6-Sigma Patent Data Quality and Applied Latent Semantic Analysis Technology in Patent Research

Andy Gibbs, President and CEO, PatentCafe.com, Inc.
S. Lata Setty, M.Sc., Esq., Vice President of Business Development, PatentCafe.com, Inc.

PURPOSE:

This paper generally discusses patents as contained in the world's various patent data repositories, the importance of patents in today's business environments, and the increased reliance on access to comprehensive patent collections upon with to conduct research. It also discusses the economic importance of obtaining reliable results from research conducted on these data, the consequences that incomplete patent data, or the inability to confidently mine patent data, can have on today's technology business leaders.

In an effort to improve product and process quality throughout their organizations, business leaders have long used a statistical approach to quality improvement, known as 6-Sigma. This paper will address the traditional, as well as state of the art standards applied to patent data quality, and the impact the transition to the targeted 6-Sigma quality of PatentCafe's ICO-GPS™ patent data can have on business processes and shareholder value.

Finally, as a component to patent data mining, whether performed on 6-Sigma patent data or traditional quality data, this paper will discuss the increase in relevance, confidence, and statistical reliability of search results sets as associated with the applied Latent Semantic Analysis natural language search technology to patent data research. Case studies of PatentCafe's ICO-GPS patent search using its LSA search technology will be presented.

PROBLEM:

Traditional patent search tools, namely those based on Boolean search technology, applied to poor quality patent data, produce poor results. Even newer technology tools applied to poor quality patent data, and older technology search tools applied to very high quality patent data, still produce poor results. Business and legal decisions are becoming increasingly important, while at the same time, the consequences of decisions made with poor or incomplete data carry an unacceptably high risk. In some cases, the consequences are catastrophic, resulting in patent infringement awards approaching the billions of dollars, or companies being forced to terminate operations.

SOLUTION:

The use of PatentCafe's ICO Global Patent Search, with its Latent Semantic Analysis *plus* traditional Boolean search technology, along with its WIPO standardized, 6-Sigma targeted patent data, can reduce risk, increase profitability and reliability, speed patent research, and improve the relevancy and quality of the typical patent search results set. Advanced research techniques not available using traditional Boolean technology can also be employed to stretch the capabilities of patent searching into complex text such as chemical compounds, occurrences of multi-lingual characters, and mathematical formulae.

BACKGROUND:

It has been estimated that patent documents contain 80% of the world's accumulated technical knowledge.[1] It therefore is understandable that the more than 35 million patent documents that reside in repositories around the world should be heavily relied upon as a primary research data source. This incredible knowledgebase is proving itself a highly important asset in the business environment that increasingly relies upon reliable mining of patent data to enhance competitive positioning and stakeholder value.

In fact, intangible assets of companies competing in technology-based industry segments now contribute more to shareholder value than traditional tangible assets – in many cases, 85% or more[2]. Currently, less than half (and possibly as little as one third or less) of the market value of securities can be accounted for by hard assets … The rest of the value must, necessarily, be coming from organizational and human capital, ideas and information, patents … "[3]

Therefore, while patent research has traditionally resided in as a "legal support" process reporting within the legal or R&D structure with an organization, patent data, when *converted to business information*, now pervades every cross functional organization within an enterprise. Increasingly, legal, financial and R&D investment looks to reliable patent search results as another metric upon which to gauge the investment risks.

Yet, while the demands on more, higher quality patent-based information increase, patent researchers have had few options than accessing traditionally available patent databases and typical Boolean-based search tools. "[T]he typical database search that underlies most critical decisions misses more than 50% of the relevant art. While these approaches may be suitable for simple matters, reliance on this level of due diligence is wholly inappropriate for critical decisions"[4]

Anthony Trippe, Senior Staff Investigator at Vertex Pharmaceuticals, recently introduced ***patinformatics***; the science of analyzing patent information to discover relationships and trends represented by large collections of patent date. Patinformatics encompasses all forms of analyzing patent information, including … patent intelligence, patent mapping, patent citation analysis … The key underlying property in each of these diverse [Patinformatics] areas is the analysis step.[5]

The authors believe that the software tools available to map and visualize patent data increase the ability of patent, as well as non-patent professionals to identify and understand trends, groupings, and time line based patent activity. However, these visualization results can only be relied upon as macroscopic in nature, since visualization of large patent data sets generally mask the occasional missing patent, patents that are missing vital data, or patents containing data that has been incorrectly catalogued. The key underlying property of reliable patinformatics more accurately therefore, is the quality of the patent data set upon which the charts and maps, and subsequently critical business decisions, are based.

