IADIS International Journal on Computer Science and Information Systems
Vol. 3, No. 2, pp. 51-65
ISSN: 1646-3692

AUTOMATIC KNOWLEDGE EXTRACTION AND MATCHING: APPLICATION TO MANAGEMENT ENGINEERING DIAGNOSIS

Lamia Hadrich Belguith. *LARIS-MIRACL laboratory. Faculty of Economic Sciences and Management of Sfax* - B.P. 1088, 3018 Sfax –TUNISIA

l.belguith@fsegs.rnu.tn

Jamel Kolsi. *LARIS-MIRACL laboratory. Faculty of Economic Sciences and Management of Sfax* - B.P. 1088, 3018 Sfax –TUNISIA

amel.kolsi@isgis.rnu.tn

Abdelmajid Ben Hamadou. *LARIS-MIRACL laboratory. Faculty of Economic Sciences and Management of Sfax* - B.P. 1088, 3018 Sfax –TUNISIA

Abdelmajid.BenHamadou@isimsf.rnu.tn

ABSTRACT

This paper proposes an approach for automatic knowledge extraction and matching that we apply to an aid system of management engineering diagnosis called SADIM (Système d'Aide au Diagnostic d'Ingénierie de Management). SADIM aims to detect the dysfunctions related to the enterprise management. It allows the knowledge acquisition from textual data related to the diagnosis, the matching and the assignment of witness sentences to the corresponding key ideas. SADIM can also serve as a decision aid system which helps experts and socio-economic management consultants to take decisions that would make enterprises reach the required standards through council interventions.

KEYWORDS

Natural Language Processing (NLP), decision aid system, automatic terminology extraction, knowledge acquisition, socio-economic diagnosis.

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 this objective 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 interest is located at the level of diagnosis. This diagnosis represents an analysis method that is applied in two steps: a qualitative diagnosis detecting the dysfunctions and a quantitative one attributing costs to these dysfunctions which we call hidden costs (Savall and Zardet 2004).

The data relative to this diagnosis are formed of witness sentences fitting key ideas and could hide useful knowledge, dependences or correlations.

Given the important volume of these data, the socio-economic consultants face a problem in their treatment and in the extraction of knowledge from them, particularly in the right matching of witness sentences to the corresponding key ideas.

The available tools proposed in this domain are characterised by the non-automation of the extraction process and the matching of knowledge. This is the case of most systems such as SEGESE (SocioEconomic Management Expert System). The major problem with this system is the key idea redundancy. This 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 key ideas that correspond to the witness sentence. Thus, this situation insides the expert to insert other key ideas.

To solve these problems, we propose an approach for (i) automatic extraction of knowledge from textual data relative to the management engineering diagnosis (i.e. term extraction of witness sentences and key ideas) and (ii) and automatic knowledge matching (i.e. matching of witness sentences with the corresponding key ideas which are related to specific topics and subtopics).

We also present in this paper our system called SADIM (Système d'Aide au Diagnostic d'Ingénierie de Management) which is based on the proposed approach.

In what follows, we present a brief overview of related works on knowledge extraction, then we propose our approach of aid for management engineering diagnosis and finally we present the implementation and the assessment of our system.

2. RELATED WORKS ON KNOWLEDGE EXTRACTION

Automatic knowledge extraction is a very important research domain, including a variety of works that can be classified in two basic categories. The first category includes methods of terminology extraction and the second category includes methods of knowledge acquisition.

Knowledge acquisition is known as the process of acquiring knowledge from a human expert for an expert system, which must be carefully organized into IF-THEN rules or some other form of knowledge representation. However, terminology extraction (also called term extraction) aims to automatically extract relevant terms from a given corpus. These terms are

particularly useful for conceptualizing a knowledge domain or for supporting the creation of a domain ontology. Thus, terminology extraction is very useful for knowledge acquisition.

2.1 Terminology Extraction Methods

Typically, approaches to automatic term extraction make use of linguistic processors (i.e. part of speech tagging, phrase chunking, etc.) to extract terminological candidates (i.e. syntactically plausible terminological noun phrases). Terminological entries are then filtered from the candidate list using statistical and machine learning methods. Once filtered, because of their low ambiguity and high specificity, these terms are particularly useful for conceptualizing a knowledge domain or for supporting the creation of a domain ontology.

The construction of a terminology can not economize a linguistic analysis especially when the terminology expert work is founded on the processing of large corpora, where the linguistic nature of terms is an essential data (Rousselot and Frath , 2002).

Terminology extraction methods could be classified into three main approaches: the structural approaches, the non structural approaches and the mixed approaches (Oueslati 1999).

