HEURISTIC BOOLEAN PATENT SEARCH
COMPARATIVE PATENT SEARCH QUALITY / COST
EVALUATION “SUPERBOOLEAN” VS. LEGACY BOOLEAN
SEARCH ENGINES

BY ANDY GIBBS

www.PantrosIP.com

White Paper Update - September 2012:
Former Company Name: PatentCafe®
Former Product Name: ICO™ Patent Search
Broken URLs Redirected as Available

Keywords: Patent search, Boolean, Heuristic, relevancy rank, expert system, artificial intelligence
ABSTRACT

This paper explores the application of an expert system to Boolean patent searching. Specifically, it will introduce skilled researchers to the next technological evolution of search methodologies that apply Heuristic Boolean methods to reduce cost, increase efficiency, and enhance search results quality.

Boolean search methodology, otherwise known as “keyword searching,” only extracts documents from a database that literally match the search query, but Boolean engines have no capability of determining which of those documents are of highest interest to the researcher. In an effort to overcome this limitation, modifications to simple Boolean engines have emerged, including truncation, proximity searching, nested complex query capability, and wildcarding.

But even with these enhancements, at best, Boolean searching remains little more than an iterative process of applying a query construct to a database in order to (a) extract a reasonable number of documents within (b) a reasonable amount of time, in order to (c) produce the most relevant documents supporting the search objective.

At worst, the restrictive nature of Boolean search methods inordinately increase direct and indirect search costs, and establishes a false confidence in search results quality that increase exposure to long term legal and commercial risks. The patent documents that Boolean engines inherently miss, sometimes discovered years later by other researchers, often establish the true (high) costs attributable to Boolean searches.

This paper examines how next generation Heuristic Boolean search methods can more quickly yield the most relevant documents, mitigate long-term risk associated with poor quality results, and reduce the direct, as well as hidden costs attributable to legacy keyword search engines. When the artificial intelligence of Heuristics is applied to Boolean patent searching, even novice researchers can quickly achieve reliable search results.

A future invalidity search is the ultimate quality test of today’s patentability search. Relying on the best search tools and processes today is critical. The future invalidity search performed when millions or 10s of millions of dollars are at risk, and can easily challenge and outperform the earlier patentability search because:

1) Liberal budgets for invalidity searches allow significantly more investment in search labor (higher cost), and
2) Invalidity searches rely on search technology advancements which have evolved since completing the patentability search.

CONCLUSION SUMMARY

The demands to perform a patent search that attempts to identify all of the relevant documents within the scope of available resources (time, budget, computing time, a given patent data quality) keep researchers reliant on the time-honored practice of crafting a lengthy, complex Boolean search string. But it’s been shown that such restrictions, although they produce relevant patents in a final results list, more dangerously drop an increasing number of relevant patents that should have been included in the final search report.

The application of heuristics such as a Latent Semantic Analysis / artificial intelligence expert system allows a patent researcher to use a less restrictive Boolean query, and obtain the Best-First search results list containing the highest quantity of documents more relevant to the search. Researchers are then able to manage very large search results lists without filtering the list to a more manageable quantity by using more keywords.

The results of applying heuristics to Boolean patent searching are faster time to identify the most relevant patents, but more importantly, the identification of the largest number of relevant patents that will serve as acceptable prior art.
**Introduction**

In the age of web-based access to nearly infinite information, the need to scour technical information for patentability of an invention has gone from “luxury” to “mission-critical”. The development of more advanced search engines is moving at breakneck speed.

The validity of a patent may not be tested for at least three to five years after filing an application. Although “unfair”, that’s also the time when a defendant will invest considerable time and money, using the search technology of tomorrow to identify un-cited prior art that the first patentability search missed, often invalidating the patent.

Boolean patent searching, otherwise known as “keyword” searching, has long been the trusted method of ferreting out patents that may teach the present invention. Clearly, with infringement suits being filed at the rate of more than 10-per-day, and infringement awards more frequently hitting the $1/2 billion mark, Boolean search methods as a whole no longer live up to the reputation of being a reliable, long term safeguard to patentability.

