Expert Systems with Applications 36 (2009) 5173–5183

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

A recommender system for research resources based on fuzzy linguistic modeling

C. Porcel [a], A.G. López-Herrera [a], E. Herrera-Viedma [b,*]

[a] Department of Computer Science. University of Jaén, 23071 Jaén, Spain
[b] Department of Computer Science and Artificial Intelligence University of Granada, 18071, Granada, Spain

ARTICLE INFO

Keywords:
Recommender systems
Information filtering
Fuzzy linguistic modeling
Multi-granular linguistic information

ABSTRACT

Nowadays, the increasing popularity of Internet has led to an abundant amount of information created and delivered over electronic media. It causes the information access by the users is a complex activity and they need tools to assist them to obtain the required information. *Recommender systems* are tools whose objective is to evaluate and filter the great amount of information available in a specific scope to assist the users in their information access processes. Another obstacle is the great variety of representations of information, specially when the users take part in the process, so we need more flexibility in the information processing. The *fuzzy linguistic modeling* allows to represent and handle flexible information. Similar problems are appearing in other frameworks, such as digital academic libraries, research offices, business contacts, etc. We focus on information access processes in technology transfer offices. The aim of this paper is to develop a recommender system for research resources based on fuzzy linguistic modeling. The system helps researchers and environment companies allowing them to obtain automatically information about research resources (calls or projects) in their interest areas. It is designed using some filtering tools and a particular fuzzy linguistic modeling, called multi-granular fuzzy linguistic modeling, which is useful when we have to assess different qualitative concepts. The system is working in the University of Granada and experimental results show that it is feasible and effective.

1. Introduction

A Technology Transfer Office (TTO) is responsible for putting into action and managing the activities which generate knowledge and technical and scientific collaboration, thus enhancing the interrelation between researchers at the University and the entrepreneurial world and their participation in various support programmes designed to carry out research, development and innovation activities. The main mission in this office is to encourage and help, from the University, the generation of knowledge and its spread and transfer to the society, with the aim of rapidly meeting society's needs and demands. A graphical representation of this mission is shown in Fig. 1 (The Centre for Innovation, XXXX).

To carry out its objectives, a TTO runs a number of services which we highlight the followings (The Centre for Innovation, XXXX):

- Information (R&D bulletins, R&D&I, calls, notices, projects).
- Guidance for Research and Development (R&D) and Technology Transfer funding.
- Advice in the preparation of offers (management, spread and exploitation).
- Support in the elaboration and negotiation of contracts with companies.
- Management of contacts.
- Technological offer (the elaboration of the offer, spread and promotion).
- The advice in the creation of new businesses.
- Evaluation, protection and transfer of ownership rights both intellectual and industrial.

To fulfil these objectives and manage all the services, a TTO is composed by a team of technicians that are experts in technology transfer. Each one manages a specific task, but all of them must provide information about research resources to the researchers and companies, that is bulletins, projects, calls, notices, events, congresses, courses, and so on. This task requires the selection by the expert of suitable researchers to deliver the information. In this task, we find a first problem, the large increase of research resources is contributing to that TTO experts not being able to spread the information to the suitable users (both researchers and companies) in a simple and timely manner. Then TTO experts are in need of tools to help them cope with the large amount of information available about research resources. A promising direction to improve the information access about research resources concerns

* Corresponding author.
E-mail addresses: cporcel@ujaen.es (C. Porcel), viedma@decsai.ugr.es (E. Herrera-Viedma).

0957-4174/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2008.06.038

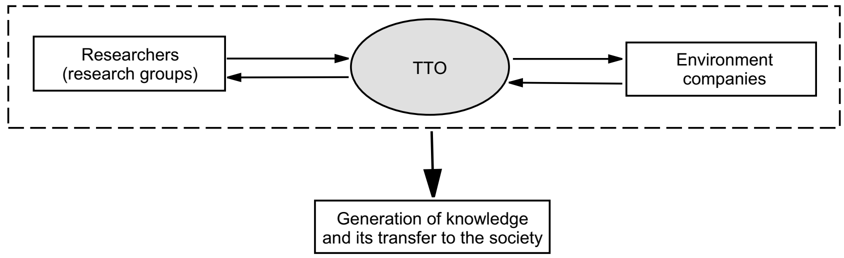

Fig. 1. Main mission in a TTO.

the way in which it is possible to filter the great amount of information available. *Recommender Systems* are tools whose objective is to evaluate and filter the great amount of information available in a specific scope to assist the users in their information access processes (Basu, Hirsh, & Cohen, 1998; Cao & Li, 2007; Hanani, Shapira, & Shoval, 2001; Hsu, 2008; Ungar, Pennock, & Lawrence, 2001; Reisnick & Varian, 1997).

Another problem is the great variety of representations and evaluations of the information. The problem becomes more noticeable when users take part in the process. Therefore, to improve the information representations and the user interface we need more flexibility in the information processing. To solve this problem we propose the use of *Fuzzy Linguistic Modeling* (FLM) (Ben-Arieh & Zhifeng, 2006; Herrera & Herrera-Viedma, 1997; Herrera, Herrera-Viedma, & Martínez, 2008; Herrera, Herrera-Viedma, & Verdegay, 1996; Herrera & Martínez, 2000; Zadeh, 1975) to represent and handle flexible information by means of linguistic labels.

In this paper, we propose SIRE2IN, a recommender system for recommending research resources based on FLM. The system allows the researchers to obtain automatically information about research resources in their interest areas and it recommends about companies or another researchers which could collaborate with them in projects (Chang, Wang, & Wang, 2007; Chen & Ben-Arieh, 2006; Herrera & Martínez, 2001; Herrera-Viedma, Cordón, Luque, López, & Muñoz, 2003; Herrera-Viedma, Martínez, Mata, & Chiclana, 2005). SIRE2IN is designed using both recommendation techniques and the multi-granular FLM to represent and handle flexible information by means of linguistic labels. To prove the system functionality we have implemented a primary version and the experimental results shows its useful and effectiveness.

The paper is structured as follows: Section 2 revises the recommendation approaches and the FLM. Section 3 presents the design of the system, analyzing its architecture, data structure and activity. Section 4 reports the system evaluation and the experimental results. Finally, we point out some concluding remarks.

