A Fuzzy Linguistic Recommender System to Advice Research Resources in University Digital Libraries

Enrique Herrera-Viedma[1], Carlos Porcel[2], Antonio Gabriel López-Herrera[3], and Sergio Alonso[1]

[1] Dept. of Computer Science and Artificial Intelligence, University of Granada
18071 - Granada, Spain
viedma@decsai.ugr.es, salonso@decsai.ugr.es

[2] Dept. of Computing and Numerical Analysis, University of Córdoba
14071 - Córdoba, Spain
carlos.porcel@uco.es

[3] Dept. of Computer Science, University of Jaén. 23071 - Jaén, Spain
agabriel@ujaen.es

Summary. As it is known the Web is changing the information access processes. The Web is one of the most important information media. Furthermore, the Web is influencing in the development of other information media, as for example, newspapers, journals, books, libraries, etc. In this chapter we analyze its impact in the development of the University Digital Libraries (UDL). As in the Web, the growing of information is the main problem of the academic digital libraries, and similar tools could be applied in university digital libraries to facilitate the information access to the students and teachers. *Filtering systems* or *recommender systems* are tools whose objective is to evaluate and filter the great amount of information available on the Web to assist the users in their information access processes. Therefore, we present a model of fuzzy linguistic recommender system to help students and researchers to find research resources which could improve the services that render the UDL to their users.

1 Introduction

In last years the new concept of digital library is growing. *Digital libraries* are information collections that have associated services delivered to user communities using a variety of technologies. The information collections can be scientific, business or personal data, and can be represented as digital text, image, audio, video, or other media. This information can be digitalized paper or born digital material and the services offered on such information can be varied, and can be offered to individuals or user communities. Internet access has resulted in digital libraries that are increasingly used by diverse communities for diverse purposes, and in which sharing and collaboration have become important social elements. As digital libraries become commonplace, as their contents and services become more varied, people expect more sophisticated services from their digital libraries [4, 10, 11, 31].

The digital libraries are composed by human resources (staff) take over handle and enable users the access to the documents more interesting for them, taking into account their needs or interest areas. The library staff searches, evaluates, selects, catalogues, classifies, preserves and schedules the digital documents access [10, 11]. Some of the main digital libraries functions are the following:

- To evaluate and select digital materials to add in its repository.
- To preserve the security and conservation of the materials.
- To describe and index the new digital materials (catalogue and classify).
- To deliver users the material stored in the library.
- Other managerial tasks.

Digital libraries have been applied in a lot of contexts. We are going to center in an academic environment. *University digital libraries* provide information resources and services to students, faculty and staff in an environment that supports learning, teaching and research [5, 27].

The exponential increase of Web sites and documents is contributing to that Internet users not being able to find the information they seek in a simple and timely manner. Users are in need of tools to help them cope with the large amount of information available on the Web [25, 28]. Therefore, techniques for searching and mining the Web are becoming increasingly vital. Furthermore, the Web is influencing in the development of many organizations, as for example, banks, companies, universities, libraries, etc. In particular, we are interested in the development of academic digital libraries. As in the Web, the exponential growing of information is the main problem of these libraries because the library staff find troubles to perform the task of information delivery to the users. We could use those tools applied successfully in the Web context to solve the new problems appeared in UDL to facilitate the tasks of library staff and therefore, the information access to the students and teachers.

A traditional search function is normally an integral part of any digital library, but however users' frustrations are increased as their needs become more complex and as the volume of information managed by digital libraries increases. Digital libraries must move from being passive, with little adaptation to their users, to being more proactive in offering and tailoring information for individuals and communities, and in supporting community efforts to capture, structure and share knowledge [4, 11, 31]. So, the digital libraries should anticipate the users' needs and recommending about resources that could be interesting for them.

In this paper we study two techniques that applied together can contribute to achieve major advances in the activities of university digital libraries in order to improve their performance:

- *Information Filtering Tools:* An important tool to improve the information access on many environments concerns the way in which it is possible to filter the great amount of information available. Information filtering is a name used to describe a variety of processes involving the delivery of information to people who need it. Operating in textual domains, *filtering systems* or *recommender systems* evaluate and filter the great amount of information available in a specific scope to assist users in their information access processes [13, 32].
- *Fuzzy Linguistic Modeling (FLM):* The great variety of representations and evaluations of the information existing in Internet is the main obstacle to the information handling from what is very important the design of appropriate communication protocol. The problem becomes more noticeable when users take part in the process. This reveals the need of more flexible techniques to the information representation and evaluation. To solve this problem we propose the use of *FLM* [16, 17, 34] to represent and handle flexible information by means of linguistic labels.

The paper is structured as follows. Section 2 revises the main aspects and models of information filtering techniques. Section 3 analyzes different approaches of FLM, the 2-tuple FLM [17, 19] and the multi-granular FLM [15, 18]. In Section 4 we present a model of fuzzy linguistic recommender systems to advice research resources in UDL. Finally, some concluding remarks are pointed out.

2 Preliminaries

2.1 Information Filtering

Information gathering on the Internet is a complex activity. Finding the appropriate information, required for the users, on the World Wide Web is not a simple task. This problem is more acute with the ever increasing use of the Internet. For example, users who subscribe to Internet lists waste a great deal of time reading, viewing or deleting irrelevant e-mail messages. To improve the information access on the Web the users need tools to filter the great amount of information available across the Web. *Information Filtering* (IF) is a name used to describe a variety of processes involving the delivery of information to people who need it. It is a research area that offer tools for discriminating between relevant and irrelevant information by providing personalized assistance for continuous retrieval of information.

IF systems are characterized by [13]:

- applicable for unstructured or semi-structured data (e.g. web documents, e-mail messages),
- based on user profiles,
- handle large amounts of data,
- deal primarily with textual data and
- their objective is to remove irrelevant data from incoming streams of data items.

