

Part V

Conclusion

In Part IV I described a user experiment to test the performance of the information need detection and information need tracking components of the probabilistic framework. The results showed that the framework chose terms and strategies that were apt and liked by experimental subjects. The evaluation also compared experimental systems that varied the amount of control subjects had over conveying relevance information, creating new query statements and deciding how to use these new statements (i.e., how they interacted with the framework and it with them). The evaluation also showed that the subjects preferred implicit relevance indications to explicit and a system that made recommendations about additional terms and strategies over systems offering intrusive forms of support (where systems act directly) or passive forms of support (where systems await searcher action). In this part I conclude this thesis and present avenues for future work; both drawn from findings obtained in the user experiment and research described throughout this thesis.

Chapter 13

Conclusions

13.1 Introduction

In this thesis I have investigated the use implicit feedback techniques to help searchers use search systems more effectively. The components introduced help searchers create new queries and help them make new search decisions about how to use these queries to find new documents or reorganise information already retrieved. In Part II I described techniques to help searchers maximise the amount of useful information they can access during a search. In Part III, heuristic-based and probabilistic implicit feedback frameworks were introduced that use this interaction to revise queries and make search decisions. The term selection parts of these frameworks (and other baselines) were evaluated with a simulation-based evaluation methodology I devised to test how well each term selection model ‘learned’ relevance and improved search effectiveness. The findings of Parts II and III motivate the development of the interfaces described in Part IV, where I present an investigation of how the implicit feedback framework should communicate with the searcher and vice versa. In this chapter I conclude and summarise the main findings and contributions of this thesis.

13.2 Content-Driven Information Seeking

In Part II I introduced new interface techniques to encourage searchers to search effectively by providing them with more information to make their decisions; I called this approach content-driven information seeking (CDIS). Unlike traditional result presentation techniques used by Web search engines such as Google, this approach shifts the focus of interaction at the results interface from documents to the information resident inside documents. To do this it uses query-relevant *Top-Ranking Sentences* extracted from top documents as an interface component to facilitate effective information access. Top-Ranking Sentences are a precision-oriented approach I devised to maximise the amount of useful information a searcher can

access. I conducted three related user studies to test the effectiveness of these sentences with real searchers in different search scenarios. In the first study, I used the ranked sentences as an alternative to document lists, shifting searcher attention from the document surrogates (i.e., titles, sentence fragments and URLs) to document content. The second used the sentences to reflect the use of two contrasting relevance feedback techniques. The third used the sentences to encourage interaction with the retrieved set, to reflect the dynamic nature of the information need and to complement, rather than replace, document lists. Each study involved human subjects and different types of information seeking scenario based around simulated work task situations. I showed that the CDIS approach, whether or not supported by additional implicit feedback techniques that reorder the sentences, can lead to effective and efficient searching.

As part of the exposition of CDIS, I also introduced the notion of ‘push’ and ‘pull’ information seeking and explained that these approaches differ in how information is presented to the searcher. Motivated by the success of the techniques in the studies described in Chapter Four, I extended the CDIS approaches in Chapter Five with the inclusion of more document representations and relevance paths that join them. Content-rich search interfaces were developed using these additional representations to encourage interaction and create more evidence for the implicit feedback frameworks introduced in Part III. In user studies of interfaces that used these additional components (Pilot Test 1 and presented in Part IV) I showed that searchers found them helpful, that they encouraged more interaction with search results and that they felt the additional interaction was beneficial to them; this benefit was more apparent when information needs were vague or the search tasks complex.

Ranking documents is a cumbersome means of result presentation. Documents may not be entirely relevant and document titles, sentence fragments and URLs may not be strictly indicative; it is the information inside documents that searchers generally seek. The CDIS approach I introduced extracts, ranks and presents potentially relevant content from the returned set, blurring inter-document boundaries and encouraging information seeking based on the pertinent document content. In the next section I describe implicit feedback frameworks that use interaction with content-rich search interfaces as evidence to help them make search decisions.