The structural approaches use two main techniques. Indeed, some structural approaches use syntactic and lexical rules and often require grammars (e.g. approach of TERMINO tool (David and Planete, 1990)). Other approaches use surface analysis and terms contexts and often use the recognition of syntactic patterns. LEXTER tool (Bourigault 1994), for example, is a terminology extraction software that allows to extract terminology from French technical documents. This tool is based on a structural approach using only local syntactic parsing techniques and not a complete syntactic analysis to extract likely terms from a corpus ((Bourigault and Charlet 2005), (Aussenac-Gilles et al, 2003)).

The non structural approaches are based on statistical and quantitative methods (e.g. approach used by MANTEX tool (Oueslati 1999)).

The statistical methods are often achieved on a large corpus and used to construct some statistical models. One of the characteristics of these methods is the use of the threshold notion to filter or to mark up information of the corpus. The quantitative methods use statistical tools; they are neither constrained by the size of a corpus nor by the use of the threshold notion. These two methods (statistical and quantitative) are used for textual data. They require less initial linguistic knowledge than the structural approaches.

The mixed approaches (e.g. approach used by SORT tool (Daille et al, 2002), (Cerbah and Daille 2006)) use the two previously described approaches (i.e. structural and non structural approaches). Some of them apply automatons to labelled corpora before performing a statistical filtering and measuring the distance variation between the simple words of a compound noun. Other approaches apply, however, a statistical location of co-occurrences, then they apply syntactic rules in order to validate or delete collocations (e.g. approach used by XTRACT tool (Smadja, 1993)).

The resulting terms of these terminology extraction approaches can be used by tools of knowledge acquisition. We describe some approaches of knowledge acquisition in the next section.

2.2 Knowledge Acquisition Methods

One relevant way to classify the knowledge acquisition methods from corpora is to oppose numerical versus symbolic techniques. Numerical approaches of acquisition exploit the frequential aspect of data, and use statistical techniques, while symbolic approaches exploit the structural aspect of data, and use structural or symbolic information (Sébillot, 2005).

Within the numerical approach, complex terms or relations between lexical units can be acquired by studying word cooccurrences, and evaluating the strength of the association with the help of a statistical score that detect words appearing together in a statistically significant way (Church and Hanks, 1989). Numerical distributional analysis methods respect a three step approach: extraction of the cooccurrents of one word, evaluation of proximity/distance between two terms, based on their shared or not shared cooccurrents (various measures are defined) and clustering into classes (using data analysis or graph techniques) (Harris et al, 1989).

Note that, numerical methods are portable and automatic but produce non-interpretable results.

Many methods have been proposed in the frame of symbolic approaches. We present in the following some of them.

2.2.1 Term Relation Based Method

This method assumes that in the framework of a construction of a terminological knowledge basis, it is necessary to dispose of not only a list of terms but also the relations maintaining one term with another. Thus, from these relations, it is possible to construct the terminological networks which models are often inspired from those of semantic networks. We can give as an example the Oueslali TERM tool (Oueslali 1999) based on this method.

2.2.2 Contextual Exploration Method

This method uses contextual exploration rules for the text analysis. These rules detect relevant linguistic indices from the context in order to build semantic representation in a progressive way. This is the case of the SEEK tool (i.e. a tool for knowledge acquisition from text) which aims to mark up the contexts of the relations which morpho-syntactic constraints are described with declarative rules (Jouis 1995). These rules do not use predefined knowledge regarding the external environment but refelect a linguistic knowledge that tends to be independent of any particular competence area.

2.2.3 Lexico-semantic Pattern Based Method

The first stage of this method consists in identifying some semantic categories and the second step extracts a schema of lexico-semantic relations. The discovery of semantic categories allows marking a schema that serves for the future extractions of the same categories. The lexico-semantic diagrams describe the lexical and semantic constraints that can denote ways of conceptual propositions of the domain.

SENNET tool (Hearst 2002) represents an example of system using this method.

2.2.4 Template Based Method

The template based method can be used to extract knowledge by considering the following steps:

- Defining templates with slots specifying important data.
- Using a system which extracts important information from a corpus (i.e. appropriate information for the slots) and fill in the templates.
- Generalizing the content of the templates on the system results.

As an example, we can cite the PALKA tool (Kim and Moldovan 1993) which is based on the template based method.

3. PROPOSTION OF A DIAGNOSIS AID APPROACH

In this section, we propose an original approach for management engineering diagnosis aid. Our approach could automatically extract terms from corpora formed of witness sentences and key ideas. The witness sentences describe particular situations or problems and they are stated spontaneously by enterprise staff (i.e. Chairmen, managers, administrative staff and workers) during interviews conducted by contributors. However, the key ideas are formulated by professional experts to represent topics and subtopics that cover all possible problems or situations that an enterprise could face.

In addition to automatic terminology extraction, our approach has also the advantage of automatically match the witness sentences with the corresponding key ideas by attributing fitness scores to each matching. This score measures how well the key idea fits the witness sentence. The aim is to help users (i.e. the stakeholders) to take the best decision regarding the affectation of the witness sentence to the correspondent key idea.

