



HAL
open science

Starting from patents to find inputs to the Problem Graph model of IDM-TRIZ

Achille Souili, Denis Cavallucci, François Rousselot, Cecilia Zanni

► **To cite this version:**

Achille Souili, Denis Cavallucci, François Rousselot, Cecilia Zanni. Starting from patents to find inputs to the Problem Graph model of IDM-TRIZ. TRIZ Future 2011, Nov 2011, Dublin, Ireland. hal-01147897

HAL Id: hal-01147897

<https://hal.science/hal-01147897>

Submitted on 2 May 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



TRIZ Future 2011, Dublin, Ireland

Starting from patents to find inputs to the Problem Graph model of IDM-TRIZ

Achille Souili^a, Denis Cavallucci^a, François Rousselot^b, Cecilia Zanni^b

^a *Laboratory for Design Engineering (LGéCo - INSA Strasbourg), France*

^b *Laboratory of Computer Sciences, Image and Remote Sensing (LSIIT), France*

Abstract

TRIZ theory is primarily based on patent's observation and one of its extensions for complex situations is known as Inventive Design Method (IDM). This paper proposes a new approach of automatic retrieval of IDM concepts from patents. It mainly consists of using generic linguistic markers to locate and extract IDM related knowledge, such as problems, partial solutions and parameters, to automatically populate IDM Ontology. Lastly, a comparison is made between respective results of manual and automatic retrieval to assess the current accuracy of our approach.

Keywords: Patent mining; Knowledge engineering; Ontology; Inventive design

1. Introduction

To design new products and methods, industries are on a perpetual quest to find innovative methods and tools to assist R&D activities. On that score, TRIZ, the Theory of Resolution of Inventive Problems which is based on logic, data, and research rather than intuition turns to be a relevant answer for these needs with its three fundamental principles [1]:

- Some problems and solutions are reoccurring in industry and science. A predictive solution to these problems can be found by categorizing the contradictions, which will be defined later, existing in each problem.
- Patterns of technical evolutions may also be recurrent across many industries.
- Creative innovation, representing these technical evolutions, generally emerges outside the field where they were developed.

TRIZ, which is primarily used for technical and physical problem solving, is now applied to solve more non technical domain problems and situations. It has become universal. [2] However, since TRIZ

lacks formalized techniques and concepts, its comprehension is a complex task; and it is difficult to construct a computational design model upon them [1]. On that score, classical TRIZ was extended to the Inventive Design Model (IDM) in order to cope with the disadvantages of classical TRIZ: the inadequacy to solve some problems belonging to different fields and the difficulty to find the most important contradiction. Ontology may be defined as the standard representation of a field or domain of the important categories of objects or concepts which exist in the field or domain, showing the relations between them. A universal ontology was created to interface IDM software for technology intelligence with the premise to evolve beyond the identification of a single contradiction and solve wider and more complex problems and situations [3][4]. This ontology is currently populated manually through discussions between different experts of the application. This process consumes time and money.

This research reports on an ongoing project, whose goal is to populate automatically or semi-automatically populate the ontology and then alleviate the work of the experts. It mainly focuses on the retrieval of three of the IDM concepts that are *problems*, *partial solutions*, and *parameters* [1].

First, we report progress on various contributions made so far on patent mining for TRIZ. Secondly, we expose our methodology based on the use of generic linguistic markers with patent analysis and text mining techniques to retrieve the IDM concepts in question. Finally, a case study on silver defects within a steel making company, Arcelor Mittal, assesses the efficiency of the approach.

Keys

...	any string of character
()	optional
[]	not optional
PREP	preposition
V	inflected form of the verb
ART	article
NC	name introducing a concept of change

2. The state of art on NLP-based patent information retrieval

Patent documents contain important research results that are valuable for product innovation. Despite several Information Retrieval (IR) systems that exist, very few are especially dedicated to patent information retrieval. These IR systems addressing patents usually use hybrid methods by combining statistics and linguistics. They typically apply a linguistic pre-processing including tokenization or sentences. They mostly contain a module of written rules made either by experts of the domain, linguists or both. Among the works dealing specifically with patent information retrieval, some research, rather linguistic, for example [6], examine the specifics related to patents documents classification. NLP is a vital medium to locate needed information in a document database. As such, Feldman's approach [7] for extracting knowledge from unstructured text applicable to patents looks promising. After a basic linguistic

pre-processing, a list of candidate words is produced to be keywords. Then a morpho-syntactic filtering and a statistical relevance filtering are applied to the list which is then presented to the user to help him build the taxonomy of the artifact.

In addition to patent processing, the work of Ghoula [8] is worth recognizing. It presents a processing chain achieving an automatic semantic patents annotation through a structural ontology and domain ontology; biology as regarding their case.

We also have tools such as GATE linguistics platform [9] that can be used for annotations to create a robust and efficient processing tool to address huge amounts of patents.

However, the problem with previous works is that they do not give access to the invention process. They do not take into account artifact improvements. PatExpert [10], a commercial tool also uses a system of rules to produce an automatic abstract from the claims section.

