Expert Systems with Applications 36 (2009) 12520–12528

Contents lists available at ScienceDirect

Expert Systems with Applications

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

A multi-disciplinar recommender system to advice research resources in University Digital Libraries

C. Porcel [a,*], J.M. Moreno [b], E. Herrera-Viedma [c]

[a] University of Jaen, Department of Computer Science, Jaen, Spain
[b] University of Murcia, Department of Computer Science, Murcia, Spain
[c] University of Granada, Department of Computer Science and A.I., Granada, Spain

ARTICLE INFO

Keywords:
Recommender systems
Fuzzy linguistic modeling
University Digital Libraries

ABSTRACT

The Web is one of the most important information media and it is influencing in the development of other media, as for example, newspapers, journals, books, and libraries. In this paper, we analyze the logical extensions of traditional libraries in the Information Society. In Information Society people want to communicate and collaborate. So, libraries must develop services for connecting people together in information environments. Then, the library staff need automatic techniques to facilitate so that a great number of users can access to a great number of resources. *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. We present a model of a fuzzy linguistic recommender system to help the *University Digital Libraries* users to access for their research resources. This system recommends researchers specialized and complementary resources in order to discover collaboration possibilities to form multi-disciplinar groups. In this way, this system increases social collaboration possibilities in a university framework and contributes to improve the services provided by a University Digital Library.

1. Introduction

In the 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 displayed on the 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 (Callan et al., 2003; Gonçalves, Fox, Watson, & Kipp, 2004; Renda & Straccia, 2005).

Digital libraries are the logical extensions of physical libraries in the electronic information society. These extensions amplify existing resources and services. As such, digital libraries offer new levels of access to broader audiences of users and new opportunities for the library. In practice, a digital library makes its contents and services remotely accessible through networks such as the Web or limited-access intranets (Marchionini, 2009).

The digital libraries are composed of human resources (staff) that take over handle and enable the users to access the documents that are more interesting for them, taking into account their needs or areas of interest. The library staff searches, evaluates, selects, catalogues, classifies, preserves and schedules the digital documents access (Gonçalves et al., 2004). 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.

Libraries offer different types of references and referral services (e.g., ready reference, exhaustive search, and selective dissemination of information), instructional services (e.g., bibliographic instruction and database searching), added value services (e.g., bibliography preparation, and language translation) and promotional services (e.g., literacy and freedom of expression). As digital libraries become commonplace and as their contents and services become more varied, the users expect more sophisticated services from their digital libraries (Callan et al., 2003; Gonçalves et al., 2004; Renda & Straccia, 2005).

A service that is particularly important is the selective dissemination of information or filtering (Morales del Castillo, Pedraza-Jiménez, Ruíz, Peis, & Herrera-Viedma, 2009; Morales del Castillo, Peis, Moreno, & Herrera-Viedma, in press). Users develop profiles

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

0957-4174/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2009.04.038

that reveals their areas of interest and as new materials are added to the collection, they are compared to the profiles and relevant items are sent to the users (Marchionini, 2009).

One interesting extension of this concept is to use the connectivity inherent in digital libraries to support collaborative filtering, where users rate or add value to information objects and these ratings are shared with a large community, so that popular items can be easily located or people can search for objects found useful by others with similar profiles (Hanani, Shapira, & Shoval, 2001; Marchionini, 2009; Reisnick & Varian, 1997).

Digital libraries have been applied in many contexts but in this paper we focus on an academic environment. University Digital Libraries (UDLs) provide information resources and services to students, faculty and staff in an environment that supports learning, teaching and research (Chao, 2002).

In this paper we propose a fuzzy linguistic recommender system to achieve major advances in the activities of UDL in order to improve their performance. The system is oriented to researchers and it recommends two types of resources: in the first place, specialized resources of the user research area, and in the second place, complementary resources in order to include resources of related areas that could be interesting to discover collaboration possibilities with other researchers and to form multi-disciplinar groups. As in (Porcel, López-Herrera, & Herrera-Viedma, 2009) we combine a recommender system, to filter out the information, with a multi-granular Fuzzy Linguistic Modeling (FLM), to represent and handle flexible information by means of linguistic labels (Chang, Wang, & Wang, 2007; Chen & Ben-Arieh, 2006; Herrera & Martínez, 2001; Herrera-Viedma, Cordón, Luque, López, & Muñoz, 2003; Herrera-Viedma, Martínez, Mata, & Chiclana, 2005; Herrera, Herrera-Viedma, & Martínez, 2008).