Given that even today’s best patent search technology in the hands of a skilled researcher will be tested with superior tools mining more data in the future, the hypothesis is that the use of traditional legacy patent search tools, specifically free and commercially available Boolean patent search tools, will result in an increasing number of invalidated or otherwise successfully challenged patents based on the later discovery of prior art that the Boolean engine missed.

In his IEEE paper “The Combinatorics of Heuristic Search Termination for Object Recognition in Cluttered Environments”[1], Grimson illustrates the total cost and performance benefits of applying an expert system to help the researcher more effectively find relevant documents in very large data collections. Heuristic processes do carry over to searching the highly-complex data contained in patent documents, but how so?

The performance metrics of research technique primarily include: search quality (discovery of the most relevant patents); speed; cost; reliability; and, durability (probability of the results surviving testing by the next generation of patent search engines).

Reasonable commercial standards are usually applied to patent searching. When is a patent search “good enough”? Usually, when a professional spends the specified time, and consumes the resources allocated to the search. However, while the searcher may find “good patents”, good patents are not as important as “good patents not found”, especially if those un-found patents actually teach more closely related prior art than those cited in the researcher’s final search report.

Boolean search engines are designed to produce positive results (you see only what your keywords you ask for), but they are incapable of returning patents it ‘knows you are looking for’.

Search strategies can be classified as “complete” or “heuristic”. In “Optimization by learning and simulation of Bayesian and Gaussian networks”[2], Larranaga, Etxeberria, Lozano, and Pena explain that the underlying idea in the complete search is the highly-complex data contained of all the possible points of the search space.

Patent data is voluminous, and as a homogeneous collection constitutes one of the world’s largest data sets. There are about 147,000 US Patent Classifications (including classes and subclasses). Pantros IP’s database of 25 million patent documents contains about 2.2 million distinct invention concepts, disclosed in about 1/4 billion pages of full text. Performing a complete Boolean search using the number of possible keyword combinations necessary to examine this volume of data to find relevant prior art is economically prohibitive.

---


Further, a complete search requires the crafting of a complex Boolean string to optimize the discovery of all relevant patent documents as instructed by Larranaga et. al., and still cannot reasonably examine all possible points in the space. With almost 100% certainty, with almost every search, one can be assured of missing important documents.

This paper will explore how Latent Semantic Analysis technology can be applied as an “expert system”, raising traditional Boolean patent searching to Heuristic, or “SuperBoolean™” searching. It examines how the cost, performance, and quality metrics overcome the inherent shortcomings of Boolean patent search engines.

**Problem: Boolean Patent Search Methods**

In order to frame the importance of considering a Heuristic Boolean search process, the critical problems with Boolean searching that hope to be overcome must first be discussed.

Let’s take a look at the fundamentals of Boolean patent searching. Interestingly, the entire process of Boolean searching is pre-disposed to only deliver the results you unwittingly ask for via your Boolean search query string - the computer simply returns the documents that match what you ask for. Clever Boolean search strategies, along with the application of proximity or truncation enhancements may occasionally produce some unanticipated results, but the results still literally correlate to your keyword request. The unanticipated results may or may not have any relevance to the subject matter being searched.

**Problem 1:** Whether or not a searcher is a subject matter expert or skilled research professional, they cannot reasonably craft a search string that examines all relevant documents in a patent database.

The detail of a Boolean search process includes, to one degree or another, follows this general sequence:

1) Review of the subject matter to be researched, and the development of search keywords or keyword strings that may find responsive documents,

2) Development of a search strategy consistent with the commercial parameters (cost, time, thoroughness, quality, or other directed metrics),

3) Executing the search - and conducting iterative combinations of planned keywords or keyword strings,

4) Obtaining a results “hits list” of sufficient breadth so as to contain documents of interest. This hits list may contain tens of thousands of documents of equally-weighted relevancy (all hits equally satisfy the literal Boolean request).

5) Narrowing of search results to a smaller hits list that can be examined by the researcher. The narrowing process requires the addition of more keywords or additional Boolean operators, filters or limiters.

6) Compiling a smaller, final search results list suitable to incorporate in a final search report.

**Problem 2:** Vagaries of Selection of Words, Phrases, and Boolean Operators

Since a Boolean engine will only return documents that the researcher asks for, the problem is not what patents match their keyword query, but rather what highly relevant patents are missed because the researcher did not use different words in the search query.

