

Towards an Agent Based Retrieval Engine (Profile - Information Filtering Project)

B.C.M. Wondergem*, P. van Bommel, T.W.C. Huibers, Th.P. van der Weide
Computing Science Institute
University of Nijmegen
The Netherlands

November 1996

Published as: B.C.M. Wondergem, P. van Bommel, T.W.C. Huibers, and Th. van der Weide. Towards an Agent-Based Retrieval Engine. In J. Furner and D.J. Harper, editors, *Proceedings of the 19th BCS-IRSG Annual Colloquium on IR Research*, pages 126–144, Aberdeen, Scotland, April 1997. Robert Gordon University.

Abstract

This article describes and analyses the retrieval component of the Profile Information Filtering Project of the University of Nijmegen. The overall structure of this project, serving as the context for the retrieval component, is stated. This component is called the Retrieval Engine and will be implemented as an intelligent retrieval agent, using sophisticated techniques from artificial intelligence. A synthesis between information retrieval and information filtering has to be found, coping with challenging problems stemming from the combination of the difficulties of both fields. The Retrieval Engine should be capable of giving an explanation of why a document was found relevant to the information need of the user. The techniques used will rely on sophisticated natural language processing. The techniques to establish relevance degrees for documents will consist of two parts: a symbolic and a numeric one. This allows for a mechanism that is both explainable and exact. Interesting approaches for obtaining this are stated.

*Dept. of Information Systems, Faculty of Mathematics and Computing Science, University of Nijmegen, Toernooiveld, NL-6525 ED Nijmegen, The Netherlands, E-mail: bernd@cs.kun.nl

Contents

1	Introduction	3
2	The Project Context	4
3	The Retrieval Engine	5
3.1	Introduction to the Retrieval Engine	5
3.2	Overall conceptual model	6
3.3	The agent space	7
4	Research Questions	9
4.1	Theoretical Foundations	10
4.2	Other components of the Profile project	10
4.3	Performance Measures	11
5	Approach	11
5.1	Underlying Frameworks	11
5.2	Approach to specific research questions	12
6	Project Management	13
6.1	Organizational topics	13
6.2	Milestones	14
7	Conclusions	14

1 Introduction

Information Retrieval (IR) is presented with new challenges stemming from the exploding growth of the amount of online available information. The sheer amount of this information as well as its dynamical nature, with respect to the place and contents, forces us to reconsider the concepts underlying our field of research.

Lately, the IR community has spent a lot of attention on Information Filtering (IF) (see e.g. [BC92] and [MS93]), which can be seen as a problem dual to IR. Both IR and IF consider an unstructured information space and users having specific information needs. But whereas IR deals with a stable information space and users having varying information needs, IF deals with a dynamic information space and users who have a relatively stable information need. The IF problem can thus be stated as follows:

Considering a number of dynamic information objects, the IF system matches the characterizations of the information objects against the user profiles, descriptions of the users' information needs, to obtain a relevance estimate of the information objects with respect to the information needs.

The information filtering project *Profile* (see [HSB⁺96]) will deal with a dynamic information space and with users who may have stable as well as dynamic information needs. It thus faces the combined difficulties encountered in the fields of IR and IF. The *Retrieval Engine*, the actual information disclosure system within the Profile project, will use results from IR and IF, whereas also new techniques will have to be developed. The approach will be embedded in a clear overall conceptual model.

As implied by previous statements, the amount of available information exceeds the capacity with which a single user can cope. The Profile system should anticipate user's interests and preferences and, by using them, allow the user a controlled access to the relevant part of the information. It should act as an intermediate between the user(s) and the information objects. Also it should be autonomous and pro active so as to spare the user in her quest for information.

Next to this, the Retrieval Engine should be capable of giving an *explanation* of why a certain information object is considered to be relevant to the user's information need. The explanation will be based on the process with which relevance is inferred or relevance estimates are computed. This process will use sophisticated techniques from artificial intelligence (AI). Techniques from this field will also be used for learning and adaptation with respect to the interests of the user and the content of the information sources.

Yet another aspect of the Profile project is the use of query and characterization languages that are more towards natural language than those of traditional approaches. We will use *noun phrases*, as opposed to keywords, as the basic building blocks for our languages. Most keyword-based retrieval systems seem to have reached their limit with respect to performance, as measured in terms of recall and precision. We believe that using a representation that is closer to natural language will improve system performance as well as user satisfaction. In Profile, natural language processing (NLP) techniques will play an important role.

Another component within the Profile project will aim at providing an effective *interface* with the user. This component consists of researchers in the field of cognitive ergonomics. User modeling is also taken care of by researchers from the corresponding field. Thus, within the Profile project a number of components are identified, each contributed by experts of that particular line of research.

The main theme of the filtering project Profile is to provide an effective intermediary between the information sources and the information users. The term information filtering is used in the context of the Profile project where information discovery more clearly describes the intuition behind it.