However, most approaches considered thus far focus on the content of the patent which implies they interpret the domain ontology, access the terms of patent or the static description: elements, sub elements and functions specific to a domain whereas, IDM focuses on a dynamic description without any distinction of domain. These approaches aim to facilitate patent surveying by experts familiar with the area of the artifact. In addition to these approaches, we have ways which are not focused on static knowledge. On the contrary, they deal with improvement oriented entities in patents documents. Goujon [6] describes a technology watch system that uses a contextual exploration method [11] to intuitively retrieve concepts related to change, for example, use and improvement. It is not based on any knowledge model. Furthermore, we have SAO-based patent analysis which controls the syntactical structure of Subject (Noun Phrase), Action (Verb Phrase) and Object (Noun Phrase) to explicitly represent relationships between the components of a patent. These approaches are closer to our work in its objectives and use of NLP tools. SAO-based approaches are intrinsically connected to the concept of function that Savranski [12] defines as the “the action changing a feature of any object”, whereas, in [13;14], Action and Subject may refer to components of a system while Action refers to functions performed by and on components. He particularly advocates functional analysis to identify a problem or generate innovative solutions. In functional analysis, a problem is broken down into its component functions, which are further divided in sub-functions and sub-sub functions, until the function level for solving the problem is reached. These interactions are expressed as Subject-Action-Objects triads [14]. In addition, SAO structures prove to be relevant for representation of knowledge related to the inventor’s domain of expertise and the patents key findings [13]. As for Yoon, he presents in [20] a method to automate the identification of TRIZ trends. On the premise that property, which refers to a specific characteristic of a system, is usually described using adjectives; and function, which indicates an action that changes a feature of an object, is typically described using verbs [21], he proposes to use the binary relations of “*adjective+noun*” or “*verb+noun*” to determine specific trends and trend phases, by measuring semantic sentences similarity. We also have [15] that describes an algorithm and a framework to extract information about the properties of a given product or product family through patent analysis for the purpose to find similar products. Of all SAO-based approaches, mentioned above [12;13;14;15; 20; 21], which deal with TRIZ, Cascini’s work [13;14] is the closest to ours. However, despite its tangible efficiency, functional analysis does not always break psychological inertia as desired in IDM-TRIZ [16].

3. IDM – TRIZ

IDM–TRIZ is an extension of TRIZ, the Theory of inventive design, developed to complete and solve the disadvantages encountered with classical TRIZ knowledge; with the need to address wider and more complex problematic situations [3]. More precisely, IDM-TRIZ deals with artifact evolution and assumes that any object created by human beings is the result of evolution guided by objective laws. Let us notice that artifacts evolve according to problems and partial solutions found in the patent document for

example. IDM-TRIZ also considers the creation of an invention as its failure to change in accordance with the above mentioned laws. One the feature is blocked by a technical or physical conflict. The method proposes to clearly formulate conflicts as contradictions that can be stated as follows:

Considering three Parameters (see later) P1, P2, and P3 of an object to be designed or a process to be accomplished, P1 can take conflicting values V_A and $V_{\bar{A}}$ if P1 takes value V_A , P2 is improved, but P3 is deteriorated; conversely when P1 takes $V_{\bar{A}}$, P2 is deteriorated while P3 is improved. V_A and $V_{\bar{A}}$ are opposite values assigned to an action parameter which will be shown later.

Using the example of a wind turbine, P1 = size of the blades; V_A = big; $V_{\bar{A}}$ = short; P2 = Power generated; P3 = Resistance to high winds [4]. Concretely speaking, once a limited set of one to three important contradictions are clearly identified, TRIZ provides the inventor with a technique and knowledge base that will enable him to generate design solutions. These solutions are called inventive, if it bypasses the main contradiction. In the case of a wind turbine, an inventive solution would be a wind turbine with a vertical axis. In other words, IDM-TRIZ is a dynamic assortment of procedures, whose purpose is “to implement a technical system or product design, starting from a domain understanding and modeling up to the synthesis of relevant Inventive Solution Concepts” [1] .

4. Linguistic knowledge base

Linguistic resources are essential for the success of many artificial intelligence tasks. Thus, for the purpose of a fine-tuned parsing at the stage of indexing, an elaborated linguistic knowledge base must be created. This knowledge base comprises lexical resources representing linguistic information about the lexical units such as the ontology and the linguistic markers.

