Interfaces
Vol. 45, No. 3, May–June 2015, pp. 271–277
ISSN 0092-2102 (print) | ISSN 1526-551X (online)
http://dx.doi.org/10.1287/inte.2014.0775

ASP, The Art and Science of Practice: How Analytics Practitioners Can Learn from Published Patents and Protect Their Work

R. John Milne
School of Business, Clarkson University, Potsdam, New York 13699, jmilne@clarkson.edu

Brian T. Denton
Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, Michigan 48109, btdenton@umich.edu

Troy White
School of Business, Clarkson University, Potsdam, New York 13699, whitet@clarkson.edu

This paper examines the high and growing rate at which the U.S. Patent and Trademark Office granted analytics patents from 2002 to 2013. We summarize examples of analytics patents and provide advice to analytics professionals on searching and reading patents and protecting their work.

Keywords: analytics patents; operations research patents; software patents.
History: This paper was refereed. Published online in *Articles in Advance* February 13, 2015.

When a company develops a valuable analytics method, it has limited alternatives for protecting its intellectual property (IP). It can take precautions to keep the method secret and hope that nobody else independently creates the method; it can copyright the underlying software; or it can patent the method. Copyrights protect the expressions of ideas (e.g., prevent copying of source code); however, they do not protect the concepts (e.g., they do not prevent copying the algorithm). For most methods, patenting is the strongest form of IP protection, and also the most expensive in terms of employee time, legal fees, and patent-filing and maintenance fees paid to government(s). Obtaining a patent in a single country often results in out-of-pocket expenses into the five figures (U.S. dollars). Extending the protection to many countries can lead to six-figure expenses. Because of the time and expense, companies patent only their most valuable analytics IP, and many inventions are patented in only a single country.

A patent provides its owner (assignee) with the right to stop others from using an invention without permission for a long period—20 years after patent application filing in the United States. This can provide companies with a competitive advantage in the marketplace and licensing opportunities. Patents can also be used as a defensive weapon: if a competing firm attacks with a patent lawsuit, the defending company may use its own patent portfolio to countersue.

According to U.S. law, a patent must be useful, novel, and nonobvious in light of prior patents and publications. Furthermore, a patent application must describe the invention in enough detail for a professional of "ordinary skill in the art" (U.S. Patent and Trade Office 2014) to implement the invention. Patent applications become publically available, thus enabling others to learn from them. In exchange, the government provides assignees with the long-term rights to exclude others from using their inventions. Patent applications must be filed prior to publication, public use, or sales activity of the invention or—depending on the country—within one year of these events. Accordingly, patents describe not only innovative work, but also recent work.

Murphy (2002) shared his observations from searching U.S. patents by operations research (OR) keywords from 1996 through January 2002. Our paper updates and extends his work by including additional keyword expressions, and identifying the growth of these patent issuances during the past 12 years. The broader scope of our search is intended to be consistent with the broader domain associated with the emerging definition of analytics as a field of expertise that subsumes OR. In addition to presenting and interpreting the search results, we provide insights and advice on patent development and reading and searching patents based on our experiences co-inventing more than three dozen analytics patents.

Demographics of Analytics Patents

We used keywords pertaining to analytics to search the U.S. Patent and Trademark Office (USPTO) website (www.uspto.gov). We use the same OR keywords as Murphy (2002); in addition, we add keywords to incorporate the broader range of methods associated with the INFORMS definition of analytics—"The scientific process of transforming data into insight for making better decisions" (INFORMS 2014). The original keywords that Murphy used are boldfaced in Tables 1–5. Our search covers patents granted

Descriptive analytics	Number of U.S. patents issued in three-year periods				Change from prior period (%)		
Keyword expression	2002–2004	2005–2007	2008–2010	2011–2013	2005–2007	2008–2010	2011–2013
Data mining	790	1,564	2,799	5,175	98	79	85
Clustering algorithm	269	370	660	1,224	38	78	85
Discriminant analysis	248	314	502	872	27	60	74
Data visualization	155	240	344	709	55	43	106
Learning curve	299	276	402	552	−8	46	37
Online analytical processing	97	119	238	327	23	100	37
Exploratory data analysis	34	35	50	58	3	43	16

Table 1: This table shows the number of U.S. patents issued during three-year periods for descriptive analytics patents, and growth rates from each prior period. Keyword expressions used by Murphy (2002) have boldface type.

