Web Scraping to Inform Curriculum Decisions

James S. Norcross
Tarleton State University

Joseph H. Schuessler
Tarleton State University

Dennis G. Jones
Tarleton State University

This study explored the use of web scraping to examine key words on a popular job hosting site in order to inform curriculum decisions for Computer Information Systems programs. By using a job posting site and web data extraction, this study demonstrated a methodology for gathering large amounts of data and its applicability in evaluating curriculum. When the data was analyzed, it provided useful insights that could aid in curriculum design. Data was analyzed through various text analysis techniques to identify data patterns. Results will help programs focus limited resources on curriculum that benefits a larger subset of their stakeholders.

Keywords: computer information systems, curriculum design, data mining, higher education information technology, text analysis, web crawling, web scraping

INTRODUCTION

Public colleges and universities have a fiduciary responsibility when it comes to providing educational programming (Schmidt, 2014). 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 related to the knowledge and skill sets of their employees, both present and future (Schmidt, 2014). In an effort to align the production of students who graduate with the knowledge and skills that meet the needs of employers, governing bodies expect quantifiable results that can be used to measure alignment success.

BACKGROUND

As a central tenet of accrediting bodies, assessment processes are often required to help institutions demonstrate that their students are achieving minimal levels of learning to meet employers' demands (White, Longenecker, McKell, & Harris, 2008). For example, the Association to Advance Collegiate Schools of Business (AACSB), encourages institutions to directly measure student ability to effectively communicate both in writing and orally (Attaway, Chandra, Dos Santos, Thatcher, & Wright, 2011). The

Southern Association of Colleges and Schools Commission on Colleges (SACSCOC) has similar requirements (White et al, 2008). 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 improve student learning.

Traditionally, direct measures involve assessing artifacts that students produce (e,g., presentations, research projects, digital media). However, more recently, indirect measures have been more widely utilized to supplement the data gathered by institutions to add additional insight into student learning (Attaway et al., 2011). Rather than measuring the student artifacts directly, indirect measures can be used to measure students' opinions, as well as the opinion of others about the effectiveness of education at an institution. These may come in the form of (a) alumni surveys, (b) student polling, (c) market analyses, and other techniques (Lopez, 2002). Such measures can be used to supplement direct measures, as well as, to glean additional insight into the effectiveness of an institution (Palomba & Banta, 2015). 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 needed to be successful in their careers and validate appropriate alignment with curriculum.

As outlined above, collection of indirect data can come from a variety of sources. One potential approach, web scraping, is the automated and targeted collection of Internet data, that can be used for a variety of purposes (Raicu, 2019). Web scraping is different from data mining in that web scraping involves the collection of data based on search words, while data mining seeks to identify emerging patterns in large data sets (Adelola, Adewale, and Iwasokum, 2016; Raicu, 2019). Web scraping involves the collection of data and placing such data in a readable format, such as in a database or an Excel worksheet. This makes the data easier to analyze and sort for research.

According to Adelola et al., (2016), the Internet is an unlimited source of data for any type of information. This information can range from various business and political subdivision reports, social information, medical data, educational sites, and even employment data. The use of data mining and web scraping can result in an abundance of information; the difficult part is sorting through everything collected to glean a meaningful repository of data that can be analyzed (Adelola et al., 2016). One of the inherent issues with data mining or web scraping is that they collect all data that may contain the search criteria. However, these techniques are unable to evaluate the semantic intent of the information.

In addition to collecting data that is difficult to analyze, Kimmons (2017) emphasized the need for data collection to be handled in an ethical and legal manner. If a web scraping engine happened to bypass certain security protocols, the collection and use of copyrighted information without permission, or the collection and use of private information, could lead to severe consequences (Kimmons, 2017). Many Internet sources prohibit web scraping engines from accessing their sites in order to protect information and to limit the use of server bandwidth, so that individuals may still be able to utilize a site (Kimmons, 2017).

STATEMENT OF THE PROBLEM

Alumni surveys can be quite insightful but they are just 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 (2017) demonstrated how large amounts of data on Google Scholar could be used to identify relevant literature on a given subject. Similarly, in order to drive curriculum choices, sites like Indeed.com, Dice.com, and many others, can be beneficial for academic organizations attempting to identify the latest knowledge and skills in demand. The focus of this data collection can be narrowed or broadened as desired to gather relevant material for assessment.