4.1. IDM – TRIZ Ontology.

IDM-TRIZ ontology is a model of knowledge common to IDM experts and practitioners [1;17]. Its structural unit is a concept. This ontology is generic with the premise to be applicable to all patents without any distinction of area. It allows to select among parameters of an artifact which are usually numerous and complex, those which involve a possible evolution in their value. The purpose of this method is to populate the generic IDM ontology with data of a given domain to build the said domain model. IDM ontology, on the contrary of usual static ontologies, describes dynamic knowledge i.e., the impact of changes on each other. The approach tries to follow IDM expert’s one when analyzing patents in the initial phase and indexing parameters that will allow him to find the problem. He must know what the problem behind the invention is and a partial solution to the said problem. Problem as well as partial solution, parameters and elements are the basic concepts of IDM ontology. A problem expresses an unsatisfactory feature within a system. It is generally described by expressions of negative opinion while partial solutions provided by the patent are defined by expressions of improvement and change. Each problem causes one or more contradictions the patent solves. The rhetoric behind the texts of the patents is therefore used to express information such as: “*Considering this artifact*”, “*such defect was found*”, “*this patent provides an improvement which eliminates this problem*”. Considering that elements are components of the system, elements of interest are those with parameters whose values change during the improvement provided by the patent. Parameters qualify the element with certain specifics. There are two types of them: action parameter on which one can act and evaluation parameter on the change of which one can notice. Parameters are usually expressed in patent documents by names, adverbs, or object complements, whereas, partial solution is an element of change or improvement expressing a result know in the domain or proved by experience.

Therefore, depending on the information to retrieve, the following steps are involved in the approach:

- Find the problem of the artifact to which the patent proposes a solution
- Find the partial solution or the improvements provided by the invention

To do this, we must know how these patent documents express such information, by identifying regularities between informational and morpho-syntactic structures of the texts. In other words, we are interested to know how the information is organized in patent document and the linguistic structure used to express such information. An algorithm is then tested and adapted to extract different information. Finally, a hybrid approach combining a statistic method (filtering and statistical extraction) and a module of rules based on linguistic analysis is developed to perform the retrieval.

4.2. Linguistic markers

Since IDM-TRIZ ontology is generic, the regularities we are looking for are generic i.e. [1,18] independent of any domain. There is no use for a domain specific ontology. The preliminary step in the selection of linguistic markers consisted in the study of how patent documents express problems and partial solutions, by finding regularities between the information structure and the morpho-syntactic structure of the patent document. A heterogeneous corpus of 100 patents from various domains was used, for this purpose. To evaluate and validate the markers, a second corpus composed of 87 patents specific to steel making industry was constituted to test the collected markers, and adapt algorithms to extract desired information. From this study, the following observation where made.

Problems can be located through the use of linguistic markers expressing a negative notion. Such markers may be:

Verbs: *blemish, break, bug, cause, crack, damage, defect, deform, degrade, deprive, destroy, deteriorate, disadvantage, disparate, hamper, harm, hinder, impair, smash, spoil, stain, trouble, weaken, fail, degrade, worsen...*

Names: *break, blemish, bug, cause, complication, crack, damage, defect, deficiency, deformity, degradation, deprivation, destruction, deterioration, detriment, difficulty, disadvantage, drawback, drawbacks, failure, flaw, hamper, hampers, harm, impairing, imperfection, instability, limitation, prejudice, problem, spoiling, stain, trouble, weakness...*

A problem can also be located using:

Structures like: “*A problem with [...] is that [...] or “It is known that [...]”*. For instance: “*It is known that [non-metallic inclusions decrease the workability of metals, lead to surface flaws in the rolled product and impair the mechanical properties of the end product]*”

Adjectives or adverbs such as: *critical, difficult, serious, severe, unintended, weakened, undesired, disadvantageous, etc.* or sentence connectors introducing conflict, a problem or condition like *If, however* can also be used to refine the localization.

As for partial solutions, they can be identified by indexing sentences containing verbs that introduce a concept of change e.g. to change, to generate, to enable, to create, to enhance, to improve, to stabilize, to maintain, to emit, to prevent.

Some structures such as the [*Mean+function*] one, represented by *NP +Prep +V_ing* (where *NP* stands for Noun Phrase, *Prep* for preposition like “*for, of, to*”, *V_ing* for Verb in gerundive) may also be very distinctive for partial solutions retrieval. Prepositions “*for, of, to*”, in this case, must express an idea of intentionality, goals or purposes. For instance: “*A transmitting coil is used for generating an electromagnetic field/ A method of minimizing nozzle clogging by supplying inert gas into the nozzle”*.”

Furthermore, expressions like, “*The present invention (relates to | aims at)[...], this invention[...], [...]so as to[...], such as, in order to, so that ...”* are also relevant.

Example: “*In order to prevent oxidation of the surface of the molten steel in the ladle and to maintain the temperature of the molten steel, a molten slag (s) is placed thereon*”

The Claims section is where partial solutions can be found the most. These lists are far to be exhaustive. It may be useful to look for synonyms or use inflected forms to complete it the most possible.

5. Indexing and retrieval

Indexing is performed by using a corpus processing system, for example, Unitex¹. After a pre-processing consisting in transforming a PDF image file into a single file in .txt format with an OCR², the corpus is tagged in XML according to the patent sections that are abstract, claims and descriptions. A patent abstract is a brief text that does not exceed 150 words in length. Its purpose is to enable from a cursory inspection a determination of the nature and the gist of the disclosure, of the artifact. The claims section contains statements about what the invention is and what it does. Claims, together with descriptions form the specifications part of the patent. As for description part, it contains the general background information and progresses to more and more detailed information about artifact or process and its parts.