The structure of this position paper is as follows. Section 2 will outline the basic structure of the Profile project, describing its components as well as the research groups involved. Section 3 will elaborate further on the difficulties which will be encountered for the Retrieval Engine. This leads to a number of research questions which will be stated in Section 4. Section 5 deals with possible approaches to overcome the difficulties in order to answer the research questions. Section 6 presents the main project management aspects, while section 7 provides some concluding remarks.

2 The Project Context

This section describes the organizational structure of the Profile information filtering project, as well as the research groups involved.

The organizational structure of the Profile Filtering Project can be depicted as is done in Figure 1. Two research groups of the University of Nijmegen are involved:

- The Cognitive Ergonomics Group of the NICI, the Nijmegen Institute for Cognition and Information
- The Software Engineering Group of the CSI, the Computing Science Institute

NICI		CSI
Modeling (Hoenkamp)	\Rightarrow <i>profile</i> \Rightarrow	Parsing (Koster)
\uparrow <i>user behaviour</i> \uparrow	'Profile Filtering Project'	\downarrow <i>descriptors</i> \downarrow
Interaction (Schomaker)	\Leftarrow <i>documents</i> \Leftarrow	Retrieval (vd Weide)

Figure 1: Organizational structure

As Figure 1 shows, there are four main sections or phases in the project. The sections are tagged with the researcher in charge and will be briefly touched upon below. The modeling phase focuses on formulating the information need of the user in the form of phrases in natural language. Then, those phrases are parsed, resulting in a profile in the form of a collection of frame phrases. The information objects are also parsed, resulting in the characterizations of the information objects. In the retrieval phase, the profile is used to establish relevance estimates of the documents with respect to the information need of the user. The relevant documents are passed to the interaction phase to be presented to the user, after which the

user can provide relevance feedback on the presented document ranking. This information will be used to adjust the user profile, thus starting a new cycle.

3 The Retrieval Engine

This section will elaborate on the aspects of the Retrieval Engine within the Profile project. First of all, the overall model is described, followed by a detailed description of the context of the Retrieval Engine.

3.1 Introduction to the Retrieval Engine

As stated in the introduction, the Retrieval Engine will have to be an autonomous, intelligent and pro active intermediary between the information sources and the user(s). Furthermore, the system will have to be reactive, in the sense that it should be able to react upon events from the information sources or user community. For instance, upon the event of new information becoming available, the Retrieval Engine should decide whether to pass this information on to the user. Or, upon a specific user request (query) it should react in such a way as to satisfy this request and provide the user with relevant information.

The above features call directly for an *agent-based* approach, since they are considered essential to the notion of agency as described in [WJ95]. More about this in subsection 3.3. For information about intelligent agents see e.g. the aforementioned article as well as [HL96], which describes an IR application of agents, and [vL96], which provides a formal framework for intelligent agents based on modal logics.

The Retrieval Engine will therefore be developed as an intelligent retrieval agent, incorporating the above mentioned characteristics, as well as being able to communicate with other agents. Agents can be described in a top down decomposing way, resulting in a number of sub agents. Communication can also take place between sub agents. In satisfying a certain information need, agents will collaboratively act in order to diminish the amount of work that is done more than once. The agents will communicate according to a protocol for retrieval agents, which has to be developed. Later in this section we will return to retrieval agents.

Within the Profile project, we consider three main spaces of interest: an *information space*, consisting of the information sources, an *agent space*, consisting of the retrieval agents and possibly other agents, and a *user space*, modeling the users and groups of users. This, together with the communication between the three spaces, is depicted in Figure 2.

We will now shortly describe the three spaces considered. The *user space* consists of all the users of the system together with the information kept about the users. This space is of major concern to the user modeling component and also to the interaction component since the users will be able to give relevance feedback on the presented documents.

The *information space* is mainly dealt with by the parsing component of Profile. It consists of the streams of information, or, the information objects. The information sources may contain structured, semi-structured or unstructured information. Also, it will embody different layers of abstraction, obtaining a stratified hypermedia, i.e. a generalization of the two-level hypermedia proposed in [BW90]. The content of the information objects will be parsed, obtaining information on which the layers can be based.

The *agent space* will consist of the agents used in the Profile system. The Retrieval Engine will thus consist of a number of retrieval agents. As the communication arrows in the picture show, the agents form intermediates between the information sources and the users. They

