Function-based patent search: Achievements and open problems

Article *in* International Journal of Product Development · January 2014

DOI: 10.1504/IJPD.2014.060036

1 author:

Davide Russo
University of Bergamo

Function-based patent search: achievements and open problems

Davide Russo

Engineering Department,
University of Bergamo Italy,
Viale Marconi 5,
24044 Dalmine, Italy
Email: davide.russo@unibg.it

Abstract: Patents are an increasingly important source of technological intelligence that companies can use to gain strategic advantage. They can be used as a stimulus for R&D to search for whether someone somewhere has already solved the problem or a very similar one. In this way, the answer to technical questions depends on how we are able to extract crucial information from the patent corpus and translate it into knowledge. The state of the art of IR tools for patent searches is very rich and in continuous improvement; moreover, current tools are inadequate to satisfy users' expectations. This paper gives a general overview of the universal tools for knowledge management and proposes a combination of knowledge bases, design methods like TRIZ and FBS theory and physical effects for improving function-based patent searches. An example dealing with a new design of nutcracker proposes the use of keywords related to physical effects in order to search a non-clearly expressed function (or behaviour).

Keywords: functional search; patent; IR; physical effect; ontology; FBS; TRIZ.

Reference to this paper should be made as follows: Russo, D. (2014) 'Function-based patent search: achievements and open problems', *Int. J. Product Development*, Vol. 19, Nos. 1/2/3, pp.39–63.

Biographical notes: Davide Russo is an Assistant Professor at University of Bergamo, in Italy. He took his PhD in Machine Design in 2007, and he teaches Systematic Innovation for students and companies. He has over ten years of experience in the field of methods and techniques for innovation and patent searches. He has experienced creative thinking tools and methods aimed at systematising the process of understanding and structuring problems, creating a new potential generation of CAI tools, both for problem solving, for knowledge management and sustainable design. The results of his research activity are reported in over than 50 paper publications and many patents.

This paper is a revised and expanded version of a paper entitled 'Knowledge extraction from patent: achievements and open problems. A multidisciplinary approach to find functions' presented at the '20th CIRP Design Conference', Ecole Centrale de Nantes, Nantes, France, 19–21 April 2010.

1 Introduction

Patent databases have become one of the central knowledge sources of mankind, maintained by millions of contributions per year (Bonino et al., 2010). Despite the crisis, the story of 2011's World Intellectual Property Indicators report is an exceptionally positive one (WIPO, 2011). The challenge is to manage this gigantic source of knowledge by extracting structured information from patent texts and by making this information easily accessible in an automatic way by means of computers.

Much of this information is expressed in natural language texts; in order to retrieve it we need efficient tools and collaborative work from several heterogeneous fields such as computational linguistics, informatics, and engineering. Patent databases and research groups are involved in innovative projects to deliver meaningful search results, (i.e. the CPC Cooperative Patent Classification project between EPO and USPTO[1], or the European project PATExpert dealing with semantic processing of patent documents[2]). Commercial searches are also following the same trend; however, few of the real needs of full text searching have been comprehensively addressed and implemented (Adams, 2010). According to Adams (2010) an effective search engine may need to be developed which operates on entirely different models and may perhaps not be word-based at all.

The main hypothesis of this work consists of assuming that one way of improving search engines is to develop new knowledge bases for IR, such as glossaries, thesauri, or ontologies.

In the following section we have classified patent analysis tools into three categories: (a) a bibliometric and visual analysis, (b) Information Retrieval (IR) models and data mining tools, and (c) semantic searches. For each, limits and trends are given.

Section 3 is dedicated to the knowledge bases and in particular to a proposed set of ontologies coming from design but potentially useful also for patent analysis, such as the ENV model from OTSM-TRIZ (Cavallucci, and Khomenko, 2007) and the FBS model framework by Gero (1990). The aim of this work is to make people working on information extraction aware of achievements so far and open problems, providing both some personal suggestions and an inedited ontology conceived to create functional search targets. In Section 4 a survey is proposed that deals with attempts to face a still open issue relating to the function-based search, showing how a function could be extracted by overcoming linguistic problems by means of thesauri or patent classifications, verb classifications, lexical and semantic relationships, ontologies, and creativity methods. Finally in the last section we propose a software framework called KOM (Knowledge Organising Module) based on an integration of a plurality of selected ontologies, IR models, and data mining techniques and conceived to support patent search and classification of the results.

2 Patent analysis tools

2.1 Bibliometric tools and visual patent analysis

Usually, traditional patent analysis utilises bibliometric data. Bibliometrics is a set of methods to quantitatively analyse, explore and organise large amounts of information so that researchers can identify 'hidden patterns' to support their decision making (Norton, 2000). Searching bibliographic data includes the ability to search for companies and inventors, but this task is not quite as simple as typing the company or inventor name into a field. There are many reasons for this; for example the frequent changes of patent ownerships; the company suffixes or company subsidiaries; the fact that the inventor name can be spelled in a variety of fashions, with or without suffixes or initials or misspelled altogether (Litvin, 2004). Bibliometric software for patent analysis has been proposed in the last years thanks to the ability to detect errors in documents during the DB creation and to provide keyword synonyms, thesauri, and similarity tools but also thanks to a new generation of patent graphical interactions (Koch and Bosch, 2011). Indeed, reading text takes time, especially when high quality text analysis is of relevance. Hence, interactive visualisation, if well-orchestrated, offers a good opportunity for shortening analytic cycles during patent analysis, thus helping users to build their queries faster and leverage mining techniques that are not very common in patent analysis since they are difficult to understand and control without the provision of visual feedback (Bernaras et al., 1996). Techniques for visualising textual information are described in the work of Börner et al. (2003). When using only a small part of the information of a patent, however (i.e. authors, affiliations, technological field, cluster, citations, and so on), limitations in terms of their explanatory and creative capacity are evident and the scope of analysis and the richness of information are limited (Lee et al., 2009). This could be the reason why in the literature it is still difficult to find a set of standard functions based on bibliometric techniques. It is possible to believe, however, that in the future bibliometric applications will continue to grow, improving patent citation strategies as a means for key-patents identification or exploiting PC (patent classification) clusters. Indeed PC code is an interesting source for setting linguistic domains, full of keywords that are independent of the specific language of both the documents retrieved and the user (Lyon, 1999). PC description contains a very interesting set of structured keywords and expressions that have already been filtered by patent office specialists. Thus, the language of IPC descriptions reflects typical jargon of every specific technological domain, and could be used as a support to build conceptual thesauri.

2.2 IR models and text mining

Information Retrieval (IR) is ‘finding material (usually documents) of an unstructured nature (usually text) that satisfies an information need from within large collections (usually stored on computers)’. Numerous recent developments in mining techniques (text mining, TM, in particular) have extended the horizon of research to the unstructured textual data and, therefore, the whole text of a patent document. Applying TM in patent analysis allows us to handle large volumes of patent documents and extract some meaningful implications from textual data. Applications of TM techniques to assist the task of patent analysis and patent mapping are countless, and thus it may be useful to classify patent searches into four types:

- *Known item search*: when the user is searching for a target that is already known (Swartout and Tate, 1999);
- *Exploratory search or browsing*: when the user is learning about the topic but does not know what may be important to correctly define the goal (Lee et al., 2006);
- *Exhaustive search*: when all information is useful for learning about the topic.

Each goal needs a specific strategy. In order to reach this goal, many different IR models have been developed. In synthesis we can group these models, distinguishing between:

– *Boolean IR* and

– *Ranked IR*

Boolean IR is the simplest model for searching for if whether or not a document matches or not the query. All terms are weighted equally, sometimes enriched by proximity operators but always producing results that cannot be ranked.

On the contrary ranked IR provides a numeric score of how likely it is that a document will be useful to the searcher and how well it matches the query.

One of the most popular applications of ranked IT techniques is to monitor the significant frequent terms by means of classical methods such as TF/IDF (traditional Term Frequency/Inverse Document Frequency) in order to improve patent retrieval activity (Li et al., 2009).Since the 1990s, probabilistic methods have been introduced in computational linguistics. Using these methods, the documents can be ranked by their probability of being relevant to the query.

