

Intranet Try To Find Project (ITTF): An approach for the search of Relevant information inside an organization.

Christophe Jouis & Jean-Marie Ferru

Université PARIS III – Sorbonne Nouvelle
13, rue de Santeuil – 75231 Paris - France
{Christophe.Jouis, Jean-Marie.Ferru}@univ-paris3.fr

Abstract

We propose to demonstrate that from the point of view of the information flow (analyze, diffusion and/or information retrieval), an organization cannot be described with one and only one exclusive thesaurus (or structured terminology with a simple link such as "synonym"). In fact, generally, any human being organization (firm, university, company, etc.) is understood and then, represented like a pyramid. The components of the pyramid are employees connected to each other in particular by a link of subordination. A leader (or "CEO", "captain", "director", "General", etc.) is at the top of this pyramid. The leader must take the great decisions. The leader is generally assisted by a "staff" (management team). But every organizations is at the same time composed of different kinds of members (employees) having different categories of knowledges, goals and needs. If we consider an organization, the flow of informations provided to an employee inside this organization must be filtered according to his needs and only those informations. Consequently, flows of information are absolutely not proportional to the hierarchical levels of the employees. For instance, the "head" of the organization could not make a decision if she/he was informed of all the informations related to its own organization: he/she would be completely submerged! We thus deduce from it that it is necessary to distinguish various categories of employees by functional properties. Then, for each category, it is necessary to specify the real needs for information. Lastly, it becomes possible to reconstitute a thesaurus by category.

ITTF is an ongoing project organizing and filtering flow of external informations inside an organization using linguistic resources about the organization. Relevant informations are presented to a particular user by the intermediary of a simplified and personalized HTML interfaces.

1. Introduction

Project ITTF started in September 2002 in and for the university "Paris Sorbonne Nouvelle" (Paris, France). This project is an information processing system founded on Apache and GNU/Linux. It uses Linguistic Resources (LR) concerning all digitized information available on the university: administrative organizations, diaries, managements of the calendars, managements of the rooms, inscriptions, reservations, services, localizations of the services, "on line" professors lessons, resources of the professors, conference articles of the professors, reports of the boards of directors, reports of the scientific councils, virtual libraries, administrative texts, lodgings of the email boxes, lodgings of individual Internet sites, statistical, managements, accountancies, etc.

It is presented to the user by the intermediary of a simplified HTML interface, accessible with a simple WEB browser. We use a simplified HTML interface because computers and systems used in the university are different: Windows XX, Mac Os XX, Linux. Its objective is to provide, in Intranet, relevant information concerning the university "Paris Sorbonne Nouvelle" in response to a request of any internal user in the university. We took and adapted to the beginning the application "EPNadmin" which makes it possible to manage several public organizations (EPN) on a city scale for example. This application is centralized on only one site and makes it possible to define several different structures (EPN). A EPN is made up mainly of material tools and trainers: (a) multi-media rooms providing computers in self-service with accesses to internet resources, web cam, videoconference, etc., (b) initiation with those technologies. The application is written mainly in PHP scripts on a basis of MySQL data. The server is Apache.

This application can indifferently function under Windows or Linux. It is distributable and completely free according to terms of license LPG 2: <http://www.gnu.org/copyleft/gpl.html?cid=6>.

For each employee, the system must provide information corresponding to his needs and only this information. In other words, ITTF system must be able to provide, in response to the same request of two different users, two different results if they do not have the same role in the university.

The general ambition is that the project ITTF could be applied in all complex and/or heterogeneous organizations: administrations, armies, companies, etc.

2. Genesis of the project and problems to be solved

2.1. Solve the problem using simple and/or traditional approach ?

The first suggested solution seemed a simple and ordinary approach. Indeed, it relied upon a search engine of type "AltaVista". The user (an internal employee of the university) proposes a request made up of a logical keywords combination. These keywords are integrated in an index for the coded documents. The index is organized in only one hierarchically arranged thesaurus. But the encountered problems are also very traditional. The user is too often submerged by the number of answers to his request. This problem is not new (Salton, 1989).

