

Linguistic Information Processing Based on Aggregation Operator over the Internet

Li Yan, Yi Qin, Zheng Pei

*School of Mathematics & Computer Engineering, Xihua University, Chengdu, Sichuan,
610039, China
pqyz@263.net*

doi: 10.4156/jcit.vol5.issue3.14

Abstract

Much information over the Internet is expressed by natural languages. The management of linguistic information involves an operation of comparison and aggregation. In this paper, based on linguistic aggregation operator F_{lowa} , we discussed linguistic information processing in a multi-agent system over the Internet, in which, aggregating linguistic information and selecting better Web pages (or documents) that fulfill the user's query are discussed.

Keywords: *Linguistic Information Processing, Linguistic Aggregation Operator, Multi-Agent System.*

1. Introduction

By means of linguistic variables, the fuzzy linguistic approach as an approximate technique is proposed, in which, qualitative aspects are represented by linguistic values [1]. In many cases, human knowledge is directly expressed by natural language, and people use those natural language to think, to reason, to deduce conclusions, and to make decisions. To handle those natural language, Zadeh presents computing with words [2, 3]. Based on the semantically ordering relation of linguistic hedges, hedge algebra (HA) and RH-algebra for computing with words are proposed [4], lattice implication algebra of linguistic value "Truth" is discussed in [5]. In [6, 7], the evaluating linguistic predications are discussed. A framework of computing with words based on the mass assignment theory and probability theory is presented in [8, 9]. From the practical point of view, the 2-tuple linguistic computational model and aggregation operators for linguistic weighted information are proposed in [10, 11]. In [12], uncertain linguistic ordered weighted averaging (ULOWA) operator and uncertain linguistic hybrid aggregation (ULHA) operator are proposed and their applications are discussed. Based on fuzziness of the index of linguistic label and OWA operator [13], a kind of linguistic aggregation operator F_{lowa} is presented, extracting linguistic rules from numerical database and aggregating linguistic expressions based on F_{lowa} are discussed [14].

More and more people consider Internet as an important source of information. From users' point of view, most information on Internet is expressed by natural language. Naturally, users expect that retrieving information by Internet is based on natural language. However, it is difficult to obtain information which users need when key words expressed by fuzzy linguistic values are used in retrieving information. This means that tools for an effective and efficient handling of imprecise elements of natural language are needed in search engine. The management of linguistic information involves an operation of comparison and aggregation. Many interesting results on linguistic information processing have been done [14-20].

In the next Section, the indexes of linguistic values are analyzed. In Section 3, a new linguistic aggregation operator F_{lowa} based on fuzzy number index of linguistic value is reviewed, its properties are discussed. Then, in a multi-agent system over the Internet, F_{lowa} is used to aggregate linguistic information and obtain better Web pages (or documents) which fulfill the user's query.

2. The Index of Linguistic Values

From the formal point of view, a linguistic variable is defined by a quintuple $\{L, H(L), U, G, M\}$ [1], in which, L is the name of the linguistic variable, and $H(L)$ (or simple H) is its linguistic values set. A numerical variable u called the base variable, which takes values in the universe of discourse U , is associated with the linguistic variable L . G is syntactic rule, which usually takes the form of a context free grammar and generates the linguistic values of H , and M is a semantic rule which associates the meaning $M(l)$ with a fuzzy subset of U to each linguistic value $l \in H, M(l) = \{\mu_l(u) | u \in U, \mu_l: U \rightarrow [0,1]\}$, μ_l is called membership function of linguistic value $l \in H$. In real world practice, for a fixed linguistic value, there are many type μ_l to be selected, such as, triangular, trapezoidal, Gaussian functions, etc. From the practical point of view, it is not always possible for user to define a membership function for a linguistic value, many times users directly use linguistic value to represent uncertain information [21,22]. In [14], fuzzy number indexes of linguistic values are proposed to handle linguistic information, formally, fuzzy number indexes of linguistic values can be obtained by Fig.1.

