Applied Soft Computing 30 (2015) 778–791

Contents lists available at ScienceDirect

Applied Soft Computing

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

REFORE: A recommender system for researchers based on bibliometrics

A. Tejeda-Lorente [a], C. Porcel [b,*], J. Bernabé-Moreno [a], E. Herrera-Viedma [a]

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

ARTICLE INFO

Article history:
Received 2 July 2014
Received in revised form 9 January 2015
Accepted 16 February 2015
Available online 23 February 2015

Keywords:
Recommender systems
Item quality
Fuzzy linguistic modeling
Digital library

ABSTRACT

Recommender systems (RSs) exploit past behaviors and user similarities to provide personalized recommendations. There are some precedents of usage in academic environments to assist users finding relevant information, based on assumptions about the characteristics of the items and users. Even if quality has already been taken into account as a property of items in previous works, it has never been given a key role in the re-ranking process for both items and users.

In this paper, we present REFORE, a quality-based fuzzy linguistic REcommender system FOr REsearchers. We propose the use of some bibliometric measures as the way to quantify the quality of both items and users without the interaction of experts as well as the use of 2-tuple linguistic approach to describe the linguistic information. The system takes into account the measured quality as the main factor for the re-ranking of the top-N recommendations list in order to point out researchers to the latest and the best papers in their research fileds. To prove the accuracy improvement, we conduct a study involving different recommendation approaches, aiming at measuring their performance gain. The results obtained proved to be satisfactory for the researchers from different departments who took part on the tests.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Research is a key task in society nowadays. Science is going through difficult times at the moment giving the current crisis facing significant budget cuts in modern economies [47]. However research has a direct impact on society, countries that support research can achieve higher Gross Domestic Product (GDP) [17], e.g., 73% of the papers cited by U.S. industry patents are public science, authored at academic, governmental, and other public institutions [48]. Therefore, the use of scientific knowledge by setting up a sustainable industry/science cooperative environment positively affects innovation performance [37].

Nowadays, where the amount of information available is exponentially growing, the information overload represents a real challenge [19] to the academic world [58]. The spectacular growth of sources providing new information introduces noise to our method of finding relevant information for research, but what certain users consider noise could be relevant for others. For example, in the same area of research, a researcher may not be interested in a specific type of publication but another may find this publication very useful. This fact can make users let through suitable information thinking that is not relevant [42]. In the academic world, research is a field of key importance in society, where the knowledge in one single area is vast and quite specific. At present, scientific databases such as Web of Science (WoS) [3], Scopus [2] or Google Scholar [1], support the research in fields where the volume of publications is vast. However, the ever increasing number of resources in a simple area might make the information consumers misunderstand the relevance of certain research resources, at the risk of missing important information.

The number of different journals in WoS is more than 12,000 (120,000 if we speak about conference proceedings) [3]. A researcher is typically focused on more than one category, which might include up to 700 journals potentially harboring valuable information, e.g., a combination of relevant sub-categories of Computer Science, Engineer and Maths. Staying in the loop of all new papers being published in all the new journals could be troublesome and cause problems of information overload.

In this sense, it is essential to have tools that allow researchers to meet this objective: access to updated and personalized information according to our interests. When a researcher tries to obtain useful information related to his/her research through a searching tool, the retrieved results might be irrelevant and contain unnecessary information even after applying the different embedded filters available. Hence, users need easier access to the large amount of resources that are available hidden among the rest [42].

Mostly, in the Web we can find two different tools to facilitate the access to the information: information retrieval systems [34] and recommender systems [25]. The former are focused on information search in a known content repository, scientist repositories mentioned above, while the latter are focused on information discovery in partially known frameworks. A *recommender system* seeks to discover information items (movies, music, books, news, images, web pages, papers, etc.) that are valuable to the user. Recommender systems are especially useful to identify information a user was previously unaware of. Besides that, recommender systems may be considered personalized services because they have an independent profile for each

* Corresponding author.
E-mail addresses: atejeda@decsai.ugr.es (A. Tejeda-Lorente), cporcel@ujaen.es (C. Porcel), juan.bernabe-moreno@webentity.de (J. Bernabé-Moreno), viedma@decsai.ugr.es (E. Herrera-Viedma).

http://dx.doi.org/10.1016/j.asoc.2015.02.024
1568-4946/© 2015 Elsevier B.V. All rights reserved.

user taking into account the particularities of each of them. Due to these reasons, these kinds of systems are becoming popular tools for reducing information overload and to improve the conversion rate in e-commerce web sites [12,31]. Personalized recommendations rely on knowing users characteristics, which might be tastes, preferences about items as well as the ratings of previously explored items. The system has to maintain user profiles updated in order to provide good recommendations. The way of acquiring this information may vary from implicit information, that is, analyzing users behavior, or explicit information, where users directly provide their preferences. In the literature, we can differentiate between two main categories:

1. *Content-based recommender systems*. The recommendations are generated taking into account the items features and the user past experience dealing with similar items.
2. *Collaborative recommender systems*. The recommendations to a user are based on other user recommendations with similar user profiles, taking into account the ratings provided by the users.

All approaches have their advantages and disadvantages, so that a widely used solution is the *hybrid approach*. The hybrid approach consists of a combination of the best features of each approach to minimize the disadvantages while maximizing the benefits. The representation of user preferences as well as other characteristics of some items present subjectivity and uncertainty making the process a complex task. In order to deal with that, fuzzy linguistic modeling has been proved as an efficient tool to face the problem. In the literature we can find different approaches of fuzzy linguistic modeling in the problem of computing with words: *classical*, *type-2*, *ordinal* and *2-tuple*.

In [58] we presented a quality-based recommender system to disseminate information in a University Digital Library (UDL) where we combined the quality of an item with the user preferences to predict which ones are relevant for the user generating more useful and accurate recommendations. Capitalizing on the improvement we reached with our system in the scope of UDL to tackle the information overload issue, we want to expand the scope to the much larger research domain. The heterogeneity of information sources from elements coming from a much larger set of databases is going to be a bigger challenge compared with the work done for the UDL.

Hence, a first step to expand the recommendations in research based on quality is to change not the scope but the database. Instead of using a UDL we are going to use the papers indexed by WoS, which could be considered the most important bibliographic database in a UDL. A second step is to change the old concept of quality applied in [58] and how to evaluate it. With that, we will test a recommender system for researchers and the use of a way to measure the items quality to increase the accuracy of the recommendations. By the use of this new concept, the problem of cold start is reduced since the content based approach along with the filters that take into account the quality comprise a large portion of the final list. The use of a hybrid recommendation scheme proved positive in [58] combining two different approaches. For this reason, we propose a combination of both. On the one hand, content-based, where we are going to deal with the quality of items, and on the other hand, the quality of authors according to the collaborative approach. That is, we are going to compute the relevance of an item and of a user to do a re-ranking of the recommendations. As test users we are going to engage with some members of Departments of Computer Science and Civil Engineer at the University of Granada and Computer Science from TU Delft.

In this paper, we present a new recommender system which incorporates methods, algorithms and filters for all the steps in the recommendation process focused on the re-ranking stage, called quality-based fuzzy linguistic REcommender system FOr REsearchers (REFORE). REFORE allows the user to be up-to-date regarding all articles that might be relevant for him related to the topics the user manifested interest at a particular moment in a real time window. In addition to the articles that are identified to be of interest for the user, a set of high-quality articles with a certain overlap with the researcher topic are also provided to open up new possibilities and therefore enrich the on-going research. Furthermore, the main elements of the system are listed below:

- The system implements a *hybrid recommendation strategy* based on a switching hybrid approach [11], which combine a content-based recommendation approach and a collaborative one to share the user individual experience and social wisdom [56] and to present the best recent papers in each are area for each user.
- The system uses fuzzy linguistic modeling to improve the representation of user preferences and facilitate user-system interactions [8].
- The system implements a two-phase *feedback process*, which can say if a paper is relevant or not, that means evaluate the recommendation, and then when they read the paper they can evaluate their quality giving it different relevance for the collaboration step.
- The resources recommended in our system are now the *most recent research papers* from one of the most important international science database, WoS by Thomson Reuters, where we can find the best journal and conference essentials for the daily work of a researcher. The items recommended fulfill the necessities of the users since the quality and novelty of the items are guaranteed by the system.
- The system incorporates a *re-ranking* process which takes into account the estimated relevance of an item along with the item quality as well as the users quality to re-order the list of possible items to be recommended. This is done taking into account not only these factor but also others to ensure the items are interesting for the researcher but also with a small innovative percentage allowing to see the different application of one singular topic.
- The method of measuring quality in previous proposals [13,14,58] implies users interaction and is based on the feedback given by them. So, we propose a new way to evaluate the quality of research resources and researchers based on the different rankings provides in the international community by experts, in our case we use the Journal Citation Report (JCR) provided by Thomson Reuters[1] for the items, and also the h-index [5] for the quality of the users given that the system is developed for researcher.

This system improves the generated recommendations, by including more quality papers and accurate recommendations, increasing its information discovering properties in the recommendations and updating the users with the newest papers of each topic being the most relevant for the users.

The paper is structured as follows. In Section 2, the background is presented, that is the basis of recommender systems, the fuzzy linguistic modeling to represent information and the aspects of quality evaluation in research environment. Next in Section 3, we describe REFORE together with its main characteristics, functionalities, architecture and technologies used. Section 4 presents the evaluation of the system and the experimental results. Finally, some concluding remarks are pointed out in Section 5.

2. Background

The purpose of this section is to provide the background information needed to describe our system. It is divided in three parts: an introduction to recommender systems, a description of the fuzzy linguistic approach followed and the evaluation of quality in research resources and authors.

2.1. Basis of recommender systems

Recommender systems provide personalized recommendations to help users in their elections, guiding the user in a personalized way toward suitable tasks among a wide range of possible options [12]. Even without detailed product domain knowledge, RSs have proven to be very useful providing personalized items suggestions to users. Well known examples of successful use of RSs are present in e-commerce [15,36], health [20,32], learning [31,38], etc. The objective of RSs is to supply each user with the most relevant information for them [25]. Therefore, the is a personalized recommendation [12].

Building users profiles, where their preferences are reflected, is a key task in RSs. In fact, the performance of the system depends on its capacity to mirror users preferences. Hence, construction of profile determines the success of the system. Personal preferences in all all variety of forms can be taken into account depending on the topic of the RS, for example, history of purchases or demographic information. Two ways of obtaining the information about user preferences are identified [25]. On the one hand, an *implicit approach* where the user preferences are updated by detecting changes while observing the user. On the other hand, an *explicit approach* where the users express some specifications of what they desire and ratings about the explored items.

