Dealing with incomplete information in a fuzzy linguistic recommender system to disseminate information in university digital libraries

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

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

ARTICLE INFO

ABSTRACT

As in the Web, the growing of information is the main problem of the academic digital libraries. Thus, similar tools could be applied in university digital libraries to facilitate the information access by the students and teachers. In [46] we presented a fuzzy linguistic recommender system to advice research resources in university digital libraries. The problem of this system is that the user profiles are provided directly by the own users and the process for acquiring user preferences is quite difficult because it requires too much user effort. In this paper we present a new fuzzy linguistic recommender system that facilitates the acquisition of the user preferences to characterize the user profiles. We allow users to provide their preferences by means of incomplete fuzzy linguistic preference relation. We include tools to manage incomplete information when the users express their preferences, and, in such a way, we show that the acquisition of the user profiles is improved.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Digital libraries are information collections that have associated services delivered to user communities using a variety of technologies [8,15,48]. Therefore, 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 [39,50].

As digital libraries become commonplace and as their contents become more varied, the users expect more sophisticated services from them [8,15,48,50]. A service that is particularly important is the selective dissemination of information or filtering, to help the users to access interesting information for them. Users develop interest profiles and as new materials (books, papers, reports, and so on) are added to the collection, they are compared to the profiles and relevant items are sent to the users [39].

Moreover, digital libraries have been applied in a lot of contexts but in this paper we focus on an academic environment. University Digital Libraries (UDL) provide information resources and services to students, faculty and staff in an environment that supports learning, teaching and research [11].

Recommender systems are becoming popular tools for reducing information overload and to improve the sales in e-commerce web sites [7,9,35,40,49]. The use of this kind of systems allows to recommend resources interesting for the users, at the same time that these resources are inserted into the system. In the UDL framework, recommender systems [7,49] can be used to help users (teachers, students and library staff) to find out and select their information and knowledge sources [43].

Generally, in a recommender system the users' information preferences can be used to define user profiles that are applied as filters to streams of documents [7,47,49]. In [45,46] we developed some recommender systems in an academic context. For instance, in [45] we proposed a fuzzy linguistic recommender system for a technology transfer office which helps researchers and environment companies allowing them to obtain information automatically about research resources (calls or projects) in their interest areas; in [46] we proposed a fuzzy linguistic recommender system to achieve major advances in the activities of UDL, which recommends researchers specialized resources and complementary resources related with their respective research areas. The problem of both recommender systems is that users must directly specify their user profiles by providing their preferences on all topics of interest and it requires too much user effort.

In this paper, we focus on the idea of that a recommender system could be seen as a decision support system (DSS) [37,38,44], where the solution alternatives are the digital resources inserted into the library, and the criteria to satisfy are the user profiles. The proper use of these recommendation systems is essential to

---

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

provide real personalized services, and it can substantially reduce information overload and increase user satisfaction. Therefore, it has become an important area in information systems and decision support research [37,38,44]. So, the activity of a recommender system can be seen as a group decision making (GDM) problem, so we can adopt the typical representation formats used in GDM, as for example, fuzzy preference relations [19,20,28,32,41]. This representation format presents a high expressivity and some interesting properties that allow us to work easily. However, in real world problems it is common to find situations in which users are not able to provide all the preference values that are required, and then, we have to deal with *incomplete fuzzy preference relations* [1–3,25,26,41].

The aim of this paper is to present a new fuzzy linguistic recommender defined in a UDL framework which overcomes the problem of user profile characterization observed in the recommender systems defined in [45,46]. In order to improve the system performance, we propose an alternative way to obtain accurate and useful knowledge about the user preferences. This new recommender system allows users to provide their preferences by means of incomplete fuzzy linguistic preference relations [1], and in such a way, we facilitate users the expression of their preferences and, consequently, the determination of user profiles process. The recommender system is able to complete the incomplete preference relations using the tools proposed in [1,2,26]. Each user profile is composed of both user preferences on topics of interest and user preferences on collaboration possibilities with other users. Then, the recommender system is able to recommend both research resources and collaboration possibilities to the users of a UDL. As in [45,46] we define this recommender system in a multi-granular fuzzy linguistic context [10,12,22,27,32,42]. In such a way, we incorporate in the recommender system flexible tools to handle the information by allowing to represent the different concepts of the system with different linguistic label sets.

The rest of the paper is set out as follows. Section 2 presents the preliminaries necessary to develop the proposed model. Section 3 presents the new recommender system to the dissemination of knowledge in a UDL. Section 4 reports the system evaluation and the experimental results. Finally, our conclusions are pointed out in Section 5.

2. Preliminaries

2.1. Recommender systems

Recommender systems could be defined as systems that produce individualized recommendations as output, or have the effect of guiding the user in a personalized way to interesting or useful objects in a large space of possible options [6].