Given the more than 200,000,000 pages of patent text, it’s unlikely that any researcher would know all of the possible words, word sequence, or word combinations to use to discover all patents relevant to the search. This problem is compounded by patent writers who intentionally “submarine” a patent application by using obscure words, or even inventing a new lexicon.

Notwithstanding these practically insurmountable language issues, the researcher will nevertheless craft long and complex queries in the hopes of discovering all relevant patents. But the addition of more words to a query amplifies the statistical probability of missing important patents - just the opposite of what logical thinking would suggest.

For instance, a researcher who identified 10 important keywords to incorporate into a search strategy would need to perform a number of search iterations approaching 10 factorial (3,628,880 searches using all combinations of the keywords). Of course, this is unreasonable, as well as economically inefficient. This supports the thesis that a complete search can never evaluate the entire space [Grimson, 1991].
The researcher would end the search at a prescribed time (budget), or when they believed their search results list contained a sufficient number of relevant patents.

The end result is a list of documents that the researcher has carefully defined, or in other words, the researcher only sees the documents they have actually asked for. However, it’s important to understand that this process does not result in a list of relevant documents not requested by the researcher.

A paradox of Boolean searching is that as more keywords and Boolean operators are strung together as a complete “search strategy” to more accurately identify relevant patents, the more patents this process actually excludes from the list of relevant prior art.

Evidence suggests that as the keywords in a search string increase, the inaccuracies increase exponentially \[3\]. The case study saw a similar exponential increase in inaccuracies. “Inaccuracies” are defined as relevant patents that were eliminated from the search results list because of non-compliance with a larger, more restrictive Boolean query string.

This phenomenon nearly borders on an indictment of Boolean search methodology, and most professional researchers are either unaware of it, or ignore it as an inherent drawback of the process over which they have no control - but must use nevertheless.

---

\[3\] The Cost of Choosing the Wrong Model in Object Recognition by Constrained Search, W. ERIC L. GRIMSON, MIT Artificial Intelligence Laboratory, 545 Technology Square, Cambridge, MA 02139 Received May 22, 1990. Revised September 27, 1991.
Problem 3: Missing Data or Critical Patent Data Errors

The US Patent and Trademark Office indicates that “a particularly relevant document not identified by one search strategy but identified by another can be deemed a ‘critical error’ for performance and/or quality review”.

Because of errors encountered in the digital scanning and conversion to searchable text using the OCR process, many (electronic) patent documents contained in USPTO database have missing claims and other inaccurate, misplaced, or missing data. The USPTO makes its best effort to ensure accuracy of the Full-Text database, but that database is not the official electronic record. Other patent issuing authorities have similar problems (EPO, WIPO, and so forth), and this poor quality data is actually the same data distributed to its commercial patent data customers.

A Boolean search that requires certain words to appear in specific sections of the patent document will miss finding patents for which the data is missing (e.g.: if the word “device” must appear in the claims text, responsive patents that contain no claims text will not be included in the search results list). This is considered a critical error.

Patent data quality that meets 4-Sigma or 5-Sigma quality standards is unsatisfactory when considering the economic investment or risk tied to finding all relevant patents. Patent data meeting 4-Sigma will have up to 6,200 errors per 1,000,000 operations. 5-Sigma can have up to 233 errors in 1,000,000 operations. A database of 3,000,000 patents will have a statistical error volume of 600 to 18,000 errors. There is a high probability of missing one of these patent documents with a Boolean query - and the probability increases exponentially as more words and operators are added to the Boolean search string.

Although at least one word is required to initiate a Boolean query, the probability of encountering missing data (and missing relevant patents) suggest that fewer Boolean filters will find more relevant patents.

The obvious drawback to using fewer words is that the results set is extraordinarily large - perhaps 100,000 or more responsive documents. Reading all 100,000 documents is not economically viable - so what process exists wherein fewer Boolean keywords can be used in the query, yet allow the researcher to view the most relevant patents?

There are other conditions in which Boolean-only searching will fail to discover relevant patent documents. They are less likely to be overcome by the use of Heuristics and are therefore not addressed in this paper.

A Heuristic is a particular technique of directing one’s attention in learning, discovery, or problem-solving, otherwise known as an expert system.