On the design of a RSs an aspect to take into account is the way of generating recommendations. In the literature the recommendation approaches can be mainly grouped into two categories. In the first one [24,25,50] authors consider two different approaches: On one side, *Content-based systems* generate the recommendations taking into account the characteristics used to represent the items and the ratings that a user has given to them [9,16]. On the other side, *Collaborative systems* provide recommendations using explicit

[1] http://thomsonreuters.com/journal-citation-reports/.

or implicit preferences from many users, ignoring the items representation. The second comprises three approaches: *Demographic systems*, *Knowledge-based systems* and *Utility-based systems* [11,12].

Since each approach has certain advantages and disadvantages, depending on the particular setup, the most adopted solution addressed in the literature is the combination of the previous approaches in what is known as an *hybrid recommender system* [11]. The aim of this hybrid approach is to combine different approaches to reduce the disadvantages of each one and to exploit their benefits. It is proved that the use of an hybrid strategy provides more accurate recommendations [9,16,24]. Burke proposed a survey about the possible strategies used to obtain hybrid recommendations. Among the proposed strategies to combine the recommendation approaches we will pay special attention to the *mixed* one, where the recommendations from more than one technique are presented together [11].

In this work, our purpose is to focus on the recommendation of papers. Since the novelty of the items is strongly constraint in the goal of this system, collaborative approach is a technique that cannot rule the process of recommendation. The use of the *mixed hybrid strategy* is interesting for our purpose, i.e., finding the balance between novelty and importance of non recommended items that other users consider important in a time window. Due to the fact that the items can be recommended when they are inserted into the system in a determined time, the use of collaborative filters is not suitable. This technique is used in a lower number in the following round to recommend possible relevant paper that the system could not recommend due to the restriction of the number of recommendations per user.

The recommendation activity is completed with two more phases: *re-ranking* and *feedback*. The former is oriented to increase the accuracy of recommendations. In [4], authors explain re-ranking as a method to increase diversity in recommendations minimizing the accuracy loss. They assume that the predicted order given by the system would be the optimums for each user based on the *probability ranking principle* in information retrieval systems that ranks the documents in order of decreasing probability of relevance [54]. The re-ranking involves a different criteria to order the items in the final list of recommendation, i.e., popularity, average rating, likeability. Counter to this objective, in this work the use of re-ranking is oriented not to increase the diversity but the accuracy. In the latter, feedback is a cyclic process whereby the users provide the system with their evaluations about the recommended items and the system uses these evaluations to automatically update user profiles [25].

2.2. Fuzzy linguistic approach

In this section we describe the 2-tuple fuzzy linguistic modeling and multi-granular fuzzy linguistic approach used to represent the linguistic information used in REFORE as well as other techniques used in the modeling of information. In some situations, the information cannot be precisely appraised in a quantitative manner but can be qualitatively evaluated. In the literature we can find countless pieces of research about different techniques of the fuzzy modeling in the problem of computing with words and some application of them to model information in recommender systems [10,51,58]. The main methods for managing qualitative information are [28]:

- The *classical* linguistic approach, based on the use of labels whose semantics is represented by means of fuzzy sets and their associated membership functions. The aggregation over a set of linguistic labels usually does not have an associated linguistic label on the term set. Therefore, an approximation function that introduces some loss of information is needed, e.g., *round()*, that is, the retranslation problem [39,62].
- The *type-2* linguistic approach, based on the use of type-2 fuzzy sets to model linguistics assessment to maintain the uncertainty [43,44,59,60], note that this approach also suffers from the retranslation problem.
- The *ordinal* linguistic approach, based on the use of labels whose semantics is established on an ordered structure defined on the labels. The semantics of the linguistic labels in an ordinal linguistic approach is established considering the labels are symmetrically and uniformly distributed around the central assessment in the set of linguistic terms.
- The 2-tuple linguistic symbolic model is the most widely used to manage ordinal linguistic information. This model was introduced in [29] to avoid the loss of information since the linguistic domain can be treated as continuous. One of the most important parameter to determine in any fuzzy linguistic approach 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/her knowledge will have more or less terms [46]. We base our proposal in this approach.

2.2.1. The 2-tuple fuzzy linguistic approach

Usage of the fuzzy linguistic methods mentioned above results in loss of information. In order to reduce it, in [29] its proposed a continuous model of information representation based on 2-tuple fuzzy linguistic modeling. To define both the 2-tuple representation model and the 2-tuple computational model to represent and aggregate the linguistic information have to be established.

Let $S = \{s_0, \ldots, s_g\}$ be a linguistic term set with odd cardinality, where the mid term represents an indifference value and the rest of the terms are symmetric related to it. For instance, we could use the following set of terms with 5 labels: $S = \{LE; L; N; H; HE\}$, where $s_0 = LE = Lowest$, $s_1 = L = Low$, $s_2 = N = Normal$, $s_3 = H = High$ and $s_4 = HE = Highest$.

We assume that the semantics of labels is given by means of fuzzy subsets defined in the [0, 1] interval, which are described by their membership functions $\mu_{s_i} : [0, 1] \rightarrow [0, 1]$, and we consider all terms distributed on a scale on which a total order is defined, that is, $s_i \leq s_j \Leftrightarrow i \leq j$. We consider that linear triangular membership functions are good enough to capture the vagueness of those linguistic assessments, since it may be impossible or unnecessary to obtain more accurate values. This representation is achieved by the 3-tuple (a, b, c), where a is the point where the membership is 1 and b and c are the left and right limits of the definition domain of the triangular membership function. For example, the following semantics, represented in Fig. 1, can be assigned to a set of five terms via triangular membership functions:

$$HE = Highest = (1, 0.75, 1) \qquad H = High = (0.75, 0.5, 1)$$
$$N = Normal = (0.5, 0.25, 0.75) \qquad L = Low = (0, 0.25, 0.5)$$
$$LE = Lowest = (0, 0, 0.25)$$

In this fuzzy linguistic context, if a symbolic method aggregating linguistic information obtains a value $\beta \in [0, g]$, and $\beta \notin \{0, \ldots, g\}$, then an approximation function is used to express the result in S.

Definition 1. [29] 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:

- s_i represents the linguistic label of the information, and

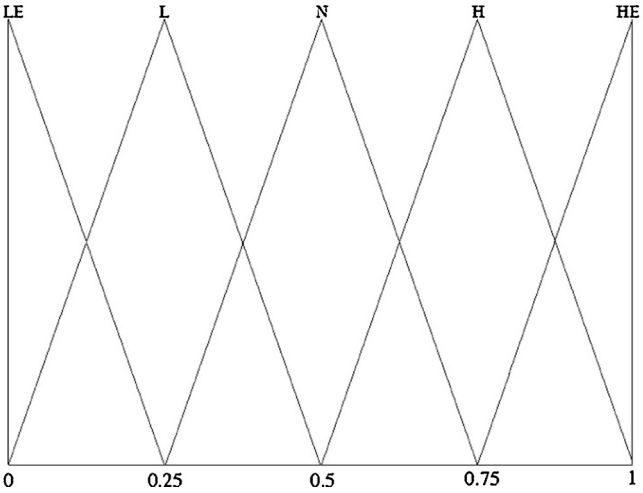

Fig. 1. A set of seven linguistic terms with its semantics.

- α_i is a numerical value expressing the value of the symbolic 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. [29] Let $S = \{s_0, \ldots, s_g\}$ be a linguistic term set and $\beta \in [0, g]$ a value representing the result of a symbolic aggregation operation, then the 2-tuple that expresses the equivalent information to β is obtained with the following function:

$$\Delta : [0, g] \longrightarrow S \times [-0.5, 0.5) \tag{1}$$

$$\Delta(\beta) = (s_i, \alpha), \quad with \quad \begin{cases} s_i & i = round(\beta) \\ \alpha = \beta - i & \alpha \in [-.5, .5) \end{cases} \tag{2}$$

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 $\Delta^{-1}(s_i, \alpha) = i + \alpha$. Moreover, 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 [7,8,29]:

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 vice versa, any of the existing aggregation operators can be easily extended to deal with linguistic 2-tuples. Some examples are:

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

$$\bar{x}^e[(r_1, \alpha_1), \ldots, (r_n, \alpha_n)] = \Delta\left(\sum_{i=1}^{n} \frac{1}{n} \Delta^{-1}(r_i, \alpha_i)\right) = \Delta\left(\frac{1}{n}\sum_{i=1}^{n} \beta_i\right). \tag{3}$$

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

$$\bar{x}^w[(r_1, \alpha_1), \ldots, (r_n, \alpha_n)] = \Delta\left(\frac{\sum_{i=1}^{n} \Delta^{-1}(r_i, \alpha_i) \cdot w_i}{\sum_{i=1}^{n} w_i}\right) = \Delta\left(\frac{\sum_{i=1}^{n} \beta_i \cdot w_i}{\sum_{i=1}^{n} w_i}\right). \tag{4}$$

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

$$\bar{x}_l^w[((r_1, \alpha_1), (w_1, \alpha_1^w)) \ldots ((r_n, \alpha_n), (w_n, \alpha_n^w))] = \Delta\left(\frac{\sum_{i=1}^{n} \beta_i \cdot \beta_{W_i}}{\sum_{i=1}^{n} \beta_{W_i}}\right), \tag{5}$$

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

2.2.2. Linguistic hierarchy to model multi-granular linguistic information

The problem of modeling the information arises when different experts have different uncertainty degrees on the phenomenon, then several linguistic term sets with a different granularity of uncertainty are necessary [46]. The use of different label sets to assess information is also necessary when an expert has to evaluate different concepts. In such situations, we need tools to manage multi-granular linguistic information [30,41].

In [30] the concept of linguistic hierarchy and a multi-granular fuzzy linguistic modeling based on a 2-tuple fuzzy linguistic approach was 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 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: $l(t, n(t)) \rightarrow l(t+1, 2 \cdot n(t) - 1)$.

Using this *LH*, the linguistic terms in each level are the following:

- $S^2 = \{a_0 = None = N, a_1 = Total = T\}$.
- $S^3 = \{b_0 = None = N, b_1 = Medium = M, b_2 = Total = T\}$
- $S^5 = \{c_0 = Lowest = LE,\ c_1 = Low = L,\ c_2 = Normal = N,\ c_3 = High = H, c_4 = Highest = HE\}$
- $S^9 = \{d_0 = None = N,\ d_1 = Very\_Low = VL,\ d_2 = Low = L,\ d_3 = More\_Less\_Low = MLL,\ d_4 = Medium = M,\ d_5 = More\_Less\_High = MLH, d_6 = High = H, d_7 = Very\_High = VH, d_8 = Total = T\}$

A graphical example of a linguistic hierarchy is shown in Fig. 2.

