From:
Sent: Monday, March 08, 2010 4:54 PM
To: patent_quality_comments
Cc: James E. Malackowski; Kristi Stathis
Subject: Comments on Enhancement of Patent Quality

Commissioner for Patents,
P.O. Box 1450, Alexandria, VA 22313–1450,
Attn: Kenneth M. Schor and Pinchus M. Laufer

In response to the notice ('Request for Comments') appearing in the Federal Register at Volume 74, number 235, December 9, 2009, pages 65093-65100 [Docket No.: PTO–P–2009–0054] please find attached a pdf document titled "WRITTEN COMMENTS ON ENHANCEMENT IN THE QUALITY OF PATENTS."

If you could please be so kind as to confirm safe receipt of this e-mail and the attached document it would be greatly appreciated. Thank you!

Best regards,
Jonathan A. Barney

JONATHAN A. BARNEY

OCEAN TOMO, LLC

4695 MACARTHUR COURT

SUITE 1141

NEWPORT BEACH, CA 92660
(949) 798-5616 PHONE
(949) 395-2535 CELL

jbarney@oceantomo.com

WRITTEN COMMENTS

ON

ENHANCEMENT IN THE QUALITY OF PATENTS

Submitted by:
Jonathan A. Barney
Managing Director Ocean Tomo, LLC
Founder & CEO PatentRatings, LLC

March 8, 2010

The comments below are respectfully submitted in response to the notice ('Request for Comments') appearing in the Federal Register at Volume 74, number 235, December 9, 2009, pages 65093-65100 [Docket No.: PTO–P–2009–0054].

By quick way of background I am a registered U.S. patent lawyer.[1] I have been preparing and prosecuting patents in the USPTO since 1989 as both an in-house patent attorney/agent and as a partner/practicioner in a major U.S. patent law firm. I am also an active inventor and have received and/or applied for dozens of patents.

About ten years ago I began investigating various approaches for analyzing the quality of U.S. patents using an objective statistical framework. One approach used multivariate regression modeling to essentially predict the probability of maintenance or abandonment of patents based on multiple quality-probative input metrics.[2] Starting from that early work I eventually developed (with others) a predictive ratings algorithm for statistically assessing patent quality for every in-force U.S. patent granted since 1982. In 2000 I founded PatentRatings, LLC which has since developed and commercialized a similar ratings algorithm under the "IPQ" brand. In 2005 I joined Ocean Tomo where I have continued my work developing and refining the IPQ ratings algorithm and related statistical tools, including the anticipated launch in 2010 of the first European version.[3]

---

[1] USPTO Reg. No. 34,292

[2] *See* J. Barney, *A Study of Patent Mortality Rates: Using Statistical Survival Analysis to Rate and Value Patent Assets*, 30 AIPLA Q.J. 317, 329 (2002); *See also* U.S. Pat. No. 6,556,992 titled *Method and system for rating patents and other intangible assets*.

[3] In partnership with Caisse des Depots et Consignations ("CDC")**.**

Given that background I very much appreciate the invitation and opportunity to provide written comments and suggestions in this important matter. In the interest of time and efficiency I have kept my comments very brief and to the point. However, I am available to expound on any of the suggestions presented below and, in general, provide any additional help or input the USPTO may desire as part of carrying out this important initiative. All of the comments and views expressed below are my own and do not neccessarily reflect the views or opinions of Ocean Tomo, CDC, Patent Ratings, or any other person or entity with whom I am currently or fomerly associated.

General Comments on Enhancing Patent Quality

One of the foundational tenets of modern management theory is that you cannot effectively manage that which you cannot measure. Performance benchmarking (organizing and driving managed resources against goal-specific objectives) is a crucial tool in exerting management control over desired performance objectives. However, identifying and maintaining accurate performance metrics appropriate to goal-specific objectives can present its own challenges. The frequently encountered conundrum is that not all things that *should* be measured *can* be measured and, conversely, not all things that *can* be measured *should* be measured.

This same old conundrum holds true in the context of patent quality management. However, the universe of things that can be measured today has greatly expanded thanks to the advent of powerful computers and the widespread availability of vast quantities of electronic information.   Such advances have enabled the development of sophisticated statistical models and predictive analytics tools for objectively benchmarking and measuring patent quality.

