
current word is natural
     -->  disaster_accident
 TOPIC NUMBER = 91
current word is weapon
     -->  unrest_conflicts_war
 TOPIC NUMBER = 91
current word is witness
     -->  trials
 TOPIC NUMBER = 91
current word is chemical
     -->  economy_business_finance
 TOPIC NUMBER = 91
current word is weapon
     -->  unrest_conflicts_war
 TOPIC NUMBER = 92
current word is illegal
     -->  demographics
 TOPIC NUMBER = 92
```

```
current word is arm
     -->  defence
 TOPIC NUMBER = 92
current word is employment
     -->  labour
 TOPIC NUMBER = 93
current word is national
     -->  defence
 TOPIC NUMBER = 93
current word is national
     -->  defence
 TOPIC NUMBER = 94
current word is crime
 TOPIC NUMBER = 94
current word is censor
     -->  politics
 TOPIC NUMBER = 94
current word is un
     -->  labour
 TOPIC NUMBER = 94
current word is terrorist
     -->  terrorism
 TOPIC NUMBER = 94
current word is anti
     -->  corporate_crime
 TOPIC NUMBER = 94
current word is drug
     -->  crime
 TOPIC NUMBER = 94
current word is crime
 TOPIC NUMBER = 95
current word is crime
 TOPIC NUMBER = 95
current word is censor
     -->  politics
 TOPIC NUMBER = 95
current word is un
     -->  labour
 TOPIC NUMBER = 95
current word is terrorist
     -->  terrorism
 TOPIC NUMBER = 95
current word is drug
     -->  crime
 TOPIC NUMBER = 95
current word is crime
 TOPIC NUMBER = 96
current word is medical
     -->  health_treatment
 TOPIC NUMBER = 96
current word is wast
     -->  environmental_issue
 TOPIC NUMBER = 96
current word is medical
     -->  health_treatment
```

APPENDIX D – Child list for Standard run (Topic 51-100)

```
Topic no = 86
ORIG WORD IS regulator
Topic no = 87
ORIG WORD IS criminal
Topic no = 87
ORIG WORD IS coup
Topic no = 87
ORIG WORD IS criminal
Topic no = 88
ORIG WORD IS oil
Topic no = 88
ORIG WORD IS crisis
Topic no = 88
ORIG WORD IS oil
Topic no = 89
ORIG WORD IS state
     --> public_finance
Topic no = 89
ORIG WORD IS oil
Topic no = 89
ORIG WORD IS minister
Topic no = 89
ORIG WORD IS meet
Topic no = 89
ORIG WORD IS state
     --> public_finance
Topic no = 90
ORIG WORD IS data
Topic no = 90
ORIG WORD IS oil
Topic no = 90
ORIG WORD IS natural
     --> earthquake
     --> tsunami
     --> flood
     --> drought
     --> avalanche
     --> landslide
     --> land_resources
     --> parks
     --> forests
     --> wetlands
     --> mountains
     --> rivers
     --> oceans
     --> wildlife
     --> energy_resources
     --> geology
     --> paleontology
     --> geography
     --> physiology
     --> botany
     --> astronomy
     --> biology
Topic no = 90
```

```
ORIG WORD IS candidat
Topic no = 90
ORIG WORD IS data
Topic no = 90
ORIG WORD IS oil
Topic no = 90
ORIG WORD IS natural
     --> earthquake
     --> tsunami
     --> flood
     --> drought
     --> avalanche
     --> landslide
     --> land_resources
     --> parks
     --> forests
     --> wetlands
     --> mountains
     --> rivers
     --> oceans
     --> wildlife
     --> energy_resources
     --> geology
     --> paleontology
     --> geography
     --> physiology
     --> botany
     --> astronomy
     --> biology
Topic no = 91
ORIG WORD IS weapon
Topic no = 91
ORIG WORD IS witness
Topic no = 91
ORIG WORD IS chemical
Topic no = 91
ORIG WORD IS weapon
Topic no = 92
ORIG WORD IS illegal
Topic no = 92
ORIG WORD IS arm
Topic no = 92
ORIG WORD IS employment
     --> labor_market
     --> job_layoffs
     --> child_labour
     --> occupations
Topic no = 93
ORIG WORD IS national
Topic no = 93
ORIG WORD IS national
Topic no = 94
ORIG WORD IS crime
     --> crime
     --> murder
     --> computer_crime
     --> theft
     --> drug_trafficking
```

```
      --> sexual_assault
      --> assault
      --> torture
      --> kidnapping
      --> arson
      --> gang_activity
      --> criminal
      --> murderer
      --> offender
      --> accused
      --> crime_victim
      --> stolen
      --> judiciary
      --> lawyer
      --> police
      --> investigation
      --> punishment
      --> prison
      --> laws
      --> justice_rights
      --> trials
      --> organized_crime
      --> international_law
      --> corporate_crime
      --> war_crime
Topic no = 94
ORIG WORD IS censor
Topic no = 94
ORIG WORD IS un
      --> oposing_group
      --> truce
      --> armed_conflict
      --> civil_unrest
      --> coup_detat
      --> terrorism
      --> massacre
      --> riots
      --> demonstration
      --> turf_war
      --> war
      --> conflict
      --> crisis
      --> weaponry
      --> bombings
      --> invasion
      --> war_victim
Topic no = 94
ORIG WORD IS terrorist
Topic no = 94
ORIG WORD IS anti
Topic no = 94
ORIG WORD IS drug
Topic no = 94
ORIG WORD IS crime
      --> crime
      --> murder
      --> computer_crime
      --> theft
```

```
      --> drug_trafficking
      --> sexual_assault
      --> assault
      --> torture
      --> kidnapping
      --> arson
      --> gang_activity
      --> criminal
      --> murderer
      --> offender
      --> accused
      --> crime_victim
      --> stolen
      --> judiciary
      --> lawyer
      --> police
      --> investigation
      --> punishment
      --> prison
      --> laws
      --> justice_rights
      --> trials
      --> organized_crime
      --> international_law
      --> corporate_crime
      --> war_crime
Topic no = 95
ORIG WORD IS crime
      --> crime
      --> murder
      --> computer_crime
      --> theft
      --> drug_trafficking
      --> sexual_assault
      --> assault
      --> torture
      --> kidnapping
      --> arson
      --> gang_activity
      --> criminal
      --> murderer
      --> offender
      --> accused
      --> crime_victim
      --> stolen
      --> judiciary
      --> lawyer
      --> police
      --> investigation
      --> punishment
      --> prison
      --> laws
      --> justice_rights
      --> trials
      --> organized_crime
      --> international_law
      --> corporate_crime
      --> war_crime
```

```
Topic no = 95
ORIG WORD IS censor
Topic no = 95
ORIG WORD IS un
     --> oposing_group
     --> truce
     --> armed_conflict
     --> civil_unrest
     --> coup_detat
     --> terrorism
     --> massacre
     --> riots
     --> demonstration
     --> turf_war
     --> war
     --> conflict
     --> crisis
     --> weaponry
     --> bombings
     --> invasion
     --> war_victim
Topic no = 95
ORIG WORD IS terrorist
Topic no = 95
ORIG WORD IS drug
Topic no = 95
ORIG WORD IS crime
     --> crime
     --> murder
     --> computer_crime
     --> theft
     --> drug_trafficking
     --> sexual_assault
     --> assault
     --> torture
     --> kidnapping
     --> arson
     --> gang_activity
     --> criminal
     --> murderer
     --> offender
     --> accused
     --> crime_victim
     --> stolen
     --> judiciary
     --> lawyer
     --> police
     --> investigation
     --> punishment
     --> prison
     --> laws
     --> justice_rights
     --> trials
     --> organized_crime
     --> international_law
     --> corporate_crime
     --> war_crime
Topic no = 96
```

```
ORIG WORD IS medical
Topic no = 96
ORIG WORD IS wast
Topic no = 96
ORIG WORD IS medical
```

APPENDIX E – Child list for Top 3 terms run (Topic 51-100)

```
Topic no = 51
ORIG WORD IS conflict
     --> peacekeeping_force
Topic no = 52
ORIG WORD IS defend
Topic no = 53
ORIG WORD IS financ
Topic no = 54
ORIG WORD IS space
Topic no = 56
ORIG WORD IS prime
Topic no = 56
ORIG WORD IS wast
Topic no = 56
ORIG WORD IS chemical
Topic no = 56
ORIG WORD IS prime
Topic no = 58
ORIG WORD IS rail
Topic no = 58
ORIG WORD IS strike
Topic no = 58
ORIG WORD IS refere
Topic no = 58
ORIG WORD IS court
Topic no = 58
ORIG WORD IS camp
Topic no = 58
ORIG WORD IS rail
Topic no = 58
ORIG WORD IS strike
Topic no = 59
ORIG WORD IS weather
     --> forecast
     --> global_change
     --> report
     --> statistic
     --> warning
Topic no = 59
ORIG WORD IS weather
     --> forecast
     --> global_change
     --> report
     --> statistic
     --> warning
Topic no = 62
ORIG WORD IS coup
Topic no = 62
ORIG WORD IS rebel
Topic no = 62
ORIG WORD IS troop
Topic no = 62
ORIG WORD IS coup
Topic no = 67
ORIG WORD IS civil
     --> revolutions
     --> rebellions
```

```
      --> political_dissent
      --> religious_conflict
      --> social_conflict
      --> protest
Topic no = 67
ORIG WORD IS prisoner
Topic no = 67
ORIG WORD IS offender
Topic no = 67
ORIG WORD IS political
Topic no = 67
ORIG WORD IS suicid
Topic no = 67
ORIG WORD IS civil
      --> revolutions
      --> rebellions
      --> political_dissent
      --> religious_conflict
      --> social_conflict
      --> protest
Topic no = 68
ORIG WORD IS health
      --> disease
      --> epidemic_plague
      --> health_treatment
      --> prescription_drugs
      --> medical_procedure
      --> therapy
      --> health_org
      --> medical_research
      --> medical_staff
      --> medicine
      --> preventative_medicine
      --> injury
      --> hospital
      --> clinic
      --> illness
Topic no = 68
ORIG WORD IS hazard
Topic no = 68
ORIG WORD IS consumer
Topic no = 68
ORIG WORD IS health
      --> disease
      --> epidemic_plague
      --> health_treatment
      --> prescription_drugs
      --> medical_procedure
      --> therapy
      --> health_org
      --> medical_research
      --> medical_staff
      --> medicine
      --> preventative_medicine
      --> injury
      --> hospital
      --> clinic
      --> illness
```

```
Topic no = 68
ORIG WORD IS hazard
Topic no = 69
ORIG WORD IS summit
Topic no = 69
ORIG WORD IS arm
Topic no = 70
ORIG WORD IS sex
Topic no = 71
ORIG WORD IS troop
Topic no = 71
ORIG WORD IS vete
Topic no = 71
ORIG WORD IS protest
Topic no = 72
ORIG WORD IS economi
Topic no = 73
ORIG WORD IS national
Topic no = 73
ORIG WORD IS fire
Topic no = 73
ORIG WORD IS national
Topic no = 74
ORIG WORD IS conflict
     --> peacekeeping_force
Topic no = 74
ORIG WORD IS crisis
Topic no = 74
ORIG WORD IS scienc
     --> applied_science
     --> engineering
     --> natural_science
     --> research
     --> scientific_exploration
     --> space_programme
     --> standards
     --> mathematics
     --> biotechnology
     --> agricultural_research_technology
     --> nanotechnology
     --> IT_computer_science
     --> scientific_institutions
Topic no = 74
ORIG WORD IS investigation
Topic no = 74
ORIG WORD IS oil
Topic no = 74
ORIG WORD IS conflict
     --> peacekeeping_force
Topic no = 76
ORIG WORD IS constitution
Topic no = 76
ORIG WORD IS constitution
Topic no = 77
ORIG WORD IS animal
Topic no = 78
ORIG WORD IS wast
Topic no = 79
```

```
ORIG WORD IS political
Topic no = 79
ORIG WORD IS parti
Topic no = 79
ORIG WORD IS political
Topic no = 79
ORIG WORD IS parti
Topic no = 80
ORIG WORD IS candidat
Topic no = 80
ORIG WORD IS candidat
Topic no = 81
ORIG WORD IS renew
Topic no = 82
ORIG WORD IS biotech
Topic no = 82
ORIG WORD IS drug
Topic no = 83
ORIG WORD IS earth
Topic no = 83
ORIG WORD IS drought
Topic no = 84
ORIG WORD IS plant
Topic no = 84
ORIG WORD IS nuclear
Topic no = 84
ORIG WORD IS plant
Topic no = 85
ORIG WORD IS corruption
Topic no = 85
ORIG WORD IS anti
Topic no = 85
ORIG WORD IS embezzlement
Topic no = 85
ORIG WORD IS prosecutor
Topic no = 85
ORIG WORD IS corrupt
Topic no = 85
ORIG WORD IS bribe
Topic no = 85
ORIG WORD IS corruption
Topic no = 86
ORIG WORD IS regulator
Topic no = 87
ORIG WORD IS criminal
Topic no = 87
ORIG WORD IS coup
Topic no = 87
ORIG WORD IS criminal
Topic no = 88
ORIG WORD IS oil
Topic no = 88
ORIG WORD IS crisis
Topic no = 88
ORIG WORD IS oil
Topic no = 89
ORIG WORD IS state
     --> public_finance
```

```
Topic no = 89
ORIG WORD IS oil
Topic no = 89
ORIG WORD IS minister
Topic no = 89
ORIG WORD IS meet
Topic no = 89
ORIG WORD IS state
     --> public_finance
Topic no = 90
ORIG WORD IS data
Topic no = 90
ORIG WORD IS oil
Topic no = 90
ORIG WORD IS natural
     --> earthquake
     --> tsunami
     --> flood
     --> drought
     --> avalanche
     --> landslide
     --> land_resources
     --> parks
     --> forests
     --> wetlands
     --> mountains
     --> rivers
     --> oceans
     --> wildlife
     --> energy_resources
     --> geology
     --> paleontology
     --> geography
     --> physiology
     --> botany
     --> astronomy
     --> biology
Topic no = 90
ORIG WORD IS candidat
Topic no = 90
ORIG WORD IS data
Topic no = 90
ORIG WORD IS oil
Topic no = 90
ORIG WORD IS natural
     --> earthquake
     --> tsunami
     --> flood
     --> drought
     --> avalanche
     --> landslide
     --> land_resources
     --> parks
     --> forests
     --> wetlands
     --> mountains
     --> rivers
     --> oceans
```

```
      --> wildlife
      --> energy_resources
      --> geology
      --> paleontology
      --> geography
      --> physiology
      --> botany
      --> astronomy
      --> biology
Topic no = 91
ORIG WORD IS weapon
Topic no = 91
ORIG WORD IS witness
Topic no = 91
ORIG WORD IS chemical
Topic no = 91
ORIG WORD IS weapon
Topic no = 92
ORIG WORD IS illegal
Topic no = 92
ORIG WORD IS arm
Topic no = 92
ORIG WORD IS employment
      --> labor_market
      --> job_layoffs
      --> child_labour
      --> occupations
Topic no = 93
ORIG WORD IS national
Topic no = 93
ORIG WORD IS national
Topic no = 94
ORIG WORD IS crime
      --> crime
      --> murder
      --> computer_crime
      --> theft
      --> drug_trafficking
      --> sexual_assault
      --> assault
      --> torture
      --> kidnapping
      --> arson
      --> gang_activity
      --> criminal
      --> murderer
      --> offender
      --> accused
      --> crime_victim
      --> stolen
      --> judiciary
      --> lawyer
      --> police
      --> investigation
      --> punishment
      --> prison
      --> laws
      --> justice_rights
```

```
     --> trials
     --> organized_crime
     --> international_law
     --> corporate_crime
     --> war_crime
Topic no = 94
ORIG WORD IS censor
Topic no = 94
ORIG WORD IS un
     --> oposing_group
     --> truce
     --> armed_conflict
     --> civil_unrest
     --> coup_detat
     --> terrorism
     --> massacre
     --> riots
     --> demonstration
     --> turf_war
     --> war
     --> conflict
     --> crisis
     --> weaponry
     --> bombings
     --> invasion
     --> war_victim
Topic no = 94
ORIG WORD IS terrorist
Topic no = 94
ORIG WORD IS anti
Topic no = 94
ORIG WORD IS drug
Topic no = 94
ORIG WORD IS crime
     --> crime
     --> murder
     --> computer_crime
     --> theft
     --> drug_trafficking
     --> sexual_assault
     --> assault
     --> torture
     --> kidnapping
     --> arson
     --> gang_activity
     --> criminal
     --> murderer
     --> offender
     --> accused
     --> crime_victim
     --> stolen
     --> judiciary
     --> lawyer
     --> police
     --> investigation
     --> punishment
     --> prison
     --> laws
```