In order not to lose information in the way of representing multi-granular linguistic information through linguistic hierarchies, in [30] a family of transformation functions was presented among labels from different levels to combine them.

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

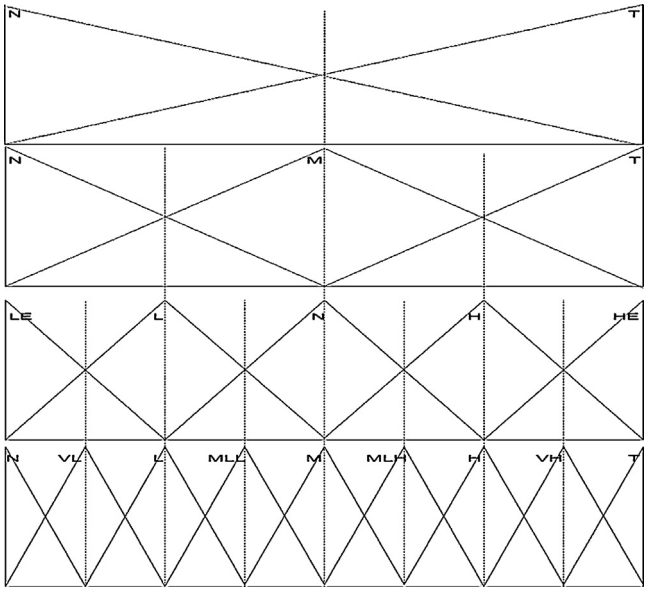

Fig. 2. Linguistic hierarchy of 2, 3, 5 and 9 labels.

transformation function between a 2-tuple that belongs to level t and another 2-tuple in level $t' \neq t$ is defined as:

$$TF_{t'}^{t} : l(t, n(t)) \longrightarrow l(t', n(t')) \tag{6}$$

$$TF_{t'}^{t}(s_i^{n(t)}, \alpha^{n(t)}) = \Delta\left(\frac{\Delta^{-1}(s_i^{n(t)}, \alpha^{n(t)}) \cdot (n(t') - 1)}{n(t) - 1}\right) \tag{7}$$

As it was pointed out in [30] this family of transformation functions is bijective. This result guarantees the transformation between levels of a linguistic hierarchy is carried out without loss of information. To define the computational model, we select a level to equalize the information (for instance, the highest granularity level) and then we can use the operators defined in the 2-tuple fuzzy linguistic approach.

2.3. On evaluation of quality of research resources and authors

Research resources is a data set with a continuous growing development and change. The proliferation of different scientific repositories provides access to this research resources gathering the most important database in different areas: e.g., Scopus[2] and WoS [3]. All of this repositories provide different search tools to find the information sought.

To measure the quality of a scientific publication is a key task for our purpose; for that reason we assume some bibliometric indicators which are commonly employed in different fields for the quality measure of them [18]. The most widely adopted method to this task is to use Garfield's Impact Factor (IF): the average number of times the published papers are cited up to two years after publication [22]. Despite IF having been criticized for its only dependency on citation counts [55], impacts of publications, journal or conferences are usually the major consideration for subscription or for submission decision-making [63]. Other metrics have been proposed to rank science journals or conferences [33,49,52].

Based on citations we can find also in the literature metrics more oriented to measure the quality of authors. Even though, we can consider that the quality of a paper is directly related with the IF of the journal where it is published, for authors there are different measures that allow us to analyze them individually. Some of the most important indicators are the *h-classics* [40].

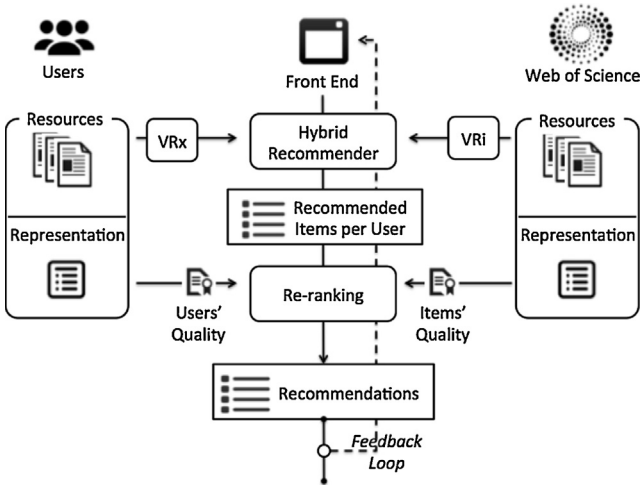

Fig. 3. Structure of the system.

As we can see, the measures of quality in the research world is a wide studied field. The measure is oriented to obtain different ranks, but very few studies actually apply the existing rankings as criteria to recommend resources. Our proposal tries to fill the gap in this area.

3. REFORE

REFORE[2] is a new platform which combines a hybrid recommender system which takes into account users and items quality when it comes to the re-ranking stage. Besides that, the system provides a Web page so that users are able to manage their information and recommendations. The system delivers personalized emails monthly to each user with a selection of the best latest papers for him. It can be used by the research community to facilitate the management of information overload at the routine investigation, as the registration and use is for free.

REFORE allows us to manage the problem of information overload in a research environment. Although this approach was originally developed to deal with UDL, we extended it in REFORE managing bigger databases and focusing more on the relevance of the quality of the items and users when working with a recommender system.

Now, we change our focus to researchers specifically and we move from UDL to an open world where all researchers can use the system based on the international databases of the best and ranked resources. We change the main idea of the system to create awareness of the newer elements among users and not only the items that are already stored in the database. For this purpose we set an alert system which sends emails each time period previously established with the best items in the system since the latest round, keeping all users up to date with the last trends in their interest topics.

The basics steps to build REFORE are shown in Fig. 3 which represents the structure of the system and can be seen below:

1. Building an administration and informative Web for users and administrator to manage the information in an easy way.
2. Acquiring items from the WoS database on a monthly basis. We obtained the meta data exposed by WoS when you download a paper, to do that we used the tool presented in [6]. We can highlight the following data: *Title, abstract, authors, keywords, number*

[2] Accessible in: http://sci2s.ugr.es/sapluweb/refore.

of cites, journal, bibliography, type, topic or information about a journal: *quartil, impact factor, cites, topic*. The system represents internally the items based on the built database.
3. Acquiring users profile. We obtained the profile for users in two steps, users papers and users interest.
4. Building the hybrid recommendation process based on the users and items quality.
5. Setting up the re-ranking process focus on the quality measure established. The system aggregates the estimated relevance of an item and a user and its quality score in a single score as well as different filter to improve the accuracy of recommendations.
6. Sending emails to users with their recommendations each month asking for their feedback.

As aforementioned, we work with a multi-granular fuzzy linguistic approach [30] to model the user-system communication 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., $S_i \in LH$. 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, based on the paper keywords, which is assessed in S_1.
- **Similarity degree** among resources or among users, which is assessed in S_2.
- Predicted **relevance degree** of a resource for a user, which is assessed in S_3.
- **Satisfaction degree** expressed by a user to evaluate the recommendation of one resource, which is assessed in S_4.
- **Quality degree** expressed by a user to evaluate a recommended resource, which is assessed in S_5.

The selected granularity must be sufficiently low to avoid imposing an excessive precision in the information you want to express and high enough to get a discrimination of the assessments in a limited number of degrees. Usually, the cardinality used in the linguistic models is an odd value, such as 7 or 9, not exceeding 11 labels. These classical values are based on Miller's observation line about human capacity [45], which indicated that an average person can handle and remember about 7 or 9 terms. We make an exception to represent if a recommendation is relevant or not for a user, using 2 terms.

Following the linguistic hierarchy shown in Fig. 2, we use the level 4 (9 labels) to represent the similarity degrees and predicted relevance degrees ($S_2 = S^9$ and $S_3 = S^9$), the level 3 (5 labels) are used for the importance degrees ($S_1 = S^5$) the level 2 (3 labels) to represent the quality degree given by a user over a paper ($S_5 = S^3$) and for the satisfaction degree we use the level 1 ($S_4 = S^2$). As the importance degrees are extracted automatically from a paper, all of them take the intermediate label of S^5 to facilitate them the characterization of papers. The similarity and relevance degrees are computed automatically by the system, we use the set of 9 labels which presents an adequate granularity level to represent the results. Using this *LH*, the linguistic terms in each level are the ones given below:

- $S^2 = \{a_0 = None = N,\ a_1 = Total = T\}$.
- $S^3 = \{b_0 = None = N,\ b_1 = Medium = M,\ b_2 = Total = T\}$
- $S^5 = \{c_0 = Lowest = LE,\ c_1 = Low = L,\ c_2 = Normal = N,\ c_3 = High = H,\ c_4 = Highest = HE\}$
- $S^9 = \{d_0 = None = N,\ d_1 = Very\_Low = VL,\ d_2 = Low = L,\ d_3 = More\_Less\_Low = MLL,\ d_4 = Medium = M,\ d_5 = More\_Less\_High = MLH,\ d_6 = High = H,\ d_7 = Very\_High = VH,\ d_8 = Total = T\}$

In the following subsections, we describe the REFORE recommender system. Firstly, the structure of items and users is analyzed in detail. Secondly, the architecture of REFORE and the description of the recommendation process with the re-ranking steps are described.

3.1. Resources representation

The considered resources are papers imported from WoS. The journals are selected from the subcategories of the JCR of ISI WoS. In the Table 1 we can see the characteristics we use to represent an item, i.e., a paper (where all the characteristics stored match with the name given to the different properties, i.e., *Title, Authors, Journal, Year* and *time cited* among others). Additionally, we consider extra characteristics at journal level, like as the Quartile where the journal is indexed in JCR, the impact factor, eigenfactor or characteristics referred to the authors, such as city, email, university, etc.

Once the system has imported all available information about a resource, an internal representation based on the keywords of the paper itself is created. We use the *vector model* [34] to represent the resource structure and the classification of keywords; i.e., a paper *i* is represented as a vector:

$$VR_i = (VR_{i1}, VR_{i2}, \ldots, VR_{in}),$$

where each component $VR_{ij} \in S_1$ is a linguistic assessment of the importance degree of the keywords. Due to the singularity of the keywords in an article, all the keywords are set with the label c_4 of S^5 that corresponds to the highest importance – being that by the own definition of a keyword. These importance degrees are assigned automatically by REFORE when new resources are imported. In the event where a paper does not have any keywords, the system takes the internal keywords assigned by WoS.