Relating to or using a problem-solving technique in which the most appropriate solution of several is found by alternative methods, heuristics are applied at successive stages of a program for use in the next step of the program.

Batali explains that the word “heuristic” is not used only to describe cases where a solution might not be found, but to describe cases where we want to find the best solution (according to some way to measure bestness). The measure of “bestness”, and the assessment of a heuristic technique, is going to be relative to the domain, and to the specific job that problem solving is going to be applied to in that domain.

An expert system, also known as a knowledge based system, is a computer program that contains some of the subject-specific knowledge approximating the equivalency of a human subject matter expert assuming it trained on the subject matter. This class of program was first developed by researchers in artificial intelligence during the 1960s and 1970s and applied commercially throughout the 1980s.

The most common form of expert systems is a program made up of a set of rules that analyze information (in the current analysis, supplied by a Semantic database) about a specific class of problems, as well as providing analysis of the problem(s), and, depending upon their design, recommend a course of user action in order to implement corrections. It is a system that utilizes reasoning capabilities to reach conclusions.

As it relates to patent searching, heuristics call on an expert system as an alternative to “Boolean-only” searching. Applied to a Boolean search, heuristics assist a researcher in most efficiently reducing the results set to identify the most relevant documents responsive to a search - without adding more keywords that increase the number of missed-but-relevant patents.

What constitutes and expert system? In the patent search world, it is a database that broadens, yet qualifies the patent search results. Expert Systems currently used in association with patent searching include:

- Thesauri or synonym database
- Semantic Analysis Concept Space / Artificial Intelligence

In briefly comparing these two expert systems, we can say that the Thesauri or synonym database relies on a human to create a lookup table of words, then match those words to related words. The reliability of the lookup table will depend on the attentiveness of the person maintaining the database. Unfortunately, patent writers are permitted to invent new words to describe their invention (lexicon), and use these words to intentionally frustrate discovery by Boolean search engines. The person maintaining a synonym database will not even know what words they don’t know, so there is always a probability that even a thesaurus or synonym lookup table will be incomplete.

On the other hand, semantic engines incorporate artificial intelligence that actually learns a new lexicon on the fly. It has been proven to be more knowledgeable about documents that are closely related to the search query, even though none of the important keywords in the semantic query are contained in the most relevant documents.

The “problems” to be solved in the patent research process are many, depending on the researcher’s objectives. Heuristics can address each of these problems, although some problems will be better solved than others. The problems, defined at the highest level, include:

1) Obtaining the highest quality search results (not missing relevant patents because of search process limitations),
2) Completing the research project while consuming less than the total available resources (time, cost).

Heuristics allow the researcher to obtain the Best-First Search results - a key process component to capturing the maximum number of relevant documents, while ensuring that the fewest number of relevant documents are missed. As we have discussed in the previous section, adding keywords to a complex Boolean search query increases exponentially the number of documents that will be missed by the researcher.

Conversely, using shorter Boolean search queries, comprised of fewer keywords, widens the potential search results “hits”. A larger results set significantly increases the capture rate of relevant documents. (Refer to Fig. 1. above)

As most patent researchers know, a search that results in 10s of 100s of thousands of hits is quite useless. That is, unless an expert system is applied to the results set to bring the most relevant patent documents to the top of the results list.

Pantros IP’s patent database has been indexed using a Latent Semantic Analysis (“LSA”) search engine. Through the indexing process, the LSA engine has learned more than 2.5 million distinct invention concepts, and has mapped every patent against these concepts. The mapping involves the development of N-Dimensional Vectors for each concept - some vectors being long with many variants of the concept positioned along its length, with other vectors being shorter to express an invention concept containing fewer variants. The index is best defined as a database “Concept Space” containing only mathematical expressions of each concept, and no human-readable words (as are found in traditional patent databases).

Without expounding on the details of LSA or Pantros IP’s Semantic database\(^7\), suffice to say that the LSA database, when applied to a Boolean search process, serves as the trained *expert system*.

Fig. 3. When compared to Fig. 1. above, shows an exponentially larger search results set when only 2 keywords (fewer keywords) are used in the Boolean query. Obviously, the number of “hits” can be extraordinarily large since very few Boolean filters were applied to the database.