The purpose of such tagging is to know the most relevant sections in retrieving problems and partial solutions.

One must be able to work with this parser at the level of morphology, grammar, lexicon, and syntax. IR process includes 4 steps which consist of:

- Division in sentences: Using the linguistic parser, the text is divided into parts on the basis of the dot “.” as the sentence delimiter. It is possible to find very long sentences of more than 500 words in a patent text, but all of them end with a dot “.”
- Sub tagging: Using linguistic indicators, the text tagged in XML with annotation graphs.
- Post processing: A second processing is done using the previous results to further refine them. An HTML file is generated after with the relevant sequences which are given a color to enable the expert to validate the results.
- Expert validation: An expert of the domain validates the candidate list and ranks them from 0 to 5, 5 being the most relevant sequence.

5.1. Problem retrieval

According to TRIZ ontology, a problem can be defined as a situation where an obstacle prevents progress, an advance, or the achievement of what has to be done [1]. It is reduced to a single idea [18] and expressed as a sentence:

<subject>+<verb>+<complement>

For example, “*Alumina tends to adhere and accumulate on the surface of the bore of the nozzle*”.

However, it is worth noticing that problems do not always exist in patent documents in this form. More readily, it is formulated in a complex way and with complex sentences. A problem expression must first have reached its maximum decomposition. For instance, “*If the sticking layer is formed in the casting nozzle as mentioned above to cause nozzle clogging, the casting process is interrupted, thereby deteriorating the casting yield.*” can be reformulated into several sentences complying with IDM ontology rules.

¹ Unitex is a corpus processing system, based on automata-oriented technology. <http://igm.univ-mlv.fr/~unitex/>

² Optical Character Recognition

- 1st sentence: *The sticking layer is formed in the casting nozzle as mentioned above to cause nozzle clogging*
- 2nd sentence: *The casting process is interrupted.*
- 3rd sentence: *The casting yield is deteriorated.*

In other words: “*If the sticking layer is formed in the casting nozzle as mentioned above to cause nozzle clogging, the casting process is interrupted, thereby deteriorating the casting yield*” gives “*If [the sticking layer is formed in the casting nozzle as mentioned above to cause nozzle clogging], [the casting process is interrupted], thereby [deteriorating the casting yield]*” hence the need to detect semantic dependences by identifying a predicate in the text and assigning case-roles which can be defined as spatially-distinguished parts of a process.

5.2. Partial solution retrieval.

Partial solutions, as well as, problems have their own syntax complying with IDM ontology [1]. This syntax is:

To <Verb in its infinitive form > + <Complement>

Partial solutions describe a known and verified by experience result in the domain. This is their simplest form stating a progress or an improvement. A solution is robust when it involves the smallest number of new resources available. For instance, “*It is generally known in the steel making industry to add aluminum to molten steel as a deoxidizer to remove oxygen from the molten steel*” when taken as a partial solution would be: “*To<add >+<aluminum to molten steel as a dioxide to remove oxygen from molten steel>*”.

Clearly, the verbs play an important role here since they are semantic centers of sentences. Broadly speaking, partial solution and problems are concepts, in which TRIZ experts need to complete the knowledge base, when solving a problem.

5.3. Problem and partial solution sub-ontology

Any problem stated in the chain gives rise to a partial solution, see Fig. 1. This relation can be expressed by, *this problem “is partially solved by” this* A chain of successive problems can then be created using “implies” and “is decomposed of” links to know the different implications of a problem and/or its decompositions. Such decomposition aims to remove ambiguities which may occur during a generic description containing several sub-problems. Also, let us notice that a partial solution can sometimes “lead to” another problem. However, the relationships between different problems and partial solutions are not always easy to find. A partial solution could lead to another and a problem is not always linked to its implications and/or partial solutions.

Links between problems and partial solutions can be identified through rhetorical structures and sentence connectors such as adverbs and conjunctions. For example, sentences connectors expressing a contrast or undesirable results like, “*in contrast, conversely, yet, despites this, however, as a consequence, etc.*” may provide connection between a partial solution and a problem it induces. Other structures like, “*to prevent the above mentioned, there is proposed ...*”, “*this can be solved...*” may be used to determine the relation between a problem and the partial solution proposed to solve it.

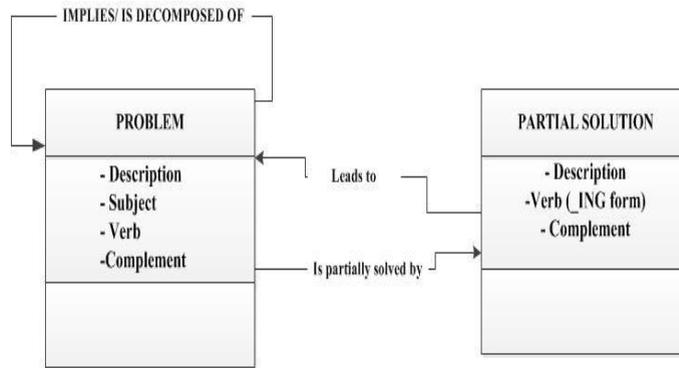


Fig. 1. Problem and partial solution sub-ontology

5.4. Parameters retrieval

Parameters are of two sorts. These are Action Parameters (AP), [1] on which the designer can act and Evaluation Parameters (EP), useful for evaluating the results of a design choice. Parameters can be retrieved using verbs introducing a concept of change like, “change, generate, enable, create, Enhance, improve, stabilize, maintain, emit”. Therefore, depending on the verb used, the parameter placement within the sentence will change. Taking the verbs in the gerundive form, the following results are obtained, see table 1.