13.3 Implicit Feedback Frameworks

In Part III two novel implicit feedback frameworks were introduced: one heuristic-based and one probabilistic. The frameworks estimate current information needs and estimate changes in those needs as a searcher interacts with the results of their retrieval. The frameworks presented support searchers by passively observing their search behaviour and choosing new query terms and retrieval strategies to help them locate relevant information. They aim to help those who are unable or unwilling to communicate relevance information directly or simply may be struggling to find what they want.

Motivated by the success of interface components and the implicit feedback techniques described in Part II the implicit feedback frameworks I created approximate searcher interests through interaction with representations of top-ranked documents and interactive paths that join them. This differs from traditional potentially unreliable sources of implicit feedback such as document reading time, scrolling and other such measurable search behaviours within the full-text of potentially relevant documents. In rich information seeking environments like those created by CDIS techniques, searchers can view information to a fine level of detail and the information they view can be used to approximate their interests. The frameworks performed this function well and provided a means through which searcher intentions could be inferred implicitly, without the need for direct searcher involvement in providing relevance information.

As I established in Part I, information needs are not static and can change during a search on exposure to new information. The implicit feedback frameworks contain components that allow them to predict when, and by how much, the topic of a search has changed based on short-term, within search session, interaction histories. Depending on the degree of the change the frameworks can pick retrieval strategies that will be useful to searchers. That is, the level of interface support offered by systems that implement these frameworks depends on the extent to which information needs are estimated to change. There are four possible strategies the framework can follow: no action (for small changes), reorder top-ranking sentence list (for small-moderate changes), reorder document list (for moderate-large changes) and re-search (for large changes). I conducted a study of topic similarity measures that demonstrated the effectiveness of correlation coefficients for predicting the extent of the difference between search topics. The results show that measures based on the level of correlation between topics concords highly with general subject perceptions of search topic similarity and that these coefficients may be useful to predict search topic change. As a

result, the Spearman and Pearson correlation coefficients were used as tools to estimate changes in searcher interests and select appropriate retrieval strategies.

Two user experiments involving a total of 72 different subjects (i.e., Pilot Test 1 and the experiment in Part IV) have shown that the heuristic-based and probabilistic implicit feedback frameworks choose new query terms and make decisions about query use that are appropriate and liked by experimental subjects. The techniques discussed in this thesis have the potential to alleviate some of the problems inherent in traditional RF – where searchers are directly involved in the provision of relevance information – whilst preserving the benefits that underlie the approach. The initial query is still modified to become attuned to a searcher's need based on an iterative process of feedback. However, searchers do not have to explicitly assess and mark documents as relevant and the way the new query is used depends on the extent to which the information need is estimated to have changed (i.e., the search systems do not only re-search the document collection). In the next section I describe the simulation-based evaluation methodology I developed to test the term selection models that in the implicit feedback frameworks. This methodology is used as a formative evaluation technique to select the best-performing model for implementation in the search interfaces described in Part IV.

13.4 Simulation-Based Evaluation Methodology

A novel simulation-based evaluation methodology was used to test the performance of the term selection components of implicit feedback frameworks (called implicit feedback models) in different simulated contexts. This methodology is less time consuming and costly than experimentation with human subjects, allows environmental and situational variables to be more strictly controlled and complex searcher interactions to be modelled. It allowed me to compare and fine-tune a number of potential implicit feedback models before the best performing model was deployed in an interactive search system. Simulations of this nature could be a powerful formative evaluation tool for the designers of search interfaces, especially those that do not conform to traditional forms of search result presentation (i.e., ranked lists of documents). Designers can test a prototype interface with one implicit feedback model to remove potentially problematic interactions or, as I have described in this thesis, test many models for a given search interface to choose the most effective model.

The implicit feedback models tested were ostensive in nature and use the exploration of the retrieved information and the viewing of document representations as an indication of relevance. Six implicit feedback models were tested in total, all using an ostensive paradigm

but each employing a different term selection stratagem. The methodology tested those models in different search situations.