2. Preliminaries

2.1. Recommender systems

Information gathering in Internet is a complex activity. Find the appropriate information, required for the users, on the 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. Recommender systems can provide information services by delivering the information to people who need it. It is a research area that offers tools for discriminating between relevant and irrelevant information by providing personalized assistance for continuous retrieval of information (Reisnick & Varian, 1997).

The recommender systems can be characterized because they (Hanani et al., 2001; Reisnick & Varian, 1997):

- are applicable for unstructured or semi-structured data (e.g. Web documents or e-mail messages),
- the users have long time information needs that are described by means of 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.

Traditionally, recommender systems have fallen into two main categories (Good et al., 1999; Hanani et al., 2001; Popescul et al., 2001; Reisnick & Varian, 1997). *Content-based recommender systems* recommend the information by matching the terms used in the representation of user profiles 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. *Collaborative recommender systems* use explicit or implicit preferences from many users to recommend documents to a given user, ignoring the representation of documents. These recommender systems tend to fail when little is known about a user, or when he/she has uncommon interests (Popescul et al., 2001). 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 a user are based on another users' recommendations with similar profiles. The construction of accurate profiles is a key task and the system's success will depend on a large extent on the ability of the learned profiles to represent the user's preferences (Quiroga & Mostafa, 2002). Moreover, we can use a hybrid approach to smooth out the disadvantages of each one of them and to exploit their benefits (Basu et al., 1998; Claypool, Gokhale, & Miranda, 1999; Good et al., 1999; Popescul et al., 2001).

On the other hand, we should point out that the *matching process* is a main process in the activity of the recommender systems. The two major approaches followed in the design and implementation of recommender systems to do the matching are the statistical approach and the knowledge based approach (Hanani et al., 2001). In our system, we have applied the statistical approach. This approach 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 a 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 (Korfhage, 1997).

The recommendation 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 profile (Hanani et al., 2001; Popescul et al., 2001; Reisnick & Varian, 1997).

Another important aspect that we must have in mind when we design a recommender system is the method to gather user information. In order to discriminate between relevant and irrelevant information for a user, 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 (Hanani et al., 2001; Quiroga & Mostafa, 2002), *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 a 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 (Hanani et al., 2001; Popescul et al., 2001; Reisnick & Varian, 1997).

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 (Zadeh, 1975) and it has proven to be useful in many problems, e.g., in decision making (Herrera, Herrera-Viedma, & Verdegay, 1996; Herrera et al., 1996; Herrera, Herrera-Viedma, & Verdegay, 1998; Xu, 2006), quality evaluation (Herrera-Viedma, Pasi, López-Herrera, & Porcel, 2006; Herrera-Viedma & Peis, 2003; Herrera-Viedma, Peis, Morales del Castillo, Alonso, & Anaya, 2007), information retrieval (Herrera-Viedma, 2001; Herrera-Viedma, 2001; Herrera-Viedma & López-Herrera, 2007; Herrera-Viedma, López-Herrera, Luque, & Porcel, 2007; Herrera-Viedma, López-Herrera, & Porcel, 2005), political analysis (Arfi, 2005), etc. It is a tool based on the concept of *linguistic variable* proposed by Zadeh (1975). Next we analyze the two approaches of FLM that we use in our system.

2.2.1. The 2-tuple fuzzy linguistic approach

The 2-tuple FLM (Herrera & Martínez, 2000) is a continuous model of representation of information that allows to reduce the loss of information typical of other fuzzy linguistic approaches (classical and ordinal Herrera & Herrera-Viedma, 1997; Zadeh, 1975). 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 $S = \{s_0, \ldots, s_g\}$ be a linguistic term set with odd cardinality, where the mid term represents a indifference value and the rest of the terms are symmetric relate to it. We assume that the semantics of labels is given by means of triangular membership functions and consider all terms distributed on a scale on which a total order is defined, $s_i \leqslant s_j \iff i \leqslant j$. In this fuzzy linguistic context, if a symbolic method (Herrera & Herrera-Viedma, 1997; Herrera et al., 1996) aggregating linguistic information obtains a value $\beta \in [0, g]$, and $\beta \notin \{0, \ldots, g\}$, then an approximation function is used to express the result in S.

Definition 1. *(Herrera & Martínez, 2000).* Let β be the result of an aggregation of the indexes of a set of labels assessed in a linguistic term set S, i.e., the result of a symbolic aggregation operation, $\beta \in [0, g]$. Let $i = round(\beta)$ and $\alpha = \beta - i$ be two values, such that, $i \in [0, g]$ and $\alpha \in [-.5, .5)$ then α is called a *Symbolic Translation*.

The 2-tuple fuzzy linguistic approach is developed from the concept of symbolic translation by representing the linguistic information by means of 2-tuples (s_i, α_i), $s_i \in S$ and $\alpha_i \in [-.5, .5)$:

- s_i represents the linguistic label of the information, 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.

Definition 2. *(Herrera & Martínez, 2000).* Let $S = \{s_0, \ldots, s_g\}$ be a linguistic term set and $\beta \in [0, g]$ a value representing the result of a symbolic aggregation operation, then the 2-tuple that expresses the equivalent information to β is obtained with the following function:

$$\Delta : [0, g] \to S \times [-0.5, 0.5),$$

$$\Delta(\beta) = (s_i, \alpha), \quad \text{with} \begin{cases} s_i & i = \text{round}(\beta), \\ \alpha = \beta - i & \alpha \in [-.5, .5), \end{cases}$$

where round$(\cdot)$ is the usual round operation, s_i has the closest index label to "β" and "α" is the value of the symbolic translation.

For all Δ there exists Δ^{-1}, defined as $\Delta^{-1}(s_i, \alpha) = i + \alpha$. On the other hand, it is obvious that the conversion of a linguistic term into a linguistic 2-tuple consists of adding a symbolic translation value of 0: $s_i \in S \Rightarrow (s_i, 0)$.