The semantics of the relationships between concepts (i.e., for each relation, the number and types of its arguments , its algebraic properties, etc.) are often too vague. From our point of view, the need to structure knowledge and

then to validate the representations obtained is fundamental.

Moreover, the problem was not yet solved by the development of a more complete thesaurus (by adding more precise keywords, a more structured representation of the concepts and a reduced linguistic analysis) (Andriaans, 1996, Leloup 1998). The relationships between concepts are often reduced to the distinction established by standards ISO 704 (1987) and ISO 187 (1990) between hierarchical relationships (genus-species relationships and art/whole relationships) and non-hierarchical relationships (“time, space, causal relationships, etc.”).

2.2. Solve the problem using a logical system for semantic relationships ?

One possible approach to this problem consists in organizing the relationships in a typology based on logical properties. However, the project remains founded on terminological knowledge bases, for instance in: WordNet (Miller, 1990). Recent works applying terminological relationships to information retrieval, in particular to the construction of thesauri, tries to better specify the properties of link between concepts and to extend non-hierarchical relationships (Molholt, 1996, Green 1996, Bean, 1996).

Other recent work aims to integrate into the terminological models theories arising from linguistics (semantics, for example) and artificial intelligence, in particular the modeling of knowledge for the design of knowledge-based systems (KBS) and “ontologies”, as defined, for example, by Sowa (1984, 2000). In all these disciplines, the need to structure knowledge and then to validate the representations obtained is fundamental.

With a view to better designing the knowledge structures underlying the concepts of an organization, and more specifically, the indexing of documents and/or information retrieval, we try to use structured set of relationships, based on a linguistic model.

We indeed use an organization of terminologies founded on a semantic and logic model. This model proposes a semantic and logical concept organization using linguistic links: an enriched terminology. (Jouis, Mustafa 1995, ... 2002). The semantic and logical organization of the terminologies is founded on a semantic model of language processing: Applicative and Cognitive Grammar (ACG, Descles 90), and an extension of this model to Terminology (Jouis, Mustafa 1995, 1996, ...Jouis 2002), The ACG postulates three levels of representation of languages especially the Cognitive Level. At this level, the meanings of lexical predicates are represented by semantic cognitive schemes. In this perspective we propose a set of semantic concepts, which defines an organized system of meanings.

Relations are a part of a specification network built on a general terminological relation schema (i.e. a coherent system of meanings of relations). The general schema of relation (“ an entity X is in relation with an entity Y ”) is

further specified according to the algebraic properties in more precise relations that are axiomatically attributed to them. In this system, a relation may be specified in other more precise relations in terms of its properties:

- (1) its functional type (the semantic type of arguments of the relation);
- (2) its algebraic properties (reflexivity, symmetry, transitivity, etc.);
- (3) its combinatorial relations with other entities in a same context (the part of the text where a concept is defined for instance).

Within the cognitive level, we distinguish four categories of primitives:

- (1) Elementary semantic types of terminological units: individualizable (i.e. separately accountable or which can be shown by pinpointing) entities, Boolean entities (expressions denoting a value of truth), massive entities, collective classes, distributive classes¹, places, etc. ;
- (2) Formation operators which create more complex types from elementary types (lists, arrays, functional types², etc.);
- (3) Fundamental static Relations between terminological units. They are binary relations. Static relations permit the description of some states (static situations) related to an area of knowledge. We distinguish more than twenty relations, especially: identifications (or equivalence between two entities), incompatibility among entities, measures, cardinality, comparisons, inclusions (among distributive classes), belonging of one individualizable entity to another distributive class, localizations of one entity in one place (interiority, exteriority, boundary and closure of a location, boundary of a locality, orientations, etc.), relations part/whole among collective

¹ Lesniewsky (1886-1939) proposed a general theory of wholes and parts (mereology), in response to the problems of set theory (Cantor, 1932, 1962). A detailed analysis of mereology was carried out by Miéville (1984). Regarding a class as an object, a whole is seen as the “accumulation” of elements that constitute it, disjoint or not. Lesniewsky (1886-1939) arrives at the conclusion that the notion of class contains two features: the distributive one and the collective one.