These vectors will experiment modifications once the re-ranking process is run based on the quality of each them, as well as the final score of each singular recommendation.

3.2. User profiles representation

The *profiles* used in REFORE are based on the characteristics of users items stored. The properties of items are generated based on the information extracted from WoS, that is, papers from proceedings and international journals. Users profiles are based on their own papers and papers they are interested in, along with keywords and their importance leading the users interests.

The introduction of the users preferences is usually a problem due to the effort required to provide them. To reduce this effort and to make easier the process of acquiring the user preferences, REFORE tries to build an automatic profile based on the users papers. The idea of this first step is just to facilitate users following steps about indicating their actual preferences based on their papers inasmuch as the preferences of researchers can be different in each time period. To do that, REFORE automatically tries to identify the resources where the user appears as author or co-author performing a search in WoS *papers* to then import them into the system. The five most important ones according to the quality estimated are then selected. The way of calculating the quality is explained below. After that, from the five selected papers of a user, the *keywords* are extracted and flagged as important keywords for this user with the label c_2 of S^5 pending for the user revision.

Table 1
Papers structure

Paper	Title	Authors	Journal	Journal metadata	DOI	Authors Data	Scope	Year	Time_cited	Type

The first time users log into REFORE, they are asked to complete / change / enhance their profiles, showing them their papers in a pre-established order. Users can upload more papers to the system by their DOI or entering manually their preferences in the fields available for this purpose. It's possible for them changing the pre-selected 5 papers, which summarizes their current interest, as we can see in Fig. 4.

Furthermore, the most important part of users profile are the keywords that determine the users preferences at any time. In Fig. 5 we see how users can set their actual preferences based on labels of S_1 as well as other personal data.

In order to obtain **user preferences, i.e., user preference vector**, we represent a user i as a vector:

$$VU_i = (VU_{i1}, VU_{i2}, \ldots, VU_{in}),$$

where each component $VU_{ij} \in S_1$ is a linguistic assessment that represents the importance degree of keyword j for the user i. Firstly, the system builds the vector with the keywords from the user profile and their labels, κ_k. Secondly the system aggregates these previous keywords and their labels with the keywords of the user's profile papers, that is κ_p, going trough the five papers selected by the users. Hence, these keywords also represent their preferences. The aggregation consists in a union of both domain of keyword where the labels established by the user prevail in the new set when some one is repeated. If the keyword is already in the keyword list, the system maintains the label given by the user specifically to the keyword. Otherwise that means that the keyword is not yet in the vector, therefore it is added with the label c_1. Hence, all the keywords in κ_p has c_1 label. So that, the user vector is represented as follows:

$$VU_i = Aggregation(K_{ki}, K_{pi}) \begin{cases} VU_{ij} = K_{kij}, & \text{if} \exists K_{kij} \in K_{pij} \\ VU_{ij} = K_{pij}, & otherwise \end{cases} \quad (8)$$

where K_{kij} is the keyword j for a user $i \in \kappa_k$ and K_{pij} is the keyword j for a user $i \in \kappa_p$.

3.3. Hybrid recommendation approach

In this phase the system filters the information stored and applies different steps to fit items and user in the delivery. As stated before, we implement a *hybrid recommendation* strategy, which combines a content-based recommendation approach and a collaborative one. According to the classification composed by different strategies given by Burke [11] on how the different techniques are combined, our proposal is based on a *mixed hybrid approach*. Thus, REFORE relies on recommendations from both recommender approaches following some criteria. The system uses the content-based and the collaborative one cooperating both in different percentage. The combination of both occurs after the re-ranking phase.

The process of filtering is based on a matching process developed by similarity measures, such as Euclidean Distance or Cosine

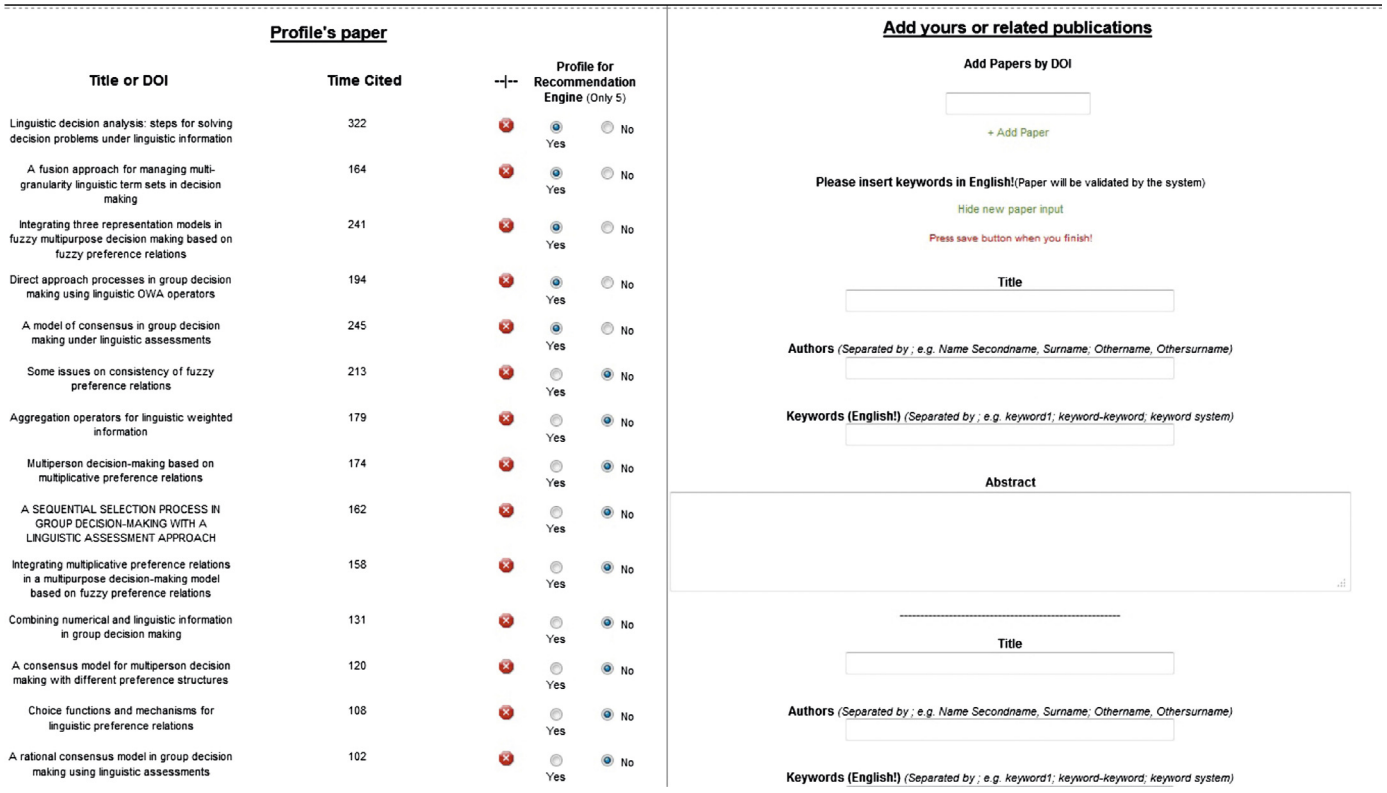

Fig. 4. User papers.

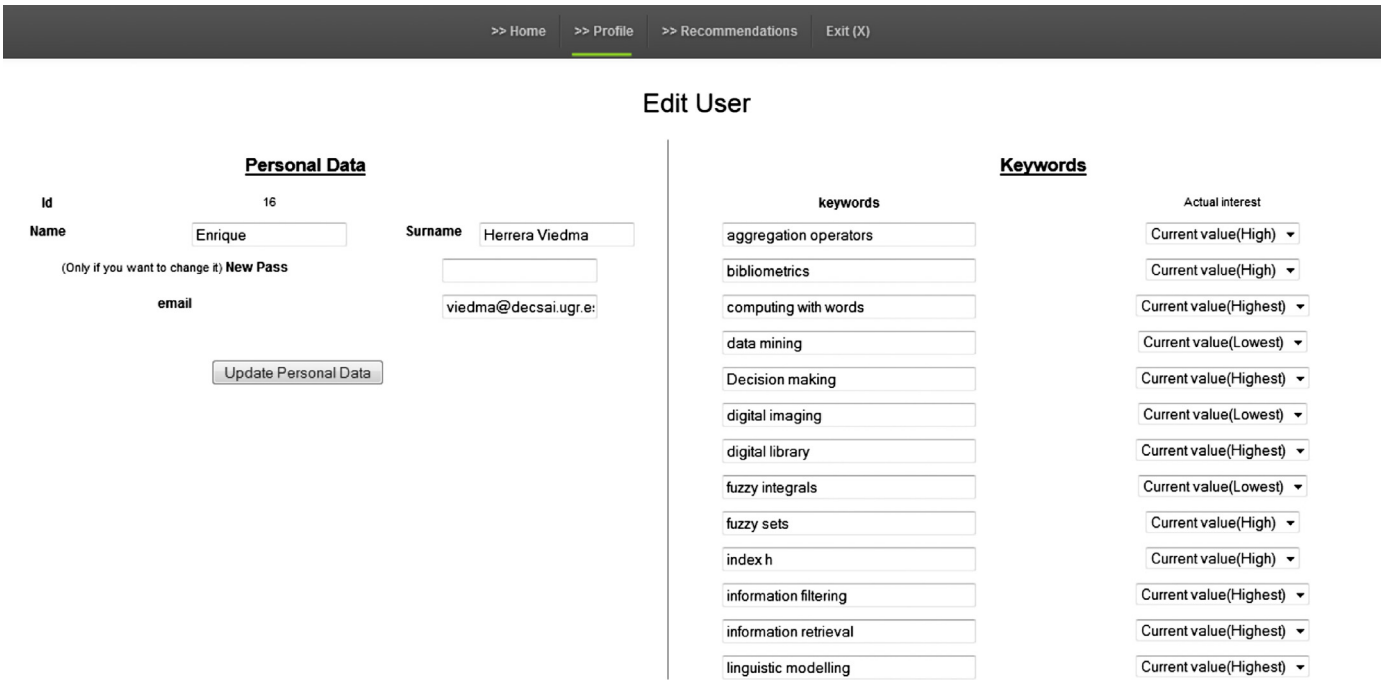

Fig. 5. User preferences.

