

# Examining the Evolution and Distribution of Patent Classifications

Daniel O. Kutz

School of Library and Information Science

Indiana University

1320 E. 10th St., LI 011

Bloomington, IN 47405

dokutz@indiana.edu

## ABSTRACT

Today, more so than ever, patents play an important role in helping inventors and organizations protect their intellectual property. With a 150% increase in the number of patents granted over the last 20 years, a rich source of data is presented that can be examined to better understand the change and development of intellectual borders. In this paper we discuss visualization techniques that examine the classification of more than 2.5 million patents in an attempt to improve understanding of how an assignee's patent portfolio, and the patent space as a whole, is partitioned and evolves over time.

## Keywords

Patents, intellectual property, information visualization, data mining, treemap, scatter-plot

## INTRODUCTION

Organizations and businesses now rely strongly on patents to protect their intellectual property and business models. As market competition has increased, there has been a subsequent increase in pressure toward patenting innovation and intellectual property. Do to the increased number of patents being filed and accepted (Figure 1), it has become important to develop tools and methodologies that better help to understand current and past patent landscapes.

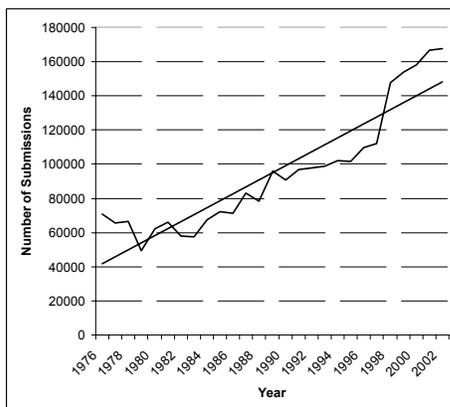


Figure 1. Number of Patents Granted for 1978-2002

Much of the research in the patent domain has focused on citation analysis. Francis Narin pioneered the use of scientometric methods to better understand the patent space and to identify patents of high value (Narin, Harhoff et al.

1999). While institutional differences have been identified in the patent citation process, as compared to traditional journal based citation (Meyer 2000), the practice still allows for the identification of high value patents, especially if a patent's citation to outside scientific sources are also taken into consideration (Narin 2003).

There has been extensive work done in visualizing knowledge domains (Börner, Chen et al. 2003; Börner, Maru et al. 2003; Boyack 2004); but this is work that has traditionally focused on scientific domains and not patents. McLean has researched using visualizations to facilitate understanding of search results in patent retrieval (McLean 2000) and Morris discusses the use of self-organizing maps (Morris, Wu et al. 2001) to present patent landscapes; but in general most of the available visualization tools and methods are in the form of proprietary software packages. For example, VxInsight (VisWave), ThemeScape (MicroPatent), and Anacubis IP Analysis (Anacubis) are software products to help analyze patent datasets (Anacubis; MicroPatent; VxInsight). These programs use textual analysis or citation mapping to provide an overview of a subset of the patent space.

However, we have not found any examples attempting to map the classification structure of the patent space. Breaking up the patent space at the classification level offers an ideal resolution for better understanding how the intellectual borders are established and change over time. In this paper we use data mining and visualization methodologies to analyze classification metrics of patents in order to better understand these evolving intellectual borders.

The paper is organized as follows: A discussion of the United States Patent Office (USPTO) classification and patent structure will be given; examples and visualizations of change in classification categories over the last 30 years will be presented and visualization techniques will then be used to compare the patent portfolio's of various assignees' in comparison to their classification structure. Finally, for a case study, we will look at the specific patent portfolios of Apple Computer and Lemelson Medical, Education & Research Foundation in an attempt to identify the classification categories of specific patent assignees.

## METHOD

The source data for this study is the USPTO patent archive containing all patent filings from January 1, 1976 to December 31, 2002. The data contains information on what day the patent was granted, title, abstract, patent number, author and coauthor, class and subclass under which the patent was filed, and citation references of each patent.