It is a research area that offers tools for discriminating between relevant and irrelevant information by providing personalized assistance for continuous information accesses [43,49]. Automatic filtering services differ from retrieval services [23,24,29–31] in that in filtering the corpus changes continuously, the users have long time information needs (described by means of user profiles) instead of introducing a query into the system, and their objective is to remove irrelevant data from incoming streams of data items [17,39,49]. A result from a recommender system is understood as a recommendation, an option worthy of consideration, while a result from an information retrieval system is interpreted as a match to the user's query [7]. However both systems present some analogies, and in this sense they could be considered a DSS [44]. In both cases, the solution alternatives would be the documents to recommend or retrieve and the criteria to satisfy would be the user profiles and user queries, respectively.

A variety of techniques have been proposed as the basis for recommender systems [7,17,40,49]; all of these techniques have benefits and disadvantages. The use of an hybrid approach is proposed to smooth out the disadvantages of each one of them and to exploit their benefits [5,13,16]. 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 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 preferences [47].

The recommendation activity is followed by a relevance feedback phase. *Relevance feedback* is a cyclic process whereby the users feed back into the system decisions on the relevance of retrieved documents and the system uses these evaluations to automatically update the user profiles [17,49].

2.2. The 2-tuple fuzzy linguistic approach

The fuzzy linguistic modeling (FLM) is a tool based on the concept of *linguistic variable*[52] which has given very good results for modeling qualitative information in many problems, e.g., in decision making [20], quality evaluation [33,34], models of information retrieval [23,24,29–31], political analysis [4], etc.

The 2-tuple FLM [21] 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 [18,52]).

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 symmetrically related 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, $s_i \leqslant s_j \Longleftrightarrow i \leqslant j$. In this fuzzy linguistic context, if a symbolic method [18,20] 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. β is represented by means of 2-tuples (s_i, α_i), $s_i \in S$ and $\alpha_i \in [-.5, .5)$ where 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 2-tuple representation model defines a set of transformation functions between numeric values and 2-tuples $\Delta(\beta) = (s_i, \alpha)$ and $\Delta^{-1}(s_i, \alpha) = \beta \in [0, g]$ [21].

The computational model is defined by presenting a negation operator, comparison of 2-tuples and aggregation operators [21]. 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 operators can be easily extended for dealing with linguistic 2-tuples. Some examples are

Definition 1 (*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,

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

Definition 2 (*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

$$\bar{x}^w[(r_1, \alpha_1), \ldots, (r_n, \alpha_n)] = \Delta\left(\frac{\sum_{i=1}^{n} \Delta^{-1}(r_i, \alpha_i) \cdot w_i}{\sum_{i=1}^{n} w_i}\right)$$
$$= \Delta\left(\frac{\sum_{i=1}^{n} \beta_i \cdot w_i}{\sum_{i=1}^{n} w_i}\right). \quad (2)$$

Table 1
Linguistic hierarchies.

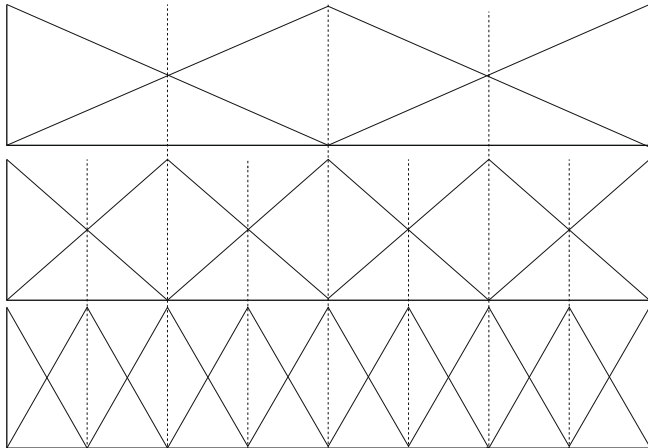|	Level 1	Level 2	Level 3
l(t,n(t))	l(1,3)	l(2,5)	l(3,9)
l(t,n(t))	l(1,7)	l(2,13)	

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

Definition 3 (*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 $\overline{x}_l^w$ is

$$\overline{x}_l^w\left[\left((r_1, \alpha_1), (w_1, \alpha_1^w)\right) \ldots \left((r_n, \alpha_n), (w_n, \alpha_n^w)\right)\right] = \Delta\left(\frac{\sum_{i=1}^n \beta_i \cdot \beta_{W_i}}{\sum_{i=1}^n \beta_{W_i}}\right), \tag{3}$$

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

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. When different experts have different uncertainty degrees on the phenomenon or when an expert has to assess different concepts, then several linguistic term sets with a different granularity of uncertainty are necessary [22,32]. In [22] a multi-granular 2-tuple FLM based on the concept of linguistic hierarchy is proposed.

