Visualizing the Landscape of U.S. University Patents at Twenty Patenting Intensive Universities

Margaret M. Clements
Indiana University
201 N. Rose Ave., SOE 4228
Bloomington, IN 47405
mclements6@yahoo.com
001-812-330-8697

Todd Holloway
Indiana University
1320 E. 10th St., LI 011
Bloomington, IN 47405
tohollow@cs.indiana.edu
001-812-856-7034

HyunSeung Koh
Indiana University
1320 E. 10th St., LI 011
Bloomington, IN 47405
hskoh@indiana.edu
001-812-856-7034

Adity Mutsuddi
Indiana University
1320 E. 10th St., LI 011
Bloomington, IN 47405
amutsudd@cs.indiana.edu
001-812-856-7034

ABSTRACT

In this paper, we present a visualization of the demography of U.S. University patenting activities between 1969 and 2000 at twenty universities producing the highest number of patents during the same time period. Extracting patents granted to U.S. universities from the United States Patent and Trademark Office [USPTO] PATSIC, CONAME and Inventor datasets, we analyze the data to explore the relationships between university assignee, and classification structures at sixty-five of the most highly funded universities with regard to total research expenditures as reported by the National Science Board (1). For the purposes of this analysis, we visualize three different time frames: 1969-1980; 1981-1990; and 1991-2000 in order to understand how time and policy initiatives have impacted patenting activities at these twenty research intensive universities.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*clustering, relevance feedback*

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—*graphical user interface, interaction styles, screen design*

I.4.10 [**Image Processing and Computer Vision**]: Image Representation—*hierarchical*

I.5.3 [**Pattern Recognition**]: Clustering—similarity measures

J.1 [**Administrative Data Processing**]: Education

General Terms

Algorithms, Management, Economics, Human Factors.

Keywords

Higher Education, Patent, Analysis, Visualization.

1. INTRODUCTION

Since the passage of the 1980 Bayh-Dole Act, significant changes have occurred in the privatization of university research activities. Prior to 1980, it was not customary for universities to patent publicly funded research. The Bayh-Dole Act, however, was enacted to encourage university partnerships with industry by permitting government grantees and contractors the right to patent federally funded research. Prior to 1980, university patents accounted for less than one-half of 1 percent of all U.S. patents. Currently, university patents comprise approximately 5 percent of all U.S. patents [1]. The annual number of patents granted to universities has risen from 250 in the early 1970's to more than 3,200 in 2003 [1]. Furthermore, both federal funding and patents have been historically concentrated at twenty universities.

Despite the growing participation of U.S. universities in the marketplace, little is known about the actual relationships that result in patents granted to universities. This study sets out to explore relational factors such as researcher clustering, researcher interconnectivity, assignee clustering, assignee interconnectivity, resource concentrations, patent class clustering and other relational factors that describe the emerging landscape of the academic scientific and engineering enterprise. Ultimately, we present our results by visualizing the data into a geographically interpretable and relevant form. This study will represent the data utilizing multiple maps of the university patenting domain utilizing VxInsightTM and ARIST.

2. RELATED WORK

Great strides have been made recently in the analysis of patent databases using visualization techniques. Specifically, a recent initiative directed by Dr. Katy Borner at Indiana University has resulted in important visualizations and analyses of knowledge domains, co-authorship teams, evolution and distribution of patent classifications. Likewise, Dr. Borner's collaboration with Dr. Kevin Boyack and others at Sandia National Laboratories has resulted in analyses of certain patent classes in a time series analysis. Additionally, Daniel Kutz's "Examining the Evolution and Distribution of Patent Classifications" analyzes the patent space as a whole and how it evolves over time using Treemaps.

The geographic component to the visualization was inspired by a number of works, including the globes of Ingo Gunther [Appendix A], and the cartographic maps of Andre Skupin [Appendix A]. These visualizations of patent databases have laid the groundwork for the current project.