Since all responsive documents are equally weighted, no patent stands out as more relevant to satisfying the researcher’s objective than any other.

This “problem” has been historically solved by continuing with a sequence of search iterations, each iteration beginning with an expanded Boolean search string. Each iteration defines more restrictive Boolean filters that results in fewer hits - until the number of patent documents is small enough so that the research can begin the manual review of the documents to identify those most relevant.

The application of heuristics results in an organization of the Boolean “hits” based on relevance to the Semantic query. Of course, all of the hits also satisfy the literal Boolean restrictions as well.

This heuristic approach is illustrated in the following Fig. 4.

---

In this Case Study, the number of hits from the first search result exceeded 14,000 (only two Boolean keywords were used). Contained in the results set were highly relevant, somewhat relevant, and the expected inclusion of a large number of “junk” patent documents. A results list of 14,000 documents is, of course, too large for the researcher to begin manually viewing and qualifying the more appropriate patents - so the researcher would begin the iterative process of expanding the Boolean search strategy by adding more keywords of other restrictions.

Following the application of heuristics to the first results set as outlined above in Fig. 4., the process shows the speed and efficiency in narrowing the results to a final prior art list.

While it’s shown that heuristics increases the overall quality of the search results, the process also shows that even with heuristics, the increase in the number of Boolean keywords in each successive iteration still results in some relevant documents being dropped from the final list, albeit fewer drops than with Boolean alone.

**Case Study**

In the following case study, one can readily see that the application of LSA Heuristics found the most relevant documents and brought them to the top of the 14,000 hits, even when only a small number of keyword terms were used.

The problems solved with the addition of heuristics were:

A. **Faster search**: by percolating the most relevant documents to the top, the researcher found the best documents within minutes. The Best-First Search achieved results that otherwise could have literally taken hours,

B. **Reduced number of missed patents** that were relevant to the search: by using fewer keywords (shorter Boolean query), more documents were included in the search results that would have otherwise been eliminated by a more extensive Boolean query.

But even with the additional heuristic qualifiers, some relevant patents were eliminated from the results set each time another Boolean keyword filter was introduced. Nevertheless, heuristics reduced the number of good patents that were eliminated by keeping the Boolean queries very short, and by reducing the total number of iterations.

**Case Study Objective**: The task called for the executing of a search process to identify the most relevant patents that would constitute a Prior Art objection to the patentability of an invention related to an improved nozzle design for a digital ink jet printer. This required the researcher to:

A) Develop a natural language search query (heuristic) that will be applied to the Semantic database. This description defines the actual invention, and constitutes the instructions for the heuristic process. The query used was:

> The improved nozzle geometry of a digital ink jet printer, such nozzle improvement providing for a more precise control of the dispersion pattern of the liquid ink. This improved control of the dispersion spray pattern provides for the printing of high resolution graphics, specifically including high definition photographic prints.

B) Develop a keyword list that could be used in the crafting of the Boolean search strategy. The large number of possible words contained in the relevant documents were assessed, and the keywords that were simple and most obvious were selected as the starting point. The objective was to obtain the Best-First Search results list with the fewest number of relevant documents being eliminated from the results set.

C) Conduct the search, initially using only two broad keywords + the heuristic query, and refine the results by expanding the Boolean query string with additional keywords.

D) Refine the results list by adding another keyword to the Boolean query during each iteration.
Case Study Details:

The following search was conducted to illustrate the correlation between the increase in the number of keywords used in a Boolean search string with the rate at which highly relevant patents are dropped from the list or search results.

**Keywords considered:** A partial list of other relevant keywords discovered during the case study, but which were not used in the search process.

- diameter
- deposit dots
- orifice
- ejection stream
- compensation
- output device
- sharp edge
- dot spread
- volume
- pixel printing
- nozzle bore
- chimney
- shamfered edge
- vertical banding
- deflection angle
- venture spittoon
- orificeless
- shaped
- channel
- inkdrop generator
- spitting
- architecture
- pixel printing
- shaped
- ejection stream
- orificeless
- volume
- pixel printing
- channel
- compensation
- output device
- chimney
- vertical banding
- deflection angle
- shape
- ejection stream
- orificeless
- volume
- pixel printing
- compensation
- output device
- chimney
- vertical banding
- deflection angle
- shape
- ejection stream
- orificeless
- volume
- pixel printing
- compensation
- output device
- chimney
- vertical banding
- deflection angle
- shape
- ejection stream
- orificeless
- volume
- pixel printing
- compensation
- output device
- chimney

*Examination of all possible keyword combinations would be $N$ Factorial variations (worst case).

