Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE @ DIRECT®

ELSEVIER International Journal of Approximate Reasoning 34 (2003) 221–239

INTERNATIONAL JOURNAL OF APPROXIMATE REASONING

www.elsevier.com/locate/ijar

A model of fuzzy linguistic IRS based on multi-granular linguistic information ☆

E. Herrera-Viedma [a,*], O. Cordón [a], M. Luque [a], A.G. Lopez [a], A.M. Muñoz [b]

[a] *Department of Computer Science and A.I., University of Granada, Library Science Studies School, Granada 18071, Spain*

[b] *Department of Library Science Studies, University of Granada, Granada, Spain*

Received 1 February 2003; accepted 1 July 2003

Abstract

An important question in IRSs is how to facilitate the IRS-user interaction, even more when the complexity of the fuzzy query language makes difficult to formulate user queries. The use of linguistic variables to represent the input and output information in the retrieval process of IRSs significantly improves the IRS-user interaction. In the activity of an IRS, there are aspects of different nature to be assessed, e.g., the relevance of documents, the importance of query terms, etc. Therefore, these aspects should be assessed with different uncertainty degrees, i.e., using several label sets with different granularity of uncertainty.

In this contribution, an IRS based on fuzzy multi-granular linguistic information and a method to process the multi-granular linguistic information are proposed. The system accepts Boolean queries whose terms can be simultaneously weighted by means of ordinal linguistic values according to three semantics: a symmetrical threshold semantics, a relative importance semantics and a quantitative semantics. In the three semantics, the linguistic weights are represented by the linguistic variable "Importance", but assessed on different label sets S^1, S^2 and S^3, respectively. The IRS evaluates weighted queries and obtains the linguistic retrieval status values of documents represented by the linguistic variable "Relevance" which is expressed on a different label set S'.

☆ This research has been supported by CICYT under project TIC2002-03276.
* Corresponding author. Tel.: +34-95824-4258; fax: +34-95824-3317.
E-mail address: viedma@decsai.ugr.es (E. Herrera-Viedma).

0888-613X/$ - see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/j.ijar.2003.07.009

Keywords: Information retrieval; Linguistic modelling; Multi-granular linguistic information

1. Introduction

Information Retrieval (IR) is a research field related to the storage and retrieval of textual information [1,12,16]. The main activity of an IRS is the gathering of the pertinent filed documents that better satisfy user information requirements (queries). Both documents and user queries must be formally represented in a consistent way, so that the IRS can satisfactorily develop the retrieval activity. Basically, IRSs present three components to carry out their activity:

(1) *A database*: an archive of documents and their representation obtained by means of the indexing procedure.
(2) *A query formulation module*: which allows the users to formulate their queries by means of a query language.
(3) *A query evaluation module*: which evaluates the documents for a user query. It presents an inference procedure that establishes a relationship between the user query and the documents stored in the database to determine the relevance of each document to the user query.

Most of the existing IRSs are based on the Boolean retrieval model [15,18]. In such IRSs, the database represents the documents as sets of index terms, the query subsystem represents the user queries as Boolean combinations of index terms, and the evaluation subsystem uses a total matching mechanism between documents and queries as the inference procedure. These IRSs present many limitations [15], mainly the lack of flexibility and precision for representing document contents, for describing user queries and for characterizing the relevance of the documents retrieved for a given user query. These drawbacks may be overcome by incorporating weights in the three levels of information representation of an IRS:

(1) *Document representation level.* By computing weights of index terms, the system specifies to what extent a document matches the concept expressed by the index terms.
(2) *Query representation level.* By associating weights to a query, a user can provide a more precise description of his/her information needs or desired documents.
(3) *Evaluation representation level.* By comparing the query term weights with the index term weights, the evaluation mechanism procedures a so called “Retrieval Status Value” (RSV).

In this paper, we mainly focus on the query and evaluation representation levels by proposing several tools to facilitate the IRS-user interaction.

The use of linguistic variables [22] to represent the input and output information in the retrieval process of IRSs considerably improves the IRS-user interaction (see [4,5,10,11,14]). The most linguistic IRSs assume that users provide their information needs by means of Boolean queries whose terms are weighted by linguistic values represented by the linguistic variable "Importance" assessed on a label set S. Then, the activity of the IRS involves evaluating the linguistic weighted queries and providing the linguistic RSVs of documents represented by the linguistic variable "Relevance", which is also assessed on S. The drawback is that the use of the same label set to express the inputs and outputs of the IRS diminishes the communication possibilities in the IRS-user interaction. Furthermore, given that both linguistic variables, "Importance" and "Relevance" represent different concepts, it seems necessary to use different label sets in their linguistic modelling, i.e., to apply a multi-granular linguistic modelling [9]. This means to use label sets with different granularity and/or semantics to represent the different information kinds that appears in the retrieval process.

The aim of this contribution is to present a model of IRS that uses multi-granular linguistic information to carry out its activity. The multi-granular linguistic information is modelled using a fuzzy ordinal linguistic approach [5,10,11]. We define a method to process the multi-granular linguistic information in an IR context. The weighted Boolean queries and the linguistic RSVs of documents are assessed on label sets with different granularity and/or semantics. The query terms can be simultaneously weighted by three linguistic weights associated with three different semantics: a symmetrical threshold semantics, an relative importance semantics and a quantitative semantics. The Boolean operators AND and OR are modelled by means of the OWA aggregation operator [21]. The OWA operator is an "and-or" operator, and this property allows us to introduce a soft computing in the evaluation of queries. The retrieved documents are arranged in linguistic relevance classes, which are identified by ordinal linguistic values assessed on a label set that is different to those used to assess the query weights.

This contribution is set out as follows. Section 2 reviews the fuzzy ordinal linguistic approach, the concept of multi-granular linguistic information, and the OWA operator. Section 3 presents the IRS based on multi-granular linguistic information. Finally, some concluding remarks are pointed out.

2. Preliminaries

In this section, we review some tools of fuzzy linguistic information processing that will be used in the development of our IRS.

2.1. Fuzzy ordinal linguistic approach

The *fuzzy linguistic approach* is an approximate technique appropriate to deal with qualitative aspects of problems. It models linguistic values by means of *linguistic variables* [22]. Its application is beneficial because it introduces a more flexible framework for representing information in a more direct and adequate way when it is not possible to express it accurately.

