Cognitive Filtering of Textual Information Agents Based Implementation

Omar Nouali[1], Nadia Nouali-Taboudjemat[1], and Bernard Toursel[2]
[1]Theory and computer engineering, DTISI, CERIST, Algeria
[2]University of Sciences and Technologies of Lille, France

Abstract : *The study presented in this paper has multiple objectives. The first objective is to automate the information filtering process by taking into account the relative importance of information and resources needed for its treatment. The second one is to demonstrate the applicability and contribution of an agents based implementation to automatic information filtering. The third one is to show how learning can improve the effectiveness of filtering and that automatic learning is necessary in the design of automatic information filtering systems. We propose an open, dynamic and evolving solution that offers to the filtering process the opportunity to learn, exploit the learned knowledge and adapt itself to the application environment. We have adopted agents to improve the response time compared to a sequential algorithmic solution. To validate our filtering approach, we led a set of experiments to evaluate performances of the techniques and tools we have developed.*

Keywords: *Information filtering, machine learning, linguistic agents, and filtering criteria.*

Received March 25, 2009; accepted August 4, 2009

1. Introduction

Being a universal means for information dissemination, the Internet has become a huge source of resources and material difficult to access, evaluate and especially to exploit. Users equipped with only classical information retrieval tools, such as search engines, can not face the flow of generated information.

The permanent increase of information amount available in electronic format induces, for users, new information access needs. In order to save precious time for useful information retrieval, the use of new tools, which assist user in his information quest, becomes inevitable. This need has resulted in investigations and research for the development of new mediators between information sources and users, including the information filtering systems.

These systems, which are thus positioned as a "third party" in the communication process between the information source and the final consumer, must have the methods and skills needed to carry out, evaluate, filter, access and extract relevant information to facilitate the task for different users.

The current filtering systems lacke precision because of information representation and ignorance of semantic side. A paper represented by a vector of key words (classical lexical representation) raises the question of relevance, i.e., what criteria ensure that the representation is relevant for a given document? The main techniques used in most filtering systems nowadays are based on a superficial contents scan [8, 13, 14]. They are generally based on a lexical property, the presence or absence of keywords that user must indicate to the system in the form of logical filtering rules based on keywords. This requires the user to frequently update the rules because of the information dynamic nature.

In addition, some systems rely on static closed architectures. They do not fit themselves or hardly do to extensions or possible upgrades. Indeed, the addition of new treatment requires the recompilation of all programs.

In this context, we propose a solution that has the following characteristics:

- Open and dynamic, this is very important. New treatments (filtering criteria) can be added over time, and the filtering system must be able to adapt itself to integrate these new treatments in order to increase the overall effectiveness without modifying what already exists.
- Evolutionary, offering the filtering process the opportunity to learn, to exploit such knowledge and adapt itself to the application environment.
- Necessity of utilizing linguistic resources and treatments to improve the performance of filtering.
- Modelling with agents to provide a time saving compared to a sequential algorithmic solution.

To evaluate our approach in terms of feasibility and performance, we conducted a set of experiments.

2. State of Art

The main models used in filtering area can be classified into the following main axis:

- Cognitive filtering: this is about content evaluation;
- Economic filtering: concerns the consumed resources (size, consulting time, *etc.*,).
- Collaborative (or social) filtering: based on other users opinions.
- Hybrid filtering.

In the frame of this work, we are interested, especially, in cognitive filtering. Two dominant approaches exist for this kind of filtering: statistical approach and symbolic approach.

The statistical approach is based on the occurrence of a set of keywords to identify or recognize the pertinent information [2, 3, 6]. Different statistical methods have been proposed using the principle of matching the representation of objects being filtered (e.g., paper) with the profiles one. These are methods that implement statistical concepts, they are thus based in their analysis on the frequency (or presence) of the words which constitute the user profile in the objects to filter. We can mention full-text, vocabulary, grouping, boolean, vectorial, probabilistic, connecting and by expert system filtering. The advantage of this statistical approach relies primarily on its simplicity, but it is based on an unrealistic assumption, which is that all words are completely independent. Indeed, the most popular systems, participating in TREC (reference conference) did not take into account the order of the words appear in, and the dependency relationship between the existing linguistic elements (words, combinations, chunks, phrases, *etc*...). Moreover, the relevance criteria is solely based on the presence or absence of keywords in the Treated text. Any analysis (finding relevant segments, *etc.*,) carried on these bases can't avoid vagueness. For example, the relevant documents whose representation does correspond approximately to the profile will not be selected. The texts contents representation by keyword (automatically or manually extracted) possibly weighted, is particularly poor, because it does not take into account the linguistic structure of the manipulated texts. A set of keywords preserves a small fraction of the original text meaning.

