

Skills Taught vs Skills Sought: Using Skills Analytics to Identify the Gaps between Curriculum and Job Markets

Alireza Ahadi
University of Technology
Sydney
PO Box 123 Broadway, Ultimo,
2007, Australia
alireza.ahadi@uts.edu.au

Marian-Andrei Rizoiu
University of Technology
Sydney
PO Box 123 Broadway, Ultimo,
2007, Australia
marian-
andrei.rizoiu@uts.edu.au

Kirsty Kitto
University of Technology
Sydney
PO Box 123 Broadway, Ultimo,
2007, Australia
kirsty.kitto@uts.edu.au

Katarzyna Musial
University of Technology
Sydney
PO Box 123 Broadway, Ultimo,
2007, Australia
katarzyna.musial-
gabrys@uts.edu.au

ABSTRACT

Higher education often aims to create job-ready graduates. Thus, the skills and knowledge taught in professional degrees are expected to align with the needs of the labor market. However, the dynamic nature of the job market makes it challenging to ensure that this alignment occurs. In this study, we show how Skills Analytics can be used to identify critical skills in the workforce, mapping these to the curriculum offerings of a university. This enables us to identify *skill gaps* between what is taught and what is needed in the job market. Methods are presented that allow universities to test the alignment of their curriculum offerings with the job market. Where gaps are identified, this would enable universities to update their curriculum more rapidly to produce graduates equipped with up-to-date skills required by the local job market. Our contributions include: a new method for ranking skills in curricula based on their relative importance in the job market; and proof of concept methods to find skills gaps between curriculum offerings and an identified job market that can lead to curriculum redesign and enhancements.

Keywords

Curriculum analytics, Skills Analytics, Skills gap

1 Introduction and Prior Research

As modern society increases in complexity, higher education is often seen as a necessity for achieving success in the workplace [7]. While this claim is sometimes challenged [5], tertiary training is still often regarded as the entry-level qual-

ification for a professional job [21]. While many claim that higher education should be aiming not just to develop work readiness but also to support people in becoming critical thinkers able to contribute to the betterment of our society [11], we increasingly see both government and employers pushing universities to prepare graduates who can make an immediate contribution in the workplace [16, 19].

How do we know that a university degree adequately prepares a student for the labor market? Over the years, a wide range of computational models have been developed to identify and/or catalog the skills and knowledge associated with a course or degree program. One popular method arises from the curation of ontologies describing a set of skills identified as critical by professional associations, employers, or some form of project. For example, the Association for Computing Machinery (ACM) has been recommending curriculum for various computer science disciplines for decades. The most recent Computing Curriculum (CC2020), includes recommendations about curriculum covering: Computer Engineering; Computer Science; Cybersecurity; Information Systems; Information Technology; Software Engineering; and an emerging Data Science category. These standards traditionally represent the curriculum for a job role as a Body of Knowledge (BoK) consisting of a set of knowledge areas (KAs), both of which contain about ten more specialized knowledge units (KUs) described by a short document. Such formal ontologies and mappings of KAs are often applied in EDM to support various Intelligent Tutoring Systems (ITS) [10].

Ontologies are often leveraged by semantic web technologies to represent curriculum [22, 12, 1], an approach that can help to ensure interoperability of curriculum data between institutions [9]. Gasmı and Bouras [13] demonstrate that it is possible to compare the competencies required by industry with those taught by education using a semantic web ontology and then propose an inference engine that can be used to match students or curriculum to jobs. However, to the best of our knowledge their proposed system was never deployed at scale.

An alternative, less manual way in which to link curriculum

A. Ahadi, K. Kitto, M.-A. Rizoiu, and K. Musial. Skills taught vs skills sought: Using skills analytics to identify the gaps between curriculum and job markets. In A. Mitrovic and N. Bosch, editors, *Proceedings of the 15th International Conference on Educational Data Mining*, pages 538–542, Durham, United Kingdom, July 2022. International Educational Data Mining Society.

© 2022 Copyright is held by the author(s). This work is distributed under the Creative Commons Attribution NonCommercial NoDerivatives 4.0 International (CC BY-NC-ND 4.0) license.
<https://doi.org/10.5281/zenodo.6853121>

to skills involves the use of Natural Language Processing (NLP). A number of different papers have extracted skills from curriculum documents using simple term frequency-inverse document frequency (TF-IDF) approaches [23, 14, 17]. More advanced methods have used simplified-supervised Latent Dirichlet Allocation (ssLDA) to compare computer science (CS) related curriculum offerings across 10 different universities in the United States [20]. This method was used to model a complete curriculum pathway by finding the center of all points corresponding to a syllabus in a degree program. However, while interesting computational approaches, these methods only work because the detailed ACM curriculum documentation has been carefully curated. This is a significant impost, and is not one that has been repeated by many other sectors. As such, it is not a method that scales to an entire university curriculum.

