

WEB SCRAPING TO INFORM CURRICULUM DECISIONS: AN EXPLORATORY STUDY

Joseph H. Schuessler and Dennis Jones

Tarleton State University
Box T-0170
Stephenville, Texas 76402, USA
schuessler@tarleton.edu
djones@tarleton.edu

ABSTRACT

This study explores the use of web scraping software to examine key words on popular job hosting sites in order to inform curriculum decisions for a computer information systems department. By using common common job posting sites and web scraping software, this study demonstrates a methodology for gathering large amounts of data which when analyzed, can provide useful insights and aid in curriculum decisions. Data, once gathered, is analyzed through various text analysis techniques in order to identify patterns in the data. Results will help programs focus limited resources on curriculum that benefits a larger subset of their stakeholders.

INTRODUCTION

Public colleges and universities have a fiduciary responsibility when it comes to providing educational programming. Not only are students paying for a substantial portion of their education but taxpayers shoulder a significant portion of the expense through state spending on higher education. Other stakeholders such as area and regional employers have specific needs as it relates to the knowledge and skill sets of their employees, present and future. In an effort to align the production of students that graduate with such knowledge and skills with the needs of employers, it should not be surprising that governing bodies expect verifiable, even quantifiable results that can be used to measure the success or lack of success from such alignment.

A central tenet of many accrediting bodies, assessment processes are often required that help institutions demonstrate that they are hopefully achieving minimal levels of achievement and thus, demonstrating that their students are graduating with the necessary skills and knowledge to meet employer's demands. For example, AACSB for colleges of business, encourages institutions to directly measure student ability to effectively communicate both in writing and orally (CITE). SACSCOC has similar requirements (CITE). When students score highly on such assessments, the claim that their instruction has been effective can be made. However, when students perform poorly, the results can provide useful information to the program about how to go about improving student instruction. Traditionally, direct measures have been emphasized where direct measures included directly measuring artifacts that students produce such as presentations, research projects, etc. However, more recently, indirect measures have been more widely seen as a way to supplement the data gathered by institutions as a way to add additional

insight into student learning (AACSB, 2013). Rather than measuring the student artifacts directly, indirect measures can be used to measure students' opinions as well as others about the effectiveness of education at an institution. These may come in the form of alumni surveys, student polling, market analyses, etc. Such measures can be used to supplement direct measures in place as well as to glean additional insight into the effectiveness or lack thereof of an institution. For example, indirect measures can help gauge the value of a degree by identifying the average salaries of alumni or identifying the knowledge and skills that alumni required to be successful in their careers.

Use of alumni surveys in this manor can be quite useful but they are snapshots in time looking backwards. More timely data can help programs identify knowledge and skills that are currently in demand. This is where gathering data from online job sites can be useful. Such sites provide access to large amounts of semi-structured data. Stephens and Young (2016) demonstrated how large amounts of data on Google Scholar could be used to identify relevant literature on a given subject. Similarly, sites like Indeed.com, Dice.com, and so on can be useful for academic sites attempting to identify the latest knowledge and skills in demand in order to drive curriculum choices.

This research proposes to apply this technique to harvest data from indeed.com and dice.com, both sites that include a large number of jobs relevant to a computer information systems department. In order to methodically identify relevant jobs, the geographic scope will be limited to the nearest major metropolitan area and use search terms based on job titles from the Occupational Information Network (www.onetonline.org):

Code	Occupation
15-1121.00	Computer Systems Analysts
15-1122.00	Information Security Analysts
15-1131.00	Computer Programmers
15-1132.00	Software Developers, Applications
15-1133.00	Software Developers, Systems Software
15-1134.00	Web Developers
15-1141.00	Database Administrators
15-1142.00	Network and Computer Systems Administrators
15-1143.00	Computer Network Architects
15-1143.01	Telecommunications Engineering Specialists

15-1151.00	Computer User Support Specialists
15-1152.00	Computer Network Support Specialists
15-1199.00	Computer Occupations, All Other
15-1199.01	Software Quality Assurance Engineers and Testers
15-1199.02	Computer Systems Engineers/Architects
15-1199.03	Web Administrators
15-1199.06	Database Architects
15-1199.07	Data Warehousing Specialists
15-1199.08	Business Intelligence Analysts
15-1199.09	Information Technology Project Managers
15-1199.10	Search Marketing Strategists
15-1199.12	Document Management Specialists

References

- AACSB.edu. (2013). *Assurance of Learning Standards*. [online] Available at: <http://www.aacsb.edu/~media/AACSB/Publications/white-papers/wp-assurance-of-learning-standards.ashx> [Accessed 26 Sep. 2017].
- Stephens, P. and Young, J., Automating the Literature Review Process for Academic Research Studies, SWDSI 2016
- YouTube. (2013). *Glenn Gutmacher Outwit Hub Demo*. [online] Available at: https://www.youtube.com/watch?v=VSIU_KpxbIo&feature=youtu.be [Accessed 26 Sep. 2017].