While statistical benchmarks cannot offer absolute precision in measuring patent quality, they do provide a way to at least probe the issue and draw statistically informative conclusions using objective criteria. For example, significant insights have been gained from analysing past renewal decisions of patent owners. The natural attrition effect of the maintenance fee system discourages the renewal of less valuable, poorer-quality patents by placing substantial recurring costs on all patents. By examining the characteristics of patents that were renewed relative to those that were abandoned, one can begin to make certain predictive assessments about the quality and likely value of other patents that share statistically similar attributes.

Similar models could be developed and used by the USPTO for internal or external quality measurement and communication purposes. For example, a multivariate regression model could be developed that would assess the probability and timing of patent allowance/rejection based on the quality of a submitted application text, initial claim set and prior art search. Scores or ratings calculated according to such a model could be used to more effectively monitor internal USPTO resources. It could also be used to communicate valuable information to the applicant as to the quality of the submitted application, amount and quality of prior art uncovered and probability and expected timing of key prosecution events (allowance, rejection, issuance, etc.). Of course, these are very high-level concepts that could benefit from further elaboration.

The remaining comments are more specific and primarily address suggested approaches for measuring and improving the quality of prior art searching. These comments are organized according to the specific categories as outlined in the notice.

Category 1 — Quality measures used

A. Identification of the key items, i.e., the activities and actions that are carried out by the USPTO, by the applicant, or by both, that bear on quality. What is the nature of activity, action, or conduct that increases quality, and why is it believed to do so?

Thoroughly and accurately identifying material prior art is the foundation of all further steps carried out in the patent examination process and, therefore, must be a primary focal point for any patent quality enhancement initiative. While most experienced examiners do a good job searching for and identifying relevant prior art, search quality is not consistent across all art units and all examiners, particularly newer examiners and examiners who may be assigned to a new or unfamiliar art unit.

There are also certain systemic data challenges which significantly hamper the ability of examiners, professional searchers and search tools in identifying and retrieving the best prior art. Specific suggestions for improvement here include:

- Early publication of a detailed search report indicating cited prior art according to its degree of materiality to the claimed invention
- Permanent on-line archiving of web pages and other web resources cited as prior art (perhaps in partnership with the Internet Archive)
- Indexing and archiving of material non-patent literature cited as prior art
- Increased infrastructure and bandwidth for public search resources

B. Identification of indicia of the presence of the desired quality items. How do the proposed indicia show that the desired activities and actions were indeed carried out, and show the quality or effectiveness of that activity performed by the USPTO and/or the applicant?

Primary indicia of quality prior art searching include:

- Relatively lower rate of OPQA compliance failure (IPR or Allowance/Final Action Compliance) based on initially undiscovered prior art
- Relatively lower rate of reissue/reexamination based on previously undiscovered prior art
- Relatively lower number of previously undiscovered novelty-destroying references found in a subsequently prosecuted foreign counterpart application
- Relatively lower number of judicially invalidated patents based on previously undiscovered prior art

Each quality indicia above can be readily measured and compared for a given cohort of patents grouped, for example, by class, date, art unit or examiner. New and more direct search quality indicia can also be developed such as: i) randomly selecting cases for duplicate or repeat searching at the pre-examination/search stage by a non-assigned examiner and comparing search results; ii) using one or more automated search tools to generate and compare search results; and/or iii) a new OPQA compliance program specifically deployed at the pre-examination/search stage and designed to evaluate search quality compliance.

C. What metric(s) should the USPTO use to measure each indicium, and what is the nexus between the measured indicium and the metric(s) used (why is the existence of the indicium proved by the metric)? Based on that nexus, why is the proposed metric believed to provide a practical combination of reliability and efficiency?

There are a number of probative metrics that can be derived from the statistical data to measure or make predictions about patent search quality. Depending on the particular quality indicium selected, suitable metrics may include, for example:

- Comparative number of cited prior art references
- Comparative age of cited prior art references
- Relative crowdedness of the technology field
- Ranking search results according to one or more automated search tools
- Various relevance metrics evolved from the patent citation network [4]

One recommendation would be to develop a multivariate regression model to optimally weight multiple input metrics to predict a desired quality indicium (e.g., likelihood of OPQA compliance). Such a model could provide a practical and efficient

---

[4] *See* below discussion new tools under Category 6.

tool for statistically monitoring patent search performance while helping identify statistical outliers.

