MNK Publication

International Journal of Latest Research in Science and Technology **ISSN (Online):2278-5299**
Vol.1,Issue 4 :Page No.427-433 ,November-December 2012
https://www.mnkpublication.com/journal/ijlrst/index.php

SURVEY ON SEMANTIC INFORMATION REPRESENTATION AND MANAGEMENT MODELS

[1]Benina Velázquez Ordoñez, [2], Miguel Patiño Ortiz, [3]Jesús M. Olivares Ceja
[1] Escuela Superior de Ingeniería Mecánica Y Eléctrica, Unidad Zacatenco- IPN, bvelazquez@ipn.mx
[2] Escuela Superior de Ingeniería Mecánica Y Eléctrica, Unidad Zacatenco- IPN, mpatino@ipn.mx
[3]Jesús M. Olivares Ceja, Centro de Investigación en Computación, CIC-IPN, jesuso@acm.org

Abstract* : *A challenge in the representation and management of automated information and knowledge, independent of language and domain, is to identify the structures and processes that allow their proper storage, processing and recovery. In this paper, different models of representation and management of information and knowledge are reviewed, with the purpose of identifying their characteristics, advantages and disadvantages in order to help the user choose the best model to fit the problem.

Keywords – models, representation, knowledge

I. INTRODUCTION

In the last decades, there has been an accelerated growth in the managing of information by electronic means. Among the sources with the most quantity of information is the web internet. Currently, in order to make a search in the web, the web search[1] is used. This search deploys links to the content that the user has to click and review to find the information required, which is time consuming.

In the web there is information in different formats: text, images, videos, sounds, among others. In this paper the focus is on the kind of representation and textual managing. In the representation and managing of the information and textual knowledge, models were found based in terms, n-grams, data bases, semantic nets, ontology, logics forms, among others.

Models of representation and managing of information and knowledge are useful in different systems. Currently, the information systems mainly take advantage of data bases and ontology. The expert systems prefer the logic forms. The recovery of documents and processing of texts have considered as useful the models based on terms, n-grams and ontology.

The rest of this article is structured in the following way. In section 1 introduction. In section 2 shows the models used with most frequency in the representation and managing of information and knowledge. In each case, characteristics, advantages and disadvantages are indicated. In section 3 conclusion trends in the usage and ways of representation and managing knowledge are discussed.

II. REPRESENTATION OF INFORMATION AND KNOWLEDGE

For the representation of the knowledge there are different models, in which, different techniques of extraction are used. Some of the models under study to generate a new alternative are described as following.

The following sentence will be taken as an example for the representation of the different models.

"Intel launched its micro-processors core of the third generation Ivy Bridge".

A) MODEL BASED IN TERMS.- It is determined with the representation of terms of the text or the corpus corresponding directly to a set of words (Bag of Words), and consists in the representation of each word [21].

The statistician processing of the natural language (Manning, 1999), represents and characterizes the system of recovery basic model. In it, each document is described by a set of words named index; its focus is characterized in a simple way and is based in what is named bag of words, where all the words of a document are treated through indexes, assigning also the specific weight to each term in function of its importance, and determining the frequency of the appearance inside the document. Therefore, it does not take the order, structure and meaning of the words. The model has the two following stages:

1.- The processing of the documents consists in preparing the documents for its parametrization, the remaining documents are eliminated

2.- Parametrization: consists in making a quantification of the terms of the documents [23].

The example consists in assigning a weight to each of the relevant terms associated with a document. The weight of a term is normally calculated by its frequency of appearance in the document and indicates the importance of the mentioned term as a describer of the content of that document.

[1] www.google.com
www.yahoo.com

Publication History
Manuscript Received : 25 December 2012
Manuscript Accepted : 28 December 2012
Revision Received : 30 December 2012
Manuscript Published : 31 December 2012

Processors	2	miniaturization	1
multiprocessors	1	circuit	1
Intel	1	tecnology	1
launched	1	transistors	1
core	3	new	1
third	2	wait	1
generation	3	to offer	1
Ivy Bridge	3	greater	3
Sandy bridge	1	performance	1
finally	1	cycle	1
release	1	watch	1
its	1	speed	1
familly	2	efficiency	1
name	1	it counts	1
code	1	design	1
chips	1	core	3
characterizes	1	improved	1
firsts	1	respect	1
made	1	packages	1
under	1	four	1
process	1	cores	1
manufacturing	1	X86-64	1
nanómeters	1	instructionsn	1
help	1	Aux	1

Figure 1. Relevant terms linked to a document.

