Recommending Biomedical Resources: A Fuzzy Linguistic Approach Based on Semantic Web

J. M. Morales-del-Castillo,[1,*] Eduardo Peis,[2,†] Antonio A. Ruiz,[2,‡]
E. Herrera-Viedma[3,§]
[1]*Department of Library and Information Science, Carlos III University, Madrid, Spain*
[2]*Department of Library and Information Science, University of Granada, Granada, Spain*
[3]*Department of Computer Science and Artificial Intelligence, University of Granada, Granada, Spain*

One of the key issues in dynamic research areas, such as that of biomedical sciences, is the development of tools capable to retrieve and provide users relevant resources from large repositories according to their information needs. In this paper, we present a filtering and recommender system that applies Semantic Web technologies and fuzzy linguistic modeling techniques to provide users valuable information about resources that fit their interests. To carry out the recommendation process, we have defined three software agents (interface, task, and information agents) that are distributed in a five level hierarchical architecture. The system is also capable of to deal with incomplete information to define enriched user profiles and, therefore, soften the problem of *cold start*. A simple evaluation has been carried out, and the experimental outcomes reveal a reasonable good performance of the system in terms of precision and recall.

1. INTRODUCTION

The exponential growth of the specialized literature in dynamic and very productive domains has become the main handicap for information systems. Far are the days when a researcher could select by hand documents of his interest using an abstract bulletin, and now providing relevant resources to information to consumers has become a difficult task. Biomedical Sciences are not strangers to this situation and keeping up with the latest researching trends and breakthroughs on a specific

*Author to whom all correspondence should be addressed; e-mail: jmdel@bib.uc3m.es.
†e-mail: epeis@ugr.es.
‡e-mail: aangel@ugr.es.
§e-mail: viedma@decsai.ugr.es.

INTERNATIONAL JOURNAL OF INTELLIGENT SYSTEMS, VOL. 25, 1143–1157 (2010)

View this article online at wileyonlinelibrary.com. • DOI 10.1002/int.20447

specialty is time consuming and requires researchers to make a personal effort. Although the Web provides an easier and more comprehensive access to information resources than *physical* environments do, current web services have shown their inability to provide an accurate and efficient response to users' requirements, since information in the Web is basically represented using natural language and machines aren't capable to interpret and contextualize it. Therefore, it is becoming necessary to develop systems for searching and mining the Web that allow improving the access to information in an efficient way.

Traditionally, this problem has been faced by the development of filtering and recommendation services, which are based on the application of different techniques that manage a series of processes that are oriented to provide users just the information that meets their needs or is of interest to them. In textual domains these services are usually developed using multiagent systems to evaluate and filter resources normally represented in XML or HTML format,[1,2] and assist people in search and retrieval tasks.[3–5]

Basically, these systems can be classified in two main categories[5,6]: content-based and collaborative recommendation systems. Content-based recommendation systems filter information and generate recommendations by comparing a set of keywords defined by the user with the terms that represent the content of documents, ignoring any information given by other users. On the other hand, collaborative filtering systems use the information provided by several users to recommend documents to a specific user, ignoring the different ways the content is represented. Nevertheless, these systems show a low performance when there are few recommendations or deal with uncommon likes. The current trends to soften this handicap are the development of hybrid systems,[7] which combine both recommender approaches, and knowledge-based recommender systems[8] that require users to define their preferences to obtain a personal profile. At this moment, there are several recurrent technologies that help developing these value-added services. Among them we can find:

- *Intelligent software agents*[9] that can actively process and exchange information with another agents in the Web,[10] and also assist users in information retrieval tasks.[11–12]
- *Fuzzy linguistic techniques* that can help allow representing qualitative phenomena from a quantitative approach[13] and even deal with incomplete information.[14]
- *Semantic Web technologies*[15] that can be used as common syntactic and data model framework for representing information and enabling software agents to access and process resources at a semantic level. Examples of knowledge-based applications developed using these technologies in the field of biomedical sciences are *Biogateway Portal*[16] and the *National Cancer Institute Thesaurus*.[17]

Our proposal consists of the development of a multiagent filtering and recommender system specialized in biomedical resources that jointly apply Semantic Web technologies and fuzzy linguistic modeling techniques to provide researchers a better access to resources of their interest.

The paper is structured as follows. In Section 2, we briefly discuss the theoretical background of the system (such as Semantic Web technologies, fuzzy linguistic modeling, and the management of information incomplete). In Section 3, the main features, structure, and functional modules of the system are presented. The outcomes of a simple experiment to evaluate the system are presented in Section 4, and finally some conclusions are pointed out in the latter section.

2. THEORETICAL BACKGROUND

The system here proposed is conceived as a multiagent model defined to ease users the access to specialized information they required by recommending a selection of the latest (or more interesting) resources published in a specific domain (in this case, biomedicine) according to users' profiles that has been generated from a partial expression of their information needs.

