An Information Retrieval System with Weighted Querying Based on Multi-Granular Linguistic Information

E. Herrera-Viedma, F. Herrera, O. Cordón, M. Luque, A.G. López
Dept. of Computer Science and A.I.
University of Granada, 18071 - Granada
viedma, herrera, ocordon@decsai.ugr.es

Abstract

In this contribution, an information retrieval system (IRS) based on fuzzy multi-granular linguistic information is proposed. The user queries and IRS responses are modelled using different linguistic domains or label sets with different cardinalities and/or semantics. We present a method to process the multi-granular linguistic information in the retrieval activity of the IRS. The system accepts Boolean queries whose terms can be simultaneously weighted by means of ordinal linguistic values according to two semantics: a symmetrical threshold semantics and an importance semantics. In both semantics the linguistic weights are represented by the linguistic variable "Importance", but assessed on different label sets S^1 and S^2, respectively. The IRS evaluates the weighted queries and obtains the linguistic retrieval status values (RSV) of documents represented by a linguistic variable "Relevance" expressed also on a different label set S'. The advantage of this linguistic IRS with respect to others is that the use of the multi-granular linguistic information facilitates the expression of information needs and improves the latter issue.

Keywords: linguistic modelling, information retrieval, multi-granular linguistic information.

1 Introduction

Information Retrieval (IR) is a research field refered the storage and retrieval of textual information [7, 9]. An important question in the IRSs is how to facilitate the IRS-user interaction. The use of linguistic variables [13] to represent the input and output information in the retrieval process of IRSs considerably improves the IRS-user interaction (see [2, 5, 6, 8]).

Usually, 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 than the above linguistic variables represent different concepts, it seems necessary to use different label sets in their linguistic modelling, i.e., to apply a multi-granular linguistic modelling [4]. This means to use label sets with different granularity and/or semantics to represent the different information in the retrieval process.

The aim of this contribution is to present a linguistic weighted IRS which is designed using multi-granular linguistic information. The weighted Boolean queries and the RSVs of documents are assessed on label sets with different granularity and/or semantics, called S^1 and S^2, respectively. The query terms can be simultaneously weighted according to two different semantics: a threshold semantics and an importance semantic. The Boolean operators AND and OR are modelled by means of the OWA aggregation operator [11]. 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 different label set S'.

In order to do that, the contribution is structured 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. And finally, some concluding remarks are pointed out.

2 Preliminaries

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* [13]. 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 [3, 5]. An ordinal fuzzy linguistic approach is defined by considering a finite and totally ordered label set $S = \{s_i\}, i \in \{0, \ldots, \mathcal{T}\}$ in the usual sense ($s_i \geq s_j$ if $i \geq j$) and with odd cardinality (7 or 9 labels) as in [1]. The mid term representing an assessment of "approximately 0.5" and the rest of the terms being placed symmetrically around it [1]. The semantics of the label set is established from the ordered structure of the label set by considering that each label for the pair $(s_i, s_{\mathcal{T}-i})$ is equally informative. The semantics of the labels is given by fuzzy numbers defined on the [0,1] interval, which 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) $\mathrm{Neg}(s_i) = s_j,\ j = \mathcal{T} - i$.

2) $\mathrm{MAX}(s_i, s_j) = s_i$ if $s_i \geq s_j$.

3) $\mathrm{MIN}(s_i, s_j) = s_i$ if $s_i \leq s_j$.

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 information. The cardinality of S must be small enough so as not to impose useless precision on 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. Then, we need tools of management of multi-granular linguistic information to model these situations. Different proposals can be found in [4, 10].

2.3 The OWA Operator

The Ordered Weighted Averaging (OWA) is an aggregation operator of information which acts taking into account the order of assessments. It was defined by Yager in [11] as follows:

Definition 1 [11]. *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*

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

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 $a_{\sigma(j)} \leq a_{\sigma(i)}\ \forall\ i \leq j$, with σ being a permutation over the set of labels A.

The OWA operator is an "or-and" operator [11]. This property allows the OWA operator to carry out a soft computing in the modelling of MAX and MIN operators. We use this good characteristic in our linguistic IRS to evaluate the Boolean queries.

In order to classify OWA operators in regard to their localization between "and" and "or", Yager [11] introduced a *orness measure*, associated with any vector W as follows

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

Fixed a weighting vector W, then the nearer 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 to zero. Generally, an OWA operator with much on nonzero weights near the top will be an orlike operator (orness(W) ≥ 0.5), and when much of the weights are nonzero near the bottom, the OWA operator will be an andlike.

3 The IRS Based on Multi-Granular Linguistic Information

In this section we shall present an IRS that accepts linguistic weighted Boolean queries expressed using multi-granular linguistic information and models the Boolean operators in a flexible way.