```
      --> justice_rights
      --> trials
      --> organized_crime
      --> international_law
      --> corporate_crime
      --> war_crime
Topic no = 95
ORIG WORD IS crime
      --> crime
      --> murder
      --> computer_crime
      --> theft
      --> drug_trafficking
      --> sexual_assault
      --> assault
      --> torture
      --> kidnapping
      --> arson
      --> gang_activity
      --> criminal
      --> murderer
      --> offender
      --> accused
      --> crime_victim
      --> stolen
      --> judiciary
      --> lawyer
      --> police
      --> investigation
      --> punishment
      --> prison
      --> laws
      --> justice_rights
      --> trials
      --> organized_crime
      --> international_law
      --> corporate_crime
      --> war_crime
Topic no = 95
ORIG WORD IS censor
Topic no = 95
ORIG WORD IS un
      --> oposing_group
      --> truce
      --> armed_conflict
      --> civil_unrest
      --> coup_detat
      --> terrorism
      --> massacre
      --> riots
      --> demonstration
      --> turf_war
      --> war
      --> conflict
      --> crisis
      --> weaponry
      --> bombings
      --> invasion
```

```
     --> war_victim
Topic no = 95
ORIG WORD IS terrorist
Topic no = 95
ORIG WORD IS drug
Topic no = 95
ORIG WORD IS crime
     --> crime
     --> murder
     --> computer_crime
     --> theft
     --> drug_trafficking
     --> sexual_assault
     --> assault
     --> torture
     --> kidnapping
     --> arson
     --> gang_activity
     --> criminal
     --> murderer
     --> offender
     --> accused
     --> crime_victim
     --> stolen
     --> judiciary
     --> lawyer
     --> police
     --> investigation
     --> punishment
     --> prison
     --> laws
     --> justice_rights
     --> trials
     --> organized_crime
     --> international_law
     --> corporate_crime
     --> war_crime
Topic no = 96
ORIG WORD IS medical
Topic no = 96
ORIG WORD IS wast
Topic no = 96
ORIG WORD IS medical
```

APPENDIX F – TREC output Listings

F1 Standard run

Topic 51-100

	baseline	pf baseline	pfl1	rf baseline	rfl1
num_ret	49964	49495	49820	49856	49848
num_rel	16386	16386	16386	16386	16386
num_rel_ret	408	1036	1197	570	963
map	0.010	0.033	0.033	0.018	0.021
gm_map	0.000	0.002	0.003	0.001	0.002
Rprec	0.024	0.062	0.062	0.038	0.047
bpref	0.027	0.065	0.073	0.040	0.062
recirank	0.212	0.510	0.461	0.060	0.443
0	0.245	0.530	0.492	0.600	0.458
0.1	0.028	0.112	0.117	0.059	0.066
0.2	0.014	0.065	0.046	0.004	0.020
0.3	0.000	0.009	0.013	0.000	0.002
0.4	0.000	0.000	0.000	0.000	0.000
0.5	0.000	0.000	0.000	0.000	0.000
0.6	0.000	0.000	0.000	0.000	0.000
0.7	0.000	0.000	0.000	0.000	0.000
0.8	0.000	0.000	0.000	0.000	0.000
0.9	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000
5	0.132	0.316	0.320	0.304	0.248
10	0.124	0.276	0.322	0.224	0.204
15	0.116	0.256	0.275	0.191	0.175
20	0.103	0.238	0.251	0.174	0.155
30	0.084	0.207	0.213	0.156	0.133
100	0.047	0.116	0.115	0.078	0.086
200	0.031	0.074	0.079	0.049	0.060
500	0.015	0.037	0.041	0.022	0.032
1000	0.008	0.021	0.024	0.011	0.019

Topic 101-150

	baseline	pf baseline	pfl1	rf baseline	rfl1
num_ret	50000	49925	49996	49980	49978
num_rel	11645	11645	11645	11645	11645
num_rel_ret	486	1343	1446	865	1117
map	0.012	0.042	0.039	0.033	0.025

gm_map	0.002	0.012	0.015	0.011	0.010
Rprec	0.038	0.088	0.087	0.063	0.066
bpref	0.040	0.094	0.105	0.070	0.083
recirank	0.340	0.595	0.586	0.835	0.598
0	0.369	0.637	0.621	0.835	0.613
0.1	0.035	0.160	0.151	0.099	0.086
0.2	0.012	0.054	0.026	0.027	0.015
0.3	0.000	0.013	0.005	0.004	0.000
0.4	0.000	0.000	0.000	0.000	0.000
0.5	0.000	0.000	0.000	0.000	0.000
0.6	0.000	0.000	0.000	0.000	0.000
0.7	0.000	0.000	0.000	0.000	0.000
0.8	0.000	0.000	0.000	0.000	0.000
0.9	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000
5	0.204	0.388	0.400	0.488	0.340
10	0.164	0.352	0.336	0.354	0.258
15	0.147	0.313	0.305	0.299	0.232
20	0.133	0.287	0.275	0.264	0.208
30	0.113	0.244	0.240	0.218	0.177
100	0.063	0.150	0.136	0.104	0.098
200	0.039	0.102	0.094	0.068	0.068
500	0.018	0.049	0.049	0.033	0.038
1000	0.010	0.027	0.029	0.017	0.022

Topic 151-200

	baseline	pf baseline	pfl1	rf baseline	rfl1
num_ret	49995	49924	49998	49991	49981
num_rel	9805	9805	9805	9805	9805
num_rel_ret	491	1069	1423	827	1203
map	0.032	0.076	0.064	0.067	0.051
gm_map	0.003	0.014	0.023	0.013	0.015
Rprec	0.066	0.115	0.117	0.095	0.087
bpref	0.067	0.122	0.135	0.108	0.116
recirank	0.365	0.639	0.619	0.821	0.583
0	0.410	0.667	0.666	0.826	0.614
0.1	0.095	0.281	0.231	0.199	0.166
0.2	0.049	0.117	0.082	0.095	0.066
0.3	0.021	0.059	0.032	0.050	0.027
0.4	0.009	0.031	0.017	0.029	0.014
0.5	0.009	0.016	0.010	0.017	0.011
0.6	0.000	0.013	0.010	0.012	0.005
0.7	0.000	0.011	0.006	0.009	0.002

0.8	0.000	0.000	0.000	0.008	0.000
0.9	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000
5	0.240	0.468	0.468	0.504	0.396
10	0.220	0.410	0.420	0.376	0.342
15	0.189	0.369	0.373	0.309	0.292
20	0.177	0.339	0.332	0.275	0.266
30	0.157	0.294	0.280	0.228	0.221
100	0.084	0.153	0.152	0.113	0.118
200	0.047	0.092	0.097	0.069	0.077
500	0.020	0.042	0.049	0.032	0.042
1000	0.010	0.021	0.029	0.017	0.024

Topic 201-250

	baseline	pf baseline	pfl1	rf baseline	rfl1
num_ret	50000	49996	49999	49989	49987
num_rel	6503	6503	6503	6503	6503
num_rel_ret	533	733	982	490	813
map	0.024	0.076	0.068	0.064	0.055
gm_map	0.002	0.012	0.018	0.013	0.014
Rprec	0.055	0.109	0.116	0.091	0.092
bpref	0.065	0.117	0.133	0.098	0.111
recirank	0.376	0.667	0.662	0.771	0.620
0	0.404	0.679	0.675	0.776	0.640
0.1	0.084	0.211	0.212	0.168	0.160
0.2	0.021	0.132	0.117	0.083	0.092
0.3	0.011	0.060	0.050	0.055	0.032
0.4	0.000	0.043	0.024	0.045	0.021
0.5	0.000	0.041	0.022	0.039	0.021
0.6	0.000	0.036	0.020	0.020	0.020
0.7	0.000	0.000	0.000	0.000	0.000
0.8	0.000	0.000	0.000	0.000	0.000
0.9	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000
5	0.208	0.424	0.388	0.412	0.324
10	0.166	0.314	0.302	0.294	0.258
15	0.143	0.259	0.252	0.227	0.208
20	0.123	0.219	0.228	0.197	0.176
30	0.105	0.180	0.187	0.161	0.143
100	0.056	0.090	0.098	0.068	0.076
200	0.037	0.055	0.063	0.040	0.051
500	0.020	0.028	0.033	0.019	0.028

1000	0.011	0.015	0.020	0.010	0.016

Topic 251-300

	baseline	pf baseline	pfl1	rf baseline	rfl1
num_ret	49290	49997	50000	49984	49274
num_rel	5524	5524	5524	5524	5524
num_rel_ret	201	390	481	278	394
map	0.016	0.043	0.056	0.057	0.052
gm_map	0.001	0.002	0.003	0.002	0.003
Rprec	0.034	0.062	0.083	0.076	0.074
bpref	0.040	0.069	0.088	0.080	0.084
recirank	0.191	0.362	0.355	0.581	0.450
0	0.209	0.384	0.397	0.585	0.459
0.1	0.060	0.172	0.128	0.167	0.151
0.2	0.028	0.067	0.095	0.110	0.087
0.3	0.025	0.046	0.063	0.093	0.077
0.4	0.002	0.014	0.049	0.055	0.038
0.5	0.002	0.014	0.044	0.036	0.036
0.6	0.000	0.006	0.036	0.000	0.004
0.7	0.000	0.001	0.016	0.000	0.004
0.8	0.000	0.000	0.004	0.000	0.003
0.9	0.000	0.000	0.003	0.000	0.003
1	0.000	0.000	0.003	0.000	0.003
5	0.068	0.224	0.240	0.268	0.216
10	0.058	0.174	0.186	0.168	0.148
15	0.056	0.152	0.139	0.132	0.119
20	0.055	0.131	0.125	0.110	0.102
30	0.051	0.104	0.098	0.082	0.083
100	0.030	0.050	0.050	0.038	0.042
200	0.017	0.033	0.032	0.023	0.026
500	0.008	0.015	0.017	0.011	0.013
1000	0.004	0.008	0.010	0.006	0.008

F2 Top 3 run

Topic 51-100

	baseline	pf baseline	pfl1	rf baseline	rfl1
num_ret	49964	49755	49820	49856	49848
num_rel	16386	16386	16386	16386	16386
num_rel_ret	408	644	1197	469	963
map	0.010	0.024	0.033	0.015	0.021
gm_map	0.000	0.001	0.003	0.001	0.002
Rprec	0.024	0.043	0.062	0.029	0.047
bpref	0.027	0.044	0.073	0.033	0.062
recirank	0.212	0.462	0.461	0.590	0.443
0	0.245	0.492	0.492	0.597	0.458
0.1	0.028	0.095	0.117	0.039	0.066
0.2	0.014	0.030	0.046	0.002	0.020
0.3	0.000	0.000	0.013	0.000	0.002
0.4	0.000	0.000	0.000	0.000	0.000
0.5	0.000	0.000	0.000	0.000	0.000
0.6	0.000	0.000	0.000	0.000	0.000
0.7	0.000	0.000	0.000	0.000	0.000
0.8	0.000	0.000	0.000	0.000	0.000
0.9	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000
5	0.132	0.264	0.320	0.308	0.248
10	0.124	0.228	0.322	0.222	0.204
15	0.116	0.204	0.275	0.172	0.175
20	0.103	0.187	0.251	0.144	0.155
30	0.084	0.159	0.213	0.115	0.133
100	0.047	0.098	0.115	0.062	0.086
200	0.031	0.056	0.079	0.038	0.060
500	0.015	0.025	0.041	0.017	0.032
1000	0.008	0.013	0.024	0.009	0.019

Topic 101-150

	baseline	pf baseline	pfl1	rf baseline	rfl1
num_ret	50000	49963	49996	49979	49978
num_rel	11645	11645	11645	11645	11645
num_rel_ret	486	907	1446	677	1117
map	0.012	0.032	0.039	0.026	0.025
gm_map	0.002	0.010	0.015	0.009	0.010

Rprec		0.038	0.071	0.087	0.054	0.066
bpref		0.040	0.069	0.105	0.056	0.083
recirank		0.340	0.595	0.586	0.825	0.598
	0	0.369	0.630	0.621	0.829	0.613
	0.1	0.035	0.121	0.151	0.086	0.086
	0.2	0.012	0.032	0.026	0.011	0.015
	0.3	0.000	0.000	0.005	0.000	0.000
	0.4	0.000	0.000	0.000	0.000	0.000
	0.5	0.000	0.000	0.000	0.000	0.000
	0.6	0.000	0.000	0.000	0.000	0.000
	0.7	0.000	0.000	0.000	0.000	0.000
	0.8	0.000	0.000	0.000	0.000	0.000
	0.9	0.000	0.000	0.000	0.000	0.000
	1	0.000	0.000	0.000	0.000	0.000
	5	0.204	0.400	0.400	0.412	0.340
	10	0.164	0.336	0.336	0.300	0.258
	15	0.147	0.280	0.305	0.244	0.232
	20	0.133	0.255	0.275	0.209	0.208
	30	0.113	0.216	0.240	0.175	0.177
	100	0.063	0.129	0.136	0.091	0.098
	200	0.039	0.079	0.094	0.057	0.068
	500	0.018	0.034	0.049	0.026	0.038
	1000	0.010	0.018	0.029	0.014	0.022

Topic 151-200

		baseline	pf baseline	pfl1	rf baseline	rfl1
num_ret		49995	49966	49998	49989	49981
num_rel		9805	9805	9805	9805	9805
num_rel_ret		491	839	1423	647	1203
map		0.032	0.057	0.064	0.053	0.051
gm_map		0.003	0.013	0.023	0.010	0.015
Rprec		0.066	0.100	0.117	0.076	0.097
bpref		0.067	0.099	0.135	0.084	0.116
recirank		0.365	0.662	0.619	0.784	0.583
	0	0.410	0.695	0.666	0.803	0.614
	0.1	0.095	0.185	0.231	0.133	0.166
	0.2	0.049	0.081	0.082	0.073	0.066
	0.3	0.021	0.036	0.032	0.026	0.027
	0.4	0.009	0.012	0.017	0.009	0.014
	0.5	0.009	0.011	0.010	0.009	0.011
	0.6	0.000	0.010	0.010	0.009	0.005
	0.7	0.000	0.008	0.006	0.008	0.002

0.8	0.000	0.000	0.000	0.006	0.000
0.9	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000
5	0.240	0.456	0.468	0.472	0.396
10	0.220	0.378	0.420	0.350	0.342
15	0.189	0.324	0.373	0.280	0.292
20	0.177	0.292	0.332	0.238	0.266
30	0.157	0.253	0.280	0.193	0.221
100	0.084	0.132	0.152	0.094	0.118
200	0.047	0.080	0.097	0.057	0.077
500	0.020	0.033	0.049	0.025	0.042
1000	0.010	0.017	0.029	0.013	0.024

Topic 201-250

			baseline	pf baseline	pfl1	rf baselinerfl1	
num_ret		50000	49986	49999	49983	49987
num_rel		6503	6503	6503	6503	6503
num_rel_ret		533	645	982	483	813
map		0.024	0.059	0.068	0.052	0.055
gm_map		0.002	0.009	0.018	0.011	0.014
Rprec		0.055	0.093	0.116	0.080	0.092
bpref		0.065	0.101	0.133	0.087	0.111
recirank		0.376	0.621	0.662	0.761	0.620
