Linguistic Variation in Information Retrieval and Filtering

A. Arampatzis*, P. van Bommel, C.H.A. Koster, Th.P. van der Weide

Technical Report CSI-R9701, January 1997

Abstract

In this paper, a *natural language* approach to Information Retrieval (IR) and Information Filtering (IF) is described. Rather than keywords, *noun-phrases* are used for both document description and as query language, resulting in a marked improvement of retrieval precision. Recall then can be enhanced by applying normalization to the noun-phrases and some other constructions. This new approach is incorporated in the Information Filtering Project *Profile*. The overall structure of the Profile project is described, focusing especially on the Parsing Engine involved in the natural language processing. Effectiveness and efficiency issues are elaborated concerning the Parsing Engine. The major contributions of this research include properties of grammars and parsers specialized in IR/IF (properties such as coverage, robustness, efficiency, ambiguity), normalization of noun-phrases, and similarity measures of noun-phrases.

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

Contents

1 Motivation

The tremendous increase of networked information has led to a new challenge in "information seeking". Currently, users everyday confront themselves with large amounts of information in the form of news, mail messages, and especially World-Wide Web pages. Although users of this electronic information have access to a rich body of information, only a small fraction of this is actually relevant to the interest of any particular user. In order to reduce the effort of a user determining which information is relevant to his needs, an automatic solution seems indispensable. Assuming specific long-term interests of a user, and taking into account that the dynamic and unstructured information sources have a high modification rate, this information filtering problem differs from the classical information retrieval problem ([BC92]). However, many of the techniques used for information retrieval can easily apply to filtering and vice versa.

Due to the fact that the largest amount of this information consists of text documents, many approaches have been seen in text filtering which all have in common four basic components:

- a technique for representing documents
- a technique for representing the information need (*profile or query*)
- a way of comparing profiles/queries to document representations
- ways of using the results of comparison (rendering, presentation, interaction and feedback)

Figure 1 illustrates the representation and comparison process implemented by text filtering systems. A framework for the text filtering problem can be found in [OM96].

The state-of-the-art text filtering/retrieval systems are mostly based on use of keywords, both in representing information objects and as a basis for the retrieval language (expressing the information need). The possibility for further improvement in precision and recall based on keywords is rather marginal. Besides, the use of keywords is inadequate for more inflected language than English. Citing C.J. van Rijsbergen [Rij79]:

> *"A big question, that has not yet received much attention, concerns the extent to which retrieval effectiveness is limited by the type of document description used. The use of keywords to describe documents has affected the way in which the design of an automatic classification system has been approached. It is possible that in the future, documents will be represented inside a computer entirely differently."*

In the information filtering project *Profile* ([HSB+96]) linguistic techniques will be used for characterizing documents and for formulating user profiles and queries. For other filtering approaches based on natural language see e.g. [Ram91] and [Ram92]. The approach described in [SO95] also comprises a natural language processor. Our approach is based on the use of *noun phrases* (NPs) instead of keywords. For previous work related to the use of NPs in information retrieval, refer to e.g [AT96], [ATK96]. The same linguistic techniques incorporated in Profile can easily be adapted to other information retrieval and filtering systems.

In order to solve the problem of accessing natural media like audio or video, the presence of textual annotation of these media is assumed. This may be in the form of text summaries or abstracts. Such media will be called *annotated natural media*, and by using annotation the general information filtering problem is related to the text filtering problem.

