

# Are Analytics Skills in Curriculum Aligned with Skills in Job Description? A Knowledge Graph Approach

Emma Cherrington<sup>1</sup> and Yong Seog Kim<sup>2\*</sup>

<sup>1</sup>Senior Data Analyst, Revere Health, USA

<sup>2</sup>Professor, Data Analytics and Information Systems Department, Utah State University, USA

\*Corresponding Author: Yong Seog Kim

Received: 30-11-2024

Revised: 14-12-2024

Accepted: 30-12-2024

## ABSTRACT

Recently, many academic institutions in the Intermountain West Region of USA update their traditional IS programs with various data analytics and data sciences courses. To evaluate whether newly incorporated contents are appropriate to prepare students as data scientists or data analysts, we compare core analytical skills emphasized in current data analytics and information systems (DAIS) curriculum with analytical skills in 11,225 job advertisements for data analysts or scientist positions. We found that job descriptions source presents the importance of application of specific skills to various business settings while curriculum source focuses on the conceptual and theoretical skills along with fundamental skills such as SQL. Based on our findings, we suggest that courses should address not only technical skills, but also soft skills to make informed decisions and turn them into business decisions through real application scenarios.

**Keywords**— Knowledge Graph, Data Analytics Curriculum, Data Scientist Job Description

## I. INTRODUCTION

Recently, many Information and Communication Technology (ICT) or Management Information Systems (MIS) program changes its name as Data Analytics and Information Systems or other similar names (denoted as DAIS from now on for notational convenience) with newer curriculum focusing on data analytics. Replacing old legacy technology platforms and associated educational curriculum is considered very necessary to improve its own sustainability and successful career planning for students. However, there have been disputes over the right direction of changes that would determine core teaching curriculum and research themes (Agarwal and Lucas 2005; Benbasat and Zmud 2003; Müller et al. 2016).

For example, one group of DAIS scholars claims that DAIS research community has focused and should devote more times and efforts to research and teaching on managerial, methodological, and how individuals and markets interact with IT-based systems (Benbasat and

Zmud 2003). In contrast, another stream of IS scholars insists that IS research community should stay on more macro studies of IT to capture the transformational power of the technology (Agarwal and Lucas 2005). Another group of IS scholars argues that DAIS discipline should adopt digital technology disruption as an inevitable phenomenon and develop core teaching curriculum centered around emerging digital technologies so that students acquire appropriate technical skill sets for new business models, products, and services upon their graduation (Case et al. 2019; Müller et al. 2016).

In particular, the pervasive expansion of data volume due to the proliferation of the web, social media, and mobile devices brings topics such as data mining and data analytics to IS educators as being the next big thing (Jafar et al. 2017). In fact, during the past two decades, data mining and data analytics courses in IS curriculum has evolved from appearing as an elective course to a minor area of study, a co/dual-major, or even as a fully-independent degree program under various names of data mining, data science, data analytics, business analytics or business intelligence. While some business analytics or business intelligence courses tend to emphasize the application of data analysis to solve business problems by revealing patterns and extracting insights from the data, other courses focus on statistical, mathematical or algorithmic properties of data analysis.

These updates seem to be partially successful in a sense that a growing number of students are interested in these courses. However, the authors believe that the innovation of DAIS curriculum should be evaluated not only by measuring whether new technologies and principles are incorporated into a new DAIS curriculum but also by measuring whether newly incorporated contents in a DAIS curriculum are appropriate in terms of depth and breadth for practical application of such skills.

To this end, in this paper, we like to measure whether newly incorporated contents in data analytics curriculum are aligned with highly demanded skills from industrial organizations. Specifically, we aim to identify core analytical skills that current DAIS curriculum intends

to offer for future data scientists or data analysts from course descriptions. At the same time, we also like to identify core analytical skills that industrial employers identify critical for positions as data scientist or data analyst from job descriptions. Once such skill sets are identified, we like to compare them to see if there exists mutual consensus or disparity between skill sets obtained from course and job descriptions and make recommendations for changes in DAIS curriculum to bridge the gaps if necessary.