Not only is reliable, high quality patent data a necessary component to any analytics tool, it's also a key component to supporting the whole of innovation system within an enterprise. Innovation must be novel, of high value, affordable, and result in products that capture a market share that can be protected through the enforcement of patents. A crucial determinant of the *success of an innovation system is the quality, variety and availability of knowledge to recombine*. The more, and more varied, the knowledge at the fingertips of a would-be innovator, the greater his scope for (technically and economically) successful innovation.[6] *(emphasis added)*

An average of 10 infringement suits are filed every business day. Patents are coming under increased competitive challenge, often resulting in invalidity, injunctive relief or extraordinarily costly infringement awards. The confidence of patents to adequately protect a market is declining.

Patent applications and R&D decisions can no longer afford to rely on outmoded search technology or poor quality patent data if companies expect to achieve sustainable competitive advantage.

PATENT DATA:

About 35 million patents reside in various patent databases and repositories worldwide. Patent data exists in a variety of formats, including paper copies, Microfiche, and in digital form (computer tapes, hard disks, DVDs and other). Combing through the immense volume of patent data alone would create a significant challenge to any professional researcher. The fact that the data resides in many disparate repositories, and that the data is not maintained against any single standardized form and format, virtually guarantees incomplete or inaccurate patent search results – even by the most skilled researchers.

Improving patent data quality is not only a politically motivated goal, the requirement of higher patent quality is a practical one. In order to compete effectively in a global economy, the companies accessing best patent data solutions will excel. Fast and accurate access to complete and comprehensive patent data is paramount.

The patent issuing authorities have heard the message, loud and clear. But, history has taught us that certain quality requirements in commerce can often differ from those of government agencies. This is not a reflection of unresponsiveness by any patent authority - it simply states the obvious: aggressive business in commerce drives the development of products and processes that provide competitive advantages.

The urgency of advances in commerce are not necessarily shared by public agencies. The urgency in the intellectual property industry is the rise in competition-driven infringement litigation, patent re-examinations, patent invalidity awards and more, these legal and business tactics often rooted in poor legal strategy years ago, but are increasingly rooted in decisions made on poor quality patent data.

Nevertheless, Herculean strides are being undertaken at many patent offices worldwide to increase the standardization and quality of patent data (i.e.: US Patent & Trademark Office, as well as the European and Japanese Patent Offices, and to a lesser extent, other patent offices of developed countries). Each issuing authority typically implements a high level process involving a few sequential steps:

(i) convert all patent data to a single digital format (often involves OCR scanning of decades worth of paper patents – the "back file"),
(ii) normalize the database containing digital data (in both a user and technical sense),
(iii) implement a process that creates future digital files conforming to the data format.

Even this process is not without problems. Use of older OCR technology usually resulted in incredibly poor paper to digital conversion quality – on the order of 87 – 93% accuracy. Even newer OCR technology, while capable of higher quality data conversion, is still quality- impeded by the inaccurate or incomplete information contained on the paper files.

Tough economic conditions worldwide have also frustrated patent offices from actually completing well thought out multi-year programs. Each time the initiatives go through the start/stop process, files are partially converted, or new standards come into play during the re-start mode, causing even more "formats" to evolve.

Because of the disparate nature of the data and the repositories in which the data reside, researchers have routinely been frustrated with the different results sets that each of the databases will produce, even though the identical search query was used. When identical search results cannot be achieved from performing the identical search upon different databases that claim to contain the same data sets from the same issuing authorities, the researcher comes to the only logical conclusions:

(a) the data are not identical, and/or
(b) the database search technology is not equivalent.