Various nations and even the private sector, have attempted to address the problem of field specificity by developing national standards for the skills, competencies, knowledge and abilities required for the entire workforce. Examples include: the USA based Occupational Information Network Program (O*NET); Singapore’s Skills Framework; and the classification of European Skills, Competences, Qualifications, and Occupations (ESCO). In the same vein, numerous companies are working to provide technical solutions that support people in becoming more employable, from large companies like Microsoft who are using the knowledge graph acquired from their acquisition of LinkedIn, to start ups and Small to Medium Enterprises (SMEs) such as Faethm, who have recently been acquired by Pearson. Similarly, EMSI-Burning Glass technologies (EBG) collect data by web scraping over 40,000 distinct job boards and company websites, and provides detailed daily information about labor and skill demand posted online.

Multiple studies have taken advantage of these skill taxonomies to explore labor market dynamics and the skills demanded by employers. For example, Clemens et al. [6] investigated whether minimum wage increases result in substitution from lower-skilled to slightly higher-skilled labor. Brüning and Mangeol [4] investigated the geographical variation of employer demand for graduate skills within and among occupations. Interestingly, these skills taxonomies are sometimes provided with services to parse documents and return a list of skills. This type of functionality was leveraged in a recent LAK paper by Kitto et al. [18] that defined a notion of *Skills Analytics* to support the Recognition of Prior Learning (RPL) between two institutions, but this same type of mapping is in principle possible between a student and a job, and in an EDM poster by the same group to perform some preliminary profiling of the entire curriculum offered across one university [15]. However, a problem remains: not all skills contribute equally to the employability of a graduate. Which skills are the most important?

Intriguingly, Dawson et al. [8] used Burning Glass data, coupled with a measure from labor economics, Revealed Comparative Advantage (RCA) [3, 2] to calculate the relative importance of skills that are associated with job ads of different occupations. This index is based on Ricardian trade theory and is widely used in international economics for calculating the relative advantage or disadvantage of a certain country in a certain class of commodity as evidenced by trade flows. RCA works for skills in the same way, enabling the calculation of the *advantage* conferred by a skill to a per-

son in a particular context. As such, the RCA provides us with a promising method for analyzing the relative importance of a specific skill to a job, or to a curriculum offering, or indeed, to a graduate when they attempt to find a job. However, to the best of our knowledge this metric has yet to be applied to the problem of finding gaps in an institution’s curriculum that would reduce a graduate’s potential future employability. This is the gap in the literature addressed by our paper, which proceeds by asking the following Research Question:

RQ: *How can we identify skills gaps between content taught at an institution (i.e. its curriculum offerings) and the requirements of a local job market?*

We adopt an approach based upon *Skills Based Curriculum Analytics* [18] to link university curricula with job market data. Extending the methodology proposed by Dawson et al. [8], we present an approach that enables the comparison of the skills taught within a specific degree program offered by a university with the skills sought by employers for a set of occupations targeted by that degree program. A proof of concept is presented, which contributes: (i) The use of a measure from labor market economics, Revealed Comparative Advantage (RCA), to weight skills according to the advantage they confer to a specific context; (ii) A method for evaluating the skills gaps that occur between a curriculum offering (i.e. identified degree program) and specific occupations targeted by that degree program; (iii) A method for measuring the criticality of any skills gaps, which can be used by university decision makers to prioritize curriculum redevelopment.

2 Methods

2.1 Definitions

For the sake of clarity, we will define here the precise meaning of different terms used across this paper. We will denote a *course* as a “course of study” or a degree or credential program offered by a University (e.g. a Bachelor of Science or a Master of Information Technology). A *subject* is taken to be a specific unit of study that a student undertakes during a course (e.g. Introduction to Programming, Introductory Calculus). A *job* reflects a one-on-one relationship between an employee and an employer hence a job ad represents a single vacancy. An *occupation* is a group of jobs that share very similar characteristics. For example, “teacher” is an occupation, but there are many different types of teachers, such as special education teachers and biology teachers. We use the term *skill* to denote skills, capabilities and knowledge components that are associated with a subject, course, job, or occupation interchangeably. In this study, we will represent subjects, courses, jobs and occupations as bags of skills where each bag of skills represents the skill set taught in a subject, attained by completing a course, or required for a job or an occupation, respectively.