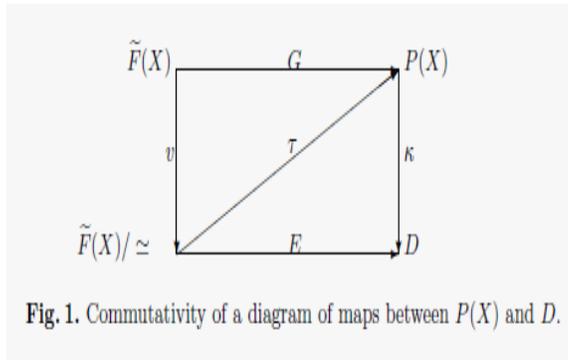


Fig. 1. Commutativity of a diagram of maps between $P(X)$ and D .

In Fig.1, v is the natural map of $\tilde{F}(X)$ (all fuzzy set on X) to the quotient set $\tilde{F}(X)/\cong$, E is extension principle, $\kappa = E\tau^{-1}$. If granule A is crisp, then $\mu_A(x)$ is the characteristic function of classical set A , and κ is also a characteristic function. Using the natural numbers theory, the indexes of granules is the natural numbers, i.e., $E(A \in x / \cong) = i \in \mathbb{R}^+$. If granule A is fuzzy, then $E([\mu_A(x)]) = \kappa$ is a fuzzy number on \mathbb{R}^+ . Let μ_A be triangular membership function of fuzzy granule A , then κ is a triangular membership function.

Based on the concept of information granule (IG)[3], we can modify the index of linguistic label as the following way: Let $X \subseteq \mathbb{R}$ be a universal set, all granules of X are denoted by $P(X)$, and $\mathbb{R}^+ = \{x / x \geq 0\}$, define a linear function $F: X \rightarrow \mathbb{R}^+$; Let μ_A be a membership function of a granule $A \in P(X)$. Let $\tilde{F}(X)$ be the collection of membership function on X such that $\forall \mu_A(x) \in \tilde{F}(X)$, $\mu_A(x)$ be a membership function of a granule $A \in P(X)$. Let $\tilde{F}(X)$ be the collection of membership functions on X such that $\forall \mu_A(x) \in \tilde{F}(X)$, $\mu_A(x)$ be a membership function of a granule $A \in P(X)$;

Define $G: \tilde{F}(X) \rightarrow P(X)$. An equivalence relation " \cong " on $\tilde{F}(X)$ can be obtained: $\mu_A(x) \cong \mu_B(x) \leftrightarrow G(\mu_A(x)) = G(\mu_B(x))$, and each equivalence class is denoted by $[\mu_A(x)] \in \tilde{F}(X)/\cong$; Select a representative element $\mu_A(x)$ of $[\mu_A(x)]$, based on F and the extension principle of fuzzy set, one-to-one mapping can be obtained: $E: \tilde{F}(X)/\cong \rightarrow D = \{x | x: \mathbb{R}^+ \rightarrow [0,1], E([\mu_A(x)]) = E(\mu_A(x)) = x\}$

3. Aggregation of linguistic values

Based on fuzzy number indexes of linguistic values and existed aggregation operators, we have the following linguistic aggregation operators.

Definition 1. [14] Let $S = \{s_i | i=1,2,\dots,T\}$ be a finite set, $A = \{a_{j_1}, a_{j_2}, \dots, a_{j_n}\} \subseteq S$ be a finite set of labels to be aggregate ($n \leq T$). $W = \{w_1, w_2, \dots, w_n\}$ be a weighting vector such that $\forall s \in \{1,2,\dots,n\}, w_s \in [0,1]$, and $\sum_{s=1}^n w_s = 1$.