The symbolic approach tries to make an analysis of the relationship between the documents contents (objects to filter) and the user interests [5, 10, 15, 16]. To do this, we need a semantic model that allows representing the user interests and understanding the content of the document that needs a language study. This is an interesting approach and provides an effective filtering, but difficult to apply to any texts covering various fields. The automatic texts analysis is booming but still difficult [2, 4, 6, 15, 16]. The analysis of all information stored in a text is a very complex process, since it involves many parameters. The interpretation of a free text statement requires not only linguistic knowledge, but also extra linguistics knowledge (knowledge of the world, conventions, and so on.). All this knowledge is difficult to encode in the automated analysis systems because of their complexity and their quantity. Currently, the automatic texts analysis requires the mobilization of significant linguistic resources (dictionary) and tools for natural language processing (parser, grammar, textual representation). It is therefore restricted to a very limited area, or to a very basic understanding. The hope to develop robust analyzers capable of handling free texts in depth, has led many researchers to implement varied analyzers in recent years. These analyzers vary in terms of strategy (determinism vs. non determinism, partial analysis vs. full analysis), and in terms of theoretical bases (statistic analyzers vs. symbolic analyzers, *etc.*,). Of course, each of the analyzers is more or less dedicated to specific tasks and no analyzer can now claim to perform a complete analysis of all the sentences in a free text corpora.

Statistical methods, although promising, are not themselves sufficient to address all aspects of automatic language processing. The symbolic grammars are also required to obtain a reliable and accurate representation of the semantics, which is crucial for many applications of natural language processing.

The coupling between statistical and symbolic methods (quantitative / linguistic) ensures, in our opinion, a more effective analysis of textual documents and, therefore, a more accurate filtering. Indeed, as documents passing on the Internet are often little linguistically correct, it is interesting to combine the two types of methods to maintain the advantage of both; simple and purely statistical treatments and treatments based on a strong linguistic knowledge.

In this context, we propose a partial analysis of documents contents using multiple levels of analysis that can generate linguistic information about both the structure and the content of documents. By the use of linguistic information in the filtering field, we mainly target the following objectives:

- Define properly discriminating and unambiguous profiles.
- Extended representation of documents to be filtered.
- To connect different wording but semantically close in order to increase the chances of matching a profile and a filtered document.

In addition, a filtering system must learn automatically to improve its knowledge.

3. Filtering Architecture

Our preferred filtering architecture is mainly based, referring to existing filtering systems, on the need to use linguistic resources or treatments. In this context, we want to show and we believe that the use of knowledge and language processing can improve the performance of a filtering system. On one hand, it's about extracting the maximum amount of characteristics of the information being filtered that will be used to improve its performance. On the other hand, to extend the representation by taking into account the semantic aspect.

The extraction of a maximum of characteristics and taking into account the semantic aspect in the filtering process are, in our opinion, the most promising to reach an effective filtering. In fact, when we analyze a given document, we find also, in addition to the lexical features, other features that seem to be interesting. These additional features are a certain set of indices that we add to the classical representation (lexical representation).

Our filtering approach uses thus a set of basic linguistic treatments to extract a set of characteristics, and build internal representation of each document. These treatments are independent from the domain of application. We classified them into several levels: lexical, architectural, structural, syntactic, enunciate and pseudo semantic.