Obtaining different patent search results from various, but purportedly similar, patent databases is a long established expectation by professional patent researchers. As a result, patent researchers routinely employ the use of multiple databases and their various research tools to create search results sets that can then be compared "off line". Using a parallel search – then off line comparison approach to patent research, some researchers have indeed become highly proficient in obtaining results that they believe to be reliable. This multiple database search process, considered by many as the "state of the art", is more appropriately defined as "status quo".

WORLD INTELLECTUAL PROPERTY ORGANIZATION (WIPO)

In order to address the "data not identical" problem, the World Intellectual Property Organization (WIPO) implemented a global standard in 2002 to which member nations agreed to comply. In Standard 32, WIPO established a complex digital data classification and management structure that would ensure that all states would be able to comply with a "least common denominator" definition.

While the immediate goal of this specification is to support E-PCT applications, the Trilateral Offices intended to use it as the basis for their own national electronic applications for a variety of industrial-property types and recommend that it would be the basis for an eventual WIPO standard for use by other Offices. With that in mind, the DTDs created for E-PCT will be constructed in components for element definitions and from which the Trilateral Offices and others can derive elements and DTDs for their needs in a consistent and compatible manner.[7]

More precisely, WIPO ST.32 defines the complex International Standard ISO 8879:1986, Information processing - Text and office systems - Standard Generalized Markup Language (SGML) tag scheme as applied exclusively to patent data. This format also requires patent data to follow a specified Document Type Definition (DTD) per WIPO variations of ISO ISO 8879, ISO 646 and others.

To illustrate some of the document data specifications outlined by WIPO, we'll start with a brief overview of the treatment of chemical compounds and formulae.[8] Provision is made for non-protected variables included in compound definitions, and in other cases, the inclusion of the compound in the description (or abstract or claims). Note the superscripts provided on 8.b. below – necessitating adherence to ISO 8879 and 646 in order for the formula text to be properly represented in the HTML or SGML version of the patent data:

8. PROVIDED THAT at least one of Ra, Rb and Rc represents an unprotected group; and, if required, the following steps, in any order:
   a. removing any protecting group, to give a compound of formula (I), and,
   b. if required, converting any group represented by R^1, R^2 or R^3 to any other group so represented, and,
   c. if required, converting a compound where R4 represents a hydrogen atom and R5 represents a cyano group to a compound where R4 represents a cyano group and R5 represents a hydrogen atom, or vice versa.

DETAILED DESCRIPTION OF THE INVENTION

9. To prepare the (+) enantiomer of the title compound, the reaction was run under the same conditions except that (+)–tramadol as the free base was used instead of the (-)-tramadol to yield 2.8 g of the (+) enantiomer of O-desmethyl tramadol (mp. 242-3°C)_25 = +32.2°
(C=1, EtOH). D

10. Alternatively a route to compounds 1 where r > 1 is available by condensation of compounds of structure 4 with an amino biphenyl methyl amine such as 51 . 4 may be prepared by heating the amino pyrazine carboxylic acid with excess acid chloride or anhydride. The precursor 2 may be prepared by heating 4 with ammonium carbonate to give the 2-substituted-pyrazino[2,3-d]pyrimidinones 2

In addition to teaching how mathematical and chemical characters are treated, ISO 8879 further prescribes the method of coding many other characters in order to accommodate language variations in a standardized SGML / XML format, as shows in the small sampling below.

```
<!ENTITY % ISOlat1 PUBLIC “ISO 8879:1986//ENTITIES Added Latin 1//EN”>
%ISOlat1;

<!ENTITY aacute SDATA “[á]”--=small a, acute accent-->
<!ENTITY Aacute SDATA “[Á]”--=capital A, acute accent-->
<!ENTITY acirc  SDATA “[â]”--=small a, circumflex accent-->
<!ENTITY Acirc  SDATA “[Â]”--=capital A, circumflex accent-->
<!ENTITY agrave SDATA “[à]”--=small a, grave accent-->
<!ENTITY Agrave SDATA “[Ë]”--=capital A, grave accent-->
<!ENTITY aring  SDATA “[å]”--=small a, ring-->
<!ENTITY Aring  SDATA “[Å]”--=capital A, ring-->
<!ENTITY atilde SDATA “[ã]”--=small a, tilde-->
<!ENTITY Atilde SDATA “[Ã]”--=capital A, tilde-->
<!ENTITY auml   SDATA “[ä]”--=small a, dieresis or umlaut mark-->
<!ENTITY Auml   SDATA “[Ä]”--=capital A, dieresis or umlaut mark-->
<!ENTITY aelig  SDATA “[æ]”--=small ae diphthong (ligature)-->
<!ENTITY AElig  SDATA “[Æ]”--=capital AE diphthong (ligature)-->
```