B).- N-GRAMAS MODEL. This Consists in the identification of the words written together (compound words, names, etc.) in a document, with the purpose to be seen as a single concept. It estimates the probability that when two words appear together with a certain frequency, they compound a single term [23]. The N-GRAMS are models that provide the probability of transition of the following previous words N-1. The model is defined as following:

$$P\ (wi\ wi= N+1;\ldots;wi-2;\ wi-1)$$

With this model, an accurate prediction of the words can be expected, from one depending word up to a unique word connectivity that can be represented. On the other hand, the number of estimated parameters, for example, the number of combinations of transitions of words is V^n in a vocabulary V. V^n increases according to the exponential "N", in other words, it is possible to estimate the efficiency of each transition probability of difficult words under a big "N".

N-Grams are proposed to solve the problem of a wide quantity of parameters in a N-Grams word. In the category of N-Grams, the probabilities of transition of the following word N-1 is given with the formula:

$$p(Ci\ Ci=N+1;\ldots;\ Ci-2;\ Ci-1)p(wi\ Ci)$$

Where Ci represents the Word category to which the word wi belongs.

The decomposition of texts in n-grams has many applications [2]. The idea is that two words with the same root but with a different suffix, at the moment they are decomposed in n-grams, a series of equal n-grams will be produced (related to the root part), plus some others different n-grams (related to the suffix part). In the recovery model based on n-grams, the words belonging to the same group of information should produce a certain similarity.

For example: suppose that the result of a de-codification assigns similar probabilities to the phrases:

Phrase: " Intel launched its micro-processors core of the third generation Ivy Bridge". Representation: combination of N consecutive words:

{Intel launched its], {launched its processors}, {its core processors},{core processors of}, {core of third},{of third generation},{third generation Ivy Bridge}.

The following example is by skip of N-Grams, combination of N-words in order of appearance in the phrase (trigrams).Representation, using the same phrase mentioned before.

{Intel processor00},{launched core processors}, {its core of third},{core generation Ivy},{core Ivy Bridge}.

C).- MODEL AHL (ANALOG HYPERSPACE TO THE LANGUAGE).

This model allows identifying syntactic patterns inside any sequential structure. This model is ideal for the communication of animals, and can have a determined language as well. Also, it can explore the usage of the created rules [1].

This model captures the concepts using recurring text information, and is combined with the ALS model (Analysis of the latent semantic). (Deewester, Dumains, Furnas, Landa uer, Harshman, 1990; Foltz, Laham, 1998), Here is where a word is expressed as a vector of documents that contains the word. In this kind of semantic rooms, the representation based on vectors makes the comparison between the words easier. AHL is useful for the identification of patterns.

D).- MODEL BASED IN ONTOLOGY.- The extraction of text information is linked directly with ontology. The learning of ontology is defined as the acquisition of the terms representing a domain based on texts and has as an objective the creating of a hierarchy of concepts. There are different modalities of ontology learning, depending of the types of inputs used for the learning, texts, dictionary base of knowledge, semi-structured data and data bases. [21].

In contrast to the verbal models based on ontology logic, the models based on ontology are defined formally and have a capacity for formal reasoning.

The logic formal models are based on different types of descriptive logics [18] that is a family of logic based on a formalism of representation of knowledge and the reasoning related to the terminology and ontology. It is characterized mainly by a set of constructors that allow the construction of complex concepts and atomic functions.

The logic based on ontological models is carried out in the area of the development, through the semantic web [18].

Example of representation: Considering the Phrase: " Intel launched its micro-processors core of the third generation Ivy Bridge", the chart of the figure 2 is gotten.

Figure 2 describes an ontological model, regarding the concepts and relations. The names of the instances are also observed.