A *linguistic hierarchy (LH)*, is a set of levels $l(t, n(t))$, where each level t is a linguistic term set with 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 $l(t, n(t)) \rightarrow l(t + 1, 2 \cdot n(t) - 1)$. Table 1 shows the granularity needed in each linguistic term set of the level t depending on the value $n(t)$ defined in the first level (3 and 7, respectively).

A graphical example of a linguistic hierarchy is shown in Fig. 1.
In [22] a family of transformation functions between labels from different levels was introduced:

Definition 4. 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′≠t is defined as

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

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

As it was pointed out in [22] 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.

2.3. Incomplete fuzzy preference relations

Definition 5. A fuzzy preference relation P on a set of alternatives $X = \{x_1, \ldots, x_n\}$ is a fuzzy set on the product set $X \times X$, i.e., it is characterized by a membership function $\mu_P : X \times X \rightarrow [0, 1]$.

When cardinality of X is small, the preference relation may be conveniently represented by the $n \times n$ matrix $P = (p_{ij})$, being $p_{ij} = \mu_P(x_i, x_j)$ ($\forall i, j \in \{1, \ldots, n\}$) interpreted as the preference degree or intensity of the alternative x_i over x_j, where

- $p_{ij} = 1/2$ indicates indifference between x_i and x_j,
- $p_{ij} = 1$ indicates that x_i is absolutely preferred to x_j,
- and $p_{ij} > 1/2$ indicates that x_i is preferred to x_j.

However, as we have mentioned, our system integrates the multi-granular FLM based on 2-tuples, so we must define a linguistic preference relation as follows:

Definition 6. Let $X = \{x_1 \ldots, x_n\}$ a set of alternatives and S a linguistic term set. A linguistic preference relation $P = p_{ij}(\forall i, j \in \{1, \ldots, n\})$ on X is

$$\mu_P : X \times X \rightarrow S \times [0.5, 0.5), \tag{4}$$

where $p_{ij} = \mu_P(x_i, x_j)$ is a 2-tuple which denotes the preference degree of alternative x_i regarding to x_j.

As aforementioned, in many real world GDM problems the experts are often not able to provide all the preference values that are required. In order to model these situations, we use incomplete fuzzy preference relations [1–3,25,26,41].

Definition 7. A function $f : X \rightarrow Y$ is *partial* when not every element in the set X necessarily maps onto an element in the set Y. When every element from the set X maps onto one element of the set Y, then we have a *total* function.

Definition 8. A two-tuple fuzzy linguistic preference relation P on a set of alternatives X with a partial membership function is an *incomplete two-tuple fuzzy linguistic preference relation*.

3. A recommender system for the dissemination of information in UDL using incomplete linguistic preference relations

The UDL staff manage and spread many information about research resources such as electronic books, electronic papers, electronic journals, official dailies and so on [8,48]. 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. On the other hand, UDL users need tools to help them to insert their preferences to form accurate profiles.

In this section we present a new fuzzy linguistic recommender system in which the user profiles are obtained from user preferences represented by incomplete fuzzy linguistic preference relations [1]. This proposal contributes with some advantages with
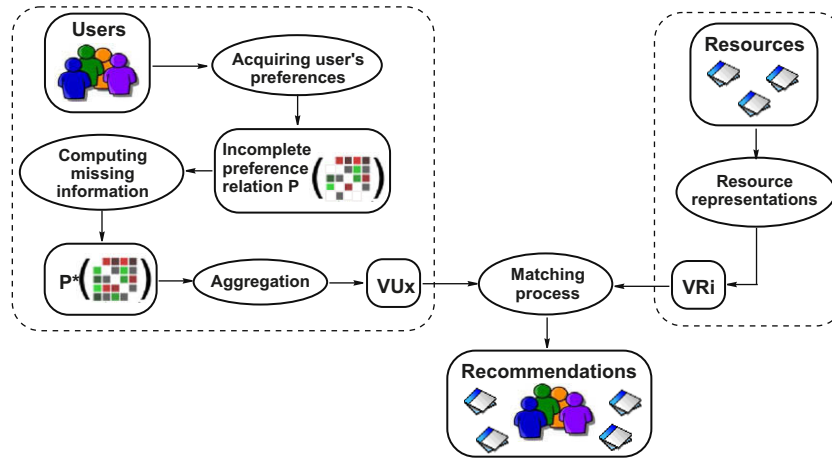
Fig. 2. Operating scheme.

regard to previous systems [46,45] because it facilitates the expression of their preferences to the users and reduces the user effort to characterize their user profiles. It is applied to advise UDL users on the best research resources that could satisfy their information needs in UDL. Moreover, the system recommends collaboration possibilities to meet another researchers of related areas which could collaborate with them in projects or interest works. In such a way, this new recommender system improves the services that a UDL provides to the users, because it is easier to obtain the knowledge about the users and it allows to decrease the time cost to establish the user profiles.

In Fig. 2 we can see the basic operating scheme, which is explained in the following subsections.

