

Hierarchical Presentation of Expansion Terms

Hideo Joho, Claire Coverson, Mark Sanderson, and Micheline Beaulieu
Department of Information Studies, University of Sheffield
Western Bank, Sheffield, S10 2TN, UK
{h.joho, m.sanderson, m.beaulieu}@sheffield.ac.uk

ABSTRACT

Different presentations of candidate expansion terms have not been fully explored in interactive query expansion (IQE). Most existing systems that offer an IQE facility use a list form of presentation. This paper examines an hierarchical presentation of the expansion terms which are automatically generated from a set of retrieved documents, organised in a *general to specific* manner, and visualised by cascade menus. To evaluate the effectiveness of the presentation, a user test was carried out to compare the hierarchical form with the conventional list form. This shows that users of the hierarchy can complete the expansion task in less time and with fewer terms over those using the lists. Relations between initial query terms and selected expansion terms were also investigated.

Keywords

Information retrieval, interactive query expansion, concept hierarchies

1. INTRODUCTION

The increasing interest in providing online information available via the Internet has heightened the need for information retrieval (IR) systems that enable users to access heterogeneous resources that meet their information needs. The means of interaction between users and such an IR system to achieve a meaningful search is of particular interest and complexity. The potential benefit of interactive query expansion (IQE) has generated wide interest in making IR systems more adaptive as opposed to automatic query expansion (AQE), where users preferences are generally ignored.

An early form of interaction was through relevance feedback [18], where users judge the relevance of retrieved documents as indicative information about their interest given to the system. Harman [9] examined a more interactive approach by presenting a list of candidate expansion terms to users. The users then selected the terms of interest from the list to add to the initial queries. This interactive approach has been a standard way of presenting IQE and has been adopted by most researchers [1, 17, 7], regardless of the techniques used for extracting, or ranking the candidate terms.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SAC 2002, Madrid, Spain

Copyright 2002 ACM 1-58113-445-2/02/03 ...\$5.00.

The study of the user interface is also considered as an important aspect of IQE. Beaulieu [2] evaluated three different interface design based on the same IR system. In her study, a character-based design and two different graphical user interfaces were compared. This experiment highlights the importance of the close relationship between system functionality and interface design in supporting query expansion. Koenemann and Belkin [12] also showed that giving the user more control over how terms were added to the query could increase retrieval effectiveness and user satisfaction. A third study [13] compared the potential retrieved performance of IQE by simulating experienced and naive users.

As can be seen, researchers have addressed several aspects concerning the interaction between users and the IR systems in the context of IQE. However, little attention has been paid to the effectiveness of different presentation of candidate expansion terms. In fact, most experimental designs discussed so far merely use a form of list for the presentation.

The motivation of exploring a more effective presentation of expansion terms may resemble the studies of visualising search results. The existing IR systems including search engines on the Web present the search results as a list of document titles with some additional information such as URLs, size of documents, or a short summary. Users therefore have to spend some time scanning the list to locate relevant documents, and they are often reluctant to examine the documents ranked further below. Visualisation of the search results aims to help users to locate the documents of their interest regardless of ranking.

A number of techniques have been developed in this area [11, 24, 4], and Hearst [10] suggested giving users an overview as a means of navigation. She stated that 'an overview can help users get started, directing them into general neighbourhoods, after which they can navigate using more detailed descriptions' (p. 268). One of the types of overview is 'category hierarchies associated with the documents of a collection'. Subject directories exemplified by Yahoo!¹ are such an example where topical concepts are organised in a hierarchical manner. Similarly, this type of navigation may be ideal for IQE.

In this paper, we examine the effectiveness of different presentation of expansion terms. The comparison is made between a hierarchical organisation and lists.

2. ORGANISATION AND VISUALISATION OF EXPANSION TERMS

In the previous section, it has been argued that existing IR systems have not paid much attention to *how to present expansion terms* [6], and the form of presentation is dominated by lists. Fur-

¹<http://www.yahoo.com>

thermore, it is anticipated that a hierarchical presentation will provide users with an overview of candidate expansion terms, and thus, can be a promising alternative form.

Although manually constructed thesauri such as MeSH or IN-SPEC have been integrated into systems to assist in query formulation [20, 15], those resources are inevitably limited in the range of vocabulary and are not necessarily applicable to all domains.