2.2 Data Preparation

To prepare the skill-based presentation of the entire curricula of our university, we took every subject offered at the University of Technology Sydney (UTS) and extracted the following information from the Curriculum Information System: a general description of each subject, any information available about content taught, learning objectives, gradu-

ate attributes, and a high-level overview of every assessment. Next, this extracted textual data was sent to a parsing service offered by EMSI-Burning Glass Technology (EBG) (i.e. the curriculum parser tool) through API calls for skill tagging. EBG provides a taxonomy of more than 33,000 skills that are curated from job advertisement data. For each API call, the curriculum parser uses NLP to extract a list of skills which are probabilistically correlated to the subject information. This method returned skill lists for 2,747 subjects, resulting in a grand total of 138,455 skill tags. The occupational data used in this pilot study includes a relatively small set of job advertisements posted during 2020 (N = 144k) that represents a total number of 612 occupations covering 78k unique job titles. The mapping of the job ads to the occupation titles and their underlying skills (1.4M skill tags) was carried out by EBG.

2.3 Finding skill gaps between courses and the occupations they target

The procedure used to identify skills gap between curricula and job market is as follows:

- Step 1: Calculate the importance of skills for each subject
- Step 2: Calculate the importance of skills for each job ad
- Step 3: Calculate the pair-wise co-occurrence of skills for each subject
- Step 4: Calculate the pair-wise co-occurrence of skills for each job ad
- Step 5: Calculate the averaged importance of the skills associated with the job ads of each occupation
- Step 6: Calculate the averaged importance of the skills associated with the subjects of each course
- Step 7: Compute the pairwise similarity matrix of all courses and all occupations

For details about the calculations themselves, and how the RCA was implemented, the paper by Dawson et al. [8] should be consulted.

3 Results

Having used the RCA to weight the relative importance of skills, and constructed a similarity matrix, we can now work to identify skills gaps between the occupations targeted by various courses, and the skills that those courses teach. In this section we will demonstrate how these methods might be used in an institutional setting to extract actionable insights capable of informing curriculum development.

3.1 The RCA downgrades common skills

Firstly, we note that the RCA can help us to reduce the relative importance of very general skills. For example, at UTS ‘Creativity’ is returned by the EBG tagger 1,496 times across all courses that return at least one skill (see Table 1 for more examples). While possessing these general skills is *expected*, having them on a resume is unlikely to *distinguish* graduates in the job market. The RCA scoring method downgrades common skills with lower scores. Cross-checking these skills with the O*NET categorization of skills confirms that they tend to belong to the basic skill category. On the other side, skills with a high RCA score were not present in the basic skills category and so represent more *specialized* skills that distinguish individuals with respect to specific jobs. This in-

Table 1: Highly common skills in UTS curricula, with their counts across the 2,747 subjects, and the % column giving the relative frequency.

Skill	Count	%
Research	2368	86%
Teamwork / Collaboration	1788	65%
Writing	1647	59%
Creativity	1496	54%
Planning	1419	51%
Communication Skills	1401	51%
Problem Solving	1356	49%
Project Management	1337	48%
Customer Service	1240	45%
Organizational Skills	1218	44%
Building Effective Relationships	1181	42%
Budgeting	1169	42%
Presentation Skills	1160	42%
Detail-Oriented	1098	39%
Written Communication	1077	39%
Scheduling	1053	38%
Microsoft Excel	1050	38%
Biologics Development	1045	38%
Microsoft PowerPoint	1026	37%
Multi-Tasking	994	36%

icates that the RCA is largely behaving as expected in its ranking of skills. It is important to note that some analyses might not want to downgrade common skills in this manner, in which case a different weighting method could be used.

3.2 Skill gaps across a collection of courses

With the methods described in Section 2 it is now possible to explore how various courses offered by UTS align with the occupations that they are attempting to prepare their students for. We have selected 10 courses which aim to prepare their students for professional careers (and so should align well to the needs of the job market), and examined their curriculum descriptions to extract target occupations (e.g. “The course prepares students to participate in a variety of emerging careers with the growth of data science — *Data Scientist, Data Engineer...*”). We also selected a small number of occupations not being targeted by any course in our sample. We then performed the analysis represented by Steps 1 to 7 to calculate the average similarity between the bag of skills returned for each course and each of the selected occupations. The results of this analysis are presented in Figure 1. According to this heatmap, the group of Information Technology courses show a relatively small skills gap for the IT related occupations, and a correspondingly high skills gap for the occupations that they are not claiming to prepare the students for. For example, these courses perform fairly poor in preparing graduates for pharmacy related occupations. On the other side, it is evident that the pharmaceutical occupations in fact show a small skills gap with the Master of Pharmacy. One interesting finding of this presentation is the fact that some courses in fact represent a low skills gap with some occupations that they are not targeting. For instance, the Master of Forensic Sciences shows a relatively small skills gap with *Data Scientist* occupation. This is likely due to a strong emphasis upon analytical thinking and some data analysis skills which are used to provide insights into the elements of a crime. This brings another application of the proposed method; to identify potential alternative career pathways for students