These standards effect the way in which thousands of characters appearing in patent data are (a) read, (b) recorded, (c) searched, and (d) displayed. The issues involve how to store and retrieve special characters including scientific and chemical symbols, German & French umlauts, mathematical symbols and so forth, from a patent database.[9]

Finally, WIPO standards help retain table data within the patent document text as shown in the following EXAMPLE TABLE[10]

15. The pigment base was diluted by mixing (not grinding) with a much larger quantity of the opaque white bleach base, so as to eliminate any minor differences of gloss and hue. In each of these comparisons, 4g of a pigment grind base (formulated as shown above) were mixed with 4g of water and 32g of the above bleach base. The results are listed in Table I below.

TABLE I				
Bleach Test1				
	J–678		GA–1	
	Density	Gloss3	Density	Gloss3
Yellow	0.66	60.2	0.66	63.1
Rubine	0.81	56.5	0.82	57.3
Blue	1.11	58.9	1.11	60.5
Black	0.95	68.7	0.95	67.1
1. Printed with #7 meyer bar on Printkote® Board. 2. Cosar Pressmate 102 Densitometer used. 3. Gloss Guard II Glossmeter, 60_.				

LATENT SEMANTIC ANALYSIS ("LSA") PATENT SEARCH TECHNOLOGY

One can readily see that under the WIPO XML standard, literally hundreds of symbols, characters and data formatting structures are retained in the digital version of the document.

Searching these symbols, in most cases, is impossible using traditional Boolean search technology. Not only does Boolean require the researcher to enter the precise text they are searching for, there is usually no provision for entering superscripts, non-English characters, or table data.

Of course, Boolean search technology, because of its precision, inherently misses important patent documents simply because the researcher cannot know, nor can s/he anticipate every possible combination of searchable keywords that *may* appear in a patent document. However, Boolean search technology limitations are not a focus of this paper.

Latent Semantic Analysis (natural language) search technology, as deployed in ICO-GPS, can indeed search, understand and return relevant results corresponding to a "natural language' query that may contain characters as outlined in ISO 8879. Combined with the Boolean filtering capability, even combinations of traditional text queries AND variables, such as the non-protected "R" variables above that may appear in Markush formulae, can be extracted from global patent data, depending of course, on the skill of the researcher to employ advanced techniques that "push" the ICO's Concept + Boolean technology toward the creation of an algebraic equation. (Note: Markush searching is not a prescribed use of ICO Global Patent Search. However, experiments conducted by PatentCafe indicate that relevant results can be obtained by searching Markush formulae as noted above. The results lack reliable consistency to be considered a robust Markush search tool.)

However, in order for ICO-GPS' LAS technology to find relevant documents, it is critical that ICO-GPS patent database not only contain all of the data currently structured under the WIPO standard, but to ensure that that data, and corresponding data structure, is precisely maintained. Without the stringent requirements imposed by WIPO and ISO standards, the creation of a standardized data structure across all authorities would not have been possible.

Therefore, in a very real sense, the implementation of the WIPO standard is what made possible the application of the advanced natural language search technology to patent data.

With the adoption of WIPO ST.32 (or approved variations thereof) by the Trilateral and other issuing authorities, one can expect a certain level of data structure and document type compliance in newly issued patents. As indicated earlier, there remains a problem in addressing the majority of the 35 million patents contained in the back files – previously issued patents that were not required to conform to the new standard.