In this section, we first present the corpora that we used to define and to evaluate our approach, then we describe the basic idea of our approach and finally we detail its different stages.

Table 1. Examples of witness sentences and their corresponding key ideas

Witness sentences	Key ideas
« L'ordinateur est dans mon bureau, et avec plusieurs utilisateurs je me retrouve obligé d'accepter l'entrée et la sortie des gens ce qui perturbe énormément mon travail avec bien sûr le bruit de l'imprimante. » The computer is in my office, and with many users, I am obliged to accept the entrance and leaving of people which strongly disturb my work in addition to the printer noise.	*«Un matériel pour plusieurs utilisateurs est une source de perturbation»* One equipment for many users is a source of disturbance
"Avant j'avais un agent permanent, maintenant il est occasionnel et tous les autres services le demandent pour d'autres missions. » Before, I had a permanent agent, now he is occasional and all the other departments require him for other missions.	*«Le manque de personnel est une source de plusieurs dysfonctionnements»* The lack of staff is a source of many dysfunctions

3.1 Our Corpora

Our work is based on two corpora. The first corpus is used to conduct a linguistic study of the witness sentences and the key ideas in order to determine how to extract important terms from them and also to find a way to match them together. This corpus is formed of 700 witness sentences and 295 key ideas. The second corpus is used to evaluate our approach (see section 4.3) and it contains 990 witness sentences and 390 key ideas. All witness sentences of the two corpora are collected by contributors who conducted many interviews with different enterprise staff (i.e. Chairmen, managers, administrative staff and workers). All key ideas of both corpora are elaborated by professional experts. Table 1 presents some witness sentences and their corresponding key ideas.

Note that the key ideas are organised by topic and subtopic. The topics are related to the six main areas: work condition, time management, communication, coordination, concertation, training and strategic implementation. The sub-topics give more precision and hierarchy to the topics, so that they facilitate their reading and their affectation by stakeholders.

3.2 Basic Idea of our Approach

Our approach for automatic management diagnosis aid allows automatic extraction and matching of knowledge from textual data. It is based on natural language processing (NLP) techniques for term extraction from witness sentences and key ideas. Thus, it takes into accounts the following lexical and morphological phenomena to improve the term extraction quality (Kolsi et al, 2006a):

- Abbreviations (e.g. Sté, SGBD, MEO, PC).
- Inflectional paradigms (e.g. "I work", "she works",… have the same lemma "work").
- Derivational paradigms (e.g. "nation" , "nationalisé", "nationaliser", … have the same root "nation").
- Suffixations (e.g. "assembler" + "age" = "assemblage", "exécuter" + "ion"+"exécution")
- Préfixations (e.g. "de"+"faire"="défaire", "re"+"faire"="refaire").
- Compound nouns (e.g. "mise en œuvre", "mise à niveau", "company manager", "ressource management", "business structure", "company structure").

Our approach also proposes the construction of a restricted domain ontology which defines specific semantic relations such as:

- "Sorte de" (i.e. kind of): joins heteronyms to hyponyms (e.g. computer material / printer)
- "Partie de" (i.e. Part of): joins an element to a whole (e.g. diagnosis/ management engineering) .
- "Relation causale" (i.e. Causal relation): establishes a causal relation (e.g. dysfonctionnement/déficit).
- "Action/Objet" (i.e. Action/Object): defines the relation between the action and the object (e.g. crisis/economy).
- "Objet/Propriété" (i.e. Object/Feature): defines the relation between the object and the feature (e.g. inflation/rate).
- "Objet/Procédé" (i.e. Object/Process): defines the relation between the object and the process (e.g. enterprise/State to rank).

Moreover, our approach suggests the use of a statistic technique to rank the different candidate key ideas that are likely to fit the witness sentences. Thus, it attributes scores to the

different key ideas. The key idea that has the best score would be considered as the best candidate to match the considered witness sentence.

3.3 Description of our Approach

Our approach for management engineering diagnosis is based on five steps (see figure 1): 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 each witness sentence, classification of key ideas with reference to each witness sentence and linking witness sentence to key idea.

3.3.1 Step 1: Pre-treatment of Witness Sentences and Key Ideas

The aim of this step is the pre-treatment of witness sentences and key ideas in order to prepare them for the next step. The pre-treatment consists in deleting the separators, the empty words and the multiple spaces (see figure 2):

– Deleting the separators: The system replaces all following separators (: ; . ? ! () [] {} = " * +, ...) except the hyphen (-) by a space since it could belong to compound terms. The objective of this task is to have only one separator that is the space character.

