Retrieval through explanation: an abductive inference approach to relevance feedback

C. J. van Rijsbergen and I. Ruthven, University of Glasgow
M. Lalmas, Queen Mary and Westfield College

A. Summary

Selecting good query terms to represent an information need is difficult. The complexity of verbalising an information need can increase when the need is vague [DDRvR99], when the document collection is unfamiliar, [Roc71], or when the searcher is inexperienced with information retrieval (IR) systems, [CPBKN96]. It is much easier, however, for users to assess which documents contain relevant information.

Relevance feedback (RF) techniques make use of this fact to automatically modify a query representation based on the documents users consider to be relevant. RF has proved to be relatively successful at increasing the effectiveness of retrieval systems in certain types of search, [Har92], and RF techniques have gradually appeared in operational IR systems and even some Web search engines. However, the traditional approaches to RF do not consider the *behavioural* aspects of information seeking and making relevance assessments. The standard RF algorithms consider only *which* documents users have marked as relevant; they do not consider *why* the users have assessed relevance. For RF to become an effective support to information seeking it is imperative to develop new models of RF that are capable of incorporating *why* users make relevance assessments.

The underlying assumption of the vast majority of RF theories, [Har92], is that terms occurring more frequently in relevant documents than non-relevant documents tend to be good for retrieving more relevant documents [RSJ76, Roc71]. However, it has been demonstrated in a number of studies, e.g. [BS98], that why users mark documents as relevant is as important as which documents they mark relevant, in deciding what further documents to retrieve. This means taking into account how information is used within documents. In [RL99a], we developed a successful RF approach which not only considers which terms are used in documents: it also incorporates how the terms are used.

In addition to analysing why users mark documents as relevant we also have to consider *how* users assess relevance. Current RF algorithms lack the qualitative reasoning ability necessary to tackle aspects of relevance assessments such as developing information needs [Cam95], consistency of relevance assessments [BI97] and users' strategies for assessing relevance [FM95]. The passive nature of standard RF techniques does not provide a basis for incorporating such user behaviour into RF.

We view RF as a process of *explanation,* [Bel88, Caw92] An RF theory should provide an explanation of why a document is relevant to an information need. Such an explanation can be based on how information is used within documents.

We will use *abductive logic*, [JJ94] to provide a framework for an explanation-based account of RF. Abductive logic is specifically designed as a technique for generating explanations of complex events, and has been widely used in a range of diagnostic systems. Such a framework will produce a set of possible explanations for why users marked a number of documents relevant at the current iteration. These explanations will be based on how information is used within relevant documents. From the set of possible explanations, one explanation, known as the best possible explanation, will be selected to reformulate the query. The choice of the best possible explanation is guided by a number of factors, the main factor being the previous search history. Our complete methodology for an explanation-based account of RF is presented in Section B.

Recent research has indicated that the success of an RF technique is related to how well users understand how RF works, e.g. [CPBKN96]. An explanation-based system of RF can be used to aid the users' understanding of the effects of marking a document relevant. In addition to a rigorous evaluation of the effectiveness of our framework in a real search environment, we will investigate how explanations affect the users' searching behaviour. Our evaluation strategy is presented in Section C.

The development of our framework and its evaluation are strongly related: our model of explanation-based RF has a strong behavioural element and the explanations it generates can be used at the interface to aid users' understanding of the feedback process.

This need for RF techniques that promote better understanding of RF by the user and improved consideration of the users' role in RF was identified by the Mira Working Group on IR evaluation [DDRvR99] as one of the most pressing requirements of interactive IR.

B. Methodology

RF can be viewed as a cycle of activity: an IR system presents users with a set of retrieved documents, the user then indicates those documents that contain relevant information, the system uses this relevance information to modify the query, and a new set of documents based on this new representation are retrieved and displayed to the user.

However, this basic model and the ones derived from it, [Har92], suffer from a number of limitations that restrict their potential in supporting information seeking. These can be classified into three main groups: *expressiveness of indexing language*, *incorporation of behavioural aspects of searching*, and the *usability of RF*. We demonstrate how these three elements of RF provide the basis of an explanation-based view of RF.

B.1 Expressiveness of indexing language

In [RL99a] we described how the simple term-based approach to document indexing can be extended to include information on how individual terms are used within documents. We defined a number of weighting schemes - *term characteristics* - to describe aspects of the use of a term in a document, e.g. *theme* - a characteristic which describes the degree to which the term is likely to be the main topic of a document, *context* - a characteristic which describes the dependence of one query term on another.