3.1. Information representation

In the proposed system, the user-system communication is carried out by using a multi-granular fuzzy linguistic approach [22,32], in order to allow a higher flexibility in the communication processes of the system. 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. We take into account the following concepts that can be assessed in the system:

- **Importance degree** of a discipline with respect to a resource scope or user preferences (S_1).
- **Relevance degree** of a resource for a user (S_2).
- **Compatibility degree** between two users (S_3).
- **Preference degree** of a resource regarding another one (S_4).

Following the linguistic hierarchy shown in Fig. 1, in our system we use the level 2 (5 labels) to assign importance and preference degrees ($S_1 = S^5$ and $S_4 = S^5$), and the level 3 (9 labels) to assign relevance and compatibility degrees ($S_2 = S^9$ and $S_3 = S^9$). Using this LH, the linguistic terms in each level are

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

3.1.1. Resources representation

The considered resources are journal articles, conference contributions, book chapters, books or edited books. Once the library staff insert all the available information about a new resource, the system obtains an internal representation mainly based in the resource scope. We use the *vector model* [36] to represent the resource scope. Thus, to represent a resource i, we use a classification composed by 25 disciplines (see Fig. 3). In each position we store a linguistic 2-tuple value representing the importance degree of the resource scope with respect to the discipline represented by that position:

$$VR_i = (VR_{i1}, VR_{i2}, \ldots, VR_{i25}). \tag{5}$$

Then, each component $VR_{ij} \in S_1$, with $j = \{1, \ldots, 25\}$, indicates the linguistic 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.

3.1.2. User profiles

The user profiles are composed of two kinds of user preferences:

(1) User preferences on topics of interest, and
(2) User preferences on collaboration possibility with other users.

The main contribution of this proposal is how users provide their preferences on topics of interest used to represent the source resources. In previous proposals [45,46] we represented such user preferences using the vector model [36]. The problem is that the users must insert or edit all the features corresponding to the disciplines, i.e., in our case 25 categories. Thus, in previous proposals we worked with vectors composed of 25 positions (each one corresponding to a discipline), but there could exist cases in which this number could be greater. In such a way, users have to perform a great effort to provide their preferences about topics of interest. To reduce this effort and make the process for acquiring the user preferences easier, in this model we propose an alternative method to obtain the user preferences on topics of interest.

We ask users to provide their preferences on some research resources, usually a limited number of resources, four or five. The choice of research resources is made by the personal staff tanking into account the relevance supplied by the users. As in [41] we propose users to represent their preferences by means of incomplete fuzzy linguistic preference relations. Then, the system presents users only a selection of the most representative resources, and

Fig. 3. Interface to define the disciplines of the resource scope.

the users provide their preferences about these resources by means of an incomplete fuzzy preference relation. Furthermore, according to results presented in [2], it is enough that the users provide only a row of the preference relation. Then, we use the method proposed in [2] to complete the relations. Once the system completes the fuzzy linguistic preference relation provided by the user, it is possible to obtain a vector representing the user preferences on the topics of interest. Next, we explain this process in detail:

(1) **Acquiring the user preferences on a limited number of research resources**: At the beginning, the main goal is to help the users to provide their preferences assuring that these preferences are as consistent as possible. The system shows users the five most representative resources, $R = \{r_1 \dots, r_5\}$, and asks them to express their preferences by means of an incomplete fuzzy linguistic preference relation (see Fig. 4). The users only fill those preferences that they wish, assigning labels of S_4. In the preference relation, each preference value p_{ij} represents the linguistic preference degree of resource i over the resource j according to the user feeling. As aforementioned, the simplest case would be to provide a relation with only one row of preference values:

$$P = \begin{pmatrix} - & p_{12} & p_{13} & p_{14} & p_{15} \\ x & - & x & x & x \\ x & x & - & x & x \\ x & x & x & - & x \\ x & x & x & x & - \end{pmatrix}. \tag{6}$$

Then, the system completes the preference relation P using the method proposed in [2], and obtains the relation P^*:

$$P^* = \begin{pmatrix} - & p_{12} & p_{13} & p_{14} & p_{15} \\ p_{21}^* & - & p_{23}^* & p_{24}^* & p_{25}^* \\ p_{31}^* & p_{32}^* & - & p_{34}^* & p_{35}^* \\ p_{41}^* & p_{42}^* & p_{43}^* & - & p_{45}^* \\ p_{51}^* & p_{52}^* & p_{53}^* & p_{54}^* & - \end{pmatrix}, \tag{7}$$

where $p_{1j} \in S_4$ are the degrees inserted by the user about the preferences of the resource x_1 with respect to x_j, p_{ii} represents indifference, and each p_{ij}^* is the estimated degree for the user about his/her preference of the resource x_i with respect to x_j.