	0	0.404	0.639	0.675	0.772	0.640
	0.1	0.084	0.177	0.212	0.154	0.160
	0.2	0.021	0.085	0.117	0.084	0.092
	0.3	0.011	0.047	0.050	0.032	0.032
	0.4	0.000	0.022	0.024	0.023	0.021
	0.5	0.000	0.022	0.022	0.022	0.021
	0.6	0.000	0.020	0.020	0.020	0.020
	0.7	0.000	0.000	0.000	0.000	0.000
	0.8	0.000	0.000	0.000	0.000	0.000
	0.9	0.000	0.000	0.000	0.000	0.000
	1	0.000	0.000	0.000	0.000	0.000
	5	0.208	0.348	0.388	0.340	0.324
	10	0.166	0.274	0.302	0.240	0.258
	15	0.143	0.219	0.252	0.192	0.208
	20	0.123	0.195	0.228	0.163	0.176
	30	0.105	0.157	0.187	0.137	0.143
	100	0.056	0.077	0.098	0.062	0.076
	200	0.037	0.047	0.063	0.040	0.051

	500	0.020	0.023	0.033	0.019	0.028
	1000	0.011	0.013	0.020	0.010	0.016

Topic 251-300

		baseline	pf baseline	pfl1	rf baseline	rfl1
num_ret		49290	49994	50000	49316	49274
num_rel		5524	5524	5524	5524	5524
num_rel_ret		201	302	481	240	394
map		0.016	0.053	0.056	0.045	0.052
gm_map		0.001	0.002	0.003	0.002	0.003
Rprec		0.034	0.075	0.083	0.067	0.074
bpref		0.040	0.075	0.088	0.069	0.084
recirank		0.191	0.357	0.355	0.552	0.450
	0	0.209	0.384	0.397	0.556	0.459
	0.1	0.060	0.134	0.128	0.138	0.151
	0.2	0.028	0.087	0.095	0.069	0.087
	0.3	0.025	0.076	0.063	0.063	0.077
	0.4	0.002	0.055	0.049	0.038	0.038
	0.5	0.002	0.049	0.044	0.036	0.036
	0.6	0.000	0.024	0.036	0.000	0.004
	0.7	0.000	0.004	0.016	0.000	0.004
	0.8	0.000	0.003	0.004	0.000	0.003
	0.9	0.000	0.003	0.003	0.000	0.003
	1	0.000	0.003	0.003	0.000	0.003
	5	0.068	0.204	0.240	0.208	0.216
	10	0.058	0.174	0.186	0.120	0.148
	15	0.056	0.140	0.139	0.088	0.119
	20	0.055	0.120	0.125	0.075	0.102
	30	0.051	0.095	0.098	0.058	0.083
	100	0.030	0.042	0.050	0.032	0.042
	200	0.017	0.026	0.032	0.020	0.026
	500	0.008	0.012	0.017	0.009	0.013
	1000	0.004	0.006	0.010	0.005	0.008

F3 5 terms, 20 Documents

Topic 51-100

	baseline	pf baseline	pfl1	rf baseline	rfl1
num_ret	49964	49674	49819	49856	49848
num_rel	16386	16386	16386	16386	16386
num_rel_ret	408	736	1143	503	963
map	0.010	0.027	0.033	0.018	0.021
gm_map	0.000	0.001	0.002	0.001	0.002
Rprec	0.024	0.049	0.061	0.034	0.047
bpref	0.027	0.049	0.070	0.037	0.062
recirank	0.212	0.483	0.451	0.594	0.443
0	0.245	0.502	0.481	0.598	0.458
0.1	0.028	0.108	0.122	0.056	0.066
0.2	0.014	0.039	0.047	0.006	0.020
0.3	0.000	0.000	0.007	0.000	0.002
0.4	0.000	0.000	0.000	0.000	0.000
0.5	0.000	0.000	0.000	0.000	0.000
0.6	0.000	0.000	0.000	0.000	0.000
0.7	0.000	0.000	0.000	0.000	0.000
0.8	0.000	0.000	0.000	0.000	0.000
0.9	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000
5	0.132	0.288	0.324	0.300	0.248
10	0.124	0.254	0.320	0.234	0.204
15	0.116	0.233	0.288	0.187	0.175
20	0.103	0.213	0.262	0.160	0.155
30	0.084	0.182	0.217	0.131	0.133
100	0.047	0.102	0.122	0.070	0.086
200	0.031	0.063	0.080	0.042	0.060
500	0.015	0.028	0.040	0.019	0.032
1000	0.008	0.015	0.023	0.010	0.019

Topic 101-150

	baseline	pf baseline	pfl1	rf baselinerfl1	
num_ret	50000	49956	49994	49980	49978
num_rel	11645	11645	11645	11645	11645
num_rel_ret	486	990	1396	769	1117
map	0.012	0.033	0.036	0.029	0.025
gm_map	0.002	0.009	0.013	0.010	0.010

Rprec		0.038	0.073	0.085	0.058	0.066
bpref		0.040	0.072	0.101	0.062	0.083
recirank		0.340	0.572	0.480	0.835	0.598
	0	0.369	0.624	0.528	0.835	0.613
	0.1	0.035	0.115	0.149	0.081	0.086
	0.2	0.012	0.032	0.029	0.019	0.015
	0.3	0.000	0.010	0.003	0.005	0.000
	0.4	0.000	0.000	0.001	0.000	0.000
	0.5	0.000	0.000	0.000	0.000	0.000
	0.6	0.000	0.000	0.000	0.000	0.000
	0.7	0.000	0.000	0.000	0.000	0.000
	0.8	0.000	0.000	0.000	0.000	0.000
	0.9	0.000	0.000	0.000	0.000	0.000
	1	0.000	0.000	0.000	0.000	0.000
	5	0.204	0.388	0.360	0.448	0.340
	10	0.164	0.330	0.320	0.332	0.258
	15	0.147	0.280	0.273	0.273	0.232
	20	0.133	0.244	0.249	0.230	0.208
	30	0.113	0.213	0.215	0.187	0.177
	100	0.063	0.132	0.134	0.096	0.098
	200	0.039	0.083	0.095	0.060	0.068
	500	0.018	0.038	0.049	0.029	0.038
	1000	0.010	0.020	0.028	0.015	0.022

Topic 151-200

		baseline	pf baseline	pfl1	rf baseline	rfl1
num_ret		49995	49950	49998	49990	49981
num_rel		9805	9805	9805	9805	9805
num_rel_ret		491	861	1313	674	1203
map		0.032	0.063	0.056	0.055	0.051
gm_map		0.003	0.013	0.015	0.011	0.015
Rprec		0.066	0.104	0.106	0.080	0.097
bpref		0.067	0.104	0.124	0.089	0.116
recirank		0.365	0.674	0.550	0.807	0.583
	0	0.410	0.708	0.587	0.820	0.614
	0.1	0.095	0.214	0.209	0.168	0.166
	0.2	0.049	0.090	0.077	0.075	0.066
	0.3	0.021	0.039	0.028	0.030	0.027
	0.4	0.009	0.012	0.016	0.010	0.014
	0.5	0.009	0.012	0.011	0.009	0.011
	0.6	0.000	0.010	0.011	0.009	0.005
	0.7	0.000	0.009	0.007	0.009	0.002
	0.8	0.000	0.000	0.000	0.006	0.000

0.9	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000
5	0.240	0.456	0.364	0.492	0.396
10	0.220	0.398	0.336	0.346	0.342
15	0.189	0.351	0.303	0.280	0.292
20	0.177	0.308	0.277	0.242	0.266
30	0.157	0.267	0.237	0.201	0.221
100	0.084	0.133	0.132	0.097	0.118
200	0.047	0.081	0.086	0.059	0.077
500	0.020	0.034	0.046	0.026	0.042
1000	0.010	0.017	0.026	0.014	0.024

Topic 201-250

	baseline	pf baseline	pfl1	rf baseline	rfl1
num_ret	50000	49991	49997	49988	49987
num_rel	6503	6503	6503	6503	6503
num_rel_ret	533	659	916	468	813
map	0.024	0.066	0.056	0.052	0.055
gm_map	0.002	0.013	0.014	0.012	0.014
Rprec	0.055	0.101	0.097	0.080	0.092
bpref	0.065	0.108	0.117	0.087	0.111
recirank	0.376	0.688	0.584	0.761	0.620
0	0.404	0.700	0.602	0.770	0.640
0.1	0.084	0.193	0.170	0.130	0.160
0.2	0.021	0.110	0.084	0.076	0.092
0.3	0.011	0.051	0.036	0.036	0.032
0.4	0.000	0.030	0.024	0.023	0.021
0.5	0.000	0.022	0.023	0.022	0.021
0.6	0.000	0.020	0.020	0.020	0.020
0.7	0.000	0.000	0.000	0.000	0.000
0.8	0.000	0.000	0.000	0.000	0.000
0.9	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000
5	0.208	0.392	0.344	0.348	0.324
10	0.166	0.296	0.264	0.238	0.258
15	0.143	0.240	0.215	0.197	0.208
20	0.123	0.204	0.186	0.169	0.176
30	0.105	0.163	0.157	0.144	0.143
100	0.056	0.085	0.084	0.060	0.076
200	0.037	0.052	0.056	0.039	0.051
500	0.020	0.024	0.031	0.018	0.028
1000	0.011	0.013	0.018	0.009	0.016

Topic 251-300

	baseline	pf baseline	pfl1	rf baseline	rfl1
num_ret	49290	49995	49999	49628	49274
num_rel	5524	5524	5524	5524	5524
num_rel_ret	201	324	454	256	394
map	0.016	0.053	0.053	0.051	0.052
gm_map	0.001	0.003	0.005	0.002	0.003
Rprec	0.034	0.072	0.080	0.071	0.074
bpref	0.040	0.077	0.085	0.074	0.084
recirank	0.191	0.351	0.375	0.571	0.450
0	0.209	0.377	0.415	0.575	0.459
0.1	0.060	0.129	0.124	0.152	0.151
0.2	0.028	0.078	0.069	0.094	0.087
0.3	0.025	0.066	0.061	0.077	0.077
0.4	0.002	0.064	0.054	0.039	0.038
0.5	0.002	0.036	0.054	0.036	0.036
0.6	0.000	0.023	0.024	0.000	0.004
0.7	0.000	0.003	0.004	0.000	0.004
0.8	0.000	0.003	0.003	0.000	0.003
0.9	0.000	0.003	0.003	0.000	0.003
1	0.000	0.003	0.003	0.000	0.003
5	0.068	0.224	0.244	0.240	0.216
10	0.058	0.170	0.182	0.134	0.148
15	0.056	0.137	0.151	0.096	0.119
20	0.055	0.116	0.123	0.080	0.102
30	0.051	0.094	0.099	0.061	0.083
100	0.030	0.044	0.044	0.033	0.042
200	0.017	0.027	0.029	0.020	0.026
500	0.008	0.012	0.016	0.010	0.013
1000	0.004	0.007	0.009	0.005	0.008

F4 10 terms, 20 documents

Topic 51-100

		baseline	pf baseline	pfl1	rf baseline	rf1
num_ret		49964	49635	49820	49856	49848
num_rel		16386	16386	16386	16386	16386
num_rel_ret		408	890	1188	537	963
map		0.010	0.032	0.032	0.017	0.021
gm_map		0.000	0.002	0.002	0.001	0.002
Rprec		0.024	0.055	0.061	0.035	0.047
bpref		0.027	0.058	0.072	0.037	0.062
recirank		0.212	0.505	0.428	0.607	0.443
	0	0.245	0.525	0.466	0.610	0.458
	0.1	0.028	0.100	0.118	0.058	0.066
	0.2	0.014	0.058	0.047	0.004	0.020
	0.3	0.000	0.002	0.012	0.000	0.002
	0.4	0.000	0.000	0.000	0.000	0.000
	0.5	0.000	0.000	0.000	0.000	0.000
	0.6	0.000	0.000	0.000	0.000	0.000
	0.7	0.000	0.000	0.000	0.000	0.000
	0.8	0.000	0.000	0.000	0.000	0.000
	0.9	0.000	0.000	0.000	0.000	0.000
	1	0.000	0.000	0.000	0.000	0.000
	5	0.132	0.332	0.324	0.296	0.248
	10	0.124	0.292	0.298	0.218	0.204
	15	0.116	0.265	0.264	0.176	0.175
	20	0.103	0.240	0.239	0.170	0.155
	30	0.084	0.205	0.205	0.147	0.133
	100	0.047	0.110	0.119	0.075	0.086
	200	0.031	0.071	0.080	0.046	0.060
	500	0.015	0.033	0.041	0.021	0.032
	1000	0.008	0.018	0.024	0.011	0.019

Topic 101-150

		baseline	pf baseline	pfl1	rf baseline	rf1
num_ret		50000	49958	49996	49980	49978
num_rel		11645	11645	11645	11645	11645
num_rel_ret		486	1231	1424	847	1117
map		0.012	0.040	0.035	0.034	0.025
gm_map		0.002	0.012	0.013	0.012	0.010
Rprec		0.038	0.086	0.084	0.065	0.066
bpref		0.040	0.087	0.101	0.070	0.083
recirank		0.340	0.629	0.484	0.835	0.598
	0	0.369	0.675	0.532	0.835	0.613

0.1	0.035	0.160	0.150	0.098	0.086
0.2	0.012	0.038	0.028	0.017	0.015
0.3	0.000	0.012	0.014	0.004	0.000
0.4	0.000	0.000	0.000	0.000	0.000
0.5	0.000	0.000	0.000	0.000	0.000
0.6	0.000	0.000	0.000	0.000	0.000
0.7	0.000	0.000	0.000	0.000	0.000
0.8	0.000	0.000	0.000	0.000	0.000
0.9	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000
5	0.204	0.416	0.360	0.488	0.340
10	0.164	0.346	0.312	0.362	0.258
15	0.147	0.301	0.280	0.311	0.232
20	0.133	0.278	0.261	0.270	0.208
30	0.113	0.233	0.231	0.230	0.177
100	0.063	0.143	0.136	0.108	0.098
200	0.039	0.096	0.092	0.068	0.068
500	0.018	0.046	0.049	0.032	0.038
1000	0.010	0.025	0.029	0.017	0.022

Topic 151-200

		baseline	pf baseline	pfl1	rf baseline	rf1
num_ret		49995	49943	49998	49991	49981
num_rel		9805	9805	9805	9805	9805
num_rel_ret		491	897	1372	764	1203
map		0.032	0.068	0.059	0.065	0.051
gm_map		0.003	0.013	0.020	0.012	0.015
Rprec		0.066	0.110	0.112	0.090	0.097
bpref		0.067	0.111	0.129	0.102	0.116
recirank		0.365	0.692	0.619	0.821	0.583
	0	0.410	0.716	0.652	0.826	0.614
	0.1	0.095	0.236	0.211	0.214	0.166
	0.2	0.049	0.103	0.070	0.087	0.066
	0.3	0.021	0.054	0.032	0.042	0.027
	0.4	0.009	0.021	0.017	0.019	0.014
	0.5	0.009	0.011	0.011	0.015	0.011
	0.6	0.000	0.010	0.010	0.013	0.005
	0.7	0.000	0.010	0.005	0.010	0.002
	0.8	0.000	0.000	0.000	0.008	0.000
	0.9	0.000	0.000	0.000	0.000	0.000
	1	0.000	0.000	0.000	0.000	0.000
	5	0.240	0.480	0.440	0.508	0.396
	10	0.220	0.386	0.386	0.384	0.342

	15	0.189	0.336	0.336	0.317	0.292
	20	0.177	0.297	0.310	0.273	0.266
	30	0.157	0.259	0.263	0.217	0.221
	100	0.084	0.137	0.140	0.110	0.118
	200	0.047	0.081	0.092	0.065	0.077
	500	0.020	0.035	0.048	0.029	0.042
	1000	0.010	0.018	0.027	0.015	0.024

Topic 201-250

		baseline	pf baseline	pfl1	rf baseline	rf1
num_ret		50000	49988	49998	49989	49987
num_rel		6503	6503	6503	6503	6503
num_rel_ret		533	750	945	478	813
map		0.0237	0.071	0.062	0.061	0.055
gm_map		0.0024	0.012	0.016	0.013	0.014
Rprec		0.0546	0.108	0.107	0.089	0.092
bpref		0.0645	0.115	0.125	0.096	0.111
recirank		0.3764	0.639	0.601	0.761	0.620
	0	0.4042	0.652	0.623	0.773	0.640
	0.1	0.0843	0.205	0.189	0.167	0.160
	0.2	0.0209	0.120	0.096	0.080	0.092
	0.3	0.011	0.052	0.051	0.061	0.032
	0.4	0	0.042	0.024	0.045	0.021
	0.5	0	0.039	0.022	0.022	0.021
