Semantic TRIZ

Mikhail Verbitsky, Ph.D., Sc.D.

Director, Software Applications and Training Worldwide
Invention Machine Corporation
133 Portland Street, Boston, MA 02114
e-mail: verb@invention-machine.com

1. Introduction: Postulates of TRIZ and their implementation in software.

The Theory of Inventive Problem Solving, better known in the engineering world by its Russian abbreviation TRIZ (e.g., [1]) has been widely acknowledged as an effective tool to solve technical problems, especially at the conceptual design stage. Its attractiveness is based on several appealing observations. First, it was noticed, that a significant number of engineering problems can be described in a contradiction template. These contradictions are continuously recurring across different industries, and these same contradictions have frequently been solved using repeating approaches. It has also been discovered that the patterns of technical system evolution are repeated across different industries as well: specifically, systems are being developed in the directions of increased ideality, harmonization, controllability, and degree of freedom. And finally, it has been broadly recognized that the best innovative solutions seem to adopt scientific effects from diverse fields of science and engineering.

Extrapolation of these observations to a specific problem created a number of tools used by TRIZ practitioners. The Matrix of Contradictions is probably the most popular one. When an engineer formulates his problem as a contradiction, the Contradiction Matrix helps him to identify a representative contradiction and classic approaches to solve this contradiction, the so-called Inventive Principles. This process is based on a simple assumption, that if these Principles demonstrated their effectiveness in similar circumstances in the past (for the same contradiction), they

might be helpful for the current situation as well. In the software media [2], the Contradiction Matrix is implemented as is illustrated in Figure 1.
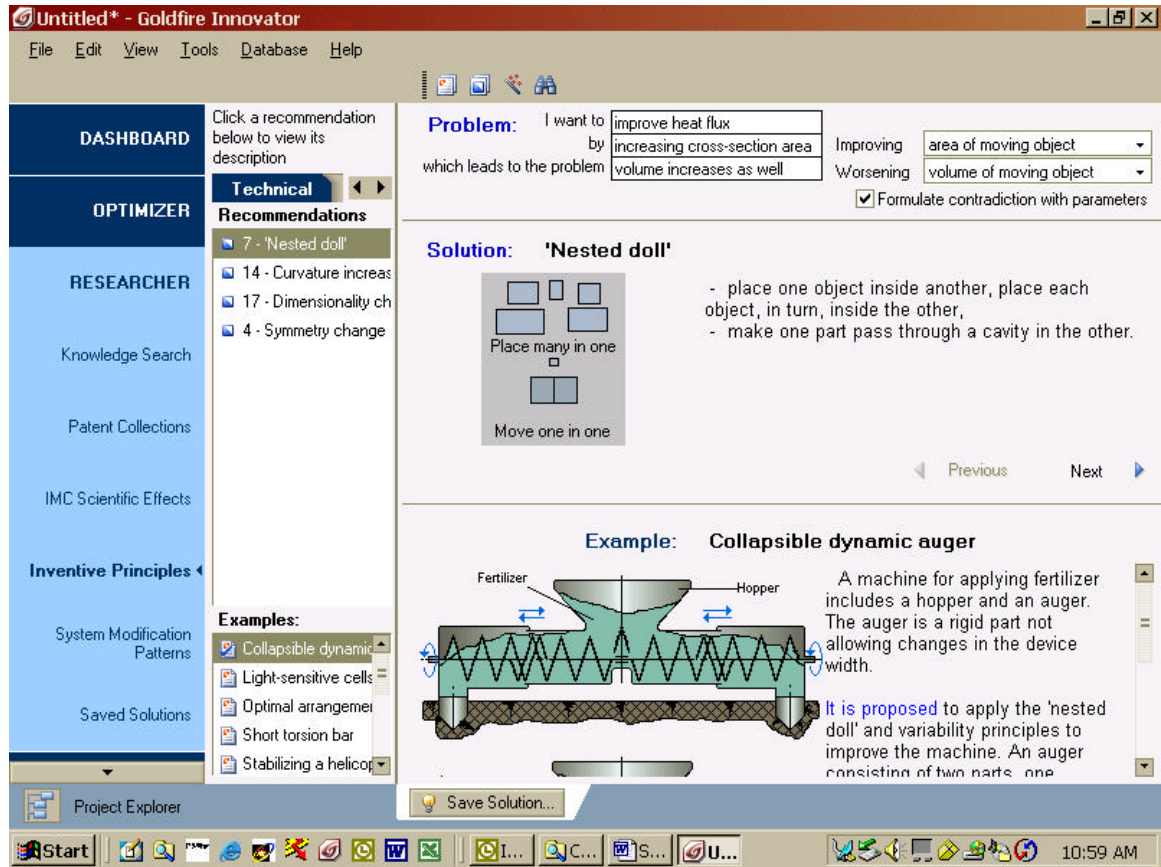

Figure 1: Contradiction Matrix in Invention Machine's Goldfire Innovator™ [2].