The paper is structured as follows. Section 2 revises some preliminaries, i.e., the concept and main aspects about recommender systems and the approaches of FLM that we use to the system design, the 2-tuple FLM and the multi-granular FLM. In Section 3 we present a multi-disciplinar fuzzy linguistic recommender systems to advice research resources in UDL. Section 4 reports the system evaluation and some experimental results. Finally, some concluding remarks are pointed out.

2. Preliminaries

2.1. Recommender systems

Recommender systems could be defined as systems that produce individualized recommendations as output or has the effect of guiding the user in a personalized way to interesting or useful objects in a large space of possible options (Burke, 2002). They are becoming popular tools for reducing information overload and for improving the sales in e-commerce web sites (Burke, 2007; Cao & Li, 2007; Hsu, 2008; Reisnick & Varian, 1997).

It is a research area that offers tools for discriminating between relevant and irrelevant information by providing personalized assistance for continuous information accesses, filtering the information and delivering it to people who need it (Reisnick & Varian, 1997). Automatic filtering services differ from retrieval services in that in filtering the corpus changes continuously, the users have long time information needs (described by mean of user profiles instead of to introduce a query into the system) and their objective is to remove irrelevant data from incoming streams of data items (Hanani et al., 2001; Marchionini, 2009; Reisnick & Varian, 1997). A result from a recommender system is understood as a recommendation, an option worthy of consideration; a result from an information retrieval system is interpreted as a match to the user's query (Burke, 2007).

A variety of techniques have been proposed as the basis for recommender systems. We can distinguish four different classes of recommendation techniques based on the source of knowledge (Burke, 2007; Hanani et al., 2001; Reisnick & Varian, 1997):

- *Content-based systems*: They generate the recommendations taking into account the terms used in the items representation and the ratings that a user has given to them (Basu, Hirsh, & Cohen, 1998; Claypool, Gokhale, & Miranda, 1999). These recommender systems tend to fail when little is known about the user information needs.
- *Collaborative systems*: The system generates recommendations using explicit or implicit preferences from many users, ignoring the items representation. Collaborative systems locate peer users with a rating history similar to the current user and they generate recommendations using this neighborhood (Good et al., 1999; Renda & Straccia, 2005).
- *Demographic systems*: A demographic recommender system provides recommendations based on a demographic profile of the user. Recommended items can be generated for different demographic niches, by combining the ratings of users in those niches (Pazzani, 1999).
- *Knowledge-based systems*: These systems generate the recommendations based on the inferences about items that satisfy the users from the information provided by each user regarding his/her knowledge about items that can be recommended (Burke, 2002).

All these techniques have benefits and disadvantages. However, 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 et al., 1999; Good et al., 1999). 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. Therefore, 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).

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; Reisnick & Varian, 1997).

2.2. Fuzzy linguistic modeling

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 (Cabrerizo, Alonso, & Herrera-Viedma, 2009; Herrera, Herrera-Viedma, & Verdegay, 1996; Mata, Martínez, & Herrera-Viedma, 2009), quality evaluation (Herrera-Viedma, Pasi, López-Herrera, & Porcel, 2006; Herrera-Viedma & Peis, 2003), models of information retrieval (Herrera-Viedma, 2001a, 2001b; Herrera-Viedma & López-Herrera, 2007; Herrera-Viedma, López-Herrera, Luque, & Porcel, 2007; Herrera-Viedma, López-Herrera, & Porcel, 2005), and political analysis (Arfi, 2005). 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 which 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 $S = \{s_0, \dots, s_g\}$ be a linguistic term set with odd cardinality, where the mid term represents a indifference value and the rest of the terms is symmetrically related to it. We assume that the semantics of the labels is given by means of triangular membership functions and we consider that all terms are 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 $\beta \notin \{0, \dots, g\}$, then an approximation function is used to express the result in S.