Measure [34]. In particular, we use the standard Cosine measure but defined in a linguistic framework:

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

with $\sigma_l(V_1, V_2) \in S_2 \times [-0.5, 0.5]$, and where g is the granularity of the term set used to express the relevance degree, i.e., S^9, n is the number of keywords and (v_{ik}, α_{vik}) is the 2-tuple linguistic value of keyword k in the vector V_i representing the resource scope or user interest topics, depending of the used filtering strategy.

In the following sections, we explain both recommendation strategies.

3.3.1. Content-based recommendations

When a new resource i is imported to the system and the time window of recommendations starts, the system computes a first content-based list of candidates to recommendations for a researcher e as follows:

1. Compute the linguistic similarity degree between VR_i and VU_e: $\sigma_l(VR_i, VU_e) \in S_2$.
2. Assuming that $S_2 = S^9$, we consider that a resource i is relevant for a researcher e if $\sigma_l(VR_i, VU_e) > (d_3, 0)$, i.e., if the linguistic similarity degree is higher than the label one step under the mid linguistic label.

All the relevant resources for a researcher e are part of the list of candidates with a relevance degree equal to linguistic similarity degree obtained in the first step.

3.3.2. Collaborative recommendations

Users in the system have set up their profiles with the information extracted from their papers or the ones they are interested in and the set of keywords which captures the preferences for them. When the system is already operating and there are at least one evaluation from other users, that means after the first round, we follow a memory-based algorithm or nearest-neighbor algorithm, to detect similar users. This algorithm generates the recommendations according to the items already evaluated of the nearest neighbors. This algorithm has been proven to provide good results [26]. In the following steps we describe the process in detail.

Given a researcher e, the recommendations to be added to the previous list of e are obtained in the following steps:

1. Identify the set of users $\aleph_e$ most similar to the user e that we are estimating in the round t, where each round is the established period between recommendations. For doing this, we calculate the linguistic similarity degree between the user preference vector of the user (VU_e) upon the vectors of all other users VU_y, $y = 1, \ldots$, where n is the number of users, that is, we calculate $\sigma_l(V_e, V_y) \in S_2$. As $S_2 = S^9$, we consider that the user y is near neighbor to e if $\sigma_l(VU_e, VU_y) > (d_4, 0)$, i.e., if the linguistic similarity degree is higher than the mid linguistic label.
2. Look for the resources stored in the system that were previously relevant to the near neighbors of e at most, two round before (t_{-1}, t_{-2}). That is, to not to break the novelty property of the recommended items, i.e., the set of resources $\kappa_y = \{1, \ldots, k\}$ such that there is a linguistic satisfaction assessment $y(j) \in S_4$, $y \in \aleph_e$, $j \in \kappa_y$, and $y(j) \geq (a_1, 0)$.
3. Due to the singularity of the problem we are tackling, researchers can have different lines of investigation. In order to avoid the problem of recommended items which are not valid for the user even if they are relevant from near neighbors, the system calculates the similarity for all the items of the last two round for all the near neighbors as described below:
   1 Compute the linguistic similarity degree between the relevant resources obtained in the step 2, VR_{ij} and VU_e: $\sigma_l(VR_{ij}, VU_e) \in S_2$, where $j \in \aleph_e$ and $i \in (t_{-1}, t_{-2})$.
   2 Assuming that $S_2 = S^9$, we consider that a resource i is relevant for a researcher e if $\sigma_l(VR_i, VU_e) > (d_3, 0)$, i.e., if the linguistic similarity degree is higher than the label one step under the mid linguistic label.

The set of relevant resources for a researcher e is part of the list of candidates with a relevance degree of $\sigma_l(VU_e, VU_y)$.

3.4. Computing the quality of research resources and users

As aforementioned, one of the most important components of the problem we are facing in this article is the novelty of the items. The resources we are dealing with in this problem are specific enough to apply the set of different measures existing in the literature for the academic world to estimate the quality of an item. So we propose to split the quality measure in two elements: papers and authors.

The former can be measure by the publishing entity, using the impact factor (IF) of the journal as well as the quartile of JCR index they are ranked in. Furthermore, the feedback given by users about papers is also taken into account to estimate the quality of an item. The final measure is the aggregation of both measures. However, as a result of the novelty characteristic of the recommended items and due to constraint we imposed of two rounds as maximum time elapsed between possible recommendable items, the metric which contributes the most to the quality estimation of an item is the IF.

Meanwhile, the latter can be measured by the h-index [5] of authors which quantifies the importance of each author within the research community.

The main reason for adopting this approach revolves around the fact that both the information of the quality of the journal and the impact factor where the paper was published are public and stored in REFORE. The users h-index is maintained up to date due to the users profile are updated each round updating the time their articles have been cited.

The problem of estimating the quality of a **resource** based on the IF of a journal is the explained in the lines below:

1. Suppose that an item belong to the best journal in one field, i.e., *Medicine*.
2. This item share some keywords with one user and fulfill the similarity degree, i.e., *Decision Making*.
3. The item under analysis deals with decision making problems but in a field where the context of the same keywords is different.
4. The item score is increased due to the quality of the journal and later recommended to the user.

In order to overcome this problem, the quality of each item i is estimated individually for each user j.

$$q(i,j) = \left(\frac{1}{Q_i} + IF_i\right) * G_j \tag{9}$$

where, we normalize the $q(i,j)$ to the range [0, 1], Q_i is the quartile where the journal i is ranked. IF_i is its impact factor of the journal i and G_j is a constant which measures how appropriate a journal for a user j is, adopting G_j the following values:

$$G_ij = \begin{cases} 1, & \text{if } \exists j \in Js_i \\ 0.75, & \text{if } \exists T(j) \in Ts_i \\ 0.5, & \text{if } \exists j \in Jns_i \\ 0.25, & \text{if } \exists T(j) \in Tns_i \\ 0.1, & \text{otherwise} \end{cases} \tag{10}$$

where Js_i is the set of journals where the selected papers of the user i are published, $T(j)$ is the function which extracts the topics of a journal j, Ts_i is the set of topics from selected journals of the user i, Jns_i are the journals non selected but part of the profile of the user i and Tns_i is the set of topics from the non selected journal of the user i.

The quality of an **author** despite is independent from the rest of qualities, is weighted with the qualities of similar authors to the user we are calculating. Before of that, they are normalized to the range [0, 1]. Therefore, the set of qualities Φ_a consist on the weighted qualities of the similar authors to j. So, the quality of an author i regarding to the user j is:

$$q_a(i,j) = H(i)/\max(Q_a(j)) \tag{11}$$

where $H(i)$ is the h-index of the author i and $Q_a(j)$ is the set with the h-index values for the authors that are near neighbor to j and $max()$ is a function to obtain the maximum element of a set.

In spite of the before mentioned procedure, REFORE does not recommend authors, just resources. The quality of the authors is used to measure how good a resource for a user is, when we are dealing with the collaborative approach.

In order to measure the quality of a resource, REFORE also leverages the feedback provided by users once they have read the paper. The rating of a resource e given by a user i is used as the quality of the resource for the system in the next rounds. The quality y is expressed in S_5, assuming that $S_5 = S^3$, so we consider that a resource i is rated with the certain quality score y for the researcher profile e.

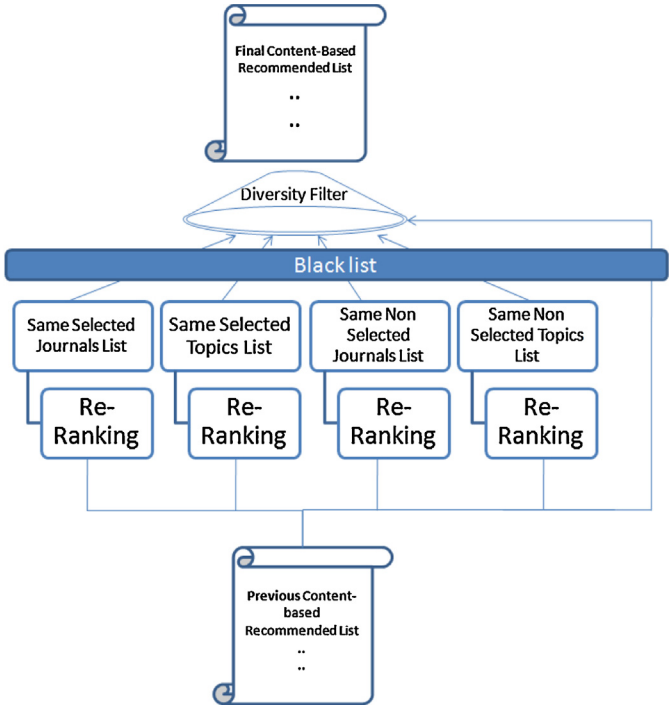

Fig. 6. Re-ranking on content-based previous list.

3.5. Re-ranking

Once the previous lists of resources have been generated, we proceed with the re-ranking process. The re-ranking phase in both approaches is independent from each other and consists of the reorganization of the previous lists aggregating the linguistic similarities with the scores of quality as we indicate below.

On the one hand we have the re-ranking of the *content-based* previous list whose process diagram is showed in Fig. 6. The previous lists of resources $\aleph_i$ that are considered relevant for a user e, with the estimated relevance degree of each resource i for e, $i(e) \in S_3$, where $S_3 = S^9$ are estimated as explained in Section 3.3.1.

The computation of the quality for the resources mentioned above is also used to support the classification of each resource i in different groups based on the selected papers by the users in their profiles. Hence, we split the previous list into four smaller lists:

1. *Same selected journals list*. Papers from the same journals of the group of the selected one for the user *e*.
2. *Same selected topics list*. Papers from the journals with the same topic of the group of the selected one for the user *e*.
3. *Same non-selected journals list*. Papers from the rest of the journals for the papers of the group of the non-selected one for the user *e*.
4. *Same non-selected topics list*. Papers from the rest of the journals with the same topic of the group of the non-selected one for the user *e*.

The aggregation of the quality scores and the relevance degree for each one leave us four candidates lists for the content-based. To do this, first we need to translate the research resource quality score into the values range in which the estimated relevance degree is defined, i.e., S_3. Then, we obtain following translated quality score, $tq(i)$ [58], as follows:

$$tq(i) = q(i) \times g \tag{12}$$

where g is the granularity of S_3, assuming $S_3 = S^9$ then $g = 8$.

In the next steps, we use a multiplicative aggregation in which the estimated relevance is multiplied by the translated quality score, as follows:

$$FinalRelevance(i) = \Delta\left(\frac{\Delta^{-1}(i(e) \times tq(i))}{g}\right) \tag{13}$$

where Δ and Δ^{-1} are the transformation functions between 2-tuples values and symbolic values defined in Section 2.2.1. To obtain the final relevance degree in S_3, we map the final relevance value to the interval $[0, g]$. The foregoing operation lets us four ordered lists of items with different characteristics.