## II. LITERATURE REVIEW

The goal of this study is to utilize a knowledge graph (KG) to explore its potential within the educational domain. The KG is a semantic network in which entities are denoted by nodes and relations between pairs of entities are represented by edges (Bordes et al. 2014; Chen et al. 2018; Paulheim 2016). In essence, a KG is a multi-relational graph composed of entities (nodes) and relations (different types of edges), where each edge is represented as a triple of the form (source entity, relation, target entity) called a fact. Its notion has garnered a great deal of attention since Google introduced its own version of a KG to improve the value of search results in 2012. Since then, a number of KGs including Yahoo's Spark, Microsoft's Satori and Facebook's entity graph have been applied to voice assistants (Singh et al. 2018), information integration and extraction (Pfaff et al. 2018) and named entity disambiguation (Lan et al. 2016). Several other notable KGs have been proposed for intelligent recommender systems, decision support systems, dialogue systems with question and answering, and financial market prediction (Elnagar and Weistroffer, 2019; Liu et al. 2019).

Due to its rich semantic information and clear structure of entities and relations among entities, KGs make it easy for AI applications developers to understand the context and the meaning behind various concepts in data. KGs are also useful for data integration and discovery of facts from seemingly unrelated events or entities. For example, relationships identified by a KG from two seemingly unrelated crimes in two different places in the UK revealed that these criminals knew each other and were affiliated to the same crime group (Ehrlinger and Wöb 2016). In addition, KGs naturally become an essential part of enterprise decision support systems (DSSs) by building scalable real-time network of knowledge edges from multiple data sources (Elnagar and Weistroffer 2019).

Several studies adopted dependency graphs in the design and evaluation of curriculum. For example, de Blas, et al., (2021) focused on identifying key nodes and their relationships to reflect the temporal sequencing and dependencies among various course contents. Similarly,

Yu, et al., (2021) analyzed the descriptions of the courses and their requirements to identify the preceding and succeeding relationship among them. Few other studies (Sibarani and Scerri 2019; Dadzie, et al. 2018) tried to identify and track the evolution of the skill network from job advertisements. All together, these studies allowed researchers to visually identify the skills in most predominant demand (Liu and Schwieger 2023). Readers are strongly recommended to refer to a recent study (Tamašauskaitė & Groth 2023) and cited references within for a review on the overall process of KG development and its key constituent steps.

Maintaining and offering up to date program contents with future high-demand skills is especially challenging in fast changing technical domains such as IT. Therefore, identifying IT job skills in the most in-demand by industry employers has garnered interests of students, academic educators and advisors, and researchers (Cummings & Janicki, 2020; Koong et al., 2002; Morris et al., 2018). For example, Sibarai and Scerri (2020) proposed a Skills Cluster Observation and Discovery (SCODIS) framework to develop a forecasting model for evolving skill networks and to predict most in-demand skills associated with mobile development. In addition, several researchers claimed that IS educators must understand and incorporate emerging IT skills to narrow the skill gap between traditional skills in academic curricula and highly dynamic emerging skills from industry (Agarwal & Ahmed, 2017; Leidig 2020). Likewise, other researchers have adopted data-driven analytics approaches to assessing IS/IT courses and curriculum (de Blas et al., 2021; Yu et al., 2021). Nevertheless, a quantitative framework is lacking for effectively assessing the emerging technology-related learning content in IS/IT courses through combining up-to-date and more accessible online job advert analytics with online course content analytics.

Our study is mostly related to a case study (Liu & Schwieger, 2023) in which the proposed KG framework represented a conceptual model of university-industry knowledge interactions via online job skill network and online course content analytics in an online mobile application development course. However, this study is very different from the prior study in terms of its scope of classes considered, analysis methods at two different levels of granularity and hence resulting implications. Most of all, in this study, we significantly expand the analysis focus on from one programming course in one institution to a more comprehensive list of 50 courses in DAIS programs of multiple institutions in the Intermountain West Region of USA. Therefore, if there were any skill gaps found between academic programs and online job skills from industry, DAIS programs across multiple institutions should carefully review their various analytics

courses and redesign/update them to better prepare students for employment opportunity.