For these files, OCR scanning is the standard of practice. Given the inherent inaccuracies of OCR technologies, it is easy to surmise that scanned patent back files will contain innumerable inaccuracies. Under the WIPO standard, these inaccuracies are unacceptable – primarily from a technical sense.

In other words, in order for an XML document to be successfully loaded into the ICO-GPS' patent database as an XML document, it *must conform* to the prescribed data structure – or it will be rejected. An underlying assumption is that the database was designed and structured *from scratch* to precisely store and retrieve XML formatted patent data. Most legacy public and commercial patent databases were never designed to handle the WIPO structured data, and therefore contain a more traditional "data dump" wherein inaccuracies in data are accepted, stored and indexed. Conversion of a legacy database is an extraordinary task – oftentimes better addressed by starting from scratch.

Based on the important, searchable WIPO parameters, the ICO-GPS patent database is structured to allow the importation of only 100% conforming data. Certain data fields may be added to the "searchable" range in the future, but at present, if there is any missing SGML tags, missing data, transposed data within fields (a common problem in transposed date fields, or different date formats contained in old paper files), or if data expected to appear between specified tags is erroneously placed within the wrong tags, as happens, for example, with CLAIMS appearing within the SPECIFICATION rather than in the CLAIMS section, the data is not allowed to load.

Adherence to the WIPO and ISO standards, as well as its own import quality checking process ensures that PatentCafe's data, even data harvested from patent offices' back files, remain on course to meet its 6-Sigma statistical patent data quality target.

Once a 6-Sigma definition appears on the landscape, the questions begin to surface, asking what is "6-Sigma", and how does that apply to patent data?"

6-SIGMA STATISTICAL QUALITY LEVEL

What is 6-Sigma?

The term sigma is a Greek alphabet letter (σ) used to describe variability. The common measurement index used in 6-Sigma is DPMO (Defects Per Million Operations) and can include anything from a component, piece of material, or line of code, to an administrative form, time frame or distance. A sigma quality level offers an indicator of how often defects are likely to occur, where a higher sigma quality level indicates a process that is less likely to create defects. Consequently, as sigma level of quality increases, product reliability improves, the need for testing and inspection diminishes, work in progress declines, cycle time goes down, costs go down, and customer satisfaction goes up.

6-Sigma has become the world standard for statistical quality sampling, having been perfected in business by Motorola, Allied Signal, and General Electric. Currently, some of the leading corporations that must deliver reliability in highly competitive markets (below) have adopted 6-Sigma performance standards throughout their organizations (insofar as 6-Sigma quality is achievable, available, and meaningful to competitive positioning and overall reliability). Now, intellectual property management can enter the domain of 6-Sigma processes with the availability of quality patent data (and the confidence and reliability that 6-Sigma creates).

<u>6-Sigma processes are important because they have proven to increase efficiency and quality of management decisions, while reducing errors, time, costs and the risks or management decisions made upon poor data.</u>

Long-Term Process Capability in Various Sigma Levels:

Sigma Level	% Good	PPM/DPMO
2	69.15	308,537
3	93.32	66,807
4	99.379	6,210
5	99.9676	233
6	99.99966	3.4

Specifically with regard to patent date, 6-SIGMA refers to an acceptable deficiency level, of 3.4 errors appearing in 1 million operations.

Companies currently employing 6-Sigma processes include:

Du Pont	Pfizer	Merck	General Motors	IBM	3M
Johnson & Johnson	Baxter	Dow CMS	Schering Plough	Ford	Motorola
Estee Lauder	Tefen USA	Wyeth	Novartis	Ecolab	Allied Signal
General Electric	Albemarle	Rohm & Haas	Procter & Gamble	Engelhard	FAA

... and hundreds of other leading innovators.

First, the cost benefits of implementing 6-SIGMA processes in traditional operations have been well documented. Typical North American companies average a 3-Sigma level. In other words, 25-40% of most companies' annual revenue gets consumed by their cost of quality. **Each improvement by 1 Sigma level will increase net income by approximately 10%!**[11]

The availability of desk-top access to major patent databases by virtually anyone with access to the Internet has seemingly created a lingering euphoria and satisfaction based primarily on the mere ease of access, even though research professionals have long known (and accepted) the "thousands", and even the "100,000" text errors reportedly appearing in the patent data of the world's leading patent issuing authorities.