We assume 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 : DxT $\rightarrow [0,1]$ is defined, called *index term weight*. 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 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 weigh index terms according to their significance in describing the content of a document in order to improve the document retrieval.

3.1 Queries Formulated with Multi-Granular Linguistic Weights

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. We assume that each term in a query can be simultaneously weighted by means of several weights [5, 6]. To facilitate users the expression of their queries, we consider that a term of a query can be weighted by means of two weights associated to the following semantics [6]:

1. *Symmetrical threshold semantics.* By associating threshold weights to terms in a query, the user is asking to see all documents sufficiently about the topics represented by such terms. 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. *Importance semantics.* This semantics defines term weights as a measure of the relative importance of each term of a query with respect to the remainders. By associating relative importance weights to terms in a query, the user is asking to see all documents whose content represents more the concept 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.

As in [2, 5, 6], we use the linguistic variable *"Importance"* to model both 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 importance weight of value "$High$" means that user requires that the meaning of t_i must have a high importance value in the computation of the set of retrieved documents. Consequently, given that both semantics present a different interpretation, 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 of both semantics, called S^1 and S^2, respectively.

Then, a query is any legitimate Boolean expression whose atomic components (atoms) are 3-tuples $< t_i, c_i^1, c_i^2 >$ belonging to the set, $\mathrm{T} \times S^1 \times S^2$; $t_i \in \mathrm{T}$, $c_i^1 \in S^1$ is a value of the linguistic variable *"Importance"*, modelling the threshold semantics, and $c_i^2 \in S^2$ is a value of the linguistic variable *"Importance"*, modelling the importance semantics. Therefore, the set of legitimate linguistic weighted Boolean queries Q is defined by the following syntactic rules:

1. $\forall q =< t_i, c_i^1, c_i^2 >\in \mathrm{T} \times S^1 \times S^2 \to q \in \mathrm{Q}$. These queries are called atoms.

2. $\forall q, p \in \mathrm{Q} \to q \wedge p \in \mathrm{Q}$.

3. $\forall q, p \in \mathrm{Q} \to q \vee p \in \mathrm{Q}$.

4. $\forall q \in \mathrm{Q} \to \neg(q) \in \mathrm{Q}$.

5. All legitimate linguistic weighted Boolean queries $q \in \mathrm{Q}$ are only those obtained by applying rules 1-4.

3.2 Evaluating Multi-Granular Linguistic Weighted Queries

Usually, the 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 Boolean combinations of 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 linguistic queries. This method evaluates documents in terms of their relevance to queries by satisfactorily supporting the two semantics associated to query weights and simultaneously by managing the multi-granular linguistic weights. Furthermore, given that the concept of relevance is different from the concept of importance, we use to provide the relevance values of documents a label set S' different from those used to express the queries (S^1 and S^2).

To manage the multi-granular linguistic weights of queries we follow a procedure for decision making presented in [4]. The idea involves making uniform the multi-granular linguistic information before processing queries. To do that, we have to choose a label set as the uniform representation base, called *basic linguistic term set (BLTS)*, and 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. It must be the label set used to express the output of the IRS (relevance degrees of documents), i.e., S'.

Then, 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 and put in either conjunctive normal form (CNF) or disjunctive normal form (DNF), in such a way that all its subexpressions must have more than two atoms.

2.- Evaluation of atoms with respect to the symmetrical threshold semantics.

We only consider the terms appearing in the queries, and thus the absent terms are not considered in the evaluation. According to a symmetrical threshold semantics, a user may search for documents with a minimally acceptable presence of one term in their representations (as in [8]) or documents with a maximally acceptable presence of one term in their representations. 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*, the user 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 by a term t_i is (are) with the value *Low Importance*, the user would not reject a document with an F value less than *Low*. Given a request $< t_i, c_i > \in \mathrm{T} \times S^1$, this means that the query weights that imply the presence of a term in a document $c_i \geq s^1_{\mathcal{T}/2}$ (e.g. *High, Very High*) must be treated differently to the query weights that imply the absence of one term in a document $c_i < s^1_{\mathcal{T}/2}$ (e.g. *Low, Very Low*). Then, if $c_i \geq s^1_{\mathcal{T}/2}$ the request $< t_i, c_i >$, is synonymous with the request $< t_i, at\ least\ c_i >$, which expresses the fact that the desired documents are those having F values as high as possible; and if $c_i < s^1_{\mathcal{T}/2}$ is synonymous with the request $< t_i, at\ most\ c_i >$, 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 $g : D \times \mathrm{T} \times S^1 \rightarrow S^1$ [6]. Given an atom $< t_i, c_i > \in \mathrm{T} \times S^1$ and a document $d_j \in D$, g establishes how well the index term weight $F(d_j, t_i)$ of document d_j satisfies the request expressed by the linguistic weight c_i of atom $< t_i, c_i >$ according to the following expression: $g(d_j, < t_i, c_i >) =$