The computational model is defined by presenting the following operators:

1. Negation operator: $Neg((s_i, \alpha)) = \Delta(g - (\Delta^{-1}(s_i, \alpha)))$.
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. 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

Definition 3. *Arithmetic mean: Let $x = \{(r_1, \alpha_1), \ldots, (r_n, \alpha_n)\}$ be a set of linguistic 2-tuples, the 2-tuple arithmetic mean $\bar{x}^e$ is computed as*

$$\bar{x}^e[(r_1, \alpha_1), \ldots, (r_n, \alpha_n)] = \Delta\left(\sum_{i=1}^{n} \frac{1}{n} \Delta^{-1}(r_i, \alpha_i)\right) = \Delta\left(\frac{1}{n} \sum_{i=1}^{n} \beta_i\right).$$

Definition 4. *Weighted average operator: Let $x = \{(r_1, \alpha_1), \ldots, (r_n, \alpha_n)\}$ be a set of linguistic 2-tuples and $W = \{w_1, \ldots, w_n\}$ be their associated weights. The 2-tuple weighted average $\bar{x}^w$ is*

$$\bar{x}^w[(r_1, \alpha_1), \ldots, (r_n, \alpha_n)] = \Delta\left(\frac{\sum_{i=1}^{n} \Delta^{-1}(r_i, \alpha_i) \cdot w_i}{\sum_{i=1}^{n} w_i}\right) = \Delta\left(\frac{\sum_{i=1}^{n} \beta_i \cdot w_i}{\sum_{i=1}^{n} w_i}\right).$$

Definition 5. *Linguistic weighted average operator: Let* $x = \{(r_1, \alpha_1), (\ldots, (r_n, \alpha_n)\}$ *be a set of linguistic 2-tuples and* $W = \{(w_1, \alpha_1^w), \ldots, (w_n, \alpha_n^w)\}$ *be their linguistic 2-tuple associated weights. The 2-tuple linguistic weighted average* $\bar{x}_l^w$ *is*

$$\bar{x}_l^w[((r_1, \alpha_1), (w_1, \alpha_1^w)) \ldots ((r_n, \alpha_n), (w_n, \alpha_n^w))] = \Delta\left(\frac{\sum_{i=1}^{n} \beta_i \cdot \beta_{W_i}}{\sum_{i=1}^{n} \beta_{W_i}}\right),$$

with $\beta_i = \Delta^{-1}(r_i, \alpha_i)$ *and* $\beta_{W_i} = \Delta^{-1}(w_i, \alpha_i^w)$.

2.2.2. 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. 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 (Herrera & Martínez, 2001; Herrera-Viedma et al., 2005). The use of different labels 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 (Herrera-Viedma et al., 2003). In such situations, we need tools for the management of multi-granular linguistic information. In Herrera & Martínez (2001) is proposed a multi-granular 2-tuple FLM based on the concept of linguistic hierarchy (Cordón, Herrera, & Zwir, 2001).

A *Linguistic Hierarchy, LH,* is a set of levels $l(t, n(t))$, i.e., $LH = \bigcup_t l(t, n(t))$, where each level t is a linguistic term set with different granularity $n(t)$ from the remaining of levels of the hierarchy (Cordón et al., 2001). The levels are ordered according to their granularity, i.e., a level $t+1$ provides a linguistic refinement of the previous level t. We can define a level from its predecessor level as: $l(t, n(t)) \rightarrow l(t+1, 2, \ldots, n(t)-1)$. 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). A graphical example of a linguistic hierarchy is shown in Fig. 2.

In Herrera & Martínez (2001) 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:

Definition 6. Let $LH = \bigcup_t l(t, n(t))$ be a linguistic hierarchy whose linguistic term sets are denoted as $S^{n(t)} = \{s_0^{n(t)}, \ldots, s_{n(t)-1}^{n(t)}\}$. The transformation function between a 2-tuple that belongs to level t and another 2-tuple in level $t' \neq t$ is defined as:

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

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

As it was pointed out in Herrera & Martínez (2001) 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 computational model,

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)	

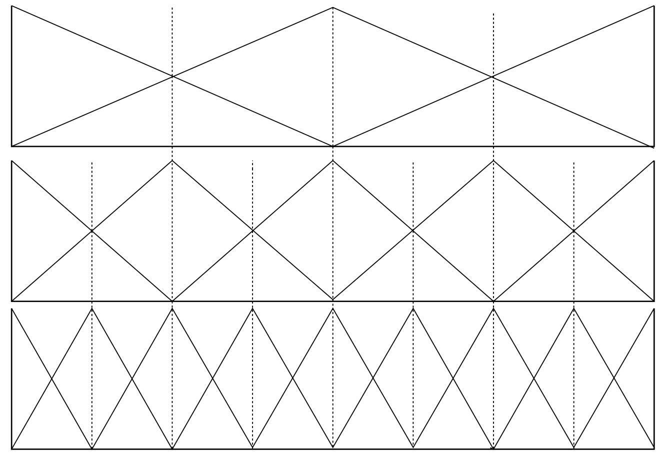

Fig. 2. Linguistic Hierarchy of 3, 5 and 9 labels.

we select a level to make uniform the information (for instance, the great granularity level) and then we can use the operators defined in the 2-tuple FLM.

3. SIRE2IN, a Recommender system for research resources

In this section, we present SIRE2IN, a recommender system based on multi-granular FLM.

As we said in the introduction, the TTO technicians manage and spread a lot of information about research information such as calls or projects. Nowadays, this amount of information is growing up and the experts are in need of automatic tools to filter and spread the information in a simple and timely manner. Because of this, our system incorporates in its activity a filtering process that follows the content-based approach. Moreover, to improve the representation of the information in the system we use multi-granular linguistic information, that is, different label sets to represent the different concepts to be assessed for different users in the filtering activity.

Then, SIRE2IN filters the incoming information stream and generates useful recommendations to the suitable researchers in accordance with their research areas. For each user the system generates an email with a summary about the resources, its relevance degrees and recommendations about collaboration possibilities.