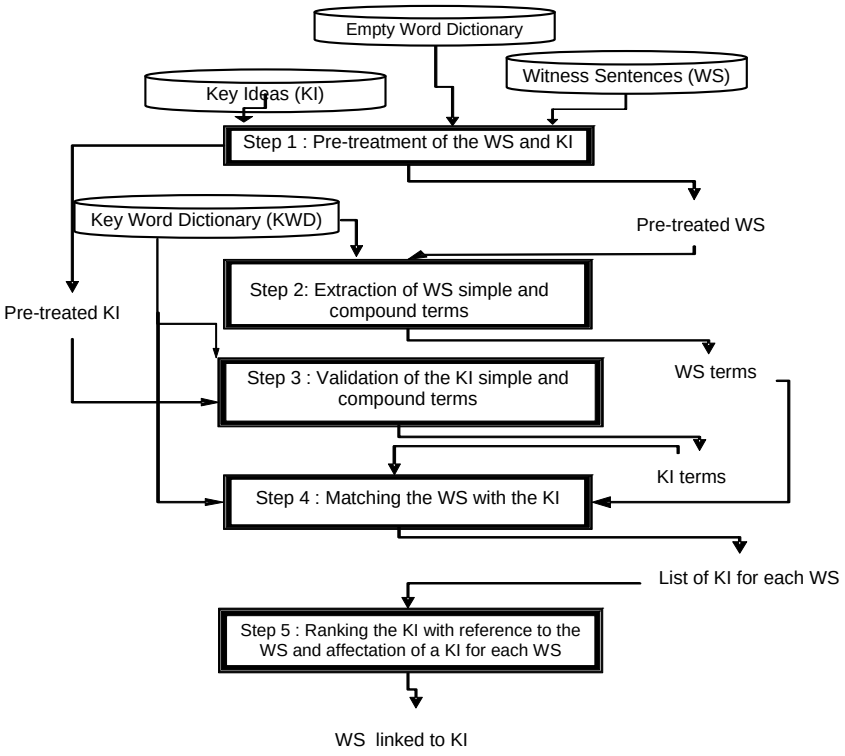

Figure 1. Our approach steps

- 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. These words can be determiners (e.g. "le", "la", "les"), prepositions (e.g. "mais", "pour"), coordination conjunctions (e.g. "et", "ou") and subordination conjunctions (e.g. "que", "qui"). This task deletes all empty words that can occur in witness sentences or in key ideas, with the exception of some prepositions (i.e. "de", "en", "d'","à") since the latter could be a part of compound terms (e.g. "bases de données" databases, "main d'oeuvre" workers). The goal of this task is to reduce the terms number of the key ideas and of the witness sentences in order to facilitate the matching process.
- Deleting the multiple spaces: To unify the word separators by only one space, this task replaces all multiple spaces of the witness sentences and the key ideas by only one space.

3.3.2 Step 2: Extraction of WS Simple and Compound Terms

This step aims to extract all simple terms and also compound terms of witness sentences by using a Key Words Dictionary (KWD). The extraction is based on the following treatments (see figure 3):

- Regrouping three consecutive words from the pre-treated WS term list. If this word grouping exists in the KWD, consider it as a compound term, add it to the list of terms, skip three words and do the same treatment again, otherwise pass to the second treatment. This treatment aims to determine compound terms formed of three words such as "Mise en oeuvre" (i.e. implementation) and "Bases de données" (Databases).

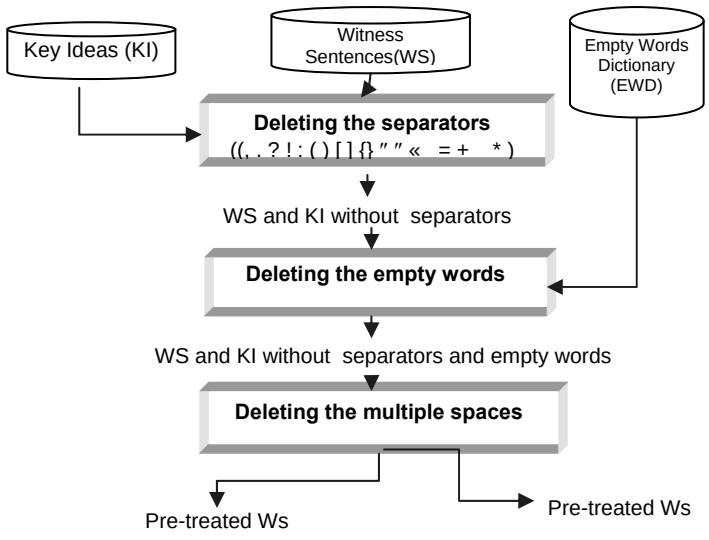

Figure 2. Pre-treatment of witness sentences and key ideas (Kolsi et al, 2006b)

- Regrouping two consecutive words from the pre-treated WS term list. If this word grouping exists in the KWD, consider it as a compound term, add it to the list of terms, skip two words and do the treatment 1 again, otherwise pass to the third treatment. This

treatment aims to determine compound terms formed of two words such as "Double face" and "Coupe courant" (Electricity interruption).