We can find some of the above features in Information Retrieval (IR) systems, but IF differs from traditional IR in that the users have long-term information needs that are described by means of user profiles, rather than ad-hoc needs that are expressed as queries posed to some IR system [2]. Traditionally IR develops storage, indexing and retrieval technology for textual documents. An user describes his information need in the form of a query to the IR system and the system attempts to find items that match the query within a document store. The information need is usually very dynamic and temporary, i.e., an user issues a query describing an immediate need. Furthermore, IR systems tend to maintain a relatively static store of information. Unlike IR systems, IF systems generally operate on continuous information streams, and always maintain a profile of the user interests needs throughout many uses of the system. As a result, IF systems tend to filter information based on more long-term interests.

Traditionally, these IF systems or recommender systems have fallen into two main categories [13, 29, 32]. *Content-based filtering systems* filter and recommend the information by matching user query terms with the index terms used in the representation of documents, ignoring data from other users. These recommender systems tend to fail when little is known about user information needs, e.g. when the query language is poor. *Collaborative filtering systems* use explicit or implicit preferences from many users to filter and recommend documents to a given user, ignoring the representation of documents. These recommender systems tend to fail when little is known about an user, or when he/she has uncommon interests [29]. In these kind of systems, the users' information preferences can be used to define user profiles that are applied as filters to streams of documents; the recommendations to an user are based on another user's recommendations with similar profiles. Many researchers think that the construction of accurate profiles is a key task and the system's success will depend to a large extent on the ability of the learned profiles to represent the user's preferences [30]. Several researchers are exploring hybrid content-based and collaborative recommender systems to smooth out the disadvantages of each one of them [3, 6, 12, 29].

On the other hand, we should point out that the *matching process* is a main process in the activity of filtering systems. The two major approaches followed in the design and implementation of IF systems to do the matching are the statistical approach and the knowledge based approach [13]. In our system, we have applied the statistical approach. This kind of filtering systems represents the documents and the user profiles as weighted vectors of index terms. To filter the information the system implements a statistical algorithm that computes the similarity of a vector of terms that represents the data item being filtered to an user's profile. The most common algorithm used is the Correlation or the Cosine measure between the user's profile and the document's vector.

The filtering activity is followed by a relevance feedback phase. Relevance feedback is a cyclic process whereby the user feeds back into the system decisions on the relevance of retrieved documents and the system then uses these evaluations to automatically update the user profiles.

Another important aspect that we must have in mind when we design a IF system is the method to gather user information. In order to discriminate between relevant and irrelevant information for an user and to provide him/her personalized information, we must have some information about this user, i.e. we must know the user preferences. Information about user preferences can be obtained in two different ways [13], *implicit* and *explicit mode*, although these ways not be mutually exclusive. The implicit approach is implemented by inference from some kind of observation. The observation is applied to user behavior or to detecting an user's environment (such as bookmarks or visited URL). The user preferences are updated by detecting changes while observing the user. On the other hand, the *explicit* approach, interacts with the users by acquiring feedback on information that is filtered, that is, the user expresses some specifications of what they desire. This last approach is very used. In [9] the personalization in digital libraries is studied. They conclude that the technology is still premature, but the next step of digital libraries services should be oriented towards the automation of the process of constructing of user profiles.

2.2 Fuzzy Linguistic Modeling

There are situations in which the information cannot be assessed precisely in a quantitative form but may be in a qualitative one. For example, when attempting to qualify phenomena related to human perception, we are often led to use words in natural language instead of numerical values. In other cases, precise quantitative information cannot be stated because either it is unavailable or the cost for its computation is too high and an "approximate value" can be applicable. The use of Fuzzy Sets Theory has given very good results for modeling qualitative information [34]. *FLM* is a tool based on the concept of *linguistic variable* [34] to deal with qualitative assessments. It has proven to be useful in many problems, e.g., in decision making [16], quality evaluation [24], models of information retrieval [20, 21], clinical decision making [8], political analysis [1], etc.

Next we analyze two FLM that we use in our system, i.e., the 2-tuple FLM [17, 19] and the multi-granular FLM [15, 18, 23].

The 2-Tuple Fuzzy Linguistic Modeling

The *2-tuple FLM* [17, 19] is a kind of fuzzy linguistic modeling that mainly allows to reduce the loss of information typical of other fuzzy linguistic approaches (classical and ordinal [16, 14, 34]). Its main advantage is that the linguistic computational model based on linguistic 2-tuples can carry out processes of computing with words easier and without loss of information. To define it we have to establish the 2-tuple representation model and the 2-tuple computational model to represent and aggregate the linguistic information, respectively.

Let $S = \{s_0, ..., s_g\}$ be a linguistic term set with odd cardinality ($g + 1$ is the cardinality of S), where the mid term represents an assessment of approximately 0.5 and with the rest of the terms being placed symmetrically around it. We assume that the semantics of labels is given by means of triangular membership functions represented by a 3-tuple (a, b, c) and consider all terms distributed on a scale on which a total order is defined $s_i \leq s_j \iff i \leq j$. In this fuzzy linguistic context, if a symbolic method [14, 16] aggregating linguistic information obtains a value $\beta \in [0, g]$, and $\beta \notin \{0, ..., g\}$, then an approximation function is used to express the result in S. To do this, we represent β as a 2-tuple (s_i, α_i), where:

- s_i represents the linguistic label, and
- α_i is a numerical value expressing the value of the translation from the original result β to the closest index label, i, in the linguistic term set ($s_i \in S$).

This model defines a set of transformation functions between numeric values and 2-tuples: $\Delta(\beta) = (s_i, \alpha)$ y $\Delta^{-1}(s_i, \alpha) = \beta \in [0, g]$ [17].