Yet, until higher quality patent data, and a more precise technology to search that data became available, acceptance of those errors was simply "status quo". No matter how you slice or dice poor patent data, older technology tools searching 5-Sigma patent data still produce results that at best contain a known error rate of nearly 250 per million records. Computer parlance popularized during the 1980s expresses this syndrome well: Garbage In = Garbage Out, or "GIGO".

While patent researchers have been shown to defend a false sense of confidence based on self-proclaimed skill or aptitude in creating "acceptable" comprehensive search queries and search strategies, the fact remains that even if a researcher was capable of creating search methodologies that approached "6-Sigma" (if such a metric existed), the underlying data and search results are still based on 5-Sigma quality at best (3-Sigma or 4-Sigma quality at worst). All the while, the critical processes happening throughout other functional areas of most of their employers indeed follow 6-Sigma statistical quality and continual improvement process.

With the exception of licensing activity tied to a patent portfolio, patent research is largely a *cost center*. Therefore, the increased costs of using 5-Sigma verses 6-Sigma patent quality systems, run straight to the company's bottom line as a (loss). **If the (loss) associated with a cost center is the inverse of, but analogous to the "net income" associated with revenue, the loss based on each 1 Sigma quality increase could possibly approach 10% of the company's total annual investment in patent driven R&D, researchers' salaries, patent legal costs, and associated overhead.** The loss in market protection anticipated by patents earned under 5-Sigma could be incalculable – possibly on the order of multiples of the total investment in 5-Sigma patent systems. The authors have found no data supporting this assumption.

Largely based on the proven increases of net income by companies adopting 6-Sigma process standards, one must question the hard costs of relying on older 5-Sigma patent research systems. Has the poor quality and efficiency of currently deployed 5-Sigma patent research systems:

a) Supported development or technologies or patents that ultimately resulted in re-examination or invalidation of any of the company's patents? If so, what was the total R&D, production and legal investment, and revenue reduction (losses) that resulted from the challenge?

b) Caused delay in discovering prior art that would have quickly determined patentability? If so, how many man-hours were spent (salaries lost) using less efficient research tools?

c) Missed patents that would have invalidated or caused to be re-examined, or could have been used in affirmative defense, any patents which were missed using traditional search technology upon a 5-Sigma (or worse) patent database?

d) Missed recognizing valuable licensing opportunities? The issuance of patents that describe your inventive matter, yet which were classified in non-obvious classifications, or which were lacking the "obvious" keywords that Boolean search technology requires, most likely missed identifying R&D activities in other industry segments that could have benefited from licensing your technologies.

e) Caused management decisions to be made on erroneous data presented by patent visualization tools (*patinformatics*)? If so, what was the economic loss associated with those decisions?

It's clear that as continually improved quality or patent data and research tools evolve, such as the movement toward 6-Sigma patent quality systems, reliance on 5-Sigma, or lower, quality levels, increase the likelihood that patent data related decisions will result in significant financial or technology loss. The business risk is unacceptably high.

PATENTCAFE'S 6-SIGMA PROCESS

Taking the long term process approach to creating 6-Sigma patent data quality, PatentCafe incorporates a number of processes, beginning with the original data extraction from various patent repositories, and continuing through a number of internal processes until the final, acceptable patent data set is entered into the ICO Global Patent database.

PatentCafe's 6-Sigma quality initiative begins with the re-formatting of ALL patent data to the previously mentioned WIPO ST.32 XML standards, regardless of the original format provided by respective patent offices. The errors encountered are many - and expectedly so, since WIPO standards became effective only in late 2002. Consequently, when we say "current digital patent collections", we are generally referring to the wide variety or patent data formats. The errors, regardless of the type or root cause, need to be corrected - a process that requires significant manual labor.

The result is ICO Global Patent Database - in fact, a new benchmark in a single global patent database containing patents meeting ISO standardized format, data quality processes targeting 6-Sigma quality, and accessibility of the data by an advanced search technology (but even the traditional Boolean search functionality incorporated into ICO-GPS benefits from the higher patent quality levels).

