SADIM: AN AID SYSTEM FOR MANAGEMENT ENGINEERING DIAGNOSIS USING KNOWLEDGE EXTRACTION AND MATCHING TECHNIQUES

Jamel Kolsi
LOGIQ – LARIS Laboratory, Sfax- TUNISIA

Lamia Hadrich Belguith, Mansour Mrabet, Abdelmajid Ben Hamadou
Faculty of Management and Economic Sciences, LARIS Laboratory, Sfax- TUNISIA

Keywords: Engineering management, dysfunction, term extraction, knowledge acquisition, witness sentences, key ideas, socio-economic diagnosis.

Abstract: This paper describes an aid system of management engineering diagnosis "Système d'Aide au Diagnostic d'Ingénierie de Management " SADIM, the aim of which is to detect the dysfunctions related to the enterprise management. This system allows the acquisition of knowledge based on textual data (given in French) related to the diagnosis, the matching and the assignment of witness sentences to the key ideas that correspond to them. SADIM can also serve as a part of a decision aid system as it includes carrying out diagnosis which can helps experts and socio-economic management consultants to take decisions that would make enterprises reach the required standards through council interventions.

1 INTRODUCTION

Looking for a better industrial productivity is the major concern of the organization in a context where priority is given to the decision making, reduction of the production cycle time, flexibility for risk facing, best quality, etc.

One of the means used to reach these objectives is the application of the management engineering process. The latter is applied in enterprises and organizations in four steps namely the socio-economic diagnosis, socio-economic innovation project, implementation and result evaluation (Savall and Zardet, 1989).

Our assignment is located at the level of diagnosis. The data relative to this diagnosis are formed of witness sentences fitting key ideas. These data can hide useful knowledge, dependences or inter-relations.

The socio-economic expert conducts semi-directive interviews with executives, mastery agents, workers… These interviews will represent the dysfunctions in the form of witness sentences.

Given the important number of collected witness sentences, the expert finds it hard to synthesize these sentences into key ideas. This synthesis can be made easier if the expert starts from a basis of key ideas that she/he has collected through several diagnoses.

The automatic tools proposed in this domain are characterized by the non-automation synthesis of witness sentences into key ideas. This is the case of SEGESE system (SocioEconomic Management Expert System) (Savall and Zardet, 2004). This system presents a problem of key idea redundancy. The problem is related to the significance of the key ideas rather than to the way they are formulated. This situation is due to a difficulty that the expert meets in the research of the dysfunction key ideas that correspond to the witness sentences. This situation incites the expert to insert other key ideas.

By adopting an extraction approach and an automatic manipulation of textual data relative to the management engineering diagnosis and in order to solve the problems of SEGESE, we propose a system baptized SADIM.

In what follows, we present a brief overview of previous works on knowledge extraction, then we propose our method of aid for management engineering diagnosis and finally we expose an assessment of our method.

Kolsi J., Hadrich Belguith L., Mrabet M. and Ben Hamadou A. (2006).
SADIM: AN AID SYSTEM FOR MANAGEMENT ENGINEERING DIAGNOSIS USING KNOWLEDGE EXTRACTION AND MATCHING TECHNIQUES.
In *Proceedings of the Eighth International Conference on Enterprise Information Systems - AIDSS*, pages 331-334
DOI: 10.5220/0002488003310334

2 A BRIEF OVERVIEW OF WORKS RELATED TO KNOWLEDGE EXTRACTION

The work related to automatic knowledge extraction could be classified in two basic categories: the methods of terminological extraction and the methods of knowledge acquisition.

2.1 Terminological Extraction Methods

We could distinguish three main approaches of automatic terminological extraction : the structural approaches, the non structural approaches and the mixed approaches (Chevallet, 2003).

The structural approaches use two kinds of techniques: some use the syntactic and lexical rules and often require grammars (TERMINO tool of (David & al, 1990)), the others adopt surface analysis and terms contexts (LEXTER tool of (Bourigault, 1994)) and often use the recognition syntactic patterns.

The non-structural approaches use some statistical and quantitative methods (MANTEX tool of (Oueslati, 1999)).

The mixed approaches (SORT tool of (Daille, 2002)) use the two previously described approaches.

The outcoming terms of these terminological extraction approaches can be used by tools of knowledge acquirement. We describe some of them in what follows.