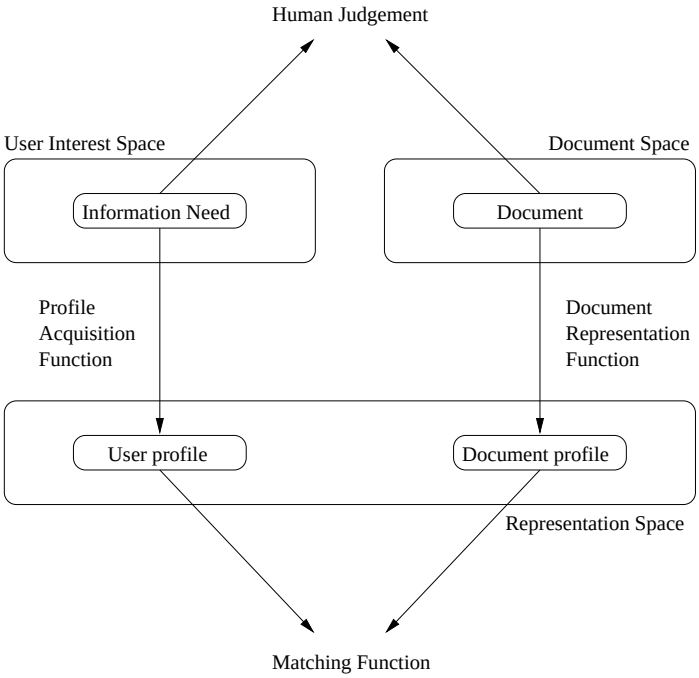

Figure 1: Text filtering model

2 The Information Filtering Project "Profile"

2.1 Organizational Structure

The information filtering project Profile is conducted in cooperation between two research groups of the Nijmegen University:

- The Software Engineering group of the Computer Science Institute (CSI), and
- The Cognitive Ergonomics group of the Nijmegen Institute for Cognition and Information (NICI).

The collaboration of these groups may be visualized as in figure 2. The figure splits the Profile

NICI		CSI
Modeling	$\Rightarrow$ *profile* $\Rightarrow$	**Parsing**
$\Uparrow$ *user behavior* $\Uparrow$	Filtering project **Profile**	$\Downarrow$ *descriptors* $\Downarrow$
Interaction	$\Leftarrow$ *documents* $\Leftarrow$	**Retrieval**

Figure 2: Organizational and Functional structure

project into four sub-projects and also delimits the different areas of research.

2.2 Functional Structure

Goals and interests of an individual user or a group of users are used to build a *profile*. Profiles are utilized to support users in formulating queries, to better understand the meaning of them, and for comparison with documents in order to filter out irrelevant information. The *User Modeling* module is responsible for providing these profiles, by building a mechanism to infer information needs from goals and interests, as well as a generator to translate information needs into vectors of noun phrases. This information need is described using natural language phrases, since users are able to verbalize easily their need.

In order to extract useful information, profiles have to be parsed. The *Parsing* module transduces noun phrases to *phrase frames* (also called *information descriptors*) which are used for retrieval.

The *Retrieval* module ([WBHW96]) employs autonomous intelligent information agents to collect documents from information sources, which are also reduced to information descriptors. The matching process selects documents to be presented to a user.

The *User Interaction and Rendering* module displays the information in the right way, logging users' reactions about presented information. This can implicitly and/or explicitly give relevance feedback to the user modeling for updating and refining profiles, adjusting the filters better to users' need. The interaction module provides the interface between users and the system, whenever this is needed, in a user-friendly and easy-to-understand way.

2.3 Profiles

Profiles are important and used in every part of the project. A typical *user profile* describes the information need of a user. It contains some general information about the user, like his/her name, address, etc. and vectors of information descriptors representing the information need. Each vector or *topicality* represents one particular information need (e.g. wind-surfing) and it is enhanced with *situational factors* and *constraints*, which could be:

- **User expertise**. As the user gets more knowledge about a topic, more expert information has to be retrieved, so documents already received in former searches must be taken into account. Prefered language is yet another constraint.
- **Amount of documents needed**. In some cases only one document is enough to satisfy a user's need. For instance, a user might want to know the date of J.F. Kennedy's death.
- **Time constraints**. These determine how soon the user needs the answer, and define constraints about document creation/modification dates.
- **Document characteristics**. The user might be interested only in a particular type of documents (e.g. TEX-files, html-files etc.) or a particular structure (e.g. letters). Document quality can also be considered in terms of text quality, reliability and fanciness (for instance, the proportion of pictures to text or animated images). Another constraint can be the size of files in bytes or words.
- **Sources of Information**. The user might be interested only in academic sources or in other particular information providers.
- **Financial constraints**. For WWW sites that provide information under payment, the cost of retrieved information must be within the users' budget.

The most important situational factors and constraints are mentioned above, but many others can be considered, such as popularity and intended use of documents, etc.