3.1. Architecture of SIRE2IN

The architecture of SIRE2IN (Fig. 3) has three main components:

- *Resources management.* This module is the responsible one of management the information sources from which the TTO experts receive all the information about research resources. It obtains an internal representation of these items. Examples of information sources are Internet, news bulletins, distribution lists, forums, etc. To manage the items, we represent them in accordance with its scope using the *UNESCO terminology* for the science and technology (The UNESCO terminology, XXXX). 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 sup-

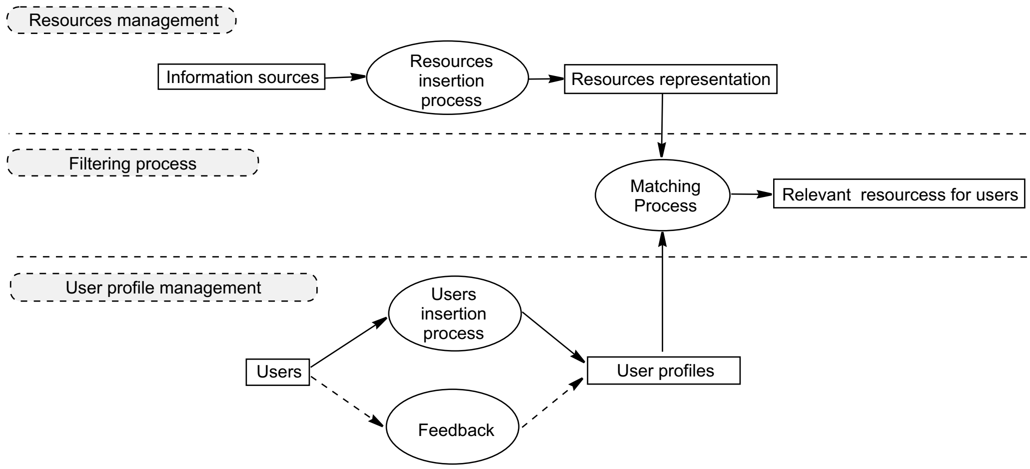

Fig. 3. Structure of SIRE2IN.

ply 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 of the University or employees of the environment companies. In both cases, the system operates with an internal representation of the user's preferences or needs, that is, the system represents each user through an user profile. To define a user profile we use the basic information about the user and his/her topics of interest, represented also by the UNESCO terminology (The UNESCO terminology, XXXX), i.e. each user have a list of UNESCO codes according to his/her information needs or interests. Both research groups and companies have assigned a set of UNESCO codes that define their research activity. So, initially the system assigns to each user the UNESCO codes of his/her research group or company and afterwards, users can update their profiles in a feedback phase in which the users can express some explicit specifications of their preferences.
- *Filtering process.* This component filters the incoming information to deliver it to the fitting users. The filtering process is based on a matching process. As our system is a content-based recommender 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 data structures.

3.2. Data structures

In this subsection, we are going to discuss the data structures that 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 does not work with explicit user queries.

To characterize a research resource, we use the following information:

- titular,
- abstract,
- text,
- date,
- source,
- link: when the system sends the users information about a resource, it does not send all the information but summarized information and the link to access the resource,
- target: this field indicates the kind of users which is oriented the resource, that is researchers, companies or anybody,
- minimum and maximum amount: it indicates the minimum and maximum amount that the user can solicit,
- scope: the system manages the resources in accordance with their scope. To represent the resource scope we use the *vector model* where for each resource the system stores a vector VR, i.e., a ordered list of terms. To build this vector we follow the UNESCO terminology (The UNESCO terminology, XXXX), 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 a 2-tuple linguistic value which represents the importance degree for the resource scope of the UNESCO code represented in that position.

To set up a user profile we use the following 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 University researcher or the company if the user is a company employee,
- address,
- phone number, mobile phone and fax,
- email: elemental information to send the resources and recommendations,
- research group: only if the user belongs to a research group. We use a code which is a string composed by six digits, three characters indicating the research area and three numbers identifying the group,
- collaboration preferences: if the user want to collaborate with other researchers of a distinct group, with companies, with anybody or with nobody,
- minimum and maximum amount: the users define the interval in which they have interested in solicit a call,
- topics of interest: to represent the topics of interest 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 (The UNESCO terminology, XXXX), 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 a 2-tuple linguistic value which represents the importance degree for the user's topic of interest of the UNESCO code represented in that position.

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 $(S_1, S_2, \ldots)$ to represent the different concepts to be assessed in its filtering activity. These label sets S_i are chosen from those label sets that compose 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 a UNESCO code with respect to a resource scope or user preferences,*
- *relevance degree* (S_2) *of a resource for a researcher or for a company,*
- *compatibility degree* (S_3) *between a researcher and a company, between researchers of different groups and between different companies.*

Following the linguistic hierarchy shown in Fig. 2, in our system 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 LH the linguistic terms in each level are

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

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

$$VR_i = (VR_i[1], VR_i[2], \ldots, VR_i[248]),$$

where each component $VR_i[j] \in S_1$, with $j = 1, \ldots, 248$, stores a linguistic 2-tuple indicating the importance degree of the UNESCO code j with regard to the resource i. These 2-tuples are assigned by the TTO technicians.

To represent the topics of interest 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:

$$VU_x = (VU_x[1], VU_x[2], \ldots, VU_x[248]),$$

where each component $VU_x[y] \in S_1$, with $y = 1 \ldots 248$, stores a linguistic 2-tuple indicating the importance degree of the UNESCO code y with regard to the preferences of the user x. These 2-tuples are also assigned by the experts, but the users can edit them when they want.

3.3. Activity of SIRE2IN

The system activity can be described briefly in three steps:

- An expert receives or finds information about a research resource and inserts it into the system.
- Then, the system runs the matching process to determinate the fitting users to receive the information and send them an email with the information, the calculated relevance degree and recommendations about possible collaborations with other users.
- Once the users have received the information, they can change the kind of information they want to receive in the future, by updating their user profiles.

In Fig. 4 we can see the main page of the system once the user is logged. Users can access the different options depending on the permissions they have assigned (user, expert or administrator).

3.3.1. Users insertion process

This process consists in to incorporate users' data into the system. It presents a form where the users insert their personal information, collaboration preferences and preferences about the resources. Users are invited to define their topics of interest and choose importance degrees (assessed in S_1) associated with them.