In a nutshell, this model is developed by the application of Semantic Web technologies[15,18] to improve *user–agent* and *agent–agent* interaction and to settle a semantic framework where software agents can process and exchange information using Web ontologies[19,20] (or simpler semantic structures like conceptual schemes or thesauri), and fuzzy linguistic modeling techniques,[13] which allow dealing with linguistic information that has a certain degree of uncertainty (as, for instance, when quantifying users' satisfaction in relation to a product or service or when defining their preferences).

To clarify the operation and structure of the model, its main theoretical pillars are herewith explained.

2.1. Semantic Web Technologies

The Semantic Web[15] tries to extend the model of the present Web using a series of standard languages that enable the description of Web resources to be enriched so that they become semantically accessible. To do that, the Semantic Web is based on two fundamental ideas: (i) semantic tagging of resources, so that information can be understood both by humans and computers and (ii) the development of intelligent agents[10] capable of operating at a semantic level with those resources and inferring new knowledge from them (in this way it is possible shifting from keyword search to the retrieval of concepts).

The semantic backbone of the project is the RDF (*Resource Description Framework*) vocabulary,[21] which provides a data model to represent, exchange, link, add and reuse structured metadata of distributed information sources and, therefore, make them directly understandable by software agents. The RDF structures the information into individual assertions (resource, property, and property value triples) and uniquely characterizes resources by means of uniform resource identifiers URIs, allowing agents to make inferences about them using Web ontologies[19–20] or to work with them using simpler semantic structures like conceptual schemes or thesauri.

As we can see, the Semantic Web basically works with information written in natural language (although structured in a way that can be interpreted by machines). For this reason, it is usually difficult to deal with problems that require operating with linguistic information that has a certain degree of uncertainty (such as, for instance, when quantifying the user's satisfaction in relation to a product or service).

A possible solution could be the use of fuzzy linguistic modeling techniques[13] as a tool for improving the communication between system and user.

2.2. Ordinal Fuzzy Linguistic Modeling

Fuzzy linguistic modeling[13] provides a set of approximate techniques appropriate to deal with qualitative aspects of problems. The ordinal linguistic approach[22–23] is defined according to a finite set S of linguistic labels arranged on a total order scale and with odd cardinality (seven or nine tags):

$$\{s_i, i \in H = \{0, \ldots, T\}\}$$

The central term has a value of "approximately 0.5," and the rest of the terms are arranged symmetrically around it. The semantics of each linguistic term is given by the ordered structure of the set of terms, considering that each linguistic term of the pair (s_i, s_{T-i}) is equally informative. Each label s_i is assigned a fuzzy value defined in the interval, [0,1] that is described by a linear trapezoidal property function represented by the following four-tupla $(a_i, b_i, \alpha_i, \beta_I)$, where the first two parameters show the interval where the property value is 1.0, the third and fourth parameters show the left and right limits of the distribution.

In addition, the following properties have to be specified:

(1) *The set is ordered:* $s_i \geq s_j$ if $i \geq j$.
(2) *Negation operator:* $\text{Neg}(s_i) = s_j$, with $j = T - i$.
(3) *Maximization operator:* $\text{MAX}(s_i, s_j) = s_i$ if $s_i \geq s_j$.
(4) *Minimization operator:* $\text{MIN}(s_i, s_j) = s_i$ if $s_i \leq s_j$.

In the scope of the ordinal fuzzy linguistic model for computing with words it is also possible defining aggregation operators, as the Linguistic Weighted Averaging operator (LWA),[22] which is capable of combining and operating with linguistic information with nonequal importance.

2.3. Dealing with Incomplete Information

As aforementioned, one of the main problems of recommender systems is the problem of lacking enough information about users' needs, which can lead to inaccurate recommendations (also known as *cold start* problem[24]). Therefore, an adequate representation of users' preferences could improve the performance of these services. One possibility of representing preferences about a set of items $X = \{x_1, \ldots, x_n\}$ is the specification of preference relations. Nevertheless, when the set of items to be assessed is too large this task can be tedious and time consuming for users and, as a result, many values in the preference relation could be missed[25].

Usually, a fuzzy preference relation can be characterized by a membership function μ_P: $X \times X \rightarrow [0, 1)$ and represented by an $n \times n$ matrix $P = (p_{ij})$, being

$p_{ij} = \mu_P\ (x_i, x_j)$ the degree in which item x_i is preferred regarding to $x_{j,}$ where

- $p_{ij} = 0.5$ indicates indifference between x_i and $x_{j,}$
- $p_{ij} = 1$ indicates that x_i is absolutely preferred to $x_{j,}$ and
- $p_{ij} > 0.5$ indicates that x_i is preferred to $x_{j.}$

Consequently, in a linguistic context, i.e., being $S = \{s_0, \ldots, s_T\}$ the linguistic term set used to represent the preferences, we can define a fuzzy linguistic preference relation $P = p_{ij}(\forall i, j \in \{1 \ldots n\})$ on X by means of a linguistic membership function $\mu_{P:}$: $X \times X \rightarrow S$, where $p_{ij} = \mu_P\ (x_i, x_j)$ is a linguistic label.