The last IR models, presented in this survey, deal with the vector space model. Here vectors are used to represent documents and queries. A query is treated as a small document and is transformed into a vector. Similarity between document and query is graphically visualised as an angle between two vectors, and the score is calculated according to it. This technique is useful for all types of search and in particular for arranging data into predefined groups, as in classification, or into groups that are not previously defined, as in clustering.

Working on patent searches means not only finding the most relevant documents but also deducing patterns or knowledge from data. Text mining approaches for patent analysis are more articulate and complex. They comprise:

- *Classification*: which arranges the data into predefined groups. Common algorithms include Nearest Neighbour, Naive Bayes Classifier, and Neural Networks.
- *Clustering*: which is like classification but the groups are not predefined, so the algorithm will try to group similar items together.
- *Regression*: which attempts to find a function which models the data with the smallest error. A common method is to use genetic programming.
- *Association rule learning* is a method for discovering interesting relations between variables in large databases. Some well-known algorithms are Apriori, Eclat and FP-Growth. Unfortunately they are algorithms for mining frequent itemsets found in a database but they don't generate rules from them.

This survey is not exhaustive but, regardless of the purpose and the reason why text mining is used, the results of these approaches are strongly influenced by the fact that assignees may use a huge set of different terms to describe the same thing.

The reasons why they use different terminologies are due to: (a) a lack of standard keywords: experts in different domains use dissimilar words; (b) the lack of standard names of developing technologies; (c) incorrect translation: there is not a standard translation among different languages; (d) some inventors and assignees do not want to be completely transparent even when they choose to file patents (Li et al., 2009); (e) the claims, which precisely specify the boundary of the invention and are thus valuable for TM, are generally written in arcane legalese so that it is difficult to extract technical meanings (Lee et al., 2009).

To partially overcome language barriers and to generate more relevant results, a domain thesaurus is proposed.

A *thesaurus* is a hierarchical structure which classifies domain terminologies into different concepts. It is very common for specific terms to be covered by common dictionaries but they cannot distinguish between different domains of use (meanings). Therefore a domain thesaurus is needed for machines to process a specific domain corpus to understand the meaning of these terminologies. An important source for building a patent-oriented thesaurus is the patent classification codes. Patent offices assign a code to each technology and mark each patent with one or more of the appropriate class codes. The schedules of class definitions form a controlled vocabulary of generic terminology for each category of technology. Other taxonomies and controlled vocabularies are currently set up for specific tasks like bio sequence searching, nomenclature, and chemical structure searches, such as the chemical substance registry index (Gruber, 1993) or the Derwent world patent index.

2.3 Semantic search

The semantic search is used to describe a search that goes beyond simple matching of query and document keywords. A semantic technology focuses on taking the meaning of the words into account, as opposed to 'just counting' them.

When we talk about semantics in patent search we have to distinguish two cases: (a) techniques that rely on compression or clustering of the term/document matrix, like latent semantic analysis, and (b) techniques using ontologies, thesauri, or taxonomies to augment the process of identifying and indexing potential search terms extracted from documents (aiding cross language interoperability, query expansion, and searching by concepts instead of terms, as well as broadening or narrowing of the search) (Stefanov and Tait, 2011).

One of the cornerstones of the semantic approach is syntactic analysis or parsing. It is the process of analysing a text, made of a sequence of tokens (for example, words), to determine its grammatical structure with respect to a given (more or less) formal grammar. The fundamental idea behind syntactic parsing is that groups of words may behave as a single unit or phrase (i.e. noun-phrases), called constituents. Generating constituents from a sentence allows us to model constituent facts (concepts).

By means of a syntactic parser, for example, users can break down a text into SAO triads (noun phrase – verb phrase – noun phrase), which is interesting for those looking for problem reformulation and technology transfer (Litvin, 2004). More specifically there are TRIZ software applications that use syntactic parser for automatically reformulating a problem statement in natural language or Boolean query into functional relationships (US2005/0114282).

A semantic search helps to disambiguate queries, to use natural languages instead of Boolean searches, and to formulate research according to a function-oriented approach (Anaya et al., 2010). Even if semantic searches can nowadays be considered the most advanced tool for information extraction, the results cannot be considered satisfactory if compared with potentialities, especially for clustering and content analysis. One of the most important limitations is that actions in SAO triplets can be not functional as in 'people waiting for the bus', and not all design concepts are expressed by a sequence subject-action-object. Section 4 will show how many different expressions can be used to describe a same function without using conjugated forms of verbs.

It should be noted that the history of SAO triads extraction from patents was written not by users from inside the information community or even by patent experts but by researchers working on problem solving design, mainly from the TRIZ community (Altshuller, 1984). In 1991 Universal Semantic Code was used to improve knowledge description and to solve intellectual problems (Boiko, 1991; Tsourikov, 1993), while in 1997 the first prototype working on semantic SAO model analysis appeared; it was called Techoptimizer (*US 5901068*), now known as Goldfire Innovator.

Then, the SAO model evolved rapidly thanks to new and important project works for setting up knowledge bases like domain vocabularies, taxonomies, thesauri, and especially new, dedicated ontologies (Zanni-Merk et al., 2011).

3 Knowledge bases for information searches

As Bates (Tsourikov, 1993) points out, 'the probability of two persons using the same term in describing the same thing is less than 20%'. As such, attempting to directly match user query terms against data set terms is likely to give bad results. This brings us to the need for a knowledge base that can distinguish different senses of words and can relate concepts that are semantically similar. Knowledge bases can be classified as follows:

- *A controlled vocabulary/glossary* is an alphabetical list of terms and definitions used to reduce the variability of terminology use;
- *A taxonomy* is a knowledge hierarchy where items are connected to each other by parent-child, part-of, or instance-of relationships. Classification hierarchies like the IPC are a kind of taxonomy;
- *A thesaurus* is a reference work that lists words grouped together by hierarchical, equivalence, or associative relationships. Synonym dictionaries used by patent searchers are a kind of thesaurus;
- *An ontology* is a taxonomy with precisely defined links between the items that presents knowledge as a set of concepts and their relationships. Different kinds of ontologies are suitable for different purposes.

With regard to the word 'ontology' it is useful to also cite other meanings; here a list of pragmatic definitions is given:

"Ontology identifies the basic terms and relations of a given domain, thus defining the vocabulary, and rules for combining those words and that relationship, going beyond the vocabulary itself" (Neches et al., 1991);

"Ontology is a hierarchically structured set of terms to describe a domain that can be used as a foundation for a knowledge base"(Swartout and Tate, 1999);

"Ontology is a mean for describing explicitly the conceptualisation behind the knowledge represented in this base of knowledge" (Gruber, 1993).

According to the definition of ontology and the meaning of conceptualisation, we can deduce that an ontology consists of general terms that express the main categories in which the world is organised (such as thing, entity, substance, person, physical object, etc.) or specific terms describing a particular domain of a specific application (domain ontologies). Ontology is also the definition of a term and the relationship between words.

Hence, the ontology is used for:

- a common lexicon: the description of a target domain requires a lexicon shared among the people involved. A major contribution is given by terms in an ontology;
- an explanation of what is left implicit in all human activities: there are explicit assumptions and implicit assumptions (i.e. the definition of common terms, relationships, and constraints between them, and different points of view in interpretation of phenomena);
- knowledge structuring: well-established concepts and vocabulary are required by which people describe phenomena, theories, and so on;
- an ontology, therefore, provides the backbone of the systematisation of knowledge;
- a meta-model: a model is generally an abstraction of a real object.

In this work the author's goal was to identify some inedited knowledge bases conceived for supporting IR and semantic searches in patent DB, as shown in the next section.

3.1 Design ontology

Among the enormous number of ontologies available for information extraction, it was decided to recommend for patent analysis two particular ontologies coming from the design world: the ENV and Function Behaviour Structure (FBS) models. A new version of FBS, called the Function–Behaviour–Physical Effect–Structure (FB-PE-S) model, is then proposed.