The USPTO currently has 466 classes under which a patent can be categorized (USPTO 2004). Classes cover a certain domain; for example class 455 encompasses all patents dealing with *Telecommunications*. Classes are further broken down by subclasses that have hierarchical associations. For example, class 455 features subclass 99 entitled “*with vehicle*.” This subclass is a child of subclass 95, “*mobile or portable*” which itself is a child of the mainline subclass 91, entitled “*transmitter*.” (USPTO 2004)

A patent is comprised of several claims, which describe the unique attributes of the artifact that the inventor is patenting. Each claim is categorized into a specific class/subclass, and for organization purposes, the patent as a whole gets assigned a single class/subclass (often referred to as its Original Classification, OCL). A patent with an OCL code of 455/91 means that a patent has been classified in class 455, subclass 91. This implies that the patent has made a claim about a vehicle based mobile or portable telecommunication transmitter.

In order to examine a change in patents and classes over the last three decades, the data set was examined in five-year intervals, from (inclusive) 1978-1982, 1983-1987, 1988-1992, 1993-1992, and 1998-2002. We retrieved from our dataset the original classification category of all patents that were filed in each respective five year interval. Our data was visualized using treemaps, which present a tree based hierarchy as a collection of nested rectangles, where the size and color of the rectangles can represent certain attributes (Shneiderman 1992). For our use, size shows total number of patents in each class, and color shows percentage increase (green) or decrease (red) in the number of patents from the previous interval. Change was calculated by comparing the number of patents filed in a specific class during the current interval in relation to the number of patents filed in a previous interval. Descriptive statistics were gathered in order to identify and rank classes that have the most activity and filings.

In order to develop a more understandable overview of which classes are associated with certain intellectual domains, the USPTO’s report *Classes Within the U.S. Classification System* is used to establish higher level hierarchies (USPTO 2002). This report places utility patents classes into three main hierarchies: Group I. Chemical and Related Arts; Group II. Communications, Radiant Energy, Weapons, Electrical, and Related Arts; and Group III. Body Treatment and Care, Heating and

Cooling, Material Handling and Treatment, Mechanical Manufacturing, Mechanical Power, Static, and Related Arts. The report breaks these hierarchies further down into sub-hierarchies. For example, Group I. contains sub-hierarchies for Conductors, Chemical Compositions, and Chemicals. It is within these sub-hierarchies that the USPTO has classified its specific class hierarchies and subclass ranges.

Scatterplots were generated in order to examine the number of patents filed versus the categories (classes) they were assigned to. Data was obtained by calculating the total number of patents granted to all assignee’s between 1978 and 2002, and by tracking the number of unique classes under which these patents were categorized. Custom scripts were developed for parsing the data, and scatterplots were generated with Microsoft Excel and R (Ihaka and Gentleman 1996).

To examine the patent portfolios of specific companies, all patents for Apple Computer and Lemelson Medical were retrieved from our dataset, and evolutionary treemaps were generated to show how the company’s patent portfolio have changed between 1978 and 2002.

## DISCUSSION OF RESULTS

### Class Changes

An examination of patent classes shows that classification categories provide insight about the current focus of research and development. For example, Table 1 shows the top ten classes with the most number of patents issued from 1/1/1978 to 12/31/1982:

Table 1. Top Classes 1978-1982

Top Classes 1978-1982		
Class	Title	# patents
260	Chemistry of Carbon Compounds	13544
424	Drug, Bio-Affecting and Body Treating Compositions	9621
252	Compositions	5858
128	Surgery	5204
428	Stock Material or Miscellaneous Articles	4994
73	Measuring and Testing	4855
123	Internal-Combustion Engines	4142
340	Communications: Electrical	4043
364	Electrical Computers and Data Processing Systems	3780
29	Metal Working	3664
<b>Total</b>		<b>59705</b>

This table shows that Class 260 (Chemistry of Carbon Compounds) is the most popular class, receiving 13544 patents between 1978 and 1982. Class titles indicate an emphasis toward the Mechanical (classes 29, 123, 428) and Electrical domain (classes 340,364).

