Cite as: Smith, D., & Ali, A. (2014). Analyzing computer programming job trend using web data mining. *Issues in Informing Science and Information Technology, 11,* 203-214. Retrieved from http://iisit.org/Vol11/IISITv11p203-214Smith0494.pdf

Analyzing Computer Programming Job Trend Using Web Data Mining

David Smith and Azad Ali
Indiana University of Pennsylvania, Indiana, PA, USA

david.smith@iup.edu azad.ali@iup.edu

Abstract

Today's rapid changing and competitive environment requires educators to stay abreast of the job market in order to prepare their students for the jobs being demanded. This is more relevant about Information Technology (IT) jobs than others. However, to stay abreast of the market job demands require retrieving, sifting and analyzing large volume of data in order to understand the trends of the job market. Traditional methods of data collection and analysis are not sufficient for this kind of analysis due to the large volume of job data that is generated through the web and elsewhere. Luckily, the field of data mining has emerged to collect and sift through such large data volumes. However, even with data mining, appropriate data collection techniques and analysis need to be followed in order to correctly understand the trend.

This paper illustrates our experience with employing mining techniques to understand the trend in IT Technology jobs. Data was collect using data mining techniques over a number of years from an online job agency. The data was then analyzed to reach a conclusion about the trends in the job market. Our experience in this regard along with literature review of the relevant topics is illustrated in this paper.

Keywords: Web data mining, Web analysis for job demand, Data analysis for programming job demand.

Introduction

Job seekers should use all available resources and strategies to find their next break. A good job-search begins with awareness of the technologies available today and those arriving tomorrow, as well as the skills that are in demand. Obviously, an IT pro can quickly become obsolete in today's competitive marketplace. (McLean, 2006, p.38)

Indeed job seekers need to use all available resources to stay abreast of the job market and develop skills according to market demands. The same could be said about educators who potentially graduate these job seekers and theoretically prepare them for the job market (Litecky, Aken, Ahmad, & Nelson, 2010). Keeping an eye on the job market and learning about the market demands may help better prepare the students for such market demand (McLean, 2006). However, staying abreast of the market demand faces many challenges. The data collection for these job demands is comprehensive,

sifting through this large volume of data requires substantial effort in order to analyze and tabulate. Many difficulties are faced when analyzing the job market demand and draw conclusions about such demand. Technology helps often with this matter. Web Data mining is the technology that is often referenced when it comes to dealing with this large volumes of data (Zhong, 2008). Web data mining results in extensive (and often unnecessary) amounts of data. The amount of data given by web data mining often defeats the purpose for which it is collected. Therefore, a smart way of using web data mining techniques to analyze this extensive data volume in a way to give a better picture of the trend of data collected.

This paper is about using web data mining to analyze data collected about employment requirements in the programming field. It illustrates our experience with using web data mining to understand the job trends within this field. The paper shows how we collected and sifted through data using web data mining techniques. It then shows how the analysis of data is conducted to give meaningful set of information about job trends. The remainder of this paper is divided into the following sections:

- The next section includes a literature review of the terms data mining/web data mining in order to draw an understanding of the terms and their popular uses.
- The Literature Review - Data Mining Analysis Techniques section discusses the methods of web data mining analysis and techniques. It gives a brief description of each technique and their application to searching and analyzing job data.
- This is followed by Literature Review – Job Demand and Trend, which conducts similar literature review but this time about reviewing the job demand in the market and the specifics of these demands in the Information Technology (IT) field
- Analyzing Computer Programming Jobs Using Web Data Mining elaborates the experience of a faculty using web data mining to analyze and study programming job trends in the market. The section further shows sample analysis resulted from this study.
- The final section concludes the data collected and a suggestion for future research.

Literature Review - Web Data Mining

With the rapid development of the Internet, the information content on the Web has become very rich. A new technology is urgently needed to automatically find, extract and filter information from the web resources. Web mining technology appeared. The purpose of web mining is to find and extract the potential useful model and the hidden information from the Web documents and Web Activities. It combines together the traditional data mining and the web together, and can play a role in many ways, such as the mining of search engines, the development of search engines, improving and enhancing the quality and efficiency of search engines, determining authority pages, web document classification, web log mining, intelligent enquiries and the establishment of Meta-Web data warehouse. (Zhong, 2008, p. 514)

It is clear and evident through practice that the recent technological advances (especially the Internet) have enabled the collection of vast amount of information which needs to be managed and analyzed in order for it to be effective (Berry & Linoff, 2000, Huang, 2007, Kaushik, 2007). Traditional technologies of database management system prove to be incapable of handling the massive volume of data that is generated from these technologies (Ali & Kokun 2010; Shmueli, Patel, & Bruce, 2010). The newer technologies of data mining and web data mining have emerged to remedy these issues. But a general consensus on terms “data mining” and “web data mining” has not been attained resulting in some degree of confusion. Therefore the understanding of these terminologies may help in clarifying their proper use in the analysis of various data gathered. This section explains these terms by introducing the definitions of relevant terms that are widely accepted. It will also explain the suggested proper use of these technologies.