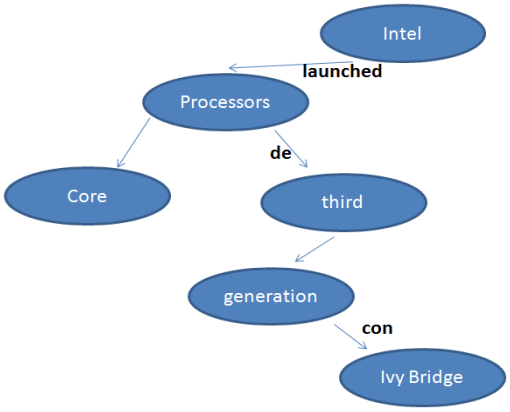

Figure 2.- Ontologic model of micro-processors.

E).- LSA MODEL.- The analysis model of the latent semantic (LSA) [9], was described by Deerwester, Dumais, Furnas, Landauer and Harshman in 1990, as a method of recovery of information. Later Landauer y Dumais in 1996 and 1997, proposed an acquisition model and knowledge representation. For its usage a text of big dimensions is processed and is represented by a matrix with rows containing all the different terms to the data base (words) and the columns represent a contextual window where those terms appear (paragraphs).The matrix contains the number of times that each term appears in a document. In this matrix the words excessively frequent are not relevant and more importance is given to the words moderately frequent [9], with the idea that the words moderately frequent are useful to modify the information of the paragraph, but not the ones of the most frequency.

This model uses the decomposed algorithm of the singular value (SVD), a variant of the factorial analysis. This algorithm is applied with the idea of reducing the number of dimensions of the original matrix to a manageable number, without losing the essential information of the original matrix [3].

In the reduction of dimensions. It is not only to improve the managing of a matrix as big as the original, but also to create a vector semantic room, in which terms of documents are represented by vectors containing only the substantial information for the concept formation.

For example: Table 2 shows an exercise with the model LSA, the results were compared.

According to the phrase to look for is: finally Intel will release its family of processors core of third generation with the code name "Ivy Bridge".

Words	Documents		Word	Documents	
	d1	d2		d1	d2
intel	1	0	Pc	0	1
Finally	1	0	Recent	0	1
Release	1	0	Estudy	0	1
Family	1	0	New	0	1
Processors	1	0	Device	0	1
Core	1	0	Increase	0	1
Third	1	0	Customer	0	1
Generation	1	0	tablets	0	1

Tabla 2 aplication of the model LSA.

F).-MODEL OF DEPENDENCIES AMONG WORDS.-

The linguistic structure can be determined in terms of dependencies among words, for this purpose the grammar of dependencies is used; in a grammar of dependencies it is said that a word is the core of a phrase and the remaining of the words can be dependent of the core or depend to another word of the phrase and this, depend of the core [34]. These dependences are: "a noun of the subject depends of the main verb of the predicate", "an adjective depends of t noun that affects", these rules that set the dependencies are the rules of production of the grammar. The dependencies of a phrase can be represented by an tree structure. The model of dependencies has as a goal to get a tree of a phrase, according to grammar of dependencies, that it is known as three of dependencies.

Table 3, shows an example of representation according to the model based on relationship of dependencies among words.

Modificator	Core	Relation of dependencies.
Intel [N]	Lanzo [N]	Proce-core
Proccesor [N]	Lanzo [N]	Ivi Bridge
Generation [N]	Lanzo [N]	Proccesor

Table 3. Relations of dependencies between words

G).- MODEL OF FUSION OF SYNTACTIC DEPENDENCIES.- [27] This model has as an objective to show the complete syntactic structure of an statement of an given input. The syntactic structure is represented by two paradigms: the grammar of dependency or a structure of grammatical phrase. While the syntactic dependency is represented differently in both paradigms, however, both have as objective to find out big quantities of syntactic units of words, for example: Phrases with explicit relations in its dependency.

Example: representation of the sentence "Intel launched its processors core of the third generation Ivy Bridge"

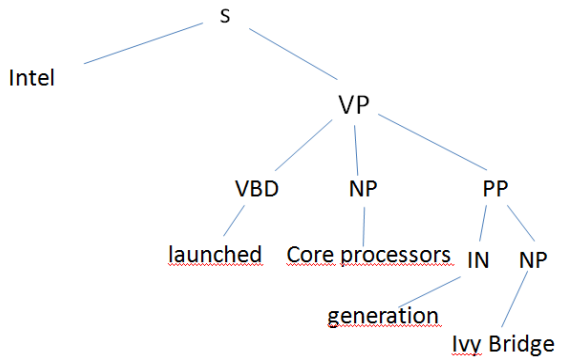

Figure 3.- Example of butterfly tree of net.