Users with common interests can create a *group profile*, which is represented in the same way as a user profile, but which 0also contains links to the users that belong to it. User and group profiles are illustrated in figure 3. More details and related work on user and group

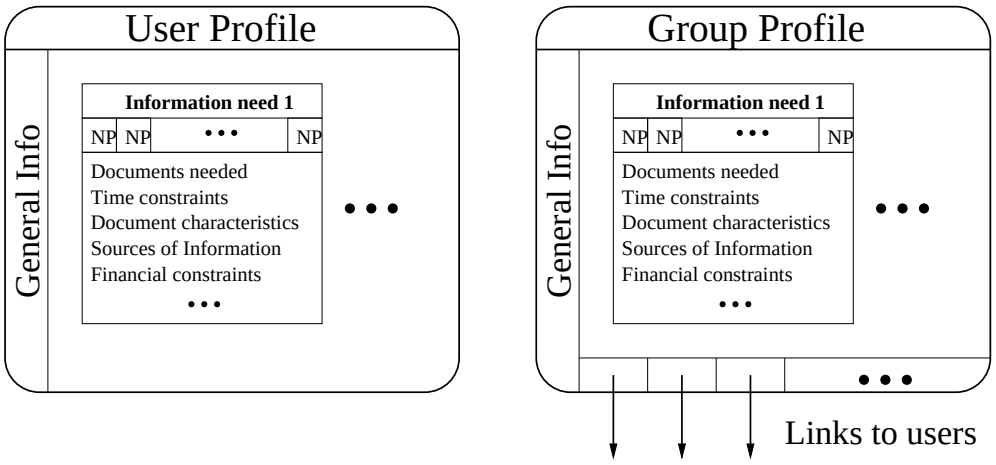

Figure 3: User and group profiles

profiling may be found in [MS94], [RIS+94] and [HSRF95].

Group profiles must also be created by the system for optimizing purposes. The system must be able to compare individual interests and classify users into groups using disjunction functions. In this respect, group profiles are stored in fast main hosts and are used for a rough matching to incoming documents. Documents that match a group profile are forwarded only to users which belong to that group and the more precise and time-consuming filtering is performed at users' workstations. Multiple levels of profiles are also possible. The flow of incoming documents and the matching to profiles is depicted in figure 4. Moving up in the hierarchical structure of profiles, information need and constraints become less specific.

3 The Parsing Engine

In the Profile filtering project, *document profiles* consisting of a set of information descriptors are used for representing documents. This makes indispensable the implementation and employment of a Parsing Engine that deals with natural language analysis. This Parsing Engine will perform the following three main tasks:

- *parsing profiles* to extract noun phrases
- *parsing documents* to extract noun phrases
- *transducing noun phrases* to *information descriptors*, by applying *normalization*.

These procedures are illustrated in figures 5a and 5b.

The parser uses a large lexicon containing terminal words-forms, and extracts noun phrases from profiles and documents. It may also create new noun phrases derived from other (e.g. verbal) constructions by applying certain transformations (see section 3.2.1). Extracted sets of

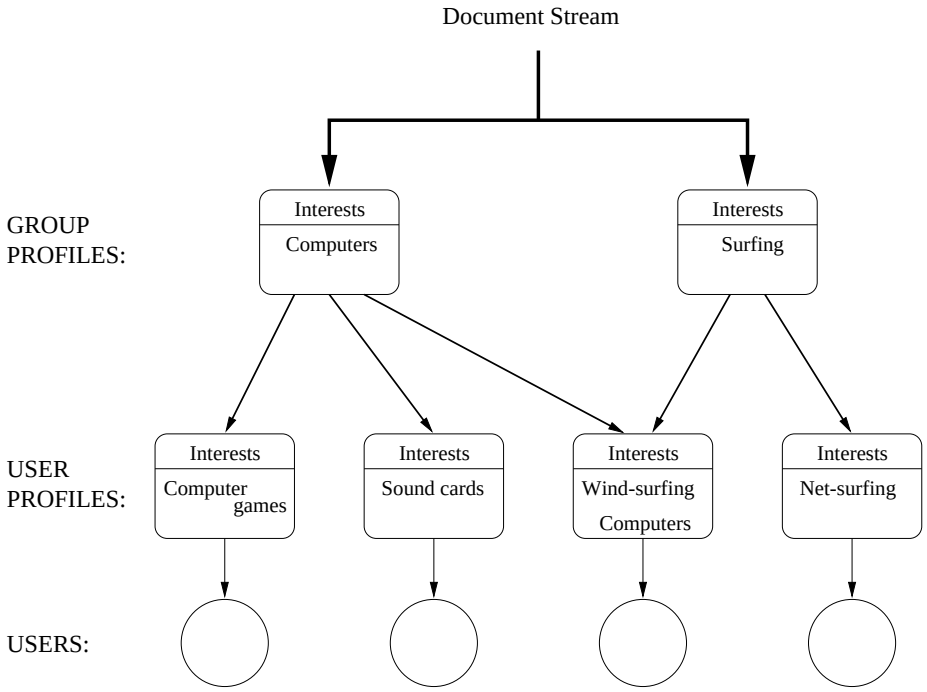

Figure 4: Information flow through hierarchical profiles

NPs are transduced to *information descriptors* by the transducer which actually runs under the parser. That means that as soon as an NP is successfully parsed, it is transduced to an information descriptor, the data structure containing only the essential information for retrieval. This is achieved by *normalization* of NPs (see section 3.2.2).