This study is also different from the prior study in a sense that we consider both hard (e.g., specific coding skills and familiarity with analysis tools) and soft analytic skills (e.g., personal traits related to collaboration, responsibility, and leadership), while the prior study only considers hard skills. In addition, we utilize the simplest KGs to present such skills at two different levels of granularity. At word or phrase level, we will create and contrast word clouds to visualize the most prominent words and phrases from a set of course and job descriptions. Then, we will extend analysis to the sentence/paragraph level at which semantic and syntactic roles of words or phrases within sentence are considered. To this end, we will identify a fact in a triple of the form from each sentence in course (or job) descriptions and present all facts in a KG.

### III. DATA SETS

Two data sets were collected, cleansed, and engineered for this study. The first data set, Course Description Data, is a small sample of undergraduate and graduate course descriptions offered by several colleges in Intermountain West Region. This data set was obtained by collecting course descriptions of 50 DAIS classes of colleges in the Intermountain West Region of USA that the authors believe provide career foundations for data analysts and data scientists through providing contents in programming, databases, data warehousing, data mining, and data analytics. To process original course descriptions for KG analysis, we used Python libraries to remove any extra spacing and lines around these headings or sections. In addition, we also removed headings and sections themselves so that they are not considered valid entities as a part of course description.

The second data set, Job Description Data, is a compiled set of 11,225 job descriptions advertised for data scientists and data analysts across the United States. This secondary data was acquired by utilizing Web scraping library in Python from Indeed's website (www.indeed.com) for data scientist or data analyst positions in US, resulting 11,225 jobs along with self-explanatory input variables such as 'JobID', 'Company', 'JobDescription', 'JobLocation', 'JobWebsite', 'Salary', and 'JobTitle.' Other additional input variables include 'JobPosted' (job posting days from web scraping date) and 'Summary' (brief summary of job description), and 'Rating' (company rating based on 1-5 scales). While it is possible to use 'Summary' variable, we used 'JobDescription' information that typically include multiple paragraphs of information on the job to build a comprehensive set of KGs.

### IV. WORD CLOUD RESULT

We first visualized the most prominent words or phrases in both course and job descriptions using word clouds, which visually present textual information weighted by their frequencies or importance using different font sizes or colors. To this end, JobDescription and CourseDescription variables from two data sets were loaded into Python using the WordCloud package.

The first word cloud in Figure 1 displays most prominent words with bigger font sizes for more prevalent words in CourseDescription variable. Note that we also had to remove several stop words (e.g., 'course', 'covers', 'teaches', and 'concepts') that appeared frequently in the course descriptions but do not deliver useful insights. According to Figure 1(a), 'Management', 'data', and 'database' were found to be the most prevalent words in course descriptions. We attribute this finding to the fact that many data analytical courses in DAIS program are still built around the usage of SQL in database management systems. Other frequent words reflect objectives of many courses toward integrating theories (e.g., 'theory' and 'principles') and applications (e.g., 'business', 'decision', 'practical' and 'application') in STEM degrees like DAIS. However, specific analytical skills that each course aim to deliver are not shown in this word cloud except general programming related words such as 'object', 'oriented', 'programming' and 'algorithm'.