A more promising technique that meets our aim has been introduced by Sanderson and Croft [19]. *Subsumption hierarchies* are designed to organise terms in a manner which is similar to existing manually constructed thesauri, or subject hierarchies in Yahoo.

Unlike other co-occurrence based techniques that measures *similarity* between terms [16], Sanderson and Croft use the co-occurrence information to identify a term that subsumes other terms. More specifically, a term, x , is said to subsume another term, y , if the documents which y occurs in are a subset of the documents in which x occurs. Given that a more frequent term tends to be more general [22], subsumption hierarchies organise terms in a 'general to specific' manner.

Although this technique was originally introduced as a means of automatic generation of concept hierarchies from a set of retrieved documents, applying it to IQE may also be valuable. For a comprehensive description of subsumption hierarchies, see [19].

The next section will describe a user test that examines the effectiveness of different presentation of expansion terms followed by the results and discussion.

3. USER TEST

The user test was carried out to examine any effects derived from different methods of presenting candidate expansion terms. In this study, a hierarchical presentation and conventional list presentation were compared.

3.1 Participants

A total of 24 subjects were recruited for the user test. The majority of the subjects (20) were students of the Department of Information Studies, University of Sheffield, and the rest were other members of the University. They consisted of 10 females and 14 males. The age of the subjects ranges from 22 to 35 with an average of 28.

3.2 Topics

Topics for the user test were taken from the TREC test collection (Topics 300-350) in the Sixth Text Retrieval Conference [23]. Among the topics, the subsumption hierarchy could not be created for topic 312, 330 and 348 as no relevance documents were retrieved in response to the queries, and therefore these topics were removed. It was also decided to remove topics which produced very small concept hierarchies. As a result, topics 316 and 327 which generated hierarchies of less than three levels or which contains less than 30 expansion terms were removed. A total of 45 topics remained and were used in the experiment.

3.3 Experimental system

INQUERY [3] was used as the IR system in this study. Candidate expansion terms were first extracted from the top 500 documents retrieved by INQUERY, in response to a query compiled from terms in the title of each topic description, then organised by the subsumption process, and finally visualised by the cascade menus. Lists were also generated using the identical set of terms included in the menus. The lists were ordered alphabetically as this was considered an arbitrary order for the presentation of the terms.

An example of candidate expansion terms for topic 302 and terms selected by a test subject is presented in the Appendix.

3.4 Procedure

Subjects were first given an explanation of the reason behind the experiment. The subjects were told that a tool that attempts to generate a summary of the retrieved documents was in the process of being developed, and that query terms extracted from the documents retrieved would be shown on screen as an indication of what the retrieved document set is about.

The subjects were invited to consider the following scenario. They had just submitted a query to a retrieval system, the system had responded by showing a set of possible terms that could be added to the query in order to improve the search results. Their task would be to select terms they deemed appropriate to expand the query. Subject were then given a demonstration of the working system using a training query and topic description (topic 327) to illustrate the procedure.

After the training session, they were asked to carry out the actual experimental expansion task with nine topics. In order to save time, all menus and lists for the 45 topics were generated in advance of the user test. Subjects were alternately assigned to a control or experimental group. The experimental group was presented with the interface containing the menus (the Menu group) and the control group was presented with the lists (the List group).

Following the completion of the expansion tasks by all the participants, selected expansion terms were added to the initial query for each topic, and the search was re-run with the expanded queries. A comprehensive description of the user test can be found in [5].

4. RESULTS AND DISCUSSION

This section describes experimental results of the user test based on the standard precision and recall measures, the number of expansion terms selected, the time taken to complete the task, and link type analysis of expansion terms.

4.1 Precision and Recall

A recall-precision graph (Figure 1) was plotted using the top 1000 retrieved documents retrieved in response to the unexpanded queries, expanded queries by lists, and by menus. The measurements of recall and precision were based on the TREC relevance assessments [23]. The graph shows the unexpanded queries retrieve documents at a higher precision for lower recall while the