So far, the research is focusing on the English language, although the usability of the system in Dutch and other languages will be considered in the near future, as well.

3.1 Noun Phrases

Noun phrases have been used in the past in IR for describing documents and as query language. These approaches have shown a marked improvement in precision at the price of a dramatical drop in recall ([AT96], [ATK96]). Noun phrases can be considered as a semantical unit that is better than text windows.

The basic premise is that words found in the same noun phrase tend to share semantic relatedness. If two or more nouns and their respective adjectives are found in a single noun phrase, then it can be claimed that these nouns share some relatedness, even without knowing what they stand for. For example in the phrases:

> *... The album "Live at the BBC" is culled from 52 BBC radio programs that the Beatles appeared on between ...*

the nouns "radio", "programs" and the proper name "BBC" which reside in the same noun phrase of the first sentence are semantically related. Therefore, searching for the programs of the famous BBC radio station with the information need described as "radio programs on BBC" — which is a noun phrase as well — a document can be retrieved containing phrases like the above and not phrases with simple co-occurrence like:

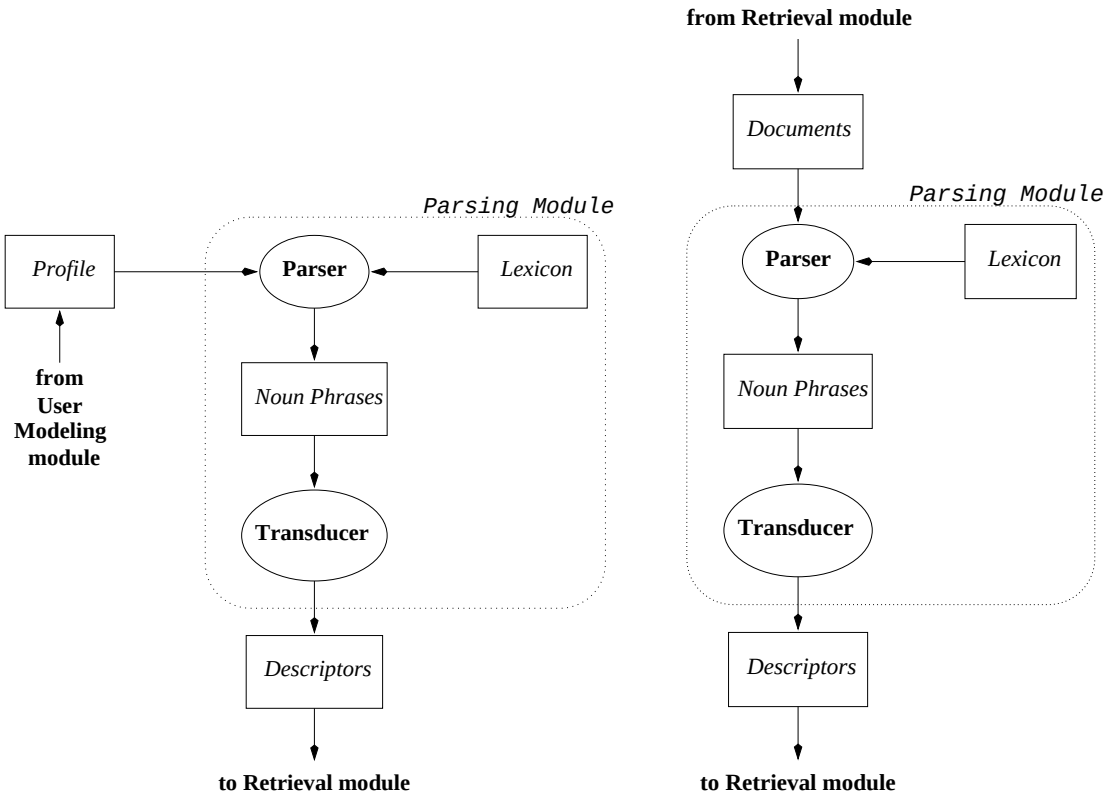

Figure 5: (a) Parsing profiles (b) Parsing documents

Ten musicians from the BBC Symphony Orchestra were interviewed in several radio programs of L.A. stations ...