Definitions and Relevant Terms

Web data mining is a relative new field, thus a concise standard definition has yet to emerge. Instead, multiple terms have been introduced and different definitions have been used to describe web data mining, the steps to be followed, and presentation of results. Das and Turkoglu (2009), for example, used the term "web usage mining" in the context of access patterns:

> Web usage mining is used to analyze web log files to discover user accessing patterns of web pages. In order to effectively manage and report on a website, it is necessary to get feedback about activity on the web servers (p. 6635).

On the other hand, the web site "Web Data Mining .net" (Web Data Mining .net, 2009) used the term "Web data mining" and looked at it from the perspective of crawling through web resources and further provided the following definition for it:

> The term **Web Data Mining** is a technique used to crawl through various web resources to collect required information, which enables an individual or a company to promote business, understanding marketing dynamics, new promotions floating on the Internet, etc. There is a growing trend among companies, organizations and individuals alike to gather information through web data mining to utilize that information in their best interest (p. 1).

Mobasher (2010) used two terms to describe the same expressions: first, "web mining" and second "web usage mining" and further elaborate on both:

> Web mining refers to the automatic discovery of interesting and useful patterns from the data associated with the usage, content, and the linkage structure of Web resources. It has quickly become one of the most popular areas in computing and information systems because of its direct applications in e-commerce, e-CRM, Web analytics, information retrieval/filtering, Web personalization, and recommender systems (p. 1).

In regards to web logs, Buzikashvili (2007) described that the goal of researching through web log data is to understand the behavior of the visitors of the site as they browse through pages or search through search engines. Corsini and Marcelloni (2006) looked at the files that store such data and noted that a web log file stores the sequence of accesses to web pages that are managed by a web server.

Although different terms are used to describe data mining we will use the terms "data mining" and "web data mining" interchangeably. We consider "Web Data Mining" be the same as "data mining" but on the web. Similarly, the steps involved in data mining may be different from case to case. Thus these steps need to be explained as well.

Steps in Web Data Mining

Collecting data is the first step in data mining and at times tends to be the easiest amongst the multiple steps involved in the process. During this step, data may be filtered, classified and categorized. The next step is data analysis and subsequent decision making. A point to be clarified here is that the data collected can be understood differently. Thus, analysis of the collected data from different perspectives needs be completed before taking additional steps. In the case of web data mining, different methods may be used for each perspective. Therefore it is important to understand various methods of data analysis in order to appropriately shed light on the correct application of each. The section that follows sheds light on the techniques followed for data mining analysis.

Literature Review - Data Mining Analysis Techniques

The enormity of the data that current technology produces can be counterproductive in many ways. First, it takes a long time to tabulate and understand. Second, it may give misleading information to the viewers. Therefore, analysis of the data is necessary in order to understand the meaning, direction and trend of the data. Different techniques have been suggested to submit the collected data through data mining and give the correct view of reporting and analysis. Shmueli et al. (2010) for example suggested traditional statistical methods that are used for such analysis like linear regression, logistic regression, discriminant analysis, and principal component analysis. Shaw, Subramaniam, Tan, and Weldge (2001) on the other hand suggested five common techniques used in web data analysis; dependency analysis, class identification, concept description, deviation analysis and keyword search.

Based on the importance of analyzing the data in web data mining, the remainder of this section explains the common techniques used in analyzing the data generated from computer data mining. It shows the meaning of each technique, it explains how they are used in general and then discusses how it could be used in studying job trends in information technology.

Dependency Analysis

Dependency analysis is practiced when the data collected reveals that one variable is dependent on another variable. For example, the demand on cold medicine is dependent on the temperature and how cold the weather is in a certain city. Demand of ski equipment is dependent on the amount of snow in a particular area. In this case, the data collected may reveal certain dependency between variables involved or it may reveal dependency on other set of variables.