The *ordinal fuzzy linguistic approach* is a special kind of fuzzy linguistic approach that facilitates the linguistic modelling [5,8,10]. An ordinal fuzzy linguistic approach is defined by considering a finite and totally ordered label set $S = s_i,\ i \in \{0, \ldots, \mathscr{T}\}$ in the usual sense ($s_i \geqslant s_j$ if $i \geqslant j$) and with odd cardinality (7 or 9 labels). The mid term representing an assessment of "approximately 0.5" and the rest of the terms being placed symmetrically around it [2]. The semantics of the label set is established from the ordered structure of the label set by considering that each label for the pair $(s_i, s_{\mathscr{T}-i})$ is equally informative. In some approaches [8,10,11], the semantics is completed by assigning fuzzy numbers defined on the $[0, 1]$ interval to the labels. These membership functions are described by linear trapezoidal membership functions represented by the 4-tuple $(a_i, b_i, \alpha_i, \beta_i)$ (the first two parameters indicate the interval in which the membership value is 1.0; the third and fourth parameters indicate the left and right widths of the distribution). Furthermore, we require the following operators:

(1) Negation operator: $\text{Neg}(s_i) = s_j,\ j = \mathscr{T} - i$.
(2) Maximization operator: $\text{MAX}(s_i, s_j) = s_i$ if $s_i \geqslant s_j$.
(3) Minimization operator: $\text{MIN}(s_i, s_j) = s_i$ if $s_i \leqslant s_j$.

2.2. Multi-granular linguistic information

In any linguistic approach, an important parameter to be determined is the *granularity of uncertainty*, i.e., the cardinality of the label set S used to express the linguistic information. The cardinality of S must be small enough so as not to impose useless precision levels to the users, and it must be rich enough in order to allow a discrimination of the assessments in a limited number of degrees.

On the other hand, according to the uncertainty degree that a user qualifying a phenomenon has on it, the label set chosen to provide his knowledge will have more or less terms. When different users have different uncertainty degrees on the phenomenon, then several label sets with a different granularity of uncertainty are necessary. In the latter case, we need tools for the management of multi-granular linguistic information to model these situations. Different proposals can be found in [9,17].

2.3. The OWA operator

The Ordered Weighted Averaging (OWA) is an aggregation operator of information which acts taking into account the order of the assessments to be aggregated. It was defined as follows:

Definition 1 [21]. Let $A = \{a_1, \ldots, a_m\}$, $a_k \in [0, 1]$ be a set of assessments to be aggregated, then the OWA operator, ϕ, is defined as

$$\phi(a_1, \ldots, a_m) = W \cdot B^{\mathrm{T}}$$

where $W = [w_1, \ldots, w_m]$, is a weighting vector, such that $w_i \in [0, 1]$ and $\Sigma_i w_i = 1$; and $B = \{b_1, \ldots, b_m\}$ is a vector associated to A, such that, $B = \sigma(A) = \{a_{\sigma(1)}, \ldots, a_{\sigma(m)}\}$, where $a_{\sigma(j)} \leqslant a_{\sigma(i)}$ $\forall i \leqslant j$, with σ being a permutation over the set of labels A.

The OWA operator is an "and-or" operator [21]. This property allows the OWA operator to carry out a soft computing in the modelling of the MAX and MIN operators. In order to classify OWA operators in regard to their location between "and" and "or", Yager [21] introduced a *orness measure*, associated with any vector W as follows

$$\text{orness}(W) = \frac{1}{m-1} \sum_{i=1}^{m} (m-i) w_i$$

Fixed a weighting vector W, then the closer an OWA operator is to an "or", the closer its orness measure is to one; while the nearer it is to an "and", the closer is the latter measure to zero. Generally, an OWA operator with much of the nonzero weights near the top will be an orlike operator ($\text{orness}(W) > 0.5$), and when the most of the nonzero weights are near the bottom, the OWA operator will be an andlike operator ($\text{orness}(W) \leqslant 0.5$). We use this good property in our linguistic IRS to evaluate the logical connectives of Boolean queries OR and AND.

3. The IRS based on multi-granular linguistic information

In this section we present a model of IRS that accepts linguistic weighted Boolean queries and provides linguistic RSVs expressed using multi-granular linguistic information. Thus, it uses multi-granular linguistic weighted queries and multi-granular linguistic RSVs. Other important property of this IRS is that it models the Boolean operators in a flexible way by means of the OWA operators [21].

Before presenting the proposal, we show the assumed framework. We consider a set of documents $D = \{d_1, \ldots, d_m\}$ represented by means of index terms $T = \{t_1, \ldots, t_l\}$, which describe the subject content of the documents. A numeric indexing function $F : D \times T \to [0, 1]$ is defined, called *index term weighting*. F maps a given document d_j and a given index term t_i to a numeric weight between 0 and 1. Thus, $F(d_j, t_i)$ is a numerical weight that represents the degree of significance of t_i in d_j. $F(d_j, t_i) = 0$ implies that the document d_j is not at all about the concept(s) represented by the index term t_i and $F(d_j, t_i) = 1$ implies that the document d_j is perfectly represented by the concept(s) indicated by t_i. Using the numeric values in $(0, 1)$ F can weight index terms according to their significance in describing the content of a document in order to improve the document retrieval.

In the following subsections, we analyze the main elements of our model of IRS, composed of, multi-granular linguistic weighted queries and the evaluation subsystem of such queries.

3.1. Defining multi-granular linguistic queries

We consider that each query is expressed as a combination of the weighted index terms which are connected by the logical operators AND ($\wedge$), OR ($\vee$), and NOT ($\neg$) and weighted with ordinal linguistic terms. Each term in a query can be simultaneously weighted by means of several weights [10,11]. Particularly, a term of a query can be weighted by means of three weights associated with different semantics. In such a way, the system gives a major support to the specification of user information needs.

3.1.1. The semantics of query weights

By assigning weights to queries, users specify restrictions on the documents that the IRS has to satisfy in the retrieval activity. We observe that a user can impose two kinds of restrictions on documents to be retrieved [10]:

(1) Qualitative restrictions: when the query weights express criteria that affect the quality of the document representation to be retrieved, i.e., constraints to be satisfied by the index term weights that appear in the retrieved document representations.
(2) Quantitative restrictions: when the query weights express criteria that affect the quantity of the documents to be retrieved, i.e., constraints to be satisfied by the number of documents to be retrieved.