These phrases ought to be rejected using the syntactic information that the three words of the query reside in different noun phrases.

From our Information Filtering point of view, an NP can be defined as[1]:

$$Determiner^* + Premodifier^* + Headword + Postmodifier^*$$

Determiners are articles (a, the), quantifiers (e.g. many), etc.; they are of little interest for our form of retrieval, and will therefore be skipped. Pre-modifiers are adjectives, non-head nouns and coordinated phrases like "*... big and black...*" The head-word, which is the last noun or (possibly) personal pronoun preceding the post-modifiers, is the central part of the NP. Post-modifiers may be prepositional phrases that post-modify the head e.g. "*from the BBC Sympony Orchestra*", relative clauses e.g. "*that the Beatles appeared on between*", or post adjectives e.g. "general" in "director general" (rare in English). Post-modifiers may recursively include other NPs. For the filtering purposes, determiners, coordination and the use of relative clauses are dropped.

[1] the asterisk denotes zero or more occurrences

3.2 Linguistic Variation

Natural languages allow the same semantic information to be described in many equivalent ways by the exploitation of linguistic variation. We will distinguish *lexicosemantical*, *morphological* and *syntactical* variation. In order to obtain as much information as possible from the documents, linguistic variation will be taken into account.

3.2.1 Normalizing Verbal Constructions

With respect to syntactical variation, documents are processed so that some verbal constructions are transformed to noun phrases minimizing the loss of information. Important transformations can be applied in cases like:

- **Predicative sentences**. For instance, "red" is a predicative adjective in the sentence

  This car is red $\rightarrow$ *this red car*

  which is transformed to a noun phrase.

- **Verb phrases**. Many verb phrases can be transformed to noun phrases by nominalizing the verb. For instance:

  to construct software for $\rightarrow$ *software construction for*

- **Adverbial constructions**. Adverbs give extra information about the verb they belong to. In some cases they are important, and wherever verb phrases are transformed to noun phrases, some kinds of adverbs must be transformed as well, e.g.

  to constructively create software $\rightarrow$ *constructive creation of software*

  "Constructively" is an adverb of *manner* and in this case it may provide useful information. The resulting NP it is not a real NP in respect to if it can be spoken or written, but it is syntactically a valid NP. Consequently, the term NP is generalized for the purpose of effectiveness.

- **Anaphora**. Noun phrases may refer to one another, and the NPs to be searched for may well have been distributed over a number of different NPs or even sentences. For example, in

  Because of its pollution, the air is dangerous to human beings

  the word "it" indicates anaphora. A technique must be found to deal with some of the more frequent cases of anaphora.

It will be investigated what syntactical transformations are possible.

3.2.2 Normalizing Noun Phrases

An NP may occur in various functions in a sentence (as a subject, object, complement or after a preposition) and therefore in different cases. In order to enhance recall, all variations of a given noun phrase have to be mapped to only one. This mapping is called *normalization*

of noun phrases and results in information descriptors. Three kinds of normalization may take place, *morphological*, *syntactical* and *semantical*.

Until now, in keyword-based systems, morphological normalization has been mostly done by means of *stemming*, which without considerable aid from a lexicon may reduce a word to a different word (executive/execute) or even to a non-word (police/polic). Stemming is rather ineffective when is applied in more inflected languages than English because of the ambiguities it introduces. Rather than using stemming for morphological normalization, verb forms are going to be reduced to the infinitive and all inflected forms of nouns to the nominative singular. This will be done by using guidance from the syntax analysis, taking into account the context of a word.

Syntactical normalization is achieved by flattening the syntactic structure. Certain constructions (such as apposition and participle constructions) should be transformed into corresponding NPs. Thus, NPs like

polluted air
air pollution
polluting the air

will all be mapped to

pollution of the air

The process of eliminating verbal constructions and transforming them to noun-phrases is also considered as syntactical normalization. These techniques are expected to lead to a major improvement in recall.