In order to gather information about users we use a hybrid approach, that is, 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. Initially a user has associated the topics of interest of his/her research group or company, but he/she can modify them. Each group or company has assigned one or more UNESCO codes, so the system assigns him/her the UNESCO codes of level 2 of his/her group or company with importance degree *Total* $((b_4, 0), with b_4 \in S_1)$. The other positions have a value *Null* $((b_0, 0), with b_0 \in S_1)$. Later, the users can update their profiles always they want, accessing to the system and editing the UNESCO codes or the importance degrees which they have assigned.

The system registers users and assigns them an identifier (email) and a password. Finally, the users receive a confirmation email with the inserted information.

Example 1. In this example, we are going to insert a new user. The user fills the form with his/her information. Let us suppose the user belongs to a group which works in Science of Nutriment, then he/she has assigned the UNESCO code 3206. Then, to define the vector of topics of interest the system assigns the user this code (3206) with degree *Total* $((b_4, 0), with b_4 \in S_1)$. With this information the user profile is represented by a vector of topics of interest with the following values:

Fig. 4. Main page of SIRE2IN.

$VU_{\mathscr{I}\mathscr{D}}[x] = (b_4, 0), \quad \text{if } x = 100$
$VU_{\mathscr{I}\mathscr{D}}[x] = (b_0, 0), \quad \text{otherwise}.$

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

The Fig. 5 shows an example of a user access.

3.3.2. Resources insertion process

This sub-process is carried out by the experts, i.e., the transfer technology technicians that receive or find information about a resource and they want to spread this information. The experts insert 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 2-tuple (b_i, α_i), with $b_i \in S_1$, to weight the importance degree of each UNESCO code of level 2 with regard to the resource scope.

Hence, when the experts are going to insert a new resource, they access to the system, insert all the information about it 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 experts decide the codes to assign to the resource scope, selecting a code of the list and assigning 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 expert receives a call i about a nutriment science research resource. Then, he/she inserts the call into the system, introducing all the available information and selecting from a list the UNESCO codes which match with the call. In this example, the expert could select the codes *3206 – Science of Nutriment* with importance degree *Total* $((b_4, 0), with b_4 \in S_1)$ and *3309 – Food Technology* with degree *Very High* $((b_3, 0), with b_3 \in S_1)$. Once the expert inserts this information, we have a vector VR_i defining the resource i with the following values:

$VR_i[j] = (b_4, 0), \quad \text{if } j = 100$
$VR_i[j] = (b_3, 0), \quad \text{if } j = 118$
$VR_i[j] = (b_0, 0), \text{otherwise}.$

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 $VR_i[118]$.

An example of resource list is shown in Fig. 6.

3.3.3. Filtering process

As we have said, we use the vector model (Korfhage, 1997) to represent the resource scope and the user's topics of interest. This vector model uses similarity calculations to do the matching process, such as Euclidean Distance or Cosine Measure. Exactly we use the standard cosine measure (Korfhage, 1997). However, as we have linguistic values, we need to introduce a new linguistic similarity measure:

$$\sigma_l(VR, VU) = \Delta\left(\frac{\sum_{k=1}^{n}(\Delta^{-1}(r_k, \alpha_{rk}) \times \Delta^{-1}(u_k, \alpha_{uk}))}{\sqrt{\sum_{k=1}^{n}(\Delta^{-1}(r_k, \alpha_{rk}))^2} \times \sqrt{\sum_{k=1}^{n}(\Delta^{-1}(u_k, \alpha_{uk}))^2}}\right),$$

where n is the number of terms used to define the vectors (i.e. the number of UNESCO codes of level 2), (r_k, α_{rk}) is the 2-tuple linguistic value of term k in the resource vector (VR), (u_k, α_{uk}) is its 2-tuple linguistic value in the user vector (VU). With this similarity measure we obtain a linguistic value to assess the similarity between a resource and a user. In the case of two users or two resources, this linguistic similarity measure can be applied in a similar way.

Following this approach, when a new resource has been inserted into the system, we compute the linguistic similarity measure $\sigma_l(VR_i, VU_j)$ between the new resource scope vector (VR_i) against all the user vectors ($VU_j, j = 1, \ldots, m$ where m is the number of users of the system) to find the fit users to deliver this information. If $\sigma_l(VR_i, VU_j) \geqslant \psi$, the user j is chosen. Previously we have defined a linguistic threshold value (ψ) to filter out the information. In this iteration, the system takes into account also the user preferences (kind of resources and amounts) to consider the user or not. The collaboration preferences are used to classify the selected users in two sets, collaborators $\mathscr{U}_C$ and non-collaborators $\mathscr{U}_N$.

Fig. 5. Example of a user access.

Abstract	Date			
FP7 Cooperation Work Programme: Theme 2 – Food, Agriculture and Fisheries, and Biotechnology	22/12/2006	Filtering	Edit	Delete
Complementary Actions	17/04/2007	Filtering	Edit	Delete
Subsidies of National Program of Environmental Sciences and Technologies	18/07/2007	Filtering	Edit	Delete

Fig. 6. Example of resources list.

After this, the system has two sets of selected users $\mathscr{U}_N$ and $\mathscr{U}_C$, and for each user it has a value $\sigma_l(VR_i, VU_j) \geqslant \psi$. The system applies to each $\sigma_l(VR_i, VU_j)$ the transformation function defined in Definition 6 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 $\mathscr{U}_N$ the resource information and its calculated linguistic relevance degree.