– Taking only one word. If this word exists in the KWD, consider it as a simple term, add it to the list of terms, skip one word and do the treatment 1 again.

3.3.3 Step 3: Validation of the Key Idea Terms

While entering the key ideas, the expert introduces for each key idea, its corresponding terms (simple and compound ones). In this stage, the following treatments are performed for each key idea in order to validate its terms:

– The extraction of the simple and compound terms (the same way the treatment is carried out in step 2).
– Validation of the extracted terms using the key words dictionary (KWD).

3.3.4 Step 4: Matching the Key Ideas with the Witness Sentences

This stage determines the similarity, the synonymy and the adherence to the same class between the witness sentence terms and the key ideas terms. Its main objective is to carry out the matching between each of the witness sentences and the different key ideas.

As shown in figure 4, this stage is achieved for each witness sentence with all the key ideas while taking witness sentence terms, key idea terms and the key word dictionary as an input.

This stage includes the following treatments:

– Similarity treatment: this treatment calculates the number of similar terms between the witness sentence and each key idea taking into account the morphological variations (i.e. "ouvrier" & "ouvrière" are considered to be similar since the second word is the feminine form of the first one).

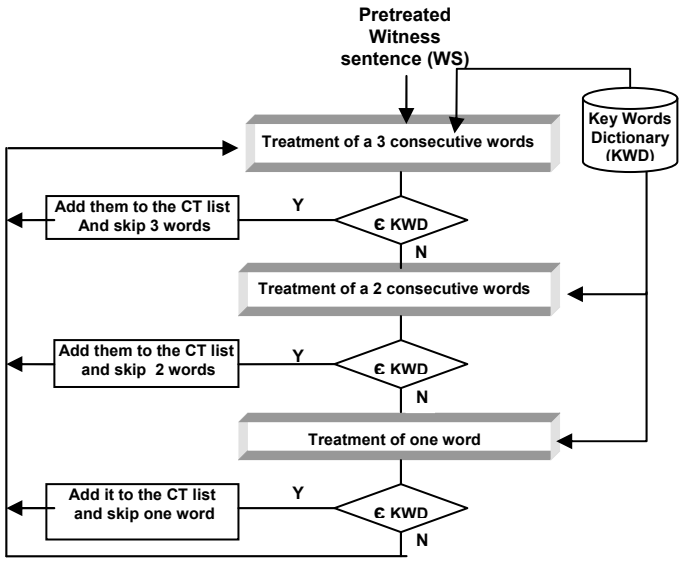

Figure 3. Extraction of simple and compound terms from witness Sentences

- Synonymy treatment: this treatment provides as a result the number of synonymous terms in witness sentence and key ideas (e.g. the terms “manqué” and”insuffisance” are synonyms and means “lack”).
- 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 (e.g. the terms "materiel informatique" and "ordinateur" computer belong to the same class).
- Matching the witness sentences with the key ideas: starting from the result of the precedent treatments (i.e., similar terms, synonym terms and terms adherence to the same class), this treatment does a matching between each of the witness sentences and the different key ideas to provide as a result the list of candidate key ideas that correspond to each of the witness sentences.

3.3.5 Step 5: Key Ideas Ranking and Affectation to Witness Sentences

- Starting from the result of the previous treatment (i.e. the list of candidate key ideas that correspond to each of the witness sentences) and to facilitate the user's choice, we can compute a grading scale (i.e., a score) of key ideas that correspond to each witness sentence.

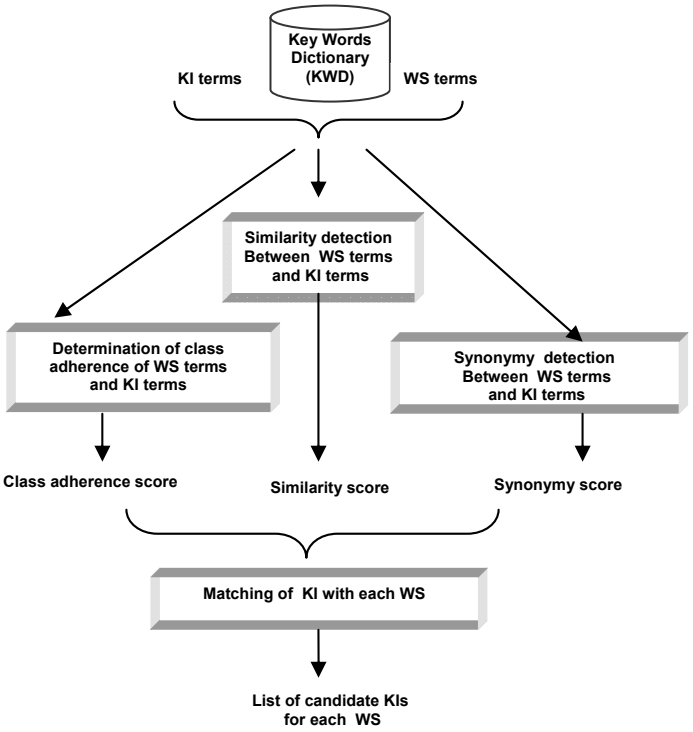

Figure 4. Matching of witness sentences to the corresponding key ideas

Thus, we obtain for each witness sentence a list of key ideas ranked according to the decreasing order of their corresponding score. This would be very helpful for the user to make a final choice. Then according to the user's choice, the witness sentence could be matched with only one key idea.