Table 2 shows classification metrics 20 years later, for the period 1/1/1998 to 12/31/2002:

Table 2. Top Classes 1998-2002

Top Classes 1998-2002		
Class	Title	# patents
514	Drug, Bio-Affecting and Body Treating Compositions	18778
438	Semiconductor Device Manufacturing: Process	17775
435	Chemistry: Molecular Biology and Microbiology	17474
424	Drug, Bio-Affecting and Body Treating Compositions	13637
428	Stock Material or Miscellaneous Articles	13314
257	Active Solid-State Devices (e.g., Transistors, Solid-State Diodes)	12924
395	Information Processing System Organization	9955
345	Computer Graphics Processing, Operator Interface Processing, and Selective Visual Display Systems	9510
359	Optical: Systems and Elements	9151
365	Static Information Storage and Retrieval	8392
<b>Total</b>		<b>130910</b>

This table shows a dramatic increase in the number of filings: there are more than twice as many patent filings for popular categories as opposed to 20 years ago. In 2002, the focus is less on the traditional mechanical/electrical domain and more oriented toward Bio-Medical (classes 514, 435, 424), Information Systems (classes 395, 345, 365), and Optics (class 359).

Treemap visualizations of the whole patent space offer further insight about the dataset. The advantage of using a treemap is that it allows one to examine and compare three data attributes at once: the total number of patents added to a class during an interval, the increase or decrease of number of patents in a class from the previous interval, and when combined with the USPTO's Classification Guide, the ability to compare these trends to the USPTO's defined hierarchy.

Figure 2 shows the patent space for 1998 to 2002. The predominance of bright green patches indicates that many classes have seen an increase in number of patents from 1993 to 1997. The treemap also shows that category II – Communications, Radiant Energy, Weapons, Electrical, is the largest category, while Group I – Chemical and Related Arts, contains three of the top ten largest classes, namely 514, 435, and 424.