Let $B = \{j_1, j_2, \dots, j_n\}$, where j_s is the center of \tilde{j}_s ,

$C = \sigma(B) = (j_{\sigma(1)}, \dots, j_{\sigma(n)})$ such that $j_{\sigma(s')} \geq j_{\sigma(s)}, \forall s' \leq s$

(C' such that $j_{\sigma(s')} \leq j_{\sigma(s)}, \forall s' \leq s$), denote

$$w = f_{owa}(B) = WC^T = \sum_{s=1}^n w_s j_{\sigma(s)}, \text{ and}$$

$$w' = f_{i-owa}(B) = WC'^T = \sum_{s=1}^n w_s j_{\sigma(s)}$$

then the new (inverse-) linguistic ordered weighted averaging operator $F_{lowa}(F_{i-lowa})$ is defined by

$$F_{lowa}((a_{j_1}, a_{j_2}, \dots, a_{j_n})) = a_{j_k}, F_{i-lowa}((a_{j_1}, a_{j_2}, \dots, a_{j_n})) = a_{j_k}$$

Where $a_{j_k}(a_{j_k})$ is defined by

$$\tilde{j}_k(w) = \max\{\tilde{j}_1(w), \tilde{j}_2(w), \dots, \tilde{j}_T(w)\} (|S|=T) \quad \tilde{j}'_k(w') = \max\{\tilde{j}_1(w'), \tilde{j}_2(w'), \dots, \tilde{j}_T(w')\} (|S|=T)$$

In Definition 1, membership function \tilde{j}_s is a triangular membership function, i.e., \tilde{j}_s is expressed by parameters (a, b, c) , where a =center, b =left half width and c =right half width σ is a permutation on B , which is equal to obtain $B = [b, b_2, \dots, b_n]^T$ of Definition 1. Due to $a_{j_s} \leftrightarrow \tilde{j}_s$, therefore, aggregating linguistic labels $\{a_{j_s} | s=1, \dots, n\}$ is transformed by aggregating the indexes of the linguistic labels.

Example 1. Let linguistic labels is $S = \{C, EL, ML, MC, IM, SC, VLC, EU, I\}$, their membership functions and the indexes are in Table 1, a figure of the membership functions and indexes is Fig.2.

Table 1. Linguistic labels, membership functions and indexes.

number	Meaning	Membership function	Index
1	C	(1, 0.125, 0)	$s_8^C = s(8,1,0)$
2	EL	(0.875, 0.125, 0.125)	$s_7^C = s(7,1,1)$
3	ML	(0.75, 0.125, 0.125)	$s_6^C = s(6,1,1)$
4	MC	(0.625, 0.125, 0.125)	$s_5^C = s(5,1,1)$
5	IM	(0.5, 0.125, 0.125)	$s_4^C = s(4,1,1)$
6	SC	(0.375, 0.125, 0.125)	$s_3^C = s(3,1,1)$
7	VLC	(0.25, 0.125, 0.125)	$s_2^C = s(2,1,1)$
8	EU	(0.125, 0.125, 0.125)	$s_1^C = s(1,1,1)$
9	I	(0, 0, 0.125)	$s_0^C = s(0,0,1)$

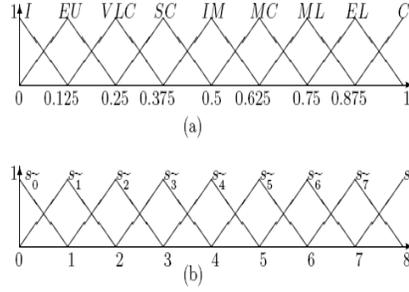


Fig. 2. (a) The membership functions of linguistic labels. (b) Indexes of linguistic labels.