REVIEW OF RELEVANT RESEARCH LITERATURE

As a research tool, (a) web scraping, (b) web crawling, and (c) data mining are becoming more common and being used for various projects in business and technology (e.g., Adelola, Adewale, & Iwasokum, 2016;

Kimmons, 2017; Raicu, 2019; Suhirman, Zain, Chiroma, & Herawan, 2014; Moretti, Brenes, & McKnight, 2014) . However, limited research was found related to using web scraping for the purposes of evaluating and updating higher education curriculum compared to the present job market needs. Therefore, a broadened approach was taken in evaluating existing literature regarding web scraping data collection as a whole; most particularly, in evaluating its current usefulness and limitations. In doing so, a methodology for data collection using web scraping was developed. This process provided for an opportunity for consistent data analysis that would be useful in determining whether current curriculum requirements met the needs of employers.

According to Suhirman, Zain, Chiroma, and Herawan (2014), the use of data mining or web scraping in education can no longer be considered a new methodology, but one that can render significant results in curriculum design if its use would only be embraced by researchers in higher education. However, it is a significantly underused methodology in curriculum design research (Moretti, Brenes, & McKnight, 2014).

METHODOLOGY

Using Outwit Hub (https://www.outwit.com), a web scraping utility engine, to apply web scraping techniques, data was harvested from Dice.com. Dice.com includes numerous jobs relevant to the field of Computer Information Systems (Roberts, 2013). In order to identify relevant jobs, the geographic scope was limited to the nearest major metropolitan area (i.e., Dallas/Fort Worth, Texas) and used search terms based on job titles from the Occupational Information Network (https://www.onetonline.org).

To more methodically evaluate job titles, the O-Net Online database (United States Department of Labor, 2018) was used to identify relevant job titles. Within the O-Net database, a broad search was performed for the career field of "information systems" which returned more than 800 available job titles. Twenty-two of the most relevant job titles were identified and selected to be used as OutWit Hub web scraping search terms. These search terms are identified in Table 1. The terms were input into the OutWit Hub web scraping engine and configured to "scrape" https://www.dice.com for matching jobs in the Dallas/Fort Worth metroplex area. The collection was performed over a 48-hour period in February 2018. There were not any significant complications encountered during the web scraping search, and no additional search terms were utilized.

TABLE 1
O*NET ONLINE SEARCH TERMS

O*Net 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

RESULTS

The OutWit Hub web scraping engine collected targeted semi-structured data including job and employer information: (a) source URL, (b) job title, (c) company, (d) city, (e) state, (f) postal code, (g) latitude, (h) longitude, and (i) job description. The job description included a variety of information including required qualifications for the job. Initial results totaled more than 11,000 individual job using all 22 occupation search terms. However, it was immediately noted through an initial, cursory review that many of the results were duplicates of the same advertised job. All results were combined in a single Excel worksheet and configured to remove all duplicate entries, thus reducing the results to 1,072 unique job postings. Initial categories of data returned were then reduced to only include (a) job title, (b) company, (c) city, (d) postal code, and (e) description with qualifications. This removed extraneous data and allowed for a more targeted mixed methodology analysis of the data.

QUALITATIVE ANALYSIS

Educational requirements were analyzed for degrees in (a) Computer Science, (b) Information Systems, or (c) another comparable educational accomplishment. Specialized certification requirements were also observed in numerous job postings. Another significant finding was the requirement for the applicant to have working knowledge of several, different computer programming languages (e.g., C, C++, C#, Python, HTML, Java, & Unix-based languages). Many of the advertised positions described jobs as having multiple responsibilities in other disciplines which resulted in the same job posting being included in multiple search results of the 22 occupation terms used in the web scraping search. As a result, it was decided to combine the results into 10 generalized job categories (see Table 2). The combining of these job categories were based on analyzing each of the 1,072 job postings, and combining the categories based on the job descriptions. Additionally, the research team relied on their combined working experience in the information systems field, to determine appropriate classifications of various postings into a single job classification.