Figure 3 represents a model of fusion of syntactic dependencies that consists of analyzing the collection of texts in a file and extracting the syntactic dependencies between the verbs and subjects.

For example: A museum shelters an impressive collection of medieval and modern art. The building combines the geometric abstraction with classic references that allude to the Roman influence in the region.

This is:

Houses_subj (museum)

Houses_sobj (collection)

Combines_subj (building)

Combines_obj (abstraction)

By the way of lemmatization "references" it is assigned to its basic form of "reference" and "combines" and "house" to "combine" and "house" respectively, what it results in:

house_subj (museum)

house_sobj (colection)

combine_subj (building)

combines_obj (abstraction)

combine_with (references)

allude_to (influence).

H).- MODEL OF SYNTACTIC REPRESENTATION.-

The grouped words in syntactic, have a determined function inside the statement, the words that compound it should keep between a determined relationship, due to the fact that by themselves they do not set something. In other words, the words that are linked between them, form statements that allow us to have communication. The relationship set in the group of words (syntactic) between them, are called syntactic functions, therefore all the functions of the statement are syntactic functions [32]. From this we define:

Sentence functions.- They are the basic functions (main) of the sentence subject and predicate. Syntactic functions.- They are internal and are located inside the syntactic. They work as complement to the nucleus of that syntactic. They are located out of the syntactic in the predicate. Therefore they are external. These functions complement the predicate establishing the relationship between all the syntactic of the sentence.

I).- MODEL BASED IN THE LOGIC.- The logic form of a sentence or expression is a formal representation of its logic structure [29], this is, the structure that it is relevant to specify a logic paper and its properties. There are a number of reasons (inter-relational) for the concession of a representation of the login form of a sentence, to show those aspects of the meaning of a sentence that they be derived from the logic function of certain terms and not from the lexicon meaning of the words.

The specification of the sentences of the logic forms of a language, must be done between the grammatical form and its logic form. The grammatical form of a phrase, specifies the syntax of the logic of predicates of first order [33] with a linguistic theory. "the formal logic" has been identified with a different point of view to the one it has the natural language, this is, the derivation of the logic forms is continued with the derivation of other syntactic representations of a phrase.

This idea was developed initially by Chomsky and Mayo, with the beginners of the generative semantic.

The levels of the syntactic representation include the deep structure, the structure of surface and the logic form (LF), where LF is a group of the syntactic structure that make the "logic form" of the tongue, derivated of the structure of the surface delmismotipode , rules of the transformation that comes from the deep structure.

The diffuse logic was created by [34], with the purpose to process and manipulate the information and those data affected by the un-accuracy. It is necessary to point that the diffuse logic go back to Aristoteles who created the law of the thought, and this is the base to develop the theory of the logic and later the mathematics.

The law of thought state that the preposition can be true or false and it not allowed another intermediate value. The Word diffuse is a term that it is understood as something wrong defined. Zadeh, introduces the logic infinite especifyng the diffuse set and the diffuse logic. He declares that the range of values of belonging of a element to a set can vary in the interval [0, 1], instead of false or true. The diffuse logic treats the uncertainty of the no-probabilistic and the no accuracy lexical.

J).- JPROPOSICIONAL LOGIC OR LOGIC OF STATEMENTS.- It is an affirmation that it can be false or true, it can be any of both but not both at the same time.

Examples:

1.- "Intel launched its processors core of third generation Ivy Bridge¨ = p^q

2.- "The tablets may be helping to the sales the PC of writing ¨= p^q

3.- "In spite of what it has been said that the tablets are finishing with the era of the PC ¨= p^r

4.- "o true is that a recent study let´s glimpse that this new device would be increasing the satisfaction of the customer PC¨= ∨ (p^q^r)

K).- SEMANTIC MODEL BASED IN GRAPHS.- The model is a finite resource connected grap[8]. The two types of nodes, of concepts and relations, are connected and related conceptually, each conceptual relation has one or more arcs, each one to them should be linked to a concept; a single concept by itself can form a conceptual graph, but each conceptual relation should be linked to a concept.