(2) In order to obtain **user preferences on topic of interest, i.e., user preference vector**, firstly we calculate the user preference degrees on each considered resource according to the preference relation P^*, and secondly, we use this preference degrees together with the vectors that represent each research resource to obtain the user preference vector. The preference degrees coincides with the dominance degrees of a linguistic preference relation [19]. To obtain them we propose the application of the arithmetic mean $\overline{x}^e$ (Definition 1). Then, the preference degree of the resource i for the expert called DG_i, is computed as follows:

$$DG_i = \overline{x}^e [p_{i1}^*, \dots, p_{i5}^*]. \tag{8}$$

Then, to obtain the user preference vector x, i.e. $VU_x = (VU_{x1}, VU_{x2}, \dots, VU_{x25})$, from the aggregation of the vectors that represents the characteristics of the chosen research resources, i.e., $\{VR_1, \dots, VR_5\}$, weighted by means of the user preference degrees $\{DG_1, \dots, DG_5\}$. To do that, we use the linguistic weighted average operator defined in Definition 3, and then each position $k = \{1, \dots, 25\}$ of the vector VU_x, is computed as follows:

$$VU_{xk} = \overline{x}_l^w [(VR_{1k}, DG_1), \dots, (VR_{5k}, DG_5)]. \tag{9}$$

On the other hand, to complete the user profile, the system asks every user to express his/her collaboration preferences, i.e., if he/she wants to receive recommendations on collaboration possibilities with others users. This could help users to develop multi-disciplinar works or participate in collaborative research projects [46]. They should respond to this question with "Yes" or "No".

3.2. Recommendation strategy

In this phase the system generates the recommendations to deliver the information resources to the fitting users. This process is based on a matching process developed between user profiles and resource representations [17,36]. To do that, we can use different kinds of similarity measures, such as euclidean distance or Cosine Measure. Particularly, we use the standard cosine measure [36]. As the components of the vectors used to represent user profiles and research resources are 2-tuple linguistic values, then we define the cosine measure in a 2-tuple linguistic context. Given two vectors of 2-tuple linguistic values:
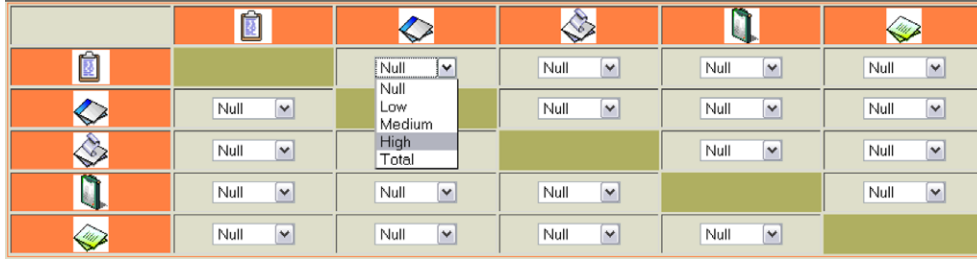
Fig. 4. Interface to define the user preferences.

$$V_1 = ((v_{11}, \alpha_{v11}), (v_{12}, \alpha_{v12}), \ldots, (v_{125}, \alpha_{v125})),$$

and

$$V_2 = ((v_{21}, \alpha_{v21}), (v_{22}, \alpha_{v22}), \ldots, (v_{225}, \alpha_{v225})),$$

then the linguistic similarity between both, called $\sigma_l(V_1, V_2) \in S_1$ is defined as

$$\sigma_l(V_1, V_2)$$

$$= \Delta \left(g \times \frac{\sum_{k=1}^{25} \left(\Delta^{-1}(v_{1k}, \alpha_{v1k}) \times \Delta^{-1}(v_{2k}, \alpha_{v2k}) \right)}{\sqrt{\sum_{k=1}^{25} \left(\Delta^{-1}(v_{1k}, \alpha_{v1k}) \right)^2} \times \sqrt{\sum_{k=1}^{25} \left(\Delta^{-1}(v_{2k}, \alpha_{v2k}) \right)^2}} \right),$$

$$(10)$$

where g is the granularity of S_1 and (v_{ik}, α_{vik}) is the 2-tuple linguistic value of term k in the vector(V_i).

When a new resource i is inserted into the system, we calculate the linguistic similarity measures, $\sigma_l(VR_i, VU_j)$, between the representation vector of this new resource (VR_i) and all the user preference vectors, $\{VU_1, \ldots, VU_m\}$, where m is the number of users in the system. These user preference vectors are obtained as we have indicated in Section 3.1.2.

Then, if $\sigma_l(VR_i, VU_j) \geqslant \psi$, the user j is selected to receive recommendations about resource i. Previously, we have defined a linguistic threshold value (ψ) to filter the output of the system. Next, the system applies to each $\sigma_l(VR_i, VU_j)$ the transformation function defined in Definition 4, to obtain the relevance degree of the resource i for the user j, expressed using a label of the set S_2.