	0.6	0	0.020	0.020	0.020	0.020
	0.7	0	0.000	0.000	0.000	0.000
	0.8	0	0.000	0.000	0.000	0.000
	0.9	0	0.000	0.000	0.000	0.000
	1	0	0.000	0.000	0.000	0.000
	5	0.208	0.396	0.360	0.388	0.324
	10	0.166	0.290	0.276	0.270	0.258
	15	0.1427	0.243	0.228	0.217	0.208
	20	0.123	0.205	0.201	0.189	0.176
	30	0.1047	0.165	0.165	0.159	0.143
	100	0.056	0.086	0.091	0.066	0.076
	200	0.0371	0.053	0.058	0.040	0.051
	500	0.0195	0.026	0.031	0.019	0.028
	1000	0.0107	0.015	0.019	0.010	0.016

Topic 251-300

		baseline	pf baseline	pfl1	rf baseline	rf1
num_ret		49290	49994	49999	49984	49274
num_rel		5524	5524	5524	5524	5524

num_rel_ret		201	323	485	254	394
map		0.016	0.047	0.057	0.053	0.052
gm_map		0.001	0.002	0.005	0.002	0.003
Rprec		0.034	0.071	0.091	0.070	0.074
bpref		0.040	0.074	0.093	0.073	0.084
recirank		0.191	0.333	0.367	0.581	0.450
	0	0.209	0.362	0.408	0.585	0.459
	0.1	0.060	0.144	0.137	0.146	0.151
	0.2	0.028	0.077	0.087	0.101	0.087
	0.3	0.025	0.060	0.060	0.075	0.077
	0.4	0.002	0.037	0.057	0.039	0.038
	0.5	0.002	0.029	0.051	0.036	0.036
	0.6	0.000	0.008	0.040	0.000	0.004
	0.7	0.000	0.002	0.020	0.000	0.004
	0.8	0.000	0.002	0.002	0.000	0.003
	0.9	0.000	0.002	0.002	0.000	0.003
	1	0.000	0.002	0.002	0.000	0.003
	5	0.068	0.220	0.260	0.248	0.216
	10	0.058	0.174	0.188	0.158	0.148
	15	0.056	0.144	0.157	0.115	0.119
	20	0.055	0.125	0.133	0.097	0.102
	30	0.051	0.097	0.102	0.071	0.083
	100	0.030	0.047	0.051	0.035	0.042
	200	0.017	0.027	0.033	0.021	0.026
	500	0.008	0.012	0.017	0.010	0.013
	1000	0.004	0.007	0.010	0.005	0.008

F5 15 terms, 20 documents

Topic 51-100

	baseline	pf baseline	pfl1	rf baseline	rf1
num_ret	49964	49609	49820	49856	49848
num_rel	16386	16386	16386	16386	16386
num_rel_ret	408	944	1215	560	963
map	0.0101	0.030	0.032	0.018	0.021
gm_map	0.0003	0.002	0.003	0.001	0.002
Rprec	0.0241	0.056	0.061	0.038	0.047
bpref	0.0267	0.059	0.074	0.039	0.062
recirank	0.2115	0.484	0.449	0.607	0.443
0	0.2449	0.504	0.485	0.610	0.458
0.1	0.0282	0.106	0.113	0.063	0.066
0.2	0.0135	0.056	0.047	0.004	0.020
0.3	0	0.002	0.015	0.000	0.002
0.4	0	0.000	0.000	0.000	0.000
0.5	0	0.000	0.000	0.000	0.000
0.6	0	0.000	0.000	0.000	0.000
0.7	0	0.000	0.000	0.000	0.000
0.8	0	0.000	0.000	0.000	0.000
0.9	0	0.000	0.000	0.000	0.000
1	0	0.000	0.000	0.000	0.000
5	0.132	0.296	0.332	0.296	0.248
10	0.124	0.264	0.314	0.218	0.204
15	0.116	0.255	0.269	0.187	0.175
20	0.103	0.233	0.256	0.169	0.155
30	0.084	0.199	0.213	0.151	0.133
100	0.047	0.110	0.119	0.077	0.086
200	0.0311	0.071	0.080	0.048	0.060
500	0.0149	0.035	0.042	0.021	0.032
1000	0.0082	0.019	0.024	0.011	0.019

Topic 101-150

	baseline	pf baseline	pfl1	rf baseline	rf1
num_ret	50000	49949	49996	49980	49978
num_rel	11645	11645	11645	11645	11645
num_rel_ret	486	1316	1420	829	1117
map	0.012	0.042	0.037	0.033	0.025
gm_map	0.002	0.012	0.013	0.012	0.010

Rprec		0.038	0.089	0.084	0.063	0.066
bpref		0.040	0.092	0.102	0.068	0.083
recirank		0.340	0.621	0.543	0.835	0.598
	0	0.369	0.655	0.586	0.835	0.613
	0.1	0.035	0.169	0.150	0.098	0.086
	0.2	0.012	0.050	0.026	0.026	0.015
	0.3	0.000	0.013	0.014	0.000	0.000
	0.4	0.000	0.010	0.000	0.000	0.000
	0.5	0.000	0.000	0.000	0.000	0.000
	0.6	0.000	0.000	0.000	0.000	0.000
	0.7	0.000	0.000	0.000	0.000	0.000
	0.8	0.000	0.000	0.000	0.000	0.000
	0.9	0.000	0.000	0.000	0.000	0.000
	1	0.000	0.000	0.000	0.000	0.000
	5	0.204	0.412	0.376	0.484	0.340
	10	0.164	0.346	0.310	0.358	0.258
	15	0.147	0.301	0.281	0.309	0.232
	20	0.133	0.274	0.257	0.270	0.208
	30	0.113	0.238	0.227	0.224	0.177
	100	0.063	0.147	0.136	0.106	0.098
	200	0.039	0.100	0.092	0.069	0.068
	500	0.018	0.049	0.049	0.031	0.038
	1000	0.010	0.026	0.028	0.017	0.022

Topic 151-200

		baseline	pf baseline	pfl1	rf baseline	rf1
num_ret		49995	49936	49998	49991	49981
num_rel		9805	9805	9805	9805	9805
num_rel_ret		491	1000	1391	789	1203
map		0.032	0.074	0.061	0.066	0.051
gm_map		0.003	0.015	0.021	0.012	0.015
Rprec		0.066	0.114	0.116	0.095	0.097
bpref		0.067	0.118	0.133	0.105	0.116
recirank		0.365	0.655	0.609	0.821	0.583
	0	0.410	0.683	0.654	0.826	0.614
	0.1	0.095	0.262	0.215	0.195	0.166
	0.2	0.049	0.108	0.077	0.096	0.066
	0.3	0.021	0.058	0.032	0.050	0.027
	0.4	0.009	0.033	0.017	0.030	0.014
	0.5	0.009	0.011	0.010	0.017	0.011
	0.6	0.000	0.011	0.010	0.012	0.005
	0.7	0.000	0.010	0.006	0.010	0.002

0.8	0.000	0.000	0.000	0.008	0.000
0.9	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000
5	0.240	0.468	0.460	0.516	0.396
10	0.220	0.410	0.410	0.378	0.342
15	0.189	0.363	0.359	0.315	0.292
20	0.177	0.325	0.320	0.275	0.266
30	0.157	0.280	0.268	0.225	0.221
100	0.084	0.147	0.146	0.111	0.118
200	0.047	0.089	0.092	0.068	0.077
500	0.020	0.039	0.048	0.031	0.042
1000	0.010	0.020	0.028	0.016	0.024

Topic 201-250

	baseline	pf baseline	pfl1	rf baseline	rf1
num_ret	50000	49989	49999	49989	49987
num_rel	6503	6503	6503	6503	6503
num_rel_ret	533	807	970	468	813
map	0.024	0.074	0.067	0.061	0.055
gm_map	0.002	0.014	0.017	0.013	0.014
Rprec	0.055	0.112	0.113	0.089	0.092
bpref	0.065	0.122	0.131	0.094	0.111
recirank	0.376	0.633	0.635	0.761	0.620
0	0.404	0.645	0.657	0.773	0.640
0.1	0.084	0.223	0.209	0.162	0.160
0.2	0.021	0.114	0.111	0.080	0.092
0.3	0.011	0.057	0.052	0.060	0.032
0.4	0.000	0.042	0.025	0.033	0.021
0.5	0.000	0.041	0.022	0.023	0.021
0.6	0.000	0.029	0.020	0.020	0.020
0.7	0.000	0.000	0.000	0.000	0.000
0.8	0.000	0.000	0.000	0.000	0.000
0.9	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000
5	0.208	0.420	0.372	0.408	0.324
10	0.166	0.322	0.290	0.276	0.258
15	0.143	0.263	0.239	0.219	0.208
20	0.123	0.232	0.220	0.188	0.176
30	0.105	0.183	0.183	0.155	0.143
100	0.056	0.094	0.095	0.065	0.076
200	0.037	0.061	0.061	0.040	0.051

500	0.020	0.030	0.032	0.018	0.028
1000	0.011	0.016	0.019	0.009	0.016

Topic 251-300

	baseline	pf baseline	pfl1	rf baseline	rf1
num_ret	49290	49989	49999	49984	49274
num_rel	5524	5524	5524	5524	5524
num_rel_ret	201	340	469	271	394
map	0.016	0.044	0.059	0.056	0.052
gm_map	0.001	0.002	0.004	0.002	0.003
Rprec	0.034	0.063	0.087	0.074	0.074
bpref	0.040	0.071	0.092	0.077	0.084
recirank	0.191	0.347	0.346	0.581	0.450
0	0.209	0.366	0.385	0.585	0.459
0.1	0.060	0.156	0.134	0.165	0.151
0.2	0.028	0.076	0.084	0.109	0.087
0.3	0.025	0.058	0.067	0.075	0.077
0.4	0.002	0.017	0.060	0.039	0.038
0.5	0.002	0.015	0.051	0.036	0.036
0.6	0.000	0.009	0.041	0.000	0.004
0.7	0.000	0.003	0.021	0.000	0.004
0.8	0.000	0.001	0.003	0.000	0.003
0.9	0.000	0.001	0.003	0.000	0.003
1	0.000	0.001	0.003	0.000	0.003
5	0.068	0.216	0.260	0.260	0.216
10	0.058	0.172	0.192	0.178	0.148
15	0.056	0.148	0.155	0.133	0.119
20	0.055	0.127	0.131	0.111	0.102
30	0.051	0.099	0.101	0.078	0.083
100	0.030	0.050	0.051	0.038	0.042
200	0.017	0.028	0.032	0.023	0.026
500	0.008	0.013	0.017	0.010	0.013
1000	0.004	0.007	0.009	0.005	0.008

F6 100 terms, 20 documents

Topic 51-100

	baseline	pf baseline	pfl1	rf baseline	rf1
num_ret	49964	47408	49820	49856	49848
num_rel	16386	16386	16386	16386	16386
num_rel_ret	408	1137	1218	695	963
map	0.010	0.037	0.034	0.020	0.021
gm_map	0.000	0.003	0.003	0.002	0.002
Rprec	0.024	0.067	0.064	0.042	0.047
bpref	0.027	0.075	0.076	0.047	0.062
recirank	0.212	0.524	0.446	0.607	0.443
0	0.245	0.543	0.489	0.610	0.458
0.1	0.028	0.152	0.126	0.065	0.066
0.2	0.014	0.057	0.043	0.005	0.020
0.3	0.000	0.025	0.012	0.000	0.002
0.4	0.000	0.000	0.000	0.000	0.000
0.5	0.000	0.000	0.000	0.000	0.000
0.6	0.000	0.000	0.000	0.000	0.000
0.7	0.000	0.000	0.000	0.000	0.000
0.8	0.000	0.000	0.000	0.000	0.000
0.9	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000
5	0.132	0.332	0.336	0.316	0.248
10	0.124	0.290	0.316	0.234	0.204
15	0.116	0.268	0.293	0.192	0.175
20	0.103	0.248	0.268	0.178	0.155
30	0.084	0.211	0.231	0.163	0.133
100	0.047	0.116	0.122	0.089	0.086
200	0.031	0.077	0.081	0.055	0.060
500	0.015	0.039	0.042	0.026	0.032
1000	0.008	0.023	0.024	0.014	0.019

Topic 101-150

	baseline	pf baseline	pfl1	rf baseline	rf1
num_ret	50000	49415	49996	49980	49978
num_rel	11645	11645	11645	11645	11645
num_rel_ret	486	1505	1417	919	1117
map	0.012	0.042	0.039	0.036	0.025
gm_map	0.002	0.012	0.015	0.012	0.010

Rprec	0.038	0.087	0.087	0.066	0.066
bpref	0.040	0.103	0.105	0.074	0.083
recirank	0.340	0.567	0.551	0.835	0.598
0	0.369	0.608	0.604	0.835	0.613
0.1	0.035	0.158	0.151	0.108	0.086
0.2	0.012	0.052	0.034	0.027	0.015
0.3	0.000	0.014	0.002	0.006	0.000
0.4	0.000	0.010	0.000	0.000	0.000
0.5	0.000	0.000	0.000	0.000	0.000
0.6	0.000	0.000	0.000	0.000	0.000
0.7	0.000	0.000	0.000	0.000	0.000
0.8	0.000	0.000	0.000	0.000	0.000
0.9	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000
5	0.204	0.412	0.360	0.500	0.340
10	0.164	0.350	0.346	0.378	0.258
15	0.147	0.316	0.317	0.327	0.232
20	0.133	0.291	0.285	0.281	0.208
30	0.113	0.240	0.252	0.231	0.177
100	0.063	0.145	0.141	0.111	0.098
200	0.039	0.098	0.094	0.070	0.068
500	0.018	0.054	0.049	0.035	0.038
1000	0.010	0.030	0.028	0.018	0.022

Topic 151-200

	baseline	pf baseline	pfl1	rf baseline	rf1
num_ret	49995	49338	49998	49991	49981
num_rel	9805	9805	9805	9805	9805
num_rel_ret	491	1152	1439	864	1203
map	0.032	0.076	0.066	0.071	0.051
gm_map	0.003	0.015	0.023	0.014	0.015
Rprec	0.066	0.119	0.120	0.097	0.097
bpref	0.067	0.127	0.136	0.108	0.116
recirank	0.365	0.639	0.627	0.821	0.583
0	0.410	0.668	0.669	0.826	0.614
0.1	0.095	0.269	0.262	0.210	0.166
0.2	0.049	0.116	0.075	0.100	0.066
0.3	0.021	0.056	0.036	0.053	0.027
0.4	0.009	0.030	0.017	0.031	0.014
0.5	0.009	0.021	0.010	0.018	0.011
0.6	0.000	0.009	0.010	0.012	0.005
0.7	0.000	0.007	0.006	0.010	0.002

0.8	0.000	0.000	0.000	0.004	0.000
0.9	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000
5	0.240	0.456	0.464	0.516	0.396
10	0.220	0.422	0.416	0.402	0.342
15	0.189	0.383	0.384	0.341	0.292
20	0.177	0.344	0.341	0.299	0.266
30	0.157	0.293	0.292	0.248	0.221
100	0.084	0.151	0.158	0.115	0.118
200	0.047	0.091	0.100	0.073	0.077
500	0.020	0.043	0.050	0.033	0.042
1000	0.010	0.023	0.029	0.017	0.024

Topic 201-250

	baseline	pf baseline	pfl1	rf baseline	rf1
num_ret	50000	49990	49999	49989	49987
num_rel	6503	6503	6503	6503	6503
num_rel_ret	533	813	998	566	813
map	0.024	0.056	0.067	0.070	0.055
gm_map	0.002	0.011	0.017	0.015	0.014
Rprec	0.055	0.091	0.114	0.099	0.092
bpref	0.065	0.106	0.132	0.108	0.111
recirank	0.376	0.554	0.654	0.771	0.620
0	0.404	0.577	0.677	0.776	0.640
0.1	0.084	0.189	0.196	0.177	0.160
0.2	0.021	0.098	0.109	0.092	0.092
0.3	0.011	0.040	0.052	0.055	0.032
0.4	0.000	0.026	0.039	0.045	0.021
0.5	0.000	0.018	0.022	0.040	0.021
0.6	0.000	0.000	0.020	0.035	0.020
0.7	0.000	0.000	0.000	0.010	0.000
0.8	0.000	0.000	0.000	0.000	0.000
0.9	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000
5	0.208	0.372	0.396	0.452	0.324
10	0.166	0.304	0.308	0.320	0.258
15	0.143	0.247	0.260	0.261	0.208
20	0.123	0.216	0.234	0.219	0.176
30	0.105	0.179	0.191	0.174	0.143
100	0.056	0.085	0.098	0.076	0.076