On the other hand the re-ranking of the *collaborative* previous list takes place as follows. The previous lists of resources from similar authors $\Im_i$ that are considered relevant for a user e with the estimated relevance degree of each resource i for e, $i(e) \in S_3$, where $S_3 = S^9$ are estimated as explained in Section 3.3.2.

The computation of the quality for the resources mentioned above is conditioned by two factors, quality of near neighboring authors and rating of resources relevant for those authors. Therefore, before aggregating quality scores and relevance degree as in the content-based approach, the final quality score must be expressed as a unique one.

To do that, we map the author quality score to values in range in which the estimated degree is defined, i.e., S_3. We obtain the translated quality score $tq(i)$ as before. As the linguistic level of the quality of the author is lower than the relevance of an item, before doing a multiplicative aggregation of all the values we change to the granularity of S_3.

Then, the multiplicative aggregation in which the estimated relevance is multiplied by the mapped quality score and by the translated author quality, is calculated as follows:

$$FinalRelevance(i) = \Delta\left(\frac{\Delta^{-1}(i(e) \times tq(i)) \times \Delta^{-1}(i_a(e))}{g}\right) \tag{14}$$

where Δ and Δ^{-1} are the transformation functions between 2-tuples values and symbolic values defined in Section 2.2.1. To obtain the final relevance degree in S_3, we translate the final relevance value to the interval $[0, g]$. The foregoing operation let us a relevance ordered list of items that where previously relevant for other users.

Before obtaining the final list or applying the diversity filter, a last step which affects the final relevance degree is applied. The black list of the combination of keywords of each user where the feedback provided was negative is compared with the keywords of each list. If the combination is found on one paper the relevance degree is penalized to half value.

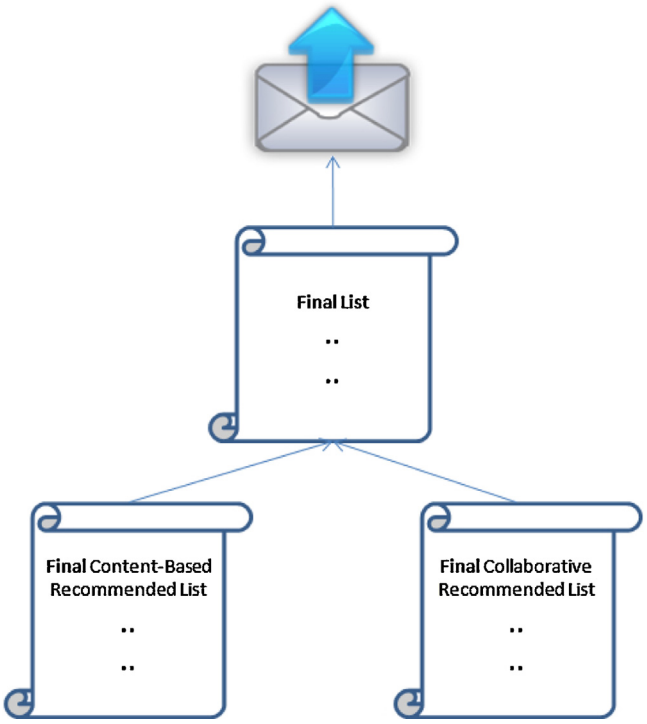

Fig. 7. Combination of both approach in one final list.

3.5.1. Diversity filter

In Fig. 7, we can see that to obtain the final list of recommendations, we previously require the definitive one of each approach, both content-based and collaborative. The first one is composed based on the preferences of the system in the matter of similarity, diversity and quality. In REFORE, the percentage which each list provide to the final list was establish as follow:

1. *Same selected journals list*: 40%.
2. *Same selected topics list*: 40%.
3. *Same non-selected journals list*: 20%.
4. *Same non-selected topics list*: [10%]. This last 10% is only used in case there are not enough relevant resources between the other lists.

With this percentage we intent to maintain the author up to date with the best and more important papers for him but we expose him to papers related with his topic in other areas.

3.5.2. Mixed hybrid combination

The final aggregation of the both lists, content-based and collaborative is made with the following percentage:

1. *Content-based*: 80%.
2. *Collaborative*: 20%.

As aforementioned, the novelty property prevails over the rest, that is, due to the peculiarity of the system, the most important resources are the new ones. Hence, we give much greater importance to the content-based approach by choosing a percentage of 80% of the total amount of items recommended. We maintain a 20% of item from the collaborative approach in order to not to lose important items that could be discarded in the two previous round.

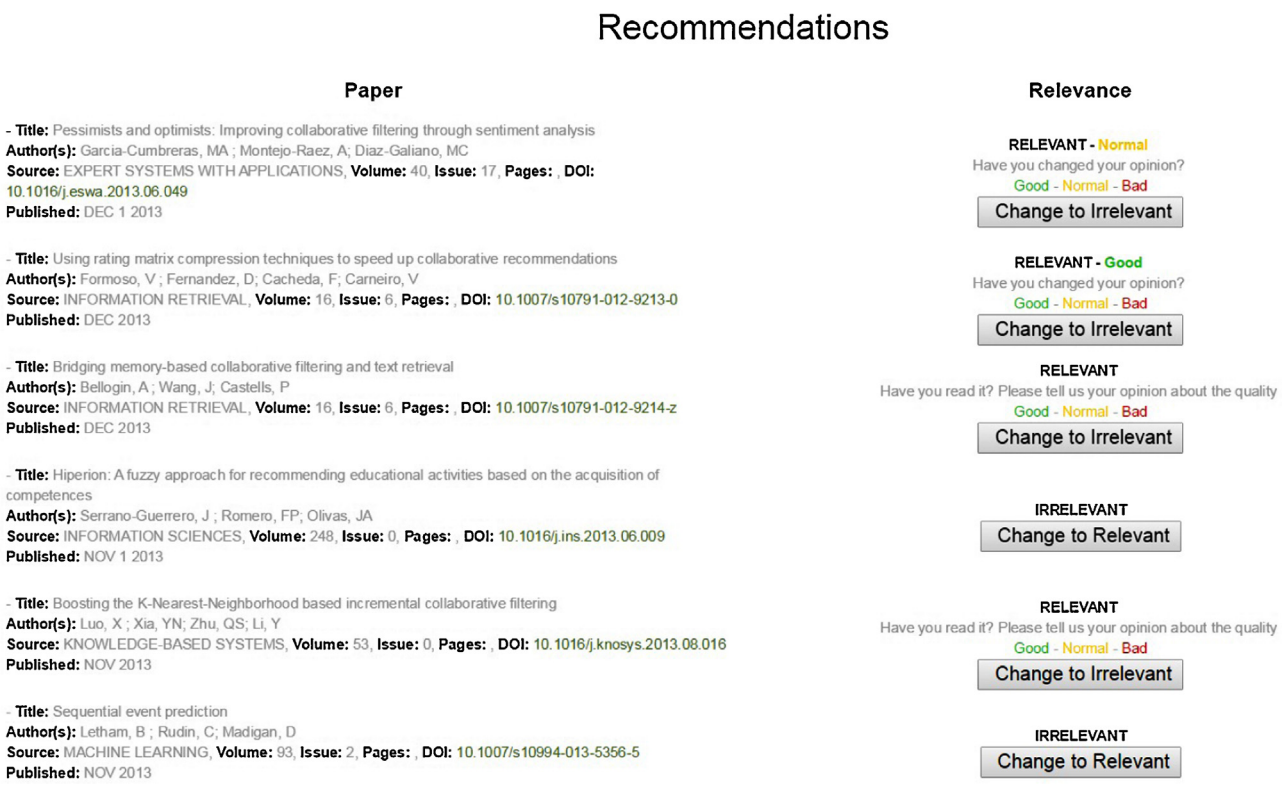

Fig. 8. Recommendations for a user.

3.6. Feedback phase

In this phase the recommender system calculates and updates the ratings of the recommended resources for the users. The system communicates the recommendations by email and registers the feedback of each user as explained below:

1. Interaction with the email content. The system recommends a resource r to the user u. In this email, the user u is asked to tell us about the resource relevance by clicking on *relevant* or *irrelevant* or by clicking on the link to the resource r – considered in this case as *relevant*. If the user does not select any of these options before the round is finished, the recommendation is considered irrelevant.
2. In the REFORE Web page, in their profiles users can mark as relevant or irrelevant the items recommended in the last round, as well as the previous one. The user communicates his/her linguistic evaluation assessment to the system, $rc_y \in S_4$.
3. In the REFORE Web page if an item was set as relevant, the user has the possibility of evaluating the quality of the paper as *Good*, *Normal* or *Bad* as is shown in the Fig. 8.

These evaluations are registered in the system for future recommendations. The idea of the collaborative approach is not only to improve the generated recommendations taking into account the users ratings, but also to palliate the possible side effect of the use of the diversity filter, where one good paper is left out of the final recommendations. That is why the interaction of users are very important for both approaches. First, the ratings provided by the user are taken to improve the quality of the next recommendations for himself. Second, the collaborative approach needs the users ratings to generate recommendations.

4. Experiments and evaluation

In this section we present the evaluation of the proposed recommender system. To do this, we could perform offline and online experiments [57]. But due to singularity of REFORE, we will only use online experiments based on the evaluation of the quality [58].

Moreover, in this case we cannot compare our method with other approaches using a standard data set. The reason is any standard data set has not this information about the preferences of users or the important phase of feedback and the adaptation of the existent data sets is not possible. We are not proposing a new recommendation strategy but an approach which combines hybrid recommendation with quality measures as bibliometrics that were widely used in many areas and whose efficiency is proved [35,61].

Consequently, in this study we only perform online experiments, i.e., practical studies where a group of researchers receive recommendations and report to the system their evaluations. When the users receive the list of recommendations per round, they provide feedback to the system rating if a recommended resource is relevant or not, i.e., they provide their opinions about the recommendation supplied by the system. If they are satisfied with the recommendation, they set as relevant recommendation. After that, they can also set the quality of a resource.

In this sense we carry out following experiment: we select the research members of different departments and introduce their data into the system. The system is tested sending every month during a year the different recommendations based on the new articles appeared in WoS. The users marked every month for each recommendation of the list if is relevant or not, and may also indicate the quality they consider for each paper. This allows us to contrast the functioning of the system and the different approaches followed in it from the analysis of user feedback.