6-Sigma is not a "test score" or an "event". Rather, 6-Sigma defines a long-term continuous improvement process, which follows a ***Measure***, ***Analyze***, ***Improve*** and ***Control*** discipline at all points of contact where corporate operational and technical functions meet patent data, or hardware on which patent data resides.

Some of these processes include:

1) Extracting patent data from databases in an XML format
2) Validating the extracted documents and data against a known chart of issued patent numbers
3) Filtering patent data through a proprietary XML import filter
4) Quarantining defective patents or patent data
5) Manually repairing defective data until it is in 100% compliance with PatentCafe's import filter.

Once patent data is verified, validated, and contained within the ICO-GPS database, it is indexed, and becomes available for searching via. ICO-GPS Concept + Boolean search technology. For more information on PatentCafe's LSA technology, please see: http://www.PatentCafe.com

CONCLUSION:

PatentCafe has built a patent database and data management system that brings together the three most crucial components necessary to raise the quality and standard of practice in high quality, patent data storage and retrieval to a new industry high:

- Ground-up development of a patent database structure optimized for very large patent data sets,
- Incorporation of WIPO's Standard ST.32 per modifications of the PCT,
- ISO standards storage and retrieval system to manage multi-language and scientific characters
- Latent Semantic Analysis search technology proven capable of fast retrieval of data meeting the new WIPO standards
- A persistent quality measurement and improvement system to ensure that data, hardware and software work as a reliable, high quality patent data delivery system.

By applying 6-Sigma systematization and the traditional cost/benefit results obtained from 6-Sigma quality initiatives deployed by the world's leading corporations we believe that a valid conclusion can be made that PatentCafe ICO Global Patent research solutions will improve patent research efficiency, discover a higher level of relevant patents upon which critical business decisions will be made, will improve the speed at which relevant documents can be discovered, and deliver an overall return on investment that could approach a 10% reduction in costs of operations, financial loss, or 10% increase in profitability for corporations that currently use patent research tools, or patent data that is now maintained at a 5-Sigma level or less.

ENDNOTES

[1] *Innovation & Technology Transfer*, Vol. 1/00, January 2000, p.15

[2] *Essentials of Patents*, Andy Gibbs and Bob DeMatteis, John Wiley & Sons (2002)

[3] *Brookings Papers on Economic Activity 1: 2000 (2000),* William C. Brainard and George L. Perry, Editors

[4] *Intellectual Property Strategy for Generic Drug Manufacturers in a Crowded, Rapidly Changing Arena,* Howard E. Davis and Bruce Rubinger Ph.D, Senior Biotech Analyst and Managing Director, Global Prior Art (GPA), Inc.

[5] *Patinformatics: Identifying Haystacks from Space*, Trippe, Searcher, Vol. 10, No. 9, October 2002

[6] Innovation Policy in a Knowledge-Based Economy, A Merit Study Commissioned By The European Commission Enterprise Directorate General (June 2000)

[7] *XML DTDs For The E-PCT Standard* prepared by the International Bureau for the purposes of consultation under Rule 89.2(b) A further revised version of document PCT/AI/1 Add.4 Prov., superseding document PCT/AI/1 Add.4 Prov. Rev.2 dated January 30, 2001.

[8] *PATENT COOPERATION TREATY (PCT) Administrative Instructions Under The Pct, Proposed Modifications Relating To The Electronic Filing And Processing Of International Applications Annex F, Appendix I XML DTDs For The E-PCT Standard Handbook On Industrial Property Information And Documentation* Ref.: Standards – ST.32 page: 3.32.107

[9] *Synopsis of the treatment of natural language characters by WIPO ST.32*.

[10] *Handbook On Industrial Property Information And Documentation*, Ref.: Standards – ST.32 page: 3.32.109

[11] *Quality Process Associates*, http://www.pqa.net/sixsigma/W06002003.html

PatentCafe.com, Inc., Main Corp
441 Colusa Avenue, Yuba City, CA 95991
Tel: 530-671-0200 • FAX 530-671-0201