TABLE 2
GENERAL COMPUTER INFORMATION SYSTEMS JOB CATEGORIES

CIS Job Categories	
Business Intel/Marketing/Documents	Cloud/Server/Database Admin
Computer/System Support	Cybersecurity & Disaster Recovery
Network Engineer/Architect/Admin	Telecommunications Engineering
Project Management/Leadership	Web Admin/Development
Software Development/Programming	Other Computer Occupations

Each of the 1,072 returns were then categorized into one of the 10 job categories by re-analyzing each job description and its respective requirements. Each job category was represented by a separate, individual Excel worksheet in which all job information was entered and tabulated for qualitative analysis (see Table 3).

TABLE 3
JOB CATEGORY TABULATION

JOB CATEGORIES	NUMBER OF JOBS
Software Development & Programming	286
Cloud/Server/Database Administration	251
Project Management\Leadership	169
Business Intel/Marketing/Documents	144
Computer/System Support	62
Cybersecurity & Disaster Recovery	55
Web Administrators & Development	49
Network Engineering/Architecture/Administration	38
Other Computer Occupations	10
Telecommunications Engineering	8

QUANTITATIVE ANALYSIS

Descriptive statistics were used to analyze the percentage of the job market for each of the 10 general job categories was computed. This analysis was used to evaluate the current undergraduate curriculum in the Computer Information Systems programs at Tarleton State University. The quantitative analyses revealed that (a) Software Development and Programming, (b) Cloud, Server, and Database Administration, (c) Project Management and Leadership, and (d) Business Intelligence, Marketing, and Document job positions were the most prevalent employment opportunities in the DFW metroplex (see Table 4).

TABLE 4
GENERAL JOB CATEGORIES QUANTITATIVE RESULTS

JOB CATEGORIES	NUMBER OF JOBS	PERCENT OF MARKET
Software Development & Programming	286	26.7%
Cloud/Server/Database Administration	251	23.4%
I.T. Project Management\Leadership	169	15.8%
Business Intel/Marketing/Document	144	13.4%
Computer/System Support	62	5.8%
Cybersecurity & Disaster Recovery	55	5.1%
Web Administrators & Development	49	4.6%
Network Engineering/Architecture/Administration	38	3.5%
Other Computer Occupations	10	0.9%
Telecommunications Engineering	8	0.7%
Total Jobs and Percent	**1072**	**100.0%**

An analysis was then performed to determine word count frequency using https://wordcounter.net to determine (a) job requirements (b) experience, (c) education, and (d) specialized certifications for candidates to be competitive for hiring consideration of the jobs advertised. Of these four areas, experience appeared to be the most required qualification for the majority of the jobs listed. Education was second, and specialized certifications was third (see Table 5).

TABLE 5
EXPERIENCE, EDUCATION, CERTIFICATIONS REQUIRED

QUALIFICATIONS REQUIRED	NUMBER OF JOBS	PERCENT
Experience	1012	94.40%
Education	853	79.57%
Certifications	799	74.53%

Each of these required qualifications were examined more closely, beginning with the experience requirement. Most of the job positions advertised indicated a need for approximately three years of experience, with the second most common time requirement being 5 years. The remaining of the advertised positions requiring experience for the job listing did not specify any particular time frame (see Table 6).

TABLE 6
EXPERIENCE REQUIRED

EXPERIENCE REQUIRED	NUMBER OF JOBS	PERCENT
3	480	44.78%
5	478	44.59%
Experience (time unspecified)	54	5.04%
Experience required for job Total	1012	94.40%

Education was the next qualification examined to determine what employers desired in a qualified candidate. The majority of hiring companies stated a requirement for a college degree, but did not specify the level (i.e. Associates, Bachelors, Masters) of the degree required; only that the candidate graduated from a college or technical school. Bachelor degrees were the next most sought after education qualification; followed lastly by a very low percentage of employers seeking a candidate with a Masters degree. There were not any jobs listed that required a doctorate (see Table 7).