Predictive analytics	Number of U.S. patents issued in three-year periods				Change from prior period (%)		
Keyword expression	2002–2004	2005–2007	2008–2010	2011–2013	2005–2007	2008–2010	2011–2013
Simulation	16,036	20,260	26,429	39,909	26	30	51
Queue	8,704	11,130	14,989	21,168	28	35	41
Neural network	2,208	2,584	3,687	5,512	17	43	49
Machine learning	410	809	2,172	4,882	97	168	125
Computer simulation	1,640	1,964	1,976	2,643	20	1	34
Principal component analysis	462	666	1,043	1,739	44	57	67
Collaborative filtering	144	398	682	1,141	176	71	67
Stochastic simulation or Monte Carlo simulation	233	386	676	1,048	66	75	55
ARIMA*	396	465	574	784	17	23	37
Markov chain	80	167	275	478	109	65	74
Stochastic process	108	185	261	346	71	41	33
Waiting line	128	124	195	261	−3	57	34
Markov process	101	122	162	254	21	33	57
Renewal process	47	68	99	145	45	46	46
Bayesian models	8	25	79	120	213	216	52
Multinomial models	1	2	9	6	100	350	−33

Table 2: This table shows the number of U.S. patents issued during three-year periods for predictive analytics patents. Growth rates from each prior period are also shown. Keyword expressions used by Murphy (2002) have boldface type.

*ARIMA: Autoregressive integrated moving average.

Prescriptive analytics	Number of U.S. patents issued in three-year periods				Change from prior period (%)		
Keyword expression	2002–2004	2005–2007	2008–2010	2011–2013	2005–2007	2008–2010	2011–2013
Optimization	18,808	23,069	31,164	49,522	23	35	59
Combinatorial and optimization	1,283	1,598	2,294	3,374	25	44	47
Shortest path	1,127	1,537	2,313	3,103	36	50	34
Decision tree	573	799	1,215	2,023	39	52	67
Maximum flow	1,210	1,182	1,227	1,607	−2	4	31
Network model	536	702	995	1,527	31	42	53
Genetic algorithm or evolutionary algorithm	432	609	946	1,470	41	55	55
Dynamic programming	425	595	931	1,465	40	56	57
Linear programming or linear optimization	375	499	931	1,497	33	87	61
Simulated annealing	446	614	856	1,135	38	39	33
Local search	128	174	396	825	36	128	108
Network flow	204	262	451	659	28	72	46
Nonlinear programming or nonlinear optimization	138	187	286	448	36	53	57
Combinatorial optimization	95	132	266	350	39	102	32
Integer programming or integer optimization	86	106	214	383	23	102	79
Lagrange multiplier	90	110	198	275	22	80	39
Game theory	9	49	120	263	444	145	119
Stochastic programming or stochastic optimization	33	49	80	154	48	63	93
Minimum spanning tree	55	114	86	118	107	−25	37
Tabu search	24	73	73	136	204	0	86
Interior point and integer	32	57	68	86	78	19	26
"Branch and bound" and integer	21	40	43	85	90	8	98
Neighborhood search	21	20	27	72	−5	35	167
Cutting plane and integer	26	23	19	59	−12	−17	211
Goal programming	6	19	34	46	217	79	35
Ant colony optimization	0	5	17	56		240	229
Column generation	15	19	25	39	27	32	56
("Branch and cut" or branch-and-cut) and integer	8	4	12	28	−50	200	133
Data envelopment analysis	0	1	6	7		500	17
("Branch and price" or branch-and-price) and integer	1	2	0	9	100	−100	
Multiattribute utility theory	0	0	1	1			0

Table 3: This table shows the number of U.S. patents issued during three-year periods and growth rates for prescriptive analytics patents.