Even if these ontologies were not conceived for IR purposes, their versatility was checked in the past in various applications: for automatic extraction of contradictions, to find invention peculiarities, to conduct supervised technological benchmarking, to build a network of inventive solutions, and finally to identify innovative technological opportunities (Cascini et al., 2011; Cascini et al., 2007; Cascini and Russo, 2007).

3.1.1 ENV model

The ENV (Element, Name of the property, Value of the property) model is a universal model proposed in OTSM-TRIZ (Cavallucci and Khomenko, 2007) for describing a system or a problem. The structure has been derived from a well-known model in artificial intelligence object-attribute-value.

The Element (E) is any kind of item in the system under analysis (both material and immaterial). The Name of the property N indicates any characteristic, feature, or variable which can be associated to the element E.

Whatever the property is, it must have at least two possible values (V); that is, the element E can assume at least two possible states distinguished by different values V1 and V2 of the property P.

3.1.2 FBS model

Gero (1990) proposes the FBS model in order to describe the design process. Variables and design choices are grouped by three classes of variables describing different aspects of a design object:

- Function (F): 'is the motivation for technical system existence', that is, what it is for (Gero and Rosenman, 1990).
- Behaviour (B): 'is defined as sequential changes of objects state governed by the laws of nature; B is the link between function and structure. Different behaviours can produce the same function, as well as different structures can be characterised by the same behaviour', that is, what it does (Gero and Rosenman, 1990).
- Structure (S): describes the components of the object and their relationships, that is, what it is.

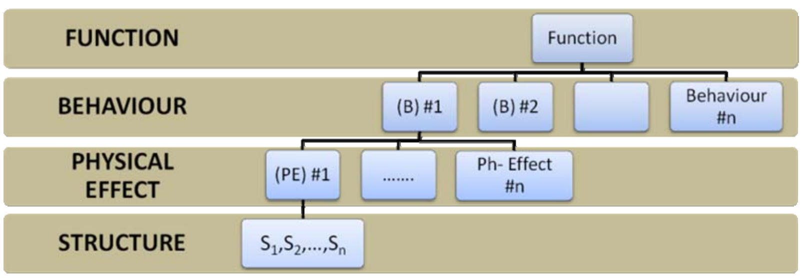
Figure 1 An example of ENV application for a motion description: 'a tool moving an object', it is a tool that changes the value of the object's (E) speed (N) from zero (V1) to a certain value (V2) measured in km/h (see online version for colours)

3.1.3 FB-PE-S model

Some of Gero's original definitions were improved (Cascini et al., 2011) with the aim of creating targets for the following patent extraction while the structural level is further specified to better classify results. Physical Effect (PE) has been separated from behaviour (B).

From the function (F) to the structure (S), the level of specification of the product increases and the relation between two consecutive levels is one to many (Figure. 2).

Figure 2 FB-PE-S architecture based on a revised version of Gero's FBS ontology (see online version for colours)



Some working definitions of the FB-PE-S ontology are proposed here:

- *Function (F)*: The function of a technical system is the motivation for or purpose of its existence, that is, what it is for. The designer specifies the requirements in terms of functional concepts. Therefore, function should represent the designer's intention, given as the requirements as in (Shimomura et al., 1998).

  Every product has a main function. According to Umeda and Chakrabarti (Chakrabarti et al., 2005; Umeda and Tomiyama, 1995), it is difficult to clearly distinguish this function from behaviour. Their definitions do not represent function independently of the behaviour. Both the function (F) and behaviour (B) of a system describe what the system does, but function is intentional and is identified at a higher level of abstraction than its behaviour ('open the nut instead of cracking it').

- *Behaviour (B)*: According to Gero, the behaviour (B) describes attributes that are derived or expected to be derived from structure (S) variables of the object, that is, what it does. This definition is difficult to turn into target keywords and use in a patent digging activity. For this purpose, it is preferable to consider behaviour (B) as a sequential change of states as defined by Umeda and Tomiyama (1995).

  The behavioural level is based on the network of alternative behaviours derived from the same functional concept. Our B-level is built by starting from the identification of the system function (F) and generating all possible ways in which it is possible to achieve the design purpose defined by the function. For example, a razor is conceived to cut hair (F), but can work via many behaviours (B), such as hair extraction, hair breaking, hair killing, inhibition of hair growth, and so on.

- *Physical effect*: To better understand this level, the concept related to the physical phenomenon must first be introduced. A physical phenomenon is the cause of a state transition from state one to state two. Thus a behaviour can be described by its initial state and a set of physical phenomena (Umeda and Tomiyama, 1995).

  The physical effects (PE) are the laws of nature governing change, so a physical phenomenon is associated with a given PE. Activation of a PE is necessary to create physical phenomena and changes of state (Chakrabarti et al., 2005). The PE can be described quantitatively by means of the physical laws governing the physical quantities involved (Pahl and Beitz, 1977). Thus the friction effect is described by Coulomb's law, $F_F = \mu * F_N$ and the expansion effect by the expansion law $\Delta l = \alpha * l * \Delta\vartheta$. If we take into account systems for noise abatement (F), the keywords identifying PE are the sound absorption, the sound insulation index, the dynamic rigidity, and so on.

  All PEs and laws are related with physical/experimental coefficients or quantities. Thus the specification of keywords for the patent digger can be collected a priori in a specific library regardless of the context in which they are used. Then a list of potentially related PEs is identified for each behaviour.

  In order to systematise this activity, more DBs of effects have been merged to offer a reference framework: fields, substances, and properties are organised by groups (mechanical, acoustic, thermal, chemical, electric, magnetic, electromagnetic, biological), and for each, a list of specific phenomena is suggested.

- *Structure (S)*: This describes the components of the object and their relationships, that is, what it is (Gero and Kannengiesser, 2004). The authors further specify this level by adding the concept of design parameters.

All transformations provided by behaviours (B) by means of PE in order to achieve the design task (F) are realised thanks to the system structure (S). The transformation is done by modifying at least one design parameter. For example, in order to increase the cutting efficiency of a razor, many design parameters can be changed, such as the blade sharpness or inclination, the number and distance of blades, and so on.

Figure 3 Function decomposition: each branch represents a different way, expressed by an action, to obtain the same result of the given main useful action (see online version for colours)

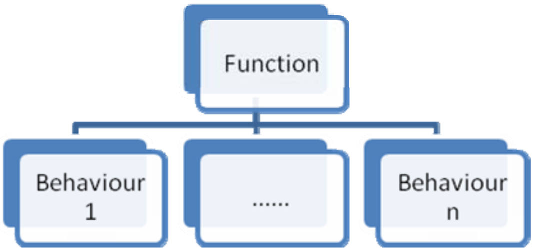
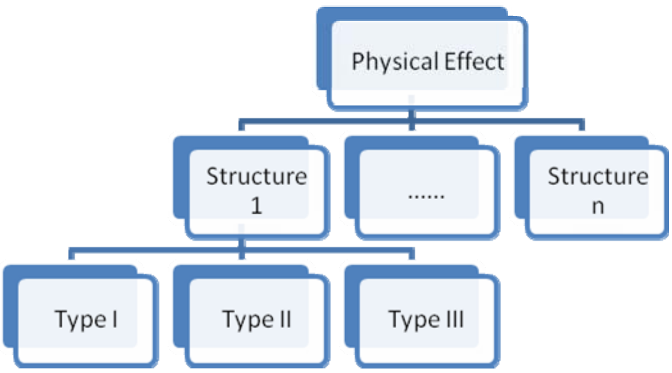


Figure 4 Level of physical effect classification: every function is decomposed into a list of physical/chemical effects by which the related function can be provided (see online version for colours)

Figure 5 Level structure classification: the ultimate leaf of each branch contains the structure using the specific phenomenon by which the function is provided (see online version for colours)



In order to better classify and specify all the workability directions, the authors have created a further classification of this level based on modified design parameters. Thus the design parameters are divided into three different types, as follows:

- *Type* 1: parameters/variables concerning the interaction between the selected object and the other elements of the system.
- *Type* 2: parameters/variables describing the object regardless of the context (system in which it is placed) and concerning design choices for manufacturing and dimensioning.
- *Type* 3: parameters/variables concerning physical properties of the object, that is, constituting the material, physical state, density, and so on.