Figure 1 shows the situation, described in the following statement: 'I want to improve the thermodynamic properties of my design by increasing its cross-section area, but there are undesirable consequences of the area increase – the volume increases as well'. The Matrix of Contradictions helps to translate this statement into a contradiction template: improving aspect – area of moving object, worsening aspect – volume of moving object; and suggests several Inventive Principles which might be helpful for this problem because they had demonstrated their effectiveness in similar situations in the past.

Similarly, the distinctive trends of technology evolution have been incorporated into a comprehensive Prediction Tree [2]. There are different techniques to solve engineering problems with its help [3] and one of them is illustrated in Figure 2.
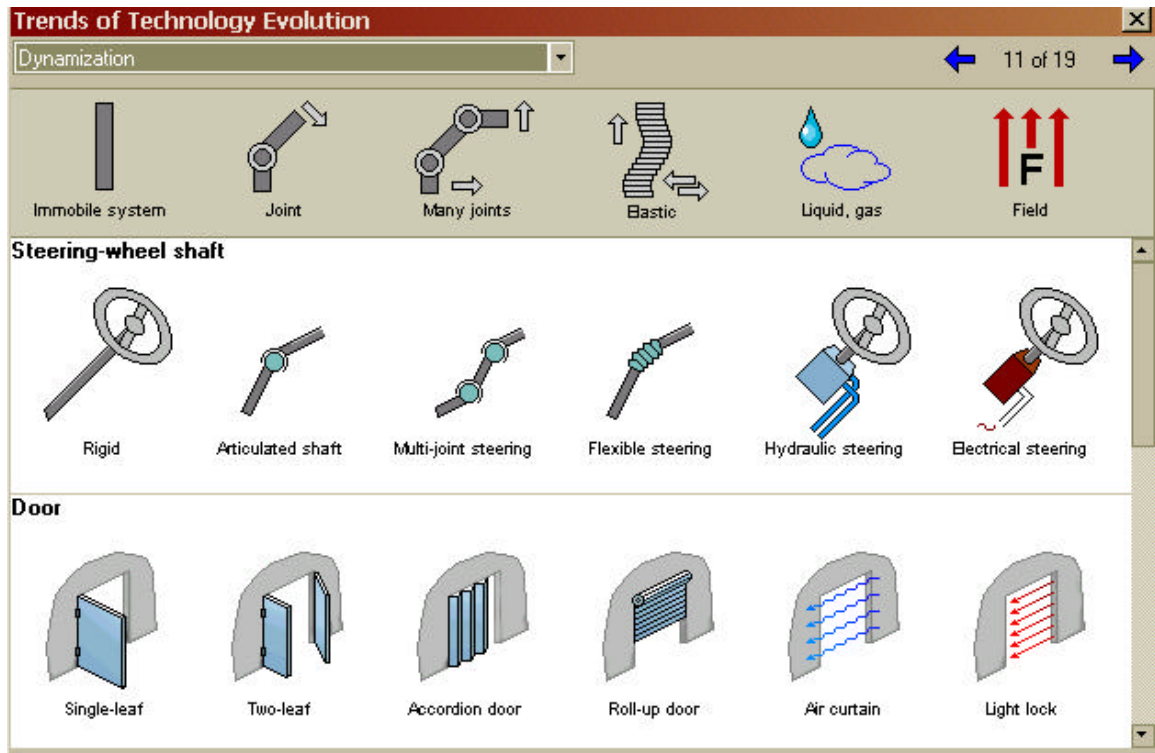
Figure 2: Dynamization trend of engineering systems evolution as described by
Invention Machine's Goldfire Innovator™ [2]

Figure 2 illustrates one of technology's evolution trends, the trend of Increased Dynamization: engineering systems generally evolve from rigid immobile systems in the direction of increased dynamization, i.e., they employ more joints, increase the component's elasticity, replace solid materials with liquid or gas, *et al*. Although the solutions presented by this method are abstracted templates of what an engineer ultimately implements, their universality serves as a stimulus for generating innovative problem-solving ideas, and leads to the conception of new system features that can improve its performance. The implicit extrapolation assumption behind this method is similar to that of Inventive Principles: other engineering systems experienced this trend, therefore it is not unlikely that the system we are currently working with may experience the same trend.