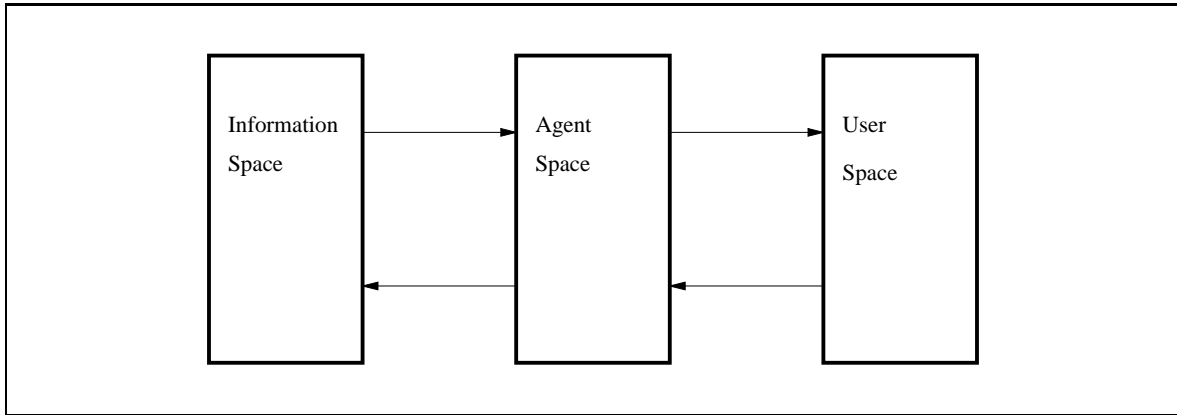


Figure 2: The three conceptual spaces

also show that, in the Profile system, no direct communication between the user space and the information space is possible.

3.2 Overall conceptual model

We will now zoom in on the three spaces and describe them in more detail, as visualized in Figure 3.

The *user space* deals with the users of the system. Each circle represents a single user. Every user has a corresponding profile, describing the interests of the user and depicted by a square attached to the user's circle. Actually, the profile will consist of a number of sub-profiles, each describing a particular field of interest. The profiles will be equipped with an indication about the amount of time or the number of times the user has been interested in the topics, thus allowing a distinction to be made between stable and short term information needs. A number of users having a common interest can be formed into a group, depicted by a rectangle around the squares of the individual user profiles. Note the visualized correspondence between users and groups of users on one hand and profiles and group profiles on the other hand. Each group has a group profile which is based on the profiles of the individual users of that group. A user can participate in any number of groups and groups may overlap. Also, several groups can be unioned to form yet another group, thus obtaining a hierarchy of groups. We thus obtain a very general notion of groups.

The *information space* contains the information sources, which may be streams of information objects or static or dynamic collections of information objects. If an information stream or a dynamic collection of information objects is considered at a specific moment in time, it can be assumed to have some properties of a static collection. The information objects can be viewed at different levels of abstraction, corresponding to different ways of characterizing the information objects. Based on the different characterizations, a stratified hypermedia is formed. At the moment, it is not clear yet how many layers will be needed in the hypermedia nor what the exact form of the layers should be.

The *agent space* consists of the agents that form the Retrieval Engine. It will thus also consist of the sub agents, but those are not modeled at this level of abstraction. Shortly, we will describe a top down decomposition method for retrieval agents. It is possible for the agents to communicate with each other. Also, agents can communicate with the entities in the information space as well as with those in the user space. For example, an agent can have