Application areas	Number of U.S. patents issued in three-year periods				Change from prior period (%)		
Keyword expression	2002–2004	2005–2007	2008–2010	2011–2013	2005–2007	2008–2010	2011–2013
Scheduling	6,958	9,798	14,667	26,436	41	50	80
Supply chain	245	853	1,826	2,834	248	114	55
Inventory management	481	743	1,184	1,996	54	59	69
Risk management	175	386	1,073	1,857	121	178	73
Portfolio management	61	127	399	580	108	214	45
Revenue management or yield management	88	160	231	350	82	44	52
Vehicle routing	53	65	48	99	23	−26	106
Reliability model	33	32	50	83	−3	56	66
Text analytics	0	0	24	78			225

Table 4: This table shows the number of U.S. patents issued during three-year periods and growth rates for a sample of application areas associated with analytics.

(issued) in four three-year periods from January 1, 2002 through December 31, 2013. In the appendix, we describe the details of executing the keyword searches.

According to World Intellectual Property Organization (2014), in about half of all U.S. patents granted in this time frame, the inventor listed first on the patent resides in the United States. Foreign inventors and their companies are motivated to file patent applications in the United States primarily because of their desire to compete in U.S. markets and with

Other keywords	Number of U.S. patents issued in three-year periods				Change from prior period (%)		
Keyword expression	2002–2004	2005–2007	2008–2010	2011–2013	2005–2007	2008–2010	2011–2013
Artificial intelligence	1,521	2,142	3,825	6,456	41	79	69
Analytics	171	302	976	3,264	77	223	234
Expert system	924	1,011	1,190	1,615	9	18	36
Business intelligence	90	196	555	1,218	118	183	119
Unstructured data	88	168	430	870	91	156	102
MapReduce	0	0	24	230			858
Business analytics	0	10	31	101		210	226
Big data	17	16	33	66	−6	106	100

Table 5: This table shows the number of U.S. patents issued during three-year periods for other keywords associated with analytics. MapReduce, analytics, and business analytics have particularly high growth rates.

competitors based in the United States. Most analytics patents fall into the category of software patents.

Tables 1–5 show the frequency of patents issued by keyword expression for each of the four periods and the percentage changes in frequency from the prior periods. Tables 1–3 contain keywords that are most commonly associated with descriptive, predictive, and prescriptive analytics, respectively. As Schneur (2011) notes, these analytics categories use data to provide decision-making insights into what happened (descriptive), what is expected to happen (predictive), and what should happen (prescriptive). Table 4 contains keywords associated with a sample of analytics applications. Table 5 contains other keywords associated with analytics, namely those that do not fit neatly into any category in Tables 1–4. The rows in the tables are sorted in descending sequence of a weighted average of the frequency counts of the four periods. The weights are 0.5, 0.25, 0.15, and 0.10, with the more recent periods having higher weights.

Some high-frequency counts, such as those for data mining, simulation, queue, and optimization, reflect the nonspecific nature of the keyword expression. For example, of the 10 latest patents to issue in 2013 containing optimization, all use the term optimization to refer to the general notion of doing something effectively rather than the more specific concept of a mathematical programming model with decision variables that maximize (or minimize) an objective function, subject to constraints. Yet, it is clear that mathematical optimization is commonly used, as indicated by dynamic programming and linear programming and (or) linear optimization, each having nearly 1,500 patents in the most recent three-year period. The tables reveal a strong and growing presence of patents containing analytics keywords.

Examples of Analytics Patents

Budiman et al. (2013) of Exxon Mobil Research describe the optimization of their company's blending and packaging facilities. Exxon's recent practice had been to make long- and medium-term capacity decisions and feed the resulting capacities as constraints to its production planning and inventory models. In this patent, the inventors describe an integrated model that jointly determines plans for capacities, production, and inventory levels. The language suggests an Excel interface for data entry, AIIMS for modeling, and CPLEX for solving, using linear approximations and iterations with cutting planes. When a patent, such as this one from Exxon, refers to a "preferred embodiment," one cannot be sure whether the company is using the method or planning to use the method; there is no legal requirement of either. Typically, a preferred embodiment is the best way to solve the problem; however, alternative embodiments may be described to protect against others using similarly envisioned solutions without permission.