Dealing with preference relations, missing information is a problem that needs to be addressed because it is not always possible for the experts to provide all the possible preference assessments on the set of alternatives. A missing value in a linguistic preference relation is not equivalent to a lack of preference of one alternative over another. A missing value can be the result of the incapacity of an expert to quantify the degree of preference of one alternative over another. It must be clear then that when an expert is not able to express the particular value p_{ij}, because he/she does not have a clear idea of how better alternative x_i is over alternative x_j, this does not mean that he/she prefers both options with the same intensity. To model these situations, in the following definitions, we express the concept of an incomplete fuzzy linguistic preference relation:

DEFINITION 1. *A function f: $X \times Y$ is partial when not every element in the set X necessarily maps to an element in the set Y. When every element from the set X maps to one element of the set Y then we have a total function.*

DEFINITION 2. *A fuzzy linguistic preference relation P on a set of alternatives X with a partial membership function is an incomplete fuzzy linguistic preference relation.*

Consequently, given an incomplete fuzzy linguistic preference relation P in which there is at least a linguistic missing value, we propose to obtain a complete fuzzy linguistic preference relation P^* using the algorithm suggested by Alonso et al.[26]:

$$P = \begin{pmatrix} - & p_{12} & p_{13} & p_{14} & p_{15} \\ x & - & x & x & x \\ x & x & - & x & x \\ x & x & x & - & x \\ x & x & x & x & - \end{pmatrix} \rightarrow P^* = \begin{pmatrix} - & p_{12} & p_{13} & p_{14} & p_{15} \\ p_{21}^* & - & p_{23}^* & p_{24}^* & p_{25}^* \\ p_{31}^* & p_{32}^* & - & p_{34}^* & p_{35}^* \\ p_{41}^* & p_{42}^* & p_{43}^* & - & p_{45}^* \\ p_{51}^* & p_{52}^* & p_{53}^* & p_{54}^* & - \end{pmatrix}$$

3. ARCHITECTURE AND MODULUS OF THE SYSTEM

In this section, we present a new multiagent model that uses three software agents (interface, task, and information agents) whose activity is distributed in a hierarchical architecture composed by five levels, to develop the filtering and recommendation process:

Level 1: User level. In this level, users interact with the system by defining their preferences, providing feedback to the system, etc.

Level 2: Interface level. This is the level defined to allow interface agent developing its activity as a mediator between users and the task agent. It is also capable of carry out simple filtering operations on behalf of the user.

Level 3: Task level. In this level where the task agent (normally one per interface agent) carries out the main load of operations performed in the system such as the generation of information alerts or the management of profiles and RSS feeds.

Level 4: Information agents level. Here is where several information agents can access system's repositories, thus playing the role of mediators between information sources and the task agent.

Level 5: Resources level. In this level are included all the information sources the system can access:

- A set of RSS feeds in RDF format[27] that play the role of current-awareness bulletins. The structure of these feeds comprises two areas: a first one where the feed is described by a series of basic metadata and another area containing the representation of each resource to be recommended (i.e., items including basic information as its title, author/s, URL, abstract, and so on). Besides, because of the widespread of RSS technologies on the Web, items can directly be checked out in ordinary Web browsers (since RSS readers have become a common feature in the vast majority of them).
- A user profile repository that stores the user profiles generated in the *user profiles generation module* (see below the description of the system's modules). Each profile contains an area of personal/ID information, an area of weighted preferences representing the interests of the user, and a *recommendations log* where the different recommendations provided by the user about any document are recorded.
- A test thesaurus in SKOS[28] format, which has been developed taking as a model the *National Cancer Institute Thesaurus*,[17] is an open content license tool, which includes more than 34,000 concepts from different cancer research domains and that is structured into 20 taxonomic trees. Our test thesaurus contains 11 main classes (which have been defined merging and discarding some of the original taxonomic trees of the NCI thesaurus), and around 2000 concepts distributed in six depth levels.
- A document repository (in this case we have opted for using the public database PubMed[29]), which is described by a set of nonweighted subjects extracted from the system thesaurus.

The underlying semantics of the different elements that make up the system (i.e., their characteristics and the semantic relations defined among them) are defined through several interoperable Web ontologies described using the OWL vocabulary.[30] Furthermore, since the communication processes carried out among agents in this model involve natural language information and fuzzy linguistic tags, we have chosen to use the adaptation of the FIPA agent communication language[31] proposed by Willmott et al.[32], which is based on XML syntax and RDF/OWL as content language.