Usually, most of the classical fuzzy query languages [3,4,6,14] do not allow users to build weighted queries according to different semantics simultaneously. In [10,11], we proposed different query languages that allow users to build weighted queries according to different semantics simultaneously. As was done in [10] we propose to use the following three semantics to represent the meaning of the weights of a query term:

(1) *Symmetrical threshold semantics* [10]. By associating threshold weights to terms in a query, the user is asking to retrieve all documents sufficiently about the topics represented by such terms [6,7,13]. Usually, a threshold semantics requires to reward a document whose index term weights F exceed the established thresholds with a high RSV, but allowing some small partial credit for a document whose F values are lower than the thresholds. Then, the query weights indicate presence requirements, i.e., they are presence weights. A symmetrical threshold semantics is a special threshold semantics which assumes that a user may use presence weights or absence weights in the formulation of weighted queries. Then, it is symmetrical with respect to the mid threshold value, i.e., it presents the usual behaviour for the threshold values which are on the right of the mid threshold value (presence weights), and the opposite behaviour for the values which are on the left (absence weights or presence weights with low values).

(2) *Relative importance semantics* [3,19]. This semantics defines term weights as a measure of the relative importance of each term of a query with respect to the remainder. By associating relative importance weights to terms in a query, the user is asking to see all documents whose content represents to a higher degree the concepts associated to the most important terms than to the less important ones. In practice, this means that the user requires that the computation of the RSV of a document is dominated by the more heavily weighted terms.

(3) *Quantitative semantics*. This semantics defines query weights as measures of the quantity of documents that users want to consider in the computation of the final set of documents retrieved for each query term. By associating quantitative weights with the terms in a query, the user is asking to see a set of retrieved documents in which the terms with a greater quantitative weight contribute with a higher number of pertinent documents. In practice, the use of this quantitative semantics has two beneficial consequences with respect to the classical existing system:

- The RSVs are calculated using a restricted number of document determined for each query term by its quantitative weight. With this weight, a user can choose those documents that best satisfy the concepts represented by the term, or most documents that satisfy the concepts, or some documents that satisfy the concepts, etc. Hence, we may perform a refinement or tuning of the output documents of IRS. In our case, this semantics helps us to refine the relevance classes of documents in the output of the IRS.

- A soft control on the total number of retrieved documents that is performed query term by query term.

We should point out that the chosen semantics are consistent and complementary between one another in the following sense: (i) consistent means that the information needs expressed by some semantics do not contradict those expressed by the others; and (ii) complementary means that the users can express all or the larger part of their informations needs using the chosen semantics.

3.1.2. Rules for formulating multi-granular linguistic weighted queries

As in [10], we use the linguistic variable "*Importance*" to model every semantics, but with different interpretations. For example, a query term t_i with a threshold weight of value "*High*" means that the user requires documents whose content t_i should have at least a high importance value. However, the same query term t_i with a quantitative weight of value "*High*" means that the user wants a set of documents in which the term t_i contributes with a higher number of pertinent documents; and the same query term t_i with an importance weight of value "*High*" means that the user requires that the meaning of t_i must have a high importance value in the computation of the set of retrieved documents. Therefore, the problem in such model [10] is that different linguistic weights associated with a term are assessed on the same label set, S. To solve this problem, we propose to represent the linguistic weights using multi-granular linguistic information, i.e., assuming label sets with different cardinality and/or semantics to assess the weights associated with the three semantics, called S^1, S^2 and S^3, respectively.

Then, we assume that a query is any legitimate Boolean expression whose atomic components (atoms) are 4-tuples $\langle t_i, c_i^1, c_i^2, c_i^3 \rangle$ belonging to the set, $T \times S^1 \times S^2 \times S^3$; $t_i \in T$, $c_i^1 \in S^1$ is a value of the linguistic variable "*Importance*" modelling the symmetrical threshold semantics, $c_i^2 \in S^2$ is a value of the linguistic variable "*Importance*" modelling the quantitative semantics, and $c_i^3 \in S^3$ is a value of the linguistic variable "*Importance*" modelling the relative importance semantics. Therefore, the set of legitimate Boolean queries is a set of multi-granular linguistic weighted queries Q which is defined by the following syntactic rules:

1. $\forall q = \langle t_i, c_i^1, c_i^2, c_i^3 \rangle \in T \times S^1 \times S^2 \times S^3 \rightarrow q \in Q$. These queries are called atoms.
2. $\forall q, p \in Q \rightarrow q \wedge p \in Q$.
3. $\forall q, p \in Q \rightarrow q \vee p \in Q$.
4. $\forall q \in Q \rightarrow \neg(q) \in Q$.
5. Every legitimate Boolean query $q \in Q$ is only obtained by applying rules 1–4.

3.2. Evaluating multi-granular linguistic weighted queries

The goal of the evaluation subsystem consists of evaluating documents in terms of their relevance to a weighted query according to three different semantics. Usually, evaluation methods for Boolean queries act by means of a constructive bottom-up process, i.e., in the query evaluation process, the atoms are evaluated first, then the Boolean combinations of the atoms, and so forth, working in a bottom-up fashion until the whole query is evaluated. Similarly, we propose a constructive bottom-up evaluation method to process the multi-granular linguistic weighted queries. This method evaluates documents in terms of their relevance to queries by supporting the three semantics associated with the query weights simultaneously and by managing the multi-granular linguistic weights satisfactorily. Furthermore, given that the concept of relevance is different from the concept of importance, we use a label set S' to provide the relevance values of documents, which is different from those used to express the queries (S^1, S^2 and S^3).

To manage the multi-granular linguistic weights of queries we develop a procedure based on the multi-granular linguistic information management tool defined in [9]. This procedure acts making uniform the multi-granular linguistic information before processing queries. To do so, we have to choose a label set as the uniform representation base, called *basic linguistic term set* (BLTS), and then we have to transform (under a transformation function) all multi-granular linguistic information into that unified label set BLTS. In our case, the choice of the BLTS is easy to perform. It must be the label set used to express the output of the IRS (relevance degrees of documents), i.e., $\text{BLTS} = S'$.

The method to evaluate a multi-granular linguistic weighted query is composed of the following steps:

(1) *Preprocessing of the query*. The user query is preprocessed to put it into either conjunctive normal form (CNF) or disjunctive normal form (DNF), in such a way that every Boolean subexpression must have more than two atoms. Weighted single-term queries are kept in their original forms. Then, if we have a query q_w with $\mathscr{I}$ subexpressions and $\mathscr{N}$ atoms, it can appear in any of the forms illustrated graphically in Fig. 1, i.e., as OR/Weighted-AND or as AND/Weighted-OR trees.