In terms of analyzing job search data, dependency analysis is helpful to measure the relationship between one variable (like years of experience) and another (like education level). Such analysis may reveal existence of a relationship and may lead to further study to assess which education level is preferable for certain jobs. In either case, dependency analysis could reveal the existence of a link between one job search variable and another.

Deviation Detection

Deviations are useful for the discovery of anomaly and changes. Anomalies are things that are different than normal. For example, compare a group of similar people and identify those who stand apart from the average, either in a positive or negative way.

This kind of analysis is used in job search where they have some expectation about job trend and there is a shift in one direction or another. For example, the search for COBOL job increased during the Y2K years, this created deviation from normal job demand. Detecting these deviations could be helpful to know especially in times of high demand for certain jobs to inform decision making.

Keyword Indexing

At the core of using web data mining is the keyword search. This works by classifying the search according to certain keywords. It searches for the occurrences of the keyword(s) specified and either shows a count of each keyword or it shows the context in which it was showing. Keyword indexing can be used in marketing for example to see the demand for certain products or certain brand. One of the limitations of relying on keyword search is that it produces large volume of results most of which unrelated. For example, a search for a software name Access could result in large number of unrelated use of the word Access as in entry or contact.

In job search analysis, it can be used to search for a certain job title, tools used or any specific keywords. The keyword search also can be used in conjunction with any of the above techniques to find more information about the data pattern in the data searched. Sobhi and Son (2010) advocated the use of keyword in studying the job trend in the industry. They suggested using a hierarchy of related keywords or a simple keyword to find such trend. They added that both methods are beneficial for finding common trends of job demand in the IT industry.

Literature Review – Job Demand and Trend

The rate and expansion of the IT industry has resulted in an abundance of job titles. The causes are multi-faceted, stemming from a technology explosion, coupled with the introduction of new programming languages, tools, software development paradigms, work practices and job requirements. There may also be, however, underlying factors, particularly when treating IT as a profession. (Donohue & Power, 2012, p.6)

Although recognizing patterns in job demand is important to all professions, identifying and following job trends takes special importance for IT profession. Prabhakar et al. (2005) noted that "IT Professionals know they must keep their skills up to date, but knowing so requires knowing what skills on demand" (p. 91). For IT educators, this kind of knowledge plays a significant importance because they (the educators) have the task of preparing the students for this fast changing environment. Reviewing trends in the job market is one way to get the educators informed of such job demand. In addition, being aware of the job trends in the market can bring additional benefit to the educators and their departments in different ways. The remainder of this sections explains how staying abreast of the job demand in the IT market can bring these benefits.

Close the Expectation Gap

There is a general sense among various experts that a gap exists between the skills that IT jobs market demand and the training being offered at IT programs. This gap is explained in different studies. For example, a comprehensive study conducted by Gallivan, Truex, and Kvansy (2002) provides an analysis of the changing patterns for IT professionals. This study makes the following notes about IT education and what is offered at colleges and IT programs:

- The IT education is criticized as incapable of producing qualified employable professionals in the field.
- Universities teach obsolete or irrelevant technologies
- Academic programs are out of touch with the reality of the job market
- They neglect to provide skills that are on demand.

This "curriculum gap" is due to the variance between what is taught at the curriculum and what employers expect from the IT college graduates. Scott-Bracey (2013) on the other hand called it the "skill gap" and pointed particularly to the absence of teaching badly needed "soft skills" in IT programs. She noted that majority of employers are demanding soft skills as a pre-condition for their employment. McGill (2009) called it the *expectation gap* referring the difference between what is considered important skills by the industry and those taught in academia.

As explained above, there are abundant claims of gaps between academia and the need of the IT industry. Staying abreast of the trends of jobs in the market is one way of recognizing this gap. Reviewing, studying, classifying and analyzing job demands can give many clues about what is expected in the job market in terms of both soft and hard skills. This in turn can be used by teaching departments to shape curriculum. A comparison between what is asked for and what is being offered at the department will results in identifying and closing this gap (Surakka, 2007). This gap can sometimes be closed (or narrowed down) by offering some of what is demanded in the

job market. For example, offering a section on "cloud computing" will provide skills currently being demanded in the job market (Burns, 2012).

The Legacy Job Titles

Unlike some other fields, the IT has job titles tend sometimes to be obsolete after certain period of time. The term "Legacy jobs" is often used to describe jobs that require the use of outdates technologies or technologies that not as commonly used in the market.