**Search Time - 1 hour, approximately:** for purposes of the case study, the 48 hits shown in Column F. were obtained in about one hour. A researcher would have taken considerably more time in qualifying the appropriate keywords and Boolean search string, but the one hour to conduct the case study search resulted in a good reportable results set.

**Patents Reviewed:** The SuperBoolean search allowed us to focus only on the top 100 results at each results stage, those patents in the results list being based on relevancy to the Heuristic Query.

**Patents Eliminated:** Upon obtaining a new results list / column with each additional keyword, those patents that were dropped from the current column (when compared to the preceding column) were highlighted in green. All eliminated patents were then individually examined to determine whether they were actually relevant to the search. Dropped, but still relevant patents are shown with a red patent number.

**Case Study Search Results**
<table>
<thead>
<tr>
<th>Baseline NO.</th>
<th>Total US Database</th>
<th>A. (2 words)</th>
<th>B. (3 words)</th>
<th>C. (4 words)</th>
<th>D. (5 words)</th>
<th>E. (6 words)</th>
<th>F. (7 words)</th>
</tr>
</thead>
<tbody>
<tr>
<td>US Pat #</td>
<td>% US Pat #</td>
<td>% US Pat #</td>
<td>% US Pat #</td>
<td>% US Pat #</td>
<td>% US Pat #</td>
<td>% US Pat #</td>
<td></td>
</tr>
<tr>
<td>4010916</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td></td>
</tr>
<tr>
<td>7066564</td>
<td>99.6</td>
<td>99.6</td>
<td>99.6</td>
<td>99.9</td>
<td>99.8</td>
<td>98.7</td>
<td></td>
</tr>
<tr>
<td>4380017</td>
<td>97.0</td>
<td>97.7</td>
<td>97.7</td>
<td>99.3</td>
<td>99.3</td>
<td>98.2</td>
<td></td>
</tr>
<tr>
<td>6612685</td>
<td>97.1</td>
<td>97.1</td>
<td>98.0</td>
<td>98.0</td>
<td>98.0</td>
<td>98.0</td>
<td></td>
</tr>
<tr>
<td>4196006</td>
<td>96.3</td>
<td>96.3</td>
<td>97.7</td>
<td>97.5</td>
<td>97.5</td>
<td>98.3</td>
<td></td>
</tr>
<tr>
<td>6056822</td>
<td>95.5</td>
<td>95.5</td>
<td>95.5</td>
<td>95.7</td>
<td>95.7</td>
<td>95.7</td>
<td></td>
</tr>
<tr>
<td>4967208</td>
<td>94.8</td>
<td>94.8</td>
<td>94.8</td>
<td>94.5</td>
<td>94.5</td>
<td>94.5</td>
<td></td>
</tr>
<tr>
<td>4429315</td>
<td>94.5</td>
<td>94.5</td>
<td>94.5</td>
<td>94.4</td>
<td>94.4</td>
<td>94.4</td>
<td></td>
</tr>
<tr>