In frame of this work, we are not trying to make a deep and comprehensive analysis of documents contents, but rather, a partial analysis using several levels of analysis that can identify linguistic characteristics (or properties) which should enable distinguish between different documents, locate a document in relation to others, and have a better quality documents filtering. In addition, to improve the filtering performance, we offer an evolving solution allowing the different modules of the architecture to learn from data, to exploit such knowledge and adapt themselves to the nature of documents carried out. The global architecture is composed of principal following modules as shown in Figure 1:

- A pre treatment module which aim is to normalize each document and prepare it for subsequent analysis steps (document part).
- A linguistic processing module that analyzes the document and delivers as output its associated representation. It uses a set of basic linguistic treatments classified into several levels: lexical, architectural, structural, syntactic, enunciate and pseudo semantic (criteria part). The first five treatments would retrieve, from document being filtered, linguistic features characterizing it, while pseudo semantic processing can be used to extend the obtained representation.
- A Filtering module that allows to compare a document with the different user profiles (profile part).
- A learning module that improves the knowledge, efficiency and performance of various processing modules and filtering module (learning part).

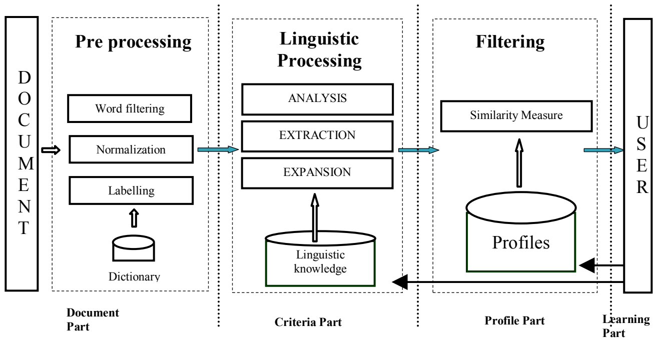

Figure 1. Filtering architecture.

4. Agent Based Modelling

We propose to our cognitive filtering architecture an agent based modelling. We describe the different types of agents (modules specialized in specific and independent treatments), enabling a reduction in processing time, a better maintainability, and therefore, a greater evolution capacity in the perspective of new uses development.

4.1. Agents Based Filtering Approach

An interesting solution to the filtering is the distribution of treatment. The idea is to structure the problem in the form of a group of specialized entities, each with a degree of autonomy and able to cooperate, coordinate, negotiate, and so on, with other entities. It would therefore be interesting to show the applicability and adaptability of the multi agents approach to cognitive filtering.

The use of a multi agents approach to drive the filtering process meets well the openness, complexity and the dynamics of our filtering architecture.

Indeed, the use of a multi agent approach offers among others the following possibilities:

- Modelling solutions using independent entities, each with a specific filtering task to perform. This offers a real time saving compared to a sequential algorithmic solution.
- Have an open and dynamic system, which is principal. Indeed, the need for openness is because new treatments can be added over time, and the system must be able to adapt itself to integrate these new treatments in order to increase the overall efficiency, without changing what already exists. The system is dynamic because the treatment agents can be created and destroyed dynamically.

The multi agents approach seems to us a very interesting way to tackle the problem of information filtering. Its decomposition into several entities specializing in specific treatments can provide an open and dynamic solution. Such an approach allows the filter system to be more efficient by providing a parallel resolution of the various tasks, which consequently reduces the response time of the system.

4.2. Agents Based Design

Our filtering architecture consists of several different types of agents that can be grouped under two main categories:

- Permanent agents are those whom after creation reside in the system. Each permanent agent manages a system module;
- Non-permanent agents are agents that are created as needed and will be destroyed at the end of their mission. They are created to perform a specific task. Once this task accomplished, the non-permanent agent is destroyed.

In our context, an agent is an autonomous entity which performs a specific filtering task using its knowledge.
The system is mainly composed of 3 main types of agents as shown in Figure 2:

- Document agent which is created for each arrival of a new document into the system. It is responsible for driving and coordinating the document analysis and filtering operations.
- Criterion agent that implements a specific treatment on the content of the document. It will analyze the document and extract a set of properties that characterize it.
- Profile agent is in charge of the pre-filtering process which consists in the elimination, at first, the documents that have different characteristics from those expected by the user, and then launch the filtering operation for the retained documents after pre-filtering operation.