Donohue and Power (2012) conducted a study that searched job titles in the IT field as compared with three other professions: Medicine, Accounting and Engineering. The study found that IT job titles often refer to tasks that are short termed and extracted from job requirements that no longer exist in the present time. Although, the study concludes also that there is still demand for some technology that is deemed "legacy". The study referred to the requirements for COBOL programming and similar other technologies that it deemed them as "legacy"

The common impression is that having courses that teach "legacy" technologies is considered a liability to the program that teaches it. After all who will take a course about a technology that is deemed obsolete, some might think. However, taking a closer look, teaching a legacy course may be beneficial at times when the legacy technology still is being widely used. The example is the case of teaching COBOL, if many organizations still use COBOL and programs still cancel their courses that teach COBOL, this creates a shortage of graduates with COBOL background. Continuing this trend may increase the demand on the graduates of the few remaining programs that teach COBOL. A continuous study and analysis of job demands reveals more about the availability of such jobs and in turn could help in determining the feasibility of keeping such courses or not.

Filling the Scarcity of Talents

There is a general notion that when a new technology emerges to the market, the market shifts focus on the emerging technology. Thus, more companies switch to the new technology and they will be in need to hire new people who are trained on the new technology. This creates what is termed as "scarcity of talents" (Bhatt et al., 2010) or a scarcity in the supply of graduates who are trained on the new version of the technology. For example, when Microsoft updated their Visual Studio version 6 (VS6) to Visual Studio .NET (VS .NET), this created a scarcity of talents and demand for people who are trained on the .NET platform (Ali & Wood, 2004; Prabhakar et al., 2005). In this case, the academic programs that first started teaching to VS .NET, were able to fill this gap and their students had a better chance of finding jobs in these new emerging technologies.

According to Burns (2012) hiring for individuals with cloud computing skills get paid 20% higher than peers who is hired with other set of skills. The increase in pay for these "cloud computing" jobs is because these are considered one of the jobs that face "scarcity of talents" in the market. Fewer people have this knowledge than in some other jobs. For this reason, demand increase for people with these jobs and they tend to get paid higher than in others. Taking it a step further, the demand for the academic programs that offer studies about these jobs tend to be higher. Departments thus can benefit from learning about these jobs through job analysis and may provide opportunities for their student to fill these jobs with the "scarcity of talents".

A study of job trends in the market place could help to understand this scarcity of talents and my give a reason for updating the program. This may be accomplished through studying the job market in general or it could be accomplished through classifying, tabulating or summarizing the results. In either case, the scarcity of talents can be observed through such monitoring. Many other

indicators can be studied to understand this scarcity of talents and the ways that it fulfilled to benefit the students in their program.

Internship Opportunities

McLean (2006) conducted a study to assess the employable skills that starts students in the industry and what termed as "A foot in the door" for IT professionals. McLean emphasized many factors enhance the chance of applicants get a starting job like the having contacts, and learning about the industry. However, the study stressed the importance of experience that supersedes all other factors. The study suggested that academic programs attempt to find such experience for their students (Harris et al., 2012).

Periodic review of the job market helps identify internship opportunities. It helps also reshape existing internship opportunities. Many employers advertise their need for interns in the job advertisement. Many other job opportunities can be better fulfilled through simple internship instead of committing a job for it. Jobs like contract programming, system conversion or training on new technology can be better fulfilled through internships if faculty knows about them. In other words, learning about the job market can uncover different potentials for internship opportunities to the students.

Employable Skills

Harris et al. (2012) conducted a study to assess the information systems job market in the 1970s and most recently. They reviewed job ads from careerberbuilder.com in five major cities Atlanta, Chicago, Los Angeles, Nashville and New York. The study found out that employers are asking for a longer lists of skills in their ads. For example, in 1978/1979, employers were asking for an average of 2.63 skills per job ads. In 2010/2011, employers are asking an average of 7.54 skills per job ads. This explains the diversity and multi skills that employers are asking more recently. The difference shifted to include more focus on experience that required for employment in recent years. This places more demand on educators to update their curriculum in a way that integrates valuable experience in the field.

Litecky, Arnett, and Prabhakar (2004) reviewed job search literature to see the different in requirements for soft skills versus hard skills. They came up with a two stage model for IT recruiting where in the first stage, the technical (hard) skills are examined first and in the second phase, more employers examine soft skill as a condition for continuation of hiring.