Definition 1 Herrera and 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 $(s_i, \alpha_i), s_i \in S$ 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 and Martínez, 2000). Let $S = \{s_0, \dots, s_g\}$ 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:

$$\Delta : [0, g] \to S \times [-0.5, 0.5)$$
$$\Delta(\beta) = (s_i, \alpha), \quad \text{with} \begin{cases} s_i & i = round(\beta), \\ \alpha = \beta - i & \alpha \in [-.5, .5), \end{cases}$$

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 $x = \{(r_1, \alpha_1), \dots, (r_n, \alpha_n)\}$ be a set of linguistic 2-tuples, the 2-tuple arithmetic mean $\bar{x}^e$ is computed as,

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

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

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

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

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

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 determine 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 to manage multi-granular linguistic information. In (Herrera & Martínez, 2001) a multi-granular 2-tuple FLM based on the concept of linguistic hierarchy is proposed.

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 a different granularity $n(t)$ from the remaining of levels of the hierarchy. 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: $l(t, n(t)) \to l(t + 1, 2 \cdot n(t) - 1)$. 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).

A graphical example of a linguistic hierarchy is shown in Fig. 1.

Table 1
Linguistic hierarchies.

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

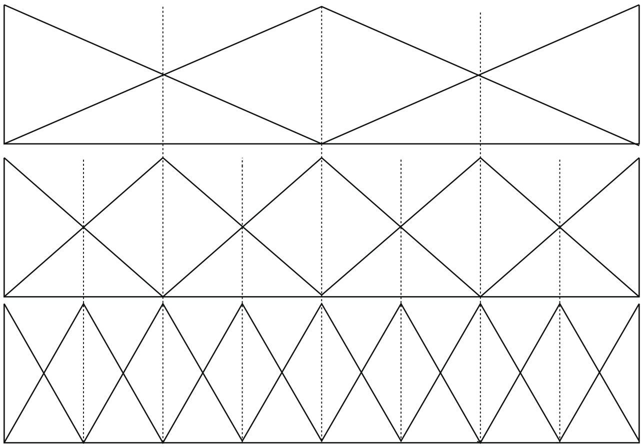

Fig. 1. Linguistic hierarchy of 3, 5 and 9 labels.

Herrera and Martínez (2001) demonstrated that the linguistic hierarchies are useful to represent 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:

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

As it was pointed out in Herrera and Martínez (2001) this family of transformation functions is bijective. This result guarantees that the transformations between levels of a linguistic hierarchy are carried out without loss of information. To define the computational model, we select a level to make uniform the information (for instance, the greatest granularity level) and then we can use the operators defined in the 2-tuple FLM.

3. A multi-disciplinar recommender system to advice research resources in UDL

In this section we present a fuzzy linguistic recommender system designed using a hybrid approach and assuming a multi-granular FLM. This system is applied to advice users on the best research resources that could satisfy their information needs in a UDL. Moreover, the system recommends complementary resources that could be used by the users to meet other researchers of related areas with the aim to discover collaboration possibilities and so, to form multi-disciplinar groups. In this way, it improves the services that a UDL could provide users.

The UDL staff manages and spreads a lot of information resources, such as electronic books, electronic papers, electronic journals, and official dailies (Callan et al., 2003; Renda & Straccia, 2005). Nowadays, this amount of information is growing up and they are in need of automated tools to filter and spread that information to the users in a simple and timely manner.

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 (Callan et al., 2003; Gonçalves et al., 2004; Renda & Straccia, 2005). So, the digital libraries should anticipate the users' needs and recommend about resources that could be interesting for them.

We present a hybrid recommender system that combines both the content-based and collaborative approaches (Burke, 2007; Hanani et al., 2001; Lekakos & Giaglis, 2006). The system filters the incoming information stream and delivers it to the suitable researchers according to their research areas. It recommends users research resources of their own research areas and of complementary areas. We use typical similarity functions based on threshold values to identify research resources of the own areas (Porcel et al., 2009). For example, we could use the threshold semantic functions defined in Information Retrieval to evaluate weighted queries (Bordogna & Pasi, 1993; Korfhage, 1997). On the other hand, to identify research resources of the complementary areas, we use Gaussian similarity functions (Bordogna & Pasi, 1993; Yager, 2007).

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 communication processes 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 of 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 discipline with respect to a resource scope or user preferences.
- **Relevance degree** (S_2) of a resource for a user.
- **Complementary degree** (S_3) between the resource scope and the user topics of interest.

Following the linguistic hierarchy shown in Fig. 1, 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 complementary degrees $(S_3 = S^9)$. Using this LH the linguistic terms in each level are:

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

The system has three main components: resources management, user profiles management and recommendation process (see Fig. 2).