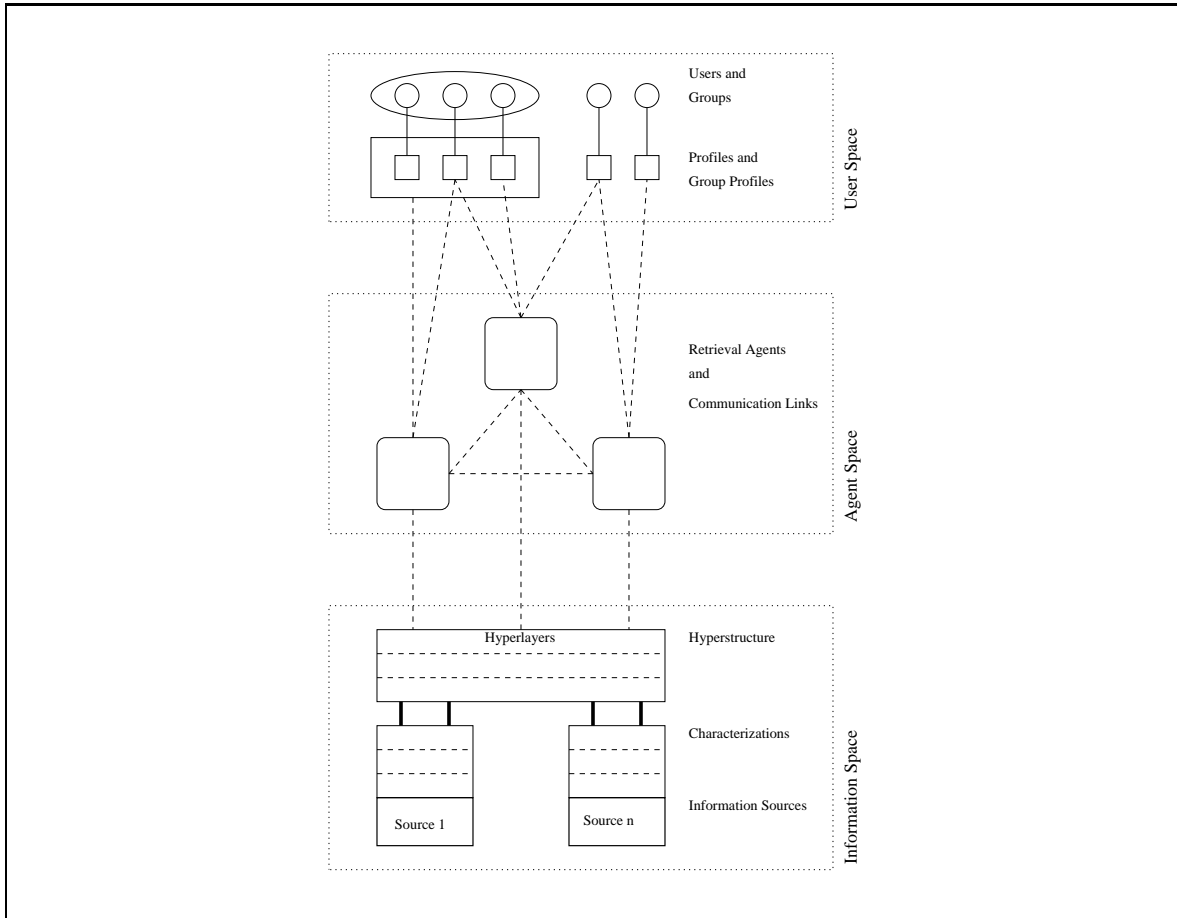


Figure 3: Overall conceptual model of an information disclosure system

noticed that some new information objects have become available and can signal this to a number of (groups of) users.

We will investigate the extent to which concepts of the information space can be projected onto the user space and vice versa. For instance, the layered structure of the documents and their characterizations can be applied to users and their profiles as well. This will lead to a clear overall conceptual structure of the 2 spaces.

Not all the aspects described here have to be available in a certain agent based information disclosure system. We merely state that the entities described here and the relations between them *could* be viable parts of the system.

3.3 The agent space

At the moment, there is no standard about the characteristics an entity should possess in order to be called an agent. We shall state five features which we consider necessary for an entity to be called an agent. They correspond to the strong notion of agency as stated in [WJ95]. We consider those features necessary for an entity to be allowed to reside in the agent space.

1. Autonomy

An agent should be an autonomous entity, capable of performing a task on its own. In performing this task, of course, the agent may have to rely on other entities.

2. Intelligence

An agent should be intelligent in the sense that its reasoning capabilities have some human characteristics. To ensure this, techniques from AI will be used, for instance in the inference process to establish relevance of a document with respect to an information need.

3. Pro activeness

We will call an agent pro active if it is able to take the initiative itself, not relying on events in the user space or information space. The direct reason for an agent to act pro actively can stem from the internal state of the agent or from communication between agents. For example, an agent can request another agent to search for specific documents, or, the clock in the internal state of the agent can indicate that the time is ready for a particular event.

4. Reactiveness

Reactiveness is the ability to act when prompted by another entity. For instance, when a user places a request, the agent should process it. Or, when an information source presents an agent with new information, the agent should decide whether or not to send this information on to the attached users.

5. Capability to communicate

An agent should be able to communicate with other agents as well as with some other entities, e.g. the information sources and users. Inter agent communication will take place conform a communication protocol for retrieval agents.

The above characteristics apply to a general notion of agents. We will now focus on retrieval agents, the task of which is to deliver relevant information objects to a user.

A number of additional attributes of agents are identified in [WJ95] that also apply to IR agents. An IR agent will adhere to the assumption of veracity, that is, it will not communicate false information deliberately. Another aspect of IR agents is rationality, i.e. a way of acting directed to achieve its goals. IR agents will also be virtually mobile, in the sense that it will appear that the agent is capable of moving around in a network. At the moment, it is not clear whether the retrieval agents conform to the principle of benevolence, that is, if they will always do what is asked for since they do not have conflicting goals.

In order to fulfill his task, a retrieval agent should first gather information objects, then match those with a number of profiles, base a relevance estimate on this match and finally pass the (most) relevant information objects on the interaction component. We can thus functionally decompose a retrieval agent into three sub modules as to the tasks it has to perform, as shown in Figure 4. We do not adapt the term sub agents for the modules since the modules do not adhere to all the characteristics of agency.

The *gathering module* supplies the matching module with information objects. Those can be newly arrived objects or an actively gained approximation of the information need of the user. In actively gathering information objects, constraints with respect to time of creation, size and ease of accessibility can be taken into account. More generally, the gathering module can exploit constraints that do not consider an interpretation of the actual content of the