Huang, Kvasny, and Joshi (2009) studied job advertisements in three sources: scholarly articles, practitioner literature and online job advertisements and to find the skills required in IT field. They found that scholarly articles lag specifics about the requirements in the job market. They also found that employers emphasize soft skill more while the online job ads specify more specific technical skills. Technical skills are most prevalent in the jobs that were advertised in online job ads.

Litecky et al. (2010) completed a study titled "Mining for Computing Jobs" in which they retrieved about quarter of a million job ads and to find a pattern of what employers are looking for in terms of skill sets. They noted

> In any discipline, and especially in a discipline with a dynamic, highly competitive technology environment, professionals should periodically review the skills sets in high demand and identify industry trends in which their skill sets might be falling behind. Where downsizing and outsourcing is common, keeping up with current skill sets is critical (p. 83).

All the studies listed above indicate the need from IT professionals to understand the employable skills in this dynamic job market. This takes special importance for students because they seek these job opportunities. It will improve the chances of the students if they know the employable skills they need to have to help them compete in the market. Studying trends in the job market is one way of learning and understanding the employable skills.

Analyzing Computer Programming Jobs Using Web Data Mining

A faculty member at the department of Computer Science (COSC) at Indiana University of Pennsylvania (IUP) conducts periodic review of job market demand in the information technology field. This faculty has concentrated on programming jobs. He used a routine and a data mining technique where a batch file is regularly used to go through various steps to extract data from the job search web site. This section explains about the experience of this faculty with using web data mining to study the job market for computer programming jobs. It then draws conclusion about certain skills required in this field and suggest plans for future study.

The faculty in this study followed the "Keyword Indexing" technique for collecting data about programming job demands. He started by supplying specific keyword and finding the occurrences of the word in job ads in the dice.com web site over a span of several years. Dice.com is job site for technical jobs that was placed into operation in the 1990's. It has gained strong reputation as was rated the best job site by PC Magazine in 2008 (DeLeo, 2008). Today around 80,000 technical jobs are posted for positions throughout the United States.

Dice.com provides a number of different predefined searches including by skill, by company, and by locations, however, these searches were to coarse grain and limited in choice. Alternatively, Dice provides search by keywords. This was found to be useful in identifying jobs in which specific skills are required. Furthermore, the returned pages clearly identified the total number of jobs listings satisfying the query. Dice.com is thus an ideal site to mine the industry demands for these skills. These queries were automated as a batch process which for a set of technology skills would record the date of the query, keyword to identify a skill, and the total number of jobs found referencing the skill. A subsequent step provides averaging of data over the course of a month to smooth fluctuations occurring on a day to day basis. Lastly, a step enables extracting of data by select technologies for use in generation of graphs.

This technology was used to gather data for years 2005 – 2007 and 2012 – present. While a gap exists, the results clearly provide picture of useful trends that can guide the curriculum decisions and content for select courses. A sampling of the found results is given in the next section.

Results

The total graph in Figure 1 provides validation that the source of data for Web mining remains strong. Over the eight year period, Dice has consistently posted around 80,000 technical positions. It is interesting to note that there are slight dips during the December to February time periods. These dips must be taken into account when assessing data of specific technologies.

A major curriculum decision is the choice of programming language in the core computer science curriculum. While academia often leads in adoption of emerging languages, the core curriculum must consider the programming language as a key skill graduates must poses upon graduation. The remaining graphs of Figure 1 provide useful insight into the choice of language. Java is the language of highest demand by industry. Near twenty percent of all technical positions in Dice.com reference Java as a desired skill. While the explosive adoption of Java has subsided, demand for Java remains strong and continues to grow relative to the total number of jobs. In

contrast, the demand for COBOL over the same time period has dropped over half. Demand for skill in COBOL approach a half percent of all jobs. The languages of Python and Ruby are gaining in demand. These are languages that have been adopted in academia and are now gaining traction in industry. Curriculums that include these will undoubtedly be of benefit to graduates as the industry progresses.

Exceeding the demand for Java is the demand for skill with SQL as shown in Figure 2. Skill with SQL has grown from around eighteen percent to over thirty percent of all posted jobs. Given this it appears imperative that a database course with SQL as a major component be cornerstone of all computer science curriculum. Additional data for specific database management systems will help in selection of such systems within these courses. Oracle is found to be the dominant database management systems; however, SQL Server has made significant gains. Open source database, such as MySQL have gained demand; nonetheless, it would behoove a computer science department to adopt use of either of the leading database vendors as this would provide graduates with valuable experience.