The concepts are described by words and mathematics symbols [31]. A graph can be related with another and a set of graphs can make a context.

The conceptual graphs count with a series of characteristics that enrich semantically and that they are used not only for the interchange of information but also for the creation of the base of knowledge and ontologies.

The scientists B.J.Garnet y E.Tsui [20], have proved theoretically that the conceptual graph is better for the representation of the knowledge than a traditional method for the representation of the knowledge, such as: semantic field,

semantic web, formal logic.The conceptual SOWA´S expresses the content of a text using the relations between the terms, the set of the terms and the set of the relations are formalized thru the application of the support, concept, to know by means of the definition of ontology that it is a description of the concept.

Example: semantic Graph, where the nodes represent the words and the arista semantic relations between them. Every graph is associated with a matrix of adjacency that represents in his income the connections presented in the graph. If the graph has n nodes, then the matrix will have dimensions of nxn.

Elements that a graph must contain:

Name: The unique ID of the node.
Type: classification of the entity (object, action).
Text: the original text.
Syn: POS label.
Sen: synonyms of the entity according to the source of knowledge.
Sem: disambiguated the meaning of the entity.
Rel: relationships with other nodes in the graph

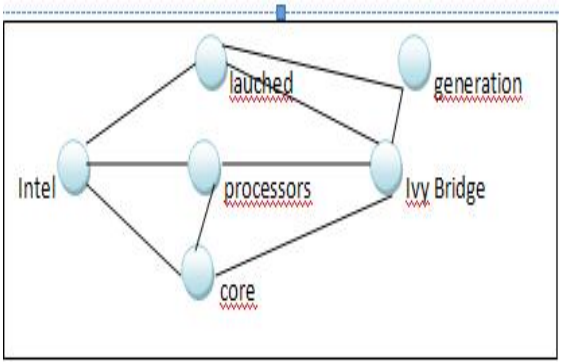

Figure 4. - Semantic Graph

L). – MODEL OF ANALYSIS OF FORMAL CONCEPT.-

It is a model created in 1982 by Rudolf Wille, which involves removing the conceptual structures of a data set, ie a "concept" is in two parts consisting of all objects belonging to that concept and intent that includes the attributes shared by those objects.

The method (FCA) is used primarily for data analysis in deriving implicit relationships between objects described by a set of attributes. The data is structured in units that are abstractions formal concepts of human thought allowing significantly understandable interpretation [7], therefore, this can be seen as a conceptual clustering technique, providing intentional descriptions of the abstract concepts or data units that it produces. This model was used to develop the system TOSCANA, which allows interactively explore databases.

Example 1: Let's consider the set of objects {p: processor, t: tablet, c: computer, or: ups, i: impress blackberry}, on that have been observed the following properties {N: electric power, mobile, M: ink, P: chassis}, obtaining the relation given by the following table:

	Electric power	Mobil	Ink	Chassis
Processor	X	X		X
Tablet	X	X		X
Computer	X			X
Ups	X			X
Printing	X		X	x

Table 4.- Relation of a set of objects.

M). - MODEL OF RECOVERY BASED ON LOCALITY
East model was proposed by Kretser and Moffat [14], it looks for the concrete positions of the text that can result relevant to the user [11]. The Recovery of Passages [35] is an intermediate approximation that it chases to identify those sections of the document - passages - relevant for the consultation. In this model, as soon as the original document has been divided in passages, these are processed and arranged by means of traditional techniques, that is to say, the recovery of the this information is based on the indexation of the simple terms, with the purpose of increasing the precision of the documents that the system returns as a response.

N).- MODEL OF ROUGH SET.- The model was developed in the eighties for Z. Pawlak (Pawlak 1991) as a tool to treat a decision process [24]. This theory implies the calculation of partitions and divisions or classes, as is needed. It is different to the theory of the statistical probability than that in the theory fuzzy set.It is possible to say that three general categories of imprecision exist in the scientific analysis:

The first category happens when an event is random by nature and the imprecision associated with this type of events can be described by the statistical probability theory.