4 Function-based search: achievements and open problems

Working on function is a crucial point for extracting every kind of information from patent documents.

Dozens of attempts are reported in the literature about how to define an accurate set of verbs to identify inventive principles. Such verbs are used as keywords constituting queries during information retrieval works.

Others function-oriented searches, based for example on SAO extraction, are numerous but in general they cannot offer good results without an accurate definition of all the keywords related to the given function. For these reasons, it is usually suggested that the best expert in the identified leading area should be found and that these activities should be supported using professional DBs.

Whatever activity is proposed, identifying and managing the right keywords to find patents dealing with a given function is a task requiring sensitivity and experience.

Indeed searching for a function does not only mean finding synonyms (verbs with the analogous meaning that, substituted for the given one in the same context, do not change the right value of the proposition) but also involves exploring a variety of linguistic forms and relationships among words in order to throw up the related implicit knowledge function.

There are many reasons for the complexity of this work:

1 Every inventor has his or her own style, and the same concept, represented by a given verb, could be expressed at a more abstract level of detail (by means of a less technical lexicon) or at a deeper level (by a very specific jargon);

2 people coming from different areas (technical fields, geographical or cultural backgrounds, etc.) use different expressions and thus different keywords to express the same concept;

3 concept represented by an action can also be expressed by other syntactic categories. Consider the case of ‘an object moving to …’. The verb ‘to move’ can be found in a text in different forms such as:
   - a functional verbal form (an object moves to..)
   - an adjective (a movable object…)

– an adverbial form (the object is mounted movably …)
– a paraphrase (an object is able to move…)
– a noun (move, motion, …) indicating the result of the action on the object.

4 Although the aim of a patent should be to share knowledge, this is not always the purpose of the inventor, who may prefer to write the patent by hiding its content. The best way to hide it is by abstracting both elements constituting the system and its function. Abstraction is also good to extend the claimed patent field.

5 When a function is obtained by a logical sequence of actions, the patent writer could totally or partially omit them, using just one instead of all, or simply citing the most abstract one. For example, 'a laser beam lighting a surface to generate both an overheating and a chemical decomposition, so causing a localised ablation' can be otherwise expressed as 'laser cutting'. This can also alter the results of the best search motor.

6 When a user explores or browses the patent DB, the search goal is not totally defined. Since it is difficult for the user to find something that he or she does not know about, creative work is needed in order to create new potential targets.

This incomplete list could be expanded but it is already sufficient to show how many things can affect the outcome. Since problems are a combination of linguistic and technical problems, solutions will also be a combination of linguistic and technical approaches, as presented in the next paragraph.

The goal of this work is to demonstrate as knowledge basis can be used to improve recall without introducing noise so the precision does not drop.

4.1 Using dictionaries and conceptual thesauri

Classical dictionaries are undoubtedly the main source of suggested words, both in the definition and in the synonyms list of the given verb. Using these words the initial query can be enlarged to include similar terms. The way to combine similar terms can be provided by several tools (it is better if this is done semantically) but query expansion is always done by combining such words through Boolean operators.

For example, using *Collins Thesaurus of the English Language* and searching for a definition of 'to move', over 100 alternatives are suggested, as shown in Table 1. Most of them do not belong to the same domain, however, and they have to be manually eliminated.

A semantic approach instead, can distinguish the domain and filter exclusively the alternatives linked to the motion domain.

Semantic dictionaries such as WordNet 3.0 (Miller et al., 1990), offer an exhaustive list of synonyms and also other related syntactic forms as shown in Table 3.

Now, once a function is given, a rich set of keywords is easily available for Boolean combination without introducing too much noise into the search results.

To get even more results it is necessary to enlarge the set of synonyms used for queries to those related to the domain of the given function. In other words a conceptual thesaurus is requested: a dictionary with words (verbs, nouns, adjectives, adverbs, etc.) describing technological areas related to the given word.

Table 1 A list of synonyms from Collins Thesaurus dictionary

From Collins Thesaurus of the English Language – 2002	
Verb: TO MOVE	
1	Transfer, change, carry, transport, switch, shift, transpose.
2	Go, walk, march, advance, progress, shift, proceed, stir, budge, make a move, change position.
3	Take action, act, do something, take steps, take the initiative, make a move, get moving, take measures.
4	relocate, leave, remove, quit, go away, migrate, emigrate, move house, flit, decamp, up sticks, pack your bags, change residence.
5	Change, shift, convert, transform, alter, diversify.
6	Progress, develop, advance, make progress, make headway.
7	Change your mind, change, shift, reconsider, budge, climb down, do a U-turn, back-pedal, do an about-turn, change your tune, do an about face.
8	Drive, lead, cause, influence, persuade, push, shift, inspire, prompt, stimulate, motivate, induce, shove, activate, propel, rouse, prod, incite, impel, drive stop, prevent, discourage, deter, dissuade.
9	Touch, affect, excite, impress, stir, agitate, disquiet, make an impression on, tug at your heartstrings (often facetious).
10	Circulate, mix, associate, go round, hang out, socialise, keep company, fraternise.
11	Propose, suggest, urge, recommend, request, advocate, submit, put forward.

Table 2 Some examples of semantic categories of the verb concepts from WordNet 2.0

Name	*Contents (verbs of)*
Body	grooming, dressing and bodily care
Change	change of size, temperature, intensity
Cognition	thinking, judging, analysing, doubting, etc.
Communication	telling, asking, ordering, singing, etc.
Competition	fighting, athletic activities, etc.
Consumption	eating and drinking
Contact	touching, hitting, tying, digging, etc.
Creation	sewing, baking, painting, performing, etc.
Emotion	feeling
Motion	walking, flying, swimming
Perception	seeing, hearing, feeling, etc.
Possession	buying, selling, owning and transfer
Social	political and social activities and events
Stative	being, having, spatial relations
Weather	raining, snowing, thawing, thundering, etc.

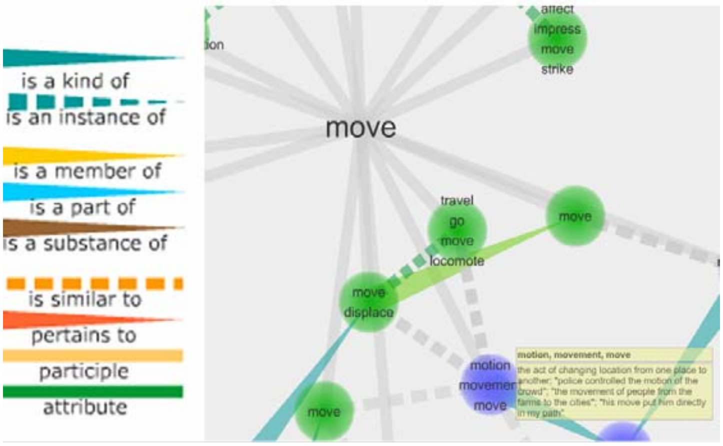
Table 3 A selected list of synonyms filtered by the most relevant synsets from Wordnet 3.0

From WordNet Search - 3.0	
Verb: TO MOVE	
<verb. Motion>S:	
(v)	travel, go, move, locomote (change location; move, travel, or proceed, also metaphorically)
(v)	move, displace (cause to move or shift into a new position or place, both in a concrete and in an abstract sense)
(v)	move (move so as to change position, perform a non-translational motion)
(v)	move (change residence, affiliation, or place of employment)
Verb: TO MOVE	
<noun. Act>S:	
(n)	move (the act of deciding to do something)
(n)	move, relocation (the act of changing your residence or place of business)
(n)	motion, movement, move, motility (a change of position that does not entail a change of location)
(n)	motion, movement, move (the act of changing location from one place to another)
(n)	move ((game) a player's turn to take some action permitted by the rules of the game)

Unfortunately, commercial software applications for building automatic conceptual thesauri are not reliable yet, but there are some, such as Visuword, that can offer some help. A picture taken from Visuword is shown in the following figure.