Category 2 — Stages of Monitoring

I agree it is wise policy to monitor patent quality at every step in the process and as close in time to when the step whose quality is being measured is performed as is feasible. In the searching context one way to achieve that goal very effectively is to randomly select cases for duplicate (parallel or repeat) searching at the pre-examination stage by a non-assigned examiner. In essence, each examiner would independently conduct his or her own search of the application. After each search is completed, the search results would be compared and an assessment made as to the relative quality of each search. With enough comparative search results it would be possible to identify consistently strong performers and consistently weak performers. Best practices could then be developed and adopted.

Category 3 — Pendency

Reducing patent pendency is really the flip side of the same coin as achieving high patent quality. Obviously, the more time and resources the USPTO spends searching and examining a patent application, the higher quality the final product is likely to be. But, it's little consolation for an applicant to receive a patent of the utmost highest quality if it comes four or five years too late! Long pendency delays are particularly painful in rapidly evolving technology areas where innovation rates can reach as high as 20% or 30% per year and product half-lives can be as short as 3-4 years. Quality cannot be the *only* goal of the USPTO, but must be balanced against other practical considerations including time and financial resources.

In theory, continuation practice should reduce pendency times by allowing applicants to secure the easier, narrower claims first while preserving applicants' rights to pursue broader claims later. In practice, however, it often has the opposite effect of prolonging overall prosecution by encouraging applicants to file multiple continuation applications instead of either appealing a final rejection or abandoning the case. It's hard to say exactly where the right balance is. The USPTO always has the option to increase filing fees or levy other special surcharges to discourage excessive use of continuations.

Another way to at least partially alleviate the long pendency problem without negatively affecting patent quality would be to offer a deferred examination option. This would allow applicants the option to keep their patent application alive for a certain period of time, and without incurring substantial prosecution costs, while they investigate commercialization opportunities. Based on observed historical patent maintenance behavior (>50% of all U.S. patents being abandoned by year 12) it is likely that a substantial portion of these unexamined patent applications would also be abandoned before any substantive examination is performed. That would save both examiner time and costs while reducing the total number of applications needing to be examined.

Category 4 — Pilot Programs

No comments.

Category 5 — Customer Surveys

No comments.

Category 6 — Tools for Achieving Objectives

One unique tool that may help the USPTO achieve its quality objectives is the Relevance Engine™ patent search tool, which is available commercially as part of Ocean Tomo's standard web subscription platform.[5] Basically, this search tool allows a searcher to find relevant patent documents (and eventually non-patent documents) simply by selecting one or more starting input documents. It also allows a user to measure the

---

[5] White paper attached (*Patent Ratings Relevance Score*); *see also* U.S. Patent Appl. Pub. No. 2007-0073748 titled *Method and system for probabilistically quantifying and visualizing relevance between two or more citationally or contextually related data objects.*

relevance (i.e., theoretical probability of citation) between any two or more patents. The search tool works by leveraging the existing citation network that directly or indirectly connects virtually all patents to statistically assess the probability of a direct citation between any one patent and any other patent within the citation universe.

One immediate advantage of the tool is ease of searching and the ability to produce a high-quality search result set that is based purely on citation analysis (extending out up to 6 generations) and not based on traditional keyword searching, classification searching or semantics analysis. Because it is based only on citations there are no inherent language limitations that would prevent the tool from finding relevant references written in languages other than English. Alternatively, the tool could be used as a final check or verification metric to assess search quality and/or to help locate possible missing or overlooked material prior art.

Category 7—Incentives

No comments.

Concluding Comments

I applaud the USPTO for taking on this important initiative to enhance patent quality. The management approach the USPTO has outlined is, I believe, the correct one and should lead to measurable quality improvement. I hope the above comments and suggestions are useful in helping the USPTO achieve success in this endeavor.

Respectfully submitted,

Jonathan A. Barney

OCEAN TOMO
INTELLECTUAL CAPITAL EQUITY

PatentRatings Relevance Score

Relevance Score Background

Corporations and entrepreneurial ventures are increasingly placing greater emphasis on identifying and quantitatively analyzing "relevant" patents in the context of the various competitive landscapes within which they exist.

The difficulty and challenge of searching and analyzing patents come from various factors such as: the sheer volume of potentially relevant patent documents (estimated at over 80 million documents worldwide), latent inaccuracies and inconsistencies in the technology classifications, the complex scientific nature of patent disclosures, the ever evolving lexicon for describing novel patented concepts, language translation issues, and use of complex legalese.