2.2 Knowledge Acquisition Methods

Knowledge acquisition methods use different techniques such as:

- Techniques of relation acquirement between terms (TERM tool of (Oueslati, 1999)).
- Techniques using rules (SEEK tool of (Jouis, 1995)).
- Techniques using lexico-semantic patterns (Oueslati, 1999).
- Techniques using templates (PALKA tool of (Kim & Moldavan, 1993))

3 PROPOSITION OF AN AID METHOD FOR THE DIAGNOSIS

To elaborate our method, we have exploited the principles of morphological analysis technique of treatment (Daille & al, 2002):

- Abréviations: "Sté, SGBD, MEO,PC,..".
- Inflectional paradigms: I work, she works,…the lemma is work.
- Derivational paradigms: "nation, nationalisé, nationaliser" the root is "nation".
- Suffixations: assembler/assemblage exécuter/exécution".
- Préfixations: "faire/défaire, faire/refaire".
- Compound noun: "mise en oeuvre, mise à niveau", company manager, resource management, business structure, company structure.

We have also used some techniques of ontology to define semantic relations as (Amarnath, 2003):

- "Sorte de", (kind of): join heteronyms to hyponyms (computer material / printer).
- "Partie de", (Part of): join an element to a whole (diagnosis/ management engineering).
- "Action/Objet", (Action/Object) : (crisis/economy).
- "Objet/propriété",(Object/property) : (inflation/rate).
- "Objet/procédé", (Object/process) : (enterprise/state to rank).
- "Relation causale", (causal Relation) : (dysfonctionnement/deficit).

4 SADIM SYSTEM

In order to test our method, we developed an aid system for management engineering diagnosis (SADIM) to create and update a data of knowledge basis which essentially includes witness sentences and key ideas.

This system involves five steps: pre-treatment of witness sentences and key ideas, extraction of the simple terms and the compound terms in witness sentences, validation of the simple terms and the compound terms of the key ideas, matching the key ideas with the witness sentences, classification of key ideas with reference to each witness sentence and linking the witness sentence to key idea. (see the following figure)(Kolsi & al, 2005).

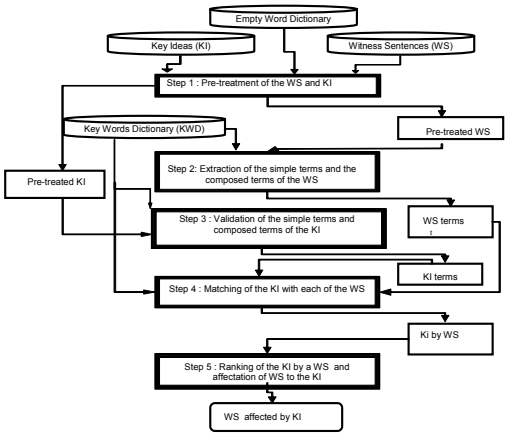

Figure 1: SADIM General architecture.

Step 1: Pre-treatment of witness sentences and key ideas

In this step, the system performs the following tasks:

- Deleting the separators: The system replaces all the following separators (:; . ? ! ( ) [] {} =" * +...) except the hyphen (-) by a space since it can exist at the level of compound terms. The objective is to have only one separator (the white space).
- Deleting the empty words: The empty words do not have a semantic content; their function is to structure the speech by a correct syntactic form that is the case of tool words. Thus we delete all empty words in witness sentences and in key ideas, with the exception of ("de, en, d',à) since the latter exist in some compound nouns.
- Deleting the multiple spaces: To unify the word separators by only one space, we replace all the existing multiple spaces in the witness sentences and in the key ideas by only one space.

Step 2: Extraction of simple terms and compound terms of witness sentences

This step consists in the extraction of simple terms and compound terms of witness sentences by using a Key Words Dictionary (KWD). This dictionary contains 950 words; each of them is represented with all its derivatives, its synonyms and the words of the same class.

In this step, the system loads all the words of the witness sentence in a table and detect all possible compound terms (i.e. composed of three words or two words) and simple terms (i.e. composed of one word).

Stage 3: Validation of the simple terms and compound terms of the key ideas

While seizing the key ideas, the expert introduces for each key idea, its corresponding simple and compound terms. In this stage, the system does the following treatments for each key idea:

- The extraction of the simple and compound terms (the same way the treatment is carried out in step 2).
- Validation of the previously mentioned terms.

Stage 4: matching of the key ideas with each of the witness sentence

This stage consists in the elaboration of a statistical table that includes some statistical data on the similarity, the synonymy and the adherence to the same class between words of a witness sentences and those of key ideas. These statistical data are going to be used to carry out the matching between each of the witness sentences and the key ideas.

This treatment can be classified as follows:

- Similarity treatment: The proposed system calculates the number of similar words in both witness sentences and key ideas taking into account the morphological variations. The results of this treatment will be stored in a statistical table.
- Synonymy treatment: This treatment provides as a result the number of synonymous words in witness sentences and key ideas. The results of this treatment will be stored in a statistical table.
- Adherence to the same class treatment: This treatment provides in the same way as an outcome the number of terms of the same class between the witness sentences and the key ideas. The results of this treatment will be stored in a statistical table.
- Matching of witness sentences with the key ideas.