The traditional approach to RF produces a new query consisting of a set of terms and frequency-based weights, for example:

index(millennium)	0.5,	
index(computer)	0.4,	**(Q1)**
index(bug)	0.1	

In our approach, the new query includes information on each of the terms role in the relevant documents, for instance:

index(millennium) 0.5, *theme(millennium)* 0.3,
index(computer) 0.4, *theme(computer)* 0.7, **(Q2)**
index(bug) 0.1, *context(millennium, bug)* 0.4

A query then consists of a set of term characteristics rather than a set of terms. Query **(Q1)** would prioritise retrieval of any document containing the three terms, *millennium, computer*, and *bug*, whereas query **(Q2)** would prioritise retrieval of documents in which the main topics were *millennium* and *computer*, and the terms *millennium* and *bug* appeared in close proximity to each other.

In [RL99a] we demonstrated that RF is not only a case of selecting *which* terms to use in a query but which *characteristics* of a term. We showed, experimentally, that it was possible to select, for a given set of relevance assessments, an optimal set of term characteristics to use in a query. This means different sets of term characteristics are better at retrieving relevant documents for different queries.

We showed, through a series of experiments on large document collections, that this technique of selecting which characteristics to use for which query not only performed well but outperformed classic RF algorithms such as the $F_{4.5}$ measure, [RSJ76]. In continuing research, [RB99], we demonstrate that this also holds for data derived from users in a real search environment.

Term characteristics, therefore, provide the basis for an indexing language for our explanation-based account of RF. Each term characteristic indicates how a term is used within a document and can be used to differentiate why users marked a document as relevant.

Objective one: We will identify other term characteristics to increase the flexibility of RF in detecting what makes a document relevant. Our approach to enhancing the document indexing has two strengths: firstly, the term characteristic information can be stored in data structures analogous to traditional index files; it does not require the design of new data structures. Secondly, it adapts existing techniques. The term characteristics we have created, and have identified for future use, are adaptations of existing retrieval techniques, such as [HP93, Kho95].

B.2 Incorporation of behavioural aspects of searching

Critics of RF techniques [Cam95, SW99, DMB98] argue that the traditional RF techniques make too many simplifying assumptions about users' role in the RF process. For example, RF techniques traditionally assume that information needs are fixed, that all relevant documents are equally relevant and that only one information need is under investigation in a search. Studies of how users make relevance assessments, e.g.[FM95], show that this is not the case: documents may only be partially relevant, different documents may satisfy different aspects of an information need and those documents that users finds relevant may change during a search. Why users assess material as relevant should then be central to RF.

We believe that understanding relevance assessments should be a process of *explanation:* explaining why users made a relevance assessment on a document. Such a process should take into account not only the current relevance assessments but also previous assessments made within the information seeking session.

Such an explanation will be based on term characteristics, and will be used as a modified query to retrieve more documents. Creating an explanation is a process of deciding which term characteristics to use and the importance of each of them in the modified query.

Abductive logic, [JJ94], provides a framework of generating such explanations and has been successfully used in IR for the construction of hypermedia and image retrieval systems [Mull99]. The use of abductive logic allows us to model RF by taking a set of relevant documents and inferring a set of possible explanations about the relevance of these documents.

In the example below, taken from [RB99], we have a query '*endowment mortgage*' and three relevant documents, **(D1)**, **(D2)** and **(D3)**. The documents are represented by the term characteristics of the query terms they contain.

theme(mortgage) **(D1)**

theme(mortgage) and
context(mortgage, endowment) **(D2)**

theme(mortgage) and
theme(endowment) and
context(mortgage, endowment) **(D3)**

Possible explanations for marking these documents relevant are:
• *theme(mortgage)* - "relevant documents contain the term *mortgage* as a main topic",
• *context(mortgage, endowment)* - "relevant documents mention *mortgage* and *endowment* in close proximity",
• *theme(mortgage)* and *context(mortgage, endowment)* - "relevant documents have *mortgage* as a main topic but will also mention *endowment* in close proximity to *mortgage*".

However, not all explanations will be equally good at explaining the relevance assessments. In addition to providing a set of possible explanations, abductive logic allows for the identification of the *best possible explanation.* Abductive logic [JJ94] provides a number of techniques to select the best possible explanation from a set of explanations. Which technique is most suitable to describe a set of relevance assessments will form a major research question in our work.