200	0.037	0.052	0.064	0.045	0.051

500	0.020	0.028	0.033	0.021	0.028
1000	0.011	0.016	0.020	0.011	0.016

Topic 251-300

	baseline	pf baseline	pfl1	rf baseline	rf1
num_ret	49290	49999	50000	49984	49274
num_rel	5524	5524	5524	5524	5524
num_rel_ret	201	338	514	361	394
map	0.016	0.033	0.051	0.059	0.052
gm_map	0.001	0.002	0.003	0.002	0.003
Rprec	0.034	0.047	0.081	0.080	0.074
bpref	0.040	0.057	0.084	0.086	0.084
recirank	0.191	0.323	0.354	0.581	0.450
0	0.209	0.359	0.388	0.585	0.459
0.1	0.060	0.114	0.140	0.179	0.151
0.2	0.028	0.040	0.083	0.084	0.087
0.3	0.025	0.029	0.050	0.074	0.077
0.4	0.002	0.029	0.045	0.058	0.038
0.5	0.002	0.007	0.032	0.036	0.036
0.6	0.000	0.006	0.025	0.000	0.004
0.7	0.000	0.005	0.005	0.000	0.004
0.8	0.000	0.004	0.004	0.000	0.003
0.9	0.000	0.004	0.004	0.000	0.003
1	0.000	0.004	0.004	0.000	0.003
5	0.068	0.180	0.236	0.288	0.216
10	0.058	0.154	0.186	0.186	0.148
15	0.056	0.121	0.141	0.141	0.119
20	0.055	0.100	0.121	0.115	0.102
30	0.051	0.077	0.099	0.085	0.083
100	0.030	0.036	0.054	0.044	0.042
200	0.017	0.023	0.034	0.029	0.026
500	0.008	0.012	0.018	0.013	0.013
1000	0.004	0.007	0.010	0.007	0.008

F7 200 terms, 20 documents

Topic 51-100

		baseline	pf baseline	pfl1	rf baseline	rf1
num_ret		49964	49848	49820	49856	49848
num_rel		16386	16386	16386	16386	16386
num_rel_ret		408	495	1218	608	963
map		0.010	0.017	0.034	0.021	0.021
gm_map		0.000	0.001	0.003	0.001	0.002
Rprec		0.024	0.033	0.064	0.041	0.047
bpref		0.027	0.034	0.076	0.043	0.062
recirank		0.212	0.440	0.446	0.538	0.443
	0	0.245	0.452	0.489	0.549	0.458
	0.1	0.028	0.054	0.126	0.072	0.066
	0.2	0.014	0.014	0.043	0.014	0.020
	0.3	0.000	0.000	0.012	0.000	0.002
	0.4	0.000	0.000	0.000	0.000	0.000
	0.5	0.000	0.000	0.000	0.000	0.000
	0.6	0.000	0.000	0.000	0.000	0.000
	0.7	0.000	0.000	0.000	0.000	0.000
	0.8	0.000	0.000	0.000	0.000	0.000
	0.9	0.000	0.000	0.000	0.000	0.000
	1	0.000	0.000	0.000	0.000	0.000
	5	0.132	0.248	0.336	0.316	0.248
	10	0.124	0.196	0.316	0.238	0.204
	15	0.116	0.168	0.293	0.200	0.175
	20	0.103	0.147	0.268	0.183	0.155
	30	0.084	0.121	0.231	0.162	0.133
	100	0.047	0.064	0.122	0.087	0.086
	200	0.031	0.040	0.081	0.052	0.060
	500	0.015	0.018	0.042	0.024	0.032
	1000	0.008	0.010	0.024	0.012	0.019

Topic 101-150

	baseline	pf baseline	pfl1	rf baseline	rf1
num_ret	50000	49978	49996	49979	49978
num_rel	11645	11645	11645	11645	11645
num_rel_ret	486	570	1417	632	1117
map	0.0124	0.021	0.039	0.028	0.025
gm_map	0.002	0.006	0.015	0.010	0.010

Rprec		0.038	0.046	0.087	0.052	0.066
bpref		0.040	0.047	0.105	0.056	0.083
recirank		0.340	0.596	0.551	0.794	0.598
	0	0.369	0.610	0.604	0.798	0.613
	0.1	0.035	0.063	0.151	0.086	0.086
	0.2	0.012	0.013	0.034	0.018	0.015
	0.3	0.000	0.000	0.002	0.000	0.000
	0.4	0.000	0.000	0.000	0.000	0.000
	0.5	0.000	0.000	0.000	0.000	0.000
	0.6	0.000	0.000	0.000	0.000	0.000
	0.7	0.000	0.000	0.000	0.000	0.000
	0.8	0.000	0.000	0.000	0.000	0.000
	0.9	0.000	0.000	0.000	0.000	0.000
	1	0.000	0.000	0.000	0.000	0.000
	5	0.204	0.340	0.360	0.444	0.340
	10	0.164	0.254	0.346	0.328	0.258
	15	0.147	0.223	0.317	0.285	0.232
	20	0.133	0.197	0.285	0.248	0.208
	30	0.113	0.164	0.252	0.201	0.177
	100	0.063	0.080	0.141	0.087	0.098
	200	0.039	0.047	0.094	0.050	0.068
	500	0.018	0.021	0.049	0.024	0.038
	1000	0.010	0.011	0.028	0.013	0.022

Topic 151-200

		baseline	pf baseline	pfl1	rf baseline	rf1
num_ret		49995	49981	49998	49990	49981
num_rel		9805	9805	9805	9805	9805
num_rel_ret		491	569	1439	647	1203
map		0.0315	0.046	0.066	0.051	0.051
gm_map		0.003	0.008	0.023	0.011	0.015
Rprec		0.066	0.079	0.120	0.087	0.097
bpref		0.067	0.079	0.136	0.088	0.116
recirank		0.365	0.581	0.627	0.669	0.583
	0	0.410	0.611	0.669	0.698	0.614
	0.1	0.095	0.134	0.262	0.173	0.166
	0.2	0.049	0.057	0.075	0.066	0.066
	0.3	0.021	0.021	0.036	0.021	0.027
	0.4	0.009	0.009	0.017	0.009	0.014
	0.5	0.009	0.009	0.010	0.009	0.011
	0.6	0.000	0.000	0.010	0.000	0.005
	0.7	0.000	0.000	0.006	0.000	0.002

0.8	0.000	0.000	0.000	0.000	0.000
0.9	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000
5	0.240	0.396	0.464	0.428	0.396
10	0.220	0.340	0.416	0.370	0.342
15	0.189	0.283	0.384	0.309	0.292
20	0.177	0.254	0.341	0.277	0.266
30	0.157	0.209	0.292	0.229	0.221
100	0.084	0.099	0.158	0.106	0.118
200	0.047	0.054	0.100	0.060	0.077
500	0.020	0.023	0.050	0.026	0.042
1000	0.010	0.011	0.029	0.013	0.024

Topic 201-250

		baseline	pf baseline	pfl1	rf baseline	rf1
num_ret		50000	49987	49999	49987	49987
num_rel		6503	6503	6503	6503	6503
num_rel_ret		533	583	998	613	813
map		0.0237	0.052	0.067	0.062	0.055
gm_map		0.002	0.010	0.017	0.013	0.014
Rprec		0.055	0.084	0.114	0.095	0.092
bpref		0.065	0.093	0.132	0.105	0.111
recirank		0.376	0.626	0.654	0.705	0.620
	0	0.404	0.645	0.677	0.719	0.640
	0.1	0.084	0.153	0.196	0.166	0.160
	0.2	0.021	0.084	0.109	0.084	0.092
	0.3	0.011	0.031	0.052	0.047	0.032
	0.4	0.000	0.020	0.039	0.036	0.021
	0.5	0.000	0.020	0.022	0.036	0.021
	0.6	0.000	0.020	0.020	0.035	0.020
	0.7	0.000	0.000	0.000	0.010	0.000
	0.8	0.000	0.000	0.000	0.000	0.000
	0.9	0.000	0.000	0.000	0.000	0.000
	1	0.000	0.000	0.000	0.000	0.000
	5	0.208	0.332	0.396	0.356	0.324
	10	0.166	0.250	0.308	0.272	0.258
	15	0.143	0.204	0.260	0.223	0.208
	20	0.123	0.171	0.234	0.189	0.176
	30	0.105	0.137	0.191	0.153	0.143
	100	0.056	0.066	0.098	0.074	0.076
	200	0.037	0.042	0.064	0.046	0.051

500	0.020	0.022	0.033	0.023	0.028
1000	0.011	0.012	0.020	0.012	0.016

Topic 251-300

	baseline	pf baseline	pfl1	rf baseline	rf1
num_ret	49290	49274	50000	49274	49274
num_rel	5524	5524	5524	5524	5524
num_rel_ret	201	247	514	290	394
map	0.016	0.051	0.051	0.052	0.052
gm_map	0.001	0.001	0.003	0.002	0.003
Rprec	0.034	0.071	0.081	0.075	0.074
bpref	0.040	0.073	0.084	0.078	0.084
recirank	0.191	0.448	0.354	0.506	0.450
0	0.209	0.455	0.388	0.512	0.459
0.1	0.060	0.146	0.140	0.162	0.151
0.2	0.028	0.083	0.083	0.083	0.087
0.3	0.025	0.077	0.050	0.077	0.077
0.4	0.002	0.038	0.045	0.038	0.038
0.5	0.002	0.036	0.032	0.036	0.036
0.6	0.000	0.003	0.025	0.000	0.004
0.7	0.000	0.003	0.005	0.000	0.004
0.8	0.000	0.003	0.004	0.000	0.003
0.9	0.000	0.003	0.004	0.000	0.003
1	0.000	0.003	0.004	0.000	0.003
5	0.068	0.216	0.236	0.248	0.216
10	0.058	0.148	0.186	0.168	0.148
15	0.056	0.117	0.141	0.131	0.119
20	0.055	0.101	0.121	0.109	0.102
30	0.051	0.082	0.099	0.085	0.083
100	0.030	0.039	0.054	0.042	0.042
200	0.017	0.022	0.034	0.026	0.026
500	0.008	0.010	0.018	0.011	0.013
1000	0.004	0.005	0.010	0.006	0.008

F8 5 documents, 20 terms

Topic 51-100

	baseline	pf baseline	pfl1	rf baseline	rf1
num_ret	49964	49664	49842	49854	49848
num_rel	16386	16386	16386	16386	16386
num_rel_ret	408	909	1131	550	963
map	0.010	0.037	0.029	0.024	0.021
gm_map	0.000	0.001	0.003	0.001	0.002
Rprec	0.024	0.059	0.058	0.039	0.047
bpref	0.027	0.064	0.071	0.042	0.062
recirank	0.212	0.513	0.460	0.572	0.443
0	0.245	0.535	0.487	0.577	0.458
0.1	0.028	0.131	0.115	0.069	0.066
0.2	0.014	0.053	0.035	0.013	0.020
0.3	0.000	0.025	0.009	0.000	0.002
0.4	0.000	0.021	0.000	0.000	0.000
0.5	0.000	0.000	0.000	0.000	0.000
0.6	0.000	0.000	0.000	0.000	0.000
0.7	0.000	0.000	0.000	0.000	0.000
0.8	0.000	0.000	0.000	0.000	0.000
0.9	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000
5	0.132	0.276	0.304	0.348	0.248
10	0.124	0.264	0.270	0.254	0.204
15	0.116	0.252	0.231	0.217	0.175
20	0.103	0.233	0.207	0.188	0.155
30	0.084	0.202	0.182	0.161	0.133
100	0.047	0.104	0.108	0.083	0.086
200	0.031	0.068	0.072	0.047	0.060
500	0.015	0.033	0.038	0.021	0.032
1000	0.008	0.018	0.023	0.011	0.019

Topic 101-150

	baseline	pf baseline	pfl1	rf baseline	rf1
num_ret	50000	49959	49990	49979	49978
num_rel	11645	11645	11645	11645	11645
num_rel_ret	486	1178	1339	666	1116
map	0.012	0.037	0.032	0.027	0.025
gm_map	0.002	0.009	0.012	0.008	0.010
Rprec	0.038	0.075	0.008	0.057	0.065

bpref	0.040	0.082	0.097	0.059	0.083
recirank	0.340	0.552	0.578	0.676	0.598
0	0.369	0.595	0.625	0.686	0.613
0.1	0.035	0.142	0.105	0.091	0.088
0.2	0.012	0.035	0.034	0.011	0.015
0.3	0.000	0.012	0.001	0.000	0.000
0.4	0.000	0.000	0.000	0.000	0.000
0.5	0.000	0.000	0.000	0.000	0.000
0.6	0.000	0.000	0.000	0.000	0.000
0.7	0.000	0.000	0.000	0.000	0.000
0.8	0.000	0.000	0.000	0.000	0.000
0.9	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000
5	0.204	0.364	0.408	0.396	0.340
10	0.164	0.352	0.326	0.296	0.258
15	0.147	0.311	0.273	0.255	0.229
20	0.133	0.267	0.242	0.226	0.207
30	0.113	0.225	0.202	0.189	0.173
100	0.063	0.133	0.122	0.096	0.097
200	0.039	0.090	0.085	0.058	0.067
500	0.018	0.043	0.046	0.026	0.038
1000	0.010	0.024	0.027	0.013	0.022

Topic 151-200

	baseline	pf baseline	pfl1	rf baseline	rf1
num_ret	49995	49927	49991	49981	49981
num_rel	9805	9805	9805	9805	9805
num_rel_ret	491	904	1322	596	1203
map	0.032	0.070	0.062	0.049	0.051
gm_map	0.003	0.011	0.018	0.009	0.015
Rprec	0.066	0.111	0.109	0.077	0.097
bpref	0.067	0.114	0.128	0.081	0.116
recirank	0.365	0.618	0.590	0.714	0.583
0	0.410	0.651	0.633	0.727	0.614
0.1	0.095	0.276	0.223	0.137	0.166
0.2	0.049	0.106	0.075	0.053	0.066
0.3	0.021	0.052	0.035	0.017	0.027
0.4	0.009	0.024	0.020	0.014	0.014
0.5	0.009	0.018	0.013	0.012	0.011
0.6	0.000	0.010	0.011	0.011	0.005
0.7	0.000	0.008	0.005	0.010	0.002
0.8	0.000	0.000	0.003	0.007	0.000

0.9	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000
5	0.240	0.468	0.452	0.444	0.396
10	0.220	0.388	0.408	0.352	0.342
15	0.189	0.344	0.339	0.296	0.292
20	0.177	0.310	0.303	0.252	0.266
30	0.157	0.265	0.251	0.196	0.221
100	0.084	0.134	0.136	0.091	0.118
200	0.047	0.081	0.090	0.053	0.077
500	0.020	0.035	0.046	0.023	0.042
1000	0.010	0.018	0.026	0.012	0.024

Topic 201-250

	baseline	pf baseline	pfl1	rf baseline	rf1
num_ret	50000	49983	49987	49989	49987
num_rel	6503	6503	6503	6503	6503
num_rel_ret	533	780	877	514	810
map	0.024	0.066	0.061	0.057	0.055
gm_map	0.002	0.012	0.016	0.011	0.013
Rprec	0.055	0.103	0.103	0.088	0.091
bpref	0.065	0.110	0.117	0.094	0.110
recirank	0.376	0.660	0.660	0.735	0.620
0	0.404	0.671	0.676	0.740	0.639
0.1	0.084	0.194	0.198	0.144	0.159
0.2	0.021	0.107	0.098	0.074	0.091
0.3	0.011	0.050	0.047	0.037	0.032
0.4	0.000	0.041	0.024	0.026	0.021
0.5	0.000	0.030	0.015	0.023	0.021
0.6	0.000	0.020	0.015	0.020	0.020
0.7	0.000	0.000	0.000	0.000	0.000
0.8	0.000	0.000	0.000	0.000	0.000
0.9	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000
5	0.208	0.380	0.380	0.360	0.324
10	0.166	0.284	0.296	0.258	0.256
15	0.143	0.235	0.231	0.208	0.208
20	0.123	0.202	0.191	0.176	0.177
30	0.105	0.165	0.158	0.146	0.143
100	0.056	0.082	0.083	0.067	0.074
200	0.037	0.051	0.055	0.042	0.050
500	0.020	0.027	0.029	0.020	0.027

1000	0.011	0.016	0.018	0.010	0.016

Topic 251-300

	baseline	pf baseline	pfl1	rf baseline	rf1
num_ret	49290	49987	49985	49984	49274
num_rel	5524	5524	5524	5524	5524
num_rel_ret	201	339	417	269	394