3.1. Resources management

This module is responsible for the management and representation of the research resources. To characterize a resource, the library staff must insert all the available information, such as the title, author(s), kind of resource (if it is a book, or book chapter, or a paper, or a journal, or a conference, or an official daily), journal

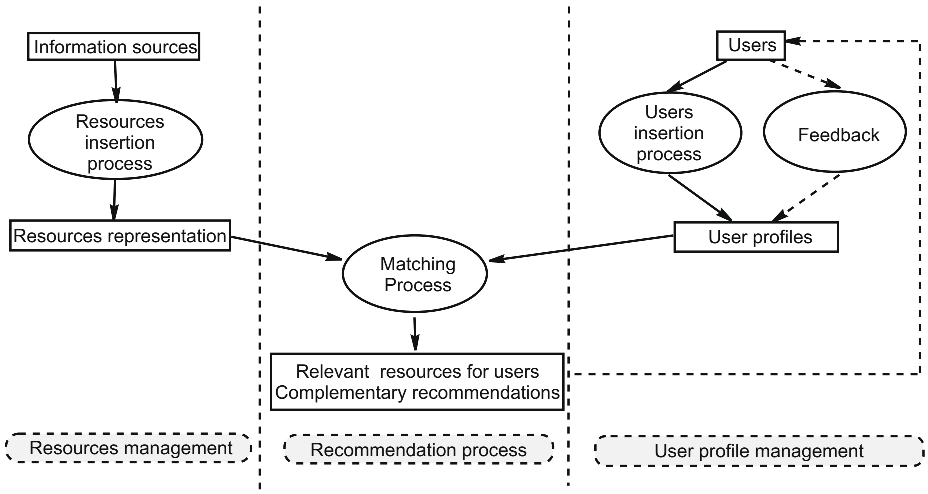

Fig. 2. Structure of the system.

(if it is part of a journal, the system stores the journal name), conference name and dates (if it is a conference), book (if it is a book chapter, the system stores the book title), official daily (if it is part of an official daily, the system stores the daily title), date, source, text, access link to the resource and its scope.

We use the *vector model* to represent the resource scope (Korfhage, 1997). Thus, to represent a resource i, we use a classification composed of 25 disciplines (see Fig. 3). In each position we store a linguistic 2-tuple value representing the importance degree for the resource scope of the discipline in that position:

$$VR_i = (VR_{i1}, VR_{i2}, \dots, VR_{i25}).$$

Then, each component $VR_{ij} \in S_1$, with $j = 1 \dots 25$, indicates the importance degree of the discipline j with regard to the resource i. These importance degrees are assigned by the library staff when they add a new resource.

3.2. User profiles management

To characterize an user, the system stores the following basic information: nickname, password (necessary to access the system), passport number, name and surname, department and center, address, phone number, mobile phone and fax, web, email (elemental information to send the resources and recommendations), research group (it is a string composed of six digits, three characters indicating the research area and three numbers identifying the group), preferences about resources (the users choose the kind of desired resources, i.e. if they want only books, or papers, etc.) and topics of interest.

We use also the vector model (Korfhage, 1997) to represent the topics of interest. Then, for a user x, we have a vector:

$$VU_x = (VU_{x1}, VU_{x2}, \dots, VU_{x25}),$$

where each component $VU_{xy} \in S_1$, with $y = 1, \dots, 25$, stores a linguistic 2-tuple indicating the importance degree of the discipline y with regard to the user x topics of interest. These 2-tuples values are also assigned by the library staff.

The system is based on a content-based approach, but this approach suffers the *cold-start* problem to handle new items or new users (Burke, 2007). New items cannot be recommended to any user until they have been rated by some one. Recommendations for new resources are considerably weaker than those for more widely rated resources. To overcome this problem, in our system, as it was done in other systems (for example in Movielens), when a new user is inserted, the first action to confirm his/her register is to access and assess more than 15 resources of all the resources in the system.

Another aspect of our system is that users can modify the threshold that defines the number of recommendations that they want to receive. So, if the system sends a lot of recommendations, the users can limit this number to N, and in the future they will receive only the N most relevant resources.