$$\begin{cases} s^1_{min\{a+\mathcal{B},\mathcal{T}\}} & \text{if } s^1_{\mathcal{T}/2} \leq s^1_b \leq s^1_a \\ s^1_{max\{0,a-\mathcal{B}\}} & \text{if } s^1_{\mathcal{T}/2} \leq s^1_b \text{ and } s^1_a < s^1_b \\ Neg(s^1_{Max\{0,a-\mathcal{B}\}}) & \text{if } s^1_a \leq s^1_b < s^1_{\mathcal{T}/2} \\ Neg(s^1_{Min\{a+\mathcal{B},\mathcal{T}\}}) & \text{if } s^1_b < s^1_{\mathcal{T}/2} \text{ and } s^1_b < s^1_a \end{cases}$$

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

$$Label(r) = Sup_q\{s^1_q \in S^1 : \mu_{s^1_q}(r) = Sup_v\{\mu_{s^1_v}(r)\}\};$$

and (iii) $\mathcal{B}$ is a bonus value that rewards/penalizes the partial RSV of d_j for the satisfaction/dissatisfaction of request $< t_i, c_i >$, which can be defined in an independent way, for example as $\mathcal{B} = 1$, or depending on the closeness between $Label(F(d_j, t_i))$ and c_i, for example as $\mathcal{B} = round(\frac{2(|b-a|)}{\mathcal{T}})$.

3.- Evaluation of subexpressions and modelling of the importance semantics

We consider that the 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 with more than two atoms.

Given a subexpression q_v with $\mathcal{I} \geq 2$ atoms, we know that each document d_j presents a partial RSV with respect to each atom $< t_i, c^1_i, c^2_i >$ of q_v, called RSV^i_j, which is assessed in S^1 and is obtained by the matching function g. 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 $\{RSV^i_j,\ i = 1, \ldots, \mathcal{I}\}$ weighted by means of the respective relative importance degrees $\{c^2_i \in S^2,\ i = 1, \ldots, \mathcal{I}\}$. Therefore, we have to develop an aggregation procedure of multi-granular linguistic information, given that the linguistic expression domains of RSV^i_j and c^2_i are different. As was aforementioned, we solve this problem using the procedure proposed in [4].

Firstly, we choose a label set BLTS to make linguistic information uniform. As said, in this case, this label set BLTS is that used to assess the relevance degrees, S'. Then, each linguistic information value is transformed into S' by means of the following transformation function:

Definition 2.-[4] *Let $A = \{l_0, \ldots, l_p\}$ and $S' = \{s'_0, \ldots, c'_m\}$ be two label sets, such that, $m \geq p$. Then, a multi-granularity transformation function, $\tau_{AS'}$ is defined as $\tau_{AS'} : A \longrightarrow \mathcal{F}(S')$*

$$\tau_{AS_T}(l_i) = \{(s'_k, \alpha^i_k)\ /k \in \{0, \ldots, m\}\},\ \forall l_i \in A$$

$$\alpha^i_k = \max_y \min\{\mu_{l_i}(y), \mu_{c'_k}(y)\}$$

where $\mathcal{F}(S')$ 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 $\tau_{S^2S'}$ we transform the linguistic values $\{RSV^i_j \in S^1,\ i = 1, \ldots, \mathcal{I}\}$ and $\{c^2_i \in S^2,\ i = 1, \ldots, \mathcal{I}\}$ into S', respectively. Therefore, the values RSV^i_j and c^2_i are represented as fuzzy sets defined on S' characterized by the following expressions $\tau_{S^1 S'}(RSV_j^i) = [(s'_0, \alpha_0^{ij}), \ldots, (s'_m, \alpha_m^{ij})]$ and $\tau_{S^2 S'}(c_i^2) = [(s'_0, \alpha_0^i), \ldots, (s'_m, \alpha_m^i)]$, 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 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 [3]: 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 [12], 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.

We use OWA operators ϕ^1 (with orness(W)≤ 0.5) and ϕ^2 (with orness(W)> 0.5) to model the AND and OR connectives, respectively. Then, according to above ideas, given a document d_j we evaluate its relevance with respect to a subexpression q_v, called RSV_j^v, as

$$RSV_j^v = [(s'_0, \alpha_0^v), \ldots, (s'_m, \alpha_m^v)],$$

where $\alpha_k^v = \phi^1(max((1-\alpha_k^1), \alpha_k^{1j}), \ldots, max((1-\alpha_k^{\mathcal{I}}), \alpha_k^{\mathcal{I}j}))$ if q_v is a conjunctive subexpression, and $\alpha_k^v = \phi^2(min(\alpha_k^1, \alpha_k^{1j}), \ldots, min(\alpha_k^{\mathcal{I}}, \alpha_k^{\mathcal{I}j}))$, if q_v is a disjunctive one. In such a way, using the OWA operator to model the AND and OR connectives we introduce a soft computing in the evaluation of queries.