4. SADIM SYSTEM: IMPLEMENTATION AND EVALUATION

In this section we present SADIM system (a system for management engineering diagnosis aid) that is based on the proposed diagnosis aid approach. We first present SADIM implementation, then we compare it to SAGESE system and finally we present the evaluation of SADIM on a real corpus of witness sentences and key ideas.

4.1 Implementation

SADIM system is developed with the PERL programming language under WINDOWS XP. The choice of PERL could be justified by the easiness that this language offers for the textual data manipulation through the use of regular expressions. SADIM presents a well adapted graphic interface and simple functions to the user. It uses an XML database composed of:

- A Key Word Dictionary (KWD) containing 1250 words. This KWD represents a restricted domain ontology defining different semantic relations between terms relations (see section 3.2).
- An empty word dictionary.
- A key ideas corpus.
- A witness sentence corpus.

SADIM interface offers the following three main options for the user:

- "Gestion des Phrases témoins" (i.e. WS treatment) : this option includes three sub options which are the WS input, the WS pre-treatment and the WS term extraction.
- "Gestion des Idées clés" (i.e. KI treatment): this option includes three sub-options which are the KI input, the KI pre-treatment and the KI term validation.
- "Affectation d'une phrase témoin à une idée clés" (i.e. affectation of a WS to a KI): this option includes two sub-options which are the matching WS to KIs and affectation of a WS to a KI.

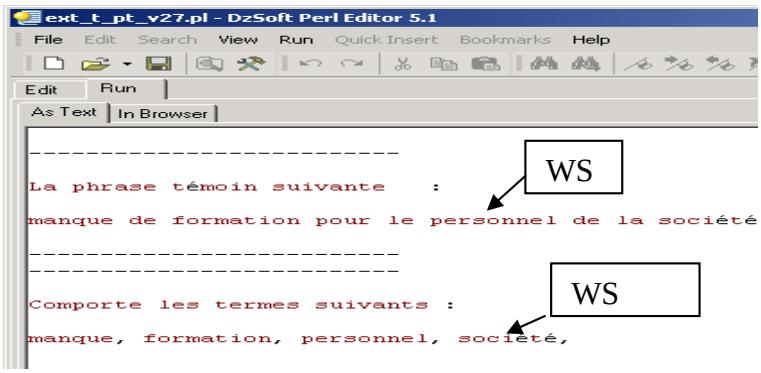

Figure 5. WS Terms extraction

An example of execution of the "WS Terms extraction" sub-option is illustrated by the screen image presented in Figure 5. Figure 6 presents the result of the "Matching the Witness sentences to the Key ideas" sub-option.

4.2 Comparison of SADIM and SEGESE Systems

SEGESE (Système Expert en GEstion Socio-Économique) is a software developed by ISEOR (Institute of Socio economics of Enterprises and Organizations) and adapted to ameliorate the Enterprise productivity and work quality by the diagnosis of global organizational dysfunctions. SEGESE is a decision aid system for consultants and intervening. It was first conceived by Henri Savall, Véronique Zardet and Nouria Harbi in 1985 and then developed by ISEOR.

Since 1987, consultants and intervening-researchers have been making diagnoses in enterprises, which permits an extension and an updating of the knowledge basis incorporated in SADIM. The data include witness sentences and the key ideas corresponding to them.

The main difference between SEGESE and our system SADIM is that SEGESE does this matching manually; however SADIM does the term extraction and the matching of witness sentences to key ideas automatically.