I introduced implicit feedback models based on Jeffrey's rule of conditioning, Binary Voting, three that use the popular *wpq* query expansion approach and a baseline that selected terms randomly. The simulated approach used to test the models assumes the role of a searcher 'viewing' relevant documents and relevance paths between different representations of documents. The simulation passes the information it viewed to the implicit feedback models, which use this as evidence of relevance to select terms to best describe this information. In the evaluation I investigated the degree to which each of the models improved search effectiveness and 'learned' what information was relevant. From the six implicit feedback models tested, the Jeffrey's Conditioning Model was most effective. As demonstrated in Chapter Eight, this model outperformed the others in a variety of different simulated search scenarios with different proportions of relevant and non-relevant information and other interaction constraints. This model was subsequently chosen as the term selection component of the implicit feedback framework tested in the experiment described in Part IV. During this experiment the ability of the model to identify information needs was re-tested with human subjects. The results of that experiment showed that the model chose terms that were relevant and useful.

The simulation-based evaluation methodology I propose is an effective way of testing the worth of implicit feedback models such as those presented in this thesis. Experimentation with human subjects can be costly and these tests can ensure that only the best models are chosen to be tested with real searchers in interactive information seeking environments. The next section discusses issues in the interface support offered to searchers.

13.5 Interface Support

The results of all user experiments described in Parts II and IV show that it is possible to get searchers to interact with more than a few search results. The approaches introduced move away from simply presenting titles to presenting alternative access methods for assessing and targeting potentially relevant information. From observations and informal post-search interviews across a series of related studies, subjects appeared to find the increased level of content shown at the results interfaces of value in their search. This is important, as the success of all experimental systems I present – especially those based on implicit feedback techniques – is dependent on the use of these interface features.

In Part IV of this thesis I investigated how implicit feedback frameworks can best communicate with searchers (and vice versa) and evaluated the implicit feedback framework chosen from the findings of the simulations in Part III. The experiment used three different types of RF interface that varied how searchers provided relevance information, how they created new queries and how they made new search decisions. Three systems were created: a Checkbox system that relied on explicit relevance assessments, provided support in creating queries and relied on searcher to select retrieval strategies; a Recommendation system, that gathered implicit relevance assessments and *recommended* query terms and strategies, and an Automatic system that gathered implicit relevance assessments and *selected* terms and strategies.

In this experiment, subjects preferred the Recommendation system and found it useful for more complex tasks where more control over the query terms was preferable. Subjects found the Checkbox system useful for low complexity tasks where the objective of the search was clear. Subjects found the Automatic system useful for complex searches where the subject did not want to be actively engaged in the information seeking process or lacked sufficient knowledge about the retrieval environment to make good decisions. The different systems were therefore useful for different types of search, although the Recommendation system (originally devised based on the feedback of subjects in Pilot Test 1) was generally most popular as it gave searchers control over query term selection and use.

The terms selected for query modification by the probabilistic framework were both useful and relevant and were accepted by searchers on different systems since they were valuable for their search tasks. The approach tracked potential changes or developments in the information need based on changes in the document representations viewed by the subject. The system communicated its estimation of these developments through the decisions it made on subjects' behalf; subjects generally felt these strategies were useful and appropriate.

Implicit relevance information is inherently uncertain. The Recommendation system worked in tandem with the searcher, making suggestions on what terms they could add or what strategies they could select. The uncertainty surrounding how implicit evidence is gathered means that it is desirable to give searchers final control over systems that use it. In the experiment in Part IV (as in Pilot Test 1) subjects wished to retain control over activities they perceived as being important for the success of their search. That is, subjects were willing to delegate control over the provision of relevance information (i.e., the inputs) as long as they could control how this information was used in constructing new queries or making new search decisions (i.e., the outputs).