The following example, borrowed from Grize (1973, p. 86), gives an idea of the difference: “A distributive class, is, to be strictly correct, the extension of a concept. If p is the concept planet, the statement that “Jupiter is a planet” is either to pose pJ or $J \in \{x / px\}$, and the transmitted information is the same one in the two writing. Thus, $p = \{\text{Mercury, Venus, Earth, Mars, Jupiter, Saturn, Uranus, Neptune, Pluto}\}$ is a distributive class. It contains nine elements and nothing else. The polar caps of Mars, the red Jupiter spots, the rings of Saturn do not belong to p. Yet all that and a thousand other things deal with the concept planet. The notion of collective class must mitigate this gap”.

² In the meaning of the Church typed “lambda-calculus” theory or typed logic theory of Curry (Curry & Feys, 1958).

classes (direct, necessary, atomic, typical, unique, quantifiable, etc.).

- (4) Fundamental dynamic relations between terminological units. Dynamic relations permit descriptions of processes or events related to an area of knowledge (terminological unit denoting dynamical situations): movements, changes of state, conservation of a movement, iterations, intensity, variation, constraints, causes, etc.³

From the set of elementary types, it is possible to define a system of more complicated types with the help of formation operator of functional types⁴.

With this model, we tried to validate conceptual representations obtained⁵. It was possible because relations were in theory defined by logico-semantic properties.

But our problem was still not solved because the consistency of our model to describe our university organization was not respected. The consistency of a representation is a necessary condition in order to validate our conceptual structures.

And the user was thus still submerged of documents. Why?

3. Our Solution

Our fundamental observation is : the diffusion of internal information into the organization is not adapted to the functional hierarchy of the employees of this organization.

Large companies or large public services (such as universities, large schools, large ships, ministries, etc.) are made up of heterogeneous human resources. The employees are systematically organized in a hierarchical system. This hierarchy is founded on the specialized tasks of each employee in the organization. Each task corresponds to an authority and responsibilities.

The "head" is at the top of the hierarchy (president, captain, minister, etc). He has the greatest authority. He

³ For more details, see Abraham (1995).

⁴ For more detail see Descles (1987, 1990) and Jouis (2002).

⁵ Representations were obtained with the interactive help of SEEK. It is a module used for semantic analysis of textual documents (specialists interviews, summaries, technical documentation, etc.) in any specialized field of knowledge in order to give a conceptual structure of this field. This approach can be used for any technical domains. Currently, SEEK's aim is to help specialists to find relations in texts. SEEK creates a visual representation by displaying graphs representing relations between entities of a given field of knowledge. Hypertext links connect parts of these graphs and parts of the analysed input text. The connections obtained by hypertext links allow a dynamic association of each type of relation found by this process with various relevant parts of the text. SEEK is implemented in the form of a knowledge-based system.

The knowledge base of SEEK regroups a database of markers (about 3300 markers classified in 240 lists) and 220 contextual exploration rules. For more detail see (Jouis, 1993).

must make decisions. He orders all the employees of the organization. But, he directly orders his staff : only his directly subordinate employees. In fact, to direct his organization, to make decisions, he should receive only necessary general informations . Those informations must be provided by his staff and nobody else. In the same way, the directly subordinate employees are themselves leaders of organization components (directors, managers, administrative, heads of services, etc). They must make decisions starting from a restricted set of informations. Moreover, they must apply orders from the directly higher person in charge. This process reproduces recursively until the base of the pyramid.

Consequently, flows of information are absolutely not proportional to the hierarchical levels of the employees. For example, the "head" of the organization could not make a decision if he was informed of all the informations related to its own organization: he would be completely submerged!