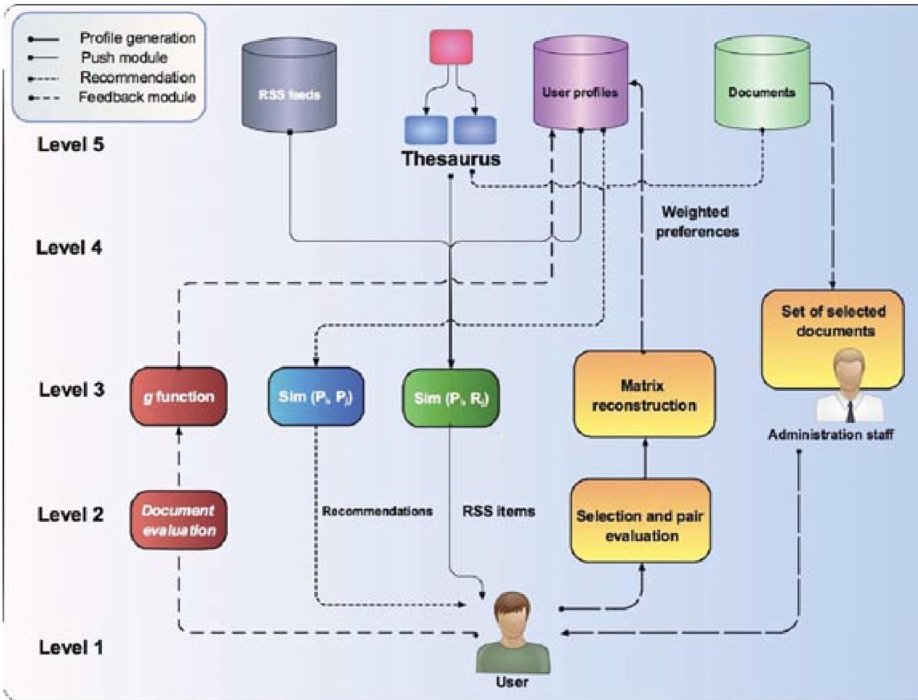

Figure 1. System levels and modules.

In the system, there are also defined five main activity modules (see Figure 1):

3.1. User Profiles Generation Module

In this module, the system generates the user profiles that are eventually stored in the profiles repository. The system needs a set of sample resources whose content can be considered as a representative within the scope of the system. These items are selected by the system's staff and will be used to define the weighted preferences to be included in each user profile. To do so, each sample resource will be represented using a set of 11 subject areas corresponding to the main categories defined in the system's thesaurus. This task is carried out by the information system's staff and consists in pondering each category according to its correspondence with the content of the document using a linguistic tag extracted from fuzzy linguistic set $S = \{s_0 = \text{very low} = \text{VL}, s_1 = \text{low} = \text{L}, s_2 = \text{indifferent} = \text{I}, s_3 = \text{high} = \text{H}, s_4 = \text{very high} = \text{VH}\}$. Eventually, we get a resource described by a vector of 11 elements $x_i = \{s_{i1}, s_{i2}, \ldots, s_{i11}\}$.

Once we got a battery of representative documents, when a user logs into the system for the first time he is presented a set of 10 abstracts, which corresponds to

10 of these selected resources and must choose from this list those five documents that better fit to his interests.

Afterward, the user must assess pairs of documents (not necessarily all the possible pairs) using a linguistic label $p_{ij} \in S$, which represents the preference degree of resource x_i regarding to resource x_j. The only imposed condition is that there have to be at least one assessment for each resource, and therefore, we assume that it could exist linguistic missing values. In such a way, the system is capable of reconstructing the whole fuzzy linguistic preference relation matrix as explained in Ref. 26.

Afterward, the linguistic preference degree (DG) of each resource for a specific user is computed applying a linguistic aggregation operator as the LOWA operator[23]:

$$DG_i = \text{LOWA}\left[\int p^*_{i\,1}, \ldots, p^*_{i\,5}\right]$$

Finally, the different linguistic degrees DG_i obtained from the preference relation are aggregated using the LWA operator,[22] thus, generating a unique set of weighted preferences that are stored in the user profile.