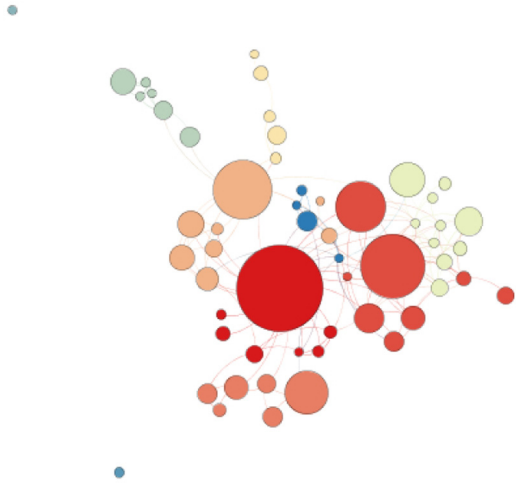

Fig. 9. Users similarities from University of Granada, weighted according to topics and quality.

4.1. Data sets

For the evaluation, we made the tests using the collaboration of around 100 researchers of the research groups: *Soft Computing and Intelligent Information Systems*[3] from University of Granada, *Delft Multimedia Information Retrieval Lab*[4] from TU Delft, some members from *Intelligent System and Data Mining*[5] and *Intelligent System Based on the Fuzzy Decision Analysis*[6] from University of Jaen and *Department of Civil Engineering*[7] from university of Granada. The similarities inside the set of users are stronger. Nevertheless, Fig. 9 shows a representation of the users from university of Granada (the larger community tested), where we can see the different clusters created where the bubbles represent users, the size of the bubble is proportional to their h-index and the similarity is represented by the line which join them, being the distance proportional to it and the lack of line the absence of similarity.

To build the set of resources 585,000 papers belonging to the year 2013 of 1731 journals related with the areas where the previous group are focused, have been downloaded from WOK using the tool described in [6]. The selected areas are the following:

Acoustics; Computer Science, Artificial Intelligence; Computer Science, Cybernetics; Computer Science, Hardware & Architecture; Computer Science, Information Systems; Computer Science, Interdisciplinary Applications; Computer Science, Software Engineering; Computer Science, Theory & Methods; Engineering, Civil; Engineering, Electrical & Electronic; Engineering, Multidisciplinary; Mathematical & Computational Biology; Mathematics; Mathematics, Applied; Mathematics, Interdisciplinary Applications; Medical; Informatics; Nanoscience & Nanotechnology; Neurosciences; Statistics & Probability.

The 585,000 resources were progressively added monthly from WoS obtaining the recent papers of journals monthly. The system filtered these resources when the recommendation round was launched and recommended them to the suitable users. To find out whether the recommendations were appropriate we based our study on the feedback given by the users, removing unusual behaviors such inactive users or users that did not update their profiles. We registered the ratings provided by users about the recommended resources to compare them with the predictions generated by our system.

The set of the users into the system is built following the steps described in Section 3.3.1. With the information provided by the users, the system sets up the user profiles. As we have mentioned, these profiles are dynamic and could be modified depending on the user preferences and on the feedback provided.

4.2. Assessing the capacity of recommendations

Commonly used measures are precision, recall and F1. They are used to evaluate whether a recommender system properly recommends items that users will consider relevant [57].

We used these measures to quantify the quality of the recommender system [58]. However, in this work we are not relying on experts who can say if the recommendations are relevant or not. Furthermore, we are implementing a temporal window to avoid overwhelming users with a excessive number recommendations. Under these conditions, we have no means to identify which items are relevant but are not recommended due to the limitation in the number of recommendations, so that, we cannot build a contingency table. For this reason, the way we measure the accuracy of our system is the **Mean Absolute Error (MAE)** [27,57], a commonly used accuracy metric which considers the average absolute deviation between a predicted relevance and the user true opinion, that is:

$$MAE = \frac{\sum_{i=1}^{n} abs(p_i - r_i)}{n} \quad (15)$$

where n is the number of cases in the test set, p_i the predicted relevance for a item, and r_i the true relevance, both binary.

As we have mentioned before, during the online study we record the ratings provided by users on the recommended resources to compare with the predictions generated by our system, and to calculate the MAE. As a rating metric for the system, we use the average of the MAE for all the users resulting into a final MAE of **0.30312**.

In our case, without the intervention of any expert participating in the process, that is, in a fully automated process, we are able to cover the whole number of users and items. That is, the coverage of the system is the 100% for each user based on the coverage definition given in [23]. The predictions generated with the new system are in line with the users preferences.

In order to test the configuration and combination of filters we choose for our system, we have tested the system with different configurations:

- No re-ranking (NRR) that is, the ten first items of the previous ordered list of items or users similar to the active user, either content-based and collaborative.
- Similarity measured based in abstract (ABS) instead of keywords, using tf-idf algorithm [53] to estimate the similarity between papers and users, either content-based and collaborative, without the use of re-ranking.
- Quality filter for the content-based based on the h-index (QH) of the authors of an article instead of the IF of the journal.
- Same quality filter in journals but without the application of the diversity filter (NDF).
- Actual Configuration (AC), that is, the structure explained above, re-ranking, quality filter based on the IF of the journal and on the h-index for an author and the use of the diversity filter.

Table 2 shows the MAE values for each configuration we used to prove that the chosen one is the optimal. Based on these values we can remark some points:

[3] http://sci2s.ugr.es/.
[4] http://dmirlab.tudelft.nl/content/members-dmir-lab.
[5] http://simidat.ujaen.es/.
[6] http://sinbad2.ujaen.es/.
[7] http://www.icivil.es/web2.0/.

Table 2
MAE for all the different configuration.

Configuration	MAE
NRR	0.39
ABS	0.72
QH	0.44
NDF	0.49
AC	0.30

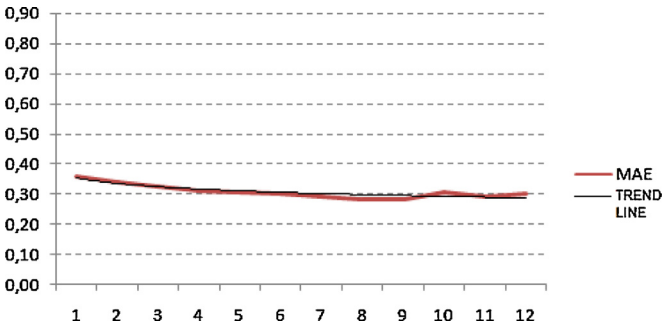

Fig. 10. Evolution of MAE during the year 2013.

- The result obtained after the non-application of the re-ranking (NRR), is lightly worst than the actual configuration (AC), so we can observe certain improvement obtained by the application of the quality as index for the re-ranking process.
- We noted that, the use of the analysis of the abstract based on the tf-idf algorithm produced poorer results than expected (ABS). Probably the fact that abstract are all written in a similar manner might explain this poor performance.
- The quality criteria of the h-index of authors (QH) of an article in the content-based approach proves that this criteria distorts the results. The reason of this behavior is that these good authors can be oriented to specific areas similar to user interest but with different approaches.
- The use of diversity filter (NDF) resulted to be an essential task to mitigate the own distortion introduced by the quality criteria, as the quality of some items could not be relevant if there is no relation with the actual profile.

In order to analyze the behavior of the system over time, we compare the MAE of the different twelve rounds corresponding with the monthly recommendation of the year 2013. We can see in the Fig. 10 the evolution of some decimals points in the monthly MAE due to the application of the collaborative filtering and the use of the black list as part of users profile. We also observe that the accuracy obtained by using combined quality of items and authors with the diversity filter outperforms the predictions computed without them or with a different approach. Specifically we achieved an improvement of 9% against *NRR*, 42% against *ABS*, 14% against *QH* and 19% against *NDF*. According to the results, the predictions generated with the system are more accurate. The users part of the trial showed themselves satisfied with the results.

5. Concluding remarks

The personalization of information related to research papers deserves special attention due to the ever growing information overload problem in the research world. Researchers need tools to assist them in their processes of getting the newest and most relevant information recently produced. Recommender systems have successfully obtained relevant information in both academic and commercial environments. Quality has been taken into account in previous works, but it relied on the supervision of experts for recommended items. Since the number of items is continuously growing, the human interaction evolved needs to be reduced to the minimum. In this paper we approached the recommendation generation process about new research papers as a hybrid recommender system. On the one hand, a content-based approach to obtain the latest papers inserted into the system, non rated by any user, and on the other hand, a collaborative approach to avoid disregarding potentially important papers for a user, rated in the last two rounds by experts in the field.

So, we have created a hybrid fuzzy linguistic recommender system based on the quality of the items and users, which has been then applied in the database extracted from WoS to assist users to access relevant papers. The system measures both items and users quality for the re-ranking process. Beside the quality of items and the quality of the users, REFORE takes into account the users profiles for the top-N recommendations. In addition, the system improves the regular feedback process in two ways. Firstly evaluating the recommendation as relevant or irrelevant, which then becomes part of the user profile. Secondly, giving the user the possibility to assess the quality of the item recommended. We have performed online studies with the proposed system and the experimental results yielded satisfactory results. The experimental study performed allowed us to contrast the behavior of these different approaches implemented in the hybrid recommender system. The proposed quality procedure in conjunction with the application of a filter based on the user profile in the re-ranking process has shown better accuracy than the use of different techniques or the absence of any. The results obtained with different methods, supports the conclusion drawn. As part of future work, we are considering enhancing the paper keywords with the automatic extraction of terms from the article abstracts for the users vectors to estimate similarity, which we think can provide promising results. One step in this direction is to use Wordnet [21] in the process of nouns extraction. Moreover, we want to expand the use of REFORE as a tool for users not only to be up to date on the newest trends but also for researchers who want to enter in a new area of research, or even to have the best personalized papers for a wider time window.

Acknowledgements

This paper has been developed with the financing of FEDER funds in FUZZYLING-II Project, TIN2010-17876, TIN2013-40658-P and Andalusian Excellence Projects TIC-05299 and TIC-5991.