With specific reference to university patents, there is a fairly large number of works examining both the patent classification system and the role of universities. Of particular interest is the work of Jerry and Marie C. Thursby: "Who is Selling the Ivory Tower? Sources of Growth in University Licensing which provides insights into the growth trends our visualization depicts.

3. DATASET

NetSci2006, May 22–25, 2006, Bloomington, IN, USA.

The primary dataset is the USPTO Patent Library. The attributes included in the database are: patent number, inventors' names, patent application submission date, date patent was granted, names of the patent examiners, address(es) of inventor(s), other relevant patent numbers referenced, patent classes explored by the patent examiners for existing patents, patent assignee (for this study, we explore only U.S. university assignees), patent assignee code, primary classification number and the state in which the assignee is located.

Our dataset extracts patents from the USPTO's PATSIC, CONAME and INVENTOR data files that were granted to U.S. universities between 1969 and 2003. During this time period, 42,625 patents were granted to 578 universities. A patent can be categorized into one of the 450+ classes in the USPTO library. Every patent is assigned a specific class/subclass. The primary classification number in our dataset captures the class categorization. The primary classification number is a 9-digit number where the first 3 digits correspond to the class number, the second 3 digits correspond to the subclass number and the last 3 digits correspond to the sub-subclass number.

3.1 Data Reduction

To aid in visualization, we focused our attention on the top 65 highest funded universities which were granted 31,055 patents during the same time period—or 72.8 percent of all patents granted to universities. These 65 universities were selected for visualization in this project based on concentration of research funding as reported by the National Science Board and the total number of patents granted since 1973 [1]. For this analysis, funding represents the amount of expenditures on R&D from all sources (Federal, state, industry, academic and other sources) for the year 2003. Once we analyzed this data, we further reduced our data to the 20 most highly funded universities to develop a prototype for our final interface.

Patents can be granted to universities under different names. That is, multiple name codes can belong to one university and there is inconsistency in the way that universities apply for patents from the USPTO. For example, patents granted to the University of California include the Regents of the University of California, University of California Office of Technology Transfer, Los Alamos National Laboratory, UCLA Medical Center and Research and Education Institute, University of California at Berkeley, etc. We have combined such patent information into "preferred assignee names" for ease of analysis. We have similarly combined the data we have on R&D funding.

To analyze the changes in patents and classes for universities over the years, we divided our dataset into three time frames based on our client's interest- 1969-1980, 1981-1990 and 1991-2000. To classify a patent in a specific time frame, we considered the issue date. We decided to discard information for 2001 to 2003 as it will skew the comparison among the three time frames. Our main goal is to find the fields in which each university is most active over the three decades. We used only the class number assigned to each patent in our dataset to perform the classification analysis. The patents in the dataset fall into 340 classes. However, these classifications are too technical. For example, class number 520 is titled "Synthetic Resins and Natural Rubbers" and class number 073 is titled "Measuring and Testing". In order to get a better overview of which classes belong with certain fields, a simpler categorization is required. For this we used the categorization schema developed by [1] at NBER (National Bureau of Economic Research).Their categorization maps each class number to one of the 6 categories- Mechanical, Chemical, Electrical and Electronics, Computers and Communications, Drugs and Medical, and Others. These categories are then divided into subcategories. For example, Drugs and Medical contains Biotechnology, Drugs, Miscellaneous-Drugs and Medical, Surgery & Medical Instruments. For our project we will consider only the top level categories.

The categorization schema by NBER is based on patent information from 1963 to 1999. As a result, some class numbers in our database were not categorized in their schema (new class numbers are added over time into the USPTO database). We categorized each of these class numbers to one of the 6 categories to the best of our understanding. Table 1, below, shows the class numbers and their categorization.