Selecting $y = F(x) = 8x$. In order to aggregate

$B = \{EU, VLC, EL, C\}$, and using the weighting vector is $W = (0.3, 0.2, 0.4, 0.1)$, one obtains using F_{lowa} and F_{i-lowa} , $w = 4.7$, $w' = 4.3$, and

$$s_{\frac{2}{5}}(4.7) = \max\{s_0(4.7), \dots, s_8(4.7)\},$$

$$s_{\frac{4}{4}}(4.3) = \max\{s_0(4.3), \dots, s_8(4.3)\}.$$

Finally, $F_{lowa}((EU, VLC, EL, C)) = s_{\frac{2}{5}} = MC$ and

$$F_{i-lowa}((EU, VLC, EL, C)) = s_{\frac{4}{4}} = IM.$$

In Example 1, there are nine Linguistic labels, and Membership function of I is $(0, 0, 0.125)$, hence, Indexes are from 0 to $T_i - 1 (T = 9)$, i.e., $y = F(x) = 8x$.

Theorem 1. (Commutativity) For F_{lowa} , we have

$$F_{lowa}((a_{\tilde{j}_1}, a_{\tilde{j}_2}, \dots, a_{\tilde{j}_n})) = F_{lowa}((a_{\tilde{j}_{\sigma(1)}}, a_{\tilde{j}_{\sigma(2)}}, \dots, a_{\tilde{j}_{\sigma(n)}})).$$

Where $\sigma(1), \dots, \sigma(n)$ is a permutation of $(1, \dots, n)$.

Theorem 2. (Boundedness) Let $W_* = (0, \dots, 0, 1)$ and $W^* = (1, 0, \dots, 0)$, then

$$1. \min\{j_1, \dots, j_n\} = W_* C^T \leq f_{owa}((j_1, \dots, j_n)) \leq$$

$$W^* C^T = \max\{j_1, \dots, j_n\}.$$

$$2. \min\{a_{\tilde{j}_1}, \dots, a_{\tilde{j}_n}\} \leq F_{lowa}((a_{\tilde{j}_1}, \dots, a_{\tilde{j}_n})) \leq$$

$$\max\{a_{\tilde{j}_1}, \dots, a_{\tilde{j}_n}\}$$

Corollary 1. Let $A = (a_{\tilde{j}_1}, a_{\tilde{j}_2}, \dots, a_{\tilde{j}_n})$ be a collection of linguistic labels, then for any $T(t-norm)$ and

$S(t-conorm)$,

$$T(a_{\tilde{j}_1}, a_{\tilde{j}_2}, \dots, a_{\tilde{j}_n}) \leq F_{lowa}(a_{\tilde{j}_1}, a_{\tilde{j}_2}, \dots, a_{\tilde{j}_n}) \leq$$

$$S(a_{\tilde{j}_1}, a_{\tilde{j}_2}, \dots, a_{\tilde{j}_n})$$

4. Information aggregation over the Internet

A multi-agent system (MAS) is a system that tries to achieve predefined goals in a complex and dynamic environment. MAS not only retrieves and filters information (in the sense of Web documents), but also handles electronic mail, news lists, FAQ lists, and so on [16, 20-25]. In this paper, a MAS over the Internet includes:

– Internet Users, who look for Web documents on the Internet by means of a weighted query where a set of terms related to the documents desired is specified.

- Interface Agents (one for the user, generally), which communicate the user's weighted query to Task Agents and filter the documents retrieved from Task Agents to give users that best satisfy their needs.
- Task Agents, which communicate the user's query to Information Agents and get those documents from every Information Agent that best fulfills the query, fusing them and resolving the possible conflicts among Information Agents.
- Information Agents, which receive the weighted query from the Task Agents, look for the information in the data sources, and give the retrieved documents back to the preceding level.
- Information Sources, consisting of all data sources within the Internet, such as databases and information repositories. In the process of information aggregation, as a response to a single-user's query on the MAS presented Fig.3, there are two parts:

- 1) On one hand, there is a communication between agents at levels 5-4 and 4-3, which is far from the user's participation and where the question that Task Agents must answer concerns which Information Agents would better satisfy the user's needs;
- 2) On the other hand, there is a communication between agents at levels 3-2 and the user, where the information element is specifically the set of retrieved Web pages (or documents) that will be analyzed and filtered by Interface Agents.