For the users in $\mathscr{U}_C$ the system performs an additional step; it calculates the collaboration possibilities between the selected users. To do it, between each two users $x, y \in \mathscr{U}_C$:

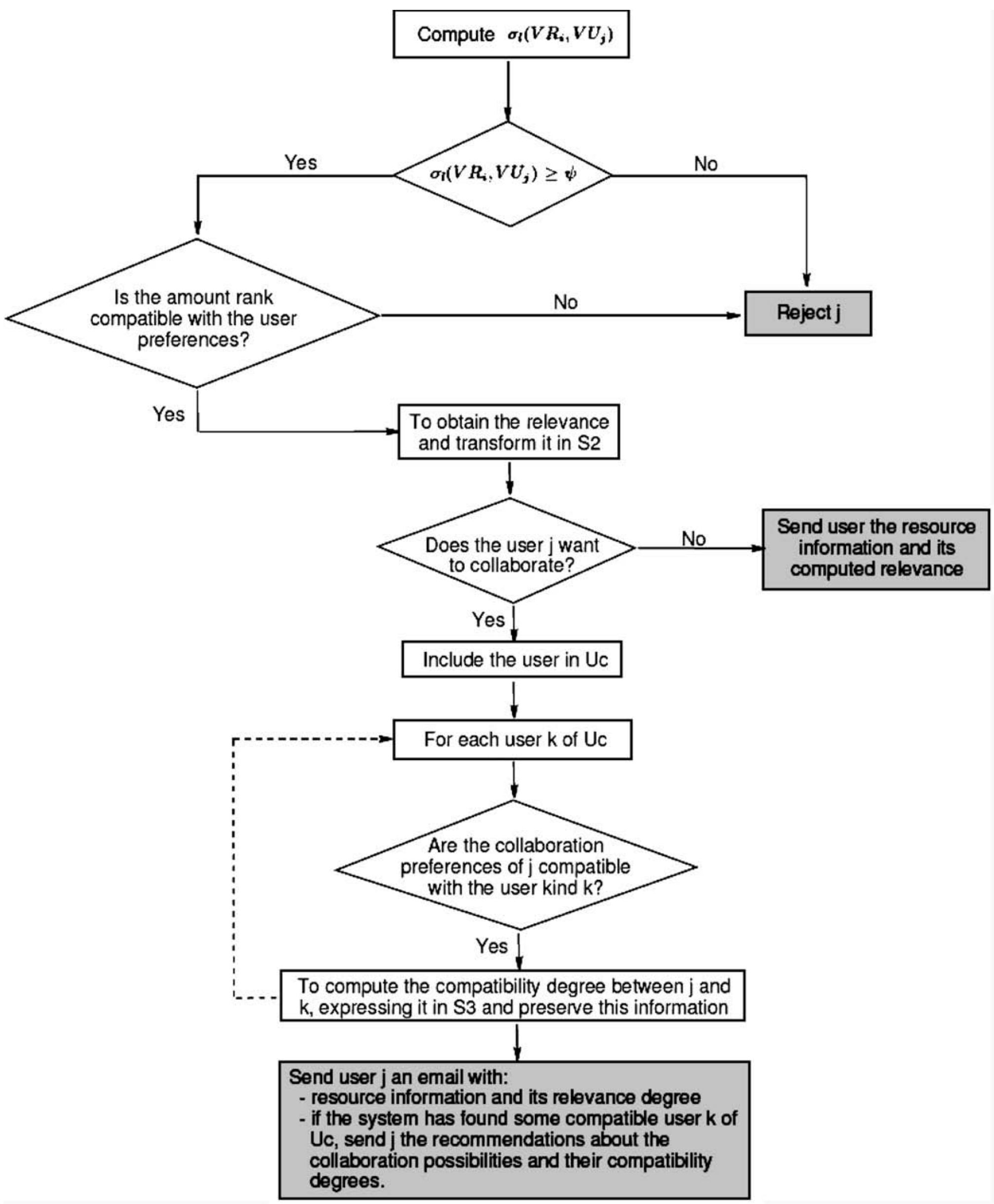

Fig. 7. Filtering process for a user j.

- to analyze if the users are researchers or company employees and take into account the users preferences about it. For example, a researcher could want to collaborate only with others researchers of different research group,
- to calculate the linguistic similarity measure between the users, $\sigma_l(VU_x, VU_y)$,
- to obtain the compatibility degree between x and y, expressing $\sigma_l(VU_x, VU_y)$ as a linguistic label in S_3 (using the transformation function defined in Definition 6) to send it to the user.

Finally, the system sends to the users of $\mathscr{U}_C$ the resource information, its calculated linguistic relevance degree and the collaboration possibilities along with a linguistic compatibility degree.

The Fig. 7 shows all the process.

3.3.4. Feedback phase

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

- The user accesses the system entering his/her $\mathscr{ID}$ and password.
- The user can do the following operations:
  - to edit his/her collaboration preferences,
  - to edit his/her preferences about minimum and maximum amount,
  - to edit his/her topics of interest:
    - to add new UNESCO codes with its importance degrees, i.e. 2-tuple linguistic (b_i, α_i) with $b_i \in S_1$ and $\alpha_i \in [-.5, .5)$,
    - to delete an existing UNESCO code,
    - to modify the importance degree (2-tuple) assigned to an existing UNESCO code.

Example 3. Going back to the Example 1, let us suppose the user $\mathscr{ID}$ wants to update his/her profile because $\mathscr{ID}$ thinks he/she should belong to the category *3309 – Food Technology*. In this case the user wants to add a new UNESCO code and assigns it a importance degree of *High* $((b_3, 0), with b_3 \in S_1)$; this code is in the position 118 of the UNESCO codes list and therefore is in the position 118 of the vector.

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

$$\begin{aligned} VU_{\mathscr{ID}}[y] &= (b_4, 0), \quad \text{if } y = 100 \\ VU_{\mathscr{ID}}[y] &= (b_3, 0), \quad \text{if } y = 118 \\ VU_{\mathscr{ID}}[y] &= (b_0, 0), \quad \text{otherwise.} \end{aligned}$$

4. Experiment and evaluation

This section presents the evaluation of SIRE2IN, which has been implemented in the TTO of University of Granada. The main focus in evaluating the system is to determinate if it fulfills the proposed objectives, that is, the recommended information is useful for the users. Now we have implemented a trial version, in which the system works only with few researchers. In a later version we will include the possibility of a free register in the system for all research community and companies.

To evaluate this primary version of SIRE2IN we have designed experiments in which the proposed system is used to recommend research resources that best satisfy the preferences of ten users that work in Information and Communication Technologies (ICT) area. For a whole evaluation we must include the collaboration recommendations, but in this initial version there aren't very much users, so we evaluate only the recommendations about research resources.