There will, however, be a number of factors to consider in deciding which is the best explanation. For example, *theme(endowment)* may be a poorer explanation than *theme(mortgage)* because *theme(endowment)* only explains the relevance of one of the three documents, whereas *theme(mortgage)* is an explanation of the relevance of all three documents.

The use of abductive inference constitutes a more qualitative reasoning approach to RF than current RF algorithms. It is qualitative in the sense that it differentiates between characteristics of a term's use. It is a reasoning approach in that it *actively* selects an optimal explanation from a number of alternative explanations.

Objective two: We will develop a theory of what constitutes a good explanation for RF based on an abductive framework.

The process of generating abductive explanations and selecting the best possible explanation also supports the modelling of *uncertainty* in information seeking, [Kuh93]. The evidence given by terms and term characteristics, as to the relevance of documents, is by nature uncertain, the development of explanations must tackle the system uncertainty of the search process.The choice of best possible explanation will be affected, amongst other things, by the likelihood of the characteristics that form the explanation. For example, a term characteristics that appears more frequently in relevant than irrelevant documents is more likely to be part of a good explanation.

Uncertainty can also arise at the user level [Kuh93]. This uncertainty can be demonstrated in different ways: the use of partial relevance assessments, the number of relevance assessments at an iteration and the similarity of information in the relevant documents.

For example, [SW99] suggests that a high number of partial relevance assessments at certain points in a search indicate a high level of uncertainty on the part of users of what they are searching for. In our approach this information can be used in a number of ways: to select an explanation that biases retrieval in favour of an explanation that covers most of the assessments or to suggest possible explanations to users.

Explanations will also take into account the temporal nature of relevance assessments, [Cam95], and the fact that user's information needs may change during a search. Our approach of generating explanations for relevance assessments makes it possible to assess how a search is changing over time. This can be achieved by comparing explanations generated at different points in a search and examining them for consistency within, and across iterations, regarding the term characteristics used in each explanation The consistency of term characteristics at different iterations can be used to direct the selection of the best possible explanation.

The quality of an explanation, then, is not just how good it is at explaining the relevant assessments made at the current iteration but also how well it explains the assessments the context of a whole search. An inference model for generating explanations will incorporate how a search is changing over time.

Traditional approaches to RF cannot incorporate this kind of reasoning because they rely on passive techniques to reformulate queries. Our abductive approach can actively select a response to the user's actions based on the relevance assessments they make.

The move towards this type of approach is necessary if RF is to take account of the phenomena described in studies of user searching behaviour.

Objective three: We will specify what constitutes a good explanation of the current relevance assessments within the context of explanations generate at previous iterations.

B.3 Usability of RF

RF was conceived as an automatic technique: its operations were to be hidden from users and not intended to be understood by users. However, [CPBKN96], and others, have shown that users' satisfaction with an RF technique is directly related to how well they understand how the RF technique works and its affect on subsequent retrieval.

The traditional approach to RF would produce a reformulated query, **(Q1)**, which, if displayed to users, would only tell them which terms were being used to search and perhaps the weights assigned to them.

index(1999)	0.5,	
index(internet)	0.4,	**(Q1)**
index(conference)	0.1	

If the users understand how IR systems work they could infer which documents are likely to be retrieved and in which order. If the users do not understand how the system works then the new query gives no meaningful information on which documents are likely to be retrieved: the users cannot tell how RF is going to affect what type of documents will be retrieved by the system.

Attempts to improve the role of users in RF, such as interactive query expansion [CPBKN96], usually rely on increasing users control over the query reformulation. The concept of explanation can extend the users' control by presenting information on how the terms are used in searching.

In our approach the re-formulated query includes information on each of the terms' role in the relevant documents. Including term characteristic information, our example query could look like query **(Q2)** or query **(Q3)**, depending on how the terms are used in the relevant documents.

index(1999)
theme(internet) **(Q2)**
context(internet, conference)

index(1999)
theme(conference) **(Q3)**
index(internet)

Query **(Q2)** is a query which will prioritise retrieval of documents whose main topic is *internet* mentioning *1999*, with *internet* and *conference* in close proximity whereas query **(Q3)** will prioritise retrieval documents about *conference*, mentioning *1999* and *internet*. The difference between what type of documents will be retrieved in response to queries **(Q2)** and **(Q3)** is hidden in **(Q1)** which only informs the users about what terms are being used to search.