TABLE 7
EDUCATION REQUIRED

EDUCATION REQUIRED	NUMBER OF JOBS	PERCENT
Degree required (level not specified)	478	44.59%
Bachelors Degree required	359	33.49%
Masters Degree required	16	1.49%
Education required for job Total	853	79.57%

Lastly, all collected job postings were reviewed for required certifications in the information systems field. Several certifications were mentioned more consistently across various job postings. The top three information systems certifications observed were through (a) Cisco Systems, (b) Amazon, and (c) Microsoft. Due to the significant number of highly specialized certifications, the majority of them were grouped according to either a common platform or organization offering the certification. Altogether, this listed 24 certifications or certification organizations that were desired by employers (see Table 8).

TABLE 8
CERTIFICATIONS REQUIRED

CERTIFICATIONS REQUIRED	NUMBER OF JOBS	PERCENTAGE
Cisco (i.e. CCNA, CCNP, CCIE)	174	16.23%
Amazon AWS	153	14.27%
Microsoft (i.e. Azure, MCSE, MCSA, MCITP)	146	13.62%
VMware (i.e. VCA, VCP)	56	5.22%
Agile Certifications (i.e. PMI-ACP, Scrum, Kanban)	51	4.76%
ITIL	49	4.57%
Splunk	23	2.15%
ISC2 (i.e. SSCP, CISSP)	23	2.15%
EC-Council (i.e. CCISO, CEH)	20	1.87%
ISACA (i.e. CISM, CRISC, COBIT)	16	1.49%
LINUX (i.e. RedHat, Centos, Oracle, LPIC)	15	1.40%
Symantec	12	1.12%
Juniper (i.e. JNCIA, JNCIS, JNCIP)	10	0.93%
GIAC certifications	9	0.84%
TOGAF	7	0.65%
Six Sigma (ranking not specified)	7	0.65%
OCSA (penetration testing certification)	7	0.65%
Palo Alto Networks (i.e. PCNSA, PCNSE)	6	0.56%
OSCP (penetration testing certification)	5	0.47%
CompTIA (i.e. Network+, Security+, Project+)	4	0.37%
Project Management Institute (i.e. PMP, CAPM)	2	0.19%
BICSI (i.e. RCDD)	2	0.19%
Aruba (wireless networking certification)	1	0.09%
Certified Wireless Network Professionals	1	0.09%
Total Jobs requiring a certification	**799**	**74.53%**

During the quantitative analysis of the collected data, it was observed that many of the employers were additionally requiring the knowledge and ability to use certain computer programming and/or platform languages. An analysis of the listed languages was performed with the assistance of the word count frequency software at https://wordcounter.net. This was done to determine the frequency in which each type of computer programming or platform language was mentioned collectively in all 1,072 job descriptions analyzed. Based on the particular languages' frequency of use in the collective job descriptions, it was assumed that this could be interpreted as the level of desire by a future employer for a candidate to have proficiency in utilization. Seventeen different computer programming and platform languages were identified and ranked according to frequency of name occurrence in the collective job descriptions. Java-based programming languages were noted to be the most frequently mentioned language, followed by SQL-based languages (see Table 9).

TABLE 9
PROGRAM/PLATFORM LANGUAGES REQUIRED

PROGRAM/PLATFORM LANGUAGES	FREQUENCY	PERCENTAGE
Java-based programming languages	783	26.96%
SQL-based programming languages	675	23.24%
Unix & Linux	324	11.16%
HTML	206	7.09%
C-based programming languages	194	6.68%
Python	118	4.06%
CSS	115	3.96%
Hadoop	100	3.44%
XML	76	2.62%
PHP	50	1.72%
Apache	41	1.41%
Powershell	40	1.38%
PERL	29	1.00%
AJAX	23	0.79%
CMS	22	0.76%
COBOL	9	0.31%
ESSCMD	2	0.069%

Comparative Analysis of Results With Current Curriculum

The results from the analyses of jobs posted were then evaluated relative to the programs offered in each of the three available undergraduate programs in Computer Information Systems (CIS) at Tarleton State University (Tarleton). The programs at Tarleton include the Bachelor of Science (BS) in CIS, the Bachelor of Business Administration (BBA) in CIS, and the Bachelor of Applied Arts and Sciences (BAAS) in Information Technology (Tarleton State University, 2019). While each program caters to a unique target market, table 10 below includes the core of five courses common to all three programs.