Sternickel et al. (2014) of CardioMag Imaging use machine learning to classify magnetic fields emitted by the electrophysiological activity of the patient's heart. They use direct kernel partial least squares and kernel ridge regression for supervised learning, and self-organizing maps for unsupervised learning. They assert the results of their methods "exceed … the quality of classification achieved by the trained experts."

Brukman et al. (2013) of Google personalize search engine results based on a user's profile. Given a sorted list of search engine results, items are moved up or down the list based on the user's search history, including clicks, time spent on Web pages, and lower weighting of older searches. The magnitude of an item's movement up (or down) the list depends on factors such as the item's popularity and the probability of a long visit to the website.

Meucci (2013) of Barclays Capital characterizes financial securities' risk factors as random variables, identifies best-fit distributions, and generates return distributions for securities and ultimately for the specified investment portfolios. Meucci commonly finds that a fat-tailed t-distribution best describes a factor's historical behavior. Monte Carlo simulations from the risk factors' distributions are used to estimate distributions for the portfolios.

Zhang et al. (2013) of Hewlett-Packard use K-means and K-means regression clustering to facilitate efficient parallel computing within a multicore processor.

Miller (2010) of Gridpoint describes linear programming and integer programming models to determine when an energy consumer should purchase electricity from the grid to store in batteries for later use. Key factors include the battery capacity and the anticipated time-varying prices and consumer demand. Other than the consumer energy context, this modeling will be straightforward for those familiar with linear programming and integer programming models used in production planning. Accordingly, this may be an example of a patent containing concepts that could be used to create exercises for students.

Searching and Reading Patents

In this paper, we used the U.S. Patent and Trademark website (www.uspto.gov), which is an original source and allows keyword-searching capabilities that work well for our purposes. However, for normal usage, we recommend using Google's patent-search website (www.google.com/patents) to search for patents and patent applications. This search engine is fast, and it provides a variety of search options, such as selecting patent office and issuance status; it also has user-friendly options for displaying patents and forwarding links to others. Other websites (e.g., www.delphion.com) also have good options for searching for patents; however, they often require an individual or group license fee.

A patent is both a technical and legal document. We offer the following approach to reading U.S. patents. The first page includes a title and abstract that help determine whether the patent is worth reading. The *date of patent* is the date the patent was granted, often two or more years after the patent application was filed. The filing date, the date the patent application was submitted to the patent office, is also on the first page of the patent, but in a smaller font. The figure on the first page is deemed by the patent office as the figure that most represents the essence of the invention; for many product patents, a glance at the figure on the first page provides a good overview of the invention. However, for software patents, at best, the figures provide an indication of the method. The first page also contains the assignee (i.e., the patent's owner at the time of patent issuance). In the United States, inventors are owners of their patented inventions by default, unless they have assigned their rights to others, typically their employers, as a condition of employment.

Some experts suggest that the current U.S. patent system, as defined by existing intellectual property law, can result in perverse incentives (Koenen and Peitz 2009, Bessen and Meurer 2012). Some inventors may disguise the precise nature of a patent through unusual nomenclature, or by having a third party as the nominal assignee. The incentives for such trickery include fooling competitors, a decreased likelihood of a patent examiner finding relevant prior art, and an increased likelihood of accidental infringement of a patent to obtain a monetary settlement.

It is not uncommon to encounter multiple patents with the same title, the same inventors, and the same technical content. This can happen when the patent office splits an application deemed to contain multiple inventions into multiple patents. In such cases, the assignee must pay fees for patenting each of the several inventions.

Following the figures is a section containing the background of the invention. The *Background* section can help the reader ascertain a partial understanding of the problem context. The *Claims* and *Summary* sections have legal importance and consequently tend to be written in a specific technical manner that often requires a patent lawyer to interpret the full meaning. To learn the algorithmic details of an analytics invention, read the *Detailed Description of the Invention*. This section may have a different title, such as *Description of a Preferred Embodiment*. It is supposed to contain the description of the invention in enough detail that professionals with ordinary skill in the area will understand how to implement the method. When reading the detailed description, refer to the figures as needed to understand the method. Number labels within the detailed description refer to specific elements of the figures.