Class Number	Class Title (USPTO)	Category
725	Interactive video distribution systems	Computer
719	Electrical computers and digital processing systems: interprogram communication or interprocess communication (ipc)	Computer
718	Electrical computers and digital processing systems: virtual machine task or process management or task management/control	Computer
717	Data processing: software development, installation, and management	Computer
716	Data processing: design and analysis of circuit or semiconductor mask	Computer
715	Data processing: presentation processing of document, operator interface processing, and screen saver display processing	Computer
703	Data processing: structural design, modeling, simulation, and emulation	Computer
398	Optical Communications	Electrical
073	Measuring and Testing	Mechanical
060	Power Plants	Mechanical

Table 1: Class numbers not categorized by NBER

We categorized all the patents in our dataset into 6 categories. For our visualization we counted the total number of patents and the number of patents in each category for each university in each time frame. We also counted the total number of patents in each class for each time frame to give us an idea about the overall distribution of classes for all 65 universities. However, this visualization includes only 20 universities selected for their funding intensity in order to develop this prototype for our final interface.

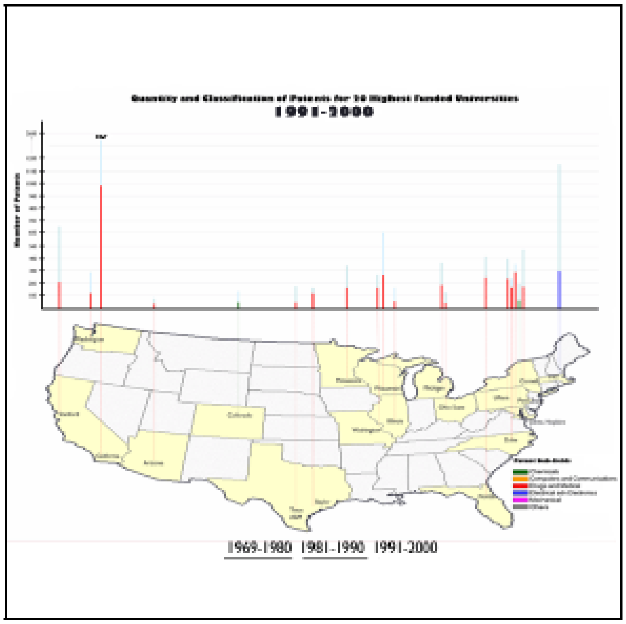

Figure 1: Overview level

3.2 Data Visualization

The purpose of our visualization is to provide our client with as much information as possible from many different perspectives. This means that we need to find the best way to facilitate the display of large amount of information within limited space. Therefore, we decided to employ the technique of generalization used in Cartography, which enables users to obtain information with different levels of details without losing context by zooming in and out. We decided to visualize our patent dataset using the map of USA at three different levels- overview, state and university.

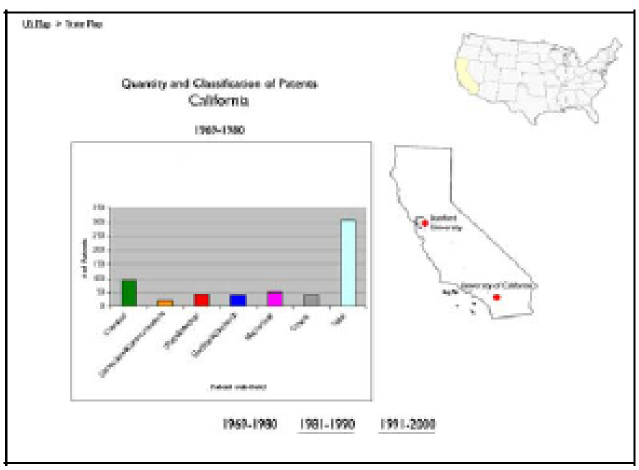

Figure 2: State Level

We chose to use a map for two reasons. First, it is an easy tool that aids in geographic visualization of concentrated resources. In this case, a map is used as an alternative tool to index which facilitates the rapid search of a single element from bulk data. Second, the map enables users to simultaneously obtain three different kinds of information: university patent information, geographic distribution of universities and university location.

We divided information into three levels of 'overview', 'state', and 'university'. In the overview level, we visualize 'total number of patents' and 'number of patents in the most active field' for each university by decade. We use standard bar graphs to present the number of patents and use the map as the x-axis. We provide buttons to view the same information for each decade. The user can click on any of the states to get more information about individual universities in a state (See Figure 1).