The TRIZ idea to search for innovative solutions in different fields of science and engineering has inspired the creation of the Scientific Effects knowledge base [2], which is currently the most comprehensive library of its kind. The organization of its 9,000 physical effects is ideally suited for problem solvers since it uses a function-oriented taxonomy. Engineers can search this knowledge base by specifying the function they want to perform (rather than topics), and retrieve the scientific phenomena that can be employed to accomplish this task. This process is illustrated by Figure 3.
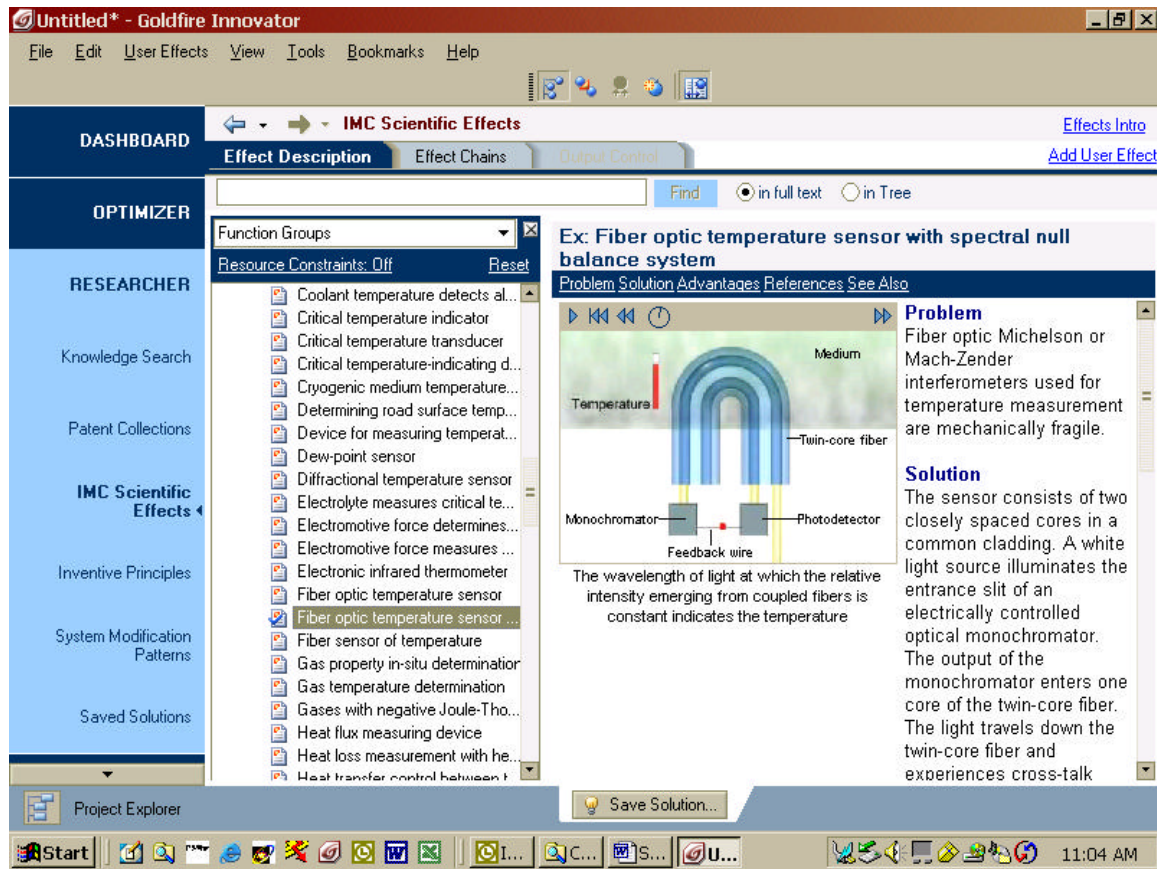
3

Figure 3: Effects database of Invention Machine's Goldfire Innovator™ [2].

It demonstrates the situation described in the following statement: 'I want to be able to measure the temperature of the substrate'. Instead of making a decision as to what branch of science should be approached to solve this problem, an engineer can simply open a folder with a name that can be associated with the problem formulated above, i.e., 'measure thermal parameters', and review a variety of scientific effects which can measure temperature.

2. A Few questions…

In science it is always a challenge to derive a general law from specific empirical observations. Those who take this challenge should be able to answer a few questions regarding the validity of the extrapolation. For TRIZ-based tools these questions are quite obvious, for example:

(i)   Is the Matrix of Contradictions statistically stable to a number of patents analyzed? Or simply speaking: If we continue to perform analysis of patents similar to that which TRIZ founder Genrich Altshuller performed, and eventually analyze all existing patents, will the Contradiction Matrix change? If it will, then how will it change?

4

(ii)   Are discovered trends of technology evolution statistically stable to a number of patents analyzed? Do they cover all existing trends in the current world of technology?

(iii)   How does one cross a chasm from a general recommendation to a specific innovative idea?

To answer these questions, we need software solutions capable of reading millions of documents (e.g., patents) and being 'intelligent' enough to 'understand' them. The solution providing such means is based on semantic indexing technology; and its application to the problem-solving practice launches a novel approach to the innovation process which eventually may become the new engineering discipline - Semantic TRIZ™.

3. Basics of Semantic Indexing Technology

Semantic Indexing Technology is based on mathematical linguistics. Linguistic analysis of the natural language text [4] is currently performed on four major levels which could be generally defined as sentence and word recognition, lexical analysis, syntactic analysis, and *semantic* analysis. The mission of the first level is obvious. Lexical analysis involves reading the input sentences, extracting individual words, and retrieving the possible word types from the databases (dictionaries). Lexical analysis is enhanced by a hidden Markov chains model, which provides the probability distribution for word type sequences and determines the most likely sequence of word types in a sentence. Syntactic analysis employs phrase-structure grammars, identifies the syntactic structure of the text and consequently determines the exact word type. Semantic analysis identifies the meaning of the text by extracting from a sentence its semantic items such as subject, action, and object.