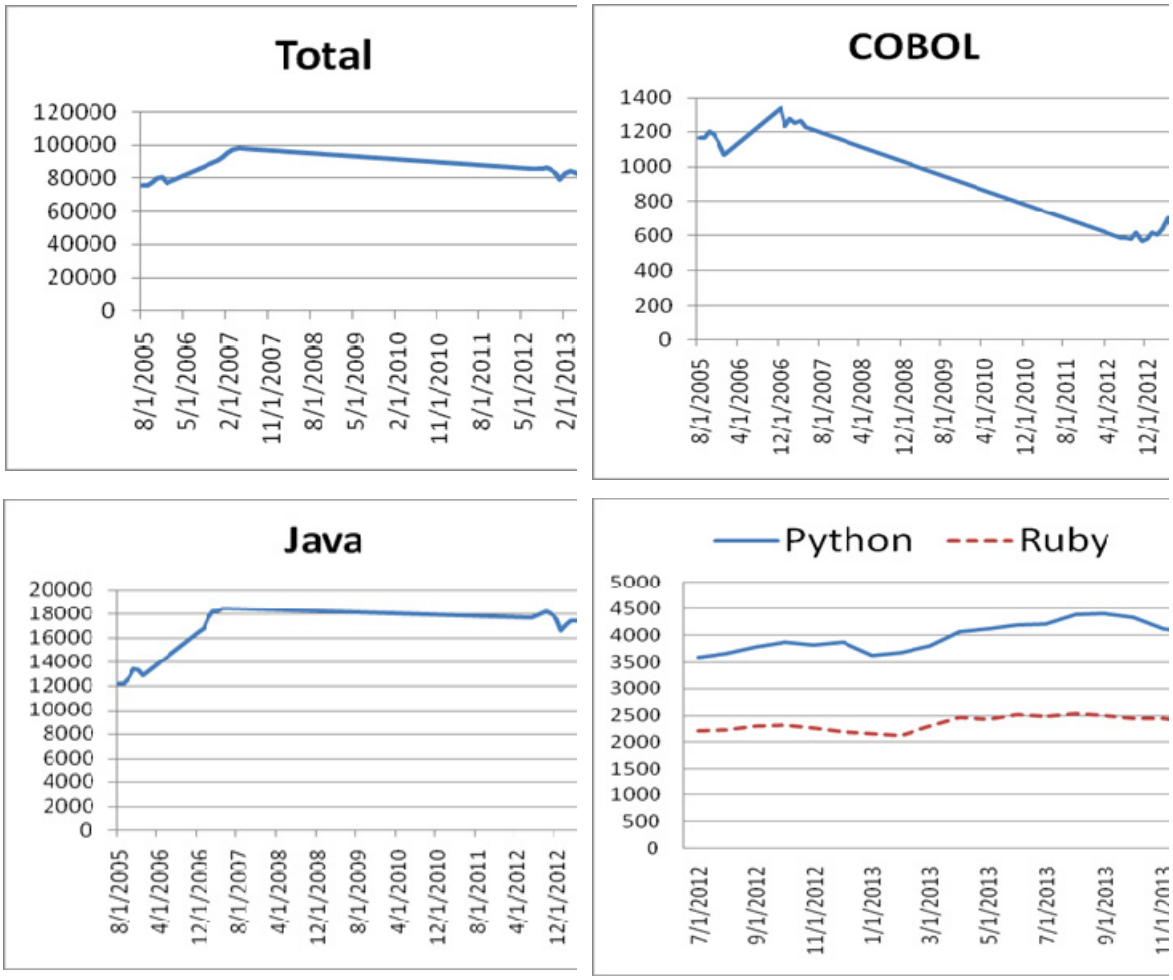

Figure 1 - Total Jobs on Dice.com and Jobs by Language

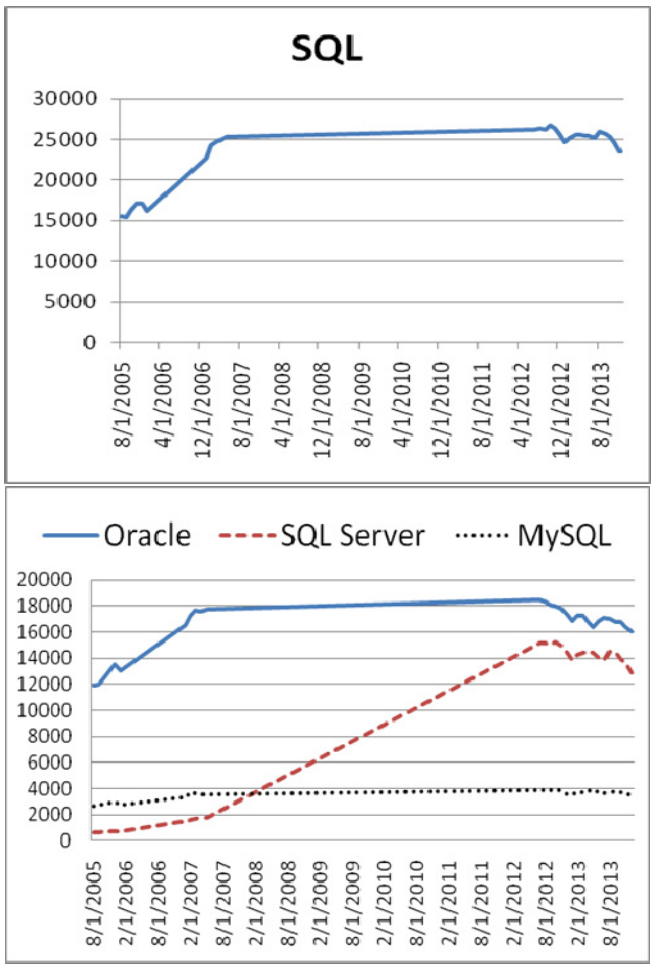

Figure 2 – SQL and Database Servers

Conclusions and Suggested Future Study

This study showed that the data mining techniques could be used by faculty and programs to work and meet the demands of Information Technology market. Through continuous data collections and through using the technology routine of data analysis, the faculty here was able to show the trend of programming jobs and will be able to provide the result of this analysis to the department for further action.

In addition to demonstrating the viability of the routines and analysis methods followed by this faculty, this study was able to make specific conclusions about the trend of market demand for programming jobs. Below is a list of specifics that we have reached after conducting this study:

- Java has been and remains the dominant programming language for near the last decade. Given this our core curriculum will continue to utilize Java.
- Demand for COBOL is clearly in decline. Due to other factors not described here, a decision to remove COBOL from the curriculum has not been reached. Ongoing collection and analysis of data using our techniques will be vital source for future decision.
- The demand for people with SQL knowledge is very strong. The results shown provide high validation of our choice to embrace Oracle in our database course.

In addition to the data listed here, data is being collected over a wide range of specific technical skills which have challenging dynamics. For example, in the area of web development new technologies; HTML5, JQuery, PHP, and so forth, are constantly emerging and others are in decline. Using our techniques we intend provide a more detail analysis of trends in these areas, this will be the subject of our future study.