The 2-tuple linguistic computational model is defined by presenting the comparison of 2-tuples, a negation operator and aggregation operators of 2-tuples:

1. Negation operator of 2-tuples: $Neg((s_i, \alpha)) = \Delta(g - (\Delta^{-1}(s_i, \alpha)))$.
2. Comparison of 2-tuples (s_k, α_1) and (s_l, α_2):
   - If $k < l$ then (s_k, α_1) is smaller than (s_l, α_2).
   - If $k = l$ then
     a) if $\alpha_1 = \alpha_2$ then (s_k, α_1) and (s_l, α_2) represent the same information,
     b) if $\alpha_1 < \alpha_2$ then (s_k, α_1) is smaller than (s_l, α_2),
     c) if $\alpha_1 > \alpha_2$ then (s_k, α_1) is bigger than (s_l, α_2).
3. Aggregation operators of 2-tuples. The aggregation of information consists of obtaining a value that summarizes a set of values, therefore, the result of the aggregation of a set of 2-tuples must be a 2-tuple. In the literature we can find many aggregation operators which allow us to combine the information according to different criteria. Using functions Δ and Δ^{-1} that transform without loss of information numerical values into linguistic

2-tuples and viceversa, any of the existing aggregation operator can be easily extended for dealing with linguistic 2-tuples. Some examples are the arithmetic mean, the weighted average operator or the linguistic weighted average operator.

The Multi-Granular Fuzzy Linguistic Modeling

In any fuzzy linguistic approach, an important parameter to determinate is the "granularity of uncertainty", i.e., the cardinality of the linguistic term set S used to express the linguistic information. According to the uncertainty degree that an expert qualifying a phenomenon has on it, the linguistic term set chosen to provide his knowledge will have more or less terms. When different experts have different uncertainty degrees on the phenomenon, then several linguistic term sets with a different granularity of uncertainty are necessary (i.e. multi-granular linguistic information) [15, 18, 23]. The use of different label sets to assess information is also necessary when an expert has to assess different concepts, as for example it happens in information retrieval problems, to evaluate the importance of the query terms and the relevance of the retrieved documents [22]. In such situations, we need tools for the management of multi-granular linguistic information, i.e., we need to define a *multi-granular FLM*. In [15] we define a proposal of multi-granular FLM based on the ordinal FLM [16], and in [18] we define other one based on the 2-tuple FLM. In this paper, we follow that defined in [18] which uses the concept of *Linguistic Hierarchies* [7] to manage the multi-granular linguistic information.

A **linguistic hierarchy** is a set of levels, where each level is a linguistic term set with different granularity from the remaining of levels of the hierarchy [7]. Each level belonging to a linguistic hierarchy is denoted as *l(t,n(t))*, t being a number that indicates the level of the hierarchy and *n(t)* the granularity of the linguistic term set of the level *t*. Usually, linguistic hierarchies deal with linguistic terms whose membership functions are triangular-shaped, symmetrical and uniformly distributed in [0,1]. In addition, the linguistic term sets have an odd value of granularity representing the central label the value of *indifference* ("approximately 0.5"). The levels belonging to a linguistic hierarchy are ordered according to their granularity, i.e., for two consecutive levels *t* and *t+1*, $n(t+1) >$ *n(t)*. Therefore, each level $t+1$ provides a linguistic refinement of the previous level t.

Generically, we can say that the linguistic term set of level *t+1*, $S^{n(t+1)}$, is obtained from its predecessor level *t*, $S^{n(t)}$ as: $l(t, n(t)) \rightarrow l(t+1, 2 \cdot n(t) - 1)$. Table 1 shows the granularity needed in each linguistic term set of the level t depending on the value *n(t)* defined in the first level (3 and 7 respectively).

In [18] was demonstrated that the linguistic hierarchies are useful to represent the multi-granular linguistic information and allow to combine multi-granular linguistic information without loss of information. To do this, a family of transformation functions between labels from different levels was defined:

Table 1. Linguistic Hierarchies.

	Level 1	Level 2	Level 3
l(t,n(t))	l(1,3)	l(2,5)	l(3,9)
l(t,n(t))	l(1,7)	l(2,13)	

Definition 1. *Let* $LH = \bigcup_t l(t, n(t))$ *be a linguistic hierarchy whose linguistic term sets are denoted as* $S^{n(t)} = \{s_0^{n(t)}, ..., s_{n(t)-1}^{n(t)}\}$*. The transformation function between a 2-tuple that belongs to level t and another 2-tuple in level* $t' \neq t$ *is defined as:*

$$TF_{t'}^{t} : l(t, n(t)) \longrightarrow l(t', n(t'))$$

$$TF_{t'}^{t}(s_i^{n(t)}, \alpha^{n(t)}) = \Delta(\frac{\Delta^{-1}(s_i^{n(t)}, \alpha^{n(t)}) \cdot (n(t') - 1)}{n(t) - 1})$$

As it was pointed out in [18] this family of transformation functions is bijective. This result guarantees the transformations between levels of a linguistic hierarchy are carried out without loss of information.

To define the multi-granular linguistic computational model we select a level to uniform the information (normally the most granularity level is selected) and then we can use the operators defined in 2-tuples model.

3 A Recommender System based on Multi-granular Fuzzy Linguistic Modeling to Advice Research Resources in University Digital Libraries

In this section we present a Recommender System (RS) designed using the *content-based filtering approach* and assuming *a multi-granular FLM*. This RS is applied to advice users on the better research resources that could satisfy their information needs in an university digital library.

The users of an university digital library are usually the students and teachers that access to its information resources. Both manage and spread a lot of information about research information such as electronic books, electronic papers, electronic journals, official dailies and so on. Nowadays this amount of information is growing up and the users of these libraries are in need of automate tools to filter and to spread the information in a simple and timely manner.

We present a RS that follows the content-based approach. Moreover to improve the filtering process we incorporate in the system the possibility to manage multi-granular linguistic information, that is, it uses different label sets to represent the different concepts to be assessed in its recommending activity. Then, the system filters the incoming information stream and delivers it to the suitable researchers or students in accordance with their research areas. The system sends the users a mail with a summarized information about the resources, the calculated relevance degrees of the resources for the users and recommendations about others researchers or students with which they could collaborate.

In that follows, we present the system architecture, the required data structures and how the system works.