TABLE 10
BBA/BS/BAAS COMMON COURSES

Class	Number	Class Name
BCIS	3332/3	Java/C# Programming
BCIS	3347	Data Communications
BCIS	3389	System Analysis and Design
BCIS	4301	Database Theory and Practice
BCIS	4350	Management Information Systems

Review of the five core courses in comparison with the data collected and examined for this study, suggests a good academic foundation in the core knowledge needed in the information systems field. A foundation in a commonly required programming language is required in the curriculum and this reflects

the most common requirement observed in the majority of the job postings. The remaining required core courses appear to support student learning in the job professions of software development or database administration (Tarleton State University, 2019). Both of these professions were previously mentioned to be the two most sought for job applicants in the DFW market. What is important to take note, is in the qualitative analysis of the majority of the job postings reviewed, knowledge of security practices appeared to be a regular requirement. However, in reviewing the core curriculum, management information systems is the only class that mentions a theoretical discussion in information systems security (Tarleton State University, 2019). It may be worth considering expanding the core curriculum to include a sixth required course in information systems security. Lastly, the third most advertised job posting in the region of this study is project management. However, this is not covered in the core curriculum either, suggesting an additional course needs to be added to this curriculum or to one of the three focused degree programs in CIS.

While each program includes the common courses described above, each program also includes unique coursework designed to address the unique interests of students and to better prepare them for their future career goals. The BS program is a highly technical program, designed for students that want a deeper understanding of technology and how it can be applied within organizations to create competitive advantage (see Table 11).

TABLE 11
UNIQUE B.S. REQUIRED COURSES

Class	Number	Class Name
BCIS	1305	Business Computer Applications
BCIS	1315	Principles of Web Design
BCIS	1317	Personal Computer Maintenance and Hardware
BCIS	3342/3	Advanced Java/Advanced C# Programming
BCIS	Electives	21 Hours Advanced

As can be immediately seen, the curriculum for the B.S. CIS program requires the student to continue with an advanced course in a popular programming language. A web design course is added to this program of study, which may lead to academic experience with other programming languages relative to this specialty. Additional courses are noted in computer applications, hardware, and maintenance, yet none of the course descriptions mention any experience with information security (Tarleton State University, 2019). While the program curriculum allows for the student select up to 21 hours of elective courses, there is no guarantee that this will include an information system security-focused course. Thus, leaving the graduating student with an underemphasized or missing critical skill. Lastly, the third most popular job postings in the region of this study was project management. Yet there are no courses listed in this curriculum to fill that knowledge gap. While this is another course that can be selected by the student as an elective, it is not a program requirement and therefore, a possible missing link in the student's complete education.

The required curriculum was evaluated for the Bachelor of Business Administration (B.B.A.) degree, specializing in CIS (see Table 12). This program targets students seeking a balanced business and technical background that could position a graduate for advancement opportunities within an organization.

TABLE 12
UNIQUE B.B.A. REQUIRED COURSES

Class	Number	Class Name
BCIS	3342/3	Advanced Java/Advanced C# Programming
BCIS	4355	Global Information Systems
BCIS/COSC	Electives	15 Hours (12 Advanced)

Like the previously discussed curriculums, the B.B.A program continues with an advanced programming language requirement that this study has suggested would be beneficial for a student. A new course that discusses information systems from a global perspective is added to the curriculum, that may be beneficial for those who may work in the international job market (Tarleton State University, 2019). However, the basic education in project management and information system security is still left to the student to select as an elective in this degree program instead of being required. Additionally, the student may select their electives not only from the advanced computer information systems courses, but also from the advanced computer science courses. This increasing the opportunity for specialized elective development, diminishes the likelihood that a student may select the needed course in information system security or project management.

Lastly, the Bachelor of Applied Arts and Sciences (B.A.A.S.) in Information Technology (I.T.) was evaluated. The B.A.A.S. program is a degree completion program that allows students to apply relevant work experience, training, and certification for 12-33 hours of credit. Ideal for students with work experience and those with Associate degrees, this degree program includes significant differences in the curriculum when compared to the previous two programs (see Table 13).