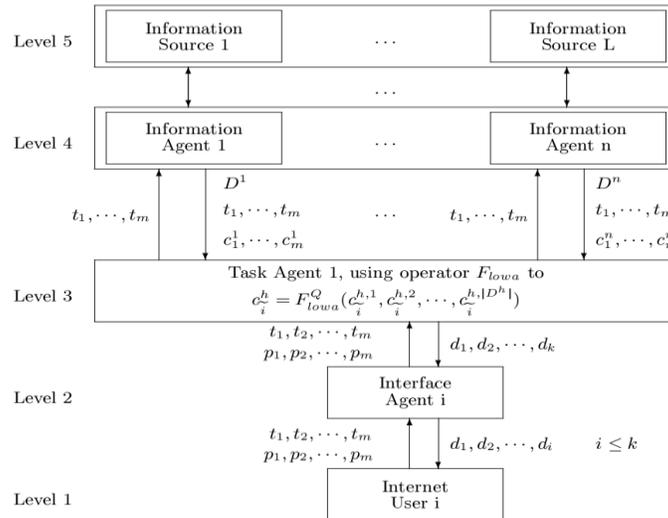


Figure 3. Overview of information flows in a single-user scheme

In this paper, the main discussion is concentrated on 1 of above two parts, and assumed that the linguistic label set S (the linguistic degrees of importance), in which every term is defined by a triangular membership function on $[0, 1]$, has been obtained. Then the process of information aggregation based on F_{lowa} in Fig. 3 can be described as follows:

- An Internet User makes a query to look for Web pages (or documents) related to the terms $\{t_1, t_2, \dots, t_m\}$, which are weighted by linguistic degrees of importance $\{p_1, p_2, \dots, p_m\}$, $p_i \in S$, every p_i is given by the user and sent to its Interface Agent.
- The Interface Agent gives these terms to its Task Agent.
- The Task Agent makes the query to all the Information Agents to which it is connected and gives them terms $\{t_1, t_2, \dots, t_m\}$.
- All Information Agents that have received the query look for the information that best satisfies it in Information Sources, and retrieve from them.

- The Task Agent receives from every Information Agent $h(h=1,2,\dots,n)$ a set of Web pages (or documents)

D^h and their linguistic degrees of importance $\{c_i^h \mid i=1,2,\dots,m\}, c_i^h \in S$, every Web page (or document) of D^h is denoted by $d_j^h (j=1,2,\dots,|D^h|)$.

Formally, every linguistic degrees of importance c_i^h can be decided as following way:

1. Let $c_i^{h,j} (j=1,2,\dots,|D^h|)$ is linguistic degrees of the Web page (or document) d_j^h about term t_i , c_i^h is aggregated by all $c_i^{h,j} (j=1,2,\dots,|D^h|)$,

2. $c_i^{h,j}$ is decided by number of times of term t_i in the Web page (or document), number of times is how many times of term t_i appear in the Webpage (or document) d_j^h

3. Let $\equiv_j^h = \sum_{i=1}^m n_j(t_i)$, where $n_j(t_i)$ be the number of times of t_i appearing in the Web page (or document)

$d_j^h \in D^h$, the membership function of $s_k \in S$ be μ_{s_k} , which is defined by experts or statistics information according to type of the Web page (or document), then

$$c_i^{h,j} = s_{\tilde{k}}, \tilde{k} = \underset{\equiv_j^h}{\arg \max} \left(\frac{n_j(t_i)}{\equiv_j^h} \right) = \max \left\{ \tilde{k} = \underset{\equiv_j^h}{\arg \max} \left(\frac{n(t_i)}{\equiv_j^h} \right) \mid s_{\tilde{k}} \in S - \{p_1, \dots, p_m\} \right\}$$

4. The Task Agent aggregates linguistic degrees of Importance $c_i^{h,j} (j=1,2,\dots,|D^h|)$ through the operator F_{lowa}^h to obtain c_i^h , i.e., where the operator F_{lowa}^Q guided by a fuzzy linguistic quantifier Q , which could be fixed by an expert or user.