Table 1. Syntactic patterns for parameters

Inflected form	Structure	Inflection
Changing, modifying	[...]+for+[V]+(ART)+[P]	Gerundive
Detecting, measuring, collecting	[...]+for+ [V]+ [P] [...]+for+[V]+[P] [...]+for+[V]+[NC]+[PREP]+(ART)+[P] [...]+for+[V]+[NC]+[PREP]+(ART)+[P](...)+and+(ART)+[P]	
Generating, enabling, creating	[...]+[V]+[...]+[PREP]+(ART)+[P] [...]+[V]+[...]+[PREP]+(ART)+[P]+(...)+and+(ART)+[P]	
Enhancing, improving, stabilizing, maintaining	[...]+[V]+(ART)+[P] [...]+[V]+(ART)+[P]+(...)+and+(ART)+[P]	
Generated, emitted	[...]+(ART)+[P]+[V] [...]+(ART)+[P]+and+(ART)+[P]+[V]	Past participle

We have for instance: *Tundishes with partition walls for dividing the tundishes into several chambers or for changing the flowing direction of the metal are already known per se* that matches with the structure [...]+for+[V]+(ART)+[P]. Here “*flowing direction*” is a parameter.

However the above structures are very efficient in retrieving parameter, they are not exhaustive. It can still be expanded patent is a specific document and every writer has his own writing style.

6. Problem graph model filling

Problem graph model is, in actuality, a representation of problems and partial solutions networks. It represents all known problems and partial solutions related to each other by its implication links. The preliminary step in the construction of the graph problem is the validation of the results by an expert. Once the results are validated, rhetorical or transitive structures such as: “however, thus, resulting in, etc.”, used contextually with the list of markers could help build the problem graph. Different types of relationships can be interpreted in the problem graph [19].

- The relationship between a space of problem and a space partial solution can be interpreted as “one can”.

Example: PBI: Thermal expansion generates an uneven roll's profile -> “One can” -> PSI: Create a concave roll in cold situations.

- The implementation of a partial solution can create new problems. This type of relationship can be interpreted as “but then”

Example: PSI: One can create a concave roll in cold situations -> “but then” -> PB2: Strip deviation is observed at start-ups.

- There is a chain of successive problems or successive partial solutions. The nature of this relationship can be expressed as “and thereafter”.

Example A : PBI: Rolls are deformed by thermal expansion -> “and thereafter”-> thermal expansion generates an uneven roll's profile

Example B: PSI: One can create a concave roll in cold situations -> “and thereafter”->PS2: One can create a convex roll in hot situations.

It is worth noticing, that the third type of relationship is to be considered with precaution before being placed in respective networks of problem and partial solutions because, the missing of an element of the chain can impair the quality of the problem graph.

The problem graph is currently built manually and our project is to automate it, in order to provide assistance to the expert. We presently succeed to partially build it automatically from the descriptions section of the patent, with the use of rhetorical and transitive structures mentioned in [5.3]. Let us consider the following example:

« The non-metallic inclusion such as alumina adhered or accumulated onto the surface of the bore is peeled off or falls down, and is entrapped in the cast steel strand, thus degrading the quality of the cast steel strand. To prevent the above-mentioned narrowing or clogging of the bore caused by the non-metallic inclusion such as alumina, there is proposed a commonly used method for preventing the non-metallic inclusion such as alumina existing in the molten steel from adhering or accumulating on the surface of the bore of the nozzle, wherein inert gas is ejected from the inner surface of the nozzle bore toward the molten steel flowing through the bore (for example, Japanese Patent Publication No.{S} Hei 6-59533/1994). However, there are problems of the above mentioned method as described below wherein the inert gas is ejected from the inner surface of the nozzle bore A large amount of the ejected inert gas causes entrapment of bubbles produced by the inert gas into the cast steel strand, resulting in defects caused by pinholes On the other hand, a small amount of the ejected inert gas can not prevent adhesion and accumulation of the non-metallic inclusion such as alumina onto the surface of the bore of the nozzle, thus causing narrowing or clogging, in the worst case, of the bore »

In the previous example, partial solutions are in orange, whereas problems are in blue. The framed words constitutes clues to automatically represent the interactions and/implications of partial solutions and problems contained in the above example. Framed words are tagged with annotation graphs and the results obtained are then processed and push into tables with Perl to be later represented as a problem graph, See Fig.2. However, the algorithms built at this stage of the research, are only able to handle problems and partial solutions belonging to the same paragraph. A good part of the process is currently done manually but we are working to automate it as much as possible.