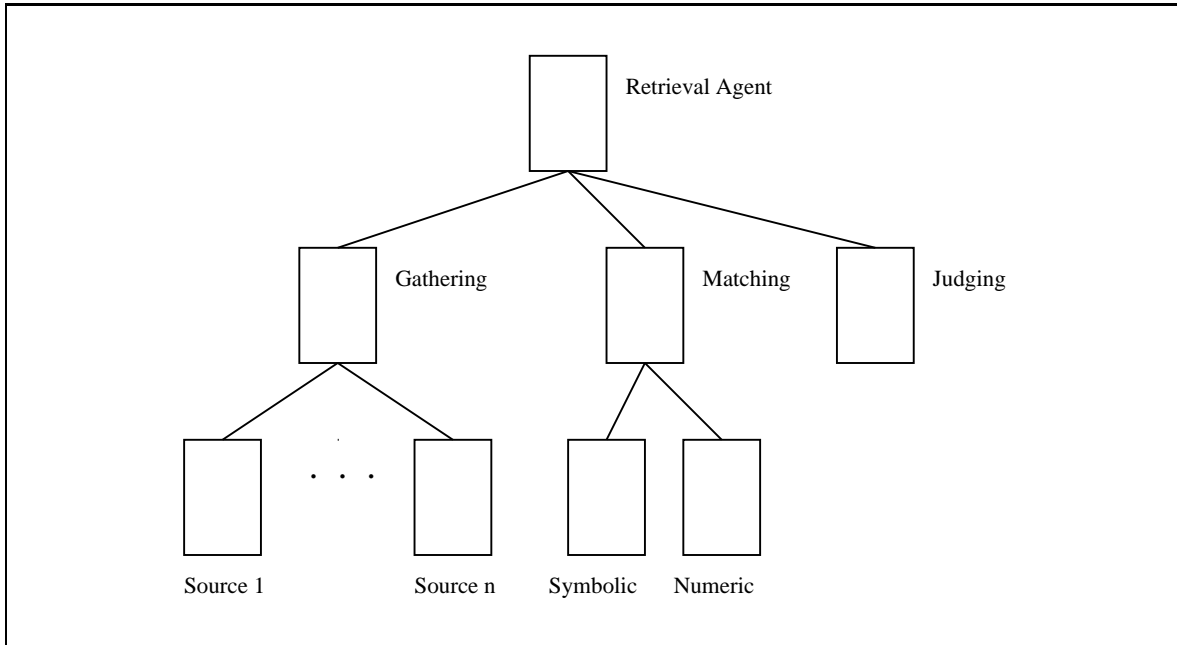


Figure 4: A retrieval agent

information object. The gathering module is further decomposed with respect to the different information sources it has access to. Gathering modules can be further decomposed to the degree of abstractness with which they access the information source. However, to avoid becoming lost in too much detail, we have not done so in Figure 4.

The *matching module* establishes a relevance estimate of the incoming information objects with respect to the user profile. The matching module has two sub modules that together form the matching process. On the one hand, there is the symbolic matching module, providing a qualitative inference process. On the other hand, the numeric matching module implements a quantitative matching process. Both matching processes are combined by the matching module to obtain an explainable and precise relevance estimate. The matching module may communicate with other entities to obtain e.g. domain specific knowledge, if available.

The *judging module* produces a ranking of the documents as to their relative relevance based on the relevance estimates computed by the matching module. The most relevant documents will be presented to the interaction component. Optionally, this module annotates documents with further information needed by the interaction component. As with the gathering modules, the judging module can be decomposed further as to the different output formats needed. The judging module can also filter out questionable information.

4 Research Questions

The previous sections described the global outline of the Profile project, and of the Retrieval Engine in particular. We will now state a number of research questions concerning the Retrieval Engine that will have to be answered during the project. The questions are divided into three categories: theoretical foundations, interaction between the different components of the project and performance measures. Section 5 will describe possible approaches to the

questions mentioned in section 4.1.

4.1 Theoretical Foundations

The main question about the implementation of the Retrieval Engine is in what way to establish the relevance of a document with respect to the information need of the user. Several techniques have been developed, for example the Vector Space Model, Logical Models and Probability based models (see e.g. [Rij75] and [Bla90]). However, at the moment, those models cannot cope with advanced underlying languages based on natural language.

With respect to the techniques to be developed, the following questions are stated:

- Which characteristics of relevance techniques can be identified and which of those are important in Profile?
- In what way can the relevance techniques be incorporated in the theory about intelligent agents, obtaining a theory for intelligent IR agents?
- What kind of protocol should the IR agents use for communication?
- How well are the techniques suited for a dynamic environment?

For instance, one characteristic of the approach taken should be that it is possible for the agent to explain why a certain document was found relevant. This demands a non-numerical approach since the process cannot be effectively explained in terms of numbers only. Another aspect is that relevance estimates should be precise and take uncertainty into account, thus demanding a numerical approach. Research should thus be put in combining both aspects.

4.2 Other components of the Profile project

The cooperation between the retrieval component and the other components (see Figure 1) gives rise to additional technical questions, which are described below.

Within the Profile project, a particular component should take the other components into account as well. The retrieval component, for instance, should wonder what kind of information is to be delivered and in what form to the interaction component. Also, it is important to know what kind of information is available from the other components and whether there is a way in which this information can be properly used. A general research question that stems from these considerations can be stated as: can one component offer what the others need and properly use what the others offer? To investigate this, a requirements analysis of the different components will be made.

On the one hand, the retrieval component should provide the interaction component with the relevant documents, possibly accompanied by a degree of relevance. Also, additional information will have to be sent along, allowing for a better representation of the relevant information. This demands capabilities for automatically annotating documents. What the exact nature and form of this information should be is another research question.

On the other hand, the Retrieval Engine has an interface with the parsing module. Consequently, the output produced by the parsing module must be dealt with appropriately. This means, for example, that all the available information should be used properly. The output of the parsing module should also match the demands from the Retrieval Engine.

Also, although there is no distinct interface with the user modeling module, the retrieval engine should be able to cope with all user properties that have been modeled. This requires knowledge of the philosophy behind the user modeling module.