Applying this analysis for the following sentence,

Electrolytic dissociation can be successfully used to measure air humidity,

the computer will determine that in this sentence

the Subject is *electrolytic dissociation*
the Action is *measure*, and
the Object is *air humidity*

These semantic items are of great importance because they contain information about what questions can be asked and what answers can be served in response. For example, if someone asks the question: 'How can I measure humidity?', the person who is asking this question in fact defines that in the possible response Action should be *'measure'* and Object should be *'humidity'*. What is unknown to him is the Subject (*what* measures humidity?). If we performed semantic analysis of all documents

available, including one with the above sentence, which means that we would extract all Subjects, Actions, and Objects, then, in response to this question, we would provide only those Subject-Action-and-Object combinations which have the Action and Object like in the questions, and this will form an exact answer to the question *(e.g., electrolytic dissociation – measure - air humidity).*

This simple example in fact defines the architecture of the semantic indexing technology which will support Semantic TRIZ™ research:

(i)   It should have a natural language interface, which is able to extract semantic entities from the question;

(ii)   It should have a searchable database of semantic entities extracted from (generally speaking) the entire universe of original documents. The process of creating such a database is called semantic processing and is quite similar to a process of educating a human: we are 'asking' the computer to read documents, understand them by means of extracting semantic items, and to memorize this understanding;

(iii)   Matching semantic items from a question with semantic items in the database provides exact answers to a question

This architecture was implemented in the Goldfire Innovator™ platform [2], where a searchable database of semantic entities was created by applying linguistic analysis to the entire worldwide collection of patents. In the following chapters, we will demonstrate that this platform is very capable of supporting the TRIZ innovation process.

4.   Semantically indexed database as answering engine platform

The most straightforward application of Goldfire Innovator™ is to ask it direct natural language questions. Figure 4 below illustrates this process.
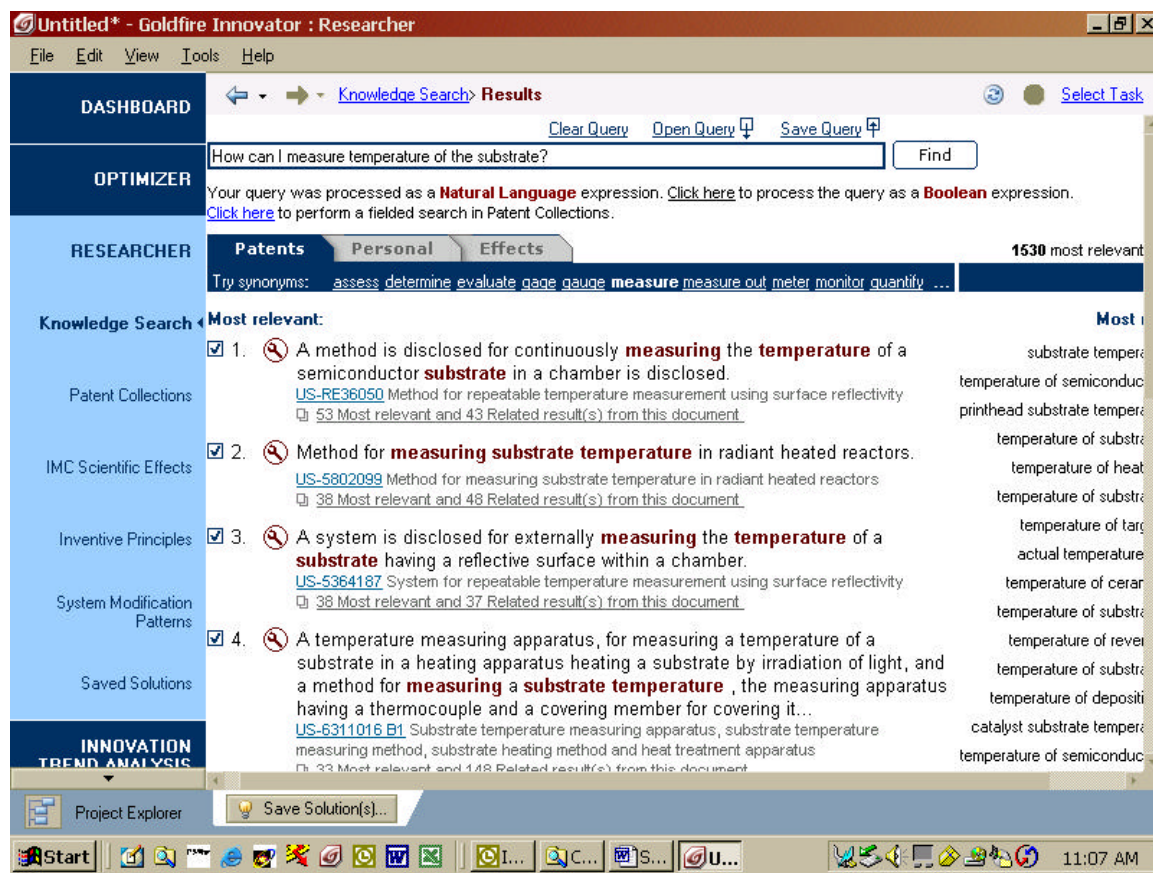
Figure 4: Goldfire Innovator™ as 'Effects-on-Demand' answering engine platform.