map	0.016	0.048	0.059	0.057	0.052
gm_map	0.001	0.002	0.004	0.002	0.003
Rprec	0.034	0.074	0.086	0.077	0.074
bpref	0.040	0.075	0.091	0.079	0.084
recirank	0.191	0.311	0.401	0.498	0.450
0	0.209	0.341	0.432	0.505	0.459
0.1	0.060	0.126	0.158	0.173	0.151
0.2	0.028	0.079	0.080	0.103	0.087
0.3	0.025	0.060	0.061	0.091	0.077
0.4	0.002	0.057	0.055	0.038	0.038
0.5	0.002	0.031	0.055	0.036	0.036
0.6	0.000	0.021	0.025	0.003	0.004
0.7	0.000	0.001	0.005	0.003	0.004
0.8	0.000	0.001	0.003	0.003	0.003
0.9	0.000	0.001	0.003	0.003	0.003
1	0.000	0.001	0.003	0.003	0.003
5	0.068	0.208	0.256	0.232	0.216
10	0.058	0.168	0.194	0.172	0.148
15	0.056	0.135	0.153	0.141	0.119
20	0.055	0.112	0.131	0.112	0.102
30	0.051	0.094	0.099	0.085	0.083
100	0.030	0.047	0.045	0.040	0.042
200	0.017	0.028	0.029	0.023	0.026
500	0.008	0.013	0.015	0.011	0.013
1000	0.004	0.007	0.008	0.005	0.008

F9 10 documents, 20 terms

Topic 51-100

	baseline	pf baseline	pfl1	rf baseline	rf1
num_ret	49964	49582	49837	49855	49848
num_rel	16386	16386	16386	16386	16386
num_rel_ret	408	1063	1465	568	1318
map	0.010	0.039	0.034	0.018	0.024
gm_map	0.000	0.002	0.004	0.002	0.003
Rprec	0.024	0.067	0.070	0.036	0.058
bpref	0.027	0.073	0.086	0.039	0.080
recirank	0.212	0.517	0.489	0.597	0.443
0	0.245	0.541	0.518	0.600	0.459
0.1	0.028	0.150	0.136	0.038	0.085
0.2	0.014	0.052	0.043	0.003	0.031
0.3	0.000	0.023	0.009	0.000	0.006
0.4	0.000	0.009	0.000	0.000	0.001
0.5	0.000	0.000	0.000	0.000	0.000
0.6	0.000	0.000	0.000	0.000	0.000
0.7	0.000	0.000	0.000	0.000	0.000
0.8	0.000	0.000	0.000	0.000	0.000
0.9	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000
5	0.132	0.320	0.332	0.324	0.248
10	0.124	0.294	0.302	0.242	0.204
15	0.116	0.267	0.261	0.207	0.177
20	0.103	0.250	0.231	0.184	0.158
30	0.084	0.211	0.197	0.158	0.137
100	0.047	0.122	0.124	0.079	0.095
200	0.031	0.078	0.085	0.047	0.070
500	0.015	0.037	0.047	0.021	0.041
1000	0.008	0.021	0.029	0.011	0.026

Topic 101-150

	baseline	pf baseline	pfl1	rf baseline	rf1
num_ret	50000	49937	49994	49979	49978
num_rel	11645	11645	11645	11645	11645
num_rel_ret	486	1291	1476	866	1116
map	0.012	0.040	0.039	0.032	0.025
gm_map	0.002	0.011	0.016	0.009	0.010
Rprec	0.038	0.085	0.085	0.067	0.065

bpref	0.040	0.090	0.105	0.071	0.083
recirank	0.340	0.578	0.602	0.704	0.598
0	0.369	0.617	0.643	0.712	0.613
0.1	0.035	0.163	0.136	0.107	0.088
0.2	0.012	0.048	0.036	0.015	0.015
0.3	0.000	0.013	0.012	0.000	0.000
0.4	0.000	0.000	0.000	0.000	0.000
0.5	0.000	0.000	0.000	0.000	0.000
0.6	0.000	0.000	0.000	0.000	0.000
0.7	0.000	0.000	0.000	0.000	0.000
0.8	0.000	0.000	0.000	0.000	0.000
0.9	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000
5	0.204	0.360	0.444	0.420	0.340
10	0.164	0.318	0.368	0.318	0.258
15	0.147	0.293	0.305	0.272	0.229
20	0.133	0.269	0.279	0.240	0.207
30	0.113	0.235	0.237	0.203	0.173
100	0.063	0.145	0.132	0.107	0.097
200	0.039	0.097	0.092	0.072	0.067
500	0.018	0.048	0.049	0.033	0.038
1000	0.010	0.026	0.030	0.017	0.022

Topic 151-200

	baseline	pf baseline	pfl1	rf baseline	rf1
num_ret	49995	49921	49993	49987	49981
num_rel	9805	9805	9805	9805	9805
num_rel_ret	491	1026	980	732	1203
map	0.032	0.075	0.031	0.055	0.051
gm_map	0.003	0.014	0.004	0.011	0.015
Rprec	0.066	0.118	0.072	0.083	0.097
bpref	0.067	0.123	0.090	0.092	0.116
recirank	0.365	0.683	0.353	0.772	0.583
0	0.410	0.709	0.391	0.782	0.614
0.1	0.095	0.285	0.103	0.148	0.166
0.2	0.049	0.125	0.033	0.058	0.066
0.3	0.021	0.056	0.015	0.025	0.027
0.4	0.009	0.033	0.008	0.018	0.014
0.5	0.009	0.016	0.007	0.013	0.011
0.6	0.000	0.012	0.006	0.012	0.005
0.7	0.000	0.010	0.004	0.011	0.002
0.8	0.000	0.000	0.003	0.008	0.000

0.9	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000
5	0.240	0.484	0.224	0.488	0.396
10	0.220	0.402	0.226	0.360	0.342
15	0.189	0.372	0.208	0.292	0.292
20	0.177	0.331	0.187	0.249	0.266
30	0.157	0.275	0.158	0.201	0.221
100	0.084	0.144	0.087	0.101	0.118
200	0.047	0.087	0.062	0.062	0.077
500	0.020	0.040	0.034	0.029	0.042
1000	0.010	0.021	0.020	0.015	0.024

Topic 201-250

	baseline	pf baseline	pfl1	rf baseline	rf1
num_ret	50000	49990	49997	49989	49987
num_rel	6503	6503	6503	6503	6503
num_rel_ret	533	832	933	548	809
map	0.024	0.077	0.068	0.064	0.055
gm_map	0.002	0.013	0.017	0.012	0.013
Rprec	0.055	0.112	0.114	0.095	0.092
bpref	0.065	0.122	0.128	0.103	0.111
recirank	0.376	0.643	0.683	0.751	0.620
0	0.404	0.661	0.707	0.756	0.638
0.1	0.084	0.224	0.198	0.159	0.159
0.2	0.021	0.127	0.109	0.085	0.092
0.3	0.011	0.062	0.058	0.054	0.032
0.4	0.000	0.040	0.039	0.040	0.021
0.5	0.000	0.038	0.016	0.040	0.021
0.6	0.000	0.036	0.015	0.020	0.020
0.7	0.000	0.000	0.000	0.000	0.000
0.8	0.000	0.000	0.000	0.000	0.000
0.9	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000
5	0.208	0.452	0.408	0.404	0.324
10	0.166	0.348	0.304	0.266	0.256
15	0.143	0.295	0.261	0.217	0.208
20	0.123	0.254	0.219	0.189	0.177
30	0.105	0.199	0.177	0.162	0.143
100	0.056	0.093	0.091	0.075	0.075
200	0.037	0.058	0.058	0.046	0.050
500	0.020	0.029	0.032	0.021	0.028

1000	0.011	0.017	0.019	0.011	0.016

Topic 251-300

	baseline	pf baseline	pfl1	rf baseline	rf1
num_ret	49290	49993	50000	49984	49274
num_rel	5524	5524	5524	5524	5524
num_rel_ret	201	386	426	243	394
map	0.016	0.051	0.056	0.060	0.052
gm_map	0.001	0.002	0.004	0.002	0.003
Rprec	0.034	0.078	0.089	0.076	0.074
bpref	0.040	0.080	0.089	0.078	0.084
recirank	0.191	0.348	0.392	0.532	0.450
0	0.209	0.375	0.413	0.538	0.459
0.1	0.060	0.143	0.151	0.145	0.151
0.2	0.028	0.093	0.083	0.098	0.087
0.3	0.025	0.063	0.062	0.088	0.077
0.4	0.002	0.053	0.054	0.053	0.038
0.5	0.002	0.031	0.049	0.053	0.036
0.6	0.000	0.020	0.036	0.023	0.004
0.7	0.000	0.000	0.016	0.003	0.004
0.8	0.000	0.000	0.000	0.003	0.003
0.9	0.000	0.000	0.000	0.003	0.003
1	0.000	0.000	0.000	0.003	0.003
5	0.068	0.212	0.248	0.240	0.216
10	0.058	0.180	0.184	0.166	0.148
15	0.056	0.151	0.140	0.131	0.119
20	0.055	0.131	0.120	0.104	0.102
30	0.051	0.103	0.095	0.079	0.083
100	0.030	0.048	0.046	0.034	0.042
200	0.017	0.030	0.029	0.020	0.026
500	0.008	0.014	0.015	0.009	0.013
1000	0.004	0.008	0.009	0.005	0.008

F10 15 documents, 20 terms

Topic 51-100

	baseline	pf baseline	pfl1	rf baseline	rf1
num_ret	49964	49532	49830	49856	49848
num_rel	16386	16386	16386	16386	16386
num_rel_ret	408	1028	1174	534	963
map	0.0101	0.037	0.032	0.0166	0.0206
gm_map	0.000	0.002	0.002	0.001	0.002
Rprec	0.024	0.066	0.063	0.035	0.047
bpref	0.027	0.069	0.074	0.037	0.062
recirank	0.212	0.514	0.456	0.591	0.443
0	0.245	0.536	0.493	0.598	0.458
0.1	0.028	0.129	0.128	0.044	0.066
0.2	0.014	0.060	0.042	0.004	0.020
0.3	0.000	0.021	0.009	0.000	0.002
0.4	0.000	0.009	0.000	0.000	0.000
0.5	0.000	0.000	0.000	0.000	0.000
0.6	0.000	0.000	0.000	0.000	0.000
0.7	0.000	0.000	0.000	0.000	0.000
0.8	0.000	0.000	0.000	0.000	0.000
0.9	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000
5	0.132	0.336	0.328	0.292	0.248
10	0.124	0.286	0.310	0.218	0.204
15	0.116	0.265	0.279	0.185	0.175
20	0.103	0.245	0.246	0.168	0.155
30	0.084	0.218	0.212	0.143	0.133
100	0.047	0.119	0.116	0.074	0.086
200	0.031	0.076	0.078	0.044	0.060
500	0.015	0.036	0.040	0.020	0.032
1000	0.008	0.021	0.024	0.011	0.019

Topic 101-150

	baseline	pf baseline	pfl1	rf baseline	rf1
num_ret	50000	49941	49996	49980	49978
num_rel	11645	11645	11645	11645	11645
num_rel_ret	486	1346	1458	860	1115
map	0.012	0.042	0.038	0.032	0.025
gm_map	0.002	0.012	0.016	0.010	0.010

Rprec	0.038	0.088	0.088	0.061	0.065
bpref	0.040	0.093	0.106	0.067	0.083
recirank	0.340	0.606	0.577	0.762	0.598
0	0.369	0.641	0.619	0.768	0.613
0.1	0.035	0.169	0.146	0.103	0.087
0.2	0.012	0.054	0.041	0.026	0.015
0.3	0.000	0.013	0.007	0.000	0.000
0.4	0.000	0.009	0.000	0.000	0.000
0.5	0.000	0.000	0.000	0.000	0.000
0.6	0.000	0.000	0.000	0.000	0.000
0.7	0.000	0.000	0.000	0.000	0.000
0.8	0.000	0.000	0.000	0.000	0.000
0.9	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000
5	0.204	0.400	0.392	0.456	0.340
10	0.164	0.328	0.356	0.348	0.258
15	0.147	0.293	0.309	0.300	0.229
20	0.133	0.261	0.283	0.261	0.207
30	0.113	0.227	0.245	0.214	0.173
100	0.063	0.147	0.141	0.110	0.097
200	0.039	0.102	0.094	0.071	0.067
500	0.018	0.050	0.049	0.033	0.038
1000	0.010	0.027	0.029	0.017	0.022

Topic 151-200

	baseline	pf baseline	pfl1	rf baseline	rf1
num_ret	49995	49921	49999	49989	49981
num_rel	9805	9805	9805	9805	9805
num_rel_ret	491	1094	1456	760	1203
map	0.032	0.076	0.065	0.060	0.051
gm_map	0.003	0.016	0.021	0.011	0.015
Rprec	0.066	0.121	0.118	0.088	0.097
bpref	0.067	0.125	0.134	0.097	0.116
recirank	0.365	0.716	0.597	0.790	0.583
0	0.410	0.732	0.636	0.799	0.614
0.1	0.095	0.269	0.233	0.169	0.166
0.2	0.049	0.125	0.095	0.074	0.066
0.3	0.021	0.055	0.035	0.051	0.027
0.4	0.009	0.033	0.017	0.028	0.014
0.5	0.009	0.016	0.013	0.016	0.011
0.6	0.000	0.011	0.011	0.012	0.005
0.7	0.000	0.010	0.010	0.011	0.002

0.8	0.000	0.000	0.003	0.008	0.000
0.9	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000
5	0.240	0.488	0.440	0.484	0.396
10	0.220	0.420	0.408	0.350	0.342
15	0.189	0.375	0.373	0.288	0.292
20	0.177	0.336	0.334	0.257	0.266
30	0.157	0.290	0.283	0.208	0.221
100	0.084	0.154	0.155	0.104	0.118
200	0.047	0.095	0.100	0.065	0.077
500	0.020	0.043	0.050	0.030	0.042
1000	0.010	0.022	0.029	0.015	0.024

Topic 201-250

	baseline	pf baseline	pfl1	rf baseline	rf1
num_ret	50000	49987	49996	49989	49987
num_rel	6503	6503	6503	6503	6503
num_rel_ret	533	785	1007	503	815
map	0.024	0.074	0.068	0.058	0.055
gm_map	0.002	0.014	0.019	0.011	0.014
Rprec	0.055	0.112	0.118	0.086	0.092
bpref	0.065	0.122	0.134	0.093	0.111
recirank	0.376	0.652	0.663	0.751	0.620
0	0.404	0.675	0.683	0.756	0.639
0.1	0.084	0.208	0.194	0.144	0.158
0.2	0.021	0.128	0.110	0.086	0.093
0.3	0.011	0.057	0.059	0.038	0.032
0.4	0.000	0.042	0.042	0.024	0.021
0.5	0.000	0.039	0.018	0.023	0.021
0.6	0.000	0.036	0.016	0.020	0.020
0.7	0.000	0.000	0.000	0.000	0.000
0.8	0.000	0.000	0.000	0.000	0.000
0.9	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000
5	0.208	0.424	0.392	0.384	0.324
10	0.166	0.322	0.300	0.264	0.258
15	0.143	0.263	0.255	0.217	0.209
20	0.123	0.227	0.224	0.188	0.176
30	0.105	0.183	0.189	0.151	0.142
100	0.056	0.090	0.095	0.066	0.075

200	0.037	0.057	0.063	0.040	0.051
500	0.020	0.029	0.034	0.019	0.028
1000	0.011	0.016	0.020	0.010	0.016

Topic 251-300

	baseline	pf baseline	pfl1	rf baseline	rf1
num_ret	49290	49998	50000	49984	49274
num_rel	5524	5524	5524	5524	5524
num_rel_ret	201	364	465	284	394
map	0.016	0.042	0.055	0.050	0.052
gm_map	0.001	0.002	0.004	0.002	0.003
Rprec	0.034	0.061	0.080	0.069	0.074
bpref	0.040	0.069	0.085	0.071	0.084
recirank	0.191	0.336	0.391	0.542	0.450
0	0.209	0.357	0.414	0.546	0.459
0.1	0.060	0.142	0.134	0.138	0.151
0.2	0.028	0.070	0.090	0.078	0.087
0.3	0.025	0.049	0.059	0.068	0.077
0.4	0.002	0.015	0.049	0.038	0.038
0.5	0.002	0.015	0.043	0.036	0.036