Conceptual thesauri can offer a set of terms to be added by Boolean operators to the initial query (or to that already expanded by synonyms).

Figure 6 Screenshot of a search about 'to move' taken from the online graphical dictionary Visuword at http://www.visuwords.com (see online version for colours)



4.2 *Using patent classification as a source for contextual synonyms extraction*

While awaiting a new generation of conceptual thesauri, another way to generate contextual synonyms has been identified. For the author a linguistic approach alone will never cover all technical expressions, especially those dealing with a very specific jargon. There are a lot of technical areas from which a patentee can draw terms to describe his invention that no dictionaries could entirely cover. In order to overcome this problem and still build reliable thesauri, patent classification codes (such as IPC, ECLA, and US classification) are taken into account.

Each full digits patent class defines a certain technological area. Although it is not always the case, in most cases a patent class circumscribes a typical jargon. Words extracted from these patent class definitions can be used as representative keywords to start building a thesaurus. From those keywords, all strategies already presented in the previous sections can be exploited and integrated.

Indeed, by browsing semantically inside patent classes it is possible to define unobvious links among clue terms (verbs, adverbs, nouns, and adjectives) having similar meanings exclusively in that specific context.

For example, no dictionary could connect the verbs 'to cut' and 'to heat' as synonyms but they can be considered synonyms in a specific context like laser cutting.

4.3 Using verbs classification as NIST

Unfortunately patent writers do not always use specific language, but sometimes prefer a more general language to better claim the idea of the patent.

A list of verbs prepared by the National Institute of Standards and Technology (NIST) of the US Department of Commerce (Hirtz et al., 2002) can be used to abstract the given function.

Instead of using a linguistic approach it is suggested that a more technical approach should be considered.

NIST developed a hierarchical taxonomy following the approach proposed by Pahl and Beitz (1977). The NIST taxonomy provides a set of terms that are atomic but also generic enough to allow modelling of a wide range of engineering systems.

The NIST list of verbs is classified into three levels with increasing degree of specification. Once the given action is fixed, a similar verb has to be chosen from those on the list. Looking at the hierarchical classification it is sometimes possible to go back to the more abstract forms of language. An example of a hierarchical organisation dealing with the verb 'to branch' is presented in Table 4.

Table 4 An example of a 'branch'-function decomposition taken from NIST – a functional basis for engineering design

Primary	*Secondary*	*Tertiary*	*Correspondents*
Branch	Separate		Isolate, sever, disjoin
		Divide	Detach, isolate, release, sort, split, disconnect, subtract
		Extract	Refine, filter, purify, percolate, strain, clear
		Remove	Cut, drill, lathe, polish, sand
Overall increasing degree of specification			

4.4 Using lexical and semantic relationships

In the previous paragraphs, different ways to collect synonyms are presented. The best dictionaries provide a plurality of definitions already classified by domain; they provide synonyms and sometimes some other related terms. Since this is not enough, it was considered useful to provide a detailed set of the most important lexical relationships for aiding patent information retrieval. The aim is to provide the minimal necessary linguistic knowledge to complete the set of keywords given by dictionaries and thesauri by ourselves.

A good thesaurus contains not only synonyms but also:

- *Antonyms*: words with opposite or nearly opposite meanings. For example: '*increase*' and '*decrease*', '*dead*' and '*alive*', '*short*' and '*tall*'. 'Antonym' has also been commonly used as a term that is synonymous with 'opposite'; however, the term also has other more restricted meanings. The antonym of a word 'x' is usually 'not x' but sometimes that may not be true. 'Rich' and 'poor' are antonyms but 'not rich' does not mean being poor.
- *Pertainym*: This relationship does not belong to verbs but only to adjectives or indirectly to adverbs. There are adjectives defined by 'meaning relating to or pertaining to' and they do not have antonyms. An adjective dealing with pertainym can be related to a noun or another adjective of this type.

Here, a set of semantic relationships follows:

- *Hypernyms* and *hyponyms* are words that refer to a general category and a specific instance of that category respectively. For example, 'vehicle' is a hypernym of 'car', and 'car' is a hyponym of 'vehicle'. Hypernym/hyponym pairs can be found in text corpora by looking for certain syntactic patterns.
- *Meronymy* is a semantic relation used in linguistics. A meronym denotes a constituent part of or member of something. That is,

  X is a meronym of Y if Xs are parts of Y(s), or

  X is a meronym of Y if Xs are members of Y(s).
- *Holonymy* is a semantic relation. Holonymy defines the relationship between a term denoting the whole and a term denoting a part of or a member of the whole. That is,

  X is a holonym of Y if Ys are parts of Xs, or

  X is a holonym of Y if Ys are members of Xs.

  For example, '*tree*' is a holonym of '*bark*', '*trunk*', and '*limb*'.
- *Entailment* is the relationship between two sentences where the truth of one (A) requires the truth of the other (B).

  A verb X entails Y if X cannot be provided without Y being provided.

  Entailment is a unidirectional relation: if X entails Y, the contrary is not true, but it does not happen if two verbs are synonyms. For example, the action of preparing a nut consists of a first stage of the cracking shell and extracting the nut. This relationship could be used to decompose the given function.
- *Troponymy* is a particular entailment relationship:

  If X is troponym of Y, X also entails Y.

If '*to limp*' is a troponym of '*to walk*', I can't limp without walking, so '*to limp'* entails '*to walk*'.

– *Causal relation* is *similar* to entailment but without time aspects.

  If X causes Y, X also entails Y, where X is a verb indicating the cause or the activity related to the verb Y.

  Entail and 'cause to' are unidirectional: Feeding causes a person to eat; the fact that a person is eating does not mean that someone is giving that person something to eat.

4.5 Using design ontologies

Even if we were able to define every kind of linguistic relationship, it could be difficult to achieve the goal. This is because in most cases patents are not written in a functional language. For such a reason only a radically different approach can recognise a specific action in a text without it being described as a verb.

The *ENV model* pushes the user to consider a function from another point of view, not only as a combination of words related to the action but as a means to change the parameters of an element.

Every system can be reduced to a minimal set of elements constituting the system and conceived in order to provide a function to an object. Due to this function the object becomes a product. This transformation can be described as a triad SAO like 'to change, increase, decrease, or maintain the value of a feature of an element of the system'. Thus, the concept of function can be recognised just by looking at specific set of verbs in combination with technical parameters.

Research exclusively based on linguistic relations around a given word is now enlarged to encompass new information dealing with the object and its transformation.

FB-PE-S ontology helps to find a function when the given function is not present in a text but is decomposed by different actions, such as in the case of relationships of 'entailment' and 'causality'.

It is also used when the given function does not appear at all in the text but is substituted by means of:

– its physical principle (PE) (i.e. 'centrifugal force' implies the action of 'twisting')
– its structure (S) by means of a specific design parameter (a 'wavelength of 260 nm' instead of 'sterilising', a 'wavelength from 315 to 400 nm' instead of 'bronzing'),
– a representative technological tool ('using a knife' substitutes for the action of 'cutting', 'X-ray' for 'scanning', 'microwave' for 'heating').

In some cases it is easier to find specific word dealing with the structure or with a specific phenomenon or a physical principle rather than looking for an action that could be missing or too general to be found (implicit knowledge). FB-PE-S offers a way to change strategy or to integrate it into the traditional function-based queries. For example, in US4358467 the claimed function is 'to cut the shell' but instead the writer prefers to use both very general terms (such as 'removing shell') and contextual synonyms of 'laser cutting' (such as 'to burn').