When the question 'How can I measure temperature of the substrate?' is asked, the system performs analysis of the question, determines what semantic items it contains (action – measure, object – temperature of the substrate), and matches it against hundreds of millions of semantic items, extracted from the worldwide collection of patents. The results of the match represent the exact answer to the question and are shown in Figure 4. While the results displayed in Figure 3 demonstrated what physics in general is available to measure temperature, the results of Figure 4 are very specific, indicating how the *substrate* temperature is measured. This is why we can call this application 'customized Effects' or 'Effects-on-Demand'.

5. Semantically indexed database as Matrix of Contradictions

The problem of Contradiction Matrix sensitivity to a number of patents analyzed was approached by a number of researchers (e.g., review in [5]). Since the questions in the contradiction template can also be addressed by a semantically indexed database, then, speaking pragmatically, Goldfire Innovator™ can transfer this sensitivity problem into historical perspective. In

Figure 1, we illustrated how Altshuller's Contradiction Matrix handles the situation, described by the following statement: 'I want to improve thermodynamic properties of my design by increasing its cross-section area, but there are undesirable consequences of the area increase – volume increases as well'. We can now ask this question as a natural language question: 'How can I increase area, but decrease volume?' and address it against the worldwide collection of patents, or any collection of internal corporate knowledge.
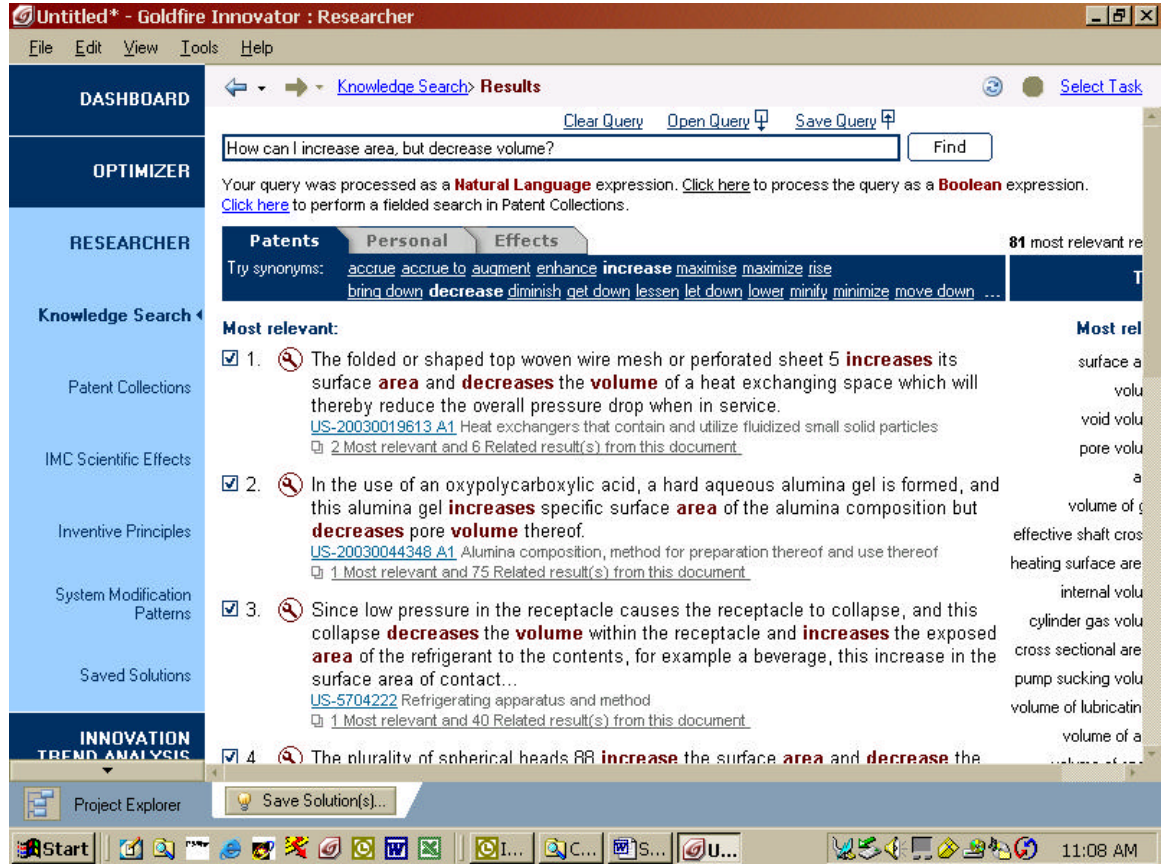

Figure 5: Goldfire Innovator™ as matrix of contradictions.

The system response is illustrated in Figure 5; it presents very explicit information about how this exact contradiction has been previously solved.