0.6	0.000	0.007	0.036	0.000	0.004
0.7	0.000	0.002	0.016	0.000	0.004
0.8	0.000	0.002	0.003	0.000	0.003
0.9	0.000	0.002	0.003	0.000	0.003
1	0.000	0.002	0.003	0.000	0.003
5	0.068	0.220	0.248	0.232	0.216
10	0.058	0.184	0.194	0.146	0.148
15	0.056	0.149	0.145	0.119	0.119
20	0.055	0.132	0.123	0.099	0.102
30	0.051	0.105	0.101	0.075	0.083
100	0.030	0.051	0.049	0.034	0.042
200	0.017	0.030	0.031	0.022	0.026
500	0.008	0.014	0.016	0.011	0.013
1000	0.004	0.007	0.009	0.006	0.008

F11 100 documents, 20 terms

Topic 51-100

	baseline	pf baseline	pfl1	rf baseline	rf1
num_ret	49964	49590	49852	49882	49816
num_rel	16386	16386	16386	16386	16386
num_rel_ret	408	749	1195	853	956
map	0.010	0.029	0.032	0.025	0.021
gm_map	0.000	0.001	0.002	0.004	0.002
Rprec	0.024	0.050	0.061	0.050	0.047
bpref	0.027	0.051	0.071	0.056	0.061
recirank	0.212	0.494	0.439	0.754	0.443
0	0.245	0.516	0.476	0.760	0.457
0.1	0.028	0.082	0.122	0.089	0.066
0.2	0.014	0.044	0.050	0.007	0.020
0.3	0.000	0.021	0.010	0.000	0.002
0.4	0.000	0.017	0.000	0.000	0.000
0.5	0.000	0.000	0.000	0.000	0.000
0.6	0.000	0.000	0.000	0.000	0.000
0.7	0.000	0.000	0.000	0.000	0.000
0.8	0.000	0.000	0.000	0.000	0.000
0.9	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000
5	0.132	0.316	0.308	0.412	0.248
10	0.124	0.268	0.282	0.320	0.204
15	0.116	0.239	0.256	0.267	0.175
20	0.103	0.219	0.227	0.232	0.155
30	0.084	0.185	0.205	0.193	0.133
100	0.047	0.091	0.120	0.103	0.086
200	0.031	0.059	0.076	0.063	0.060
500	0.015	0.028	0.041	0.031	0.031
1000	0.008	0.015	0.024	0.017	0.019

Topic 101-150

	baseline	pf baseline	pfl1	rf baseline	rf1
num_ret	50000	49933	49996	49998	49978
num_rel	11645	11645	11645	11645	11645
num_rel_ret	486	1009	1374	1130	1095
map	0.012	0.036	0.035	0.048	0.025
gm_map	0.002	0.008	0.008	0.021	0.010

Rprec	0.038	0.074	0.084	0.088	0.064
bpref	0.040	0.076	0.103	0.096	0.083
recirank	0.340	0.570	0.413	0.920	0.585
0	0.369	0.595	0.466	0.928	0.600
0.1	0.035	0.126	0.133	0.175	0.084
0.2	0.012	0.033	0.046	0.027	0.015
0.3	0.000	0.012	0.019	0.001	0.000
0.4	0.000	0.000	0.007	0.000	0.000
0.5	0.000	0.000	0.000	0.000	0.000
0.6	0.000	0.000	0.000	0.000	0.000
0.7	0.000	0.000	0.000	0.000	0.000
0.8	0.000	0.000	0.000	0.000	0.000
0.9	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000
5	0.204	0.388	0.296	0.560	0.332
10	0.164	0.326	0.262	0.450	0.260
15	0.147	0.299	0.240	0.376	0.235
20	0.133	0.270	0.229	0.324	0.210
30	0.113	0.223	0.197	0.263	0.175
100	0.063	0.124	0.128	0.137	0.096
200	0.039	0.082	0.088	0.087	0.066
500	0.018	0.038	0.049	0.041	0.037
1000	0.010	0.020	0.028	0.023	0.022

Topic 151-200

	baseline	pf baseline	pfl1	rf baseline	rf1
num_ret	49995	49909	49989	49997	49981
num_rel	9805	9805	9805	9805	9805
num_rel_ret	491	769	1189	1061	1205
map	0.032	0.048	0.044	0.088	0.051
gm_map	0.003	0.007	0.011	0.032	0.014
Rprec	0.066	0.083	0.094	0.119	0.096
bpref	0.067	0.087	0.112	0.132	0.116
recirank	0.365	0.618	0.477	0.927	0.581
0	0.410	0.624	0.534	0.929	0.612
0.1	0.095	0.158	0.168	0.270	0.166
0.2	0.049	0.076	0.044	0.139	0.066
0.3	0.021	0.023	0.022	0.058	0.027
0.4	0.009	0.012	0.011	0.036	0.014
0.5	0.009	0.012	0.009	0.018	0.011
0.6	0.000	0.010	0.008	0.014	0.005
0.7	0.000	0.007	0.006	0.009	0.002

0.8	0.000	0.000	0.002	0.008	0.000
0.9	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000
5	0.240	0.392	0.328	0.664	0.400
10	0.220	0.316	0.300	0.508	0.340
15	0.189	0.269	0.280	0.416	0.293
20	0.177	0.241	0.259	0.369	0.266
30	0.157	0.201	0.214	0.297	0.219
100	0.084	0.107	0.118	0.148	0.117
200	0.047	0.066	0.078	0.086	0.077
500	0.020	0.030	0.041	0.040	0.042
1000	0.010	0.015	0.024	0.021	0.024

Topic 201-250

	baseline	pf baseline	pfl1	rf baseline	rf1
num_ret	50000	49968	49992	49989	49973
num_rel	6503	6503	6503	6503	6503
num_rel_ret	533	620	739	651	737
map	0.024	0.054	0.045	0.077	0.055
gm_map	0.002	0.004	0.005	0.023	0.010
Rprec	0.055	0.087	0.085	0.108	0.090
bpref	0.065	0.097	0.100	0.118	0.107
recirank	0.376	0.510	0.433	0.897	0.605
0	0.404	0.512	0.466	0.900	0.619
0.1	0.084	0.139	0.130	0.174	0.170
0.2	0.021	0.085	0.076	0.098	0.092
0.3	0.011	0.054	0.046	0.063	0.032
0.4	0.000	0.044	0.024	0.044	0.021
0.5	0.000	0.024	0.016	0.042	0.021
0.6	0.000	0.020	0.015	0.036	0.020
0.7	0.000	0.000	0.001	0.000	0.000
0.8	0.000	0.000	0.001	0.000	0.000
0.9	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000
5	0.208	0.276	0.264	0.484	0.320
10	0.166	0.220	0.210	0.364	0.248
15	0.143	0.180	0.175	0.297	0.201
20	0.123	0.157	0.156	0.249	0.169
30	0.105	0.133	0.133	0.201	0.138
100	0.056	0.074	0.071	0.089	0.072
200	0.037	0.046	0.046	0.053	0.047

500	0.020	0.022	0.025	0.025	0.025
1000	0.011	0.012	0.015	0.013	0.015

Topic 251-300

	baseline	pf baseline	pfl1	rf baseline	rf1
num_ret	49290	49959	49988	50000	49274
num_rel	5524	5524	5524	5524	5524
num_rel_ret	201	318	472	364	394
map	0.016	0.047	0.040	0.066	0.052
gm_map	0.001	0.002	0.003	0.008	0.003
Rprec	0.034	0.069	0.064	0.085	0.074
bpref	0.040	0.072	0.072	0.089	0.083
recirank	0.191	0.283	0.318	0.761	0.450
0	0.209	0.313	0.343	0.761	0.459
0.1	0.060	0.149	0.096	0.179	0.150
0.2	0.028	0.076	0.059	0.097	0.087
0.3	0.025	0.061	0.047	0.067	0.077
0.4	0.002	0.040	0.037	0.044	0.038
0.5	0.002	0.038	0.034	0.023	0.036
0.6	0.000	0.021	0.012	0.003	0.004
0.7	0.000	0.001	0.005	0.003	0.004
0.8	0.000	0.000	0.004	0.003	0.003
0.9	0.000	0.000	0.004	0.003	0.003
1	0.000	0.000	0.004	0.003	0.003
5	0.068	0.168	0.192	0.340	0.216
10	0.058	0.132	0.150	0.244	0.148
15	0.056	0.116	0.125	0.195	0.119
20	0.055	0.100	0.109	0.157	0.102
30	0.051	0.079	0.084	0.122	0.083
100	0.030	0.040	0.041	0.051	0.042
200	0.017	0.027	0.028	0.031	0.026
500	0.008	0.012	0.015	0.014	0.013
1000	0.004	0.006	0.009	0.007	0.008

F12 200 documents, 20 terms

Topic 51-100

	baseline	pf baseline	pfl1	rf baseline	rf1
num_ret	49964	49664	49930	49882	49821
num_rel	16386	16386	16386	16386	16386
num_rel_ret	408	493	964	1002	901
map	0.010	0.025	0.027	0.026	0.020
gm_map	0.000	0.000	0.001	0.005	0.002
Rprec	0.024	0.040	0.049	0.055	0.045
bpref	0.027	0.040	0.059	0.061	0.058
recirank	0.212	0.360	0.362	0.761	0.451
0	0.245	0.384	0.396	0.766	0.463
0.1	0.028	0.086	0.094	0.071	0.060
0.2	0.014	0.040	0.034	0.014	0.018
0.3	0.000	0.021	0.007	0.004	0.002
0.4	0.000	0.017	0.000	0.000	0.000
0.5	0.000	0.000	0.000	0.000	0.000
0.6	0.000	0.000	0.000	0.000	0.000
0.7	0.000	0.000	0.000	0.000	0.000
0.8	0.000	0.000	0.000	0.000	0.000
0.9	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000
5	0.132	0.224	0.252	0.448	0.244
10	0.124	0.200	0.220	0.336	0.194
15	0.116	0.184	0.196	0.289	0.165
20	0.103	0.166	0.174	0.251	0.145
30	0.084	0.141	0.159	0.205	0.127
100	0.047	0.072	0.096	0.112	0.081
200	0.031	0.043	0.061	0.073	0.056
500	0.015	0.019	0.033	0.037	0.029
1000	0.008	0.010	0.019	0.020	0.018

Topic 101-150

	baseline	pf baseline	pfl1	rf baseline	rf1
num_ret	50000	49932	49996	49998	49978
num_rel	11645	11645	11645	11645	11645
num_rel_ret	486	654	1215	1212	1090
map	0.012	0.023	0.027	0.048	0.025
gm_map	0.002	0.002	0.006	0.021	0.009
Rprec	0.038	0.053	0.075	0.087	0.064
bpref	0.040	0.053	0.093	0.098	0.082
recirank	0.340	0.434	0.373	0.900	0.584

0	0.369	0.450	0.420	0.911	0.602
0.1	0.035	0.071	0.110	0.144	0.084
0.2	0.012	0.031	0.032	0.032	0.015
0.3	0.000	0.011	0.007	0.000	0.000
0.4	0.000	0.000	0.000	0.000	0.000
0.5	0.000	0.000	0.000	0.000	0.000
0.6	0.000	0.000	0.000	0.000	0.000
0.7	0.000	0.000	0.000	0.000	0.000
0.8	0.000	0.000	0.000	0.000	0.000
0.9	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000
5	0.204	0.284	0.244	0.604	0.336
10	0.164	0.226	0.226	0.468	0.260
15	0.147	0.189	0.213	0.389	0.237
20	0.133	0.169	0.198	0.341	0.209
30	0.113	0.145	0.176	0.271	0.173
100	0.063	0.084	0.109	0.141	0.096
200	0.039	0.055	0.075	0.091	0.066
500	0.018	0.026	0.042	0.044	0.037
1000	0.010	0.013	0.024	0.024	0.022

Topic 151-200

	baseline	pf baseline	pfl1	rf baseline	rf1
num_ret	49995	49931	49993	49997	49981
num_rel	9805	9805	9805	9805	9805
num_rel_ret	491	527	980	1171	1194
map	0.032	0.032	0.031	0.092	0.051
gm_map	0.003	0.002	0.004	0.035	0.013
Rprec	0.066	0.063	0.072	0.124	0.096
bpref	0.067	0.064	0.090	0.139	0.115
recirank	0.365	0.422	0.353	0.912	0.582
0	0.410	0.440	0.391	0.915	0.611
0.1	0.095	0.112	0.103	0.278	0.165
0.2	0.049	0.044	0.033	0.137	0.066
0.3	0.021	0.018	0.015	0.061	0.027
0.4	0.009	0.010	0.008	0.040	0.014
0.5	0.009	0.010	0.007	0.020	0.011
0.6	0.000	0.007	0.006	0.014	0.005
0.7	0.000	0.005	0.004	0.011	0.002
0.8	0.000	0.002	0.003	0.007	0.000
0.9	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000
5	0.240	0.256	0.224	0.652	0.396
10	0.220	0.212	0.226	0.504	0.340

15	0.189	0.184	0.208	0.428	0.292
20	0.177	0.164	0.187	0.375	0.264
30	0.157	0.134	0.158	0.310	0.219
100	0.084	0.074	0.087	0.152	0.117
200	0.047	0.047	0.062	0.090	0.076
500	0.020	0.021	0.034	0.044	0.041
1000	0.010	0.011	0.020	0.023	0.024

Topic 201-250

	baseline	pf baseline	pfl1	rf baseline	rf1
num_ret	50000	49978	49981	49997	49975
num_rel	6503	6503	6503	6503	6503
num_rel_ret	533	315	562	724	672
map	0.024	0.038	0.037	0.091	0.053
gm_map	0.002	0.001	0.003	0.035	0.008
Rprec	0.055	0.063	0.071	0.124	0.086
bpref	0.065	0.067	0.088	0.131	0.102
recirank	0.376	0.380	0.366	0.930	0.565
0	0.404	0.392	0.398	0.944	0.578
0.1	0.084	0.123	0.118	0.224	0.156
0.2	0.021	0.063	0.062	0.130	0.089
0.3	0.011	0.033	0.040	0.075	0.032
0.4	0.000	0.020	0.016	0.058	0.021
0.5	0.000	0.020	0.015	0.042	0.021
0.6	0.000	0.020	0.015	0.036	0.020
0.7	0.000	0.000	0.001	0.016	0.000
0.8	0.000	0.000	0.001	0.000	0.000
0.9	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000
5	0.208	0.236	0.224	0.560	0.304
10	0.166	0.162	0.178	0.396	0.230
15	0.143	0.125	0.143	0.328	0.191
20	0.123	0.108	0.124	0.276	0.162
30	0.105	0.085	0.103	0.218	0.131
100	0.056	0.043	0.059	0.100	0.066
200	0.037	0.026	0.038	0.059	0.044
500	0.020	0.012	0.019	0.027	0.023
1000	0.011	0.006	0.011	0.015	0.013

Topic 251-300

	baseline	pf baseline	pfl1	rf baseline	rf1
num_ret	49290	49979	49988	50000	49274

num_rel	5524	5524	5524	5524	5524
num_rel_ret	201	211	413	358	394
map	0.016	0.036	0.041	0.070	0.050
gm_map	0.001	0.001	0.001	0.013	0.003
Rprec	0.034	0.053	0.061	0.087	0.072
bpref	0.040	0.054	0.068	0.093	0.083
recirank	0.191	0.266	0.278	0.830	0.450
0	0.209	0.274	0.303	0.836	0.459
0.1	0.060	0.082	0.100	0.184	0.140
0.2	0.028	0.047	0.053	0.111	0.087
0.3	0.025	0.046	0.042	0.080	0.077
0.4	0.002	0.037	0.040	0.044	0.038
0.5	0.002	0.037	0.040	0.023	0.036
0.6	0.000	0.024	0.026	0.000	0.004
0.7	0.000	0.004	0.006	0.000	0.004
0.8	0.000	0.004	0.005	0.000	0.003
0.9	0.000	0.004	0.005	0.000	0.003
1	0.000	0.004	0.005	0.000	0.003
5	0.068	0.144	0.144	0.412	0.200
10	0.058	0.116	0.128	0.284	0.140
15	0.056	0.093	0.109	0.216	0.109
20	0.055	0.078	0.093	0.171	0.095
30	0.051	0.063	0.079	0.127	0.077
100	0.030	0.027	0.039	0.053	0.041
200	0.017	0.019	0.026	0.031	0.026