We thus deduce from that it is necessary to distinguish various categories of employees by functional properties. Then, for each category, it is necessary to specify the real needs for information. Lastly, it becomes possible to create a thesaurus by category.

Several thesaurus (organized terminologies) are needed! Each thesaurus is associated to a category of employees having a specific point of view on the same concept/object.

4. Outlines and fundamental questions

In the development of this project, we are identifying categories of employees and then we extract terminologies thus built. These categories and the terminologies are currently relatively stable: system ITTF is thus an operational prototype.

We will show finally some particularly relevant results of the residual problems of adequacies between the conceptual representations, such as the nonempty intersections between thesauri.

These nonempty intersections can generate fundamental questions. These questions are terminological, ontological⁶, encyclopedic, sociological, knowledge organization field dependent, knowledge classifications field dependent or, even of the field of economic theories?

Acknowledgements

This ongoing project is sponsored by the "Université PARIS III –Sorbonne Nouvelle". Particularly, the authors thank Mr. Pr. A. Rocchetti in charge of the Center of Information System Resources (CRI). He helped us to analyze with precision the information system organization of our University. We would like also to

⁶ These questions are tackled for example by the "OntoWeb" organization: a European Union founded project about Ontology based information exchange for knowledge management and electronic commerce. See <http://ontoweb.aifb.uni-karlsruhe.de/>.

address our sincere thanks to Claude Jouis (Ecole Polytechnique/The Highest French Polytechnic School) for his assistance.