3.1 System architecture

The system architecture is shown in figure 1. As we can see in the figure, the system has three main components:

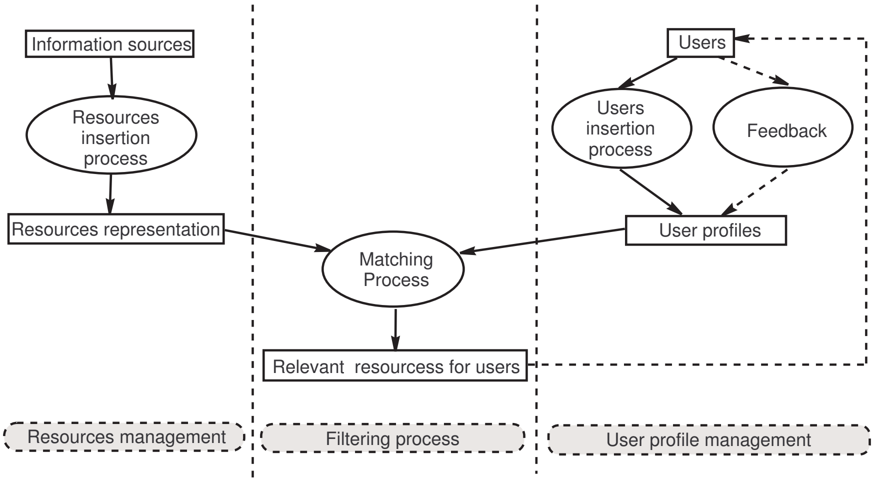

Fig. 1. Structure of the system

- *Resources management.* This module is the responsible of management the information sources from where the library staff receive all the information about research resources, and obtain an internal representation of these electronic resources. To management the resources, we represented them in accordance with their features (title, author(s), abstract, text, date, type and so on) and their scope, and to obtain this scope representation we use the *UNESCO terminology* for the science and technology [33]. This terminology is composed by three levels and each one is a refinement of the previous level. The first level includes general topics and they are codified by two digits. Each topic includes some disciplines codified by four digits in a second level. The third level is composed by subdisciplines that represent the activities developed in each discipline; these subdisciplines are codified by six digits. We are going to operate with the first and second levels, because we think the third level supply a discrimination level too much high and this could difficult the interaction with the users. Moreover, for each resource we store another kind of information that the system uses in the filtering process.

- *User profiles management.* The users can be researchers or students. In both cases, the system operates with an internal representation of the user's preferences or needs, that is, the system represents the users' preferences through user profiles. To define an user profile we are going to use the basic information about the user and his/her interest topics, defined too by the UNESCO terminology [33], i.e. each user has a list of UNESCO codes according to his/her information needs or interests. The research groups have assigned a set of UNESCO codes that define their research activity. So, initially the systems assign to each research or student the UNESCO codes of the research group which the user belongs. If the user doesn't belong to a group, the library staff assigns him/her the UNESCO codes by hand, in accordance with his/her interest areas. Afterwards the users can update their profiles by a feedback phase in which the users express some explicit specifications of their preferences.
- *Filtering process.* In this phase the system filters the incoming information to deliver it to the fitting users and this process is based in a Matching Process. As our system is a content-based filtering system, it filters the information by matching the terms used in the representation of user profiles against the index terms used in the representation of resources. Later we will study this process in detail taking into account the used data structures.

3.2 Data Structures

In this subsection we are going to discuss the data structures we need to represent all the information about the users and research resources. We must have in mind that the system stores this information because it doesn't work with explicit user queries.

To characterize a resource, we use the title, author(s), abstract, journal (if is part of a journal, the system stores the journal name), book (if is a book chapter, the system stores the book title), official daily (if is part of an official daily, the system stores the daily title), date, source, text, link (when the system send the users information about a resource, it doesn't send all the information but summarized information and the link to access the resource), kind of resource (if is a book, a paper, a journal, an official daily and so on), target (this field indicates the kind of users that is directed the resource, that is researchers, students or anybody) and scope. To represent the resource scope we use the *vector model* where for each resource the system stores a vector VR, i.e. an ordered list of terms. To build this vector we follow the

UNESCO terminology [33], specifically we use the second level. This level has 248 disciplines, so the vector must have 248 positions, one position for each discipline. In each position the vector stores the importance degree for the resource scope of the UNESCO code represented in that position.

To characterize an user we must distinguish if is a research or a student, although the system stores the same basic information: user's identity (usually his/her mail), password (necessary to access the system), dni (identity national document), name and surname, department and center (if the user is a students this information is not necessary), address, phone number, mobile phone and fax, web, email (elemental information to send the resources and recommendations), research group (is a string composed by 6 digits, 3 characters indicating the research area and 3 numbers identifying the group; if the user is a students this information is not necessary), collaboration preferences (if the user want collaborate with other researchers of other groups, with students, with anybody or with nobody), preferences about resources (the user choose the kind of desired resources, i.e. if he/she want only books, or papers, etc.) and interest topics. To represent the interest topics we use the vector model too where for each user the system stores a vector VU. To build this vector we follow the UNESCO terminology [33], specifically we use the second level. This level has 248 disciplines, so the vector must have 248 positions, one position for each discipline. In each position the vector stores the importance degree for the user research of the UNESCO code represented in that position. With all this information the system sets up the user profiles.

On the other hand, to represent the linguistic information we use different label sets, i.e. the communication among the users and the system is carried out by using multi-granular linguistic information, in order to allow a higher flexibility in the processes of communication of the system. Therefore the system uses different label sets ($S_1,\ S_2, ...$) to represent the different concepts to be assessed in its filtering activity. These label sets S_i are chosen from those label sets that composes a LH, i.e., $S_i \in LH$. We should point out that the number of different label sets that we can use is limited by the number of levels of LH, and therefore, in many cases the label sets S_i and S_j can be associated to a same label set of LH but with different interpretations depending on the concept to be modeled. In our system, we distinguish between three concepts that can be assessed:

- *Importance degree* (S_1) of an UNESCO code with respect to a resource scope or user preferences.
- *Relevance degree* (S_2) of a resource for a researcher or for a student.
- *Compatibility degree* (S_3) between a researcher and a student, between researchers of different groups and between different students.