The collaboration preferences provided by the users are used to classify the selected users in two sets, collaborators $\mathscr{U}_C$ and non-collaborators $\mathscr{U}_N$. For the users of $\mathscr{U}_N$ the system has finished the recommendation process, and therefore it sends them the resource information together with its linguistic relevance degree.

For the users in $\mathscr{U}_C$ the system calculates the collaboration possibilities. To do it, between each two users $x, y \in \mathscr{U}_C$, the system performs the following steps:

(1) Calculate the linguistic similarity measure between both users, $\sigma_l(VU_x, VU_y)$.
(2) Obtain the linguistic compatibility degree between both users, which must be expressed in S_3. To do that, we apply the transformation function defined in 4 on $\sigma_l(VU_x, VU_y)$.

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

4. Experiment and evaluation

In this section we present the evaluation of the proposed system. The main focus in evaluating the system is to determinate if

it fulfills the proposed innovations, that is, the recommended information is useful and interesting for the users, reducing the effort and making easier the process for acquiring the user's preferences. Now we have implemented a trial version, in which the system works only with a few researchers.

Table 2
Contingency table for the resources.

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

4.1. Evaluation metrics

In the scope of recommender systems, precision, recall and F1 are measures widely used to evaluate the quality of the recommendations [9,14,43,51]. We use them to compare the new proposal with previous systems. 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 (see Table 2) 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:

$$P = \frac{N_{rs}}{N_s}. \tag{11}$$

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:

$$R = \frac{N_{rs}}{N_r}. \tag{12}$$

F1 is a combination metric that gives equal weight to both precision and recall:

$$F1 = \frac{2 \times R_1 \times P_1}{R_1 + P_1}. \tag{13}$$

4.2. Experimental results

The purpose of the experiment is to test the performance of the proposed system, so we compared the recommendations made by the system with the information provided by the library staff. When the users receive a recommendation, they provide a feedback to the system assessing the relevance of the recommended resource, i.e., they provide their opinions about the recommendation supplied

by the system. If they are satisfied with the recommendation, they provide a higher value. We use that feedback information to evaluate the system, applying the measures described in the previous section.

We considered a data set with 30 research resources of different areas, collected by the library staff from different information sources. These resources were included into the system following the indications described in Section 3.1.1. Initially we limited these experiments to 6 users; all of them completed the registration process and they inserted their preferences about the five most relevant resources presented by the system (like in Fig. 4).

From this information provided by the users, the system builds the user profiles. These user profiles obtained from the provided preferences and the resources previously inserted, constituted our training data set. Then, we added 20 new resources that constituted the test data set. The system filtered these 20 resources and recommended each one to the suitable users. To obtain data to compare, the 20 new resources also were recommended using the advices of the library staff.

For example, for the user 1, the system selected 4 resources as relevant. However, from the information provided by the library staff and the user feedback, we could see that the system selected 1 irrelevant resource for user 1, and it didn't select 2 resources that library staff considered relevant for the user 1. Then, to build the contingency table, we compared the recommendations provided

by the system with the recommendations provided by the library staff and the relevance degrees inserted by the users. With this information, we build the contingency table for the recommended resources. It is shown in Table 3.

From this contingency table, we obtain the corresponding precision, recall and F1 which are shown in Table 4. The average of precision, recall and F1 metrics are 67.50%, 61.39% and 63.51%, respectively. The 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 good user satisfaction.

5. Conclusions

Digital libraries can serve as powerful tools for universities to reach out and expand their sphere of influence in the society. UDL provides effective channels for the dissemination of research information. But 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, as for example, recommender systems [43].

We have proposed a multi-granular linguistic recommender system in this research topic [46]. However, the process for acquiring the user profiles requires great effort, and sometimes, it is complicated by the great quantity of information that the user has to give to characterize their feeling on topics of interest. In this paper we have proposed a new method to overcome this problem. Users do not directly provide the user preference vectors that characterize their profiles. They provide preferences on some research resources and from this information we calculate their respective preference vectors on topics of interest. Furthermore, to facilitate the process for acquiring the user preferences on the resources we allow users to provide their preferences by means of incomplete fuzzy linguistic preference relations. The user profile is completed with user preferences on the collaboration possibilities with other users. Therefore, this recommender system acts as a decision support system that makes decisions about both the resources that could be interesting for a researcher and his/her collaboration possibilities with other researchers to form interesting working groups. The experimental results show the user satisfaction with the received recommendations.

Table 3
Experimental contingency table.

	User 1	User 2	User 3	User 4	User 5	User 6
Nrs	3	2	3	1	4	3
Nrn	2	1	3	1	2	1
Nis	1	1	1	2	1	1
Nr	5	3	6	2	6	4
Ns	4	3	4	3	5	4

Table 4
Detailed experiment results for the recommendations.