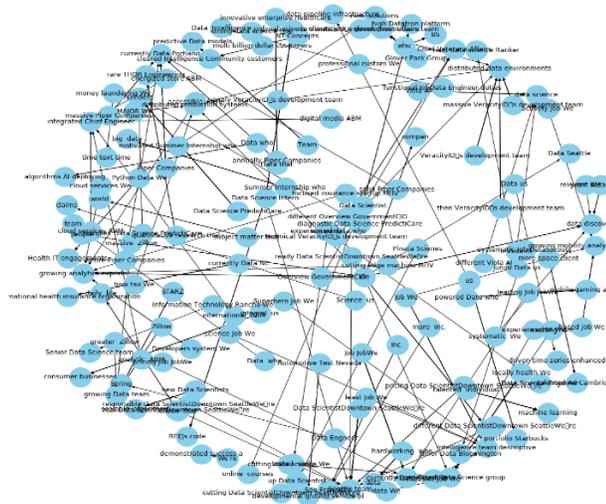


Figure 1(a): Class Description Word Cloud

Figure 1(b): Job Description Word Cloud

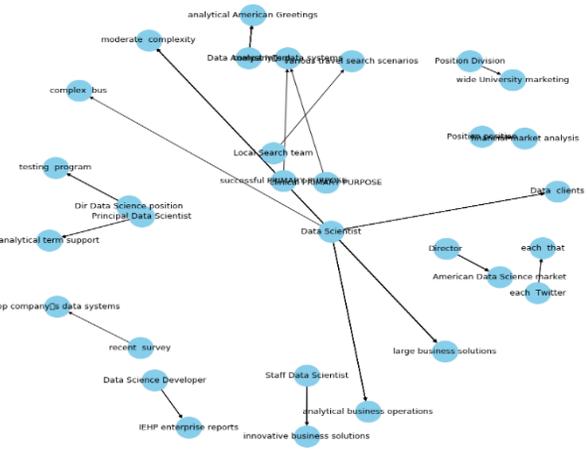
Figure 1(b) presents the most prevalent words from the 11,225 job descriptions after removing stop words such as 'together', 'brings', 'mutual', 'job', 'seen', and 'free'. Note that some stop words like 'together' and 'mutual' may reflect cultural and societal goals of job hosting companies. However, we decided to remove them from our word cloud mainly because our primary goal is to identify key analytical and technical skill mentioned in job descriptions for data scientists. As expected, the most prevalent single words are 'data scientist' with other associated words like data usage terms ('use scientifically', 'scientifically proven', 'data posting', 'looking data', 'connects Data', and 'Data Advance'). We also found





**Figure 3: 'Looking For' Relation KG**

Figure 3 presents 'Looking For' relations between source and target nodes. Visual inspections on the subgraph followed simple Python commands for network analysis revealed that target nodes such as 'Algorithms AI deploying', 'Machine Learning', 'Data Science Engineer', 'Data Discovery', and 'Growing analytics Experts' have the largest number of 'Looking For' relations from source nodes like '7 State Providence Company', 'Handworker who', 'Glover Park Group', and 'Data Engineer duties'. Thus, this knowledge subgraph confirms our initial conjecture that many companies are looking for candidates with broad ranges of data analytic skills such as 'Algorithms AI deploying', 'Machine Learning', 'Data Science Engineer', and 'Data Discovery.' However, we found that this knowledge subgraph does not identify target nodes that reflect specific analytic skills like 'Python Programming', 'R Programming', or 'Statistical Analysis' for a chosen 'Looking For' relation.



**Figure 4: 'Is Responsible' Relation KG**

Another knowledge subgraph was identified using the 'Is responsible' relation, anticipating that the knowledge subgraph would find multi-facts reading like 'the candidate is responsible for a task'. We show this knowledge subgraph with 'Is responsible' relations between source and target nodes in Figure 4. Careful investigations on this subgraph revealed that target nodes such as 'innovative business solutions', 'analytical business operations', 'complex business', and 'wide university marketing' have 'is responsible' relations from source nodes such as 'Data Scientist', 'Position Division' or 'Staff Data Scientist'. We found this observation extremely interesting considering the fact that almost all data scientist and analysts job position are supposed to be responsible for creating business [or university] solutions or operations by utilizing their skills and knowledges. This implies that many companies emphasize in their job description not only hard skills that allow analysts or developers to analyze data sets but also soft skills that utilize information and knowledge extracted from data sets to ultimately generate business values. We also found that data scientists are also responsible for specific missions such as 'companies data systems', 'analytical team support', 'enterprise reports' or 'testing program.'