6. Trends of evolution research with semantically indexed database

Goldfire Innovator™ enables researchers to investigate trends of evolution for any industry, any technology, any product, design, material, or, generally speaking, it can build a time dependence of answers to any natural language question. This process is presented in Figure 6. It shows that, for example, the question 'How can we detect a gas leak?' can be asked in a specific time domain. Asking this question recurrently, we will see how answers to this question evolve in time. Answer-to-question (or solution-to-problem) time

8

dependence is nothing else but a specific technology evolution trend. Results for the question 'How can we detect a gas leak?' asked in 5-year intervals are shown in Figure 7.
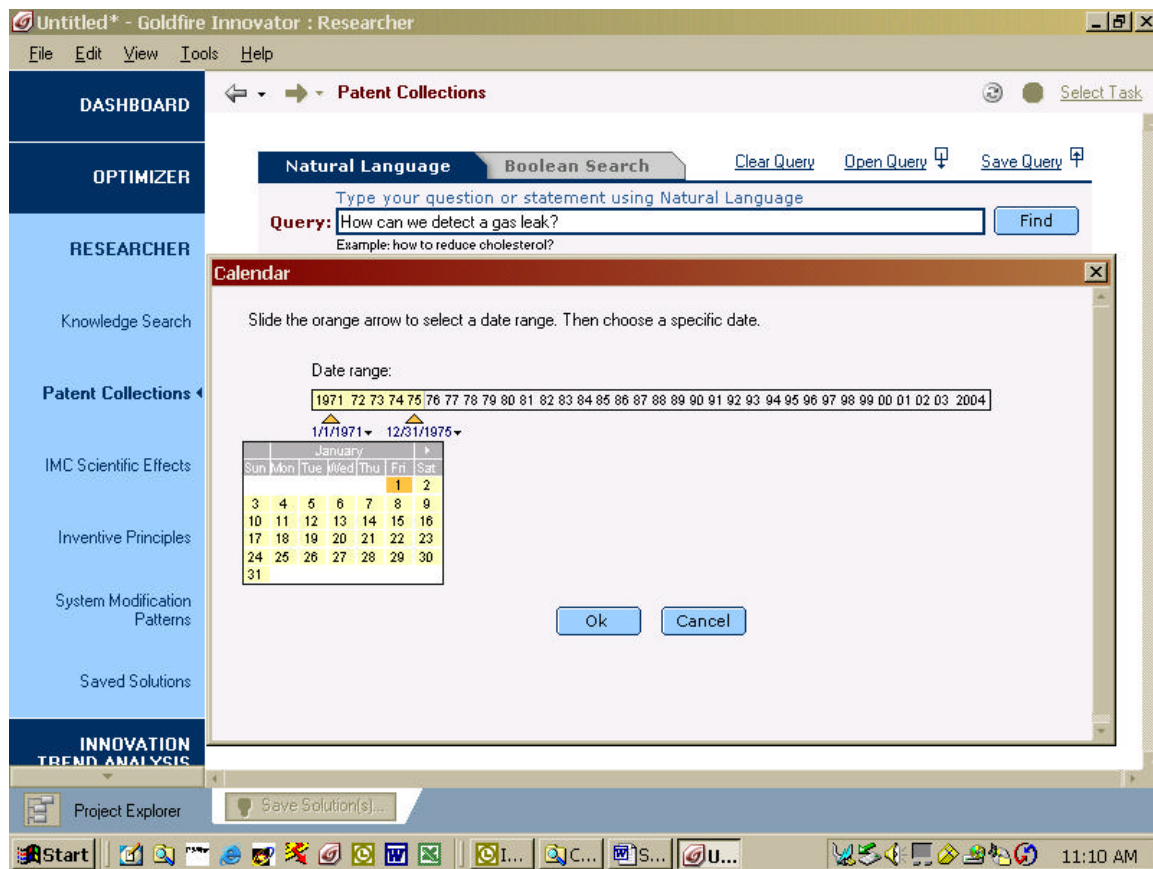

Figure 6: Asking questions in a specific time domain.

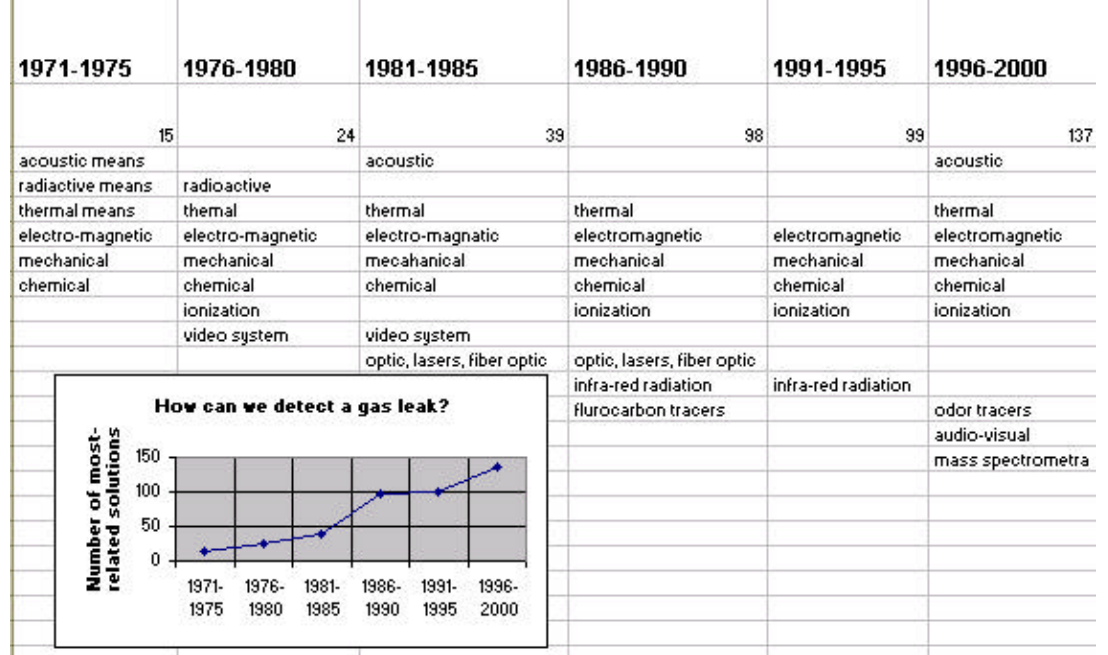| 1971-1975	1976-1980	1981-1985	1986-1990	1991-1995	1996-2000
15	24	39	98	99	137
acoustic means		acoustic			acoustic
radiactive means	radioactive				
thermal means	thermal	thermal	thermal		thermal
electro-magnetic	electro-magnetic	electro-magnatic	electromagnetic	electromagnetic	electromagnetic
mechanical	mechanical	mecahanical	mechanical	mechanical	mechanical
chemical	chemical	chemical	chemical	chemical	chemical
	ionization		ionization	ionization	ionization
	video system	video system			
		optic, lasers, fiber optic	optic, lasers, fiber optic		
			infra-red radiation	infra-red radiation	
			flurocarbon tracers		odor tracers
					audio-visual
					mass spectrometra