TABLE 13
B.A.A.S. INFORMATION TECHNOLOGY CURRICULUM

Class	Number	Class Name
BCIS	3332/3	Java/C# Programming
BCIS	3347	Data Communications
BCIS	3389	System Analysis and Design
BCIS	4301	Database Theory and Practice
BCIS	4350	Management Information Systems
BCIS	4359	Strategic Application of Information Systems
BCIS	Electives	9 Hours Advanced
	Electives	15 Hours Advanced may come from BCIS, ARTS, COSC, ENGT, ENGR, ECON, ACCT, BLAW, FINC, MGMT, BUSI, and/or MKTG
	Electives	0-21 Hours from BCIS, ARTS, COSC, ENGT, ENGR, ECON, ACCT, BLAW, FINC, MGMT, BUSI, MKTG (number of hours will depend on how many hours awarded for transfer credit or previous technical training)

(Tarleton State University, 2019).

The B.A.A.S. program does maintain a basic programming language course, along with the core curriculum first presented in this section. It is noted that project management and information system security courses remain absent in the curriculum. While this particular degree program is designed to allow credit transfer, there is no guarantee that the student has had the previously needed education in these missing areas. Twenty-four hours of elective course are available to fill this gap, but this continues to gamble on the student making appropriate decisions that will help them in the future job market.

DISCUSSION

Major contributions of this study were to demonstrate an innovative way to collect relevant employment data and to use those results to inform curriculum decisions. All of the curriculums reviewed appear to be well developed. Some adjustments could be made to better align with the findings of this study. All of the curriculums ensure a programming language requirement is included in the requirements and this coincides well with the study results. Further, every curriculum plan had at least a basic education in database administration coinciding with the second most sought for job position in the DFW area. However, it was mentioned in the qualitative analysis that most of the job postings wanted the applicant to be able to utilize current security practices as part of their responsibilities. However, no information system security courses are required in any of the curriculum designs. Also, project management was observed to be the third most common posted job in the region. Yet, no project management courses were required in any of the curriculums reviewed.

What was important to mention is the elective hours available in each degree program that would allow students to select courses in areas of interests. This would also allow students to select courses in areas of needed development, that have not been assigned as required in the existing degree plans. This may be an area in which Tarleton's curriculum design team may have to consider adding required courses. Another approach may consider conducting a search similar to the one performed in this study and posting the results for students to help guide their own curriculum decisions.

LIMITATIONS OF THIS STUDY

The research in this study is limited to the geographical region and job market for the Dallas/Fort Worth metroplex in north Texas. While this region may initially appear large in scope, it is limited as it cannot necessarily predict the same results that may be obtained in other regions of Texas or the United States in general. Further, the web scraping collection was performed over a two-day period at the end of February in 2018, representing a snapshot in time. Subsequent data collection did not occur thus preventing any analysis of changing job requirements over time. Additionally, the information collected is static and does not lend itself to ascertaining any trends in the job market for the limit region for other times in the course of the same year or years following. Lastly, the data collected yielded a .csv document that contained more 300,000 words which estimates to approximately almost 30 hours of reading time required to review the material. With this mind, the possibility of human error is plausible in manually reviewing the raw data collected for this research.

CONTRIBUTION OF THIS RESEARCH STUDY

This study contributed two very important findings that should be expanded upon in further research later. The first contribution was that it demonstrated an efficient way of gathering large quantities of relevant, semi-structured data for examination. Secondly, it provided an effective method to validate or identify curriculum alignment issues relative to current and future job market prospects for future graduates.

RECOMMENDATIONS FOR FURTHER STUDY

This initial study shows promise in developing an innovative approach to collect relevant job data that can inform curriculum design. Further study should be pursued and extended by first using the data collected from this study, analyze the graduate curriculums in the computer information systems and computer science programs. Secondly, repeating similar studies for the DFW region at different times to document and evaluate dynamic employment trends. Lastly, replicate study in other regions of Texas and in other regions of other states in the United States and compare results.