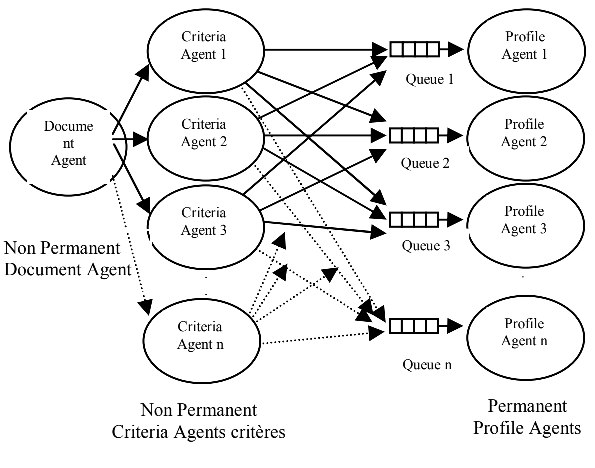

Figure 2. Multi-agent filtering architecture.

4.3. Document Agent

Document analysis previously requires a labelling phase and a normalization step. Labelling phase allows the extraction of some linguistic indices (syntactic, for example). The standardization phase reduces the words morphological variations in a common form (return verbs into infinitive form, delete the plurals, *etc*.,), often called term. After these labelling and normalization steps, document agent calls the criteria agents who are responsible for analyzing and extracting a set of features (or properties) that characterize the currently treated document. Each criterion agent implements a set of specific characteristics. Indeed, for the content treatment, the document agent transmits (broadcast) it to criteria agents. The goal is to analyze the document and retrieve properties that characterize its contents. This operation is performed in two stages: a pre-filter stage launching "advanced" criteria agent and a filtering stage by launching the various linguistic criteria (lexical, architectural, structural, syntactic and enunciated). The analysis results of the various criteria agents are sent to different profiles agents that compare them and return filtering decisions. The document agent is awaiting responses from different profiles (document fits the profile or does not). It destroyed itself once it has received all the answers (profiles). This agent is non permanent.

4.4. Criteria Agents

The basic criteria agents are used to analyze the document, extract different characteristics, and build the internal representation of each document. These agents are grouped into following types:

- "Advanced criteria" Agent to identify the document language, the author, the Internet address… These advanced criteria are necessary for the pre-filtering operation to eliminate some irrelevant documents. Lexical Agent to identify document lexical properties.
- Architectural Agent to identify document architectural properties.
- Structural Agent to identify structural properties of the document.
- Syntactic Agent for identifying syntactic properties of the document.
- Enunciate Agent to identify document enunciate properties.
- Pseudo semantics Agent to expand and improve the representation of the document.

All these agents are created by the document agent except pseudo semantic agent which is run by the profile agent to eventually complete the initial representation issued by all agents. Each criterion agent disseminates its results (linguistic characteristics) to profiles agents then it is destroyed. They are non permanent agents.

4.5. Profile Agent

The main task of this agent is composed of two parts: a pre-filter part and a filtering one. The pre-filtering is to decide on the continuation or no of the current document analysis. Indeed, after advanced criteria identification step carried by concerned agent, the profile agent decide according to the user requirements to reject or not the document being analyzed. Figure 3 describes the diagram sequence associated with the protocol of interaction between agents for pre-filtering operation.

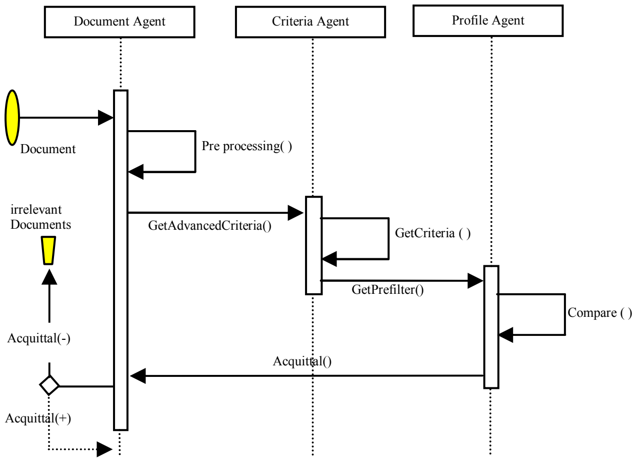

Figure 3. Sequences diagram of the pre-filtering protocol.