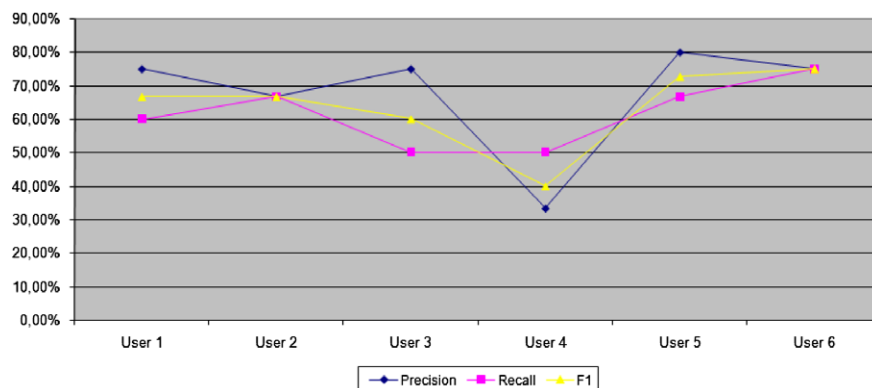|	Precision (%)	Recall (%)	F1 (%)
User 1	75,00	60,00	66,67
User 2	66,67	66,67	66,67
User 3	75,00	50,00	60,00
User 4	33,33	50,00	40,00
User 5	80,00	66,67	72,73
User 6	75,00	75,00	75,00
Average	67,50	61,39	63,51

Fig. 5. Experiment results.

References

[1] S. Alonso, F.J. Cabrerizo, F. Chiclana, F. Herrera, E. Herrera-Viedma, Group decision-making with incomplete fuzzy linguistic preference relations, International Journal of Intelligent Systems 24 (2) (2009) 201–222.

[2] S. Alonso, F. Chiclana, F. Herrera, E. Herrera-Viedma, J. Alcalá-Fdez, C. Porcel, A consistency-based procedure to estimating missing pairwise preference values, International Journal of Intelligent Systems 23 (2008) 155–175.

[3] S. Alonso, E. Herrera-Viedma, F. Chiclana, F. Herrera, Individual and social strategies to deal with ignorance situations in multi-person decision making, International Journal of Information Technology and Decision Making 8 (2) (2009) 313–333.

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

[5] C. Basu, H. Hirsh, W. Cohen, Recommendation as classification: using social and content-based information in recommendation, in: Proceedings of the Fifteenth National Conference on Artificial Intelligence, 1998, pp. 714–720.

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

[7] R. Burke, Hybrid web recommender systems, in: P. Brusilovsky, A. Kobsa, W. Nejdl, (Eds.), The Adaptive Web, LNCS 4321, 2007, pp. 377–408.

[8] J. Callan et al., Personalisation and recommender systems in digital libraries, Joint NSF–EU DELOS Working Group Report, 2003.

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

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

[11] H. Chao, Assessing the quality of academic libraries on the web: the development and testing of criteria, Library & Information Science Research 24 (2002) 169–194.

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

[13] M. Claypool, A. Gokhale, T. Miranda, Combining content-based and collaborative filters in an online newpaper, in: Proceedings of the ACM SIGIR-99 Workshop on Recommender Systems-Implementation and Evaluation, 1999.

[14] C.W. Cleverdon, E.M. Keen, Factors determining the performance of indexing systems. Test results, vol. 2. ASLIB Cranfield Research Project, Cranfield, Bedford, England, 1966.

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

[16] N. Good, J.B. Schafer, J.A. Konstan, A. Borchers, B.M. Sarwar, J.L. Herlocker, J. Riedl, Combining collaborative filtering with personal agents for better recommendations, in: Proceedings of the Sixteenth National Conference on Artificial Intelligence, 1999, pp. 439–446.

[17] U. Hanani, B. Shapira, P. Shoval, Information filtering: overview of issues, research and systems, User Modeling and User–Adapted Interaction 11 (2001) 203–259.

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

[19] F. Herrera, E. Herrera-Viedma, Choice functions and mechanisms for linguistic preference relations, European Journal of Operational Research 120 (2000) 144–161.

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

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

[22] L. Martínez, F. Herrera, 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) (2001) 227–234.

[23] E. Herrera-Viedma, 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) (2001) 460–475.

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

[25] E. Herrera-Viedma, S. Alonso, F. Chiclana, F. Herrera, A consensus model for group decision making with incomplete fuzzy preference relations, IEEE Transactions on Fuzzy Systems 15 (5) (2007) 863–877.

[26] E. Herrera-Viedma, F. Chiclana, F. Herrera, S. Alonso, Group decision making model with incomplete fuzzy preference relations based on additive consistency, IEEE Transactions on Systems, Man and Cybernetics. Part B: Cybernetics 37 (1) (2007) 176–189.