In our system we use a linguistic hierarchy of three levels. Specifically we use the level 2 (5 labels) to assign importance degree ($S_1 = S^5$) and the level 3 (9 labels) to assign relevance degrees ($S_2 = S^9$) and compatibility degrees

($S_3 = S^9$). Using this linguistic hierarchy the linguistic terms in each level are:

- $S^3 = \{a_0 = Null = N,\ a_1 = Medium = M,\ a_2 = Total = T\}$.
- $S^5 = \{b_0 = Null = N,\ b_1 = Low = L,\ b_2 = Medium = M,\ b_3 = High = H,\ b_4 = Total = T\}$
- $S^9 = \{c_0 = Null = N,\ c_1 = Very\_Low = VL,\ c_2 = Low = L,\ c_3 = More\_Less\_Low = MLL,\ c_4 = Medium = M,\ c_5 = More\_Less\_High = MLH,\ c_6 = High = H,\ c_7 = Very\_High = VH,\ c_8 = Total = T\}$

Therefore, for a resource i we have a vector representing its scope:

$$VR_i = (VR_{i1}, VR_{i2}, ..., VR_{i248}),$$

where each component $VR_{ij} \in S_1$, with $j = 1..248$, stores a linguistic label indicating the importance degree of the UNESCO code j with regard to the resource i. These linguistic labels are assigned by the library staff when they add a new electronic resource.

To represent the interest topics in the user profiles we follow the same method, using a vector VU for each user of the system. Then, for the user x, we have a vector:

$$VU_x = (VU_{x1}, VU_{x2}, ..., VU_{x248}),$$

where each component $VU_{xy} \in S_1$, with $y = 1..248$, stores a linguistic label indicating the importance degree of the UNESCO code y with regard to the preferences of the user x. These linguistic labels are assigned by the library staff too, but the users can edit it when they want.

3.3 Operation of Recommender System

Any university digital library must provide the next two kind of services [10]:

- *User registration.* The users accesses to the system to solicit the services offered by the university digital library. The system present a form where the users introduce their personal information, their collaboration preferences and their preferences about the kind of resources they want to receive. Finally the users define their interest topics setting up the UNESCO codes and the importance degrees. If the user belongs to a research group, the system shows him/her the UNESCO codes of the group, and the user can edit (add, delete, or assign new degrees) these codes to adjust them to his/her interest areas. The system registers the user and assigns him/her an identifier (usually it uses the mail address) and a password. To conclude the registration process, the system sends the user an email to confirm the inserted information.

- *Information and documents access services.* Once the users have their identifies and passwords, they can use the digital library services. Therefore they can performs their information access processes taking into account their profiles.

Next we describe the users insertion process, the resources insertion process, the filtering process and the feedback phase.

Users Insertion Process

In order to gather information about users we use a hybrid approach between the explicit and implicit approach. When we insert a new user we use implicit information to generate the profile and afterwards the users can update their profiles following the explicit approach.

So, to add a new user into the system, it shows a form that the user must fill in introducing his/her personal information, collaboration preferences, preferences about the kind of resources he/she want to receive and so on. Then the system defines the user interest topics using the UNESCO codes of the research group which the user belongs. Each group or company has assigned one or more UNESCO codes, so when the system is inserting a new user, it assigns him/her the UNESCO codes of level 2 of the group which the user belong, with importance degree *Total ($b_4 \in S_1$).* The other positions have a value *Null ($b_0 \in S_1$).* The system presents this information to the users who can edit it if they want. The users who don't belong to a research group, must define their profiles manually, that is, they select the UNESCO codes and their importance degrees ($b_i \in S_1$) to establish their interest topics. Later the users can update their profiles always they want, accessing to the system and editing the UNESCO codes or the linguistic labels (in S_1) which they have assigned.

With this information the system defines and updates the user profiles which will use to filter information when a new resource arrives to the system.

Example 1. In this example we see the process of insertion of a new user. The user inserts all the information about him/her together with the user's identity $\mathcal{ID}$ and a password. Next, the system defines his/her interest topics. Let us suppose the user belong to a group which works in *Science of Nutriment*, because of this it has the UNESCO code *3206*; remember the group could have more UNESCO codes. Then, to define the vector of interest topics the system assigns the user this code (*3206*) with degree *Total ($b_4 \in S_1$).* With this information the user profile is represented by a vector of interest topics with the following values:

$$\begin{array}{l} VU_{\mathcal{ID}}[x] = b_4, \text{ if } x = 100 \\ VU_{\mathcal{ID}}[x] = b_0, \text{ otherwise.} \end{array}$$

Remark. The UNESCO code 3206 is in the position 100 of the list so it is stored in $VU_{\mathcal{ID}}[100]$.

Resources Insertion Process

This sub-process is carried out by the library staff that receive or find information about a resource and they want to spread this information. The experts introduce the interesting resources into the system and it automatically sends the information to the suitable users along with a relevance degree and collaborations possibilities.

As we said in the previous section, the system stores the general information about the resource and its scope. The scope is represented by a vector of UNESCO codes whereby to insert the resource the experts decide the UNESCO codes to assign it. Moreover, to manage the linguistic information, the experts also decide a linguistic label in S_1 to weight the importance degree of each UNESCO code of level 2 with regard to the resource.

Hence, when the library staff are going to insert a new resource, they access to the system, insert all the information about it, i.e. title, author(s), abstract, date, source, book name, journal name, daily name, link, text, kind of resource, target and finally they assess the importance degree of each UNESCO code of level 2 with regard to the resource. To do this, the system shows a list of UNESCO codes of level 2 and the library staff decide the codes to assign to the resource scope, selecting a code of the list and assign it a linguistic label to assess its importance degree. Then they accept and can either add another UNESCO code or finally the resource insertion.