The Profile agent receives specifications sent by the «advanced criteria» agent and compares them with those chosen by the user (as defined in its profile). If the characteristics are identical then the concerned document will be retained by the pre filtering operation. In this case, the agent profile sends a positive acquittal to the issuer Document agent. Otherwise, it sends a negative acquittal. If all Profile agents send a negative acquittal, the document will not be accepted and therefore it will be ignored. The filtering is to compare the document with the profile. Indeed, after linguistic analysis stage conducted by criteria agent, the profile agent initially launches the pseudo semantic agent to complement the representation of the document issued by the criteria agents and measure the similarity to decide whether the document being treated is corresponding or not. The profile agent looked for each created document the results of criteria agents: "advanced criteria," lexical, architectural, structural, syntactic and enunciates. He lanches pseudo semantics agent and is awaiting a response. Then it measures similarity and decides whether the treated document correspond to itself or not. Finally, it sends a reply to the document agent. This agent is permanent.

Following Figure 4 describes the sequence diagram associated with the protocol of interaction between agents for the filtering process. At the arrival of a positive acquittal from a profile agent, the document agent initiates the filtering protocol by calling criteria agents to linguistically analyze its content. Criteria agents able to handle the document, send the output resulting from analysis to profiles agents. Each profile agent comes in interaction with the associated semantic agent to try to expand the representation of the document. He then compares this new representation to the characteristics of the profile. According to the calculated value, it responds by either a positive or a negative acquittal.

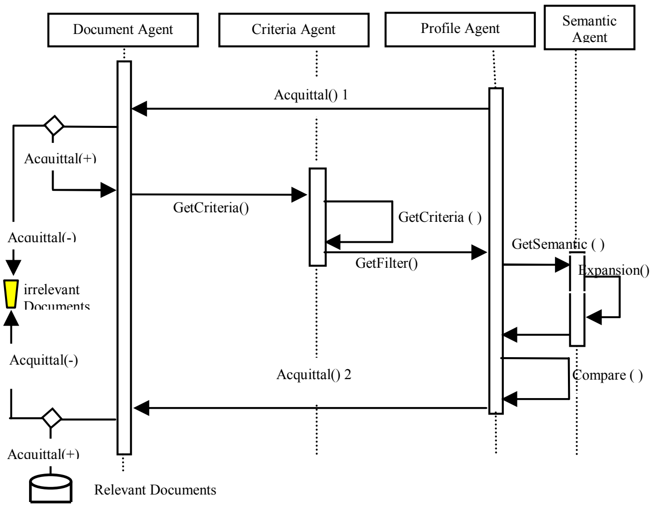

Figure 4. Sequences diagram of the filtering protocol.

5. Evaluation

We conducted experiments to:

- Measuring the importance of using linguistic resources and treatments for representing documents and measure the reaction of this new representation on the effectiveness of filtering.
- Measuring the contribution of an agent based implementation.

To measure the performance of the approach, we choose the use of the two standard performance metrics in information retrieval: measures of precision and recall. Scores of silence and noise, which on are based the precision and recall, are simply calculated by taking the difference between the observed responses and the expected ones. The recall rate is calculated by the ratio of the number of properly filtered documents (or ordered referring to profiles) on the total number of relevant documents in the collection. While the rate of precision is measured by the ratio of the number of documents properly filtered on the total number of documents filtered (correctly and incorrectly).

5.1. Using Linguistic Resources and Treatments

We present and discuss, in what follows, the results of performance, during a quantitative evaluation of our filtering approach, in several configuration cases:

- Performance depending on the simple lexical characteristics.
- Performance depending on the composed lexical characteristics.

- Performance based on the number of linguistic features.
- Measuring the importance and role of semantic information.