4.3 Performance Measures

Finally, an interesting question is which approaches for obtaining relevance estimates aimed at in subsection 4.1 result in a significant improvement of performance. In order to be able to answer this question, one should first decide what performance measures to use. The traditional notions used in IR, recall and precision, will not be directly applicable since they are not suitable for dynamic environments. Also, the generally accepted definitions of recall and precision do not allow for a non-binary notion of relevance. Recall and precision assume that a document is either relevant or not, whereas degrees of relevance, i.e. a non-binary notion, will provide a mechanism better capable of approximating the granularity in the user's information need. Research on evaluation frameworks for interactive multimedia information retrieval applications is also described in [Gre].

The techniques to be developed should result in a performance improvement. Different performance measures will have to be developed that can cope with a dynamic environment and a non-binary notion of relevance. Those might be based on the idea that the preferences which reside in the cognitive search model of the user impose an ordering on the set of information objects (see e.g. [BH96] and [Won96]). Performance measures should then establish a similarity measure between the relevance estimates produced by the Retrieval Engine and this ordering which is obtained by the user modeling component.

5 Approach

This section describes the underlying frameworks we will consider for implementing the Retrieval Engine. Also, approaches for the explicit research questions formulated in section 4.1 are described.

5.1 Underlying Frameworks

One of the characteristics of the Profile project is that natural language processing (see e.g. [Sme92]) will play an important role. For more information on the use of natural language for information filters see e.g. [Ram92]. The underlying query and characterization language will be based on noun phrases, a concept more towards natural language than keywords. This will be of major concern to the parsing component of the Profile project (see [AvBKvdW97]).

The techniques for determining the degree of relevance of information objects with respect to an information need will consist of a symbolic (qualitative) as well as a numeric (quantitative) part in order to combine the virtues of both approaches. The numerical part of the matching process should provide granularity in the (binary) inference process of the qualitative part.

We will consider two formal symbolic frameworks:

- Preferential Structures

Preferential Structures were proposed as a meta theory for non monotonic logics in [KLM90] and exploited for IR in e.g. [Bru96], [BH95b], [WHHW96] and [Won96]. They

are based on a set of possible worlds, which will be seen as documents, and a preference relation on those worlds. A formal proof system can be defined with them, inferring whether a document is relevant or not. Preferential Structures are well suited to model user preferences on information items and exploit them via plausible inference. Those preferences are to be obtained by the user modeling and interaction components of the Profile project.

- Situation Theory

Situation Theory (see e.g. [BE90], [Bar92]) has been proposed as a formal theory for information. Contrary to most formal logics, its most basic notion is information, rather than to truth. Situation Theory has been adapted for IR in e.g. [HLR96], [BH95a] and [HB94]. Lalmas has taken the approach of combining Situation Theory with a numerical component in [Lal96]. Huibers describes an axiomatic theory of IR using Situation Theory in [Hui96].

As numeric frameworks we will consider:

- Bayesian Probability Theory, adapted to noun phrases, can be used to compute the probability that a document is relevant given a certain profile or request.
- Conceptual graphs (see e.g. [Sow84]) can be used for the modeling of noun phrases. After this has been done, known similarity measures for conceptual graphs (see e.g. [MLL92] and [Mah93]) can be used to define a similarity measure between noun phrases and thus for similarity measures between documents and a query or profile.
- A Phrase Space, an adaptation of the Vector Space Model to noun phrases, can be developed to obtain a measure of similarity between vectors of noun phrases. Note that this framework hinges on a similarity measure between noun phrases. Research must show whether this framework can effectively be combined with a symbolic mechanism.

5.2 Approach to specific research questions

This section elaborates further on the research questions stated in section 4.1 and will indicate possible directions for finding solutions to them.

Lalmas has stated a number of characteristics a logic, or a formal framework, for IR should have (see [Lal96]). We believe those are very useful characteristics and will describe them briefly. *Imprecision* concerns the meaning of a disjunction. In IR a disjunction should only be true if it is known that one of the components is true but it is not known exactly which. The *flow of information* is about inferring or deriving implicitly available information. *Partiality* concerns the fact that the concepts we actually work with are only partial descriptions of the corresponding real world entities. For example, the characterization of a document is only a partial description of the original information content. *Uncertainty* is a natural feature in reasoning about or with information. The framework which specifies the inference process should therefore be able to specify and properly deal with (degrees of) uncertainty. *Intensionality* means that the meaning of an information item can be context dependent. *Significance* allows us to specify that certain information items are more important than others. *Informative Relationship* means that only information items that are in some way related should be combined. Next to the mentioned characteristics the formal framework should be expressive

enough as well as intuitively understandable to build an explanation mechanism on top of it. In addition, it should be possible to make the framework suitable for the use of noun phrases.

One of the aims of the Profile project is to implement the techniques for determining the relevance of a document in the form of an intelligent IR agent. It should thus be investigated to what extent these techniques can be integrated in an agent based approach. The techniques should not interfere with the necessary characteristics for agency.