Example 2. Now let us suppose the digital library receives a paper i about an Science of Nutriment Conference. Then, he/she inserts the paper into the system, introducing all the available information and selecting from a list the UNESCO codes which match with the resource scope. In this example, the library staff could select the codes *3206 - Science of Nutriment* with importance degree *Total ($b_4 \in S_1$)* and *3309 - Food Technology* with degree *Very High ($b_3 \in S_1$).* Once the expert inserts this information, we have a vector VR_i defining the resource i with the following values:

$$\begin{array}{l} VR_i[j] = b_4, \text{ if } j = 100 \\ VR_i[j] = b_3, \text{ if } j = 118 \\ VR_i[j] = b_0, \text{ otherwise.} \end{array}$$

Remark. The UNESCO codes 3206 and 3309 are in the positions 100 and 118 of the list so they are stored respectively in $VR_i[100]$ and $VR_i[118]$.

Filtering Process

As we have said, we are going to use the vector model [26] to represent the resources scope and the user interest topics. This vector model uses sophisticated similarity calculations to do the matching process, such as Euclidean Distance or Cosine Measure. Exactly we are going to use the Cosine Measure we described next.

The *Cosine Measure* is a similarity measure that is developed from the cosine of the angle between the vectors representing the scope resource (VR) and the user interest topics (VU), or between the vectors representing two users interest topics or between the vectors representing two scope resources. Its definition is [26]:

$$\sigma(VR, VU) = \frac{\sum_{k=1}^{n}(r_k \times u_k)}{\sqrt{\sum_{k=1}^{n}(r_k)^2} \times \sqrt{\sum_{k=1}^{n}(u_k)^2}}$$

where n is the number of terms used to define the vectors (i.e. the number of UNESCO codes of level 2), r_k is the value of term k in the resource vector and u_k is its value in the user vector. In mathematical terms this is the inner product of the resources and users vectors, normalized by their lengths. Using this cosine transforms the angular measure into a measure ranging from 1 for the highest similarity to 0 for the lowest. In the case of two users or two resources, this cosine measure is applied of the same way.

Angular measures representing a view of the resources and users items space from a fixed point, the origin. In addition, an angular measure does not consider the distance of each item from the origin, but only the direction. Hence two items that lie along the same vector from the origin will be judged identically, despite the fact that they may be far apart in the document space. This means that a one-paragraph announcement and an extensive, detailed paper about a topic might be judged to be equally relevant to a query. For example, suppose there are three notices, each described by the same two terms, with resource vectors:

$$VR_1 =< 1, 3 >,$$
$$VR_2 =< 100, 300 >, \ and$$
$$VR_3 =< 3, 1 > .$$

By the cosine measure, $\sigma(VR_1, VR_2) = 1.0$ and $\sigma(VR_1, VR_3) = 0.6$. The cosine measure views R_2 as more similar to R_1 than is R_3. It can be argued that in R_1 and R_2 the two terms have the same relative importance; that is, that the ratio of their values is the same.

Following this approach when a new resource has been inserted into the system, we compute the cosine measure $\sigma(VR_i, VU_j)$ between the new scope resource vector (VR_i) against all the user vectors (VU_j, $j = 1..m$ where m is the number of users of the system) to find the fit users to deliver this information. If $\sigma(VR_i, VU_j) \geq \alpha$, the system select the user j. Previously we have defined a threshold value (α) to filter out the information. In this iteration, the system takes into account too the user preferences (kind of resource) to consider or not the user. The collaboration preferences are used to classify the selected users in two sets, the selected users that don't want to collaborate $\mathcal{U}_{\mathcal{S}}$ and the selected users arranged to collaborate $\mathcal{U}_{\mathcal{C}}$.

After this, the system has two sets of selected users $\mathcal{U_S}$ and $\mathcal{U_C}$ and for each user it has a value $\sigma(VR_i, VU_j) \geq \alpha$. The system apply to each $\sigma(VR_i, VU_j)$ the transformation function defined in definition 1 to obtain the relevance degree of the resource i for the user j, expressed in the set S_2. Then, the system sends to the users of $\mathcal{U_S}$ the resource information and its calculated relevance degree by a linguistic label more effective than a number.

For the users in $\mathcal{U_C}$ the system performs other step; it calculates the collaboration possibilities between the selected users. To do it, between each two users $x, y \in \mathcal{U_C}$:

- to analyze if the users are researchers or students and take into account the users preferences about it. For example a researcher could want to collaborate only with others researches of different research group.
- to calculate the cosine measure between the users, $\sigma(VU_x, VU_y)$,
- to obtain the compatibility degree between x and y, expressing $\sigma(VU_x, VU_y)$ as a linguistic label in S_3 (using the transformation function defined in definition 1) to send it to the user.

Finally the system sends to the users of $\mathcal{U_C}$ the resource information, its calculated relevance degree and the collaboration possibilities along with a compatibility degree. All the process is shown in the figure 2.

Feedback Phase

This phase is related to the activity developed by the *filtering system* once users have taken some of the resources delivered by the system. As we said, user profiles represents the users' long-term information needs or interests and a desire property for user profiles is that they should be adaptable since users' needs could change continuously. Because of this, the system allows the users update their profiles to improve the filtering process with the needs of each one. In our system this feedback process is developed in the following steps:

- The users access the system entering their $\mathcal{ID}$ and password.
- The users can do the following operations:
  - to edit their collaboration preferences,
  - to edit their preferences about kind of desired resources,
  - to edit their interest topics:
    - to add new UNESCO codes with its importance degrees, i.e. linguistic labels $b_i \in S_1$.
    - to delete an existing UNESCO code.
    - to modify the importance degree (linguistic label $b_i \in S_1$) assigned to an existing UNESCO code.