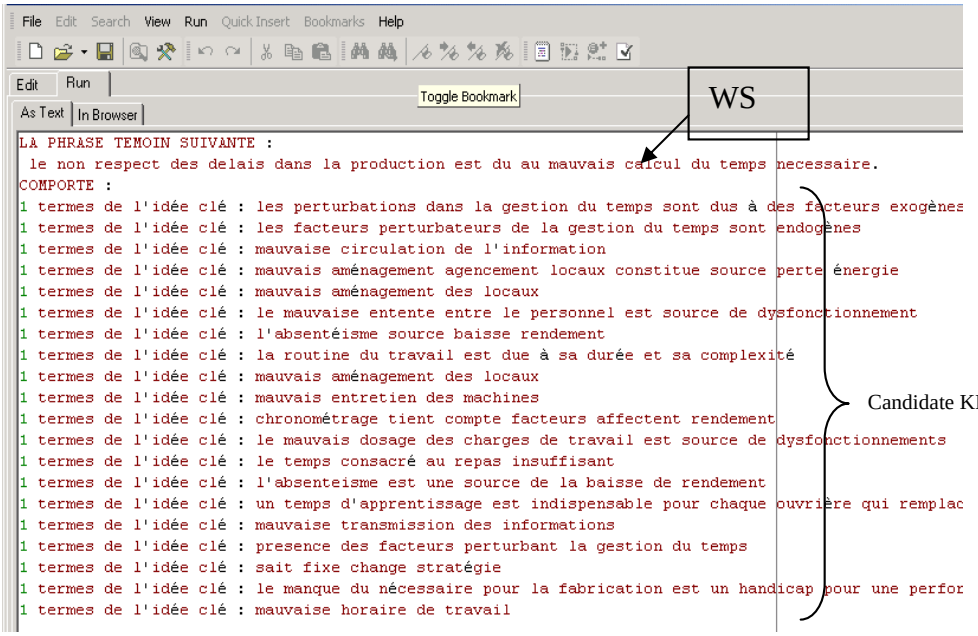

Figure 6. Matching key ideas to the witness sentences

Another difference is that the knowledge base of SEGESE is larger then that of SADIM since the latter is recently developed (compared to SEGESE) and till now it does not cover all management engineering process steps. Table 2 presents a detailed comparison between these two systems.

The important volume of the textual data (i.e. witness sentences and key ideas) may trigger some problems such as the difficulty of making diagnosis and the redundancy of the key ideas.

In this framework, SADIM suggests a solution to those problems by the use of new technologies of knowledge extraction from data. SADIM is based on our proposed approach for term extraction, and matching of witness sentences to the corresponding key ideas.

Table 2. SEGESE and SADIM comparision

Methods	SEGESE	SADIM
Automatic management of witness sentences	Yes	Yes
Automatic management of key ideas	Yes	Yes
Automatic matching of the witness sentences to the key ideas	No	Yes
Treatment of all management engineering process steps	Yes	No
Term extraction	No	Yes
Knowledge extraction from the witness sentences and the key ideas	No	Yes
Automatic matching of the key ideas with each witness sentences and affectation of witness sentence to the corresponding key idea	No	Yes

4.3 SADIM Assessment

To evaluate SADIM, we carried out an assessment based on two types of evaluation corpus:

- Type 1: Presence of a Common term (Example: machines, machine) between key idea terms and those of witness sentences.
- Type 2: Absence of common terms between terms of witness sentences and the key ideas but there can be some semantic links between these terms.

Our evaluation corpus is composed of 990 witness sentences and 390 key ideas.

In this assessment we determine the *recall*, the *precision* and the *F-score* measures that are the most used in the natural language processing and information research domains. Thus, we adapted these measures to our context in the following way:

$$\textbf{Recall} = \frac{\sum_i \text{Number of KI correctly matched to WSi by SADIM}}{\sum_i \text{Number of KI matched to WSi by the expert}}$$

$$\textbf{Precision} = \frac{\sum_i \text{Number of KI correctly matched to WSi by SADIM}}{\sum_i \text{Number of KI matched to WSi by SADIM}}$$

F-score = F-Score = 2* Precision * Recall / (Precision+Recall)

As shown in table 3, the obtained results are very encouraging and prove the high performances of SADIM. Indeed, the overall recall, precision and F-score are relatively 79%, 70% and 74.22%.

Table 3. SADIM evaluation

Corpus	Recall	Precision	F-Score
Type 1	94%	90%	91,95%
Type 2	64%	50%	56,14%
Type1+Type2	79%	70%	74,22%

Note that the failure cases of matching between WS and KI are mainly due to the following problems:

- Missing terms in the KWD,
- Missing term relations in the KWD,
- Complex syntax of some witness sentences (e.g. use of negation forms).

5. CONCLUSION

In this paper we proposed an original approach for management engineering diagnosis. Our approach allows automatic knowledge extraction and matching from textual data. It suggests the use of some natural language processing and statistical techniques for that purpose.

We also presented in this paper our system called SADIM which is based on the proposed approach and aims to detect the dysfunctions related to the enterprise management. Indeed, SADIM allows the knowledge acquisition from textual data related to the diagnosis, the matching and the assignment of witness sentences to the corresponding key ideas. SADIM can serve as a decision aid system which helps experts and socio-economic management consultants to take decisions that would make enterprises reach the required standards through council interventions.

We conducted a comparison between SADIM and SEGESE systems and we proceeded to an experimentation of SADIM in order to give evidence to our approach contribution.