References

Ali, A. & Kokun, F. (2010). The use of web log analysis in academic journals – Case study. *Issues in Information Systems, 12*(20) 612-619.

Ali, A. & Wood, D. (2004). Transition from teaching VB6 to VB.NET. *Issues in Information Systems,* 5, 1-7.

Berry, M. & Linoff, G. (2000). *Mastering data mining*. New York: Wiley Publishing.

Bhatt, G., Emdad, A., Roberts, N., & Grover, V. (2010). Building and leveraging information in dynamic environments: The role of IT infrastructure flexibility as enabler of organizational responsiveness and competitive advantage. *Information & Management, 47*(7-8), 341-349.

Burns, J. (2012, October 8). Top 10 Cloud Related Job Titles. *Network World*

Buzikashvili, N. (2007). Sliding window technique for the web log analysis. *Proceedings of the 16th international conference on World Wide Web*, 1213-1214. Retrieved May 13, 2010, from ACM Digital Library http://www.acm.org/dl

Corsini, P., & Marcelloni, F. (2006). A fuzzy system for profiling web portal users from web access log. *Journal of Intelligent & Fuzzy Systems, 17*, 503–516.

Das, R., & Turkoglu, I. (2009). Creating meaningful data from web logs for improving the impressiveness of a website by using path analysis method. *Expert Systems with Applications, 36*(3), 6635-6644.

DeLeo, J. L (January 22, 2008). 10 sites to help you land a tech job. *PC Magazine.*

Donohue, P., & Power, N. (2012, May). Legacy job titles in IT: The search for clarity. *Proceedings of the 50th Annual Conference on Computers and People Research* (pp. 5-10). ACM.