Example 3. Assuming information given in 1, let us suppose the user $\mathcal{ID}$ wants to update his/her profile because $\mathcal{ID}$ thinks he/she should belong to the

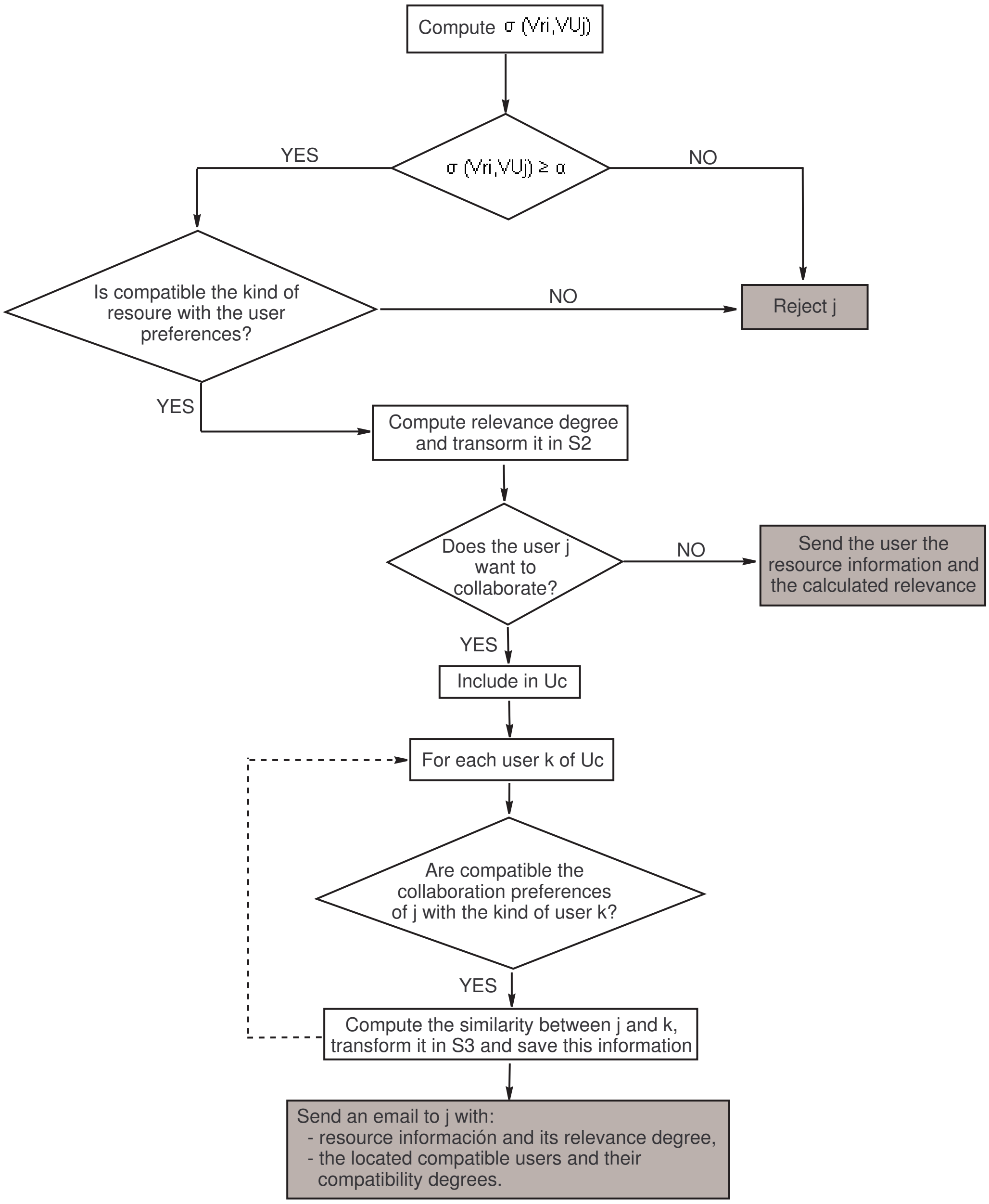
Compute σ (Vri,VUj)
σ (Vri,VUj) ≥ α
YES
NO
Is compatible the kind of resoure with the user preferences?
NO
Reject j
YES
Compute relevance degree and transorm it in S2
Does the user j want to collaborate?
NO
Send the user the resource information and the calculated relevance
YES
Include in Uc
For each user k of Uc
Are compatible the collaboration preferences of j with the kind of user k?
YES
Compute the similarity between j and k, transform it in S3 and save this information
Send an email to j with:
- resource información and its relevance degree,
- the located compatible users and their compatibility degrees.

Fig. 2. Matching process for an user j

category *3309 - Food Technology.* In this case the user wants to add a new UNESCO code and assigns it an importance degree of *High* ($b_3 \in S_1$).

After this, the user $\mathcal{ID}$ has a new profile represented by a new vector with the following values:

$$\begin{array}{l} VU_{\mathcal{ID}}[y] = b_4, \text{ if } y = 100 \\ VU_{\mathcal{ID}}[y] = b_3, \text{ if } y = 118 \\ VU_{\mathcal{ID}}[y] = b_0, \text{ otherwise.} \end{array}$$

4 Concluding Remarks

The exponential increase of Web sites and electronic documents is contributing to that Internet users not being able to find the information they seek in a simple and timely manner. The impact of the new digital technologies in others organizations is causing the apparition of problems similar to the Web ones, as for example it happens in UDL. Hence, users of UDL need tools to assist them in their processes of information gathering because of the large amount of information available on these systems. We have presented two techniques that could contribute to solve this problem, the information filtering tools and multi-granular FLM. Then, we have defined a model of fuzzy linguistic recommender system to spread research resources in UDL using both techniques. The proposed system is oriented both researchers and student and advice them research resources that could be interesting for them. In particular, it is a personalized system based on both content-based filtering tools and the multi-granular FLM. The system filters the incoming information stream to spread the information to the fitting users and recommends them about collaboration possibilities. The multi-granular FLM has been applied in order to improve the users-system interaction and the interpretability of the system activities. Moreover, the system brings a extra value, that is, on the one hand it sends the users a linguistic relevance degree to justify the information mailing and on the other hand it recommends the user the collaboration possibilities with other users. However we think the system could improve, incorporating some features, such as incorporate a module to define the resources scope automatically, or apply new techniques that have been used in development of the recommender systems.

References

1. Arfi B., Fuzzy decision making in politics. A linguistic fuzzy-set approach (LFSA). Political Analysis, 13 (1), 23-56, 2005.
2. Baeza-Yates R. and Ribeiro-Neto B., Modern Information Retrieval. Addison-Wesley, 1999.