Apart from communication between agents, the agents can also communicate with entities from the user and information space. The user modeling and interaction phase will be majorly concerned with agent-user interaction. The Retrieval Engine will mostly be concerned with communication between agents, for which a protocol should be designed. This protocol should be easily linkable with the matching process. Also, it should have a formal basis, enabling certain properties to be rigorously proven. Those can concern user privacy and copyright issues as well as issues related to conflicts between agents. The communication between agents will be lead by the agent's knowledge and beliefs and establishes knowledge about the other agent's knowledge. The agent's reasoning capabilities should thus include a meta notion of knowledge.

Since the project should eventually lead to an application on the World Wide Web, which is a very dynamic collection of information objects, the techniques that are developed should be able to cope with this feature. The dynamics of an information object can be seen in two ways: its place may change as well as its contents. Also, new information objects may appear just as existing ones may disappear. Note that a changing contents of an information object may also require a new characterization to be computed.

6 Project Management

This section will elaborate on the organizational topics involved in the Profile project and will state a number of milestones on which a schedule is based.

6.1 Organizational topics

The researchers involved in the Profile project are divided into 3 groups. The 'Stuurgroep', consisting of the researchers in charge of the different components, will take care of setting the directions and will look after the project as a whole. The PhD students involved constitute the research team and will do the actual research, both as a team and as individual researchers. Last, a supporting team of additional researchers will support the research team in their day to day progress.

The different groups will attend to their corresponding meetings and provide feedback to the other groups as well as to other interested persons. The stuurgroep's meetings will take place once every 3 months. The research team will have an official meeting every week, next to informal meetings when considered necessary. Plenary meetings will be held every 6 months.

Another point of concern is language problems. The members of the research team come from different countries and therefore cannot communicate in their native language. The different backgrounds of the researchers will also cause language problems, since different terminology is used. Therefore, intensive communication between the members of the research team is needed, requiring good abilities to communicate field specific research problems in terms that are understood by the other researchers.

6.2 Milestones

Figure 5 shows a timetable with milestones for the development of the Retrieval Engine in the Profile project. Each period will consist of 6 months.

Period	Product
1	Position Paper Requirements Specification for the Retrieval Engine Refined Overall Project Plan (cooperation) Demonstrator (cooperation)
2 & 3	First version of Retrieval Engine satisfying minimal requirements Experiments and testing
4	Evaluation and reconsideration of requirements Experiments Plan for adaptation of the Retrieval Engine Selection of level of ambition from Requirements Specification
5 & 6	Final version of the Retrieval Engine Experiments and testing
7	Evaluation and last bits
8	Finishing and integration thesis

Figure 5: Preliminar Schedule

The period of 4 years that the project will take is roughly divided into 2 major sections. At the end of the first 2 years, the initial version of the Retrieval Engine will be ready. In order to do so, a position paper is written first, clearly defining the problem area and stating possible approaches. Also, in cooperation with the other researchers, the overall project plan is refined, based on the individual position papers. In the first period of 6 months, the requirements specification for the Retrieval Engine will be stated leading to a demonstrator which will be constructed in cooperation with the other researchers. After that, the first version of the Retrieval Engine, satisfying minimal requirements, will be developed and tested. The first major period ends with an evaluation of the initial version of the Retrieval Engine, based on extensive experiments and testing. From this, the requirements will be reconsidered and a plan for adaptation will be produced including the choice of a level of ambition from the requirements specification.

The second major period will result in the final version of the Retrieval Engine, based on the first version and the renewed requirements specification. This version will also be extensively tested and evaluated. Finally, the results and considerations of the project will be integrated in a thesis.

7 Conclusions

The Information Filtering Project Profile at the University of Nijmegen explores the combined areas of IR and IF. A highly dynamic information space and varying user interests are complicating factors as well as great challenges in this research project.

The Retrieval Engine, the actual information disclosure system within the Profile project, will be embedded in the theory of intelligent agents. For its implementation, a combination

between symbolic and numeric mechanisms is sought. The Retrieval Engine will be implemented using techniques from Artificial Intelligence which have to be equipped with a notion of noun phrases.

The Profile project involves two research departments of the University of Nijmegen. Barriers will have to be broken in order to obtain the right environment for cooperative work. We are convinced that the structure of the Profile project depicts a promising combination of fields of research.