4.1. Evaluation metrics

For the evaluation of recommender systems precision, recall and F1 are measures widely used to evaluate the quality of the recommendations (Cao & Li, 2007; Cleverdon et al., 1966; Sarwar, Karypis, & Konstan, 2000). To calculate these metrics we need a contingency table to categorize the items with respect to the information needs. The items are classified both as relevant or irrelevant and selected (recommended to the user) or not selected. The contingency table (Table 4.1) is created using these four categories.

Precision is defined as the ratio of the selected relevant items to the selected items, that is, it measures the probability of a selected item be relevant:

$$P = \frac{N_{rs}}{N_s}$$

Recall is calculated as the ratio of the selected relevant items to the relevant items, that is, it represents the probability of a relevant items be selected:

$$R = \frac{N_{rs}}{N_r}.$$

F1 is a combination metric that gives equal weight to both precision and recall (Cao & Li, 2007; Sarwar et al., 2000):

$$F1 = \frac{2 \times R \times P}{R + P}.$$

4.2. Experiment result

The purpose of the experiments is to test the performance of the proposed recommender system, so we take into account the recommendations made about the research resources. We consider a data set with 25 research resources of different areas collected by the TTO experts from different information sources about research resources. These resources are included into the system following the indications described above and the system recommends these resources to the suitable users of the ICT area. The system considers that nine resources in all 25 resources are interesting for researchers of the ICT area. Therefore the system recommends nine resources to the users. In particular, 10 researchers use our experimental recommender system and evaluate the relevance of the recommended resources. The contingency table for each one is shown in Table 4.2.

The corresponding precision, recall and F1 are shown in Table 4.3. The average of precision, recall and F1 metrics are 51.11%, 67.67% and 57.62%, respectively. The Fig. 8 shows a graph with the precision, recall and F1 values for each user. These values reveals a good performance of the proposed system.

Table 4.1
Contingency table

	Selected	Not selected	Total
Relevant	Nrs	Nrn	Nr
Irrelevant	Nis	Nin	Ni
Total	Ns	Nn	N

Table 4.2
Experimental contingency table

	User1	User2	User3	User4	User5	User6	User7	User8	User9	User10
Nrs	5	6	3	4	5	6	5	3	4	5
Nrn	3	2	2	2	3	2	2	1	3	2
Nis	4	3	6	5	4	3	4	6	5	4
Nr	8	8	5	6	8	8	7	4	7	7
Ns	9	9	9	9	9	9	9	9	9	9

Table 4.3
Detailed experiment result

	Precision (%)	Recall (%)	F1 (%)
User1	55.56	62.50	58.82
User2	66.67	75.00	70.59
User3	33.33	60.00	42.86
User4	44.44	66.67	53.33
User5	55.56	62.50	58.82
User6	66.67	75.00	70.59
User7	55.56	71.43	62.50
User8	33.33	75.00	46.15
User9	44.44	57.14	50.00
User10	55.56	71.43	62.50
Average	51.11	67.67	57.62

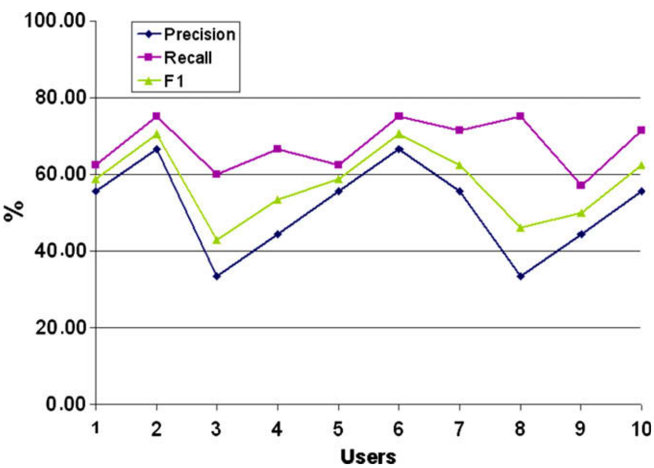

Fig. 8. Experiment result.

5. Concluding remarks

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. Because of this, users are in need of tools to assist them cope with the large amount of information available on the Web and they receive by email. In this paper, we have studied a particular case of information access and we have presented SIRE2IN, a recommender system using both information filtering tools and FLM. The proposed system is oriented to researchers of the University and environment companies and allows them to obtain automatically information about research resources interesting for them. In particular, it is a 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 FLM has been applied in order to improve the experts-system interaction and researchers-system interaction. Experimental results have shown the useful and effectiveness of our systems.

Acknowledgement

This paper has been developed with the financing of SAINFO-WEB Project (TIC00602), FUZZYLING Project (TIN2007-61079) and PETRI Project (PET2007-0460).

References

Arfi, B. (2005). Fuzzy decision making in politics: A linguistic fuzzy-set approach (LFSA). *Political Analysis, 13*(1), 23–56.

Basu, C., Hirsh, H., & Cohen, W., (1998). Recommendation as classification: Using social and content-based information in recommendation. In *Proceedings of the fifteenth national conference on artificial intelligence*. (pp. 714–720).

Ben-Arieh, D., & Zhifeng, C. (2006). Linguistic labels aggregation and consensus measure for autocratic decision-making using group recommendations. *IEEE Transactions on Systems, Man, and Cybernetics Part A – Systems and Humans, 36*(3), 558–568.

Cao, Y., & Li, Y. (2007). An intelligent fuzzy-based recommendation system for consumer electronic products. *Expert Systems with Applications, 33*, 230–240.

Chang, S. L., Wang, R. C., & Wang, S. Y. (2007). Applying a direct multigranularity linguistic and strategy-oriented aggregation approach on the assessment of supply performance. *European Journal of Operational Research, 177*(2), 1013–1025.

Chen, Z., & Ben-Arieh, D. (2006). On the fusion of multi-granularity linguistic label sets in group decision making. *Computers and Industrial Engineering, 51*(3), 526–541.

Claypool, M., Gokhale, A., & Miranda, T., (1999). Combining content-based and collaborative filters in an online newpaper. In *Proceedings of the ACM SIGIR workshop on recommender systems-implementation and evaluation*.