Figure 2. – Treemap Visualization of 98-02 Patent Space

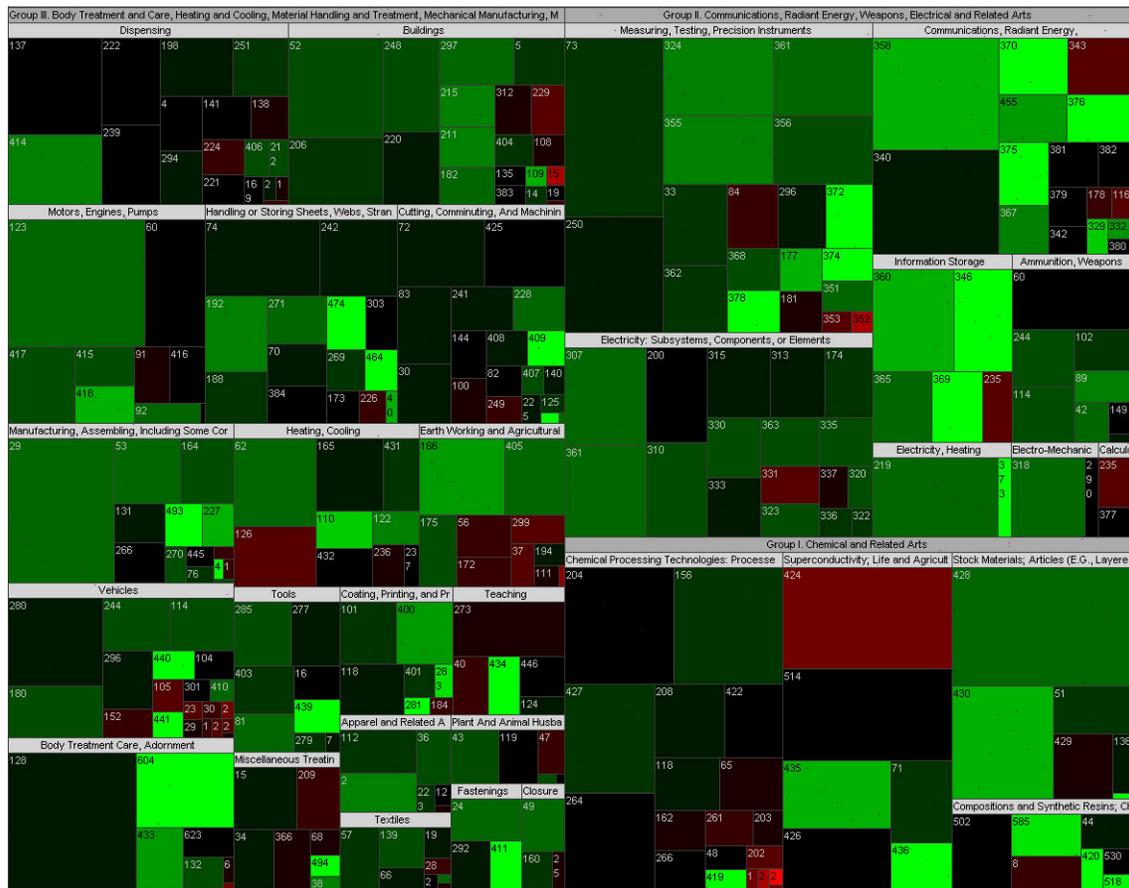


Figure 3. – Treemap Visualization of 83-87 Patent Space

A treemap of the patent space for 1988 shows a different configuration (Figure 3). Group III is the most active hierarchy, and the super-conductivity sub-hierarchy is not showing as much growth. Decline or minimal expansion exists in other categories, such as Chemical Processing, Vehicles, and Teaching, but strong growth is taking place in Communications, and Information Storage.

### Patents vs. Classes

By examining the categories of patent classes for an organization, one can get another view of the patent space while also gaining further insight into the patent portfolio of specific assignees. Figure 4. shows a scatter-plot of a count of total number of patents held by all assignees, for the whole dataset, in relation to a count of the total number of classes an assignees' patents cover.

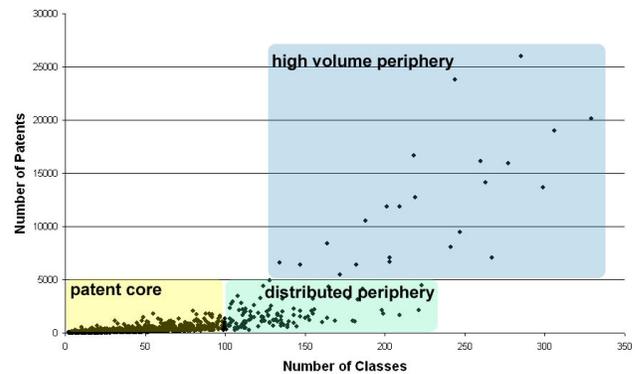


Figure 4. Patents vs. Class Scatterplot

This plot shows that in the last 30 years, the majority of assignees, which make up the “patent core,” have received less than 5000 patents distributed across less than 100 classes. One also has a wide periphery, which we refer to as the *high volume periphery*, made up of assignees who have received 5000 or more patents (Table 3).

Table 3. Top Patent Holders

Top 10 Patent Holders 1978-2002		
Name	# Classes	# Patents
International Business Machines Corporation (IBM)	285	25970
Canon Kabushiki Kaisha	244	23773
General Electric Company (GE)	329	20107
Hitachi Ltd	306	18987
NEC Corporation	218	16699
Kabushiki Kaisha Toshiba	260	16127
Mitsubishi Denki Kabushiki Kaisha	277	15917
Eastman Kodak Company	263	14147
Matsushita Electric Industrial Co Ltd	299	13683
Sony Corporation	227	13484

Figure 5. shows a close-up of the high-volume periphery, made up of only 36 organizations. While IBM may lead with total number of patents, the plot shows that General Electric's patent portfolio is more diverse as opposed to Canon's (this is expected, as Canon's business holding is less diverse than that of General Electric). Patent assignee's in the high-volume periphery share common organizational attributes, namely the majority of them are fairly large, located in Asia, and have a strong technology focus.