3.2. Information Push Module

This module is responsible for generating and managing the information alerts to be provided to users (so it can be considered as the service core). The similarity between user profiles and resources is measured according to the hierarchical exponential operator defined by Oldakowsky and Byzer,[33] which takes into account the position of the concepts to be matched in a taxonomic tree. Selective dissemination of information service performance is based on generating passive queries to RSS feeds about the preferences stored in the user's profile without the need of an explicit request from the user (an information delivery technique known as information push). In such a way, users are alerted to new resources fitting their information requirements without having to request them each time they access the system. This process is developed as follows:

- *Step 1*: Users must provide their login and password to get authenticated access to the system.
- *Step 2:* Once the user is identified, the task agent proceeds to match the user's preferences with the content descriptors of the n items in the RSS feed, thus identifying those resources that better fit to user's specific information needs. In this case, instead of using a lexical matching of the strings of both terms, the task agent measures their semantic similarity. To do this, we use the semantic similarity function defined by Oldakowsky and Byzer[33] that allows measuring the distance between two concepts in a taxonomy (or thesaurus) described as an RDF graph. The similarity *sim* between two concepts c_1 and c_2 in the thesaurus is defined as follows:

$$\text{sim}_c(c_1, c_2) = 1 - d_c(c_1, c_2)$$

The distance d_c between two concepts represents the path to be followed to get from one to another through their closest common parent (*ccp*). This distance is measured as follows:

$$d_c(c_1, c_2) = d_c(c_1, ccp) + d_c(c_2, ccp)$$

$$d_c(c, ccp) = milestone\ (ccp) - milestone\ (c)$$

where each concept in the taxonomy is assigned a milestone value that depends on the depth of the concept in the taxonomic tree. This milestone can be measured by applying both a linear or exponential function (depending on the characteristics and requirements of our system). In our case, we have chosen the *exponential milestone calculator*, which is defined as follows[34]:

$$milestone(n) = 1/2\ k^{(n)}$$

where n is the depth of the concept in the thesaurus and $k$is a factor larger than 1 that defines the rate at which the milestone values decrease along the hierarchy (this value can be modified depending on the depth of the thesaurus if a fine-tuning of the system is required). The exponential factor provides a better measurement of the similarity since it makes the operator *sensitive* to the *kinship* of concepts (i.e., given a specific concept it is more similar to a "*child concept*" than to a "*sibling concept*") and their depth in the hierarchical tree (the similarity among broad concepts is smaller than among specific ones).

Once the similarity between preferences and topic terms is defined, the relevance of resources or profiles is calculated according to the concept of *semantic overlap*. This concept tries to ease the problem of measuring similarity using taxonomic operators since all the concepts in a taxonomy are related in a certain degree and therefore the similarity between two of them would never reach 0 (i.e., we could find relevance values higher than 1 that can hardly be normalized). The underlying idea in this concept is determining areas of maximum semantic intersection between the concepts in the taxonomy. To obtain the relevance of profiles according to other items (resources or profiles of other users), we define the following function:

$$\mathrm{Sim}\,(P_i, P_j) = \frac{\sum_{k=1}^{\mathrm{MIN}(N,M)} H_k(\mathrm{Sim}\,(\alpha_i, \delta_j))\left(\frac{\omega_i+\omega_j}{2}\right)}{\mathrm{MAX}(N, M)}$$

where H_k (Sim (α_i, δ_j)) is a function that extracts the k maximum similarities defined between the preferences of$P_i = \{\alpha_1, \ldots, \alpha_N\}$ and $P_j = \{\delta_1, \ldots, \delta_M\}$, ω_i, ω_j are the corresponding associated weights with α_i and δ_j, and N and M are the number of preferences of the profiles P_i and P_j, respectively. When matching profiles $P_i = \{\alpha_1, \ldots, \alpha_N\}$ and items $R_j = \{\beta_1, \ldots, \beta_M\}$, since the subjects that describe the resources are not weighted, we will take into account only the weights associated with preferences so the function in this case is slightly different:

$$\mathrm{Sim}\,(P_i, R_j) = \frac{\sum_{k=1}^{\mathrm{MIN}(N,M)} H_k(\mathrm{Sim}(\alpha_i, \beta_j))\omega_i}{\mathrm{MAX}(N, M)}$$

- *Step 3*: User is provided with a set of RSS items that are semantically close to their interests and that can be checked out in any Web browser with an integrated RSS reader. In addition, each item links to a document in the document repository of the system and is displayed together with a recommendation value (see the *Collaborative Recommendation Module*).
- *Step 4*: Once the user access a document through the link he can find in the RSS item of his interest, he is asked to assess the checked document using a linguistic label (see the *User profiles updating module*) whose value is stored in the *recommendations log* of his profile (which will be useful later on in the *Collaborative Recommendation module*).