Explanation used in this way does not require users to understand more of the retrieval implementation because it is directly related to the information users are searching for. We will investigate in our evaluation stage how reformulated queries provided by explanations can best be presented to users to help them understand the effects on RF on their search.

In addition to helping users understand what the system is searching for, explanations can also be used to help users understand how RF works. The strength of a reasoning approach to RF is that it can be used to present explanations of why the system made decisions on how to reformulate a query.

For example we could explain to users that the system is using the characteristic *theme(internet)* to search for more documents because the previously marked documents all had *internet* as the main topic. We will investigate how this type of information can be used to help users understand the RF process.

There are then two uses for explanation which we will investigate: aiding users in searching and aiding users in understanding how to search.

Objective four: We will investigate the use of explanations at the interface. In particular we will look at whether the use of explanations aids users in searching and in understanding RF.

C. Evaluation

We will perform a rigorous evaluation of our approach, to examine the following three aspects: **i.** the effectiveness of the term characteristics approach in RF, **ii.** the effectiveness of abductive inference in representing dynamic information seeking and **iii.** an investigation of the use of explanation as a means of increasing users' understanding of RF. Each aspect will be investigated separately.

• **Aspect i.** We will test whether retrieval effectiveness is improved by taking into account how words are used within documents. Our early results, [RL99a], indicate that this is so but we require to test the abductive approach. This will be accomplished using a standard test collection methodology.

• **Aspect ii.** We will evaluate whether our system of abductive inference can take into account dynamic information seeking. Standard IR evaluation measures such as Recall and Precision are not suitable for evaluating this aspect of our system as they assume a fixed set of relevant documents known in advance. As users's information need may change during a search, the set of relevant documents may change. The degree to which this occurs is not something which can be assessed algorithmically [DDRvR99]; it must be tested with users tackling real information seeking problems in conjunction with standard evaluation techniques such as semi-structured interview, observational studies and questionnaires.

One local source of users with existing information needs is the Glasgow Medical School. This school has recently moved to a task-based model of teaching in which students are given clinical tasks to solve such as 'a patient has the following symptoms [...] diagnose what illnesses the patient may be suffering from' or 'write an essay on the physiology of liver cells'. Students are given a variety of information sources, including IR systems, which they can use to complete the task. The aim of this teaching method is to teach the students by developing their information seeking and self-learning skills rather by presenting information in the more traditional lecture and tutorial style. These students are typical of many users of IR systems in that they have real, time-limited information requirements, have varying knowledge of the domain, and varying experience of on-line information seeking tools.

We already have strong links with the Medical School and the IR group have used its students in previous experiments. This aspect of our evaluation will be users-centred evaluation, assessing how well the IR system supports users in completing tasks.

• **Aspect iii.** We will investigate how users' understanding of the RF process is enhanced by explanations based on our abductive model. For this, we will look at different methods of presenting explanation at the interface. This will be carried out complementary to our examination of the dynamic nature of information seeking.

E. Beneficiaries

The ultimate aim in developing an RF model is to build a system that will automatically improve an information search by adapting to users' information needs. This project is a step on the way: it will investigate an approach that seeks to explain why users marks documents as relevant in terms of how information is used within the documents.

The internet and digital libraries are major application areas for our work. With the huge growth of the internet and digital libraries, issues such as the variety of available document types, delays in downloading documents, and the number of available documents mean that users require more information about the documents they are retrieving. There is empirical evidence that users will not spend time in creating complex queries, even if these would help their search [DMB98].

Techniques such as the ones described in this proposal are not only intended to increase the retrieval effectiveness of search engines but also to increase the information users receive, at the interface, about the retrieval process. For users of IR systems, the main advantage of this approach is to increase the flexibility of system in detecting what users are searching for. This offers improved support to novice users in particular, who may have little experience of IR systems.

The development of a formal system of information manipulation also supports *reasoning* about users searching behaviour. That is, users' searches can be analysed to elicit information on how users are searching for information. An example of this kind of formal approach to modelling the interaction between users and an IR system has been developed using channel theory, [ML99]. The techniques developed by our proposal can provide the data necessary to facilitate such detailed analyses of searching behaviour.

The work proposed in this project will also benefit many IR and information science groups that are concerned with providing effective interaction between IR systems and users, and other basic research into users' information seeking patterns [SW99].