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

[28] E. Herrera-Viedma, F. Herrera, F. Chiclana, A consensus model for multiperson decision making with different preference structures, IEEE Transactions on Systems Man and Cybernetics Part A: Systems and Humans 32 (3) (2002) 394–402.

[29] E. Herrera-Viedma, A.G. López-Herrera, A model of information retrieval system with unbalanced fuzzy linguistic information, International Journal of Intelligent Systems 22 (11) (2007) 1197–1214.

[30] E. Herrera-Viedma, A.G. López-Herrera, M. Luque, C. Porcel, A fuzzy linguistic IRS model based on a 2-tuple fuzzy linguistic approach, International Journal of Uncertainty Fuzziness and Knowledge-Based Systems 15 (2) (2007) 225–250.

[31] E. Herrera-Viedma, A.G. López-Herrera, C. Porcel, Tuning the matching function for a threshold weighting semantics in a linguistic information retrieval system, International Journal of Intelligent Systems 20 (9) (2005) 921–937.

[32] E. Herrera-Viedma, L. Martínez, F. Mata, F. Chiclana, A consensus support system model for group decision-making problems with multi-granular linguistic preference relations, IEEE Transactions on Fuzzy Systems 13 (5) (2005) 644–658.

[33] E. Herrera-Viedma, G. Pasi, A.G. López-Herrera, C. Porcel, 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) (2006) 538–549.

[34] E. Herrera-Viedma, E. Peis, Evaluating the informative quality of documents in SGML-format using fuzzy linguistic techniques based on computing with words, Information Processing and Management 39 (2) (2003) 233–249.

[35] M.H. Hsu, A personalized English learning recommender system for ESL students, Expert Systems with Applications 34 (2008) 683–688.

[36] R.R. Korfhage, Information Storage and Retrieval, Wiley Computer Publishing, New York, 1997.

[37] T.-P. Liang, Recommendation systems for decision support: an editorial introduction, Decision Support Systems 45 (2008) 385–386.

[38] J. Malinowski, T. Weitzel, T. Keim, Decision support for team staffing: an automated relational recommendation approach, Decision Support Systems 45 (2008) 429–447.

[39] G. Marchionini, Research and development in digital libraries. <http://ils.unc.edu/march/digital_library_R_and_D.html>. (accessed 04.09).

[40] L. Martínez, M. Barranco, L.G. Pérez, M. Espinilla, A knowledge based recommender system with multigranular linguistic information, International Journal of Computational Intelligence Systems 1 (3) (2008) 225–236.

[41] L. Martínez, L.G. Pérez, M. Barranco, M. Espinilla, Improving the effectiveness of knowledge based recommender systems using incomplete linguistic preference relations, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 16 (2) (2008) 33–56.

[42] L. Martínez, L.G. Pérez, M. Barranco, A multi-granular linguistic based-content recommendation model, International Journal of Intelligent Systems 22 (5) (2007) 419–434.

[43] A.W. Neumann, Recommender Systems for Information Providers: Designing Customer Centric Paths to Information, Physica-Verlag, Heidelberg, 2009.

[44] G. Pasi, Modeling users' preferences in systems for information access, International Journal of Intelligent Systems 18 (7) (2003) 793–808.

[45] C. Porcel, A.G. López-Herrera, E. Herrera-Viedma, A recommender system for research resources based on fuzzy linguistic modeling, Expert Systems with Applications 36 (2009) 5173–5183.

[46] C. Porcel, J.M. Moreno, E. Herrera-Viedma, A multi-disciplinar recommender system to advice research resources in university digital libraries. Expert Systems with Applications, doi:10.1016/j.eswa.2009.04.038.

[47] L.M. Quiroga, J. Mostafa, An experiment in building profiles in information filtering: the role of context of user relevance feedback, Information Processing and Management 38 (2002) 671–694.

[48] M.E. Renda, U. Straccia, A personalized collaborative digital library environment: a model and an application, Information Processing and Management 41 (2005) 5–21.

[49] P. Reisnick, H.R. Varian, Recommender systems, Special issue of Communications of the ACM 40 (3) (1997) 56–59.

[50] L. Ross, P. Sennyey, The library is dead, long live the library! The practice of academic librarianship and the digital revolution, The Journal of Academic Librarianship 34 (2) (2008) 145–152.

[51] B. Sarwar, G. Karypis, J. Konstan, J. Riedl, Analysis of recommendation algorithms for e-commerce. in: Proceedings of ACM E-Commerce 2000 Conference, 2000, pp. 158–167.

[52] L.A. Zadeh, The concept of a linguistic variable and its applications to approximate reasoning, Part I. Information Sciences 8 (1975) 199–249;
L.A. Zadeh, The concept of a linguistic variable and its applications to approximate reasoning, Part II, Information Sciences 8 (1975) 301–357;
L.A. Zadeh, The concept of a linguistic variable and its applications to approximate reasoning, Part III, Information Sciences 9 (1975) 43–80.