In the United States, patent applications are published 18 months after the filing date. Some of these applications may not have yet been patented. Some applications may never become patents. This can happen when the patent office rejects the work's implied assertions of being novel and nonobvious. It can also happen when the assignee decides the anticipated value of the patent is no longer worth the continued expense and time of debating with the patent office and paying the eventual patent issuance fee. Prior to issuance, the assignee may have discovered a better way of solving the problem.

Patenting Analytics Work

The process of creating a patent involves describing your work in detail, reviewing how others have addressed the problem, differentiating your work from theirs, and establishing that the underlying ideas can be implemented in practice. The first step is to begin documenting the information required for patenting as early as you can in the analytics solution development process. A concise statement of the problem being solved and the attributes of an ideal solution will be helpful in justifying your design. Second, searching existing patents for how others have addressed the problem—and documenting the best prior solutions—is useful both for patenting and for providing ideas on solution approaches. Third, writing down essential steps of your method will be helpful prior to the creation of well-documented program code. We have typically found that the final patent is better if you do as much of the writing as you can the first time, before having a patent attorney or agent work on the patent. However, we have found it more efficient to let the patent attorney or agent write the initial drafts of the claims, summary, and technical field of invention of the patent application.

The claims of a patent define the boundaries of the intellectual property covered by the patent. For a software algorithm with seven essential steps, an initial draft of a first claim may contain the essence of three or four of those essential steps. The assignee of the patent wants the claims to be as broad as possible; however, to protect the public, the patent office wants the claims to be as narrow as fairly allowed. Months (or years) after the submission of your patent application, it is normal for a patent office examiner to initiate correspondence with your attorney challenging the breadth of the claims. The examiner does this through an *office action* (OA) indicating why the claims are not valid based on prior patents the examiner has found that bear similarities to elements of your invention. When you receive the OA, read the sections pertaining to the analytics aspects of the OA (leaving the patent attorney to handle any legal formalities) and write to your attorney. For each analytics item of the OA, explain to your attorney any reason why you think the patent examiner is mistaken or provide additional features of your work not suggested by the prior work that the examiner has identified. The attorney will use this information in drafting a response to the patent office. The attorney's response will often dispute some of the patent examiner's assertions, while conceding to some narrowing of the claims. After up to a few iterations of this debate, the patent will be granted or the examiner will issue a final OA rejecting the patent application, indicating the process should stop. In that event, further debate (and further narrowing of the claims) is possible through appeals or a request for continued examination. Along the way, your organization may decide that the claims have become so narrow that the patent is no longer worth the fight. More commonly, however, the patent issues within two or three iterations of debate with the examiner.

Concluding Remarks

The demographic analysis of patents can extend far beyond frequency assessment. Abbas et al. (2014) provide a literature review and taxonomy on text-mining and visualization approaches for patent analysis; the methods can help with forecasting technology trends, competitor analysis, detecting infringement, and identifying promising patents and patent vacuums. Avasarala et al. (2013) describe an interactive system for analyzing patent landscapes. Huang et al. (2013) take a broader view by examining patent performance and technology interactions of universities, industries, and governments. Ideas from Klavans and Boyack (2009) toward establishing a consensus map of science could be extended to patent analysis.

In some cases, the patent office grants a patent that it should have rejected as an obvious use of prior methods; some of these patents are invalidated in court (e.g., Supreme Court of the United States 2014). As a result, readers should approach the patent literature with some caution. Nevertheless, patents are a rich (and growing) source of information on leading-edge analytics practice.

Appendix. Details of Patent Searches

We conducted our patent searches by using the following Web page within the U.S. Patent and Trademark Office website: http://patft.uspto.gov/netahtml/PTO/search-adv.htm. This page may also be reached from the U.S. Patent and Trademark website (http://www.uspto.gov) by clicking on "patent search" and then clicking on the "advanced search" feature of issued patents.

As an example, corresponding to Table 2, we use the following search expression to find a frequency of 233 issued patents containing "stochastic simulation" or "Monte Carlo simulation" between 2002 (inclusive) and 2004 (inclusive):

(ISD/1/1/2002->12/31/2004) AND
(("stochastic simulation") or
("Monte Carlo simulation")).