3. Basu C., Hirsh H. and Cohen W., Recommendation as classification: Using social and content-based information in recommendation. In Proc. of the Fifteenth National Conference on Artificial Intelligence, 714-720, 1998.
4. Callan J., et. al., Personalisation and Recommender Systems in Digital Libraries. Joint NSF-EU DELOS Working Group Report. May 2003.
5. Chao H., Assessing the quality of academic libraries on the Web: The development and testing of criteria. Library & Information Science Research, 24, 169-194, 2002.
6. Claypool M., Gokhale A. and Miranda T., Combining content-based and collaborative filters in an online newpaper. In Proc. of the ACM SIGIR-99 Workshop on Recommender Systems-Implementation and Evaluation, 1999.
7. Cordón O., Herrera F. and Zwir I., Linguistic modelling by hierarchical systems of linguistic rules. IEEE Transactions on Fuzzy Systems, 10 (1), 2-20, 2001.
8. Degani R. and Bortolan G., The Problem of Linguistic Approximation in Clinical Decision Making. Int. J. of Approximate Reasoning, 2, 143-162, 1988.
9. Frias-Martinez E., Magoulas G., Chen S. and Macredie R., Automated user modeling for personalized digital libraries. International Journal of Information Management 26, 234-248, 2006.
10. Garcia E. and Garcia L.A., La Biblioteca Digital. Arco Libros S.L., 2001.
11. Gonçalves M. A., Fox E. A., Watson L. T. and Kipp N. A., Streams, structures, spaces, scenarios, societies (5s): A formal model for digital libraries. ACM Trans. Inf. Syst. 22, 2, 270-312, Apr. 2004.
12. Good N., Schafer J.B., Konstan J.A., Borchers A., Sarwar B.M., Herlocker J.L. and J. Riedl., Combining collaborative filtering with personal agents for better recommendations. In Proc. of the Sixteenth National Conference on Artificial Intelligence, 439-446, 1999.
13. Hanani U., Shapira B. and Shoval P., Information Filtering: Overview of Issues, Research and Systems. User Modeling and User-Adapted Interaction, 11, 203-259, 2001.
14. Herrera F. and Herrera-Viedma E., Aggregation operators for linguistic weighted information. IEEE Trans. on Systems, Man and Cybernetics, Part A: Systems, 27, 646-656, 1997.
15. Herrera F., Herrera-Viedma E. and Martínez L., A Fusion Approach for Managing Multi-Granularity Linguistic Term Sets in Decision Making. Fuzzy Sets and Systems, 114, 43-58, 2000.
16. Herrera F., Herrera-Viedma E. and Verdegay J.L., Direct approach processes in group decision making using linguistic OWA operators. Fuzzy Sets and Systems, 79, 175-190, 1996.
17. Herrera F. and Martínez L., A 2-tuple fuzzy linguistic representation model for computing with words. IEEE Transactions on Fuzzy Systems, 8 (6), 746-752, 2000.
18. Herrera F. and Martínez L., A model based on linguistic 2-tuples for dealing with multigranularity hierarchical linguistic contexts in multiexpert decision-making. IEEE Transactions on Systems, Man and Cybernetics. Part B: Cybernetics, 31(2), 227-234, 2001.
19. Herrera F. and Martínez L., The 2-tuple linguistic computational model. Advantages of its linguistic description, accuracy and consistency. Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems, 9, 33-48, 2001.

20. Herrera-Viedma E., Modeling the retrieval process of an information retrieval system using an ordinal fuzzy linguistic approach. J. of the American Society for Information Science and Technology, 52(6), 460-475, 2001.
21. Herrera-Viedma E., An information retrieval system with ordinal linguistic weighted queries based on two weighting elements. Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems, 9, 77-88, 2001.
22. Herrera-Viedma E., Cordón O., Luque M., López A.G. and Muñoz A.M., A Model of Fuzzy Linguistic IRS Based on Multi-Granular Linguistic Information. International Journal of Approximate Reasoning, 34 (3), 221-239, 2003.
23. Herrera-Viedma E., Martínez L., Mata F. and Chiclana F., A Consensus Support System Model for Group Decision-making Problems with Multi-granular Linguistic Preference Relations. IEEE Trans. on Fuzzy Systems, 2005. To appear.
24. Herrera-Viedma E. and Peis E., Evaluating the informative quality of documents in SGML-format using fuzzy linguistic techniques based on computing with words. Information Processing & Management, 39(2), 195-213, 2003.
25. Kobayashi M. and Takeda K., Information retrieval on the web. ACM Computing Surveys, 32(2), 148-173, 2000.
26. Korfhage R.R., Information Storage and Retrieval. New York: Wiley Computer Publishing, 1997.
27. Lau Noriega J., Bibliotecas universitarias: Su importancia en el proceso de acreditacin. Gaceta Universitaria, 149, p. 1, 2002.
28. Lawrence S. and Giles C., Searching the web: General and scientific information access. IEEE Comm. Magazine, 37 (1), 116-122, 1998.
29. Popescul A., Ungar L.H., Pennock D.M. and Lawrence S., Probabilistic models for unified collaborative and content-based recommendation in sparce-data environments. In Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence (UAI), San Francisco, 437-444, 2001.
30. Quiroga L.M. and Mostafa J., An experiment in building profiles in information filtering: the role of context of user relevance feedback. Information Processing and Management, 38, 671-694, 2002.
31. Renda M.E. and Straccia U., A personalized collaborative Digital Library environment: a model and an application. Information Processing and Management, 41, 5-21, 2005.
32. Reisnick P. and Varian H.R., Recommender Systems. Special issue of Comm. of the ACM, 40 (3), 56-59, 1997.
33. Clasificacin UNESCO. Ministerio de Educación y Ciencia. http://www.mec.es/ciencia/jsp/plantilla.jsp?area=proyectos/invest&id=53
34. Zadeh L.A., The concept of a linguistic variable and its applications to approximate reasoning. Part I. Information Sciences, 8, 199-249, 1975. Part II, Information Sciences, 8, 301-357, 1975. Part III, Information Sciences, 9, 43-80, 1975.

5. Web Intelligence

6. Computer Vision