References

- [AvBKvdW97] A.T. Arampatzis, P. van Bommel, C.H.A. Koster, and Th.P. van der Weide. Linguistic Variation in Information Retrieval and Filtering. Technical report CSI-R9701, University of Nijmegen, Nijmegen, The Netherlands, January 1997.
- [Bar92] J. Barwise. Constraints, channels and the flow of information. In *Situation Theory and its Applications*, volume 3 of *CSLI Lecture Note Series*. CSLI, 1992.
- [BC92] N.J. Belkin and W.B. Croft. Information filtering and information retrieval: Two sides of the same coin? *Communications of the ACM*, 35(12):29–38, December 1992.
- [BE90] J. Barwise and J. Etchemendy. Information, Infons, and Inference. In R. Cooper, K. Mukai, and J. Perry, editors, *Situation Theory and its Applications*, volume 1 of *CSLI Lecture Note Series*, pages 33–78. CSLI, 1990.
- [BH95a] F.C. Berger and T.W.C. Huibers. A framework based on situation theory for searching in a thesaurus. *Journal of Text and Document Management*, 1:253–276, 1995.
- [BH95b] P.D. Bruza and T.W.C. Huibers. How nonmonotonic is aboutness? Technical Report UU-CS-1995-09, Department of Computer Science, Utrecht University, The Netherlands, March 1995.
- [BH96] P.D. Bruza and T.W.C. Huibers. A Study of Aboutness in Information Retrieval. *Artificial Intelligence Review*, 10:1–27, 1996.
- [Bla90] D.C. Blair. *Language and Representation in Information Retrieval*. Elsevier, 1990.
- [Bru96] P.D. Bruza. Intelligent Filtering using Nonmonotonic Inference. In *Proceedings of the 1th Australian Document Computing Symposium*, pages 1–7, Melbourne, Australia, March 1996. Department of Computer Science, RMIT.
- [BW90] P.D. Bruza and Th.P. van der Weide. Two Level Hypermedia - An Improved Architecture for Hypertext. In A.M. Tjoa and R. Wagner, editors, *Proceedings of the Data Base and Expert System Applications Conference (DEXA 90)*, pages 76–83, Vienna, Austria, 1990. Springer-Verlag.

- [Gre] IMAG Grenoble. MIRA - Evaluation Frameworks for Interactive Multimedia Information Retrieval Applications. Technical report. Available from: <http://www.docs.dcs.gla.ac.uk/mira/>.
- [HB94] T.W.C. Huibers and P.D. Bruza. Situations, a General Framework for Studying Information Retrieval. Uu-cs-1994-15, Utrecht University, March 1994.
- [HL96] T.W.C. Huibers and B. Linder. Formalising Intelligent Information Retrieval Agents. In *Proceedings of the 18th British Computer Society Annual Information Retrieval Colloquium*, pages 125–143, Manchester, England, 1996. Manchester Metropolitan University.
- [HLR96] T.W.C. Huibers, L. Lalmas, and C.J. van Rijsbergen. Information retrieval and situation theory. Technical Report 1, New York, New York, 1996.
- [HSB⁺96] E. Hoenkamp, L. Schomaker, P. van Bommel, C.H.A. Koster, and Th.P. van der Weide. Profile - A Proactive Information Filter. Technical Note CSI-N9602, Computing Science Institute, University of Nijmegen, Nijmegen, The Netherlands, 1996.
- [Hui96] T.W.C. Huibers. *An Axiomatic Theory of Information Retrieval*. PhD thesis, Department of Computer Science, Utrecht University, November 1996.
- [KLM90] S. Kraus, D. Lehmann, and M. Magidor. Nonmonotonic Reasoning, Preferential Models and Cumulative Logics. *Artificial Intelligence*, 44:167–207, 1990.
- [Lal96] M. Lalmas. *Theories of Information and Uncertainty for the modelling of Information Retrieval: an application of Situation Theory and Dempster-Shafer's Theory of Evidence*. PhD thesis, Department of Computer Science, University of Glasgow, April 1996.
- [Mah93] P.E. Maher. A Similarity Measure for Conceptual Graphs. *International Journal of Intelligent Systems*, 8:819–837, 1993.
- [MLL92] S.H. Myaeng and A. Lopez-Lopez. Conceptual graph matching: a flexible algorithm and experiments. *Journal of Experimental Theoretical Artificial Intelligence*, 4:107–126, 1992.
- [MS93] M. McElligott and H. Sorensen. An emergent approach to information filtering. *Computer Science Journal*, 1(4), December 1993.
- [Ram92] Ashwin Ram. Natural language understanding for information-filtering systems. *Communications of the ACM*, 35(12):80–81, December 1992.
- [Rij75] C.J. van Rijsbergen. *Information Retrieval*. Butterworths, London, United Kingdom, 1975.
- [Sme92] A.F. Smeaton. Progress in the Application of Natural Language Processing to Information Retrieval Tasks. *The Computer Journal*, 35, June 1992.

- [Sow84] J.F. Sowa. *Conceptual Structures: Information Processing in Mind and Machine*. Addison-Wesley, Reading, Massachusetts, 1984.
- [vL96] B. van Linder. *Modal Logics for Rational Agents*. PhD thesis, Department of Computer Science, Utrecht University, The Netherlands, june 1996.
- [WHHW96] B.C.M. Wondergem, W. van der Hoek, T.W.C. Huibers, and C. Witteveen. Preferential Semantics for Query by Navigation. In K. van der Meer, editor, *Informatiewetenschap 1996*, Delft, The Netherlands, December 1996.
- [WJ95] M. Wooldridge and N.R. Jennings. Intelligent Agents: Theory and Practice. *Knowledge Engineering Review*, 10(2):115–152, 1995.
- [Won96] B.C.M. Wondergem. Preferential Structures for Information Retrieval. Master’s thesis inf-scr-96-21, Department of Computer Science, Utrecht University, Utrecht, The Netherlands, 1996.