At the state level, we visualize 'total number of patents' and 'number of patents in each of the 6 categories' in a state by decade. In this case, we use standard bar graphs where the x-axis represents patent classification category. Universities are depicted by red dots on the state map. The user can click on the red dots to view more information about a university. To help the user maintain context while zooming to more detailed information, we maintain a visualization of the U.S. map in the corner at the same time we shade the state selected by the user (See Figure 2).

At the university level, we visualize the change in the 'number of patents in each of the 6 categories' for a single university. IN this case, we use standard line graphs which show the changes in the number of fields over the three decades very clearly (See Figure 3).

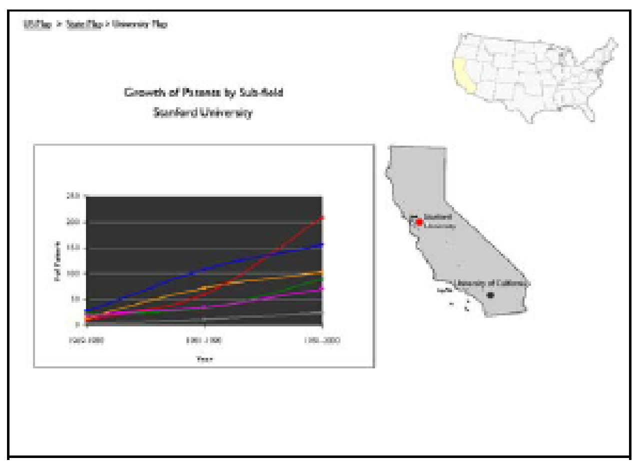

Figure 3: University level

At this level, the selected university is represented by a red dot on the state while the other universities retreat into the background by changing color to dark gray.

4. LIMITATIONS

Several concerns emerged as we were visualizing this dataset. Below is a discussion of the issues we confronted and how we resolved them.

4.1 Scaling

To enable users to compare data across three decades, we use the same scale on the Y-axis for each decade. However, because the large difference between university patenting activity before and after 1980 (when the Bayh-Dole Act was ratified), the range of patents between the decades varies between 5 and 2,500. Utilizing the same scale for each decade presents difficulty for accurately presenting small numbers. To solve this problem, we adjusted our scale to present numbers as high as 1200. For numbers larger than 1200, we marked the Y-axis with '…' and wrote the exact number on the graph (see Figure 1).

4.2 Universities in Close Proximity

Most highly funded universities are located in the eastern United States. Due to their close proximity, mapping and displaying the data results in the overlapping of information. To solve this problem, we slightly distort the map by stretching it horizontally to create more space between closely located universities.

4.3 Map Size

The map itself occupies too much space. However, the map we chose was the only map with sufficient space to display university names. We resolve this problem by squeezing the map vertically in order to crate more space for the graph above the map.

4.4 Prototype for Patent Visualization

The visualization prototype we have designed can be developed into a tool for analyzing quantity and classification of patents over time for all universities in the United States. Currently, our visualization is capable of providing information for the top 20 highest funded universities. At each level our client can obtain the following information:

Overview Level

- The number of patents received by universities
- The most active field in individual universities
- The universities with the highest (or the lowest) number of patents in comparison to others
- The number of patents in the dominant field compared to the total number of patents in an university (proportional information)
- Changes in the number of patents granted to the universities over time
- Changes in the most active field over time

State level

- The number of patents received by universities in a single state
- The number of patents in the 6 fields
- Which field in which state was more active than the other fields

University level

- Trends in the 6 individual fields across three decades for a single university

4.5 Validation

This data visualization has been validated with data available through the National Science Board and the USPTO.

5. CONCLUSION AND FUTURE WORK

The major challenge we faced with this project was the need to draw the bar graphs manually because we couldn't find a proper tool which could automate drawing bar graphs from a specific location on a map (where the map is the x-axis) to another specific point on the Y-axis. Therefore, one of the extensions of this project could be to program this process.