☐ Agriculture, animal breeding and fishing	☐ Vegetal and animal biology and ecology
☐ Biotechnology, molecular and cellular biology and genetics	☐ Food science and techonology
☐ Materials science and techonology	☐ Earth science
☐ Social science	☑ Computers science and techonology
☐ Law	☐ Economy
☐ Energy and combustibles	☐ Pharmacology and pharmacy
☐ Philology and philosophy	☐ Physics and space sciences
☐ History and art	☐ Civil engineering, transportations, construction and architecture
☐ Industrial, mechanics, naval and aeronautic engineering	☐ Mathematics
☐ Medicine and veterinary	☐ Environment and environmental technology
☐ Multi-disciplinar	☐ Scientific policy
☐ Psychology and education sciences	☐ Chemistry and chemistry technology
☐ Telecommunications, electric engineering, electronics and automatics	

Fig. 3. Disciplines to define the resource scope.

3.3. Recommendation strategy

In this phase the system generates the recommendations to deliver the information to the fitting users. We use the following recommendation strategies:

- When a new resource is inserted into the system, it recommends this information to the users. In this case, the system follows the content-based approach.
- When a new user is inserted into the system, he/she receives information about resources, previously inserted, interesting for him/her. Now, the system follows the collaborative approach.

Both the processes are based on a Matching Process among the terms used in the users and resources representations (Hanani et al., 2001; Korfhage, 1997). We use the vector model (Korfhage, 1997) to represent both the resource scope and the users 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:

$$\sigma_l(V_1, V_2) = \Delta\left(g \times \frac{\sum_{k=1}^{n}(\Delta^{-1}(v_{1k}, \alpha_{v1k}) \times \Delta^{-1}(v_{2k}, \alpha_{v2k}))}{\sqrt{\sum_{k=1}^{n}(\Delta^{-1}(v_{1k}, \alpha_{v1k}))^2} \times \sqrt{\sum_{k=1}^{n}(\Delta^{-1}(v_{2k}, \alpha_{v2k}))^2}}\right),$$

where g is the granularity of the used term set, n is the number of terms used to define the vectors (i.e. the number of disciplines) and (v_{ik}, α_{vik}) is the 2-tuple linguistic value of term k in the user or resource vector (V_i). With this similarity measure we obtain a linguistic value in S_1 to assess the similarity among the two resources, two users, or a resource and a user.

When a new resource has been inserted into the system, the linguistic similarity measure $\sigma_l(V_i, V_j)$ is computed among the new resource scope vector (V_i) against all the stored resources in the system $(V_j, j = 1, \ldots, m$ where m is the number of resources). If $\sigma_l(V_i, V_j) \geqslant \alpha$ (linguistic threshold value to filter out the information), the resource j is chosen. Next, the system searches for the users which were satisfied with these chosen resources (previously they have rated the resource as good) and takes into account the user preferences (kind of resources) to consider the user or not. To obtain the relevance of the resource i for a selected user x, the system aggregates (using the arithmetic mean defined in Definition 3) the $\sigma_l(V_i, V_j)$ with the assessments previously provided by x about the similar resources and with the assessments provided by others users. To aggregate the information we need to transform the value $\sigma_l(V_i, V_j)$ in a linguistic label in S_2, using the transformation function in Definition 6.

Finally, if the calculated relevance degree is greater than a linguistic threshold μ, then, the system sends the resource information and its calculated linguistic relevance degree (label of S_2) to the selected users. If not, the system proceeds to estimate if the resource could be interesting as a complementary recommendation.

To obtain the complementary recommendations, the system calculates the linguistic similarity measure $\sigma_l(V_i, V_x)$ among the resource i and the user x (for all users). Then, it applies a multi-disciplinar function to the value $\sigma_l(V_i, V_x)$. This function must give greatest weights to similarity middle values (near 0.5), because values of total similarity contribute with efficient recommendations but are probably known for the users. Same, null values of similarity show a null relationship between areas. To establish this function we can use the centered OWA operators in which the OWA weights are generated from a Gaussian type function (Yager, 2007). In the proposed system we use a triangular function (Fig. 4):

$$g(x) = \begin{cases} 2x & \text{for } 0 \leqslant x \leqslant 1/2, \\ 2 - 2x & \text{for } 1/2 < x \leqslant 1. \end{cases}$$

Next, if the obtained multi-disciplinar value is greater than that of a previously defined linguistic threshold γ, the system recommends the complementary resource. To express multi-disciplinar values as a linguistic label in S_3, the transformation function in Definition 6 is used. Finally, the system sends the resource information and its estimated linguistic complementary degree (label of S_3) to the appropriate users.