Table 5 Two examples of a patent analysis made by decomposing the patent according to FB-PE-S ontology. The terms in bold are those automatically selected by the author's software framework

Patent number #	*FB-PE-S*
US4358467 – […] The removal of shells from hard shelled nuts, particularly macadamia nuts, is accomplished by rotating the nut in the path of a high power cw laser beam, such as a CO_2 laser beam, so as to burn a path around the shell which separates the shell into parts which can readily be removed from the nut.	F: to cut shell B-PE: Thermal field: to heat/ to burn S: Laser/ laser beam / Co2 laser
JP56001849 – [..] In the method, the coffee beans are suspended in a liquid and then the suspension is irradiated by means of ultrasound with a *frequency* of between 13 and 100 kHz, with sufficient energy to give rise to cavitation phenomena.	F: to consume shell B-PE: Mechanical field: cavitation S: Ultrasound/ frequency/hertz/
CN 86205713 – The utility model discloses a husking device for hard and crisp husk nuts of pine nut, acorn, hazelnut, sunflower seed, etc. The machine is designed by applying the collision theory, which mainly comprises a hopper, a husking breaking unit, a kernel dividing unit, a husk out groove and a kernel out groove	F: to brake shell B-PE: mechanical field: collision/hit S: hammer/…
FR2607670 A1 – The subject of the invention is a method for opening shellfish and the apparatus for implementing the method. It is composed of a frame 12 which can receive a plate 2 on which the shellfish are arranged, a means acting as a lid 9 forms, in a leak tight manner, the space 8 which is subjected to vigorous depressurisation and/or ultrasound either through a vacuum pump 6 or by an emitter 13	F: to channel B-PE: mechanical field: differential pressure S: vacuum

4.6 Using creativity methods

In browsing and exploratory searching, creativity methods can be introduced to create new targets to search (Russo et al., 2011) and organise alternative systems.

'Alternative systems' means all variants of a given technical system that are able to achieve the same goal. Every alternative system changes at least one element of the FB-PE-S model, but the most radical alternatives deal just with modification of function and/or behaviour.

Instead of searching for variants, as is done in traditional patent search, the aim of the creative approach is to support the user in inventing all the ways the goal can be reached by the system we want to find. This approach can generate a bigger set of functional keywords in order to increase the recall.

For example a patentee that looks for 'cracking a nut' can use the creative approach for expanding the query with 'levering', 'drilling' or 'cutting'. These variants are not directly linked with 'cracking', therefore they cannot be contained inside synonyms KB, so they are difficult to find. A research in A47J43/26 (the nutcracker patent class) has been conducted to show how many patents can be missed without considering the actions of 'cutting', 'levering' or 'drilling'. Patents dealing with these actions are 856 and more than 70 patents cannot be found by searching only with 'cracking' and/or 'cracking' expansion (i.e. 50 out of 320 patents dealing with cutting do not contain crack or any crack expansion, 15 out of 400 dealing with levering and finally 8 out of 136 about drilling). Variants can also be exploited for classifying results. According to FBS, 'drilling, cracking, cutting and levering' are different behaviours of a same function: 'to open'. They can be organised in a hierarchical diagram and then used for example for searching when a freezing effect (PE) is used for cracking a nut instead of levering it.

A combination of three different creativity approaches is proposed for systematically inventing new F/B/PE-based variants as shown in detail in (Russo and Montecchi, 2011b).

In order to briefly show how the procedure works, a case study about a nutcracker is provided: "*We want to shell walnuts to be sold; to do that, it is essential to keep intact their kernels. How can we extract the kernel keeping it intact?*"

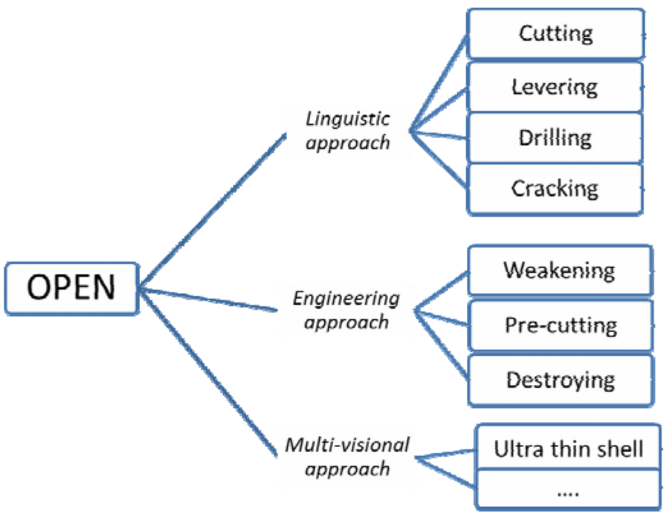
In this case it is essential to know if somebody has already patented an idea for solving this problem and which alternatives are present at the state of the art (inside or/and outside patent domain dealing with nutcrackers).

The first step of the procedure is to fix the minimal description at the highest abstraction level using a minimal SAO triplet: 'the nutcracker – opens – a nut'.

Then, by using three creative methods shown in (Russo and Montecchi, 2011b) the user has to imagine how it is possible to open a nut by (a) browsing lexical and semantic relationships of the verb 'to open', (b) by applying a list of Inventive Standard Solutions (solution paths) derived from TRIZ, and finally (c) identifying new points of view about the given problem situation by abstracting the situation at a higher level and by changing its temporal dimension.

Partial results are shown in Figure 7. This list is not exhaustive because 'to open' is a function that can be realised by a huge numbers of behaviours, depending on how much great is our creativity.

Figure 7 Tree-diagram represents the functional decomposition of a nutcracker. Targets on the right leaves are generated by applying three different creativity approaches (see online version for colours)



After alternative ways for opening a nut are found at the behavioural level, a specification of these directions is possible, by adding information about which PE can be used with them.

Using the effects DB it is possible to specify these directions according to the typology of interaction with the nut. Therefore, at general level, the behaviour of cracking can be performed mechanically, acoustically, thermally, chemically, etc.

Every branch of the tree is finally composed of a list of keywords containing behaviours at different levels of specificity, an expansion of these behaviours and a list of keywords related to the effect used.

An algorithm automatically elaborates a complex query for each branch, combining the target keywords belonging to the same branch in order to obtain the most pertinent results by means of a patent search.

The exploratory search on the nutcracker is thus turned into a list of functional and behavioural targets; all queries are generated according to this functional decomposition. The deeper the decomposition is, the more finely classified are the results.

5 Case study

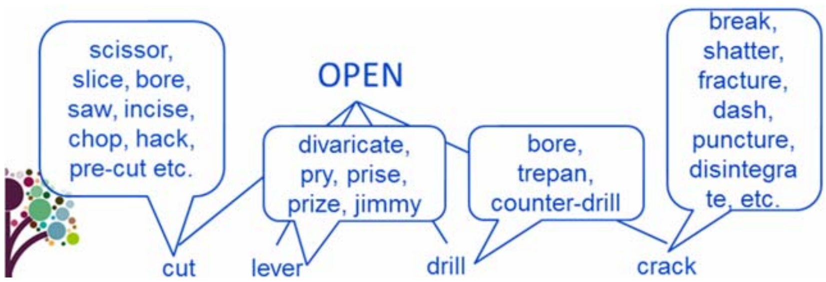
In order to demonstrate how the proposed approach can help a patent searcher, a case study using patent documents as the research data is provided. In this case study, patents related to nutcracker are selected for experimental purposes.

The research was carried out on a pool of 1,302 patent families pertaining to the cracking of nuts. This pool was manually created and it contains only documents in English full-text belonging to the US, EP, WO, GB patent offices and A47J43/26 or A23N5 IP classes.

Searches have been conducted using a patent search engine called KOM (Russo et al., 2012; Russo and Montecchi, 2011a). It is a concept-based search engine, using FBS to perform functional patent analysis, and integrating creativity methods and all KBs above mentioned.

Figure 8 gives an example of the functional decomposition of the nutcracker into function (to open a nut) and linguistic behaviours (cut, lever, drill and crack). Engineering and multi-visional approaches are not shown for the sake of brevity.

Figure 8 Alternative situations (B) of opening walnuts, and their partial expansion (see online version for colours)



In Table 6 the expansion of 'to crack' is proposed. It has been created by combining synonyms with other lexical relationships.