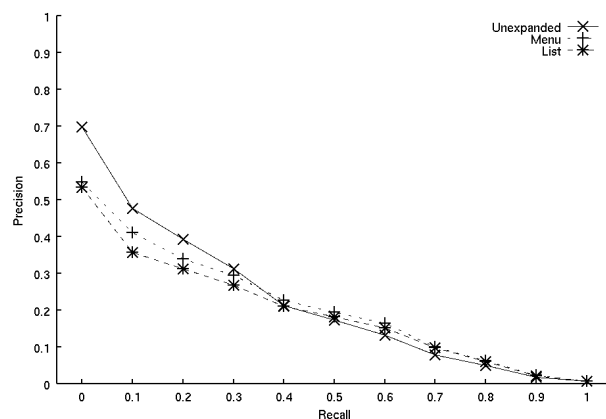


Figure 1: Precision-Recall graph

expanded queries produce a higher precision for higher recall after the 0.4 point. The graph also shows that the queries expanded by the menus constantly perform slightly better than the lists. No statistical significance was, however, found between them.

Table 1 shows precision at the top one, five, ten and twenty retrieved documents. As indicated in the precision-recall graph, the unexpanded queries retrieve more relevant document at all the levels shown than the expanded queries. Similarly, precision of the expanded queries by the Menus are higher than the List except at level one.

Table 1: Precision at one, five, ten and twenty

	Unexp.	Expanded		Residual	
		List	Menu	List	Menu
PREC-AT-1	0.6117	0.4078	0.4078	0.2816	0.2330
PREC-AT-5	0.4252	0.3612	0.3767	0.2447	0.2485
PREC-AT-10	0.3903	0.3097	0.3408	0.2223	0.2223
PREC-AT-20	0.3393	0.2937	0.2966	0.2155	0.2044

The residual precision (Fifth and Sixth columns in Table 1) is calculated by a set of documents retrieved by the expanded queries but removing the relevant documents that the corresponding initial query already retrieved in the top 20. The lower ranked documents were then promoted. This simulates the performance of the expanded queries where a user is supposed to identify all the relevant documents from initial search results in the top 20. This also reveals the extent to which new relevant documents were retrieved by the expanded queries.

The data shows that both expanded queries cause new relevant documents to be retrieved in the top 20. However, there is little difference between the lists and menus in terms of the standard retrieval effectiveness shown above.

4.2 Number of expansion terms and time to complete task

The standard retrieval effectiveness was shown in the previous section. However, the standard method may not be fully adequate for evaluating the line of research presented here [21]. This section presents data concerning efficiency to complete the expansion task. Number of expansion terms selected by subjects and time to complete the whole expansion task are shown in Table 2. They can indicate the effectiveness of term selection tasks frequently carried out in IQE.

Table 2: Efficiency of expansion task

	List	Menu
	No. of selected expansion terms	16.87
Time to complete expansion task (sec.)	203.28	168.40

As can be seen, the subjects in the Menu group completed the task with over four terms fewer than the List group on average, and this significantly shortened the time to complete the task. These two performance figures are statistically significant ($p < 0.001$).

4.3 User perceptions

In order to gain an indication of the ability of the interfaces to provide an overview of the retrieved documents, subjects in both groups were asked if, after exploring the expansion terms, they felt that they had a better idea of the contents of the retrieved documents. 80% of the subjects in the Menu group stated that they had

a better idea of the contents. This is significantly higher than the List group where only 42% felt that exploring the list gave them a better idea of the contents. One subject in the Menu group also argued that the hierarchies gave them an idea of whether or not they were going to find a decent answer to their query. This seem to support our approach to provide an adequate summarisation of documents retrieved in response to a query and that the subsumption hierarchies are meaningful.

4.4 Link types

The semantic and/or topical relations between initial query terms and expansion terms have been of interest in IQE since this indicates an aspect of user's searching behaviour [7, 8].

The relations used in our link type test were based on the ones defined by WordNet [14], but an additional relation, *conceptually related* (or contextually related), was also devised. A conceptually related term was defined more broadly than the related term (RT) in an existing thesaurus. An example of such a relation is *tooth* and *dentist*. This type of relation is not defined in WordNet but can be useful in the context of query expansion [20].

Table 3: Link types between initial query terms and expansion terms selected by the Menu group (N=131)

Relation	Portion (%)
Hyponym	8
Meronym	2
Hypernym	9
Holonym	0
Coordinate Sister	2
Synonym	13
Antonym	0
Conceptually Related	65
Other/Don't know	1
Total	100

Table 3 shows the link types between initial query terms and expansion terms selected by the Menu group. This reveals that far more than half of the expansion terms selected by the subjects were terms conceptually related to the initial query terms. Synonyms were the second largest portion among the relations. A total of 10% of the selected expansion terms were of aspects of corresponding initial query terms (Hyponym and Meronym). A similar portion was also found in the parental relation (Hypernym) with the initial query terms.