500	0.008	0.008	0.014	0.014	0.013
1000	0.004	0.004	0.008	0.007	0.008

APPENDIX G – Single value results

G1 – Recall results (varying number of terms parameter)

Topic	#terms	Baseline	Pf	Pf with ont	% diff	Rfb	Rfb with ont	% diff
topic 51-100	5	0.025	0.045	0.070	55.6	0.031	0.059	90.3
	10	0.025	0.054	0.073	35.2	0.033	0.059	78.8
	15	0.025	0.058	0.074	27.6	0.034	0.059	73.5
	20	0.025	0.063	0.073	15.54	0.073	0.059	-19.55
	100	0.025	0.069	0.074	7.2	0.042	0.059	40.5
	200	0.025	0.030	0.074	146.7	0.037	0.059	59.5
	Top3	0.025	0.039	0.073	87.3	0.029	0.059	102.7
topic 101-150	5	0.042	0.085	0.120	41.2	0.066	0.096	45.5
	10	0.042	0.106	0.122	15.1	0.073	0.096	31.5
	15	0.042	0.113	0.122	8.0	0.071	0.096	35.2
	20	0.042	0.115	0.124	7.67	0.074	0.096	29.13
	100	0.042	0.129	0.122	-5.4	0.079	0.096	21.5
	200	0.042	0.049	0.122	149.0	0.054	0.096	77.8
	Top3	0.042	0.078	0.124	59.2	0.058	0.096	65.4
topic 151-200	5	0.050	0.088	0.134	52.3	0.069	0.123	78.3
	10	0.050	0.091	0.140	53.8	0.078	0.123	57.7
	15	0.050	0.102	0.142	39.2	0.080	0.123	53.8
	20	0.050	0.109	0.145	33.12	0.084	0.123	45.47
	100	0.050	0.117	0.147	25.6	0.088	0.123	39.8
	200	0.050	0.058	0.147	153.4	0.066	0.123	86.4
	Top3	0.050	0.086	0.145	68.8	0.066	0.123	85.9
topic 201-250	5	0.082	0.101	0.141	39.6	0.072	0.125	73.6
	10	0.082	0.115	0.145	26.1	0.074	0.125	68.9
	15	0.082	0.124	0.149	20.2	0.072	0.125	73.6
	20	0.082	0.113	0.151	33.97	0.075	0.125	65.92
	100	0.082	0.125	0.153	22.4	0.087	0.125	43.7
	200	0.082	0.090	0.153	70.0	0.094	0.125	33.0
	Top3	0.082	0.099	0.151	52.5	0.074	0.125	68.9
topic 251-300	5	0.036	0.059	0.082	39.0	0.046	0.071	54.3
	10	0.036	0.058	0.088	51.7	0.046	0.071	54.3
	15	0.036	0.062	0.085	37.1	0.049	0.071	44.9
	20	0.036	0.071	0.087	23.33	0.050	0.071	41.73
	100	0.036	0.061	0.093	52.5	0.065	0.071	9.2
	200	0.036	0.045	0.093	106.7	0.052	0.071	36.5
	Top3	0.036	0.055	0.087	58.3	0.043	0.071	65.9

G2 – Recall results (varying number of docs parameter)

Topic	#docs	Baseline	Pf	Pf with ont	% diff	Rfb	Rfb with ont	% diff
topic 51-100	5	0.025	0.055	0.069	24.4	0.034	0.059	75.1
	10	0.025	0.065	0.089	36.9	0.035	0.080	128.6
	15	0.025	0.063	0.072	14.3	0.033	0.059	78.8
	20	0.025	0.063	0.073	15.54	0.073	0.059	-19.5
	100	0.025	0.046	0.073	58.7	0.052	0.058	11.5
	200	0.025	0.030	0.059	96.7	0.061	0.055	-9.8
	Top3	0.025	0.039	0.073	87.3	0.029	0.059	102.7
topic 101-150	5	0.042	0.101	0.115	13.7	0.057	0.096	67.6
	10	0.042	0.111	0.127	14.4	0.074	0.096	29.7
	15	0.042	0.116	0.125	7.8	0.074	0.096	29.7
	20	0.042	0.115	0.124	7.67	0.074	0.096	29.1
	100	0.042	0.087	0.118	35.6	0.097	0.094	-3.1
	200	0.042	0.056	0.104	85.7	0.104	0.094	-9.6
	Top3	0.042	0.078	0.124	59.2	0.058	0.096	65.4
topic 151-200	5	0.050	0.092	0.135	46.2	0.061	0.123	101.8
	10	0.050	0.105	0.100	-4.8	0.075	0.123	64.0
	15	0.050	0.112	0.148	32.1	0.078	0.123	57.7
	20	0.050	0.109	0.145	33.12	0.084	0.123	45.5
	100	0.050	0.078	0.121	55.1	0.108	0.123	13.9
	200	0.050	0.054	0.100	85.2	0.119	0.122	2.5
	Top3	0.050	0.086	0.145	68.8	0.066	0.123	85.9
topic 201-250	5	0.082	0.120	0.135	12.4	0.079	0.125	57.6
	10	0.082	0.128	0.143	11.7	0.084	0.124	47.6
	15	0.082	0.121	0.155	28.1	0.077	0.125	62.3
	20	0.082	0.113	0.151	33.97	0.075	0.125	65.9
	100	0.082	0.095	0.114	20.0	0.100	0.113	13.0
	200	0.082	0.048	0.086	79.2	0.111	0.103	-7.2
	Top3	0.082	0.099	0.151	52.5	0.074	0.125	68.9
topic 251-300	5	0.036	0.061	0.075	23.0	0.049	0.071	44.9
	10	0.036	0.070	0.077	0.1	0.044	0.071	61.4
	15	0.036	0.066	0.084	27.3	0.051	0.071	39.2
	20	0.036	0.071	0.087	23.33	0.050	0.071	41.7
	100	0.036	0.058	0.085	47.2	0.066	0.071	7.6
	200	0.036	0.038	0.075	97.4	0.065	0.071	9.2
	Top3	0.036	0.055	0.087	58.3	0.043	0.071	65.9

G3 – MAP results (varying number of terms parameter)

Topic	#terms	Baseline	Pf	Pf with ont	% diff	Rfb	Rfb with ont	% diff
topic 51-100	5	0.010	0.027	0.033	20.7	0.018	0.021	13.2
	10	0.010	0.032	0.032	2.5	0.017	0.021	20.5
	15	0.010	0.030	0.032	6.9	0.018	0.021	16.4
	20	0.010	0.033	0.033	-1.5	0.018	0.021	16.4
	100	0.010	0.037	0.034	-9.6	0.020	0.021	4.0
	200	0.010	0.017	0.034	94.3	0.021	0.021	-1.4
	Top3	0.010	0.024	0.033	34.3	0.015	0.021	40.1
topic 101-150	5	0.012	0.033	0.036	8.2	0.029	0.025	-15.0
	10	0.012	0.040	0.035	-11.3	0.034	0.025	-26.5
	15	0.012	0.042	0.037	-11.7	0.033	0.025	-25.4
	20	0.012	0.042	0.039	-7.6	0.033	0.025	-24.3
	100	0.012	0.042	0.039	-7.1	0.036	0.025	-30.4
	200	0.012	0.021	0.039	89.4	0.028	0.025	-10.8
	Top3	0.012	0.032	0.039	21.1	0.026	0.025	-5.3
topic 151-200	5	0.032	0.063	0.056	-10.6	0.055	0.051	-6.9
	10	0.032	0.068	0.059	-13.6	0.065	0.051	-21.0
	15	0.032	0.074	0.061	-17.2	0.066	0.051	-22.5
	20	0.032	0.076	0.064	-16.4	0.067	0.051	-23.2
	100	0.032	0.076	0.066	-13.6	0.071	0.051	-27.4
	200	0.032	0.046	0.066	43.6	0.051	0.051	0.8
	Top3	0.032	0.057	0.064	10.8	0.053	0.051	-2.7
topic 201-250	5	0.024	0.066	0.056	-14.8	0.052	0.055	5.2
	10	0.024	0.071	0.062	-12.6	0.061	0.055	-10.5
	15	0.024	0.074	0.067	-9.7	0.061	0.055	-9.9
	20	0.024	0.076	0.068	-9.9	0.064	0.055	-14.5
	100	0.024	0.056	0.067	19.6	0.070	0.055	-22.2
	200	0.024	0.052	0.067	28.7	0.062	0.055	-11.5
	Top3	0.024	0.059	0.068	15.4	0.052	0.055	4.8
topic 251-300	5	0.016	0.053	0.053	0.6	0.051	0.052	1.6
	10	0.016	0.047	0.057	20.7	0.053	0.052	-1.9
	15	0.016	0.044	0.059	33.0	0.056	0.052	-6.6
	20	0.016	0.043	0.056	29.1	0.057	0.052	-9.2
	100	0.016	0.033	0.051	55.0	0.059	0.052	-11.1
	200	0.016	0.051	0.051	0.4	0.052	0.052	0.6
	Top3	0.016	0.053	0.056	6.1	0.045	0.052	15.3

G4 – MAP results (varying number of docs parameter)

Topic	#docs	Baseline	Pf	Pf with ont	% diff	Rfb	Rfb with ont	% diff
topic 51-100	5	0.010	0.037	0.029	-21.7	0.024	0.021	-12.7
	10	0.010	0.039	0.034	-12.8	0.018	0.024	32.2
	15	0.000	0.002	0.002	0.0	0.001	0.002	53.8
	20	0.010	0.033	0.033	-1.5	0.018	0.021	16.4
	100	0.010	0.029	0.032	9.2	0.025	0.021	-16.3
	200	0.010	0.025	0.027	8.0	0.026	0.020	-25.0
	Top3	0.010	0.024	0.033	34.3	0.015	0.021	40.1
topic 101-150	5	0.012	0.037	0.032	-13.3	0.027	0.025	-7.4
	10	0.012	0.040	0.039	-4.5	0.032	0.025	-22.4
	15	0.012	0.042	0.038	-8.6	0.032	0.025	-23.1
	20	0.012	0.042	0.039	-7.6	0.033	0.025	-24.3
	100	0.012	0.036	0.035	-3.0	0.048	0.025	-48.8
	200	0.012	0.023	0.027	17.2	0.048	0.025	-48.2
	Top3	0.012	0.032	0.039	21.1	0.026	0.025	-5.3
topic 151-200	5	0.032	0.070	0.062	-12.1	0.049	0.051	4.0
	10	0.032	0.075	0.031	-58.7	0.055	0.051	-6.0
	15	0.032	0.076	0.065	-14.6	0.060	0.051	-14.5
	20	0.032	0.076	0.064	-16.4	0.067	0.051	-23.2
	100	0.032	0.048	0.044	-8.1	0.088	0.051	-41.9
	200	0.032	0.032	0.031	-3.7	0.092	0.051	-44.3
	Top3	0.032	0.057	0.064	10.8	0.053	0.051	-2.7
topic 201-250	5	0.024	0.066	0.061	-7.8	0.057	0.055	-3.5
	10	0.024	0.077	0.068	-12.5	0.064	0.055	-14.9
	15	0.024	0.074	0.068	-8.8	0.058	0.055	-5.7
	20	0.024	0.076	0.068	-9.9	0.064	0.055	-14.5
	100	0.024	0.054	0.045	-16.0	0.077	0.055	-29.2
	200	0.024	0.038	0.037	-3.6	0.091	0.053	-42.0
	Top3	0.024	0.059	0.068	15.4	0.052	0.055	4.8
topic 251-300	5	0.016	0.048	0.059	21.8	0.057	0.052	-9.2
	10	0.016	0.051	0.056	9.2	0.060	0.052	-12.8
	15	0.016	0.042	0.055	31.1	0.050	0.052	3.4
	20	0.016	0.043	0.056	29.1	0.057	0.052	-9.2
	100	0.016	0.047	0.040	-13.1	0.066	0.052	-21.0
	200	0.016	0.036	0.041	15.7	0.070	0.050	-28.0
	Top3	0.016	0.053	0.056	6.1	0.045	0.052	15.3

G5 –BPref results (varying number of terms parameter)

Topic	#terms	Baseline	Pf	Pf with ont	% diff	Rfb	Rfb with ont	% diff
topic 51-100	5	0.027	0.049	0.070	42.6	0.037	0.062	66.9
	10	0.027	0.058	0.072	23.2	0.037	0.062	66.9
	15	0.027	0.059	0.074	25.3	0.039	0.062	57.5
	20	0.027	0.065	0.073	12.3	0.040	0.062	54.8
	100	0.027	0.075	0.076	1.2	0.047	0.062	30.5
	200	0.027	0.034	0.076	120.3	0.043	0.062	42.6
	Top3	0.027	0.044	0.073	67.7	0.033	0.062	86.7
topic 101-150	5	0.040	0.072	0.101	40.4	0.062	0.083	34.6
	10	0.040	0.087	0.101	16.3	0.070	0.083	18.3
	15	0.040	0.092	0.102	10.5	0.068	0.083	21.4
	20	0.040	0.094	0.105	11.8	0.070	0.083	19.1
	100	0.040	0.103	0.105	1.8	0.074	0.083	11.7
	200	0.040	0.047	0.105	120.9	0.056	0.083	48.0
	Top3	0.040	0.069	0.105	51.6	0.056	0.083	47.0
topic 151-200	5	0.067	0.104	0.124	19.1	0.089	0.116	31.1
	10	0.067	0.111	0.129	16.2	0.102	0.116	13.7
	15	0.067	0.118	0.133	12.2	0.105	0.116	11.3
	20	0.067	0.122	0.135	10.3	0.108	0.116	7.4
	100	0.067	0.127	0.136	6.8	0.108	0.116	7.5
	200	0.067	0.079	0.136	70.7	0.088	0.116	32.8
	Top3	0.067	0.099	0.135	36.6	0.084	0.116	39.1
topic 201-250	5	0.065	0.108	0.117	8.7	0.087	0.111	27.0
	10	0.065	0.115	0.125	8.1	0.096	0.111	15.4
	15	0.065	0.122	0.131	7.5	0.094	0.111	18.6
	20	0.065	0.117	0.133	13.3	0.098	0.111	13.0
	100	0.065	0.106	0.132	24.4	0.108	0.111	2.8
	200	0.065	0.093	0.132	42.5	0.105	0.111	5.4
	Top3	0.065	0.101	0.133	32.1	0.087	0.111	27.7
topic 251-300	5	0.040	0.077	0.085	10.1	0.074	0.084	12.5
	10	0.040	0.074	0.093	24.8	0.073	0.084	14.5
	15	0.040	0.071	0.092	28.3	0.077	0.084	7.9
	20	0.040	0.069	0.088	26.6	0.080	0.084	4.0
	100	0.040	0.057	0.084	49.2	0.086	0.084	-3.2
	200	0.040	0.073	0.084	15.8	0.078	0.084	7.5
	Top3	0.040	0.075	0.088	17.4	0.069	0.084	21.0

G6 –BPref results (varying number of docs parameter)

Topic	#docs	Baseline	Pf	Pf with ont	% diff	Rfb	Rfb with ont	% diff
topic 51-100	5	0.027	0.064	0.071	10.9	0.042	0.062	47.7
	10	0.027	0.073	0.086	18.9	0.039	0.080	103.0
	15	0.212	0.514	0.456	-11.4	0.591	0.443	-25.0
	20	0.027	0.065	0.073	12.3	0.040	0.062	54.8
	100	0.027	0.051	0.071	38.6	0.056	0.061	10.3
	200	0.027	0.040	0.059	45.2	0.061	0.058	-5.4
	Top3	0.027	0.044	0.073	67.7	0.033	0.062	86.7
topic 101-150	5	0.040	0.082	0.097	18.2	0.059	0.083	41.2
	10	0.040	0.090	0.105	15.7	0.071	0.083	16.0
	15	0.040	0.093	0.106	13.4	0.067	0.083	23.8
	20	0.040	0.094	0.105	11.8	0.070	0.083	19.1
	100	0.040	0.076	0.103	35.8	0.096	0.083	-14.2
	200	0.040	0.053	0.093	75.7	0.098	0.082	-16.6
	Top3	0.040	0.069	0.105	51.6	0.056	0.083	47.0
topic 151-200	5	0.067	0.114	0.128	12.8	0.081	0.116	43.4
	10	0.067	0.123	0.090	-26.9	0.092	0.116	26.4
	15	0.067	0.125	0.134	7.7	0.097	0.116	20.1
	20	0.067	0.122	0.135	10.3	0.108	0.116	7.4
	100	0.067	0.087	0.112	28.1	0.132	0.116	-11.9
	200	0.067	0.064	0.090	39.6	0.139	0.115	-17.1
	Top3	0.067	0.099	0.135	36.6	0.084	0.116	39.1
topic 201-250	5	0.065	0.110	0.117	5.5	0.094	0.110	16.6
	10	0.065	0.122	0.128	4.7	0.103	0.111	7.4
	15	0.065	0.122	0.134	9.8	0.093	0.111	19.1
	20	0.065	0.117	0.133	13.3	0.098	0.111	13.0
	100	0.065	0.097	0.100	3.2	0.118	0.107	-9.0
	200	0.065	0.067	0.088	31.6	0.131	0.102	-22.6
	Top3	0.065	0.101	0.133	32.1	0.087	0.111	27.7
topic 251-300	5	0.040	0.075	0.091	20.9	0.079	0.084	5.8
	10	0.040	0.080	0.089	12.1	0.078	0.084	6.6
	15	0.040	0.069	0.085	24.3	0.071	0.084	17.1
	20	0.040	0.069	0.088	26.6	0.080	0.084	4.0
	100	0.040	0.072	0.072	-0.4	0.089	0.083	-6.6
	200	0.040	0.054	0.068	25.7	0.093	0.083	-11.5
	Top3	0.040	0.075	0.088	17.4	0.069	0.084	21.0