5. According to linguistic degrees of importance $\{p_1, p_2, \dots, p_m\}$ that are given by the user, for every Information Agent h , defines $\lambda^h = G((p_1, p_2, \dots, p_m), (c_1^h, \dots, c_m^h))$, where G is a measure function, e.g., Euclidean measure. Intuitively, λ^h expresses similarity between $\{p_1, p_2, \dots, p_m\}$ and (c_1^h, \dots, c_m^h) . Information Agent that best fulfills the user's query is selected by λ^h . Formally, the

Task Agent orders λ^h decreasingly and obtains the vector $\sigma(\{\lambda^1, \dots, \lambda^n\}) = \{\lambda^{\sigma(1)}, \dots, \lambda^{\sigma(n)}\}$, where σ is a permutation over the set of labels $\{\lambda^1, \dots, \lambda^n\}$ such that $\lambda^{\sigma(j)} \leq \lambda^{\sigma(i)}, \forall i \leq j$

- The Interface Agent receives an ordered list of Web pages (or documents) $\{D^{\sigma(h)} \mid h=1,2,\dots,n\}$ from the Task Agent.

- The Interface Agent filters these Web pages (or documents) to give to the user only those Web pages (or documents) that best fulfill his or her needs. To implement above contents via the Internet, two aspects need to be considered, one is a query language, i.e., SQL, the other is WWW. Due to fuzzy linguistic terms, a precise query language is needed to extend so as to allow for the use of fuzzy linguistic terms, i.e., fuzzy querying. On the other hand, an integration of a DBMS with WWW is needed. In an implementation of fuzzy querying over the Internet, Java applets and JavaScript may be effectively and efficiently employed for implementing advanced, multi-platform and user-friendly fuzzy querying, which will be our future research tasks.

5. Conclusion

Based on linguistic ordered weighted averaging operators F_{lowa} , we discuss linguistic information aggregation over the Internet, its advantages are 1) user makes natural linguistic expression directly. Moreover, linguistic degrees of importance of terms can be obtained by agents, all of these linguistic information can be handled by F_{lowa} ; 2) The operator F_{lowa} has a greater flexibility in the

information gathering process, reducing the effect of the low importance of some terms within the overall performance of Information Agents.

6. Acknowledgement

This work is partly supported by the research fund of key laboratory of the radio signals intelligent processing of Xihua University (XZD0818-09) and technique support project of sichuan province (2008GZ0118).