Arguably, this result can be seen as an echo of the findings from Efthimiadis's experiment [7] where 44% of selected expansion terms were not considered as a type of thesaurus-like relations but as *new ideas* by the users. A similar tendency was found in our test. The reason for this tendency is not clear [7], but it is speculated that many terms which users choose as further descriptions of their information needs can be more complex (or flexible) relations to initial queries than those in a thesaurus.

This also highlights the advantage of the subsumption hierarchies as a means of hierarchical presentation of candidate expansion terms, as opposed to a manually constructed thesaurus which provides well organised but limited range of relations.

5. CONCLUSION

This paper examined the effectiveness of the hierarchical presentation of candidate expansion terms in comparison of a conventional list form. Although no significant difference in terms of

precision-recall between them was found, the users in the hierarchical presentation group completed the expansion task in significantly shorter time than users in the list presentation group. From these results, it is concluded that different presentations of expansion terms do effect the expansion task, and a more structured presentation can improve the selection of expansion terms. This also suggests that more attention on presenting expansion terms should be made in the research of IQE.

6. ACKNOWLEDGMENTS

Tim Gollins helped in postprocessing the experimental data. This work was carried out as a part of the CiQuest project, which is funded by the Library and Information Commission (now re:source, the Council for Museums, Archives and Libraries). Any opinions, findings, and conclusions described here are the authors and do not necessarily reflect those of the sponsor.