References

[1] Google scholar. http://scholar.google.com/intl/en/scholar/about.html
[2] Scopus. http://www.elsevier.com/online-tools/scopus
[3] Web of science, information. http://wokinfo.com/
[4] G. Adomavicius, Y. Kwon, Toward more diverse recommendations: Item re-ranking methods for recommender systems, in: Workshop on Information Technologies and Systems, 2009.
[5] J.E. Hirsch, An index to quantify an individual's scientific research output, Proc Natl Acad Sci U S A. 102 (46) (2005) 16569–16572.
[6] S. Alonso, F.J. Cabrerizo, E. Herrera-Viedma, F. Herrera, Wos query partitioner: a tool to retrieve very large numbers of items from the web of science using different source-based partitioning approaches, J. Am. Soc. Inform. Sci. Technol. 61 (8) (2010) 1582–1597.
[7] S. Alonso, F. Chiclana, F. Herrera, E. Herrera-Viedma, J. Alcalá-Fdez, C. Porcel, A consistency-based procedure to estimating missing pairwise preference values, Int. J. Intell. Syst. 23 (2008) 155–175.
[8] S. Alonso, I.J. Pérez, F.J. Cabrerizo, E. Herrera-Viedma, A linguistic consensus model for web 2.0 communities, Appl. Soft Comput. 13 (1) (2013) 149–157.
[9] 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.
[10] P. Bedi, P. Vashisth, P. Khurana, et al., Modeling user preferences in a hybrid recommender system using type-2 fuzzy sets, in: IEEE International Conference on Fuzzy Systems (FUZZ), 2013, IEEE, 2013, pp. 1–8.
[11] R. Burke, Hybrid recommender systems: survey and experiments, User Model. User-Adapt. Interact. 12 (2002) 331–370.

[12] R. Burke, Hybrid web recommender systems, in: P. Brusilovsky, A. Kobsa, W. Nejdl (Eds.), The Adaptive Web, LNCS 4321, 2007, pp. 377–408.
[13] F. Cabrerizo, J. López-Gijón, A. Ruíz-Rodríguez, E. Herrera-Viedma, A model based on fuzzy linguistic information to evaluate the quality of digital libraries, Int. J. Inform. Technol. Decis. Mak. 9 (3) (2010) 455–472.
[14] H. Chao, Assessing the quality of academic libraries on the web: the development and testing of criteria, Libr. Inform. Sci. Res. 24 (2002) 169–194.
[15] P.-Y. Chen, S.-Y. Wu, Does Collaborative Filtering Technology Impact Sales? Empirical Evidence From amazon.com, 2007.
[16] 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, pp. 714–720.
[17] A. Czarl, M. Belovecz, Role of research and development in the 21st century, in: Proceedings of the Eight International Conference on Informatics in Economy, 2007, pp. 497–502.
[18] V. Durieux, P.A. Gevenois, Bibliometric indicators: quality measurements of scientific publication 1, Radiology 255 (2) (2010) 342–351.
[19] A. Edmunds, A. Morris, The problem of information overload in business organizations: a review of the literature, Int. J. Inform. Manag. 20 (2000) 17–28.
[20] B. Esteban, A. Tejeda-Lorente, C. Porcel, M. Arroyo, E. Herrera-Viedma, Tplufibweb: a fuzzy linguistic web system to help in the treatment of low back pain problems, Knowl.-Based Syst. 67 (2014) 429–438.
[21] C. Fellbaum, WordNet, Wiley Online Library, 1999.
[22] E. Garfield, Citation indexes for science. a new dimension in documentation through association of ideas, Int. J. Epidemiol. 35 (5) (2006) 1123–1127.
[23] M. Ge, C. Delgado-Battenfeld, D. Jannach, Beyond accuracy: evaluating recommender systems by coverage and serendipity, in: Proceedings of the Fourth ACM Conference on Recommender Systems, ACM, 2010, pp. 257–260.
[24] N. Good, J. Schafer, J. Konstan, A. Borchers, B. Sarwar, J. 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.
[25] U. Hanani, B. Shapira, P. Shoval, Information filtering: overview of issues, research and systems, User Model. User-Adapt. Interact. 11 (2001) 203–259.
[26] J. Herlocker, J. Konstan, J. Riedl, An empirical analysis of design choices in neighborhood-based collaborative filtering algorithms, Inform. Retr. 5 (2002) 287–310.
[27] J. Herlocker, J. Konstan, L. Terveen, J. Riedl, Evaluating collaborative filtering recommender systems, ACM Trans. Inform. Syst. 22 (1) (2004) 5–53.
[28] F. Herrera, S. Alonso, F. Chiclana, E. Herrera-Viedma, Computing with words in decision making: foundations trends and prospects, Fuzzy Optim. Decis. Mak. 8 (4) (2009) 337–364.
[29] F. Herrera, L. Martínez, A 2-tuple fuzzy linguistic representation model for computing with words, IEEE Trans Fuzzy Syst. 8 (6) (2000) 746–752.
[30] F. Herrera, L. Martínez, A model based on linguistic 2-tuples for dealing with multigranularity hierarchical linguistic contexts in multiexpert decision-making, IEEE Trans. Syst. Man Cybern. B: Cybern. 31 (2) (2001) 227–234.
[31] M. Hsu, A personalized English learning recommender system for ESL students, Expert Syst. Appl. 34 (2008) 377–408.
[32] A.S. Hussein, W.M. Omar, X. Li, M. Ati, Efficient chronic disease diagnosis prediction and recommendation system, in: 2012 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES), IEEE, 2012, pp. 209–214.
[33] P. Katerattanakul, B. Han, S. Hong, Objective quality ranking of computing journals, Commun. ACM 46 (10) (2003) 111–114.
[34] R. Korfhage, Information Storage and Retrieval, Wiley Computer Publishing, New York, 1997.
[35] R. Kostoff, R. Tshiteya, K. Pfeil, J. Humenik, G. Karypis, Power source roadmaps using bibliometrics and database tomography, Energy 30 (5) (2005) 709–730.
[36] C. Long-Sheng, H. Fei-Hao, C. Mu-Chen, H. Yuan-Chia, Developing recommender systems with the consideration of product profitability for sellers, Inform. Sci. 178 (4) (2008) 1032–1048.
[37] I. Macho-Stadler, D. Perez-Castrillo, R. Veugelers, Licensing of university inventions: the role of a technology transfer office, Int. J. Ind. Org. 25 (2007) 483–510.
[38] C.D. Maio, G. Fenza, M. Gaeta, V. Loia, F. Orciuoli, S. Senatore, Rss-based e-learning recommendations exploiting fuzzy {FCA} for knowledge modeling, Appl. Soft Comput. 12 (1) (2012) 113–124.
[39] O. Martin, G.J. Klir, On the problem of retranslation in computing with perceptions, Int. J. Gen. Syst. 35 (6) (2006) 655–674.
[40] M. Martí nez, M. Herrera, J. López-Gijón, E. Herrera-Viedma, H-classics: characterizing the concept of citation classics through h-index, Scientometrics 98 (2014) 1971–1983.
[41] S. Massanet, J.V. Riera, J. Torrens, E. Herrera-Viedma, A new linguistic computational model based on discrete fuzzy numbers for computing with words, Inform. Sci. 258 (2014) 277–290.
[42] G. Meghabghab, A. Kandel, Search Engines, Link Analysis, and User's Web Behavior, Springer-Verlag, Berlin Heidelberg, 2008.
[43] J. Mendel, Historical reflections on perceptual computing, in: Proceedings of the 8th International FLINS Conference on Computational Intelligence in Decision and Control, World Scientific, 2008, pp. 181–186.
[44] J.M. Mendel, An architecture for making judgments using computing with words, Int. J. Appl. Math. Comput. Sci. 12 (2002) 325–335.
[45] G. Miller, The magical number seven or minus two: some limits on our capacity of processing information, Psychol. Rev. 63 (1956) 81–97.
[46] J. Morente-Molinera, I. Perez, M. Ureña, E. Herrera-Viedma, On multi-granular fuzzy linguistic modelling in group decision making problems: a systematic review and future trends, Knowl.-Based Syst. 74 (2015) 49–60.
[47] A. Moro-Martín, Spanish changes are scientific suicide, Nature 482 (7385) (2012), 277–277.
[48] F. Narin, K.S. Hamilton, D. Olivastro, The increasing linkage between u.s. technology and public science, Res. Policy 26 (3) (1997) 317–330.
[49] S. Nerur, R. Sikora, G. Mangalaraj, V. Balijepally, Assessing the relative influence of journals in a citation network, Commun. ACM 48 (11) (2005) 71–74.
[50] A. Popescul, L. Ungar, D. Pennock, S. Lawrence, Probabilistic models for unified collaborative and content-based recommendation in sparce-data environments, in: Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence (UAI), San Francisco, 2001, pp. 437–444.
[51] C. Porcel, E. Herrera-Viedma, Dealing with incomplete information in a fuzzy linguistic recommender system to disseminate information in university digital libraries, Knowl.-Based Syst. 23 (2010) 32–39.
[52] E. Rahm, A. Thor, Citation analysis of database publications, ACM Sigmod Rec. 34 (4) (2005) 48–53.
[53] S. Robertson, Understanding inverse document frequency: on theoretical arguments for IDF, J. Doc. 60 (5) (2004) 503–520.
[54] S.E. Robertson, The probability ranking principle in ir, J. Doc. 33 (4) (1977) 294–304.
[55] S. Saha, S. Saint, D.A. Christakis, Impact factor: a valid measure of journal quality? J. Med. Libr. Assoc. 91 (1) (2003) 42.
[56] R. Schenkel, T. Crecelius, M. Kacimi, T. Neumann, J. Parreira, M. Spaniol, G. Gerhard Weikum, Social wisdom for search and recommendation, Data Eng. Bull. 31 (2) (2008) 40–49.
[57] G. Shani, A. Gunawardana, in: F. Ricci, L. Rokach, B. Shapira, P.B. Kantor (Eds.), Recommender Systems Handbook, Chap. Evaluating Recommendation Systems, Springer, 2011, pp. 257–298.
[58] A. Tejeda-Lorente, C. Porcel, E. Peis, R. Sanz, E. Herrera-Viedma, A quality based recommender system to disseminate information in a university digital library, Inform. Sci. 261 (2014) 52–69.
[59] I.B. Türkşen, Type 2 representation and reasoning for CWW, Fuzzy Sets Syst. 127 (1) (2002) 17–36.
[60] I.B. Türkşen, Meta-linguistic axioms as a foundation for computing with words, Inform. Sci. 177 (2) (2007) 332–359.
[61] C.J. Williams, M. O'Rourke, S.D. Eigenbrode, I. O'Loughlin, S.J. Crowley, Using bibliometrics to support the facilitation of cross-disciplinary communication, J. Am. Soc. Inform. Sci. Technol. 64 (9) (2013) 1768–1779.
[62] R.R. Yager, On the retranslation process in Zadeh's paradigm of computing with words, IEEE Trans. Syst. Man Cybern. B: Cybern. 34 (2) (2004) 1184–1195.
[63] Z. Zhuang, E. Elmacioglu, D. Lee, C.L. Giles, Measuring conference quality by mining program committee characteristics, in: Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries, ACM, 2007, pp. 225–234.