(2) *Evaluation of atoms with respect to the symmetrical threshold semantics*. According to a symmetrical threshold semantics, a user may search for documents with a minimally acceptable presence of one term in their representations, or documents with a maximally acceptable presence of one term in their representations [10,11]. Then, when a user asks for documents in which the concept(s) represented by a term t_i is (are) with the value *High Importance*, he/she would not reject a document with an F value greater than *High*. On the contrary, when a user asks for documents in which the concept(s) represented

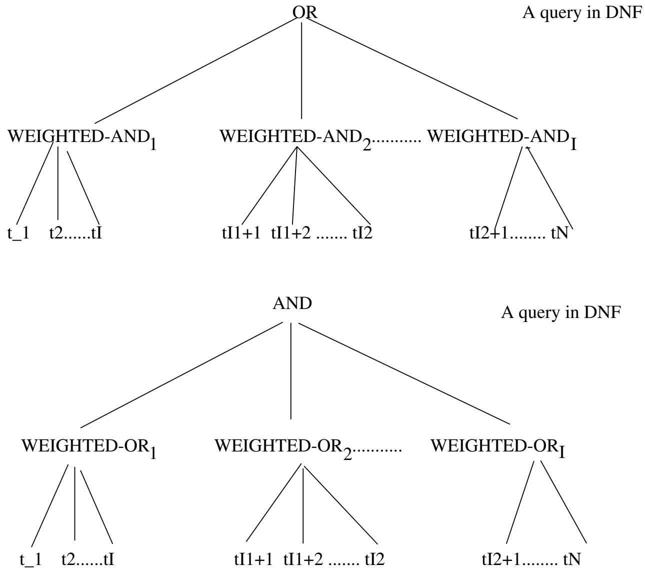

Fig. 1. Queries in normal form.

by a term t_i is (are) with the value *Low Importance*, he/she would not reject a document with an F value less than *Low*. Given a request $\langle t_i, c_i^1, c_i^2, c_i^3 \rangle$, this means that the query weights that imply the presence of a term in a document $c_i^1 \geqslant s^1_{\mathcal{T}/2}$ (e.g. *High*, *Very High*) must be treated differently to the query weights that imply the absence of one term in a document $c_i^1 < s^1_{\mathcal{T}/2}$ (e.g. *Low*, *Very Low*). Then, if $c_i^1 \geqslant s^1_{\mathcal{T}/2}$, the request $\langle t_i, c_i^1, c_i^2, c_i^3 \rangle$ is synonymous with the request $\langle t_i,$ at least $c_i^1, c_i^2, c_i^3 \rangle$, which expresses the fact that the desired documents are those having F values as high as possible; and if $c_i^1 < s^1_{\mathcal{T}/2}$, the former request is synonymous with the request $\langle t_i,$ at most $c_i^1, c_i^2, c_i^3 \rangle$, which expresses the fact that the desired documents are those having F values as low as possible. This interpretation is defined by means of a parameterized linguistic matching function $g^1 : D \times T \times S^1 \rightarrow S^1$ [11]. Given an atom $\langle t_i, c_i^1, c_i^2, c_i^3 \rangle$ and a document $d_j \in D$, g^1 obtains the linguistic RSV of d_j, called $\mathrm{RSV}_j^{i,1}$, by measuring how well the index term weight $F(d_j, t_i)$ satisfies the request expressed by the linguistic weight c_i^1 according to the following expression: $g^1(d_j, \langle t_i, c_i \rangle) =$

$$\mathrm{RSV}_j^{i,1} = g^1(d_j, t_i, c_i^1) = \begin{cases} s^1_{\min\{a+\mathcal{B},\mathcal{T}\}} & \text{if } s^1_{\mathcal{T}/2} \leqslant s^1_b \leqslant s^1_a \\ s^1_{\max\{0,a-\mathcal{B}\}} & \text{if } s^1_{\mathcal{T}/2} \leqslant s^1_b \text{ and } s^1_a < s^1_b \\ \mathrm{Neg}(s^1_{\mathrm{Max}\{0,a-\mathcal{B}\}}) & \text{if } s^1_a \leqslant s^1_b < s^1_{\mathcal{T}/2} \\ \mathrm{Neg}(s^1_{\mathrm{Min}\{a+\mathcal{B},\mathcal{T}\}}) & \text{if } s^1_b < s^1_{\mathcal{T}/2} \text{ and } s^1_b < s^1_a \end{cases}$$

such that, (i) $s^1_b = c^1_i$; (ii) s^1_a is the linguistic index term weight obtained as $s^1_a = \text{Label}(F(d_j, t_i))$, being $\text{Label} : [0,1] \rightarrow S^1$ a function that assigns a label in S^1 to a numeric value $r \in [0,1]$ according to the following expression:

$$\text{Label}(r) = \text{Sup}_q\left\{s^1_q \in S^1 : \mu_{s^1_q}(r) = \text{Sup}_v\{\mu_{s^1_v}(r)\}\right\}$$

and (iii) $\mathscr{B}$ is a bonus value that rewards/penalizes the value $\text{RSV}^{i,1}_j$ for the satisfaction/dissatisfaction of request $\langle t_i, c^1_i, c^2_i, c^3_i\rangle$, which can be defined in an independent way, for example as $\mathscr{B} = 1$, or depending on the closeness between $\text{Label}(F(d_j, t_i))$ and c^1_i, for example as $\mathscr{B} = \text{round}\left(\frac{2(|b-a|)}{\mathscr{T}}\right)$.

(3) *Evaluation of atoms with respect to the quantitative semantics*. In this step, documents go on being evaluated with regard to their relevance to individual atoms of the query, but considering the restrictions imposed by the quantitative semantics.

The linguistic quantitative weights are interpreted as follows: when a user establishes a certain number of documents for a term in the query, expressed by a linguistic quantitative weight, then the set of documents to be retrieved must have the minimum number of documents that satisfies the compatibility or membership function associated with the meaning of the label used as linguistic quantitative weight. Furthermore, these documents must be those that better satisfy the threshold restrictions imposed on the term.

We should point out that, in a fuzzy IR context, the use of a threshold semantics implies the establishment of restrictions on the membership function that characterizes the fuzzy set of documents retrieved for an index term, while the use of a quantitative semantics implies the establishment of restrictions on the support of such a fuzzy subset.