Semantical normalization can be accomplished to a rather high degree by considering word relations like *synonymy*, *is-a*, *has-a*, *part-of*, etc. Even simple relations like synonymy have been proven quite effective.

Multiple parse trees can be obtained for the same NP as a result of *structural* and *lexical ambiguities*. Structural ambiguity refers to all the possible ways of grouping words as to form an NP, and may result in an exponential increase of the number of analyses and parsing time. Lexical ambiguity occurs when a word belongs to more than one lexical category (part-of-speech). All these ambiguities will be resolved in the best possible way, keeping the most likely parse tree.

3.3 Information Descriptors

An information descriptor is a set of words, having the most important noun as head followed by its adjectives and other nouns. Head-words can be used as indices, speeding up the search process. In most cases the head noun of an NP is the last noun before the first post-modifier, which can be easily recognized by the parser.

Information descriptors can also be viewed as lattices having in one end the head-word and in the other end the noun phrase as a whole. As an example, consider the following transduction (normalization) of an NP to an information descriptor.

air pollution by stinking fumes $\rightarrow$ {`fume, stink, air, pollution`}

The head-word is on the right. Notice the morphological normalization of "fumes" as "fume" and the syntactical normalization of "stinking" as "stink". This list is represented as a lattice in figure 6.

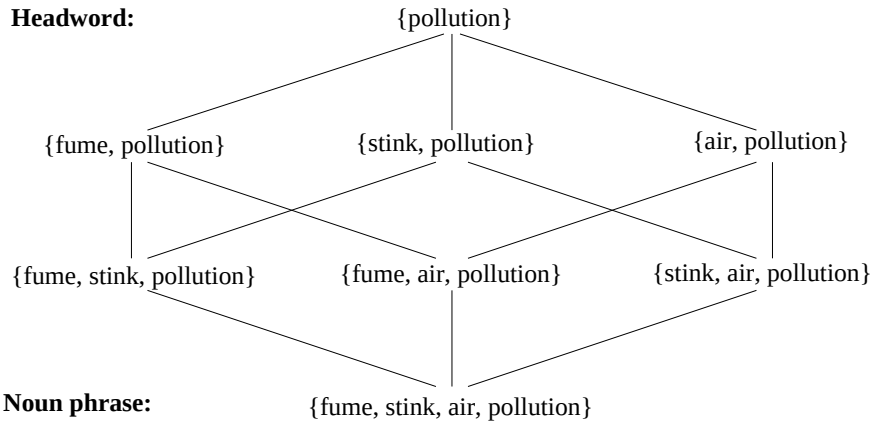

Figure 6: The lattice of an information descriptor

Sets of information descriptors augmented with extra information are used by the retrieval module of Profile in comparing profiles/queries to documents. This extra information denotes the importance of the NP which the lattice belongs to, in terms of *occurrence count*, *point of occurrence* (NPs appear in titles are regarded as more important than others, e.g. NPs in main body of text or enclosed in brackets), *emphasis* (topicalized, capitalized or not), etc.

4 Research Questions and Approach

The main research question is whether natural language techniques are beneficial and suitable in agent-based Information Retrieval and Filtering. Related work indicates that this might be true, but it must also be done in an efficient way. In order to investigate the possibilities, further research must be conducted concerning natural language parsing in relation to normalizing techniques.

4.1 IR/IF Grammar Properties

At the start of the project, relatively complete and well-validated grammars for English (ENP4IR, [Tib96]) and Dutch (AMAZON) are available. The main research questions are:

- **Coverage**. The coverage of the grammar and lexica will be investigated, that is the ability of the grammar to describe the syntax of English language. Proper name recognition is another important feature for an IR/IF grammar, considering the large amount of proper names that reside in documents. Necessary modifications of the grammar will be done as to attain the greatest coverage and to fit the IR/IF problem.

- **Robustness**. Specialized IR/IF grammars must differ from linguistic grammars due to allowance for the bad syntactic constructions and unknown words which may appear in real documents. An IR/IF grammar must be robust, that is, open to unknown words and bad syntactic constructions. Techniques which will be used for that are *wild-card* and *island* parsing.

- **Efficiency**. Applying natural language analysis can be time-consuming. The response time of the intended system must satisfy users' time limitations. This suggests a trade-

off between parsing accuracy and time consumption.

- **Ambiguities**. Syntactical and lexical ambiguities must be resolved considering an IR/IF point of view. Techniques which may be used are *under-specification*, *wild-card* and *island* parsing, *latest closure*, etc., and the impact of them on effectiveness has to be kept on low levels, minimizing the loss of semantic information.