5.1.1. Simple Lexical Characteristics

We measure the performance of filtering considering earlier several profiles exclusively consisting of simple words. The profiles are entered into the system in two different cases: manual modelling and automatic modelling:

- Manual modelling: user enters for each profile a list of criteria in the form of simple words. It associates, for each criterion, a weight which represents the degree of importance.
- Automatic modelling: user employs data acquisition and lexical knowledge retrieval tool. The tool takes care of collecting documents and modelling profiles.

The modelling is based on a self-learning from collected documents. It is to analyze and extract keywords and to assign them a weight.The results of filtering in the two cases are given in Table 1.

Table 1. Performances according to the lexical characteristics.

Manual Modelling	Automatic Modelling	
	Before modification	After modification
61,33%	71%	79%

Results obtained by an automatic modelling of profile are significantly better than manual modelling (61,33% versus 79%) which shows the difficulty of user to describe his profile. Moreover, we note that certain words correlate with certain types of considered documents, but are not statistically significant (low value). This pushed us to change the importance of different words while keeping a generic treatment without user intervention. The system assigns a high value weight to the words that are unique to each profile (relevant words), compared with those in several profiles (less relevant words). After modification, the test results were better (79%).

5.1.2. Composed Lexical Characteristics

We measure the performance of filtering by adding a set of compound words to the different basic profiles initially constituted of simple words (previous experience). For this, we used a compound words automatic recognition and extraction tool that we have developed [13, 14]. The extraction is based on learning from documents. The tool takes care to analyze and extract compound words and to assign them a weight.
The results of using multi-terms on the quality and relevance of filtering are given in Table 2.

Table 2. Performances according to the compound lexical characteristics.

Profiles Characteristics	Average Performance
Simple words	79%
Simple words + compound words	80,2%
Simple words + compound words + Weighting	85%

We do not observe a big improvement in performance. Indeed, compound words correlate with documents of considered profiles, but are statistically insignificant (low value). Then, we changed the relative importance of different compound words, by awarding them a high value on weight. Test results were better (85%).

5.1.3. Linguistic Features

We are studying additional characteristics (or properties) that we add to the lexical features of different profiles. We have defined and identified a set of automatable properties. It is a set of linguistic indicators on the document. As an example and for the Spam profile, here are a few indices added to the lexical characteristics: domain of document (.com, .gov, .edu, .com, and so on.), The length of the document, the type of content (html / txt), the language of the message, the capitalized words, abbreviations, non-alphanumeric characters ($,!, #,% * &, and so on.), numeric characters, the size of sentences, the time of creation (night / day), *etc.*

As a first step, we considered all the linguistic without restriction to represent a document. We note that many features do not correlate with the considered documents. Indeed, the representation of various documents is a sparse matrix (a lot of null values). This thus considerably degrades the performance of filtering (a measure of similarity goes to 0). We therefore tested the performance when we reduce the number of features imposing a threshold value on the characteristics to be considered for the various representations of documents as shown in Table 3.

Table 3. Performances according to the linguistic characteristics.

Performances	Only Lexical Characteristics	Lexical Characteristics + Additional Characteristics
Précision	85%	89.6%

We note that the filtering performance improved slightly. This can be explained by the fact that the documents rejected by the system (the case of lexical features only) by lack of keywords or very low value, will be accepted this time, and this because of the presence of some additional features.

Indeed, these additional features added to the lexical features can increase the chances of matching a document and a profile (89,6%). The exploitation of these criteria is not really discriminate (in reality is not always true). Nevertheless, the probability of having a spam type document, for example, is stronger when those criteria are verified. For example, if the domain (in the URL address) is "edu", "fr", the document has a low probability to be a spam. However, if the domain is "com" "net", the probability of being spam increases.

5.1.4. Semantic Information

The experiment is to introduce the system in two different cases a set of documents to be filtered in several sessions. Then measure each time performance (precision), and make assisted learning on the lexical network to measure its effectiveness and its impact on the filtering operation. The results are given in Table 4. We note that the model with lexical network improves the results of filtering. Indeed, the lexical network allows the system to link a document to a profile, even if they have no keywords in common: unknown words are replaced by other closer words, which increases the accuracy rate.

Table 4. Performances according to semantic information.