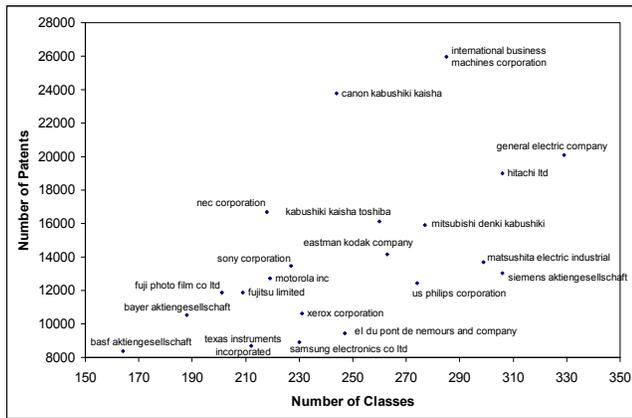


Figure 5. High Volume Periphery (Close-up of Fig. 4)

The Distributed Periphery, on the other hand, (Figure 6) is comprised of assignees that have received less than 5,000 patents that are spread over at least 150 classes. Here, the focus is on fairly mid-volume producers with a strong spread across the intellectual domain.

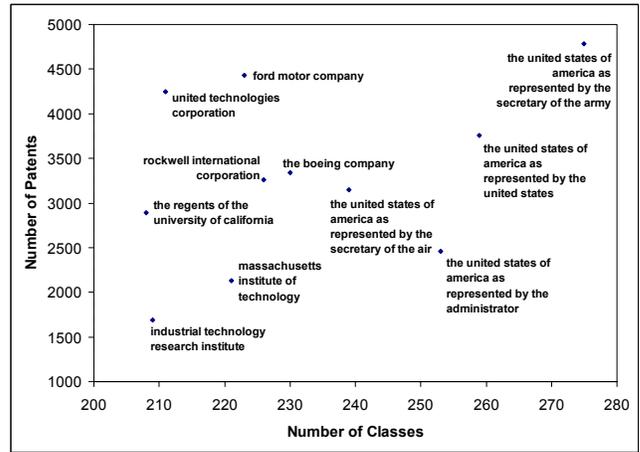


Figure 6. – Distributed Periphery (Close-up of Fig. 4)

The assignees that fall into this group are of a different organizational type than those in the high-volume periphery. Here, the majority of assignees consist of academic or national organizations focused on research and development, including four branches of the United States Government.

### Examining Assignee's Portfolios

Examining the whole patent space enables one to get a global perspective of changing intellectual borders, and scatterplots of an organizations' patent portfolio offers further insight into the relation between patents and classes. One can also use treemaps to view an assignee's specific portfolio in order to get an understanding of an organizations own intellectual borders. For example, Figure 7 uses treemaps to show how Apple Computer's portfolio has evolved in yearly increments from 1980 to 2002.