The standard input/output text interface of most conventional patent analytics tools using keyword searching, latent semantic analysis also does a poor job of displaying and communicating input/output search criteria and search results in a way that facilitates intuitive understanding and visualization of the logical relationships sought to be explored between two or more related concepts being searched. It would be of particular benefit to provide an improved search algorithm, database and user interface that would overcome or at least mitigate some or all of the above-noted problems and limitations.

As a result, patent analysis using conventional data base queries and search engines tends to be a slow and tedious process, typically producing significant numbers of irrelevant documents or "false positive hits" and often failing to turn up one or more desired relevant patent documents. The use of relational citation analysis is proven to greatly improve the ability to search, identify and categorize patent documents according to relevant subject matter. Therefore, PatentRatings Relevance method seeks to mitigate the above limitations inherent with prevailing approaches by systematically analyzing entire citational relationships that exist among all US patents.

Relevance Score Overview

The PatentRatings Relevance score is a quantitative representation of the technical relevance or relatedness of any one US patent to any other. The relevance score essentially represents the probability (on a scale from zero to one) that a first patent directly cites a second patent (i.e. that one directly cites the other) based on the network of intercitational relationships among all patents. Relevance scores are calculated by analyzing up to six generations of citational relationships (i.e. one generation includes all patents cited by or citing a reference patent) that exist among all issued patents. For example, examiners at the USPTO cite relevant prior art patents and other relevant documents based on a thorough search of the technical literature. These are the primary or first-generation citations. The PatentRatings system extrapolates this information by building a citation network and observing actual citational relationships that exist among patents dating back about 50 years to present. Predictive modeling techniques are applied to arrive at an event-specific probability of a primary citation occurring between two patents (a first generation citation) based on the entire network of intercitational relationships among all patents.

The analysis essentially determines the degree of overlap between the network of citational relationships of the two reference patents. More overlapping citational relationships indicate a statistically higher probability that the two patents are citationally related at the first generation. The resulting relevancy scores can be used to not only identify other relevant patents (both earlier and later filed) but can be used to objectively and statistically quantify the degree of relevance. Thus, it provides an improved model approach for quantitatively measuring a degree of relevance between two or more patents and/or other documents of interest.

The PatentRatings' Relevance database contains over 4.8 billion records for over 5 million US patents (as of May 2009) patent to patent Relevance relationships above a threshold value of 0.05. Typically about 1000 patents are mapped as "Relevant" to a single US issued patent in a rank order (.999 to 0). The relevance scores are pre-processed and are maintained in a database exclusively accessible through www.patentratings.com and are updated periodically.

The primary use for the PatentRatings Relevance system is the identification of relevant patents and their current owners for the purpose of acquisition, sale, licensing or competitive analysis. In the typical corporate and investment context, the identification of relevant opportunities is not only subjective, but a difficult and resource driven activity.

The Ocean Tomo Relevance Engine determines the most closely related patents and identifies their current owners based on USPTO Assignment records. This allows decision makers to quickly and objectively obtain information regarding relevant patents, technologies, potential partners, acquisitions or other strategic targets based on Patent information.

Relevance Score Construction and Theory

Relevance scores utilizes multiple regression modeling which is a statistical technique for examining the relationship between two or more predictor variables (PVs) and a criterion variable (CV). Predictor variables (or independent variables) describe or quantify certain observable characteristics of a particular patent population or other documents of interest, e.g., secondary citational relationships to other patents or related documents, etc. Criterion variables (or dependent variables) measure a selected quality of interest of a particular patent population, such as primary citational relationships. Multi-variate regression modeling allows the criterion variable to be studied as a function of the predictor variables in order to determine a probabilistic relationship between selected variables. This data, in turn, can be used to predict the presence or absence of the selected quality in other patents or related documents of interest.