References

- Abraham, M. (1995). *Analyse Sémantico-Cognitives des Verbes de Mouvement et d'Activité: Contribution Méthodologique à la Constitution d'un dictionnaire Informatique des Verbes*. Doctoral Dissertation, Ecole des Hautes Etudes en Sciences Sociales (EHESS) : Paris.
- Adriaans, P. and D. Zantinge (1996). *Data Mining*. Reading, Massachusetts, Addison-Wesley.
- Bean, C. (1996). *Analysis of non-hierarchical associative relationships among Medical Subject Headings (MeSH): Anatomical and Related terminology. Knowledge Organization and Change: Proceedings of the Fourth International ISKO Conference*, pp. 80-86.
- Cantor, G. (1962). *Gesammelte Abhandlungen* (E. Zermelo, Ed.). Hildesheim. Oms. (Original work published 1932).
- Curry, H. & Feys, R. (1958). *Combinatory Logic (Volume I)*. North-Holland.
- Descles, J.-P. (1990). *Langages applicatifs, langues naturelles et cognition*. Hermès eds., Paris, France.
- Desclés, J.-P., (1987). *Réseaux sémantiques: La nature logique et linguistique des relateurs Langages*, 87, p. 55-78.
- Green, R, Bean A, Bean and Sung Hyon Myaeng (2002). *The Semantics of Relationships, An Interdisciplinary Perspective* : Kluwer Academic Publishers: Dordrecht, The Netherlands.
- Green, R. (1998). *Attribution and relationality. Structures and Relations in Knowledge Organization: Proceedings of the fifth International ISKO Conference*, pp. 328-335.
- International Organization for Standardization (ISO). (1990). *Terminology – Vocabulary = Terminologie – Vocabulaire*. Genève: Organisation internationale de normalisation (ISO 1087-1990.).
- International Organization for Standardization (ISO). (1987). *Principes et Méthodes de la Terminologie*. Genève: Organisation internationale de normalisation (ISO 704-1987.).
- Grize, J.-B. (1973) *Logique Moderne*. (Fascicule II). Paris : Mouton/Gauthier-Villars.
- Jouis (1993). *Contributions à la Conceptualisation et à la Modélisation des Connaissances à partir d'une Analyse Linguistique de Textes ; Réalisation d'un Prototypé : Le Système SEEK*. Doctoral dissertation, Ecole des Hautes Etudes en Sciences Sociales (EHESS) : Paris.
- Jouis, C, (1995), *SEEK, un logiciel d'acquisition des connaissances utilisant un savoir linguistique sans employer de connaissances sur le monde externe*. Actes des 6ème Journées Acquisition, Validation, (JAVA 95), INRIA, pp. 159--172, Grenoble, Avril 95.
- Jouis, C, Mustafa-Elhadi, W. (1996). "Vers un nouvel outil interactif d'aide à la conception de dictionnaires électroniques spécialisés", *Lexicomatique et Dictionnaires*. Actes des IVème Journée Scientifique du réseau thématique "Lexicologie Terminologie, Traduction", Lyon, sept. 95, Beyrouth, AUPELF-UREF et FMA, pp. 255- 266.
- Jouis, C. (1998) "System of Types + Inter-Concept Relations Properties: Towards Validation of Constructed Terminologies". *Advances in Knowledge Organization*, Vol. 6, Editions ERGON Verlag: Frankfurt/Main, pp. 39-47.
- Jouis, C., (1993), *Contributions à la conceptualisation et à la modélisation des connaissances à partir d'une analyse linguistique de textes. Réalisation d'un prototype : le système SEEK*. PhD Thesis, Ecole des Hautes Etudes en Sciences Sociales, Paris.
- Jouis, C., Mustafa, W. & Rialle, V. (1998), "COGNIWEB, Modélisation hybride linguistique et numérique pour un outil de filtrage d'informations sur les réseaux ". Actes du colloque international RIFRA'98 (Rencontre Internationale sur l'extraction, le filtrage et le résumé automatiques), Sfax, Tunisie, 11-14 nov. 98, pp. 191-203.
- Jouis, C., (2002), "Logic of Relationships ", In (Bean 1996)
- Leloup, C. (1998). *Moteurs d'Indexation et de Recherche : Environnements client-serveur, Internet et Intranet*. Paris, Eyrolles.
- Miéville, D. (1984). *Un développement des Systèmes Logiques de Stanislaw Lesniewsky : Protothétique, Ontologie, Méréologie*. Bern. Frankfurt am Main, and New York P. Lang.
- Miller, G.A., (1990) : "Nouns in WordNet : a Lexical Inheritance System". In *International Journal of Lexicography*, vol. 3, n° 4, p. 245-264.
- Molholt, P. (1996). *Standardization of interconcept links and their usage. Knowledge Organization and Change: Proceedings of the Fourth International ISKO Conference*, pp. 65-71.
- Mustafa W. & Jouis, C. (1998) " Terminology Extraction and acquisition from textual data: criteria for evaluating tools and methods " In *Proceedings of the First International Conference On Language Resources and Evaluation*, Granada (Spain) : 28-30 May 1998, organized by ELRA (European Language Resources Association). Granada: ELRA, Vol. 2, pp. 1175-1180.
- Mustafa, W., Jouis C. (1997), " Natural Language Processing-based Techniques and their Use in Data Modelling and Information Retrieval ", In *Proceedings of 6th International Study Conference on Classification Research, Knowledge Organization for Information Retrieval*, 16-19 June 1997, University College of London, London, FID/CR, & ISKO. The Hague : FID, pp. 157-161.
- Mustafa-Elhadi, W. & Jouis, C. (1996), "Evaluating Natural Language Processing Systems as a Tool for Building Terminological Databases", in *Proceedings of the Fourth International ISKO Conference: Knowledge Organization and Change, Advances in Knowledge Organization*, Library of Congress, Washington D.C., Vol.5, INDEX Verlag, Frankfurt/Main, 346-355.
- Sowa, J.F. (1984). *Conceptual Structures: Information Processing In Mind And Machine*. Reading, Mass.: Addison-Wesley.
- Sowa, J.F. (2000). *Knowledge Representation: Logical, Philosophical and Computational Foundations*. Pacific Grove, CA: Brooks/Cole.
- Wielingua, B. Schreiber, A. & Breuker, J. (1992). *KADS: A modeling approach to knowledge Engineering, Knowledge Acquisition*, 4 (1), pp. 5-53.