3.3. Feedback or User Profiles Updating Module

In this module, the updating of user profiles is carried out according to users' assessments about the set of resources recommended by the system. This updating process consists of recalculating the weight associated with the preference of the user profile with the highest similarity value when matched with the topics of the RSS item (preference which is supposed to better represent the interests of the user for that query). We have defined a matching function that rewards those preferences that are present in resources positively assessed by users and penalized them, on the contrary, when this assessment is negative. We must note that this assessment should be interpreted as the perception of the user about the adequacy of the suggested resource to their interests, not as an evaluation of the quality of the content. According to these premises, let $e_j \in S'$ be the degree of satisfaction provided by the user, and $\omega_{il}^{j} \in S$ the weight of property i (in this case $i =$ "Preference") with value l. Then we define the following updating function g: $S' \times S \to S$:

$$g\left(e_j, \omega_{li}^{j}\right) = \begin{cases} S_{\mathrm{Min}(a+\beta,T)} & \text{if} \quad S_a \leq S_b \\ S_{\mathrm{Max}(o,a-\beta)} & \text{if} \quad S_a > S_b \end{cases}$$

$$s_a, s_b \in S|a, b \in H = \{0, \ldots, T\}$$

where (i) $s_{a=} \omega_{li}^{j}$, (ii) $s_{b=} e_j$, (iii) a and b are the indexes of the linguistic labels whose value ranges from 0 to T (where T is the number of labels of the set $S - 1$), and (iv) β is the value that rewards or penalizes the preference weight according to users' assessments (i.e., the weights assigned to the preferences in the *users profile generation module* are modified in an attempt to capture subtle changes in users' interests). It is defined as $\beta = round\ (2|b–a|/T)$, where *round* is the typical round function.

3.4. Collaborative Recommendation Module

The aim of this module is generating recommendations about a specific resource in base to the assessments provided by different experts with a profile similar to that of the active user. The different recommendations values for the selected document (which are sought in the *recommendations log* of experts' profile) are aggregated using the LWA operator[22] and displayed next to the link of the RSS item describing the document suggested by the system (as an aid for the user to discern whether a resource is of his interest or not). Another functionality of the system will be allowing users to explicitly know the identity and institutional affiliation data of these experts to contact them for any research purposes. This feature of the system implies a total commitment between the service and its users since their altruistic collaboration can only be achieved by granting that their data will exclusively be used for contacting other researchers subscribed to the service. Therefore, it becomes a critical issue defining privacy policies to protect those individuals that prefer to be *invisible* for the rest of users. This is the reason for which this module has not been fully implemented yet.

4. EVALUATION OF THE SYSTEM

We have set up a simple experiment to evaluate the content-based module of the system in terms of precision[35] and recall.[36] These two measures (together with the F1 score[37] are usually used in filtering and recommender systems to assess the quality of a set of retrieved resources.

In order assess the set of retrieved items (N), we have defined in our experiment three categories of resources: relevant items (Nr), selected items (Ns), and relevant suggested items (Nrs). According to these categories, precision, recall, and F1 score are defined as follows:

Precision: A ratio of selected relevant items to selected items, i.e., the probability of a selected item to be relevant.

$$P = \mathrm{Nrs/Ns}$$

Recall: A ratio of selected relevant items to relevant items, i.e., the probability of a relevant item to be selected.

$$R = \mathrm{Nrs/Nr}$$

$F1$: A combination metric that equals both the weights of precision and recall.

$$F1 = (2^* P^* R)/(P + R)$$

The goal of the experiment is to test the performance of our system in the generation of accurate and relevant content-based recommendations for the users of the system, exclusively considering the monodisciplinary search. To do so, we have asked a random sample of 10 researchers in the field of biomedicine to evaluate the results provided by the system. In addition, we require two elements:

- an RSS feed that contains 30 items extracted from the PubMed repository,[29] from which only 10 of them are semantically relevant.
- a set of user profiles.

The system is set to recommend up to 10 resources and then users are asked to assess the results by explicitly stating which of the recommended items they

Table I. Experimental Data.

	User1	User2	User3	User4	User5	User6	User7	User8	User9	User10
Nrs	5	4	6	3	4	5	6	7	4	5
Nr	6	6	8	6	6	6	9	9	5	8
Ns	10	10	10	10	10	10	10	10	10	10

Table II. Detailed Experimental Outcomes.

	User1	User2	User3	User4	User5	User6	User7	User8	User9	User10	Aver.
P	50.00	40.00	60.00	30.00	40.00	50.00	60.00	70.00	40.00	50.00	49.00
R	83.33	66.67	75.00	50.00	66.67	83.33	66.67	77.78	80.00	62.50	71.01
F1	62.50	50	66.67	37.50	50.00	62.50	63.16	73.68	53.33	55.56	57.99

All values are in percent.

consider are relevant. With these starting premises, the experiment was carried out and the results are shown in Table I.

Precision, recall, and F1 for each user are shown in Table II (in percentage) and represented in the graph in Figure 2. The average outcomes reveal a quite good performance of the system (nearly close to the 50% in terms of precision).