Profiles	Keywords	Keywords + Lexical Network	
	Without Feedback	Without Feedback	With Feedback
Profile 1	50%	50%	70%
Profile 2	66%	66%	61%
Profile 3	28%	40%	43%
Profile 4	41%	41%	72%
Profile 5	57%	60%	65%
Average Profile	48%	51%	62%

We note that the lexical network model converges towards a satisfactory filtering, but slowly (62% versus 51%). Indeed, the model requires several sessions of assisted learning to improve the quality of its results. It is therefore necessary to launch feedback learning regularly, such as after each filtering session.

5.2. Response Time

The waiting time by the user to have the result is as an evaluation criterion for an automatic system. We are therefore interested in the response time evaluation of our agents based approach. We simulate two machines (single processor) on which agents operate:

- Machine1: allows sequential execution.
- Machine2: allows running multiple agents in parallel.

Table 5 presents the various tasks of this system and the average time (estimated) for the execution of each.

Table 5. Estimated execution time per task.

Task	Average Operating Time (ms)	
Document Processing	Word's Filtering Labelling Lemmatization	0.05 0.15 0.15
	Advanced Criteria Analysis	0.1
	Lexical Analysis	0.2
	Architectural Analysis	0.1
	Structural Analysis	0.1
	syntactical Analysis	0.1
	Enunciated Analysis	0.05
	Global Time Analysis	1
Expansion	0.1	
Profile Processing	0.1	
Similarity Measure	0.05	
Communication Time between two agents	0,02	

To estimate the time, we conduct several experiments by varying the number of profiles and the number of documents as shown in Figure 5.

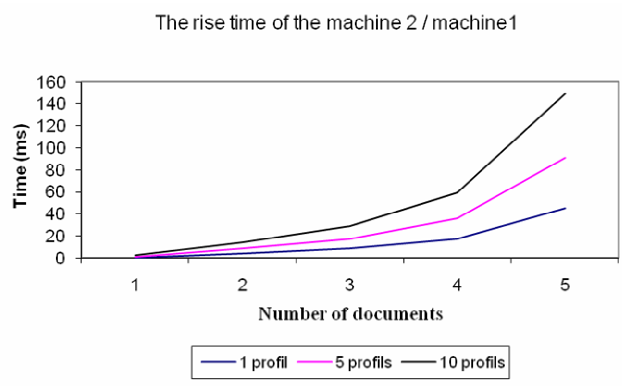

Figure 5. Time saving according to number of profiles and documents.

The service time measured in the machine1 has significantly risen with profiles and documents increasing, by contrast, in machine2 it slightly raised depending on the number of agents (documents and profiles). By this experiment, the machine2 has a better response time than machine1 (sequential filtering), agents based filtering is better suited in a parallel environment.

7. Conclusion

This paper focuses on the filtering problem and proposes an open, dynamic and evolving solution that we modelled with agents to provide a time saving compared to a sequentially algorithmic solution. Each agent has the opportunity to learn and exploit this learned knowledge to adapt itself to the nature of the task assigned to it. The filtering approach is based on the need to use linguistic resources and treatments for representing documents. Indeed, unlike most existing systems, our approach uses linguistic properties on the structure and content of the documents in order to improve the results of filtering. For the semantic aspect, our approach makes use of a co-occurrence lexical network: it gathers the words semantically close and helps improve the representation of documents to filter and therefore increases the chances of matching a document with a profile. Our filtering architecture is completely independent from the domain of knowledge. It has a modular structure, allowing it to eventually adapt to any extension and modification. The knowledge specific to different profiles are automatically generated (linguistic). Indeed, each profile is calculated by automated analysis of the content to produce a set of terms and the linguistic properties characterizing it. The results obtained through different experiments, seem interesting. They allowed us to validate the interest of linguistic knowledge and machine learning in the improvement of the performance of an information filtering system and to demonstrate the applicability and the contribution of an agents based implementation in the information filtering process. With these results, we can conclude that we can make is that the automatic learning is a must in designing and improving the performances of a dynamic information filtering system, and that linguistic methods combined with statistical methods look promising for effective cognitive filtering.