We created another knowledge subgraph using the 'encompasses' relation, anticipating that the knowledge subgraph would find multi-facts reading like 'the position encompasses this task or skill'. This knowledge subgraph shown in Figure 5 is very unique in that, within this subgraph, all source nodes have 'encompasses' relation with single target node, 'enquiry business decision making' and each of the subject nodes includes 'data science' as a part of job title.

This subgraph ultimately illustrates that data science jobs must encompass business decision making.

While this finding may seem to be obvious, its implication is significant for students who desire to be a data scientist. In essence, data scientists and analysts must take their technical skills one step further to be able to utilize the information they have gained to make ‘relevant’, ‘ongoing’, ‘casual’ or ‘statistical’ business decisions.

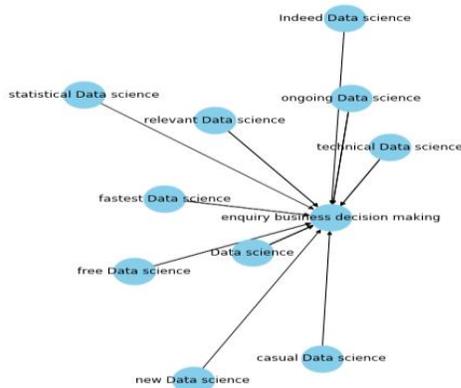


Figure 5: ‘Encompasses’ Relation KG

## VI. CONCLUSION

In this study, we compared core skills geared toward future data analysts and data scientists from two data sources: 50 DAIS course descriptions of colleges in the Intermountain West Region and 11,225 job descriptions of data scientists. According to word clouds from the two data sources, DAIS course descriptions present more general terms related to theories while job description sources illustrates more specific terms related to data usage, data types and analytical skills.

In addition, from multiple KGs based textual course and job descriptions, we found that the job descriptions source presents the importance of application of specific skills to various business settings while the curriculum source focuses on the conceptual and theoretical skills along with fundamental skills such as SQL. Therefore, we suggest that if DAIS courses truly want to prepare students for being data scientists, they should include not only technical skills, but also soft skills to make informed decisions and turn them into business decisions through real application scenarios.

We admit that our course description data does not represent national or international level, and therefore one of our immediate future research directions is to validate our findings based on more comprehensive DAIS curriculum data from multiple regions.

## REFERENCES

- [1] Agarwal, N. & Ahmed, F. (2017). Developing collective learning extension for rapidly evolving information system courses. *Education and Information Technologies*, 22(1), 7-37.
- [2] Agarwal, R. & Lucas, H. C., Jr. (2005). The information systems identity crises: Focusing on high-visibility and high-impact research. *MIS Quarterly*, 29(3), 381-398.
- [3] Benbasat, I. & Zmud, R. W. (2003). The identity crisis within the IS discipline: Defining and communicating the discipline’s core properties. *MIS Quarterly*, 27(2), 183-194.
- [4] Bordes, A. & Gabrilovich, E. (2014). Constructing and mining web-scale knowledge graphs: KDD 2014 Tutorial. In: *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1967.
- [5] Case, T., Dick, G., Granger, M. J. & Akbulut, A. Y. (2019). Teaching information systems in the age of digital disruption. *Journal of Information Systems Education*, 30(4), 287-297.
- [6] Chen, P., Lu, Y., Zheng, V. W., Chen, X. & Yang, B. (2018). KnowEdu: A system to construct knowledge graph for education. *IEEE Access*, 6, 31553-31563.
- [7] Cummings, J. & Janicki, T. N. (2020). What skills do students need? A multi-year study of IT/IS knowledge and skills in demand by employers. *Journal of Information Systems Education*, 31(3), 208-217.
- [8] Dadzie, A.S., Sibarani, E.M., Novalija, I. & Scerri, S. (2018). Structuring visual exploratory analysis of skill demand. *Journal of Web Semantics*, 49, 51-70.
- [9] de Blas, C. S., Gonzalez, D. G. & Herrero, R. C. (2021). Network analysis: An indispensable tool for curricula design. A real case-study of the degree on mathematics at the URJC in Spain. *PLoS One*, 16(3), 1-21.
- [10] Ehrlinger, L. & Wöß, W. (2016). Towards a definition of knowledge graphs. In: *Proceedings of the Posters and Demos Track of 12th International Conference on Semantic Systems – SEMANTiCS*.
- [11] Elnagar, S. & Weistroffer, H.R. (2019). *Introducing knowledge graphs to decision support systems design*. In: Wrycza S., Maślankowski J. (eds) *Information Systems: Research, Development, Applications, Education*. SIGSAND/PLAIS 2019. *Lecture Notes in*