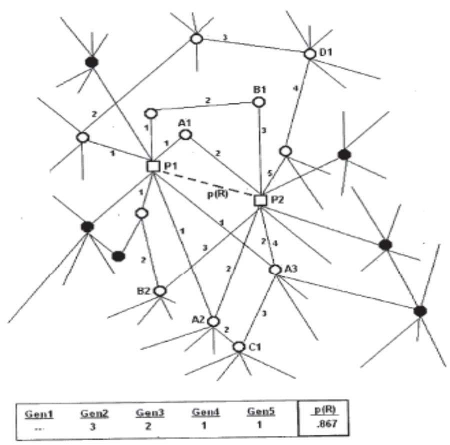
Relational citation analysis is a novel technique that exploits citational relationships ("relevance links") that may exist between two or more patent documents and/or other related documents of interest for the purpose of quantitatively measuring a degree of relevance. The primary assumption underlying the approach is that patent documents and/or other documents that are citationally related to one another at the first generation (one document directly citing the other) are "relevant" to one another. In other words, if document X directly cites (refers to or mentions) document Y (or vice versa, or both), then documents X and Y are considered to be citationally related at the first generation and are therefore deemed or assumed to be "relevant" to one another. If document X cites document Z which, in turn, cites document Y, then documents X and Y are considered to be citationally related at the second generation and are potentially relevant to one another according to a particular derived relevance probability distribution. Specifically, it has been determined that the occurrence of a citational relationship between any two documents at a second generation creates a corresponding measurable probability that the documents may be citationally related at the first generation. Thus, "relevance" between any two documents can be defined and measured in the absolute sense of an event-specific probability that the documents are citationally related at the first generation.

A relevance score can be mathematically expressed as the simple event probability that two or more documents of interest are citationally related at the first generation. Thus, for example, two documents having a direct citational relationship (one document directly citing the other document) can be described as having a relevance score of 1, indicating 100% probability of a direct citational relationship. Two documents having no citational relationship and no possible likelihood of a citational relationship at the first generation can be described as having a relevance score of 0, indicating 0% probability of a direct citational relationship. Thus, any two patent documents or other documents of interest selected from a given population can be characterized as having a certain relevance score calculated as the simple event probability that one or both documents would directly cite or reference the other.

There is a strong statistical covariance between citational relationships occurring at the first generation and citational relationships occurring at the second and higher generations. Intuitively, this makes some sense. Two patents or other documents that cite one another are also more likely to cite other contextually similar documents as well, thereby creating second generation and higher citational relationships. As a result of this strong covariance, a powerfully predictive probit or logit regression model can be constructed using the first generation citational relationship as the dependent variable (criterion variable sought to be predicted) and the second and higher generation citational relationships as independent variables (predictor variables). A suitably constructed regression model can then be optimized to calculate the event probability p(R) that a first-generation citational relationship exists between any two documents of interest by examining the number and type of citational relationships that may exist at the second generation and higher.

FIG. 1 illustrates one preferred approach for determining and measuring multi-generational citational relationships between two or more selected documents. In this case two patent documents (or other documents) P **1** and P **2** are selected for which it is desired to quantify the degree of relevance or relevance score. Each document P **1** and P **2** is citationally related to a total of 8 other documents at the first generation. This includes in each case 4 "backward" cites (document of interest citing earlier documents; illustrated depending from below) and 4 "forward" cites (later documents citing document of interest; illustrated extending from above). Preferably (though not necessarily), we ignore for now any actual citational relationship that may exist between documents P **1** and P **2** at the first generation since this is the dependent variable sought to be determined in the regression.

Figure 1: Patent Citations Map



Using basic computer database logic we extend multiple generations of citations and/or other relevance links from each document P **1** and P **2** and we identify and count the number of shared or overlapping citations at each generation. Thus, for example we identify 3 overlapping citational relationships at the second generation ("GEN2") citing common documents A **1,** A **2** and A **3.** Note that in each case A **1** -A **3,** we can count a total of 2 citational links separating document P **1** from document P **2,** corresponding to a second generation citational relationship. Similarly, we see there are a total of 2 citational relationships occurring at the third generation ("GEN3"), citing common documents B **1** and B **2.**

Finally, we see there is 1 citational relationship occurring at each the fourth and fifth generations ("GEN4" and "GEN5"), citing common documents C **1** and D **1,** respectively.

The determined count of citational relationships at each generation 2-5 are provided as input predictor variables (independent variables) to a multi-variate probit regression model. The regression model is formulated and optimally adjusted to predict the existence or absence of a first generation citational relationship between documents P **1** and P **2** (whether such relationship actually exists or not) and/or some other objective relationship based on some or all of the input predictor variables provided. The resulting probability score (and/or a mathematical derivation thereof) is an objective and repeatable probabilistic quantification of the likely relevance between documents P **1** and P **2.**

For more details on this method and technology utilized, please reference US patent application number: 11/236,965 or contact@patentratings.com. Please visit us at www.patentratings.com.