13.6 Chapter Summary

In this thesis I have presented and evaluated a set of techniques to support searchers engaged in interactive information retrieval. I have developed novel search interfaces that ‘push’ potentially relevant information toward the searcher, helping them proactively as they search. I have developed content-rich search interfaces that extend this approach to involve a greater variety of document representations and interactive relevance paths that join these representations. These interface techniques have been shown to help searchers, especially for complex search tasks. I have developed and tested heuristic-based and probabilistic implicit feedback frameworks that use interaction with these content-rich interfaces to estimate and track information needs. User experiments have shown that the frameworks select terms and retrieval strategies that subjects found appropriate and helpful.

I developed a simulation-based evaluation methodology for testing implicit feedback models with simulated searchers and benchmarked the performance of six different models in a variety of retrieval scenarios. The methodology allows complex interaction to be modelled and experimental variables to be closely controlled whilst giving system designers a formative evaluation tool to assist in the selection of RF algorithms or design of search interfaces. The best model was chosen to be part of an experiment to further test it with human subjects and with different types of interface support in feedback systems. I developed three RF search interfaces, whose design was motivated by findings in my earlier studies, each using the best performing model and each with different interface options that afforded different amounts of searcher control. The results showed that searchers are happy to delegate responsibility for relevance assessment to RF systems (through implicit feedback), but not more severe decisions such as formulating queries or selecting retrieval strategies; for such decisions searchers wanted support from the system, but ultimately control over its actions.

This research has investigated innovative techniques for interface design, implicit feedback and evaluation for interactive IR. The ramifications of this work are notable and warrant further investigation. The final chapter will outline potential avenues for such investigation in future work.

Chapter 14

Future Work

14.1 Introduction

This thesis has explored many issues in the areas of implicit feedback and interactive information retrieval. Many avenues have emerged for the research described to be taken further and in this chapter I describe some of the main opportunities and challenges that this work provides. In the same way as the previous chapter I discuss future work in a number of sections, based on the contributions made by this thesis.

14.2 Content-Driven Information Seeking

The presentation of multiple representations of search results at the results interface was promising and liked by searchers in all user studies conducted. There were minor issues with the presentation of this content, such as the occlusion of other information when viewing document summaries or sentences in context, although these could be resolved with slight modifications to interface design. The results of the experiments in Chapter Four suggested that the content-driven approaches were of most use for search tasks where a lot of information is preferred to improve topic familiarity or awareness (*background* search) or improve decision making abilities (*decision* search). However, the approach was not of as much use for ‘fact’ searches where the information need was exact. Since the content-driven approaches are not as effective in all information seeking contexts it is important to identify when the approach should be used and when it should not. That is, when should a searcher be presented with a list of Top-Ranking Sentences and other interface components, and when should they be faced with a ranked list of document surrogates. The decision of when to use content-driven information seeking techniques should ideally be taken by the search system, since searchers may not realise the potential benefits of the approach. There is future work in developing mechanisms to make these decisions.

As suggested in Chapter Five, the interfaces described in this thesis implement aspects of a polyrepresentative approach i.e., presenting multiple document representations to reduce uncertainty in implicit feedback. In future work I will analyse the interactive experiments conducted with these interfaces in Pilot Test 1 and in the experiment presented in Part IV from the perspective of polyrepresentation and use the system logs of mouse movements and clicks to allow me to better understand and interpret system usage.

14.3 Implicit Feedback Frameworks

The implicit feedback frameworks chose additional search terms that were relevant and useful for searchers. However, a larger scale empirical evaluation to improve the indicativity heuristics used in the Binary Voting Model may improve the effectiveness of the heuristic-based framework. The performance of both frameworks could also be improved if searchers could indicate what information is definitely not relevant. During the experiments some subjects suggested that they wanted control over what information the search system disregarded and excluded from the search. The issues about whether searchers are actually able to exercise the control to provide negative RF effectively has already been raised by Belkin *et al.* (1998). Potential avenues for future work could be on the development of hybrid positive/negative implicit/explicit RF systems that gather positive assessments unobtrusively and negative assessments directly from the searcher. In this thesis searcher actions such as regressing back along a relevance path, or reversing a search decision made by the search system were ignored by the frameworks and only positive assessments were considered. Further work is needed in using these indications of dissatisfaction to infer what information searchers do not want.