The *AGFL*[2] (Affix Grammars over a Finite Lattice; [Kos91]) system will be employed for the implementation of the parser and transducer.

AGFL offers a small, flexible and simple formalism in which large context-free (CF) grammars can be described very efficiently. In AGFL, context-free grammars are augmented with features for expressing agreement between parts of speech. In that respect, the formalism consists of two levels, *rules* and *meta-rules*. Rules constitute the first level of the grammar, and those are the ordinary CF part of AGFL. The second level contains the meta-rules which define the affix domain. The two levels are connected by associating affixes with non-terminals.

It is not fully proved that CF grammars (and therefore AGFLs) are adequate to describe the syntax of natural languages. There are some debates in the community of linguists on the subject which arise mostly by the fact that not all the linguistic phenomena can be easily and efficiently captured in this formalism. On the other hand, AGFLs have shown their power in the past in describing some languages in a rather utilizable, reliable and comprehensive way ([Kos96]). Although AGFLs can be transformed to CF grammars, they manage to reduce the size of the grammars and make the construction of a complicated grammar for a NL feasible and in a much more understandable way.

The AGFL formalism was chosen to be used in Profile's Parsing Engine because it is extremely fast, and it allows to separate the linguistic information from the algorithmic implementation; in this way, the linguistic knowledge remains transparent, robust and manageable. Another important reason is the availability of AGFL grammars and parsers for English which would be helpful at the start of the project.

4.2 Normalizing Noun Phrases

Normalization of noun-phrases is an important aspect of the Parsing Engine. Morphological, syntactical and semantical normalization might result in distortion of the original meaning, disorienting the retrieval. It will also be researched which of the verbal constructions can be transformed to noun phrases, and the consequences of such transformations, from a semantic point of view. The effect of such normalizations in precision and recall of retrieval must be measured. Experimentation with large corpora (e.g. from TREC) will give a guidance for adjusting the normalization techniques.

Normalizing techniques are introduced to enhance the recall of the system. They depend more on linguistic knowledge than stemming techniques. Semantical normalizations demand the use of sophisticated lexica which supply also synonyms, hypernyms, hyponyms and other word relations. The *WordNet* database will be exploited in obtaining these word relations. For more information about WordNet refer to [MBF+93]. Such innovations will be compared to standard recall enhancement techniques.

[2]The AGFL system has been developed by Department of Informatics, University of Nijmegen, The Netherlands. http:// www.cs.kun.nl/ agfl/

It will be also investigated whether the use of information descriptors allows searching over documents in another language than the query language. This can be done using multi-lingual lexica like EuroWordNet[3]. For certain applications, this *multi-lingual* aspect may be of great importance.

4.3 Similarity of Noun Phrases

An important goal of this research is a metrics for expressing *similarity* of noun-phrases (or information descriptors, since they represent noun-phrases). As soon as such a metrics is created, the comparison between user and document profiles will be feasible, allowing the choice between more or less relevant documents.

5 Planning

The total project effort is planned to be about 16 man years, divided over four years of research and developement. The research model is based on incremental refinement of the system as a whole. This means that at an early stage a prototype has to be developed, with some functions simulated or "suggested" in the user interface.

The foreseen tasks for research and implementation of the Parsing Engine are shown in table 1. Each period consists of 6 months and 2 major milestones are defined, the first at the

Period	**Task**
1	Position Paper Requirements analysis for the Parsing Engine Refine overall project plan Demonstrator
2 & 3	Adaptation of the ENP4IR grammar and lexicon First version of the Parsing Engine Experiments and testing
4	Evaluation and reconsideration of requirements Integration planning Experiments
5 & 6	Final version of the Parsing Engine Integration Experiments and testing
7	Evaluation
8	Completion of thesis

Table 1: Tasks and milestones

end of 2 years and the other in 4 years. At the end of 2 years a first integrated version of the Parsing Engine will be ready. In order to achieve that, testing and many experiments are

[3]still under development and is a joint enterprise of the University of Amsterdam, the University of Sheffield, the Instituto Linguistica Computazionella del CNR (Pisa), the Fundacion Universidad-Empresa (a cooperation of the Universities of Barcelona and Madrid) and Novell Linguistic Development in Antwerp. For more information refer to http:// www.dcs.shef.ac.uk /research /groups /nlp /funded /eurowordnet.html

required, so reconsideration of requirements can be done. Small experiments will support the ongoing research.