The second is the fact that the objects cannot belong exclusively to only one category, but can belong to several categories though in different degrees. In this case the imprecision takes the diffuse form belonging to a set, and is the object of diffuse logic.

The third category is the rough set theory that is useful when the classes in those that have to qualify the objects are vague, but nevertheless, they can come closer by means of precise sets [36]. These differences show one of the principal advantages of the rough set and it is that it does not need any additional information about the data, since it can be a distribution of probability in statistics or the degree or probability of belonging to the diffuse sets theory.

Rough set is a set of objects that cannot be characterized in a precise way in terms of the available information. If the information consists of a set of objects described by another set, it is said that a rough set is a set of objects that cannot be characterized in a precise way in terms of values of a set of attributes.

The Rough set method assumes the representation of knowledge of the objects in the shape of a table of information. The rows of the table indicate the objects, whereas the columns correspond to the attributes. The entry of the table are the values of the attribute, that is to say, the

entry in column q and in row x has the value f (x, q) Therefore, for every couple (object, attribute) a value named describer is known. Every row of the table contains describers who represent information corresponding to an object. The relationship of non-differentiation would happen if two given objects for all the attributes the describers were taking the same value.

For example: The following table shows the ontology of influenza (new computer), C = {processor, core, Ivy Bridge} is a set of attributes of condition. D = {third generation of computers} is the set of decision attributes. The significant attribute of the three sets of condition is different for the technician, if a user wishes new generation equipment.

U	Procesador	core	Ivy Bridge	Equipo de tercerageneración
1	Yes	Yes	Integrado	Yes
2	Yes	No	integrado	Yes
3	Yes	No	No	No
4	Yes	Yes	No	No

In table 5 is F = {{1, 2, 3, 6, 8}, {4, 5, 7}}. The sample division of a U set of condition attributes.

III. CONCLUSIONS

There is a great quantity of texts that appear in the documents described in natural language and therefore they are conditioned to the imprecisions and the ambiguities of the natural language: polysemy, homonymies, synonymy, amphibologies, metaphors, anaphors, etc. These prevent the identification of a significant and the meaning of a word and an important concrete and precise significant. Therefore, to identify the content of a document continues to be a complex task, in manual or in automated form, since the indexation depends not only on knowledge and control of languages and documentary technologies, but the degree of knowledge that is has on the topic which the document contains.

Therefore there are different models that allow identifying the content of a document. The development of the current article described the different models existing for text extraction. This will allow us to define which of these models is the most convenient to the needs of the user.

REFERENCES

[1] Allison B. Kaufman, California State University, San Bernardino, U.S.A. Sean R. Green, University of Buffalo, The State University of New York, U.S.A., Aaron R. Seitz and Curt Burgess University of California, Riverside, Using a Self-Organizing Map (SOM) and the Hyperspace Analog to Language (HAL) Model to Identify Patterns of Syntax and Structure in the Songs of Humpback Whales U.S.A. International Journal of Comparative Psychology,2012, 25, 237-275.Copyright 2012 by the International Society for Comparative Psychology.

[2] Carlos G. Figueroa, Ángel F. Zazo,

[3] Emilio Rodríguez Vázquez de Aldana, José Luis Alonso Berrocal, Grupo Reina, Universidad de Salamanca. Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial. Vol. VIII, No.22 (2004),pp. 135-145. ISSN: 1137-3601. © AEPIA

[4] (http://www.aepia.org/revista).

[5] Deerwester, S., Dumais, S. T., Landauer, T. K., Furnas, G. W. and Harshman, R. A. (1990). Indexing by latent semantic analysis, Journal of the Society for Information Science, 41(6), 391-407."

[6] Dowty, D., Wall, R.E., and Peters, S. (1981) Introduction to Montague Semantics. Dordrecht: Reidel. Cooper, R. (1983) Quantification and Syntactic Theory.Dordrecht: Reidel.

[7] Fernando Martínez Santiago, Miguel Ángel García Cumbreras, Universidad de Jaén, Campus Las Lagunillas. Identificación de Formas Lógicas en el Caso del Español: Propuesta de un modelo en reglas y aprendizaje automático. Procesamiento de Lenguaje Natural, num. 35 (2005), pp, 245-252.