Gallivan, M., Truex III, D. P., & Kvasny, L. (2002, May). An analysis of the changing demand patterns for information technology professionals. *Proceedings of the 2002 ACM SIGCPR Conference on Computer Personnel Research* (pp. 1-13). ACM.

Harris, A. H., Greer, T. H., Morris, S. A., & Clark, W. J. (2012). Information systems job market late 1970's-early 2010's. *Journal of Computer Information Systems*, *Fall*, 72-79.

Huang, X. (2007). Comparison of interestingness measures for web usage mining: An empirical study. *International Journal of Information Technology & Decision Making. 6*(1), 15–41.

Huang, H., Kvasny, L., Joshi, K. D., Trauth, E. M., & Mahar, J. (2009, May). Synthesizing IT job skills identified in academic studies, practitioner publications and job ads. *Proceedings of the Special Interest Group on Management Information System's 47th Annual Conference on Computer Personnel Research* (pp. 121-128). ACM.

Litecky, C., Aken, A., Ahmad, A., & Nelson, H. J. (2010). Mining for computing jobs. *Software, IEEE, 27*(1), 78-85.

Litecky, C. R., Arnett, K. P., & Prabhakar, B. (2004). The paradox of soft skills versus technical skills in IS hiring. *Journal of Computer Information Systems, 45*(1).

McLean, C. (2006). A foot in the door: IT job-search strategies. *Certification Magazine, 8*(4), 38-40.

McGill, M. M. (2009, April). Defining the expectation gap: a comparison of industry needs and existing game development curriculum. *Proceedings of the 4th International Conference on Foundations of Digital Games* (pp. 129-136). ACM.

Mobasher, B. (2010). *Web data mining*. Retrieved May 10, 2010 from http://facweb.cs.depaul.edu/mobasher/classes/ect584/syllabus.html

Prabhakar, B., Litecky, C. R., & Arnett, K. (2005). IT skills in a tough job market. *Communications of the ACM, 48*(10), 91-94.

Rubin, J. H. (2004). Log analysis pays off. (weblogs). *Network Computing, 15*(18), 76-78.

Scott-Bracey, P. (2013). *Analyzing internet job advertisements to compare IT employer soft skill demand versus undergraduate IT program curriculum programs in Texas*. Retrieved 10/15/2013 from http://www.editlib.org/noaccess/38669

Shaw, M. J., Subramaniam, C., Tan, G. W., & Welge, M. E. (2001). Knowledge management and data mining for marketing. *Decision Support Systems, 31*(1), 127-137.

Shmueli, G., Patel, N. R., & Bruce, P. C (2010). *Data mining for business intelligence, concepts, techniques, and applications in Microsoft Office Excel with XLMiner*. Hoboken, New Jersey: Wiley & Sons.

Sobhi, M. S., & Son, B. G. (2010). Content analysis of OR job advertisements to infer required skills. *Journal of Operational Research, 61*(9), 1315-1327.

Surakka, S. (2007). What subjects and skills are important for software developers? *Communication of the ACM, 50*(1), 73-78.

Web Data Mining .net. (2010). *Web data mining*. Retrieved May 10, 2010 from http://www.web-datamining.net

Zhong, S. (2008). Information intelligent system based on web data mining. *Proceedings of the 2008 International Symposium on Electronic Commerce and Security*, 514-517. Retrieved May 14, 2010 from IEEE Computer Society Digital Library.

Biographies

David T. Smith, Ph.D., Associate Professor of Computer Science – Indiana University of Pennsylvania has 12 years experience in academia and 21 years of industry experience in database systems, computer language development, and other systems programming. He holds a bachelor degree in Physics and Mathematics Education from Indiana University of Pennsylvania, an M.S. in Computer Science from University of Central Florida, and a Ph.D. in Computer Science from Nova Southeastern University. Dr. Smith is active in consultancy and has research interests in artificial intelligence, distributed object computing, data mining, and software engineering.

Azad Ali, D.Sc., Professor of Information Technology at Eberly College of Business – Indiana University of Pennsylvania has 30 years of combined experience in areas of financial and information systems. He holds a bachelor degree in Business Administration from the University of Baghdad, an M.B. A. from Indiana University of Pennsylvania, an M.P.A. from the University of Pittsburgh, and a Doctorate of Science in Communications and Information Systems from Robert Morris University. Dr. Ali's research interests include service learning projects, web design tools, dealing with isolation in doctoral programs, and curriculum.