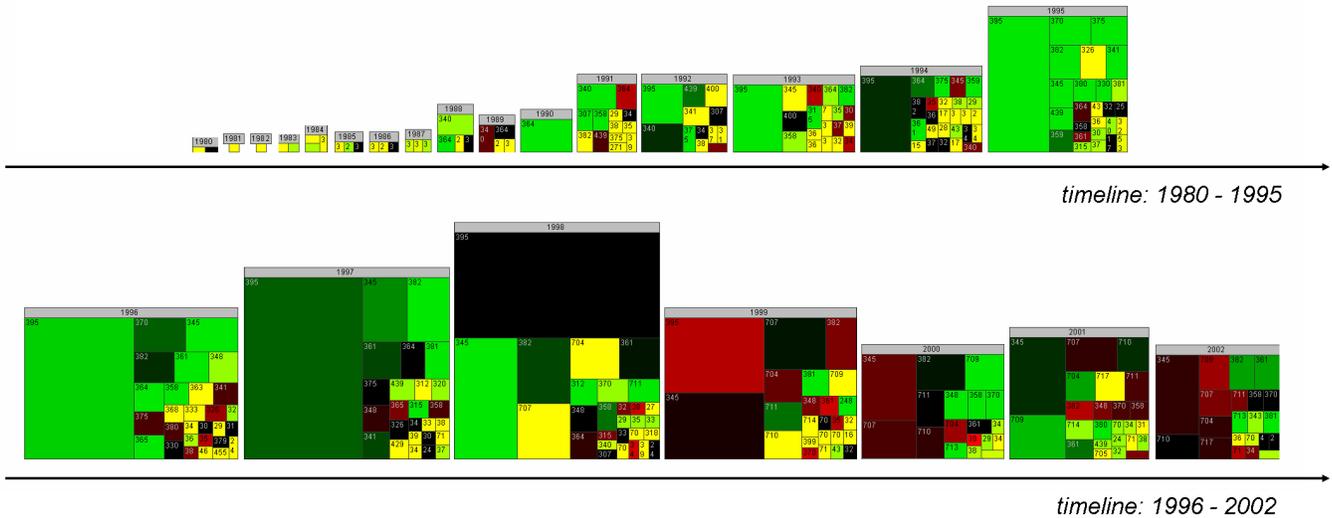


Figure 7. Apple Computer's Patent Portfolio 1980-2002

Classes that are shown in shades of yellow signify that in the previous one to five year the assignee had not been granted patents in that specific class. In this example, the treemap illustrates that in most years, more than half of Apple Computer's patent filings were placed into four classes, namely 395 (Information Processing System Organization), 345 (Computer Graphics Processing, Operator Interface Processing, and Selective Visual Display Systems), 382 (Image Analysis), and 707 (Data Processing: Database and File Management or Data Structures.) These four classes are an integral part of Apple Computer's patent portfolio, receiving patent's every year. Lack of continuity is mainly identified in smaller classes.

By placing the treemaps next to each other on a timeline, one is better able to see changes in the number of patents granted. In the case of Apple Computer, one can identify a decrease after 1998. If one were to attempt to visualize the whole patent space in a single treemap, it would require more effort to identify progression, as the default algorithm would sort based on a patent portfolio size, and not by year.

A treemap visualization of an assignee whose patent distributions are more diluted over the intellectual space would differ from the above example. An example of this would be the patent holdings of Jerome Lemelson, a prolific inventor considered by critics as someone who took advantage of the patent system and by admirers as a genius of same stature as Thomas Edison (Baker and Ertel November 2002).

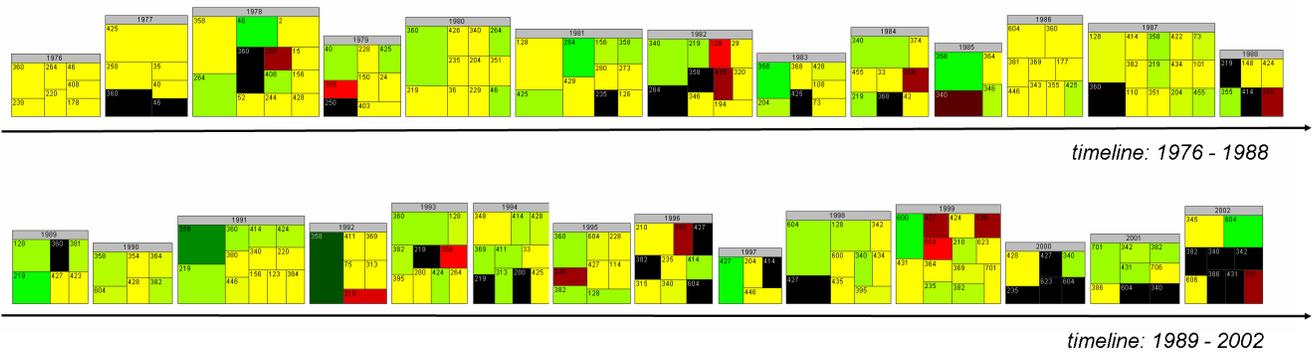


Figure 8. Jerome Lemelson's Portfolio 1976-2002

A treemap of Lemelson's patent holdings shows a more even distribution of patents over multiple classes. No class dominates over the years for patent filings. This illustrates that patents assigned to Lemelson, while fewer in number than Apple Computer, are distributed more broadly over the intellectual space, as opposed to Apple Computer, whose patent portfolio dominates four classes. The predominance of yellow also shows that there is not much