7. References

- [1] L. A. Zadeh, "The concept of a linguistic variable and its applications to approximate reasoning" (Part I, Part II, Part III). *Information Sciences*, Vol. 8, No. 9, 1975, pp. 199-249, 301-357, 43-80.
- [2] L. A. Zadeh, "Fuzzy logic = computing with words", *IEEE Trans, Fuzzy Syst*, No. 4, 1996, pp. 103-111.
- [3] L. A. Zadeh, "Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic", *Fuzzy Sets and Systems*, Vol. 90, 1997, pp. 111-127.
- [4] N. C. Ho, T. D. Khang, V. N. Huynh, "An algebraic approach to linguistic hedges in Zadehs fuzzy logic, *Fuzzy Sets and Systems*", Vol. 129, 2002, pp. 229-254.
- [5] Z. Pei, Y. Xu, "Lattice implication algebra model of linguistic variable truth and its inference", *Applied Computational Intelligence*, World Scientific, 2004, pp. 93-98.
- [6] V. Novák, "Antonyms and linguistic quantifiers in fuzzy logic", *Fuzzy Sets and Systems*, Vol. 124, 2001, pp. 335-351.
- [7] A. Dvořák, V. Novák, "From theories and linguistic descriptions", *Fuzzy Sets and Systems*, Vol. 143, 2004, pp. 169-188.
- [8] J. Lawry, "A methodology for computing with words", *International Journal of Approximate Reasoning*, Vol. 28, 2001, pp. 51-89.
- [9] J. Lawry, "A framework for linguistic modelling. *Artificial Intelligence*", Vol. 155, 2004, pp. 1-39.
- [10] F. Herrera, L. Martínez, "The 2-tuple linguistic computational model: advantages of its linguistic description, accuracy and consistency", *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, Vol. 9, 2001, pp. 33-48
- [11] F. Herrera, E. Herrera-Viedma, "Aggregation operators for linguistic weighted information", *IEEE Trans, System, Man, Cybernet, Part A, Systems Humans*, Vol. 27, 1997, pp. 646-656.
- [12] Z.S. Xu, "Uncertain linguistic aggregation operators based approach to multiple attribute group decision making under uncertain linguistic environment", *Information Sciences*, Vol. 168, 2004, pp. 171-184.
- [13] R.R. Yager, "On ordered weighted averaging aggregation operators in multicriteria decision making", *IEEE Transaction on Systems, Man and Cybernetics*, Vol. 18, 1988, pp. 183-190.
- [14] Z. Pei, Y. J. Du, L. Z. Yi, Y. Xu, "Obtaining a complex linguistic data summaries from database based on a new linguistic aggregation operator", *Computational Intelligence and Bioinspired Systems*, Lecture Notes in Computer Science, Vol. 3512, Springer-Verlag, Berlin Heidelberg New York, 2005, pp. 771-778.
- [15] Z. Pei, D. Ruan, J. Liu, Y. Xu, *Linguistic Values based Intelligent Information Processing: Theory, Methods, and Application*, Atlantis Computational Intelligence Systems, Vol. 1, Atlantis press & World Scientific, 2009.
- [16] M. Delgado, F. Herrera, E. Herrera-Viedma, "Combining linguistic information in a distributed intelligent agent model for information gathering on the internet". *Computing with Words*. New York: John Wiley and Sons, 2001, pp. 251-276.
- [17] H. Becker, "Computing with words and machine learning in medical diagnostics, *Information Sciences*", Vol. 134, 2001, pp. 53-69.
- [18] E. Herrera-Viedma, E. Peis, "Evaluating the informative quality of documents in SGML format from judgments by means of fuzzy linguistic techniques based on computing with words",

- Information Processing and Management, Vol. 39, 2003, pp. 233-249.
- [19] J. Kacprzyk, S. Zadrozny, "Computing with words in intelligent database querying: standalone and Internet-based applications", *Information Sciences*, Vol. 134, 2001, pp. 71-109.
 - [20] A. Aldea, R. Banres-Alcatara, L. Jimenez, "The scope of application of multiagent systems in the process industry: three case studies", *Expert Systems with Applications*, Vol. 26, 2004, pp. 39-47.
 - [21] L. A. Zadeh, "From computing with numbers to computing with words from manipulation of measurements to manipulation of perceptions", *IEEE Transactions on Circuits and Systems*, Vol. 45, 1999, pp. 105-119.
 - [22] L. A. Zadeh, "Precisiated natural language (PNL)", *AI Magazine*, Vol. 25, No. 3, 2004, pp. 74-91.
 - [23] N. C. Ho, T. D. Khang, V. N. Huynh, H. C. Nguyen, "Hedge algebras, linguistic valued logic and their application to fuzzy reasoning", *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, Vol. 7, 1999, pp. 341-361.
 - [24] H. C. Yang, C. H. Lee, "A text mining approach on automatic generation of web directories and hierarchies", *Expert Systems with Applications*, Vol. 27, 2004, pp. 645-663.
 - [25] G. Meghabghab, "Discovering authorities and hubs in different topological Web graph structures", *Information Processing and Management*, Vol.38, 2002, pp. 111-140.