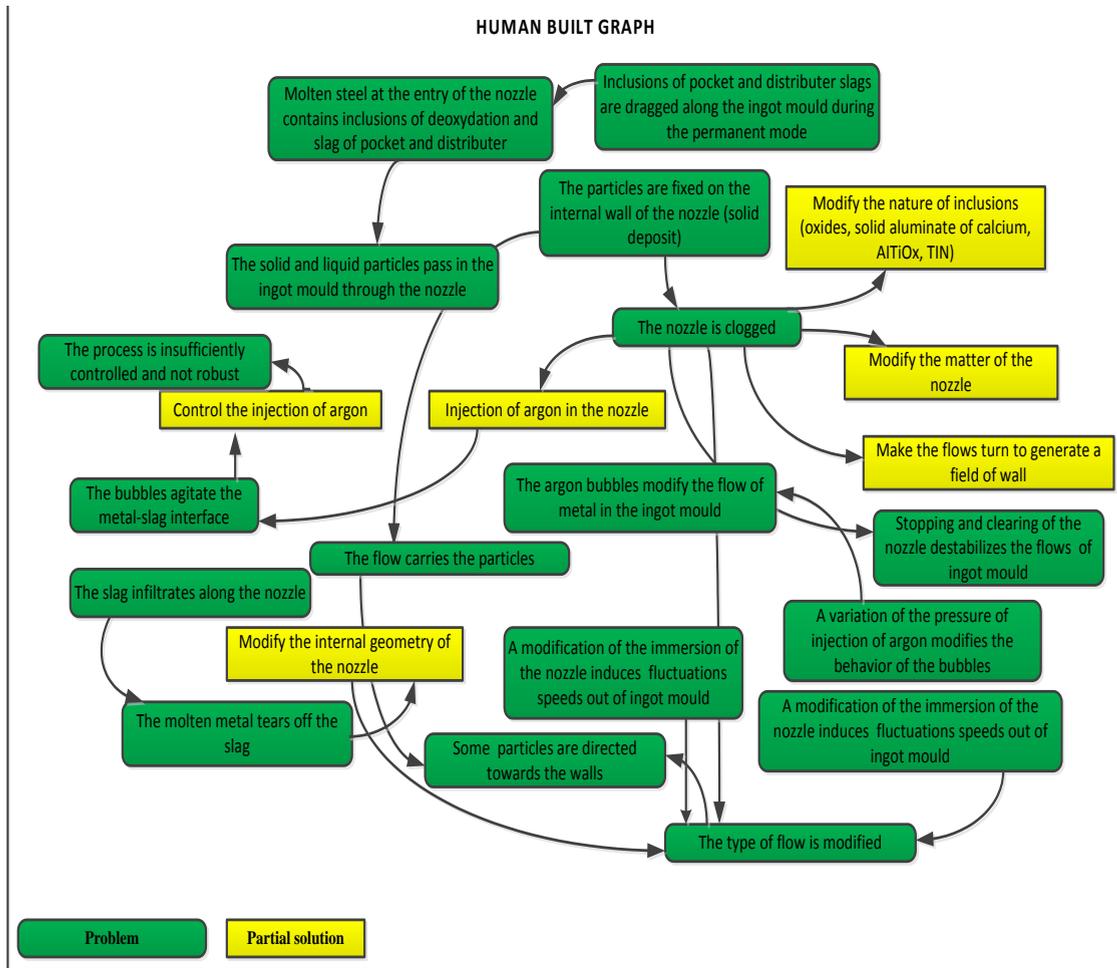


Fig. 2. Human built graph

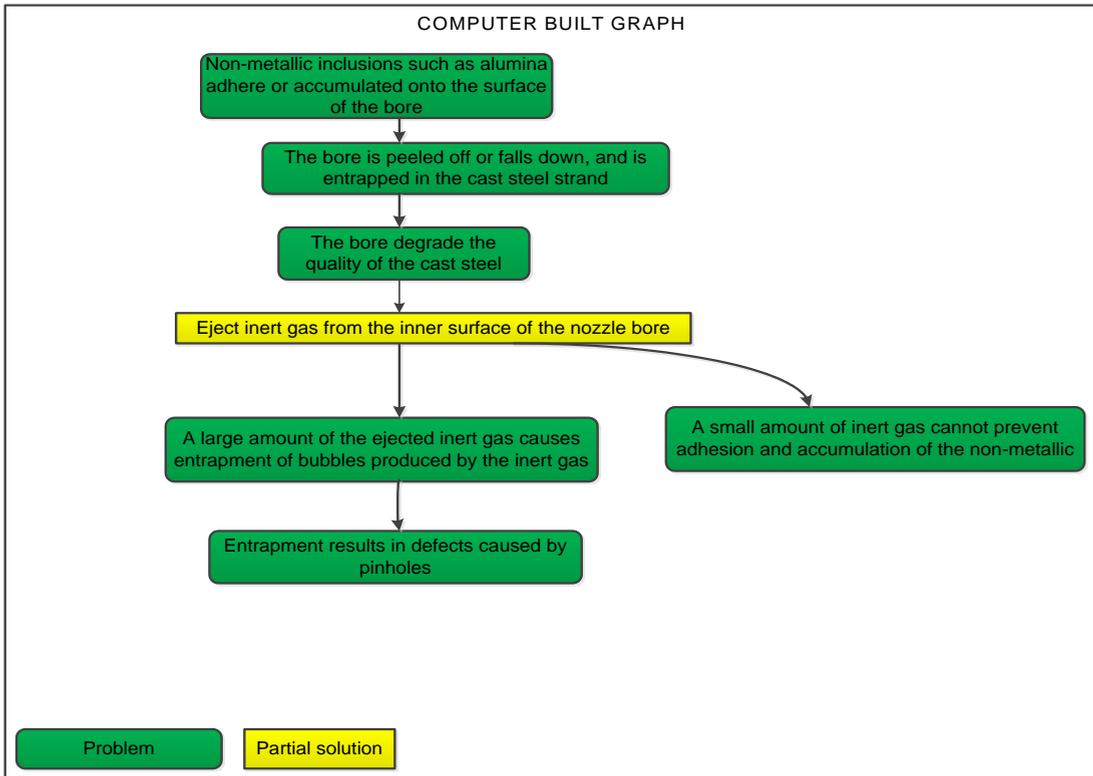


Fig. 3.computer built graph

7. Case example : silver defects in the steel making industry

The objective of this case study is to demonstrate that the markers collected from a previous corpus, constituted from patents of various domains and activities, were effectively accurate and robust enough to be applied to another corpus. This corpus is constituted of 88 patents written in English and available on www.googlepatents.com, www.patents.com, and www.worldwide.espacenet.com. And despite the number of patents it contains, it fits with our needs.

Using Unitex graphs and Perl3 scripts we extracted a primary list of 866 candidate problems and 76 candidate partial solutions. This contrast is due to the fact that we had redundant problems; whereas, for partial solutions we selected the most relevant ones located in the first claim, see table 2.

Table 2. Result according to patent sections

Concepts	Abstract	Claims	Descriptions
Problems	14	18	834
Partial solutions	112	137	27

³ Perl is a high-level, interpreted, dynamic programming language, developed by Larry Wall in 1987 as a general-purpose Unix scripting language.

Even though the quality cannot be thoroughly estimated at this stage of the assessment, it is interesting to notice that the method is efficient. Table 3 summarizes the recall and precision of concepts extracted.

- Recall is the ability of an algorithm to present all relevant concepts

$$Recall = \frac{\text{Number of relevant concept retrieved}}{\text{Number of relevant concept in collection}}$$

- Precision is the ability of an algorithm to present only relevant concept.

$$Precision = \frac{\text{Number of relevant concept retrieved}}{\text{Total number of concepts retrieved}}$$

Table 3. Precision and recall of the results

Concepts	Recall	Precision
Problems	40,29%	87,13%
Partial solutions	32,10%	73,30%
Parameters	49,10%	92,13%