Therefore, given an atom $\langle t_i, c^1_i, c^2_i, c^3_i\rangle$ and assuming that $\text{RSV}^{i,1}_j \in S^1$ represents the evaluation according to the symmetrical threshold semantics for d_j, we model the interpretation of a quantitative semantics by means of a linguistic matching function, called g^2, which is defined between the $\text{RSV}^{i,1}_j$ and the linguistic quantitative weight $c^2_i \in S^2$. Then, the evaluation of the atom $\langle t_i, c^1_i, c^2_i, c^3_i\rangle$ with respect to the quantitative semantics associated with c^2_i for a document d_j, called $\text{RSV}^{i,1,2}_j \in S^1$, is obtained by means of the linguistic matching function $g^2 : D \times S^1 \times S^2 \rightarrow S^1$ as follows

$$\text{RSV}^{i,1,2}_j = g^2(\text{RSV}^{i,1}_j, c^2_i, d_j) = \begin{cases} s^1_0 & \text{if } d_j \notin \mathscr{B}^S \\ \text{RSV}^{i,1}_j & \text{if } d_j \in \mathscr{B}^S \end{cases}$$

where $\mathscr{B}^S$ is the set of documents such that $\mathscr{B}^S \subseteq \text{Support}(\mathscr{M})$ where $\mathscr{M} = \{(d_1, \text{RSV}^{i,1}_1), \ldots, (d_m, \text{RSV}^{i,1}_m)\}$, is a fuzzy subset of documents obtained according to the following algorithm:

(1) $K = \#\text{Supp}(\mathscr{M})$
(2) REPEAT

$M^K = \{s_q \in S : \mu_{s_q}(K/m) = \text{Sup}_v\{\mu_{s_v}(K/m)\}\}.$

$s^K = \text{Sup}_q\{s_q \in M^K\}.$

$K = K - 1.$

(3) UNTIL $((c_i^2 \in M^{K+1})$ OR $(c_i^2 \geqslant s^{K+1}))$.
(4) $\mathscr{B}^S = \{d_{\sigma(1)}, \ldots, d_{\sigma(K+1)}\}$, such that $\text{RSV}^{i,1}_{\sigma(h)} \leqslant \text{RSV}^{i,1}_{\sigma(l)}, \forall l \leqslant h$.

According to g^2, the application of the quantitative semantics consists of reducing the number of documents to be considered by the evaluation subsystem for t_i in the later steps. Then, by assigning quantitative weights close to s_0, a user shows his/her preferences by considering the most representative document in $\mathscr{M}$ and by assigning quantitative weights close to $s_{\mathscr{T}}$. He/she does not make a distinction between the documents existing in $\mathscr{M}$.

Remark 1. Although in this step we have worked with different linguistic domains, there has not been however aggregation of multi-granular linguistic information. Consequently, due to this reason it has not been a need to use the label set BLTS.

(4) *Evaluation of subexpressions and modelling of the relative importance semantics*. We consider that the relative importance semantics in a single-term query has no meaning. Then, in this step we have to evaluate the relevance of documents with respect to the subexpressions of queries composed of more than two atoms.

Given a subexpression q_v with $\mathscr{I} \geqslant 2$ atoms, we know that each document d_j presents a partial $\text{RSV}_j^{i,1,2} \in S^1$ with respect to each atom $\langle t_i, c_i^1, c_i^2, c_i^3 \rangle$ of q_v. Then, the evaluation of the relevance of a document d_j with respect to the whole subexpression q_v implies the aggregation of the partial relevance degrees $\{\text{RSV}_j^{i,1,2},\ i = 1, \ldots, \mathscr{I}\}$ weighted by means of the respective relative importance degrees $\{c_i^3 \in S^3,\ i = 1, \ldots, \mathscr{I}\}$. Therefore, as $S^1 \neq S^3$, we have to develop an aggregation procedure of multi-granular linguistic information. As said, to do so, we first choose a label set BLTS to make linguistic information uniform. In this case, $\text{BLTS} = S'$ which is used to assess RSVs (relevance degrees of documents). Then, each linguistic information value is transformed into S' by means of the following transformation function:

Definition 2 [9]. Let $A = \{l_0, \ldots, l_p\}$ and $S' = \{s'_0, \ldots, c'_m\}$ be two label sets, such that $m \geqslant p$. Then, a multi-granularity transformation function, $\tau_{AS'}$ is defined as $\tau_{AS'} : A \to \mathscr{F}(S')$

$$\tau_{AS_T}(l_i) = \{(s'_k, \alpha_k^i)/k \in \{0, \ldots, m\}\}, \quad \forall l_i \in A,$$
$$\alpha_k^i = \max_y \min\{\mu_{l_i}(y), \mu_{c'_k}(y)\}$$

where $\mathscr{F}(S')$ is the set of fuzzy sets defined in S', and $\mu_{l_i}(y)$ and $\mu_{s'_k}(y)$ are the membership functions of the fuzzy sets associated to the terms l_i and s'_k, respectively.

Therefore, the result of $\tau_{AS'}$ for any linguistic value of A is a fuzzy set defined in the BLTS, S'. Using the multi-granularity transformation functions $\tau_{S^1S'}$ and $\tau_{S^3S'}$, we transform the linguistic values $\{\mathrm{RSV}_j^{i,1,2} \in S^1,\ i = 1,\ldots,\mathscr{I}\}$ and $\{c_i^3 \in S^3,\ i = 1,\ldots,\mathscr{I}\}$ into S', respectively. Therefore, the values RSV_j^i and c_i^3 are represented as fuzzy sets defined on S' characterized by the following expressions:

(1) $\tau_{S^1S'}(\mathrm{RSV}_j^{i,1,2}) = [(s'_0, \alpha_0^{ij}),\ldots,(s'_m, \alpha_m^{ij})]$, and

(2) $\tau_{S^2S'}(c_i^3) = [(s'_0, \alpha_0^{i}),\ldots,(s'_m, \alpha_m^{i})]$, respectively.

In each subexpression q_v we find that the atoms can be combined using the AND or OR Boolean connectives, depending on the normal form of the user query. The restrictions imposed by the relative importance weights must be applied in the aggregation operators used to model both connectives. These aggregation operators should guarantee that the more important the query terms, the more influential they are in the determination of the RSVs. To do so, these aggregation operators must carry out two activities [8]: (i) the transformation of the weighted information under the importance degrees by means of a transformation function h; and (ii) the aggregation of the transformed weighted information by means of an aggregation operator of non-weighted information f. As it is known, the choice of h depends upon f. In [20], Yager discussed the effect of the importance degrees on the MAX (used to model the connective OR) and MIN (used to model the connective AND) types of aggregation and suggested a class of functions for importance transformation in both types of aggregation. For the MIN aggregation, he suggested a family of t-conorms acting on the weighted information and the negation of the importance degree, which presents the non-increasing monotonic property in these importance degrees. For the MAX aggregation, he suggested a family of t-norms acting on weighted information and the importance degree, which presents the non-decreasing monotonic property in these importance degrees.