In the following, we describe the process followed when a new user is inserted into the system. A new user gives few information about the items that satisfied his/her topics of interest, so we use the collaborative approach to generate the recommendations. We follow a memory-based algorithm, which generates the recommendations according to the preferences of nearest neighbors, also known as nearest-neighbor algorithms. These algorithms present good performance as related research reported (Symeonidis, Nanopoulos, Papadopoulos, & Manolopoulos, 2008).

The first step is to identify the users most similar to the new user, using a similarity function. We use the linguistic similarity measure $\sigma_l(V_x, V_y)$ between the topics of interest vectors of the new user (V_x) against all users in the system $(V_y, y = 1, \ldots, n$ where n is the number of users). If $\sigma_l(V_x, V_y) \geqslant \delta$ (linguistic threshold value), the user y is chosen as nearest neighbor of x. Next, the system searches for the resources that satisfied these users and takes into account the user preferences (kind of resources) to consider the resource or not. To obtain the relevance of a resource i for the user x, the system aggregates (using the arithmetic mean defined in Definition 3) the $\sigma_l(V_x, V_y)$ with the assessments previously provided about i by the nearest neighbors of x. To aggregate the information, we need to transform the value $\sigma_l(V_x, V_y)$ in a linguistic label in S_2, using the transformation function in Definition 6.

Finally if the calculated relevance degree is greater than the linguistic threshold μ, then, the system recommends to the new user the resource information and its calculated linguistic relevance degree (label of S_2). If not, the system proceeds to estimate if the resource could be interesting as a complementary recommendation for the new user.

Then the system calculates the linguistic similarity measure $\sigma_l(V_x, V_i)$ among the user x and the resource i (for all resources). Then, it applies the multi-disciplinar function $g(x)$ previously shown (Fig. 4) to the value $\sigma_l(V_x, V_i)$. If the obtained multi-disciplinar value is greater than the that of linguistic threshold γ, the system recommends the resource as complementary. To express multi-disciplinar value as a linguistic label in S_3, the transformation function in Definition 6 is used.

Finally, the system sends to the new users the information of all identified resources that are interesting for them, and its estimated linguistic complementary degree (label of S_3).

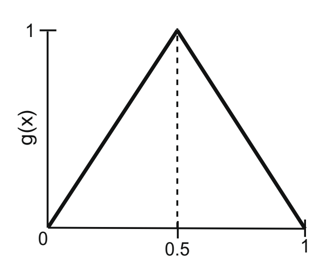

Fig. 4. Triangular function.

3.4. Feedback phase

In this phase the recommender system recalculates and updates the recommendations of the accessed resources. This feedback activity is developed in the following steps:

(1) The system recommends the user U a resource R, and then it asks the user his/her opinion or evaluation judgements about it.
(2) The user communicates the linguistic evaluation judgements, $rc_y \in S_2$.
(3) This evaluation is registered in the system for future recommendations. The system recalculates the linguistic recommendations of R by aggregating the opinions provided by other users together with rc_y provided by U. This can be done using the 2-tuple aggregation operator $\bar{x}^e$ given in Definition 3.

4. Experiment and evaluation

In this section we present the evaluation of the proposed system. The main focus in evaluating the system is to determine if it fulfills the proposed objectives, that is, the recommended information is useful and interesting for the users. At the moment, we have implemented a trial version, in which the system works only with a few researchers. In a later version we will include the system in a UDL.

To evaluate this trial version we have designed experiments in which the system is used to recommend research resources that best satisfy the preferences of 10 users.

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 & Keen, 1966; Sarwar, Karypis, Konstan, & Riedl, 2000). To calculate these metrics we need a contingency table to categorize the items with respect to the information needs. The items are classified as both relevant or irrelevant and selected (recommended to the user) or not selected. The contingency table (Table 5.1) 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 being 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 being 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. Experimental result

The purpose of the experiments is to test the performance of the proposed recommender system, so we compared the recommendations made by the system and the recommendations made by the library staff.

We considered a data set with 50 research resources of different areas, collected by the library staff from different information sources. These resources were included in the system following the indications described above. We limited these experiments to 10 users; all of them completed the registration process and evaluated 15 resources. The resources and the provided evaluations constituted our training data set. After this, we took into account other 20 resources that constituted the test data set. The system filtered this 20 resources and recommends them to the suitable users. Then, we compared the recommendations provided by the systems with the recommendations provided by the library staff, and the obtained contingency table for all users is shown in Table 5.2.