Even though the results are not satisfactory because the recall scores which are low, most extracted concepts are relevant. The precision scores are high, particularly for parameters. Such results can be explained by the choice to use strict algorithms to extract desired concepts. A solution for the improvement of the recall score may reside in a thorough study of patent document structure to built more exhaustive algorithms.

8. Discussion

The goal of our approach is to simplify experts' task, since the collection of IDM-TRIZ concepts currently consumes an exuberant amount of time and money. The evaluation indicates that the quality of the automatic retrieval is not yet high enough to replace the manual retrieval in the population of our ontology. However, a look at the automatic extraction, see Fig.3, shows that even though it is far from perfect, most of the relevant problems found are plausible candidates for IDM ontology population. Patent documents are very specific and can be very heterogeneous depending on the writer's style as sentences may sometimes be very long. Difficulty in the redundancy of sequences matched also impairs the quality of the retrieval. However, we regard this approach for IDM concepts automatic retrieval as a baseline for further work. Even though we are particularly looking to extend our research on other IDM concepts like element, value and contradiction, our concern for the moment is to refine the results and to generalize the construction of the machine built problem graph, see Fig. 3, by taking into account the claims section

which contains most of the partial solutions candidate lists, as this graph is currently built from the description section. Furthermore, although the present approach is generic, it is worth recognizing that it may not be applicable to domains not covered by patents like software companies. Moreover, the list of markers, as well as, are not exhaustive yet, and it could be necessary to extend it as much as possible.