References

Abbas A, Zhang L, Khan SU (2014) A literature review on the state-of-the-art in patent analysis. *World Patent Inform.* 37(June):3–13.

Avasarala V, Hershey JE, Varma A, Bonissone PP (2013) Semi-automated and interactive system and method for analyzing patent landscapes. U.S. Patent 8,412,659, filed May 11, 2009, issued April 2, 2013.

Bessen J, Meurer MJ (2012) The direct costs from NPE disputes. Working paper, Boston University School of Law, Boston, 12–34.

Brukman MY, Horling BC, Zamir OE (2013) Systems and methods for promoting personalized search results based on personal information. U.S. Patent 8,620,915, filed August 28, 2007, issued December 31, 2013.

Budiman BS, Krammer CP, Locke LL, Mueller MR, Lee S (2013) Method and apparatus for optimizing a performance index of a bulk product blending and packaging plant. U.S. Patent 8,447,423, filed November 19, issued May 21, 2013.

Huang M-H, Sung H-Y, Wang C-C, Chen D-Z (2013) Exploring patent performance and technology interactions of universities, industries, governments and individuals. *Scientometrics* 96(1):11–26.

INFORMS (2014) What is analytics? Accessed June 23, 2014, https://www.informs.org/About-INFORMS/What-is-Analytics.

Klavans R, Boyack KW (2009) Toward a consensus map of science. *J. Amer. Soc. Inform. Sci. Tech.* 60(3):455–476.

Koenen J, Peitz M (2013) Firm reputation and incentives to "milk" pending patents. Working Paper 4355, CESifo Group Munich, Munich, Germany.

Meucci A (2013) Systems and methods for portfolio analysis. U.S. Patent 8,548,884, filed July 2, 2008, issued October 1, 2013.

Miller CH (2010) Method for deferring demand for electrical energy. U.S. Patent 7,783,390, filed August 13, 2007, issued August 24, 2010.

Murphy FH (2002) The occasional observer ⟨www.uspto.gov⟩, a new source for what is happening in operations research practice. *Interfaces* 32(3):26–29.

Schneur R (2011) The president's desk. *OR/MS Today* 38(1).

Sternickel K, Szmanski B, Embrechts, M (2014) Use of machine learning for classification of magneto cardiograms. U.S. Patent 8,744,557, filed February 20, 2013, issued June 3, 2014.

Supreme Court of the United States (2014) Alice Corporation PTY. LTD. v CLS Bank International et al. Accessed June 24, 2014, http://www.supremecourt.gov/opinions/13pdf/13-298_7lh8.pdf.

U.S. Patent and Trade Office (2014) 2141 Examination guidelines for determining obviousness under 35 U.S.C. 103 [R-11. 2013]. Accessed September 2, 2014, http://www.uspto.gov/web/offices/pac/mpep/s2141.html.

World Intellectual Property Organization (2014) Statistical country profiles. Accessed June 8, 2014, http://www.wipo.int/ipstats/en/statistics/country_profile/countries/us.html.

Zhang B, Wu R, Hsu M (2013) Method, system, and product for performing uniformly fine-grain data parallel computing. U.S. Patent 8,620,985, filed October 14, 2010, issued December 31, 2013.

R. John Milne joined Clarkson University in 2010 after a 26-year career at IBM where he was a master inventor and senior technical staff member. His research focuses on the application of operations research to decision problems in production planning and scheduling. He is the Neil '64 and Karen Bonke Assistant Professor in Engineering Management at Clarkson.

Brian T. Denton is an associate professor in the Department of Industrial and Operations Engineering at University of Michigan, in Ann Arbor, MI. Previously he has been an associate professor in the Department of Industrial and Systems Engineering at NC State University, a senior associate consultant at Mayo Clinic in the College of Medicine, and a senior engineer at IBM. His primary research interests are in optimization under uncertainty and applications to healthcare delivery and medical decision making. He completed his PhD in management science at McMaster University, MSc in physics at York University, and BSc in chemistry and physics at McMaster University in Hamilton, Ontario, Canada.

Troy White is in the honors program at Clarkson University and on track to graduate in 2017 with a BS in engineering and management.