4.- 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 documents is achieved combining their evaluations with respect to all subexpressions using again the OWA operators ϕ^1 and ϕ^2 to model the AND and OR connectives.

Then, given a document d_j, we evaluate its relevance with respect to a query q as

$$RSV_j = \{(s'_0, \beta_0^j), \ldots, (s'_m, \beta_m^j)\},$$

where $\beta_k^j = \phi^1(\alpha_k^1, \ldots, \alpha_k^{\mathcal{V}})$, if q is in CNF, and $\beta_k^j = \phi^2(\alpha_k^1, \ldots, \alpha_k^{\mathcal{V}})$, if q is in DNF, and $\mathcal{V}$ stands for the number of subexpressions of q.

Remark 1: *On the NOT Operator.* We note that if a query is in CNF or DNF form, we have to define the negator operator only at the level of single atoms. This simplifies the definition of the NOT operator. As was done in [5, 6], the evaluation of document d_j for a negated weighted atom $< \neg(t_i), c_i^1, c_i^2 >$ is obtained from the negation of the index term weight $F(t_i, d_j)$. This means to calculate g from the linguistic value $Label(1 - F(t_i, d_j))$.

3.3 Presenting the Output of IRS

At the end of the evaluation of a user query q, each document d_j is characterized by a RSV_j, which is a fuzzy set defined in S' and obtained as above. 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 to manage. Usually, the answer of the IRS is ordered according to some criterion.

To overcome this problem we present the output of our IRS by means of ordered linguistic relevance classes, as in [5, 6]. 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 for each document d_j a label $s^j \in S'$ 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

$$s^j = MAX\{s'_l | Conf(s'_l, RSV_j) = min_k\{Conf(s'_k, RSV_j)\}\},$$

where $Conf(s'_k, RSV_j) \in [0, 1]$ is the confidence degree associated to d_j defined as

$$Conf(s'_k, RSV_j) = \sqrt{\sum_{i=0}^{k-1} (\beta_i^j)^2 + (\beta_k^j - 1)^2 + \sum_{i=k+1}^{m} (\beta_i^j)^2}.$$

4 Concluding Remarks

We have presented a linguistic IRS that is able to manage multi-granular linguistic information to express queries. 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 to apply this method in the multi-weighted query languages [5].

References

[1] P.P. Bonissone and K.S. Decker, Selecting Uncertainty Calculi and Granularity: An Experiment in Trading-off Precision and Complexity, in: L.H. Kanal and J.F. Lemmer, Eds., *Uncertainty in Artificial Intelligence* (North-Holland, 1986) 217-247.

[2] G. Bordogna and 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.

[3] F. Herrera and E. Herrera-Viedma, Aggregation Operators for Linguistic Weighted Information, *IEEE Transactions on Systems, Man, and Cybernetics* **27** (1997) 646-656.

[4] F. Herrera, E. Herrera-Viedma and L. Martnez, A Fusion Approach for Managing Multi-Granularity Linguistic Term Sets in Decision Making, *Fuzzy Sets and Systems,* **114** (2000) 43-58.

[5] 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.

[6] E. Herrera-Viedma, An Information Retrieval System with Ordinal Linguistic Weighted Queries Based on Two Weighting Elements, *Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems,* **9** (2001) 77-88.

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

[8] D.H. Kraft, G. Bordogna and G. Pasi, An Extended Fuzzy Linguistic Approach to Generalize Boolean Information Retrieval, *Information Sciences* **2** (1994) 119-134.

[9] G. Salton, *Automatic Text Processing: The Transformation, Analysis of Information by Computer* (Addison Wesley Series in Computer Science. Reading, MA: Addison-Wesley, 1989).

[10] V. Torra, Aggregation of Linguistic Labels when Semantics is Based on Antonyms, *International Journal of Intelligent Systems,* **16** (2001) 513-524.

[11] R.R. Yager, On Ordered Weighted Averaging Aggregation Operators in Multicriteria Decision Making, *IEEE Transactions on Systems, Man, and Cybernetics,* **18** (1988) 183-190.

[12] R.R. Yager, A Note on Weighted Queries in Information Retrieval Systems, *Journal of the American Society for Information Science,* **38** (1987) 23-24.

[13] 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.