The results of Pilot Test 1 and the experiment presented in Part IV suggests that the weakest part of the implicit feedback frameworks is the component to estimate changes in information needs. Developing sound techniques to track changing needs can be difficult as searchers may be unaware that any change has occurred and needs may change in different ways to different degrees. However, in future work there is a need to address the shortfall between searcher expectations of the component and its actual performance.

The frameworks currently only track information needs during a single search session. An avenue for future work would be test the effect of incorporating searchers' long term interests. These interests could be used to develop a potentially more robust formulation of the information need from which query terms could be chosen. The frameworks I have introduced use interaction with IR system result interfaces as implicit feedback. A deeper

understanding of how searchers interact with the full-text of relevant documents is needed before traditional implicit feedback metrics (e.g., viewing time or scrolling) could be used to complement these frameworks I present here.

14.4 Simulation-Based Evaluation Methodology

The simulation-based evaluation methodology allowed feedback models to be compared in an experimental setting without human subjects. Simulations of this nature can be used either after a prototype interface was built (as was the case in this thesis), or before the interface is built, to test its performance with every possible set of potential searcher interactions prior to development. Testing of this nature can assist system designers in identifying the strengths and weaknesses of the interface with a particular implicit feedback model (allowing them to eliminate interactions that could cause problems) or the strengths and weaknesses of many implicit feedback models for a given search interface (allowing them to choose a model that suits their needs). More work is necessary in developing a framework to allow simulations of this nature to be developed in a robust, generic and extensible way.

To more closely emulate the search behaviour of humans the simulations need to make decisions that resemble those that human subjects may make. In future work I will address this issue and try to make the methodology more ‘intelligent’ to allow better ⁴⁷ decisions on what information to interact with and develop a suite of simulated searchers, each with their own stereotypical search behaviour. To test a model or system interface with, for example, experienced searchers, it should be possible to select a group of simulations with the appropriate characteristics, and plug them into the methodology. There is much scope for future work in developing effective searcher simulations to model different scenarios, searcher and searching style.

Searcher simulations could also be used to mimic changes in the topic of the search and monitor how well the relevance feedback techniques adapt to this change and how well components to track changes in information needs detect these changes. Changes in the search topic are potentially difficult to estimate as information need change or development is perhaps more difficult to monitor than the information need itself, which can be approximated at the relevant document level or through decent query terms. More work is necessary in simulating different rates of change and different search strategies and tactics used during this change.

⁴⁷ To simulate novice searchers or those engaged in complex search tasks the methodology may also need to make bad decisions.

14.5 Interface Support

In Part IV I tested three experimental interfaces that gave searchers different levels of control and responsibility over aspects of the search. The interfaces were generally liked by searchers, and while they were happy to delegate responsibility for gathering relevance information to the search system they wished to retain control of query creation and retrieval strategy selection. This suggested that subjects did not trust the implicit feedback framework sufficiently to give it complete control over all search decisions. More work is therefore necessary to engender trust in system decisions by improving the effectiveness of the framework and how the decisions are communicated at the search interface. The use of explanations, such as that proposed by Ruthven (2002), may help searchers understand why certain terms were chosen and search decisions made. Further work is necessary to investigate task and situational differences in searcher control and responsibility.

Further work is also necessary on the association between the support offered by the experimental system and the complexity of the search task. The testing of interfaces in laboratory settings may not reveal problems encountered in operational settings, where IR systems are typically used. In future work a longitudinal evaluation of the systems in an operational environment is essential to test the worth of the interface approaches proposed.

14.6 Chapter Summary

This chapter has detailed opportunities to further the research presented in this thesis. The techniques proposed have fundamental implications for the design of interactive information retrieval systems and their evaluation. This work has shown that searchers respond well to the content-driven information seeking approaches and the implicit feedback frameworks that use them. The simulation-based evaluation techniques I propose provide a means through which interfaces and their underlying mechanisms can be assessed. It is vital therefore that more work is undertaken to further this imaginative research, and test these concepts in operational environments and longitudinal user experiments.