I. References

[BS98] C.L. Barry and L Schamber. Users' criteria for relevance evaluation: a cross-situational comparison. *Information, Processing and Management.* 34 (2/3). 1998.

[Bel88]. N. J. Belkin. On the nature and function of explanation in intelligent information retrieval. *Proceedings of the 11th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval.* Grenoble. pp 135 -145. 1988.

[BI97] P. Borlund and P. Ingwersen. The development of a method for the evaluation of interactive information retrieval systems. *Journal of Documentation*. 53 (5). pp 225 -250. 1997.

[Cam95] I. Campbell. Supporting information needs by ostensive definition in an adaptive information space. *MIRO '95*. electronic Workshops in Computing. Springer Verlag. I. Ruthven (ed). 1995.

[Caw92] A. Cawsey. *Explanation and interaction: the computer generation of explanatory dialogues*. MIT Press. 1992.

[CPBKN96] C. Cool, S. Park, N. J. Belkin, J. Koenemann, K. B. Ng. Information seeking behavior in new searching environment. *CoLIS 2, Proceedings of the Second International Conference on Conceptions of Library and Information Science: Integration in Perspective*. Copenhagen. pp 403-416. 1996.

[CLVR98] F. Crestani, M. Lalmas and C.J. van Rijsbergen. *Information Retrieval: Uncertainty and Logics - Advanced models for the representation and retrieval of information.* Kluwer Academic Publishers. 1998.

[DMB98] S. Dennis, R. McArthur and P. Bruza. Searching the WWW made easy? The Cognitive Load imposed by Query Refinement Mechanisms. *Proceedings of the Third Australian Document Computing Symposium.* 1998.

[DDRvR99] *Mira '99: Evaluating Information Retrieval.* electronic Workshops in Computing. S. Draper, M. Dunlop, I. Ruthven and C. J. van Rijsbergen (ed). *(to appear).*

[FM95] V. Florance and G. Marchionini. Information processing in the context of medical care. *Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval.* Seattle. pp 158 -163. 1995

[JJ94] J. R. Josephson and S. G. Josephson (ed). *Abductive Inference: Computation, Philosophy, Technology.* Cambridge University Press. 1994

[Har92] D. Harman. Relevance feedback and other query modification techniques. In: *Information retrieval: data structures & algorithms* . W. B. Frakes and R. Baeza-Yates, ed. Chapter 11. pp 241-263. 1992

[HP93] M. A. Hearst and C. Plaunt. Subtopic structuring for full-length document access. In: *Proceedings of ACM SIGIR Conference on Research and Development in Information Retrieval*. Pittsburgh, USA. ACM Press. 1993.

[Kho95] C. S-G. Khoo. *Automatic identification of causal relations in text and their use for improving precision in information retrieval.* PhD Thesis. Syracuse University. 1995.

[Kuh93] C.C. Kuhlthau. Principle for uncertainty for information seeking. *Journal of Documentation.* 49(4). pp 339-355. 1993.

[LR99] M. Lalmas and I. Ruthven. Modelling the interaction in information retrieval. *9th European-Japanese Conference on Information Modelling and Knowledge Bases (EJ-IMKB).* Iwate, Japan, 1999. *(to appear)*

[Mull99] A. Müller. *A Flexible Framework for Multimedia Information Retrieval.* In [CLVR98].

[RSJ76] S. E. Robertson and K. Sparck Jones. Relevance weighting of search terms. *Journal of the American Society of Information Science.* 27. pp 129-146.1976.

[Roc71] J. J. Rocchio. Relevance feedback in information retrieval. In: *The SMART retrieval system: experiments in automatic document processing.* G. Salton, ed. Chapter 14. pp 313-323. Prentice-Hall.

[RB99] I. Ruthven and P. Borlund. *Information use and relevance assessments.* (in preparation). 1999

[RL99a] I. Ruthven and M. Lalmas. Selective relevance feedback using term characteristics. CoLIS 3, Proceedings of the Third International Conference on Conceptions of Library and Information Science. Dubrovnik. 1999. *(to appear)*

[RL99b] I. Ruthven and M. Lalmas. Integrating information use into relevance feedback. *Submitted to European Conference on Digital Libraries (ECDL '99).* Paris. 1999.

[SW99] A. Spink, and T. D. Wilson. *Toward a theoretical framework for information retrieval (IR) evaluation in an information seeking context*. In [DDRvR99]