<td>6332662</td>
<td>94.3</td>
<td>94.3</td>
<td>94.3</td>
<td>94.2</td>
<td>94.2</td>
<td>94.0</td>
<td></td>
</tr>
<tr>
<td>6075699</td>
<td>94.0</td>
<td>94.0</td>
<td>94.0</td>
<td>94.0</td>
<td>94.0</td>
<td>94.0</td>
<td></td>
</tr>
<tr>
<td>4184881</td>
<td>93.9</td>
<td>93.9</td>
<td>93.9</td>
<td>93.9</td>
<td>93.9</td>
<td>93.9</td>
<td></td>
</tr>
<tr>
<td>6561609</td>
<td>93.9</td>
<td>93.9</td>
<td>93.9</td>
<td>93.9</td>
<td>93.9</td>
<td>93.9</td>
<td></td>
</tr>
<tr>
<td>6858334</td>
<td>93.4</td>
<td>93.4</td>
<td>93.4</td>
<td>93.4</td>
<td>93.4</td>
<td>93.4</td>
<td></td>
</tr>
<tr>
<td>6038834</td>
<td>93.3</td>
<td>93.3</td>
<td>93.3</td>
<td>93.3</td>
<td>93.3</td>
<td>93.3</td>
<td></td>
</tr>
<tr>
<td>6241333</td>
<td>92.6</td>
<td>92.6</td>
<td>92.6</td>
<td>92.6</td>
<td>92.6</td>
<td>92.6</td>
<td></td>
</tr>
<tr>
<td>6881788</td>
<td>92.0</td>
<td>92.0</td>
<td>92.0</td>
<td>92.0</td>
<td>92.0</td>
<td>92.0</td>
<td></td>
</tr>
<tr>
<td>6421559</td>
<td>91.8</td>
<td>91.8</td>
<td>91.8</td>
<td>91.8</td>
<td>91.8</td>
<td>91.8</td>
<td></td>
</tr>
<tr>
<td>6561800</td>
<td>91.4</td>
<td>91.4</td>
<td>91.4</td>
<td>91.4</td>
<td>91.4</td>
<td>91.4</td>
<td></td>
</tr>
<tr>
<td>6441556</td>
<td>91.2</td>
<td>91.2</td>
<td>91.2</td>
<td>91.2</td>
<td>91.2</td>
<td>91.2</td>
<td></td>
</tr>
<tr>
<td>6056075</td>
<td>91.2</td>
<td>91.2</td>
<td>91.2</td>
<td>91.2</td>
<td>91.2</td>
<td>91.2</td>
<td></td>
</tr>
<tr>
<td>6869795</td>
<td>91.1</td>
<td>91.1</td>
<td>91.1</td>
<td>91.1</td>
<td>91.1</td>
<td>91.1</td>
<td></td>
</tr>
<tr>
<td>6785644</td>
<td>90.9</td>
<td>90.9</td>
<td>90.9</td>
<td>90.9</td>
<td>90.9</td>
<td>90.9</td>
<td></td>
</tr>
<tr>
<td>6524744</td>
<td>90.8</td>
<td>90.8</td>
<td>90.8</td>
<td>90.8</td>
<td>90.8</td>
<td>90.8</td>
<td></td>
</tr>
<tr>
<td>6585950</td>
<td>90.7</td>
<td>90.7</td>
<td>90.7</td>
<td>90.7</td>
<td>90.7</td>
<td>90.7</td>
<td></td>
</tr>
<tr>
<td>6461583</td>
<td>90.6</td>
<td>90.6</td>
<td>90.6</td>
<td>90.6</td>
<td>90.6</td>
<td>90.6</td>
<td></td>
</tr>
<tr>
<td>6357952</td>
<td>90.5</td>
<td>90.5</td>
<td>90.5</td>
<td>90.5</td>
<td>90.5</td>
<td>90.5</td>
<td></td>
</tr>
<tr>
<td>6244687</td>
<td>90.5</td>
<td>90.5</td>
<td>90.5</td>
<td>90.5</td>
<td>90.5</td>
<td>90.5</td>
<td></td>
</tr>
</tbody>
</table>