Starting from the statistical table of the similarity, the synonymy and the adherence to the same class, the system does a matching between each of the witness sentences and the key ideas to provide as a result a table that contains a list of candidate key ideas that correspond to each of the witness sentences.

Stage 5: Ranking of the key ideas and fitting of witness sentences to the key ideas.

Starting from the result table of the previous treatment and to facilitate the user's choice, the system provides a grading scale of key ideas that correspond to each witness sentence. This could be illustrated as follows:

- If the key idea and the witness sentence share one common term, add 3 points to the score of the key idea;
- If the key idea and the witness sentence share one synonymous term, add 2 points to the score of the key idea;

- If the key idea and the witness sentence share one same class term, add 1 point to the score of the key idea;

SADIM displays the key ideas corresponding to each witness sentence on the basis of the already calculated score. Depending to the user's choice the system matches each witness sentence to a chosen key idea.

5 ASSESMENT OF SADIM

The first assessment of SADIM is based on a test corpus that contains 990 witness sentences (WS) and 390 key ideas (KI). The corpus contains sentences of two types:

- Type 1: Presence of a Common term between KI terms and those of WS.
- Type 2: Absence of common terms between terms of WS and the KI but there can be some semantic ties between these terms.

In this assessment we determine the recall and the precision measures that are extensively used in the domain of information research. We are going to adapt these measures to our diagnosis method in the following way:

$$\text{Recall} = \frac{\sum_i \text{Number of KI correctly generated with SADIM to } WS_i}{\sum_i \text{Number of KI correctly generated with the expert to } WS_i}$$

(Number of KI correctly generated with SADIM to WS/ number of KI correctly generated with the expert to WS)

$$\text{Precision} = \frac{\sum_i \text{Number of KI correctly generated with SADIM to } WS_i}{\sum_i \text{Number of KI correctly generated with SADIM to } WS_i}$$

(Number of KI correctly generated with SADIM to WS/ Number of KI correctly generated with SADIM to WS)

Results of type 1 sentences

For this type of sentences, the results are as follows:

Recall = 94% Precision = 90%

Results of type 2 sentences

For this type of sentences the results are as follows:

Recall = 64% Precision = 50%

According to the previous results we can identify a general recall and precision of the order:

General recall = 79% General precision = 70%

6 CONCLUSION

In this paper we started with a presentation of the concept of diagnosis of the management engineering, then we gave a brief overview of the methods of knowledge extraction from the textual data.

In a latter step we exposed a method of knowledge extraction that permits to solve insufficiencies of the SEGESE tool. This method has led to the emergence of the SADIM system.

We finally made an experimentation of SADIM in order to give evidence to our method contribution.

As perspectives we intend to spread the application of our approach into other domains and to integrate the training and ontological techniques in SADIM.

REFERENCES

Amarnath, G. and al., 2003. "Towards a formalization of disease-specific ontologies for neuroinformatics," Neural Netwoks, Special issue, No. 16, pp. 1277-1292.

Bourigault, D., 1994. "Terminology extraction software for the aid of text-based acquisition knewledge," Thesis deposited in the school of Higher Studies Sciences Paris, 1994.

Chevallet, J., 2003. "Language processing for Knowledge extraction for information research," Third scientific days of young research GEI03.

David S., Plante P., 1990. Termino version 1.0. Rapport de recherche du Centre d'Analyse de Textes parOrdinateurs, Université du Québec à Montréal.

Daille B., 2002. "Mixed approach for terminology automatic extraction: lexical statistics and linguistic filters," Doctorate Thesis, University of Paris VII.

Daille B., Fabre C. and Sebillot P,., 2002. "Applications of computational Morphology: many morphologies," P. BOUCHER, 2002. ISBN 1-57473-125-4.

Jouis, C. 1995. "SEEK, Knowledge acquisition software using linguistic background without using knowledge on the external world," Days concerns acquisition, validation and training, Grenoble..

Kim, J. and Moldovan D., 1993. "Palka: a system for lexical knowledge acquisition," in proceeding of the second international conference on information and knowledge management, pages 124-131, Washington, United States..

Kolsi J. , Belguith L., and Abdelmajid B., 2005. "Conception and development aid system of management engineering diagnosis," Third scientific days of young research GEI05,.

Oueslati, R., 1999."Corpus-based knowledge acquisition aid," Doctorate Thesis, University Louis PasteuStrasbourg.

Savall H. and Zardet V., 1989. "Mastering the hidden coasts and performances: periodically negociable activity contract," preface: Marc andré Lanselle, Edition Economica, edition 2, 351 P.

Savall H. and Zardet V., 2004. "Research in management science: Qualimetric, seeing complex object," preface: BOJE David (USA), Edition Economica, 432p.