From this contingency table, the corresponding precision, recall and F1 are shown in Table 5.3. The average of precision, recall and F1 metrics is 63.52%, 67.94% and 65.05%, respectively. Fig. 5 shows a graph with the precision, recall and F1 values for each user. These values reveal a good performance of the proposed system and therefore a great satisfaction by the users.

Table 5.1
Contingency table.

	Selected	Not selected	Total
Relevant	N_{rs}	N_{rn}	N_r
Irrelevant	N_{is}	N_{in}	N_i
Total	N_s	N_n	N

Table 5.3
Detailed experimental result.

	Precision (%)	Recall (%)	F1 (%)
User1	71.43	62.50	66.67
User2	66.67	57.14	61.54
User3	60.00	75.00	66.67
User4	33.33	50.00	40.00
User5	62.50	83.33	71.43
User6	62.50	71.43	66.67
User7	75.00	66.67	70.59
User8	80.00	66.67	72.73
User9	66.67	80.00	72.73
User10	57.14	66.67	61.54
Average	63.52	67.94	65.05

Table 5.2
Experimental contingency table.

	User1	User2	User3	User4	User5	User6	User7	User8	User9	User10
N_{rs}	5	4	3	1	5	5	6	4	4	4
N_{rn}	3	3	1	1	1	2	3	2	1	2
N_{is}	2	2	2	2	3	3	2	1	2	3
N_r	8	7	4	2	6	7	9	6	5	6
N_s	7	6	5	3	8	8	8	5	6	7

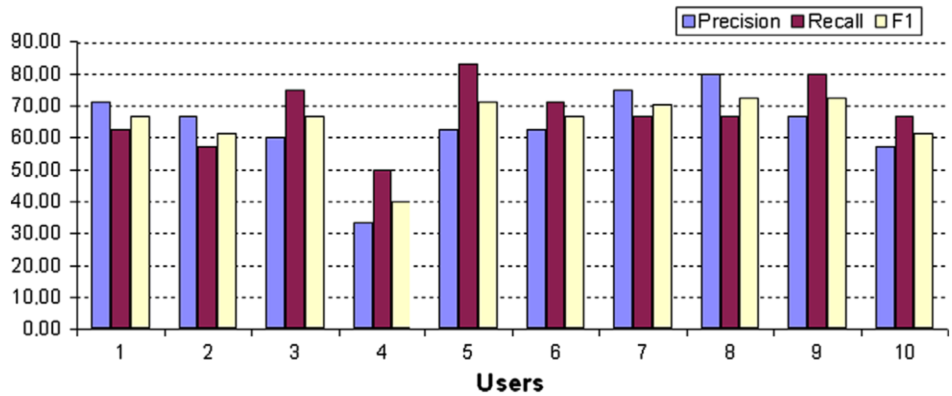

Fig. 5. Experimental result.

5. Conclusions

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. 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 a multi-disciplinar fuzzy linguistic recommender system to spread research resources in UDL. The proposed system is oriented to researchers who receive recommendations about resources that could be interesting for them. In particular, it is a hybrid recommender system that incorporates complementary recommendations. The system filters the incoming information stream to spread the information to the fitting users, and when new users are inserted into the system, they receive interesting information for them. To improve the services that a UDL provides, it additionally recommends complementary resources that allow researchers to discover collaboration possibilities with other colleagues and to form multi-disciplinar groups. The multi-granular fuzzy linguistic modeling has been applied in order to improve the users-system interaction and the interpretability of the system activities. The experimental results show great user satisfaction with the received recommendations.

Acknowledgements

This paper has been developed with the financing of SAINFO-WEB Project (TIC00602) and FEDER funds in FUZZYLING Project (TIN2007-61079), PETRI project (PET2007-0460), and Project of Ministry of Public Works (90/07).

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 *Proc. of the fifteenth national conference on artificial intelligence* (pp. 714–720).

Bordogna, G., & Pasi, G. (1993). A fuzzy linguistic approach generalizing boolean information retrieval: A model and its evaluation. *Journal of the American Society for Information Science, 44*, 70–82.

Burke, R. (2002). Hybrid recommender systems. *User Modeling and User-Adapted Interaction, 12*(4), 331–370.