Figure 7:   Time dependence of Goldfire Innovator™ generated answers: Trend of gas-leak-detection means evolution.

It can be clearly determined that gas-leak-detection systems evolved from acoustic, thermal, and mechanical designs to optical, audio-visual, and spectrographic means.

In TRIZ, it is widely believed that the assessment of the system development stage ('infancy' – fast development – maturity S-curves) can be determined when time dependence of the quantity of patents is compared against their innovation level. While number-of-patents versus time functions can be easily calculated, *automatic* innovation level evaluation is very challenging. Five levels of innovation, suggested by Altshuller, are logical, but do not provide the exact criteria for practical usage and therefore are extremely difficult to quantify. In technical intelligence (e.g., [6, 7]), patent forward and backward citations are commonly used to evaluate intellectual property competitive advantage. We propose to employ patent forward citation as an indicator (certainly, not the unique one) of its level of innovation, speculating that the break-through inventions will be extensively referenced by the followers.

The following statistical model has been adopted. Suppose we have a sequence of related (i.e., same technology) patents $P_1; P_2; ... P_i; P_{i+1}; ... P_n$, where index i increases in time. Then the probability that a single reference in P_{i+1} is directed to the patent P_i can be estimated as

(1) $c = \dfrac{1}{i}$

The probability that a single reference in P_{i+1} is not directed to the patent P_i can be estimated as

10

(2) $f = 1 - \dfrac{1}{i}$

The probability that patent P_i is not referenced at least once by patent P_{i+1}, if patent P_{i+1} has m backward references in total, can be estimated as

(3) $F = \left(1 - \dfrac{1}{i}\right)^{m}$

The number of patents K, which being combined would definitely provide at least one reference to the patent P_i, can be estimated from the equation (4):

(4) $\displaystyle\prod_{k=0}^{k=K} F_k = \left(1 - \dfrac{1}{i}\right)^{m_{i+1}} \left(1 - \dfrac{1}{i+1}\right)^{m_{i+2}} \cdots \left(1 - \dfrac{1}{i+k}\right)^{m_{i+1+k}} \cdots \left(1 - \dfrac{1}{i+K}\right)^{m_{i+1+K}} = 0$;

Then, the normally expected number of forward references received by patent P_i can be estimated as:

(5) $C \approx \dfrac{n-i}{1+K}$

We will consider a patent as being an 'outstanding' one if the number of its actual forward references exceeds C.