9. Conclusion and perspectives

From technical documentation to the World Wide Web, Natural Language Processing techniques for text mining and information retrieval are powerful tools for data mining and knowledge management technologies; which can also be applied to patent mining. Even though there are many tools addressing patent analysis, most of them are utilizing specific Ontologies. On the contrary, our approach by using generic linguistic indicators to populate IDM-TRIZ generic ontology proves to be a valuable alternative for patent processing; with the possibility to retrieve maximum information from patent documents and applying it to other patents without any restriction of domain. Although, at this stage of our research, results are not perfect, they are promising. Moreover, it is already a step forward to be able to extract the first data of problems formulation directly from the patents. The case study of Arcelor Mittal gave us the opportunity to test this approach. However, for various reasons, e.g., confidentiality of the applications and the cost of experts, it is often very difficult to carry out thorough evaluation. As for contradictions, they are the hardest to find because not all patents solve contradictions and the latter are not clearly expressed. By the evidence of these encouraging results, we are considering to increase our research to improve this method and extend it to other concepts of IDM-TRIZ, in particularly for the retrieval of contradictions.

References

- [1] Zanni-Merk C, Cavallucci D, Rousselot F, Using patents to populate an inventive design ontology, *Proceedings of the TRIZ Future conference 2010*, Bergamo, 3-5 November 2010, published by Elsevier Ltd; 2010; p.52-62.
- [2] Zlotin et al. TRIZ beyond technology: The theory and practice of applying TRIZ to non-technical areas, *Ideation International*, Detroit, MI; 2000.
- [3] Cavallucci D, Khomenko N. From TRIZ to OTSM-TRIZ, Addressing complexity challenges in Inventive design, *International Journal of Product Development*; 2007, Vol. 4, 1/2 , p.4-21.
- [4] Gábor K, Rousselot F, De Beuvron F,Cavallucci D,Wu D. Extraction de connaissances orientées évolution dans les textes techniques. *Proceedings of the Terminology & Ontology Conference (TOTh)*; Annecy, France; 2009.
- [5] Guyot B, Normand S. Le document brevet, un passage entre plusieurs mondes. *Actes du Forum pluridisciplinaire document et organisation. Semaine document numérique*, La Rochelle; 2004.
- [6] Goujon B. Extraction d'informations pour la veille technologique avec le système VIGITEXT. *Actes de RECITAL*, Cargese, France; 1999.
- [7] Feldman R, Fresko M., Hirsh H, Aumann Y, Liphstat O, Schler Y, Rajman M, Knowledge Management : A Text Mining Approach. *Proceedings .of the 2nd International Conference on Practical. Aspects of Knowledge Management*, Basel; 1998.
- [8] Ghoula N, Khelif K, Dieng-Kuntz R. Supporting Patent Mining by using Ontology-based Semantic Annotations. *Proceedings. of IEEE/WIC/ACM International Conf. on Web Intelligence*, Silicon Valley, USA; 2007.
- [9] Cunningham H, Maynard D, Bontcheva K, Tablan V. GATE : A Framework and Graphical Development Environment for Robust NLP Tools and Applications. *Proceedings of ACL'02*. Philadelphia; 2002.
- [10] Mille S, Wanner L. Making Text Resources Accesible to the Reader. The Case of Patent Claims. *Proceedings of LREC. Marakesh :Morocco* ; 2008
- [11] Desclés J-P. Systèmes d'exploration contextuelle ; *Published in Co-texte et calcul du sens (C. Guimier)*, PUC ; 1997, p.215-232.
- [12] Savransky S. Engineering of creativity: Introduction to triz methodology of inventive problem solving. *Boca Raton*. USA; 2000.
- [13] Cascini G, Fantechi A, Spinicci E. Natural language processing of patents and technical documentation. *Lecture Notes in Computer Science*, 3163; 2004, p.508–520.
- [14] Cascini G, Russo D. Computer-aided analysis of patents and search for TRIZ contradictions, *Int. J. Product Development*; 2007, Vol. 4, Nos. 1/2, p.52–67.
- [15] Moehrle M., Geritz A. Developing acquisition strategies based on patent maps. *Proceedings of the 13th International Conference on Management of Technology*; Washington, USA: R&D Management; 2004, p.1–9.
- [16] Zanni-Merk C, Cavallucci D, Rousselot F. KAID: A tool for conducting the use of inventive design in leading complex studies”, *Proceedings of SKIMA*; 2008.
- [17] Simon H. A. The structure of ill-structured problems. *Artificial Intelligence*, 4, 181-201.Also in Cross, N. (Ed.). 1984, *Developments in design methodology Chichester*, England: Wiley; 1973/1984, p. 145-166.
- [18] Cavallucci D, Zanni-Merk C, Rousselot F. Ontology for TRIZ, *Proceedings of the TRIZ Future conference 2009*, Timisoara:Romania,; 2009, p.251-260
- [19] Cavallucci D, Rousselot F, Zanni-Merk C. Procedures and Models for Organizing and Analysing Problems in Inventive Design, *Proceedings of the 19th CIRP Design Conference – Competitive Design*; 2009.
- [20] Yoon J, Kim K. An automated method for identifying TRIZ evolution trends from patents. *Experts Systems with Applications*. (38) 2011
- [21] Verhaegen P. A, D’Hondt J, Vertommen J, Dewulf S, Duflou J. R. Relating properties and functions from patents to TRIZ trends. *CIRP Journal of Manufacturing Science and Technology*; 2009, Vol. 1, 3, p.126–130