CONCLUSION

As previously mentioned, Alumni surveys and curriculum completion assessment have been widely used in academia to assess the quality of programs in many institutions of higher education. These surveys not only provide useful data for curriculum decisions but is encouraged and required by many accreditation bodies in higher education. Web scraping is receiving attention as a viable method to collect subject data in discipline research, but very little emphasis has been mentioned about using this method to assess education curriculum quality. The results of this study have demonstrated significant potential of web scraping as a useful research tool to ascertain job qualification requirements in comparison to curriculum design. Web scraping used together with survey and assessment methodologies, show promise of ensuring higher education institutions are meeting the needs of today's students and their future employers.

REFERENCES

Adelola, M.A., Adewale, O.S., & Iwasokun, G.B. (2016). Graph theoretic and genetic algorithm-based model for web content mining. *International Journal of Computer Science Issues*, *13*(6), 68-75. https://doi.org/10.20943/01201606.6875

Attaway, A.N., Chandra, S., Dos Santos, B.L., Thatcher, M.E., & Wright, A.L. (2011). An approach to meeting AACSB assurance of learning standards in an IS core course. *Journal of Information Systems Education*, *22*(4), 355-367. Retrieved from https://jise.org/

Gheorghe, M., Mihai, F-C., & Dardala, M. (2018). Modern techniques of web scraping for data scientists. *Romanian Journal of Human-Computer Interaction*, *11*(1), 63-75. Retrieved from http://rochi.utcluj.ro/rrioc/en/

Kimmons, R. (2017). Open to all? Nationwide evaluation of higher-priority web accessibility considerations among higher education websites. *Journal of Computing in Higher Education, 29*(3), 434-450. https://doi.org/10.1007/s12528-017-9151-3

Lopez, C.L. (2002). Assessment of Student Learning: Challenges and Strategies. *Journal of Academic Librarianship*, *28*(6), 356–367. https://doi-org.zeus.tarleton.edu/10.1016/S0099-1333(02)00345-2

Moretti, A., González-Brenes, J.P., & McKnight, K. (2014). Data-driven curriculum design: Mining the web to make better teaching decisions. *Proceedings of the 7th International Conference on Educational Data Mining (EDM 2014)*. Retrieved from https://files.eric.ed.gov/fulltext/ED558339.pdf

OutWit Technologies. (2019). OutWit Hub: Your own web collection engine. Retrieved from https://www.outwit.com/products/hub/

Palomba, C., & Banta, T. (2015). *Assessment Essentials: Planning, Implementing, and Improving Assessment in Higher Education* (2nd ed.). San Francisco, CA: Jossey-Bass.

Raicu, I. (2019). Financial banking dataset for supervised machine learning classification. *Informatica Economica*, *23*(1), 37-49. https://doi.org/10.12948/issn14531305/23.1.2019.04

Roberts, J. [ERE Media Videos]. (2013, December 12). *Glenn Gutmacher Outwit Hub Demo*. [Video file]. Retrieved from https://www.youtube.com/watch?v=VSIU_KpxbIo

Schmidt, B. (2014). *Governance for a New Era: A Blueprint for Higher Education Trustees*. American Council of Trustees and Alumni. Retrieved from https://www.goacta.org/images/download/governance_for_a_new_era.pdf

Stephens, P.R., & Young, J.A. (2017, March). *Automating the Literature Review Process for Academic Research Studies*. Southwest Decision Sciences Institute, Little Rock, Arkansas. Retrieved from http://www.swdsi.org/proceedings/ 2017/Papers/Papers/AB053.pdf

Suhirman, S., Herawan, T., Chiroma, H., & Zain, J.M. (2014). Data mining for education decision support: A review. *International Journal of Emerging Technologies in Learning*, *9*(6), 4-19. Retrieved from https://doi.org/10.3991/ijet.v9i6.3950

Tarleton State University. (2019). *2019-20 Catalog Marketing and Computer Information Systems*. Retrieved from http://catalog.tarleton.edu/undergrad/ collegeofbusinessadministration/marketingandcomputerinformationsystems/#text

Wang, Y. (2016). Big opportunities and big concerns of big data education. *TechTrends*, *60*(4), 381-384. Retrieved from https://doi.org/10.1007/s11528-016-0072-1

White, B., Longenecker, H., McKell, L., & Harris, A. (2008). Assessment: Placing the emphasis on learning in information systems programs and classes. *Journal of Information Systems Education*, *19*(2), 165-168.