As an example, on Figure 8 we show results of the analysis conducted with Goldfire Innovator™ for 344 patents mentioning 'fiber optic gyroscope' in their titles or abstracts. Interestingly, we can see that the number of patents evolves in a very 'classic' manner clearly exhibiting an 'infancy' – fast development – maturity cycle. As it could be also anticipated (newer patents have less chance to be referenced), the number of expected forward references decreases in time. Results also demonstrate that 'outstanding' patents are not necessarily concentrated at the beginning of the process as one could expect. This conclusion should be regarded as tentative, however, since the correlation of criteria (5) with the 'innovation level' is not linear indeed.
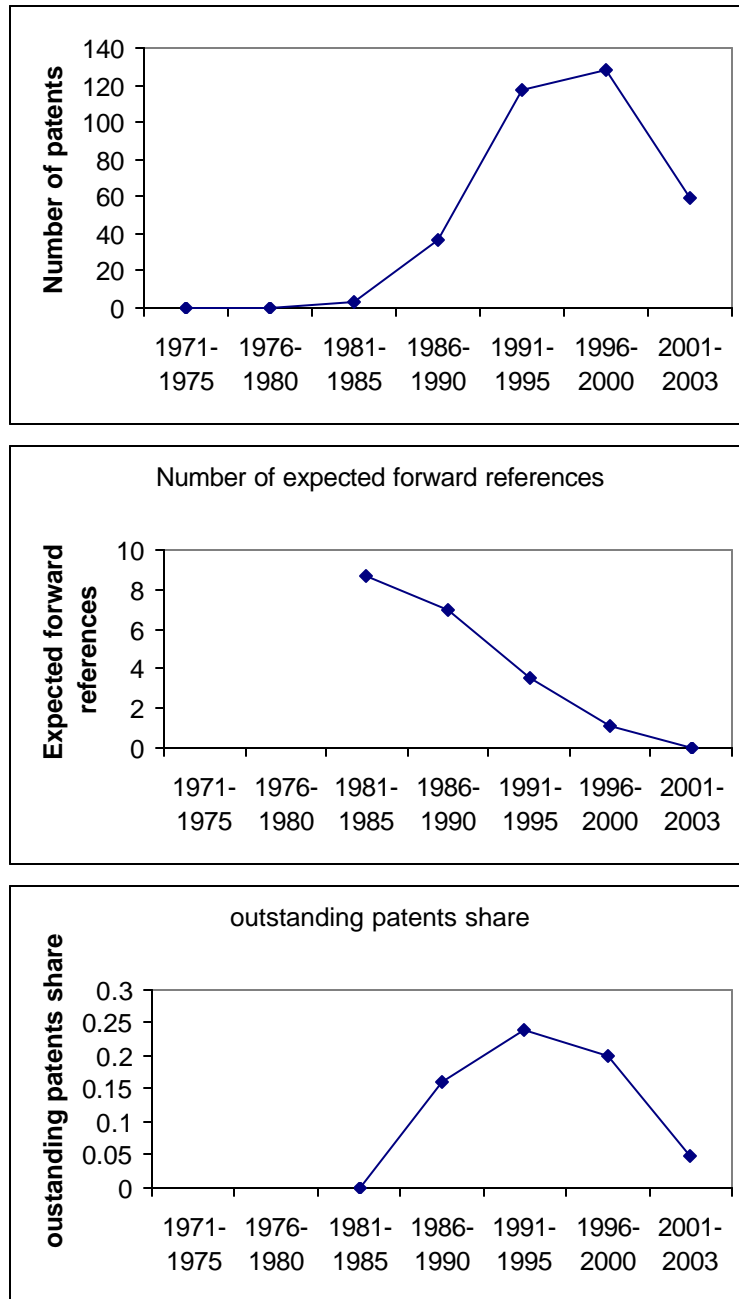
Figure 8: Goldfire Innovator™ generated evolution of fiber optic gyroscope technology.

7. Conclusions

Half a century ago, when TRIZ methodology was originated, the ability to automatically study millions of documents was absolutely inconceivable. Therefore traditional TRIZ repeated steps of many empirically derived methods: from extensive (though limited) observations to extrapolation and generalization.

Recent advances of computational linguistics allowed us to combine the classic TRIZ approach to the innovative problem solving process with the benefits of Semantic Indexing Technology. We call this enhanced process Semantic TRIZ™. By its nature, since every question is addressed to the entire worldwide collection of patents, Semantic TRIZ™ is very specific and therefore it can support many traditional TRIZ tools:

(i)   If questions are formulated directly, Semantic TRIZ™ works as a customization of a scientific effects data base;

(ii)   If questions are formulated as a contradiction, Semantic TRIZ™ works as a huge (currently $10^7 \text{x} 10^7$) matrix of contradictions providing specific answers on how this contradiction has been solved

(iii)   If questions are formulated relative to a specific time domain, Semantic TRIZ™ generates exact trends of technology evolution

References

1. Genrich Altshuller. *The Innovation Algorithm: TRIZ, Systematic Innovation and Technical Creativity.* 1999. Technical Innovation Center, 312 pp

2. Ena Arel. *Goldfire Innovator™. Volume I: Optimizer User Guide. Volume II: Patents and Innovation Trend Analysis User Guide.* 2004. Invention Machine Corporation, Boston, MA, 725 + 263 pp.

3. Ena Arel, Mikhail Verbitsky, Igor Devoino, Sergei Ikovenko. *TechOptimizer Fundamentals*, 2002, Invention Machine Corporation, Boston, MA, 134 pp.

4. *Precision in Knowledge Retrieval*. 2002. Invention Machine Corporation, Boston, MA, 5 pp. (http://www.invention-machine.com/sitemap.cfm - white paper)

5. Ellen Domb and Michael Slocum. "Introducing Matrix 2003 - New research and a new format for the popular TRIZ tool", *TRIZ Journal*, 2003, v.8

6. Mogee, M.E., "Patent analysis for strategic advantage: Using International Patent Records," Competitive Intelligence Review, v. 5 (Spring 1994): 27-35.

7. Mogee, M.E., "Patents and technology intelligence," in Technical Intelligence for Business: Keeping Abreast of Science & Technology, ed. by W.B. Ashton and R.A. Klavans, Washington, D.C.: Battelle Press, 1997.