continuity from previous years in regards to which classes patents get filed. During any given year, it would be more likely for Lemelson to be granted a patent in a new class in which he didn't receive any patents the previous year.

**CONCLUSION**

In conclusion, data mining and information visualization techniques can help to better manage and understand the

evolving patent space. By analyzing in what classes a patent is categorized, conservatively, one can get insight into current and past intellectual borders, and ideally, from this develop some understanding of where research and development is heading.

When looking at specific portfolios, and examining how patent-class distributions change, one can broadly identify an assignee's domains of interest. Visualizations facilitate awareness: if these domains are swiftly evolving and distributed, are building on past patents, or are focused on one or two specific areas. We are currently working on developing algorithms that classify assignees according to patent-class distributions and using visualizations to validate our results.

Further work also remains to be done in linking patent data with other scientific metrics in an attempt to not only understand the pragmatic technological aspects that patents codify, but also recognize how patents and other scientific knowledge domains evolve and interact with each other.

#### ACKNOWLEDGMENTS

I would like to thank Dr. Katy Börner for continued review and feedback during the development of this paper and Dr. Steven Morris and Wu Zheng at Oklahoma State University for their help in organizing and supplying the patent dataset.

#### REFERENCES

Anacubis Anacubis Desktop and Intellectual Property Analysis Add-in.

Baker, W. H. and P. D. Ertel (November 2002). "The Lemelson patents: A retrospective." Intellectual Property & Technology Law Journal **14**(11): 1-6.

Börner, K., C. Chen, et al. (2003). Visualizing Knowledge Domains. Annual Review of Information Science & Technology. B. Cronin. Medford, NJ, Information Today, Inc./American Society for Information Science and Technology. **37**: 179-255.

Börner, K., J. T. Maru, et al. (2003). "The Simultaneous Evolution of Article and Author Networks in PNAS." Presented at the Arthur M. Sackler Colloquium on Mapping Knowledge Domains, NAS' Beckman Center, Irvine, CA, May 9-11, 2003.

Boyack, K. W. (2004). "Mapping knowledge domains: Characterizing PNAS." Proc. Natl. Acad. Sci. **10.1073**.

Ihaka, R. and R. Gentleman (1996). "R: A Language for Data Analysis and Graphics." Journal of Computational and Graphical Statistics **5**(3): 299-314.

McLean, A. W. (2000). Patent Space Visualization for Patent Retrieval. Guildford Research Park, Surrey GU2 7YP, UK, Canon Research Centre Europe, Ltd.

Meyer, M. (2000). "What is Special About Patent Citations? Differences Between Scientific and Patent Citations." Scientometrics **49**(1): 93-123.

MicroPatent ThemeScape.

Morris, S. A., Z. Wu, et al. (2001). A SOM Mapping Technique for Visualizing Documents in a Database. International Joint Conference on Neural Networks Proceedings.

Narin, F. (2003). From Science Papers to Technology Patents to Company Financial Performance. National Academy of Sciences Sackler Colloquium.

Narin, F., D. Harhoff, et al. (1999). "Citation Frequency and the Value of Patented Inventions." The Review of Economics and Statistics **81**(3): 511-515.

Shneiderman, B. (1992). "Tree Visualization with Tree-Maps: 2d Space-Filling Approach." ACM Transactions on Graphics **11**(1): 92-99.

USPTO (2002). Classes Within the U.S. Classification System, United States Patent Office.

USPTO (2004). Class 455. Telecommunications.

USPTO (2004). US Classes by Number.

VxInsight, VisWave.