Following the ideas shown above, we use the OWA operators ϕ^1 (with orness$(W) \leqslant 0.5$) and ϕ^2 (with orness$(W) > 0.5$) to model the AND and OR connectives, respectively. Hence, when $h = \phi^1$, $f = \max(\mathrm{Neg}(\mathrm{weight}), \mathrm{value})$, and when $h = \phi^2$, $f = \min(\mathrm{weight}, \mathrm{value})$.

Then, given a document d_j, we evaluate its relevance with respect to a subexpression q_v, called RSV_j^v, as $\mathrm{RSV}_j^v = [(s'_0, \alpha_0^v),\ldots,(s'_m, \alpha_m^v)]$, where

(1) if q_v is a conjunctive subexpression then

$$\alpha_k^v = \phi^1\left(\max((1-\alpha_k^1), \alpha_k^{1j}),\ldots,\max((1-\alpha_k^{\mathscr{I}}), \alpha_k^{\mathscr{I}j})\right) \quad \text{and}$$

(2) if q_v is a disjunctive subexpression then

$$\alpha_k^v = \phi^2\big(\min(\alpha_k^1, \alpha_k^{1j}), \ldots, \min(\alpha_k^{\mathscr{I}}, \alpha_k^{\mathscr{I}j})\big)$$

In such a way, using the OWA operators to model the AND and OR connectives we introduce a soft computing in the evaluation of queries.

(5) *Evaluation of the whole query*. In this step, each document d_j is assigned a total RSV_j with respect to the whole query. The final evaluation of each document is achieved by combining their evaluations with respect to all the subexpressions using, again, the OWA operators ϕ^1 and ϕ^2 to model the AND and OR connectives, respectively.

Then, given a document d_j, we evaluate its relevance with respect to a query q as $\mathrm{RSV}_j = \{(s'_0, \beta_0^j), \ldots, (s'_m, \beta_m^j)\}$, where $\beta_k^j = \phi^1(\alpha_k^1, \ldots, \alpha_k^{\mathscr{V}})$, if q is in CNF, and $\beta_k^j = \phi^2(\alpha_k^1, \ldots, \alpha_k^{\mathscr{V}})$, if q is in DNF, with $\mathscr{V}$ standing for the number of subexpressions in q.

Remark 2. *On the NOT Operator*. We should note that, if a query is in CNF or DNF form, we have to define the negation operator only at the level of single atoms. This simplifies the definition of the NOT operator. As was done in [10], the evaluation of document d_j for a negated weighted atom $\langle \neg(t_i), c_i^1, c_i^2, c_i^3 \rangle$ is obtained from the negation of the index term weight $F(t_i, d_j)$. This means to calculate g^1 from the linguistic value $\mathrm{Label}(1 - F(t_i, d_j))$.

(6) *Presenting the output of the IRS*. At the end of the evaluation of a user query q, each document d_j is characterized by RSV_j which is a fuzzy set defined on S'. Of course, an answer of an IRS where the relevance of each document is expressed by means of a fuzzy set is not easy to understand, and neither to manage. To overcome this problem we present the output of our IRS by means of ordered linguistic relevance classes, as in [10,11]. Furthermore, in each relevance class we establish a ranking of the documents using a confidence degree associated to each document.

To do so, we calculate a label $s^j \in S'$ for each document d_j, which represents its linguistic relevance class. We design an easy linguistic approximation process in S' using a similarity measure, e.g., the Euclidean distance. Each label $s'_k \in S'$ is represented as a fuzzy set defined in S', i.e., $\{(s'_0, 0), \ldots, (s'_k, 1), \ldots, (s'_m, 0)\}$. Then, we calculate s^j as

$$s^j = \mathrm{MAX}\{s'_l \setminus \mathrm{Conf}(s'_l, \mathrm{RSV}_j) = \min_k\{\mathrm{Conf}(s'_k, \mathrm{RSV}_j)\}\}$$

where $\mathrm{Conf}(s'_k, \mathrm{RSV}_j) \in [0, 1]$ is the confidence degree associated to d_j defined as $\mathrm{Conf}(s'_k, \mathrm{RSV}_j) = \sqrt{\sum_{i=0}^{k-1}(\beta_i^j)^2 + (\beta_k^j - 1)^2 + \sum_{i=k+1}^{m}(\beta_i^j)^2}$.

3.3. Example of application

In this section, we present an example of performance of the proposed IRS let us suppose a small database containing a set of seven documents $D = \{d_1, \ldots, d_7\}$, represented by means of a set of 10 index terms $T = \{t_1, \ldots, t_{10}\}$. Documents are indexed by means of an indexing function F, which assigns the following weights to each of them:

$$d_1 = 0.7/t_5 + 0.4/t_6 + 1/t_7$$
$$d_2 = 1/t_4 + 0.6/t_5 + 0.8/t_6 + 0.9/t_7$$
$$d_3 = 0.5/t_2 + 1/t_3 + 0.8/t_4$$
$$d_4 = 0.9/t_4 + 0.5/t_6 + 1/t_7$$
$$d_5 = 0.7/t_3 + 1/t_4 + 0.4/t_5 + 0.8/t_9 + 0.6/t_10$$
$$d_6 = 1/t_5 + 0.99/t_6 + 0.8/t_7$$
$$d_7 = 0.8/t_5 + 0.02/t_6 + 0.8/t_7 + 0.9/t_8$$

In the same way, let us suppose the following four label sets with different cardinality and semantics to assess threshold weights, quantitative weights, relative importance weights and RSVs, respectively:

1. $S^1 = \{MI = (0,0,0,0.25), VL = (0.25,0.25,0.25,0.15),$
   $L = (0.4,0.4,0.15,0.1), M = (0.5,0.5,0.1,0.1),$
   $MU = (0.6,0.6,0.1,0.15), VM = (0.75,0.75,0.15,0.25),$
   $MA = (1,1,0.25,0)\}$.
2. $S^2 = \{N = (0,0,0,0.25), L = (0.25,0.25,0.25,0.25),$
   $M = (0.5,0.5,0.25,0.25), H = (0.75,0.75,0.25,0.25),$
   $T = (1,1,0.25,0)\}$.
3. $S^3 = \{N = (0,0,0,0), EL = (0.01,0.02,0.01,0.05),$
   $VL = (0.1,0.18,0.06,0.05), L = (0.22,0.36,0.05,0.06),$
   $M = (0.41,0.58,0.09,0.07), H = (0.63,0.80,0.05,0.06),$
   $VH = (0.78,0.92,0.06,0.05), EH = (0.98,0.99,0.05,0.01),$
   $T = (1,1,0,0)\}$.
4. $S' = \{N = (0,0,0,0), EL = (0.01,0.02,0.01,0.05),$
   $VL = (0.1,0.18,0.06,0.05), ML = (0.22,0.30,0.05,0.06),$
   $L = (0.31,0.36,0.05,0.06), M = (0.41,0.58,0.09,0.07),$
   $H = (0.63,0.70,0.05,0.06), MH = (0.71,0.80,0.05,0.06),$
   $VH = (0.78,0.92,0.06,0.05), EH = (0.98,0.99,0.05,0.01),$
   $T = (1,1,0,0)\}$.