[8] Franceschetti, D.R., Karnavat, A., Marineau, J., Mc-Callie, G.L., Olde, B.A., Terry, B.L., &Graesser, A.C. (2001). Development of physics test corpora for latent semantic analysis. Proceedings of the 23th Annual Meeting of the Cognitive Science Society (pp. 297-300). Mahwah, NJ: Erlbaum.

[9] Ganter, B., &Wille, R. (1999). Formal Concept Analysis – Mathematical Foundations. Springer Verlag.

[10] Gerad Ellis, 1995, Compiling Conceptual Graphs, IEEE Transactions on Knowledge and Data Engineering Vol. 7, No. 1, February 1995.

[11] Guillermo Jorge-botana, Ricardo Olmos, José A. León, Análisis de la Semántica Latente (LSA) y estimación automática de las intenciones del usuario en diálogos de telefonía (callrouting), Revista Faz, Año 2007, Nº 1, Pág. 57.

[12] HirofumiYamamoto, ATR SLT,2-2-2 HikaridaiSeika-cho, Soraku-gun, Kyoto-fu, Japan, yama@slt.atr.co.jp. ShuntaroI sogai, Waseda University, 3-4-1 Okubo, Shinjuku-ku, Tokyo-to, Japan,

[13] isogai@shirai.info.waseda.ac.jp, Multi-Class Composite N-gram Language Model for Spoken Language Processing Using MultipleWord Clusters.

[14] Jesús Vilares y Miguel A. Alonso, Tratamiento de la Variación Sintáctica mediante un Modelo de Recuperación Basado en Localidad. Departamento de Computación, Universidad de Coruña.

[15] José A. Alonso, J. Borrego, M.J. Hidalgo, F.J. Martín y J.L. Ruiz, Unaintroducción al Análisis Formal de Conceptos en PVS., Dpto. de Ciencias de la Computación e Inteligencia Artificial. EscuelaTécnica Superior de IngenieríaInformática. Universidad de Sevilla, Avda. Reina Mercedes, s/n. 41012 Sevilla, Spain.

[16] Kurby, C. A., Wiemer-Hastings, K., Ganduri, N., Magliano, J. P., Millis, K. K., & Mc Namara, D. S. (2003). Computerizing reading training: Evaluation of a latent semantic analysis space for science text. Behavior Research Methods, Instruments, & Computers, 35, 244-250.

[17] Kretser, O. y A. Moffat. 1999a. Effective document presentation with a locality based similarity heuristic. En Proc. of SIGIR '99, Berkeley, USA, pág.113-120.

[18] Kretser, O. y A. Moffat. 1999b. Locality based information retrieval. En Proc.of 10th Australasian Database Conference (ADC '99), Auckland, New Zealand, pág.177-188.

[19] Landauer, T.K., &Dumais, S.T. (1997). A solution to Plato Õs problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. Psychological.

[20] León, J., Olmos, R., Escudero, I., Cañas, J. & Salmerón, L. (2006). Assessing short summaries with human judgments procedure and latent semantic analysis in narrative and expository texts. Behavior Research Methods, 38(4), 616-627.

[21] Leonid Kalinichenko1, Michele Missikoff2, Federica Schiappelli2, Nikolay Skvortsov1, 1Institute for Problems of Informatics, Russian Academy of Sciences {leonidk,scvora}@synth.ipi.ac.ru, 2Institute of Systems Analysis and Computer Science (IASI), CNR missikoff@iasi.cnr.it, fed_schi@iasi.rm.cnr.it.

[22] Maedch, A.; STAAB, S. (2001): «Ontology Learning for the Semantic Web», IEEE. Intelligent Systems, Special Issue on the Semantic Web, vol. 16, num. 2, 72-79.

[23] Maosheng ZHONG, School of Information Engineering, East China Jiaotong University, Nanchang, China, Jianyong DUAN, School of Information Engineering, North China University of Technology, Beijing, China, Jian ZOU, Library, Jiangxi Justice Police Vocational

College, Nanchang, China. Indexing Conceptual Graph for Abstracts of Books. 2011 Eighth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD).