As perspectives we first intend to enrich our Key word dictionary with more terms from the domain and also to add the missing semantic relations between the terms, so that we can have a sort of ontology for our domain. Then as a first stage, we intend to apply training techniques in SADIM, particularly in to the process of matching between WS and KI. We also consider studying the possibility of achieving a distributed architecture for our approach, in order to let SADIM system available for the WEB users.

REFERENCES

Aussenac-Gilles N., Bourigault D., Teulier R., 2003. Analyse comparative de corpus : cas de l'ingénierie des connaissances. *14èmes journées francophones d'ingénierie des connaissances (IC 2003)*, Laval, pp. 67-84.

Bourigault D. and Charlet J., 2005. *Construction d'un index thématique de l'ingénierie des connaissances*, in Ingénierie des connaissances, Teulier R., Charlet J. & Tchounikine P. (Ed), Paris, L'Harmattan, pp. 29-47.

Bourigault D., 1994. *LEXTER, un Logiciel d'Extraction de Terminologie. Application à l'acquisition des connaissances à partir de textes*, PhD thesis, École des Hautes Etudes en Sciences Sociales, Paris.

Cerbah F. and Daille B., 2006. Une architecture de service pour mieux spécialiser les processus d'acquisition terminologique. *Traitement Automatique des Langues (TAL)*, 47(3), pp 39-61.

Church, K.W., and Hanks, P., 1989. Word association norms, mutual information, and lexicography, *Proceedings of ACL'89, 27th Annual Meeting of the Association for Computational Linguistics*, Vancouver, Canada, pp 76-83.

David S. and Planete P., 1990. Termino version 1.0. *Rapport du Centre d'Analyse de Textes par Ordinateur*, Université du Québec à Montréal.

Daille B., 1994. *Approche mixte pour l'extraction automatique de terminologie: statistiques lexicales et filtres linguistiques*, PhD Thesis in computer Science, Paris VII University.

Daille B., Fabre C. and Sebillot P., 2002. *Applications of Computational Morphology: Many Morphologies*, P. BOUCHER Eds, Cascadilla Press, pp 210-234.

Rousselot F and Frath P. 2002.Terminologie et Intelligence Artificielle, *Traits d'union*, Presses Universitaires de Caen, pp. 181-192.

Harris, Z., Gottfried, M., Ryckman, T., Mattick, P.(Jr), Daladier, A., Harris, T.N., and Harris, S., 1989, *The Form of Information in Science, Analysis of Immunology Sublanguage*, Boston Studies in the Philosophy of Science, 104, Kluwer Academic Publisher, Dordrecht.

Hearst M., 2002. Automatic Acquisition of Hyponyms from Large Text Corpora, *13th international Conference On Computational Linguistics (COLING)*, Nantes, pp. 539-545.

Jacquemin C., Bourigault D., 2003. *Term Extraction and Automatic Indexing*, The Oxford Handbook of Computational Linguistics, Oxford University Press, pp. 599-615.

JOUIS, C., 1995. SEEK, un logiciel d'acquisition de connaissances utilisant un savoir linguistique sans employer de connaissances sur le monde externe, *6ème Journées Acquisition Validation et Apprentissage (JAVA95)*, Grenoble, pp. 159-172.

Kim, J-T. and Moldovan D-I., 1993. PALKA: a system for lexical knowledge acquisition, *Second international conference on information and knowledge management*, Washington, United States, pp 124-131.

Kolsi J., Belguith Hadrich L. Mrabet M. and Ben Hamadou A., 2006a. SADIM: An aid System for Management Engeneering Diagnosis Through the use of Knowledge extraction and matching techniques, *8th International conference on Entreprise Information systems (ICEIS 2006)*, Paphos/Cyprus.

Kolsi J., Belguith Hadrich L., Ben Hamadou A., 2006b. Knowledge Extraction and Matching Applying a Qualitative Diagnosis System, *International Conference Applied Computing (IADIS 2006)*, San Sebastien, Spain.

Oueslati R., 1999. *Aide à l'acquisition de connaissances à partir de corpus*, PhD thesis, Louis Pasteur University, Strasbourg, France.

Savall H. and Zardet V., 1989. *Maîtriser les coûts et les performances caches : le contrat d'activité périodiquement négociable*, Economica, Paris, 368p.

Savall H. and Zardet V., 2004. *Recherche en sciences de gestion : Approche qualimétrique (observer l'objet complexe)*, Economica, Paris, 432p.

Sébillot P. 2005. Symbolic Machine Learning: A Different Answer to the Problem of the Acquisition of Lexical Knowledge from Corpora. *TripleC (Cognition, Communication, Co-operation)*, special issue: selected papers from ECAP 2005 - European Computing and Philosophy Conference 2005, 4(2):277-283.

Smadja, F. 1993. XTRACT: An Overview. *Computers and the Humanities*, N° 26, Kluwer Academic publishing, pp. 399-413.