- Business Information Processing*, vol 359. Springer, Cham.
- [12] Jafar, M. J., Babb, J. & Abdullat, A. (2017). Emergence of data analytics in the information systems curriculum. *Information Systems Education Journal*, 15(5), 22-36.
- [13] Koong, K. S., Liu, L. C. & Liu, X. (2002). A Study of the demand for information technology professionals in selected internet job portals. *Journal of Information Systems Education*, 13(1), 21-28.
- [14] Lan, M., Xu, J. & Gao, W. (2016). Ontology feature extraction via vector learning algorithm and applied to similarity measuring and ontology mapping. *IAENG International Journal of Computer Science*, 43(1), 10-19.
- [15] Leidig, P. (2020). IS 2020: Developing an ACM/AIS information systems model curriculum for undergraduate programs. In: *Proceedings of the 41st International Conference on Information Systems, ICIS 2020, Making Digital Inclusive: Blending the Local and the Global*.
- [16] Liu, J., Lu, Z. & Du, W. (2019). Combining enterprise knowledge graph and news sentiment analysis for stock price prediction. In *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 1247-1255.
- [17] Liu, X. & Schwieger, D. (2023). Developing a data-driven emerging skill network analytics framework for automated employment advert evaluation. *Information Systems Education Journal*, 21(1), 39-52.
- [18] Morris, G. J., Fustos, J. T. & Haga, W. (2018). Connecting the dots and nodes: A survey of skills requested by employers for network administrators. *Information Systems Education Journal*, 16(1), 4-12.
- [19] Müller, O., Junglas, I., Brocke, J. & Debortoli, S. (2016). Utilizing big data analytics for information systems research: Challenges, promises and guidelines. *European Journal of Information Systems*, 25(4), 289-302.
- [20] Paulheim, H. (2016). Knowledge graph refinement: A survey of approaches and evaluation methods. *Semantic Web*, 8(3), 489-508.
- [21] Pfaff, M., Neubig, S. & Krcmar, H. (2018). Ontology for semantic data integration in the domain of IT benchmarking. *Journal of Data Semantics*, 7(1), 29-49.
- [22] Sibarani, E. & Scerri, S. (2019). Generating an evolving skills network from job adverts for high-demand skillset discovery. In: *Proceedings of International Conference on Web Information Systems Engineering*, pp. 441-457.
- [23] Singh, M. P., Agarwal, P., Chaudhary, A., Shroff, G., Khurana, P., Patidar, M., Bisht, V., Bansal, R., Sachan, P. & Kuma, R. (2018). KNADIA: Enterprise Knowledge Assisted DIALOGue systems using deep learning. In: *Proceedings of 2018 IEEE 34th International Conference on Data Engineering (ICDE)*, pp. 1423-1434.
- [24] Tamašauskaitė, G. & Groth, P. (2023). Defining a knowledge graph development process through a systematic review. *ACM Transactions on Software Engineering and Methodology*, 32(1), 1-40.
- [25] Yu, X., Stahr, M., Chen, H. & Yan, R. (2021). Design and implementation of curriculum system based on knowledge graph. In: *Proceedings of IEEE International Conference on Consumer Electronics and Computer Engineering*, pp. 767-770.