© 2006 – 2012 All Rights Reserved. PANTROS IP
<table>
<thead>
<tr>
<th>Patent Numbers</th>
<th>Percent Eliminated</th>
</tr>
</thead>
<tbody>
<tr>
<td>6502912</td>
<td>90.5</td>
</tr>
<tr>
<td>6705702</td>
<td>90.4</td>
</tr>
<tr>
<td>6854829</td>
<td>90.4</td>
</tr>
<tr>
<td>6431704</td>
<td>90.4</td>
</tr>
<tr>
<td>6244687</td>
<td>90.3</td>
</tr>
<tr>
<td>7004571</td>
<td>90.2</td>
</tr>
<tr>
<td>6559621</td>
<td>90.1</td>
</tr>
<tr>
<td>6830320</td>
<td>90.0</td>
</tr>
<tr>
<td>6431704</td>
<td>89.9</td>
</tr>
<tr>
<td>5724079</td>
<td>89.9</td>
</tr>
<tr>
<td>6764744</td>
<td>89.8</td>
</tr>
<tr>
<td>6354689</td>
<td>89.8</td>
</tr>
<tr>
<td>6811565</td>
<td>89.7</td>
</tr>
<tr>
<td>6905191</td>
<td>89.0</td>
</tr>
<tr>
<td>6992828</td>
<td>89.0</td>
</tr>
<tr>
<td>6595621</td>
<td>88.9</td>
</tr>
<tr>
<td>6629752</td>
<td>88.8</td>
</tr>
<tr>
<td>6742868</td>
<td>88.7</td>
</tr>
<tr>
<td>6902274</td>
<td>88.5</td>
</tr>
<tr>
<td>6187082</td>
<td>88.3</td>
</tr>
<tr>
<td>6244687</td>
<td>88.2</td>
</tr>
<tr>
<td>7004571</td>
<td>88.0</td>
</tr>
<tr>
<td>6559621</td>
<td>87.9</td>
</tr>
<tr>
<td>6830320</td>
<td>87.8</td>
</tr>
<tr>
<td>6431704</td>
<td>87.8</td>
</tr>
<tr>
<td>5724079</td>
<td>87.6</td>
</tr>
<tr>
<td>6764744</td>
<td>87.4</td>
</tr>
<tr>
<td>6354689</td>
<td>87.3</td>
</tr>
<tr>
<td>6811565</td>
<td>87.2</td>
</tr>
<tr>
<td>6905191</td>
<td>87.0</td>
</tr>
<tr>
<td>6992828</td>
<td>87.0</td>
</tr>
<tr>
<td>6595621</td>
<td>86.9</td>
</tr>
<tr>
<td>6830320</td>
<td>86.7</td>
</tr>
<tr>
<td>6431704</td>
<td>86.5</td>
</tr>
<tr>
<td>5724079</td>
<td>86.4</td>
</tr>
<tr>
<td>6764744</td>
<td>86.2</td>
</tr>
<tr>
<td>6354689</td>
<td>86.1</td>
</tr>
<tr>
<td>6811565</td>
<td>85.9</td>
</tr>
<tr>
<td>6905191</td>
<td>85.8</td>
</tr>
<tr>
<td>6992828</td>
<td>85.7</td>
</tr>
<tr>
<td>6595621</td>
<td>85.5</td>
</tr>
<tr>
<td>6830320</td>
<td>85.4</td>
</tr>
<tr>
<td>6431704</td>
<td>85.2</td>
</tr>
</tbody>
</table>

Legend and Keyword Results Summary correlating to additional keywords appended to a Boolean query string:

- **Top 100 Heuristic Only**
- **Top 100 Boolean**
- **8% eliminated with 4 keywords**
- **2% eliminated with 3 keywords**
- **2% eliminated with 2 keywords**
- **4 relevant patents missed / dropped between A&B**
- **11 relevant patents missed / dropped between B&C**
- **9 relevant patents missed / dropped between C&D**
- **13 relevant patents missed / dropped between D&E**

© 2006 – 2012 All Rights Reserved. Pantros IP
Conclusion

We’ve acknowledged the various problems with the Boolean search process in general, and more specifically related these problems to the inordinately high legal and financial risks associated with patent documents.

The demands to perform a patent search that attempts to identify all of the relevant documents within the scope of available resources (time, budget, computing time, a given patent data quality) keep researchers reliant on the time-honored practice of crafting a lengthy, complex Boolean search string. But it’s been shown that such restrictions, although they produce relevant patents in a final results list, more dangerously drop an increasing number of relevant patents that should have been included in the final search report.

The application of heuristics, namely a synonym lookup table, or more preferably a Semantic / artificial intelligence expert system that allows the researcher to use a less restrictive Boolean query, results in the Best-First search results list that positions the most relevant documents at the top. Researchers are then able to manage very large search results lists without filtering the list to a more manageable quantity by using more keywords.

The results of applying heuristics to Boolean patent searching are faster time to identify the most relevant patents, but more importantly, the identification of the largest number of relevant patents that will serve as acceptable prior art.