Finally, consider that a user formulates the following query $q = ((t_5, MU, L, VH) \wedge (t_6, L, L, VL)) \vee (t_7, MU, L, H)$. The process applied by the IRS is shown as follows:

(1) *Preprocessing of the query*. The query q is in DNF, but it presents one subexpression with only one atom. Therefore, q must be preprocessed and transformed into normal form with every subexpression having more than two atoms. Then, q is transformed into the following equivalent query $q = ((t_5, MU, L, VH) \vee (t_7, MU, L, H)) \wedge ((t_6, L, L, VL) \vee (t_7, MU, L, H))$, which is expressed in CNF.

(2) *Evaluation of atoms with respect to the symmetrical threshold semantics*. In this step we obtain the documents represented in a linguistic form using the translation function Label:

$$d_1 = VM/t_5 + L/t_6 + MA/t_7$$
$$d_2 = MA/t_4 + MU/t_5 + VM/t_6 + MA/t_7$$
$$d_3 = M/t_2 + MA/t_3 + VM/t_4$$
$$d_4 = MA/t_4 + M/t_6 + MA/t_7$$
$$d_5 = VM/t_3 + MA/t_4 + L/t_5 + VM/t_9 + MU/t_{10}$$
$$d_6 = MA/t_5 + MA/t_6 + VM/t_7$$
$$d_7 = VM/t_5 + MI/t_6 + VM/t_7 + MA/t_8$$

Let us set the sensitivity parameter $\mathscr{B} = \text{round}\left(\frac{2(|b-a|)}{\mathscr{T}}\right)$. Then, the evaluations of atoms according to the symmetrical threshold semantics modelled by means of the function g^1 are the following:

$$\{\text{RSV}_1^{5,1} = VM, \text{RSV}_2^{5,1} = MU, \text{RSV}_5^{5,1} = VL, \text{RSV}_6^{5,1} = MA, \text{RSV}_7^{5,1} = VM\}$$
$$\{\text{RSV}_1^{6,1} = MU, \text{RSV}_2^{6,1} = MI, \text{RSV}_4^{6,1} = M, \text{RSV}_6^{6,1} = MI, \text{RSV}_7^{6,1} = MA\}$$
$$\{\text{RSV}_1^{7,1} = MA, \text{RSV}_2^{7,1} = MA, \text{RSV}_4^{7,1} = MA, \text{RSV}_6^{7,1} = VM, \text{RSV}_7^{7,1} = VM\}$$

(3) *Evaluation of atoms with respect to the quantitative semantics*. The evaluation of the atoms of the query according to the quantitative semantics modeled by g^2 are:

$$\{\text{RSV}_1^{5,1,2} = VM, \text{RSV}_6^{5,1,2} = MA\}$$
$$\{\text{RSV}_1^{6,1,2} = MU, \text{RSV}_7^{6,1,2} = MA\}$$
$$\{\text{RSV}_1^{7,1,2} = MA, \text{RSV}_2^{7,1,2} = MA\}$$

We should note that the quantitative semantics decreases the number of documents associated to be considered in each query term.

(4) *Evaluation of subexpressions and modelling the relative importance semantics*. We choose the set of eleven labels S' as BLTS to make the linguistic information uniform.

The results of the transformation functions $\tau_{S^1S'}$ and $\tau_{S^3S'}$ applied on $\mathrm{RSV}_j^{i,1,2}$ and on the relative importance degrees of terms c_i^3 are respectively as follows:

$$\tau_{S^1S'}(MU) = \{(s_0,0),(s_1,0),(s_2,0),(s_3,0),(s_4,0),(s_5,0.88),(s_6,0.85),(s_7,0.45),(s_8,0.14),(s_9,0),(s_{10},0)\}$$
$$\tau_{S^1S'}(VM) = \{(s_0,0),(s_1,0),(s_2,0),(s_3,0),(s_4,0),(s_5,0.23),(s_6,0.76),(s_7,1.0),(s_8,0.90),(s_9,0.23),(s_{10},0)\}$$
$$\tau_{S^1S'}(MA) = \{(s_0,0),(s_1,0),(s_2,0),(s_3,0),(s_4,0),(s_5,0),(s_6,0.032),(s_7,0.35),(s_8,0.73),(s_9,0.96),(s_{10},1)\}$$
$$\tau_{S^1S'}(N) = \{(s_0,1),(s_1,0.96),(s_2,0.68),(s_3,0.27),(s_4,0),(s_5,0),(s_6,0),(s_7,0),(s_8,0),(s_9,0),(s_{10},0)\}$$

and

$$\tau_{S^3S'}(VH) = \{(s_0,0),(s_1,0),(s_2,0),(s_3,0),(s_4,0),(s_5,0),(s_6,0.33),(s_7,1.0),(s_8,1.0),(s_9,0.4),(s_{10},0)\}$$
$$\tau_{S^3S'}(H) = \{(s_0,0),(s_1,0),(s_2,0),(s_3,0),(s_4,0),(s_5,0.58),(s_6,1.0),(s_7,1.0),(s_8,1.0),(s_9,0),(s_{10},0)\}$$
$$\tau_{S^3S'}(VL) = \{(s_0,0),(s_1,0.27),(s_2,1),(s_3,0.6),(s_4,0),(s_5,0),(s_6,0),(s_7,0),(s_8,0),(s_9,0),(s_{10},0)\}$$

The query q' has two subexpressions and each of the presents two atoms, $q'^1 = (t_5, MU, L, VH) \vee (t_7, MU, L, H)$ and $q'^2 = (t_6, L, L, VL) \vee (t_7, MU, L, H)$. Each subexpression is in disjunctive form, and thus we must use an OWA operator ϕ^2 with orness$(W) > 0.5$ (for example, with $(W = [1,0])$).