Cleverdon, C. W., & Keen, E. M. (1966). *Factors determining the performance of indexing systems. Test results. ASLIB Cranfield research project* (Vol. 2). Bedford, England: Cranfield.

Cordón, O., Herrera, F., & Zwir, I. (2001). Linguistic modelling by hierarchical systems of linguistic rules. *IEEE Transactions on Fuzzy Systems, 10*(1), 2–20.

Good, N., Shafer, J. B., Konstan, J. A., Borchers, A., Sarwar, B. M., Herlocker, J. L., et al. (1999). Combining collaborative filtering with personal agents for better recommendations. In *Proceedings of the sixteenth national conference on artificial intelligence* (pp. 439–446).

Hanani, U., Shapira, B., & Shoval, P. (2001). Information filtering: Overview of issues, research and systems. *User Modeling and User-Adapted Interaction, 11*, 203–259.

Herrera, F., & Herrera-Viedma, E. (1997). Aggregation operators for linguistic weighted information. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems, 27*, 646–656.

Herrera, F., Herrera-Viedma, E., & Martínez, L. (2008). A fuzzy linguistic methodology to deal with unbalanced linguistic term sets. *IEEE Transactions on Fuzzy Systems, 16*(2), 354–370.

Herrera, F., Herrera-Viedma, E., & Verdegay, J. L. (1996). A linguistic decision process in group decision making. *Group Decision and Negotiation, 5*, 165–176.

Herrera, F., Herrera-Viedma, E., & Verdegay, J. L. (1996). Direct approach processes in group decision making using linguistic OWA operators. *Fuzzy Sets and Systems, 79*, 175–190.

Herrera, F., Herrera-Viedma, E., & Verdegay, J. L. (1998). Choice processes for non-homogeneous group decision making in linguistic setting. *Fuzzy Sets and Systems, 94*, 287–308.

Herrera, F., & Martínez, L. (2000). A 2-tuple fuzzy linguistic representation model for computing with words. *IEEE Transactions on Fuzzy Systems, 8*(6), 746–752.

Herrera, F., & Martínez, L. (2001). 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.

Herrera-Viedma, E. (2001). Modeling the retrieval process of an information retrieval system using an ordinal fuzzy linguistic approach. *Journal of the American Society for Information Science and Technology, 52*(6), 460–475.

Herrera-Viedma, E. (2001). An information retrieval system with ordinal linguistic weighted queries based on two weighting elements. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 9*, 77–88.

Herrera-Viedma, E., Cordón, O., Luque, M., López, A. G., & Muñoz, A. M. (2003). A model of fuzzy linguistic IRS based on multi-granular linguistic information. *International Journal of Approximate Reasoning, 34*(3), 221–239.

Herrera-Viedma, E., & López-Herrera, A. G. (2007). A model of information retrieval system with unbalanced fuzzy linguistic information. *International Journal of Intelligent Systems, 22*(11), 1197–1214.
Herrera-Viedma, E., López-Herrera, A. G., Luque, M., & Porcel, C. (2007). A fuzzy linguistic IRS model based on a 2-tuple fuzzy linguistic approach. *International Journal of Uncertainty, Fuzziness and Knowledge-based Systems, 15*(2), 225–250.
Herrera-Viedma, E., López-Herrera, A. G., & Porcel, C. (2005). Tuning the matching function for a threshold weighting semantics in a linguistic information retrieval system. *International Journal of Intelligent Systems, 20*(9), 921–937.
Herrera-Viedma, E., Martínez, L., Mata, F., & Chiclana, F. (2005). Aconsensus support system model for group decision-making problems with multi-granular linguistic preference relations. *IEEE Transactions on Fuzzy Systems, 13*(5), 644–658.
Herrera-Viedma, E., Pasi, G., López-Herrera, A. G., & Porcel, C. (2006). Evaluating the information quality of web sites: A methodology based on fuzzy computing with words. *Journal of the American Society for Information Science and Technology, 57*(4), 538–549.
Herrera-Viedma, E., & Peis, E. (2003). Evaluating the informative quality of documents in SGML-format using fuzzy linguistic techniques based on computing with words. *Information Processing and Management, 39*(2), 233–249.
Herrera-Viedma, E., Peis, E., Morales del Castillo, J. M., Alonso, S., & Anaya, E. K. (2007). A fuzzy linguistic model to evaluate the quality of web sites that store XML documents. *International Journal of Approximate Reasoning, 46*, 226–253.
Hsu, M. H. (2008). A personalized English learning recommender system for ESL students. *Expert Systems with Applications, 34*, 683–688.
Korfhage, R. R. (1997). *Information storage and retrieval*. New York: Wiley Computer Publishing.
Popescul A., Ungar, L. H., Pennock, D. M., & Lawrence, S. (2001) 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)* (pp. 437–444). San Francisco.
Quiroga, L. M., & Mostafa, J. (2002). An experiment in building profiles in information filtering: The role of context of user relevance feedback. *Information Processing and Management, 38*, 671–694.
Reisnick, P., & Varian, H. R. (1997). Recommender systems. *Special issue of Communications of the ACM, 40*(3), 56–59.
Sarwar, B., Karypis, G., Konstan, J., & Riedl J., (2000). Analysis of recommendation algorithms for e-comerce. In *Proceedings of ACM e-commerce 2000 conference (pp. 158–167)*..
The Centre for Innovation, Research and technology transfer. The Polytechnic University of Valencia. <http://www.ctt.upv.es>.
The UNESCO terminology. Ministerio de Educación y Ciencia. <http://www.mec.es/ciencia/jsp/plantilla.jsp?area=proyectos/invest&id=53>.
Xu, Z. S. (2006). An approach based on the uncertain lowg and induced uncertain lowg operators to group decision making with uncertain multiplicative linguistic preference relations. *Decision Support Systems, 41*(6), 488–499.
L.A. Zadeh (1975). The concept of a linguistic variable and its applications to approximate reasoning. Part I, Information Sciences, 8(1975), 199–249 (Part II, Information Sciences, 8 (1975) 301–357. Part III, Information Sciences, 9 (1975) 43–80).