7. REFERENCES

- [1] J. Allan, L. Ballesteros, J. Callan, W. Croft, and Z. Lu. Recent experiments with InQuery. In D. K. Harman, editor, *Proceedings of the Fourth Text REtrieval Conference (TREC-4)*, pages 49–72, Gaithersburg, MD, 1996. NIST.
- [2] M. Beaulieu. Experiments on interfaces to support query expansion. *Journal of Documentation*, 53(1):8–19, 1997.
- [3] J. P. Callan, B. W. Croft, and S. M. Harding. The INQUERY retrieval system. In *Proceedings of the Third International Conference on Database and Expert Systems Applications*, pages 78–83, Valencia, Spain, 1992. Springer-Verlag.
- [4] H. Chen and S. Dumais. Bringing order to the Web: Automatically categorizing search results. In *Proceedings of the CHI 2000 Conference on Human factors in computing systems*, pages 145–152, The Hague Netherlands, 2000. ACM.
- [5] C. Coverson. *Query Expansion Using an Interactive Concept Hierarchy*. Master's dissertation, Department of Information Studies, University of Sheffield, 2000.
- [6] E. N. Efthimiadis. Query expansion. *Annual Review of Information Systems and Technology*, 31:121–187, 1996.
- [7] E. N. Efthimiadis. Interactive query expansion: A user-based evaluation in a relevance feedback environment. *Journal of the American Society for Information Science*, 51(11):989–1003, 2000.
- [8] J. Greenberg. Optimal query expansion (QE) processing methods with semantically encoded structured thesauri terminology. *Journal of the American Society for Information Science and Technology*, 52(6):487–498, 2001.
- [9] D. Harman. Towards interactive query expansion. In *Proceedings of the 11th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 321–331, Grenoble, France, 1988. ACM.
- [10] M. A. Hearst. User interfaces and visualization. In R. Baeza-Yates and B. Ribeiro-Neto, editors, *Modern Information Retrieval*, pages 257–323. ACM Press, New York, 1999.
- [11] M. A. Hearst and J. O. Pederson. Re-examining the cluster hypothesis: Scatter/Gather on retrieval results. In H.-P. Frei, D. Harman, P. Schable, and R. Wilkinson, editors, *Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 76–84, Zurich, Switzerland, 1996. ACM.
- [12] J. Koenemann and N. J. Belkin. A case for interaction: A study of interactive information retrieval behavior and effectiveness. In *Conference Proceedings on Human Factors in Computing Systems (CHI '96)*, pages 205–212, Vancouver, Canada, 1996. ACM.
- [13] M. Magennis and C. J. van Rijsbergen. The potential and actual effectiveness of interactive query expansion. In *Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 324–332, Philadelphia, PA, 1997. ACM.
- [14] G. A. Miller. WordNet: A lexical database for English. *Communications of the ACM*, 38(11):39–41, 1995.
- [15] S. Pollitt. Interactive information retrieval based on faceted classification using views. In *Proceedings of the 6th International Study Conference on Classification (FID/CR)*, London, UK, 1997. University College of London. Available from <http://www.hud.ac.uk/schools/cedar/dorking.htm> [Accessed: 29/03/2001].
- [16] Y. Qiu and H. P. Frei. Concept based query expansion. In R. Korfhage, E. M. Rasmussen, and P. Willett, editors, *Proceedings of the 16th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 160–169, Pittsburgh, PA, 1993. ACM.
- [17] S. E. Robertson, S. Walker, and M. Beaulieu. Laboratory experiments with Okapi: Participation in the TREC programme. *Journal of Documentation*, 53(1):20–34, 1997.
- [18] G. Salton and C. Buckley. Improving retrieval performance by relevance feedback. *Journal of the American Society for Information Science*, 41(4):288–297, 1990.
- [19] M. Sanderson and B. Croft. Deriving concept hierarchies from text. In M. Hearst, G. Gey, and R. Tong, editors, *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 206–213, Berkeley, CA, 1999. ACM.
- [20] B. R. Schatz, E. H. Johnson, P. A. Cochrane, and H. Chen. Interactive term suggestion for users of digital libraries: Using subject thesauri and co-occurrence lists for information retrieval. In *Proceedings of the 1st ACM International Conference on Digital Libraries*, pages 126–133, Bethesda, MD, 1996. ACM.
- [21] M. M. Sebrechts, J. Vasilakis, M. S. Miller, J. V. Cugini, and S. J. Laskowski. Visualization of search results: A comparative evaluation of text, 2D, and 3D interfaces. In M. Hearst, G. Gey, and R. Tong, editors, *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 3–10, Berkeley, CA, 1999. ACM.
- [22] K. Sparck Jones. A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, 28(1):11–21, 1972.
- [23] E. Voorhees and D. Harman. Overview of the sixth text retrieval conference (TREC-6). In E. Voorhees and D. Harman, editors, *NIST Special Publication 500-240: The Sixth Text REtrieval Conference (TREC 6)*, pages 1–24, Gaithersburg, MD, 1998. NIST.
- [24] O. Zamir and O. Etzioni. Grouper: A dynamic clustering interface to web search results. In A. Mendelzon, editor, *Proceedings of the Eighth International World Wide Web Conference (WWW8)*, Toronto, Canada, 1999. Elsevier Science.

APPENDIX

A. SAMPLE USER INTERFACE

The screenshot shows a Microsoft Internet Explorer browser window with the address bar displaying <http://startrec.shef.ac.uk/~mark/subsumption/menus/Claire/2/>. The browser's menu bar includes File, Edit, View, Favorites, Tools, and Help.

The main content area is divided into two columns:

- Left Column:**
 - Number:** 302
 - Query entered:** Poliomyelitis and Post-Polio
 - Information need:** Is the disease of Poliomyelitis (polio) under control in the world? Relevant documents should contain data or outbreaks of the polio disease (large or small scale), medical protection against the disease, reports on what has been labeled as "post-polio" problems. Of interest would be location of the cases, how severe, as well as what is being done in the "post-polio" area.
 - Expansion terms:** A list of terms with a dropdown menu. The menu is open, showing a list of terms: post office, Polio, vaccine, vaccinated, measles, tetanus, New York Post, Postup, Post, Postup, Lviv, polio vaccine, Salk, Salk vaccine, Jane Smith, new vaccine, Sabin, immunised, measles vaccine, AIDS vaccine, immunization program, and virulence. The terms 'vaccine', 'Salk', and 'Jane Smith' are highlighted in pink.
- Right Column:**
 - Selected term:** A list of terms with a 'Delete' link next to each:
 - Delete - polio virus
 - Delete - polio victim
 - Delete - polio cases
 - Delete - post polio syndrome
 - Delete - polio vaccine
 - Delete - immunisation

The browser's status bar at the bottom shows the Internet icon and the text 'Internet'.