Burke, R. (2007). Hybrid web recommender systems. In P. Brusilovsky, A. Kobsa, W. Nejdl (Eds.), *The Adaptive Web, LNCS* (Vol. 4321) (pp. 377–408).

Cabrerizo, F. J., Alonso, S., & Herrera-Viedma, E. (2009). A consensus model for group decision making problems with unbalanced fuzzy linguistic information. *International Journal of Information Technology and Decision Making, 8*, 1–23.

Callan, J., Smeaton, A., Beaulieu, M., Borlund, P., Brusilovsky, P., Chalmers, M., et al. (2003). *Personalization and recommender systems in digital libraries*. Joint NSF-EU DELOS Working Group Report. Available from: <http://www.dli2.nsf.gov/internationalprojects/working_group_reports/personalisation.html>.

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.

Chao, H. (2002). Assessing the quality of academic libraries on the Web: The development and testing of criteria. *Library and Information Science Research, 24*, 169–194.

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 *Proc. of the ACM SIGIR-99 workshop on recommender systems-implementation and evaluation*.

Cleverdon, C. W., & Keen, E. M. (1966). *Factors determining the performance of indexing systems* (Vol. 2), Test Results. ASLIB Cranfield Res. Proj., Cranfield, Bedford, England.

Gonçalves, M. A., Fox, E. A., Watson, L. T., & Kipp, N. A. (2004). Streams, structures, spaces, scenarios, societies (5s): A formal model for digital libraries. *ACM Transactions on Information Systems, 22*(2), 270–312.

Good, N., Schafer, 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 *Proc. 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). Direct approach processes in group decision making using linguistic OWA operators. *Fuzzy Sets and Systems, 79*, 175–190.

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. (2001a). 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. (2001b). 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). A consensus 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.
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.
Lekakos, G., & Giaglis, G. M. (2006). Improving the prediction accuracy of recommendation algorithms: Approaches anchores on human factors. *Interacting with Computers, 18*, 410–431.
Marchionini, G. (2009). *Research and development in digital libraries*. <http://ils.unc.edu/march/digital_library_R_and_D.html>. Accessed April 2009.
Mata, F., Martínez, L., & Herrera-Viedma, E. (2009). An adaptive consensus support model for group decision making problems in a multi-granular fuzzy linguistic context. *IEEE Transactions on Fuzzy Systems, 17*(2), 279–290.
Morales del Castillo, J. M., Peis, E., Moreno, J. M., & Herrera-Viedma, E. (in press). D-Fussion: A semantic SDI service for the research community in digital libraries. *Information Research*.
Morales del Castillo, J. M., Pedraza-Jiménez, R., Ruíz, A. A., Peis, E., & Herrera-Viedma, E. (2009). A semantic model of selective dissemination of information for digital libraries. *Information Technology and Libraries, 28*(1), 22–31.
Movielens. *Movie recommendations*. <http://movielens.umn.edu/login>.
Pazzani, M. (1999). A framework for collaborative, content-based and demographic filtering. *AI Review, 13*(5/6), 393–408.
Porcel, C., López-Herrera, A. G., & Herrera-Viedma, E. (2009). A recommender system for research resources based on fuzzy linguistic modeling. *Expert Systems with Applications, 36*, 5173–5183.
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.
Renda, M. E., & Straccia, U. (2005). A personalized collaborative digital library environment: A model and an application. *Information Processing and Management, 41*, 5–21.
Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2000). Analysis of recommendation algorithms for e-commerce. In *Proceedings of ACM e-commerce 2000 conference* (pp. 158–167).
Symeonidis, P., Nanopoulos, A., Papadopoulos, A. N., & Manolopoulos, Y. (2008). Collaborative recommender systems: Combining effectiveness and efficiency. *Expert Systems with Applications, 34*, 2995–3013.
Yager, R. R. (2007). Centered OWA operators. *Soft Computing, 11*, 631–639.
Zadeh, L. A. (1975). The concept of a linguistic variable and its applications to approximate reasoning. Part I. *Information Sciences, 8*, 199–249; Zadeh, L. A. (1975). The concept of a linguistic variable and its applications to approximate reasoning. Part II. *Information Sciences, 8*, 301–357; Zadeh, L. A. (1975). The concept of a linguistic variable and its applications to approximate reasoning. Part III. *Information Sciences, 9*, 43–80.