$$\mathrm{RSV}_1^1 = \{(s_0,0),(s_1,0),(s_2,0),(s_3,0),(s_4,0),(s_5,0),(s_6,0.33),(s_7,1.0),(s_8,0.9),(s_9,0.23),(s_{10},0)\}$$
$$\mathrm{RSV}_2^1 = \{(s_0,0),(s_1,0),(s_2,0),(s_3,0),(s_4,0),(s_5,0),(s_6,0.032),(s_7,0.35),(s_8,0.73),(s_9,0),(s_{10},0)\}$$
$$\mathrm{RSV}_6^1 = \{(s_0,0),(s_1,0),(s_2,0),(s_3,0),(s_4,0),(s_5,0),(s_6,0.032),(s_7,0.35),(s_8,0.73),(s_9,0.4),(s_{10},0)\}$$
$$\mathrm{RSV}_1^2 = \{(s_0,0),(s_1,0),(s_2,0),(s_3,0),(s_4,0),(s_5,0),(s_6,0.032),(s_7,0.35),(s_8,0.73),(s_9,0),(s_{10},0)\}$$
$$\mathrm{RSV}_2^2 = \{(s_0,0),(s_1,0.27),(s_2,0.67),(s_3,0.26),(s_4,0),(s_5,0),(s_6,0.032),(s_7,0.35),(s_8,0.73),(s_9,0),(s_{10},0)\}$$
$$\mathrm{RSV}_7^2 = \{(s_0,0),(s_1,0),(s_2,0),(s_3,0),(s_4,0),(s_5,0),(s_6,0),(s_7,0),(s_8,0),(s_9,0),(s_{10},0)\}$$

(5) *Evaluation of the whole query*. We obtain the document evaluation with respect to the whole query using an OWA operator ϕ^1 with orness$(W) < 0.5$ (e.g. with $(W = [0.4, 0.6])$).

$$\begin{aligned}\mathrm{RSV}_1 = \{&(s_0,0),(s_1,0),(s_2,0),(s_3,0),(s_4,0),(s_5,0),(s_6,0.15),\\ &(s_7,0.61),(s_8,0.8),(s_9,0.92,(s_{10},0)\}\\ \mathrm{RSV}_2 = \{&(s_0,0),(s_1,0.11),(s_2,0.27),(s_3,0.10),(s_4,0),(s_5,0),(s_6,0.032),\\ &(s_7,0.35),(s_8,0.73),(s_9,0),(s_{10},0)\}\\ \mathrm{RSV}_6 = \{&(s_0,0),(s_1,0),(s_2,0),(s_3,0),(s_4,0),(s_5,0),(s_6,0.013),\\ &(s_7,0.14),(s_8,0.3),(s_9,0.16),(s_{10},0)\}\\ \mathrm{RSV}_7 = \{&(s_0,0),(s_1,0),(s_2,0),(s_3,0),(s_4,0),(s_5,0),(s_6,0),\\ &(s_7,0),(s_8,0),(s_9,0),(s_{10},0)\}\end{aligned}$$

Finally, we calculate a label $s_j \in S'$ for each document d_j, which represents its linguistic relevance class, obtaining the final system output:

$$\{(d_1, MH),(d_2, MH),(d_6, MH)\}$$

4. Concluding remarks

We have presented a model of IRS based on multi-granular linguistic information. In such a way, we get that the interchange of information between the user and the IRS is carried out in a more natural way, improving the IRS-user interaction.

In the future, we think on applying this method in the multi-weighted query languages where the different elements of a query, terms subexpressions, connectives, and the whole query itself, can be weighted.

References

[1] R. Baeza-Yates, B. Ribeiro-Neto, Modern Information Retrieval, Addison, 1999.

[2] P.P. Bonissone, K.S. Decker, Selecting uncertainty calculi and granularity: An experiment in trading-off precision and complexity, in: L.H. Kanal, J.F. Lemmer (Eds.), Uncertainty in Artificial Intelligence, North-Holland, 1986, pp. 217–247.

[3] A. Bookstein, Fuzzy request: An approach to weighted Boolean searches, Journal of the American Society for Information Science 31 (1980) 240–247.

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

[5] G. Bordogna, G. Pasi, An ordinal information retrieval model, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 9 (1) (2001) 63–75.

[6] D. Buell, D.H. Kraft, Threshold values and Boolean retrieval systems, Information Processing & Management 17 (1981) 127–136.

[7] D. Buell, D.H. Kraft, A model for a weighted retrieval system, Journal of the American Society for Information Science 32 (1981) 211–216.
[8] F. Herrera, E. Herrera-Viedma, Aggregation operators for linguistic weighted information, IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems 27 (1997) 646–656.
[9] F. Herrera, E. Herrera-Viedma, L. Martínez, A fusion approach for managing multi-granularity linguistic term sets in decision making, Fuzzy Sets and Systems 114 (2000) 43–58.
[10] E. Herrera-Viedma, Modeling the retrieval process for 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.
[11] 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 (1) (2001) 77–88.
[12] R.R. Korfhage, Information Storage and Retrieval, Wiley Computer Publishing, New York, 1997.
[13] D.H. Kraft, D.A. Buell, Fuzzy sets and generalized Boolean retrieval systems, International Journal of Man–Machine Studies 19 (1983) 45–56.
[14] D.H. Kraft, G. Bordogna, G. Pasi, An extended fuzzy linguistic approach to generalize Boolean information retrieval, Information Sciences 2 (1994) 119–134.
[15] G. Salton, M.H. McGill, Introduction to Modern Information Retrieval, McGraw-Hill, New York, 1983.
[16] G. Salton, Automatic text processing: The transformation, analysis of information by computer, Addison Wesley Series in Computer Science, Addison-Wesley, Reading, MA, 1989.
[17] V. Torra, Aggregation of linguistic labels when semantics is based on antonyms, International Journal of Intelligent Systems 16 (2001) 513–524.
[18] C.J. Van Rijsbergen, Information Retrieval, second ed., Butterworth, 1979.
[19] W.G. Waller, D.H. Kraft, A mathematical model of a weighted Boolean retrieval system, Information Processing & Management 15 (1979) 235–245.
[20] R.R. Yager, A note on weighted queries in information retrieval systems, Journal of the American Society for Information Science 38 (1987) 23–24.
[21] R.R. Yager, On ordered weighted averaging aggregation operators in multicriteria decision making, IEEE Transactions on Systems, Man, and Cybernetics 18 (1988) 183–190.
[22] L.A. Zadeh, The concept of a linguistic variable and its applications to approximate reasoning. Part I, Information Sciences 8 (1975) 199–249, Part II, Information Sciences 8 (1975) 301–357, Part III, Information Sciences 9 (1975) 43–80.