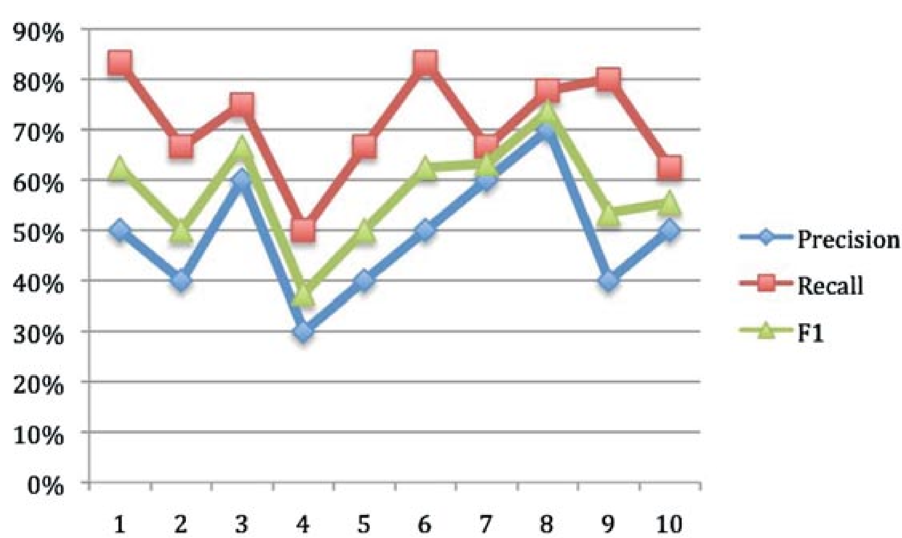

Figure 2. Precision, recall and F1.

5. CONCLUSIONS

In this paper, we have presented a multiagent filtering and recommender system (designed to be used by biomedical researchers) which provides an integrated solution to minimize the problem of access relevant information in vast document repositories.

The system combines Semantic Web technologies and a fuzzy linguistic modeling approach to define a richer description of information, thus improving communication processes and user–system interaction. Among its functionalities, the system includes a module that is capable of generating users' profiles from an incomplete fuzzy linguistic preference relation.

It has also been evaluated, and experimental results show that it is reasonably effective in terms of precision and recall, although further detailed evaluations may be necessary.

In future, we think to improve our proposal with the application of others tools related with the Web navigation[38] and clustering.[39]

Acknowledgments

This paper has been developed with financing of the Andalusian excellence project (TIC05299), FEDER funds in PETRI project (PET2007-0460) and FUZZYLING project (TIN2007-61079).

References

1. Herrera-Viedma E, Peis E. Evaluating the informative quality of documents in SGML-format using fuzzy linguistic techniques based on computing with words. Inform Process Manage 2003;39(2):233–249.
2. Herrera-Viedma E. Peis E, Morales-del-Castillo, JM, Alonso S, Anaya EK. A fuzzy linguistic model to evaluate the quality of web sites that Store XML documents. Int J Approx Reason 2007;46(1):226–253.
3. Resnick P; Varian HR. Recommender systems. Commun ACM 1997;40(3):56–58.
4. Shapira B, Hanani U, Raveh A, Shoval P. Information filtering: A new two-phase model using stereotypic user profiling. J. Intell Inform Syst 1997;8(2):155–165.
5. Porcel C, López-Herrera AG, Herrera-Viedma E A. Recommender system for research resources based on fuzzy linguistic modeling. Expert Syst Appl 2009;36(3)5173–5183.
6. Porcel, C, Moreno, J.M, Herrera-Viedma, EA. multi-disciplinar recommender system to advice research resources in University Digital Libraries. Expert Syst Appl 2009;36(10):12520–12528.
7. Burke R. Hybrid recommender systems: Survey and experiments. User Model User-Adapted Interact 2002;12(4):331–370.
8. Burke R. Knowledge-based recommender systems. Encycl Library Inform Syst 2000;69(32):180–201.
9. Ferber J. Multi-agent systems: An introduction to distributed artificial intelligence, New York: Addison-Wesley Longman; 1999.
10. Hendler J. Agents and the semantic web. IEEE Intell Syst 2001;16(2):30–37.
11. Brenner W, Zarnekow R, Witting H. Intelligent software agent, foundations, and applications. Heidelberg, Germasny: Springer-Verlag; 1998.