[24] Mari Carmen Marcos; Rafael Pedraza-Jiménez (Coords) Web Semántica y Sistemas de Información Documental, "La web semántica y el procesamiento del lenguaje natural", en Luis Codina; Ed. Trea, Gijón, pp. 155-180.

[25] Maria Fernanda Coropeso. School of Information Technology and Engineering. University of Ottawa Incorporating Syntax and Semantics in the Text Representation for Sentence Selection, 2006.

[26] Mari Vallez y Rafael Pedraza-Jiménez. El Procesamiento del Lenguaje Natural en la Recuperación de Información Textual y áreas afines [en línea]. "Hipertext.net", núm. 5, 2007.http://www.hipertext.net

[27] M.J. Segovia Vargas, J.A. Gil Fana, A. Heras Martínez, J.L. VilarZanón, Predicción de insolvencias con el método Rough Set.

[28] Dpto. de Economía Financiera y Contabilidad I, Facultad de C. Económicas y Empresariales, Universidad Complutense de Madrid, Campus de Somosaguas, 28223 Madrid. Tfno. 913942564. Fax 913942570. E-mail mjsegovia@ccee.ucm.es.

[29] Nakov,P.I., Popova, A.., and Mateev, P.(2001) Weight Functions Impact on LSA Performance, (Sofia University Press, Sofia, 2001).

[30] Philipp Cimiano Institute AIFB, University of Karlsruhe Englerstr. 11, 76131 Karlsruhe, Germany, Andreas Hotho Knowledge and Data Engineering Group, University of Kassel. Wilhelmshoher Allee 73, 34121 Kassel, Germany, Steffen Staab Institute for Computer Science, University of Koblenz-Landau Universit¨atsstr.1, 56016 Koblenz, Germany. Learning Concept Hierarchies from Text Corpora using Formal Concept Analysis.Journal of Arti_cial Intelligence Research 24 (2005) 305.339.

[31] Philipp Cimiano, Ontology Learning an Population from Text. Algorithms, Evaluation and Applications.

[32] Rafael Muñoz Terol, 2009, Representación del conocimiento textual mediante técnicas lógico-conceptuales en aplicaciones de tecnologías del lenguaje humano.

[33] Robert May, Department of Linguistics, School of Social Sciences, University of California Irvine, CA 92697 LOGICAL FORM IN LINGUISTICS, Entry for MIT Encyclopedia of Cognitive Sciences.

[34] http://kleene.ss.uci.edu/~rmay/LogicalForm.html.

[35] Santiago Márquez Solís, Web Semántica 2007, primera edición, ISBN: 978-1-84753-192-6.

[36] Sonia Ordoñez-Salinas, Alexander Gelbukh, Generación de grafos Conceptuales. Universidad Distrital F.J.C and Universidad Nacional, Colombia. Instituto Politécnico Nacional, México.

[37] Tania Pérez Terol, Lengua castellana. 4º ESO. Sintaxis: Funciones sintácticas y funciones semánticas.

[38] The Minimalist Program. Cambridge, MA: MIT Press.Partee, B. (1975) Montague Grammar and Trasnformational grammar. Linguistic Inquiry, 6, 203-300.

[39] L. A. Zadeh, Fuzzy Sets, Department of Electrical Engineering and Electronics Research Laboratory, University of California, Berkely, California, 8, 338-35, 1965.

[40] Christopher D. Manning, HinrichSchütze, Fudations of Statistical Natural Language Processing , The MIT Press Cambridge, Massachusetts, London England. Second Printing, 1999.

[41] MarcinKaskiel, Justin Zobel, Effective Ranking with Arbitrary Passages, Department of Computer Science, RMIT University, GPO Box 2476, Melbourne 3001, Australia.

[42] M.J. Segovia Vargas, J.A. Gil Fana, A. Heras Martínez, J.L. Vilar Zanón. Predicción de insolvencias con el método Rough Set. Dpto. de Economía Financiera y Contabilidad I, Facultad de C. Económicas y Empresariales, Universidad Complutense de Madrid, Campus de Somosaguas, 28223 Madrid. Tfno. 913942564. Fax 913942570. E-mail mjsegovia@ccee.ucm.es.