The final version will be ready at the end of 3rd year followed by extensive evaluation at the first period of the 4th year and completion of a thesis which will incorporate all the research.

6 Conclusions

The Profile project combines different areas of research in order to reach its objective. The research contribution, coming from areas like User Modeling, Interaction and Rendering, Linguistic Parsing, Information Retrieval and Filtering, will be integrated in a pro-active information filter.

The Parsing Engine deals with the linguistic analysis and normalization of documents and users' profiles. The linguistic techniques which will be applied in filtering are expected to result in better performance in terms of precision and recall than keyword-based approaches.

The ultimate goal of the whole ongoing research is the implementation of an intelligent retrieval/filtering system which will have the capability of continuously adapting to users' information need, will be easy to understand and attractive to use, and of course must be efficient and effective.

References

[AT96] A.T. Arampatzis and T. Tsoris. A Linguistic Approach to Information Retrieval. Master's thesis, Department of Computer Engineering and Informatics, University of Patras, Patras, Greece, June 1996. Available from: http:// www.cs.kun.nl/ ~avgerino/ LA2IR.ps.Z.

[ATK96] A.T. Arampatzis, T. Tsoris, and C.H.A. Koster. IRENA: Information Retrieval Engine based on Natural language Analysis. Technical report CSI-R9623, Computing Science Institute, University of Nijmegen, Nijmegen, The Netherlands, 1996.

[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.

[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.

[HSRF95] W. Hill, L. Stead, M. Rosenstein, and G. Furnas. Recommending and evaluating choices in a virtual community of use. http:// community.bellcore.com/ navigation/ videos.html, 1995.

[Kos91] C.H.A. Koster. Affix Grammars for natural languages. In *Attribute Grammars, Applications and Systems, International Summer School SAGA*, volume 545 of *Lecture Notes in Computer Science*, pages 469–484. Springer-Verlag, Berlin, Germany, June 1991.

[Kos96] C.H.A. Koster. AGFL Grammars for full-text Information Retrieval. Technical Report, Department of Computer Science, University of Nijmegen, Nijmegen, The Netherlands, February 1996. (to appear).

[MBF+93] G.A. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. Miller. Five Papers on WordNet. Technical report, Cognitive Science Laboratory, Princeton University, August 1993. Available for anonymous ftp from ftp:// clarity.princeton.edu /pub /wordnet /5paper.ps.

[MS94] M. Morita and Y. Shinoda. Information Filtering Based on User Behaviour Analysis and Best Match Text Retrieval. In W.B. Croft and C.J. van Rijsbergen, editors, *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 272–281. ACM Press, 1994.

[OM96] D.W. Oard and G. Marchionini. A Conceptual Framework for Text Filtering. http:// www.ee.umd.edu/ medlab/ filter/ papers/ filter.ps, 1996.

[Ram91] Ashwin Ram. Interest-based information filtering and extraction in natural language understanding systems. In *Bellcore Workshop on High-Performance Information Filtering*, Morristown, NJ, 1991.

[Ram92] Ashwin Ram. Natural language understanding for information-filtering systems. *Communications of the ACM*, 35(12):80–81, December 1992.

[Rij79] C.J. van Rijsbergen. *Information Retrieval.* Butterworths, London, United Kingdom, 2nd edition, 1979.

[RIS+94] P. Resnick, N. Iacovou, M. Suchak, P. Bergstorm, and J. Riedl. GroupLens: An Open Architecture for Collaborative Filtering of Netnews. In *Proceedings of ACM 1994 Conference on Computer Supported Cooperative Work*, pages 175–186, Chapel Hill, NC, 1994. ACM.

[SO95] H. Sorensen and A. O'Riordan. A learning personalised information filter. In *Proceedings of the AI'95 Conference*, Montpellier, France, 1995.

[Tib96] C. Tiberius. An Annotated AGFL Grammar for English. Technical report, University of Nijmegen, Nijmegen, The Netherlands, 1996.

[WBHW96] B.C.M. Wondergem, P. van Bommel, T.W.C. Huibers, and Th. van der Weide. Towards an Agent Based Retrieval Engine. Technical Report CSI-R9620, Computing Science Institute, University of Nijmegen, Nijmegen, The Netherlands, 1996.