12. Fazlollahi B, Vahidov RM, Aliev RA. Multi-agent distributed intelligent system based on fuzzy decision making, Int. J. Intell. Syst, 2000;15(9):849–858.
13. Herrera F Herrera-Viedma E. Linguistic decision analysis: Steps for solving decision problems under linguistic information. Fuzzy Sets Syst 2000;115(1):67–82.
14. Cabrerizo FJ, Pérez IJ, Herrera-Viedma E. Managing the consensus in group decision Making in an unbalanced fuzzy linguistic context with incomplete information.Knowl.-Based Syst, 2010;23(2):169–181.
15. Berners-Lee T, Hendler J, Lassila O. The Semantic Web. The Sci Am 2001;284(5):34–43.
16. Biogateway, 2009. http://www.semantic-systems-biology.org/home. Accessed on June 15, 2010.
17. National cancer institute thesaurus. Available online at http://www.mindswap.org/2003/CancerOntology/ Accessed on June 17, 2010.
18. Berners-Lee T Semantic Web road map; 2000. Available online at http://www.w3.org/DesignIssues/Semantic.html. Accessed on June 17, 2010.
19. Gruber TR. Toward principles for the design of ontologies used for knowledge sharing. Int J Hum- Comput Stud, 1995;43 (5–6):907–928.
20. Guarino N. Formal ontology and information systems. In Guarino N. editor. Formal ontology in information systems. Amsterdam, Netherlands :IOS Press; 1998; pp. 3–17.
21. Beckett D. (editor). RDF/XML syntax specification; 2004. Available online at http://www.w3.org/TR/rdf-syntax-grammar/. Accessed on June 16, 2010.
22. Herrera F, Herrera-Viedma E. Aggregation operators for linguistic weighted information. IEEE Trans Syst Man Cybern 1997;27:646–656.
23. Herrera F, Herrera-Viedma E, Verdegay JL. Direct approach processes in group decision making using linguistic OWA operators. Fuzzy Sets Sys 1996;79(2):175–190.
24. Schein AI., Popescul A, Ungar LH. Methods and metrics for cold-start recommendations. In Jarvelin K, Beaulieu M, Baeza-Yates R, Myaeng SH. editors, Proc. of the 25th annual international ACM SIGIR conference on research and development in information retrieval (SIGIR 2002). New York: ACM; 2002. pp 253–260.
25. Herrera-Viedma E, Alonso S, Chiclana F, Herrera F A. Consensus model for group decision making with incomplete fuzzy preference relations. IEEE Trans Fuzzy Syst 2007;15(5):863–877.
26. Alonso S, Cabrerizo FJ., Chiclana F, Herrera F, Herrera-Viedma E, Group decision making with incomplete fuzzy linguistic preference relations. Int J Intell Syst 2009;24(2):201–222.
27. RSS-DEV working group, RDF site summary (RSS) 1.0; 2001. Available online at http://web.resource.org/rss/1.0/spec. Accessed on June 16, 2010.
28. Isaac A, Summers E, editors. SKOS simple knowledge organization system primer; 2008. Available online at http://www.w3.org/TR/skos-primer/. Accessed on June 15, 2010.
29. PubMed; 2009. Available online at http://www.ncbi.nlm.nih.gov/pubmed/. Accessed on June 15, 2010.
30. McGuinness DL., van Harmelen F. Editors, OWL Web ontology language overview; 2004. Available online at http://www.w3.org/TR/2004/REC-owl-features-20040210/. Accessed on June 17, 2010.
31. Foundation for intelligent agents (FIPA); 2009. Available online at http://www.fipa.org. Accessed on June 16, 2010.
32. Willmott S, Peña FOF, Merida-Campos C, Constantinescu I, Dale J, Cabanillas D. Adapting agent communication languages for semantic Web service inter-communication. In Skowron, A., et.al. editors Proc. of the IEEE/WIC/ACM International Conference on Web Intelligence 2005. New York: IEEE Society; 2005. pp 405–408.
33. Oldakowsky R, Byzer C. SemMF: A. framework for calculating semantic similarity of objects represented as RDF graphs; 2005. Available online at http://www.corporate-semantic-web.de/pub/SemMF_ISWC2005.pdf. Accessed on June 17, 2010.
34. Zhong J, Zhu H, Li J, Yu Y. Conceptual graph matching for semantic search. In: The 2002 International Conference on Computational Science (ICCS2002). LNCS, Vol 2393. Berlin: Springer-Verlag; 2002. pp 92–96.

35. Cao Y, Li Y. An intelligent fuzzy-based recommendation system for consumer electronic products. Expert Syst. Appl., 2007;33(1):230–240.
36. Cleverdon CW, Mills J, Keen EM. Factors determining the performance of indexing systems, Vol. 2: Test Results. Cranfield, UK: ASLIB Cranfield Project; 1966.
37. Sarwar B, Karypis G, Konstan J, Riedl J. Analysis of recommendation algorithms for e-commerce. In Jhingran A, Mason JM, Tygar D. editors In: Proc. of ACM E-Commerce 2000 conference, New York: ACM; 2000. pp 158–167.
38. LoiaV, Pedrycz W, Senatore S, Sessa MI. Web navigation support by means of proximity-driven assistant agents. J Am Soc Inform Sci Technol 2007;57(4):515–527.
39. Loia V, Pedrycz W, Senatore S. Semantic web content analysis: A study in proximity-based collaborative clustering. IEEE Trans Fuzzy Syst 2007;15(6):1294–1312.