Table 6 **'**Cracking' expansion with stemming made by KOM framework

crack	break or shatter or fracture or dash or puncture or disintegrate or snap or fragment or mash or bray or comminute or pestle or splinter or sliver or grate or smash or grind or crunch or mill or crumb or crush or open or shell or hull or skin or peel or decorticate or dehull or deshell or husk or dehusk

Table 7 shows an example dealing with 'thermal cracking by freezing'. The function 'freezing' can be expressed by synonyms of 'freezing' (to frost, to ice, etc.), properties and standard units (temperature, Celsius, BTU, etc.) and terms dealing with cold technologies (cryogenic liquid, refrigerator, etc.).

Table 7 An example of functional patent analysis. The underlined terms are those automatically selected by KOM as FB-PE targets for the function *to freeze*

EP1145653 – Furthermore, by the shock freezing mechanical stresses are induced in the outer shell which usually cause it to crack open automatically.
WO7900982 – The embrittlement of the grains by the liquid nitrogen causes individual pieces of crushed grain to fracture quite easily in the rollers 28, which makes it easier to dislodge the fibrous fractions from the grain pieces. The grain is kept at a very low temperature during the abrasive action to facilitate the fracture between the grain and the fibrous portions.
US4436757 – Disclosed are methods for decorticating and for hulling sunflower seeds with cryogenic liquid gases such as liquid nitrogen.

Content in patent text also shows how functions and behaviours are used to clarify the purpose or the way the PE is used (e.g. the PE freezing can be used for cracking, preserving, hardening the nut, etc.). For example in the EP1145653 freezing is used to crack open the shell, while in WO7900982 the low temperature facilitates the fracture.

In order to find all pertinent patents dealing with freezing from the given corpus (100% recall) one way is to build an exhaustive thesaurus (Salton, 1968). This implies to work with several terms, sometimes very general as 'cooling' or 'temperature' that introduce noise and they drastically reduce precision.

As the table 8 shows, searching with a specific thesaurus about 'freezing' (PE) leads to a high recall value (close to 100%), while searching by a semantic combination of PE with the function (F) and/or behaviour (B) limits the irrelevant results, improving the precision from 11% to 31%.

Table 8 Recall and precision comparison between PE search and FB-PE search

	Retrieved patents ∩ pertinent patents	*Recall*	*Precision*	*Retrieved patents*
Search based only on PE thesaurus	21/21	100%	11%	195
Search based on PE+F+B	20/21	95%	31%	65

In our case this means that patentees have to manually analyse 65 patents instead of 195 for finding the 21 pertinent patents.

Precision is low in both cases, because terms associated to freezing are partially the same as the case of heating. Unfortunately, in the given corpus, patents using heating for cracking nuts are numerous. For other cases (other PEs) of the same case study, precision score is much higher.

In Figure 9, the *cracking* action is decomposed in mechanical, acoustic, thermal, chemical, electric and electromagnetic cracking. Every effect is also sub classified with greater detail; for example in this specific case the mechanical breaking is further specified by delta pressure, gravity, compression, centrifugal, collision, explosion, friction and so on.

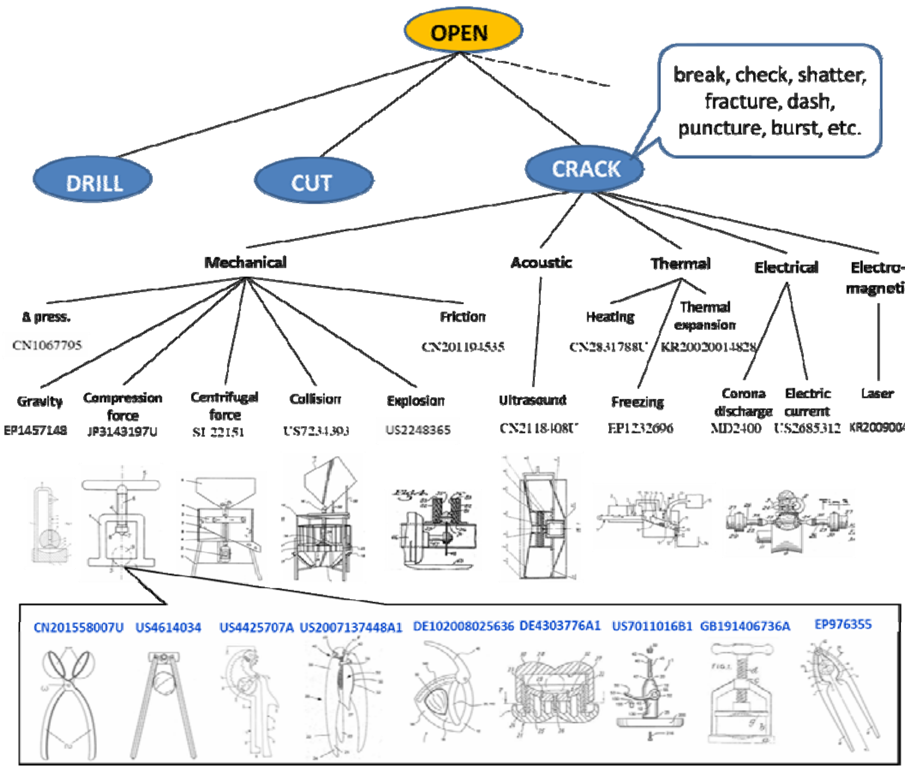
The result is a hierarchical tree that KOM uses for automatically creating the set of queries. In fact, each branch of the diagram provides a query for collecting patents containing 'different nutcracker structures using mechanical compression (PE) for cracking (B) a nut in order to open it (F)'. A list of variants for the compression force branch is given at the bottom of the figure.

In this case patent search has been carried out only inside the nutcracker domain. Setting KOM for searching outside the nutcracker domain new PEs can be discovered. KOM searches all PEs that are not yet exploited at the state-of-the-art of the nutcracker but already developed in other fields.

In fact, other interesting domains have been found dealing with devices for cracking shellfish or shell eggs or gas tank shell *(JP2005312315: using vibrations for eggs cracking)*, and machine for cracking tablets or stones, devices for grinding rice or grain.

The pool of patents, automatically classified by our functional classification can be used as support both for patent analysis, for problem solving activities and decision making.

Figure 9 Technology survey of a nutcracker realised by KOM framework (see online version for colours)



6 Conclusion

At present, working on a patent is often a pioneering activity. Indeed, despite the huge amount of information contained in several million technological inventions, only a very small amount of knowledge is actually retrievable.

Reflective reasoning about achievements and open problems dealing with function-based patent searches is proposed. The novelty of this work consists of enriching knowledge bases for IR or semantic searches with ontologies coming from the conceptual design area and problem solving, such as TRIZ.

In particular, an inedited ontology for IR applications, working on Function, Behaviour, Physical Effect, and Structure (FB-PE-S) has been conceived for searching non-clearly expressed functions.

This paper contributes to create a different way of searching functions using linguistic and creativity methods. Unlike a traditional search, the aim of a creative approach is to think/imagine in advance all the ways the goal can be reached instead of browsing patents to find what already exists.

The final result of this approach produces a greater overview of all possible functions related to the given product or technology. It can be useful for any kind of patent analysis from classical known item search to state-of-the-art assessment, but also to make personal knowledge database, technological transfer, forecasting activity and more in general for decision making.

Furthermore, a list of knowledge bases suitable for IR activities is given, such as the verb classification from NIST, patent classification indexes, and semantic ontologies that can be useful especially for expanding queries and creating exhaustive thesauri.

Finally, we propose a strategy using FB-PE considering a huge number of keywords dealing with a specific PE without losing precision.

A case study on nutcracker shows that with a search based on an exhaustive thesaurus, it is possible to obtain a high recall value (close to 100%) with a low precision, while searching with the proposed approach recall remains almost the same but precision grows up to three times more.

The FB-PE search provides a possible way to increase precision in functional searches without limiting the recall. Even when functions are not explicitly used, patent writers describe unawares inventions according to the logic FB-PE.

Nevertheless, this still needs further investigations in order to get better results.