References

[1] Burke R., "Hybrid Systems for Personalized Recommendations," *Book Chapter in Intelligent Techniques for Web Personalization*, vol. 3169, Springer, 2005.

[2] Caropreso M., Matwin S., and Sebastiani F., *A Learner-Independent Evaluation of the Usefulness of Statistical Phrases for Automatic Text Categorization*, Hershey, USA, 2001.

[3] Cattuto C., Szomszor M., Alani H., O'Hara K., Baldassarri A., Loreto V., and Servedio V., "Folksonomies, the Semantic Web, and Movie Recommendation," *in Proceedings of the European Semantic Web Conference (ESWC'07),* 2007.

[4] Copeck T., Barker K., Delisle S., and Szpakowicz S., "Automating the Measurement of Linguistic Features to Help Classify Texts as Technical," *in Proceedings of TALN*, Lausanne, pp. 101-110, 2000.

[5] Groh G. and Ehmig C., "Recommandations in taste related domains : Collaborative Filtering vs. Social Filtering," *Proceedings of the international ACM conference on Supporting group work*, 2007.

[6] Joachims T., "Text Categorization with Support Vector Machines: Learning with Many Relevant Features," *in Proceedings of 16th European Conference on Machine Learning*, pp. 137-142, 1999.

[7] Karen H., Marinho L., and Schmidt-Thieme L., "Tagaware Recommender Systems by Fusion of Collaborative Filtering Algorithms," *in Proceedings of ACM SAC'08*, pp. 1995-1999, 2008.

[8] Kilander F., Fahraeus E., and Palme J., "Intelligent Information Filtering," *Technical Report 97-002, Department of Computer and Systems Sciences*, Stockholm University, 1998.

[9] Kong F. and Wang R., "Semantic-Enhanced Personalized Recommender System," *in Proceedings of the Internationnal Conference on Machine Learning and Cybernetics*, Hong Kong, pp. 4069-4074, 2007.

[10] Lops P., Degemmis M., and Semeraro G., "Improving Social Filtering Techniques Through Word Net-Based User Profiles," *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, 2007.

[11] Nguyen A., Denos N., and Berrut C., "Modèle D'espaces De Communautés Basé Sur La Théorie Des Ensembles D'approximation Dans Un Système De Filtrage Hybride," *Actes De La 3ème Conférence en Recherche Information et Applications*, Lyon, 2006.

[12] Nouali O. and Blache P., "Filtrage Automatique de Courriels : Une Approche Adaptative et Multiniveau," *Annals of Telecommunications*, Hermes Science, Get, Lavoisier, vol. 60, no. 11-12, pp. 1466-1487, 2005.

[13] Nouali O. and Krinah A., "Improvement of a Retrieval and Filtering Systems by An Automatic Multi Words Extraction Tool," *in Proceedings of CSIT2006 4th International Conference on Computer Science and Information Technology*, Applied Science University, Jordan, pp. , 2006.

[14] Nouali O. and Krinah A., "Vectorial Information Structuring for Documents Filtering and Diffusion," *Computer Juornal International Arab Journal of Information Technology*, vol. 5, no. 1, pp. 1-6, 2008.

[15] Oufaida H. and Nouali O., "Le Filtrage Collaboratif Et Le Web 2.0, Etude De l'état de l'art", *Document Numérique*, Lavoisier, no. 1-2, 2008.

[16] Yang Y. and Pedersen J., "A comparative Study on Feature Selection in Text Categorization," *in Proceedings of International Conference on Machine Learning ICML*, Nashville, TN, USA, 1997.

Nouali Omar is a research associate at the Theory and computer engineering, DTISI, CERIST, Algeria. His research interests include artificial intelligence, expert systems, natural language processing, information filtering, computer security, and human computer interface.

Nadia Nouali-Taboudjemat is a research associate at the Theory and computer engineering, DTISI, CERIST,Algeria. Her research interests include wireless networks,ad hoc networks, distributed computing, mobile computing,Mobile databases and transactions, and security.

Bernard Toursel is professor in computer science at the Computer Science Laboratory of Lille, University of Lille-1, France. His research interests include parallel and distributed computing.