As mentioned in the earlier section, our prototype can be developed into a tool to analyze patents received by US universities. In addition, our prototype can be developed into a visualization tool that can draw standard graphs with a map as the x-axis. This will be useful in analyzing data related to geographic location. The zoom in and out function can be used to view detailed information about any specific location on a map.

Due to absence of an automated tool, we were able to visualize data for only the top 20 highest funded universities. For more accurate data analysis and visualization, it would be better to use all 65 universities or more. However, this will result in several problems. First, it will be difficult to draw lines for closely located universities in the overview level. Second, labeling the universities will also be an issue. We will have to design a legend with icons or short names for every university on the map.

6. ACKNOWLEDGEMENTS

We thank Katy Borner for her support and guidance. We also thank Weimao Ke for helping us with parsing the data.

7. REFERENCES

[1] National Science Board, 2006. Scien*ce and Engineering Indicators*, *2006*. Two volumes. Arlington, VA: National Science Foundation (volume 1, NSB 06-01; volume 2, NSB 06-01A).

[2] B.H. Hall, A. B. Jaffe, and M. Tratjenberg, *The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools*. NBER. Cambridge, MA, 2001 Available at http://www.nber.org/patents/

[3] Kevin W. Boyack†, Brian N. Wylie, George S. Davidson and David K. Johnson, *Analysis of Patent Databases Using VxInsightTM*. New Paradigms in Information Visualization and Manipulation, a Workshop at the 9th International Conference on Information and Knowledge Management (CIKM 2000), McLean, VA. November 10, 2000. Available at http://ipal.sandia.gov/vxinsight/pubs/npivm00.pdf

[4] Brewer, C. A. *Color Use Guidelines for Data Representation*. Proceedings of the Section on Statistical Graphics, American Statistical Association, Alexandria VA. pp. 55-60. 1999.

[5] Ingo Gunther, World Processor. Available at http://worldprocessor.org/

[6] Andre Skupin, *From Metaphor to Method: Cartographic Perspectives on Information Visualization*. Roth, S.F. and Keim, D.A. (Eds.) Proceedings IEEE Symposium on Information Visualization, Salt Lake City, Utah. Los Alamitos: IEEE Computer Society, pp. 91-97. 2000. Available at http://doi.ieeecomputersociety.org/10.1109/INFVIS.2000.885095

[7] Andre Skupin, *A Cartographic Approach to Visualizing Conference Abstracts*. IEEE Computer Graphics and Applications. 22 (1): 50-58. 2002. Available at http://ella.slis.indiana.edu/~tneirync/g1050.pdf

[8] Robert J. Vanderbei, Geospatial Visualization of New Jersey Gubernatorial Race. Available at http://www.princeton.edu/~rvdb/JAVA/election2005/

[9] Daniel O. Kutz, *Examining the Evolution and Distribution of Patent Classifications*. Eighth International Conference on Information Visualisation (IV'04). 983-988. 2004. Available at http://doi.ieeecomputersociety.org/10.1109/IV.2004.1320261

[10] A Agrawal and R Henderson, *Putting Patents in Context: Exploring Knowledge Transfer from MIT*. Management Science, 2002. Available at https://dspace.mit.edu/retrieve/3871/IB_Putting%2BPatents.pdf

[11] R Henderson, AB Jaffe and M Trajtenberg, *Universities as a Source of Commercial Technology: A Detailed Analysis of University Patenting, 1965–1988*. The Review of Economics and Statistics, 1998.

[12] Jerry G. Thursby and Marie C. Thursby, *Who Is Selling the Ivory Tower? Sources of Growth in University Licensing*. NBER Working Paper No. 7718. May 2000. Available at https://palm.nber.org/papers/w7718

[13] Wallmark J.T., *Inventions and patents at universities: the case of Chalmers University of Technology*. Technovation, Volume 17, Number 3, pp. 127-139+162(13). March 1997.