Acknowledgements

The author sincerely thanks Tiziano Montecchi for his help and full support and encouragement, as well as for his contribution in developing the software.

References

Adams, S. (2010) 'The text, the full text and nothing but the text: part 1 – standards for creating textual information in patent documents and general search implications', *World Patent Information*, Vol. 32, No. 1, pp.22–29.

Altshuller, G.S. (1984) *Creativity as an Exact Science: the Theory of the Solution of Inventive Problems*, Gordon and Breach Science Publishers.

Anaya, V., Berio, G., Harzallah, M., Heymans, P., Matulevičius, R., Opdahl, A.L., Panetto, H. and Verdecho, M.J. (2010) 'The unified enterprise modelling language – overview and further work', *Integration and Information in Networked Enterprises*, Vol. 61, No. 2, pp.99–111.

Bernaras, A., Laresgoiti, I. and Corera, J. (1996) 'Building and reusing ontologies for electrical network applications', *Proceedings of the 12th European Conference on Artificial Intelligence ECAI'96, John Wiley and Sons, Chichester, UK, pp.298–302.*

Boiko, I.M. (1991) 'Semantic coding and solving intellectual problem', *TRIZ Journal*, Vol. 2, No. 1, pp.43–47.

Bonino, D., Ciaramella, A. and Corno, F. (2010) 'Review of the state-of-the-art in patent information and forthcoming evolutions in intelligent patent informatics', *World Patent Information*, Vol. 32, No. 1, pp.30–38.

Börner, K., Chen, C. and Boyack, K.W. (2003) 'Visualizing knowledge domains', *Annual Review of Information Science and Technology*, Vol. 37, No. 1, pp.179–255.

Cascini, G., Frate, L.D., Fantoni, G. and Montagna, F. (2011) 'Beyond the design perspective of Gero's FBS framework', *Design Computing and Cognition'10*, pp.77–96.

Cascini, G., Rotini, F. and Russo, D. (2011) 'Networks of trends: systematic definition of evolutionary scenarios', *Procedia Engineering*, Vol. 9, pp.355–367.

Cascini, G. and Russo, D. (2007) 'Computer-aided analysis of patents and search for TRIZ contradictions', *International Journal of Product Development*, Vol. 4, No. 1, pp.52–67.

Cascini, G., Russo, D. and Zini, M. (2007) 'Computer-aided patent analysis: finding invention peculiarities', *Trends in Computer Aided Innovation*, pp.167–78.

Cavallucci, D. and Khomenko, N. (2007) 'From TRIZ to OTSM-TRIZ: addressing complexity challenges in inventive design', *International Journal of Product Development*, Vol. 4, No. 1, pp.4–21.

Chakrabarti, A., Sarkar, P., Leelavathamma, B. and Nataraju, B.S. (2005) 'A functional representation for aiding biomimetic and artificial inspiration of new ideas', *AI EDAM: Artificial Intelligence for Engineering Design, Analysis, and Manufactoring*, Vol. 19, No. 2, pp.113–132.

Gero, J.S. (1990) 'Design prototypes a knowledge representation schema for design', *AI Magazine*, Vol. 11, No. 4, pp.26–36.

Gero, J.S. and Kannengiesser, U. (2004) 'The situated function-behaviour-structure framework', *Design Studies*, Vol. 25, No. 4, pp.373–391.

Gero, J.S. and Rosenman, M.A. (1990) 'A conceptual framework for knowledge based design research at Sydney University's Design Computing Unit', *Artificial Intelligence in Engineering*, Vol. 5, No. 2, pp.65–77.

Gruber, T.R. (1993) 'A translation approach to portable ontology specifications', *Knowledge Acquisition*, Vol. 5, No. 2, pp.199–220.

Hirtz, J., Stone, R.B., Szykman, S., McAdams, D.A. and Wood, K.L. (2002) 'A functional basis for engineering deisgn: reconciling and evolving previous efforts', *Research in Engineering Design – Theory, Applications, and Concurrent Engineering*, Vol. 13, No. 2, pp.65–82.

Koch, S. and Bosch, H. (2011) 'From static textual display of patents to graphical interactions', *Current Challenges in Patent Information Retrieval*, Vol. 29, pp.217–235.

Lee, J.H., Renear, A. and Smith, L.C. (2006) 'Known-item search: variations on a concept', *Proceedings of the American Society for Information Science and Technology*, Vol. 43, No. 1, pp.1–17.

Lee, S., Yoon, B. and Park, Y. (2009) 'An approach to discovering new technology opportunities: keyword-based patent map approach', *Technovation*, Vol. 29, Nos. 6/7, pp.481–497.

Li, Y.R., Wang, L.H. and Hong, C.F. (2009) 'Extracting the significant-rare keywords for patent analysis', *Expert Systems with Applications*, Vol. 36, No. 3, pp.5200–5204.

Litvin, S. (2004) *New TRIZ-Based Tool – Function-Oriented Search*, ETRIA Conference TRIZ Future, TFC, Florence, Italy.

Lyon, M. (1999) 'Language related problems in the IPC and search systems using natural language', *World Patent Information*, Vol. 21, No. 2, pp.89–95.

Miller, G.A., Beckwith, R., Fellbaum, C., Gross, D. and Miller, K.J. (1990) 'Introduction to wordnet: an on-line lexical database', *International Journal of Lexicography*, Vol. 3, No. 4, pp.235–244.

Neches, R., Fikes, R., Finin, T., Gruber, T., Patil, R., Senator, T. and Swartout, W.R. (1991) 'Enabling technology for knowledge sharing', *AI magazine*, Vol. 12, No. 3, p.36.

Norton, M. (2000) *Introductory Concepts in Information Science*, Information Today, Inc.

Pahl, G. and Beitz, W. (1977) *Engineering Design: A Systematic Approach*, Springer Verlag.

Russo, D. and Montecchi, T. (2011a) 'A function-behaviour oriented search for patent digging', *ASME 2011 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference IDETC/CIE*, Washington, DC, USA.

Russo, D. and Montecchi, T. (2011b) 'Creativity techniques for a computer aided inventing system', *Proceedings of the 18th International Conference on Engineering Design (ICED11)*, Lyngby/Copenhagen, Denmark.

Russo, D., Regazzoni, D. and Montecchi, T. (2011) 'Methodological enhancements for concept exploration in product design', *International Journal of Product Development*, Vol. 15, No. 1, pp.26–53.

Russo, D., Montecchi, T. and Ying, L. (2012) 'Functional based search for patent technology transfer', *Proceedings of the ASME 2012 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference*, Chicago, Illinois, USA.

Salton, G. (1968) *Automatic Information Organization and Retrieval*, McGraw Hill.

Shimomura, Y., Yoshioka, M., Takeda, H., Umeda, Y. and Tomiyama, T. (1998) 'Representation of design object based on the functional evolution process model', *Journal of Mechanical Design*, Vol. 120, No. 2, pp.221–229.

Stefanov, V. and Tait, J.I. (2011) 'An introduction to contemporary search technology', *Current Challenges in Patent Information Retrieval*, pp.45–65.

Swartout, W. and Tate, A. (1999) 'Ontologies', *Intelligent Systems and their Applications, IEEE*, Vol. 14, No. 1, pp.18–19.

Tsourikov, V.M. (1993) 'Inventive machine: second generation', *AI & Society*, Vol. 7, No. 1, pp.62–77.

Umeda, Y. and Tomiyama, T. (1995) 'FBS modeling: modeling scheme of function for conceptual design', *Workshop on Qualitative Reasoning about Physical Systems*, pp.271–278.

WIPO Economics and Statistics Division (2011) *World Intellectual Property Indicators*, WIPO Publication, Vol. 941, p.214.

Zanni-Merk, C., Cavallucci, D. and Rousselot, F. (2011) 'Use of formal ontologies as a foundation for inventive design studies', *Computers in Industry*, Vol. 62, No. 3, pp.323–336.

Notes

1 CPC Project Website http://www.cooperativepatentclassification.org/index.html

2 PATExpert Project Website www.patexpert.org