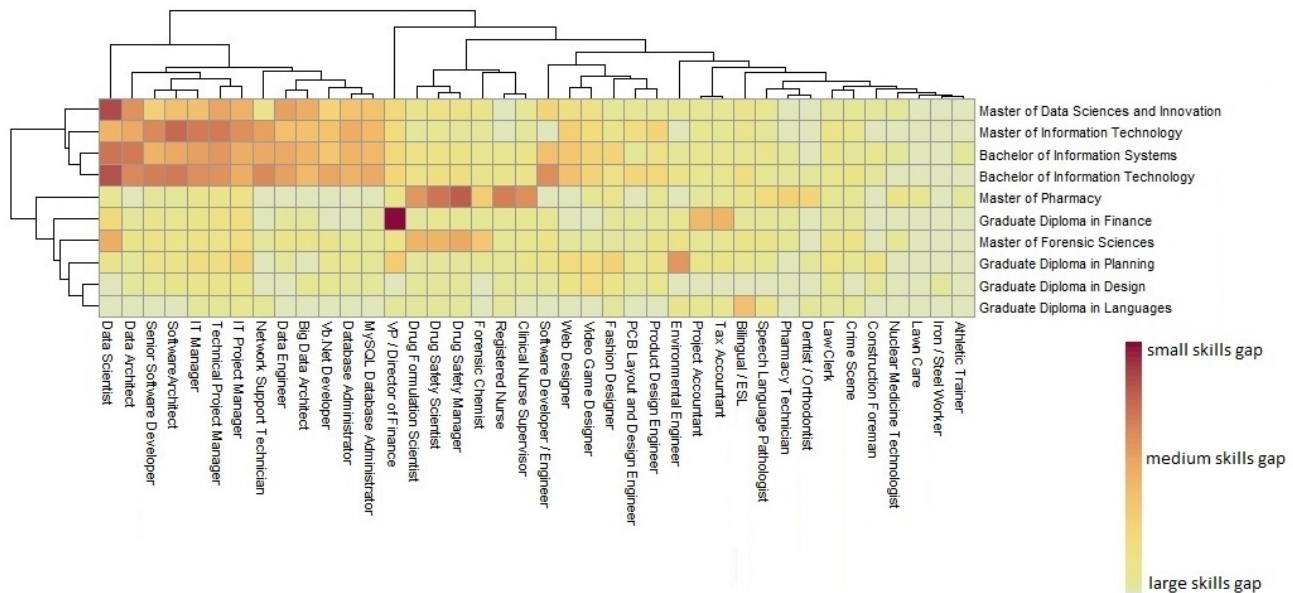


Figure 1: A heat map, representing the skills gap between 50 selected occupations and 10 courses offered at UTS.

beyond their chosen enrollment if they are dissatisfied with their existing pathway. We reserve the examination of such applications of our method for future work.

The hierarchical clustering of the occupations has successfully identified two major clusters, one including IT related occupations, and a second cluster that covers a number sub-clusters representing the other occupation categories chosen (i.e. pharmaceutical, finance, design, language, and forensic). Note in particular, that the occupations to the far right of Figure 1 show a high skill gap. This is a good sign, as they were the occupations chosen which are not targeted by any of the courses we chose to investigate. Given that the similarity calculation ranges from 0 to 1, these results are surprisingly good for two reasons. First, there are 6,684 skills in the occupation space (extremely unique skills are removed) and 3,467 skills in the curriculum space. The fact that there is such overlap between the courses and their intended occupations is a sign that the courses are indeed doing a good job in preparing their students with the skills required in the labor market. Second, as mentioned in Section 2, the occupational data used for this study includes the job advertisements of only one year (2020). This negatively impacts the results of our analysis because it reduces the accuracy of the RCA scores (as a job ad with an unusual skill profile can have a disproportionate influence over the form an occupation takes). Furthermore, the small sample means that some of the skills taught in the course that are in fact required for a given occupation are not identified in the job ads that we had access to. We expect our results to improve with access to a larger job ad dataset.

4 Discussion and Conclusions

Returning to our original research question (in Section 1), this paper has presented a new method that shows clear promise for identifying skill gaps between the content taught by an institution and the job market that it is targeting. The methods proposed in this paper are based on the assumption that relevant textual data can be represented as

a bag of skills. Keeping that notion in mind, our methods can be used in other scenarios as well. For example, this same method could potentially be used when developing a new course or subject to identify significant overlaps in the skills covered by two existing subjects, which may indicate that resources could be more effectively allocated towards the teaching of new content instead. In the same vein, this technique could be adopted for the automated identification of existing competencies, which could be used to identify students who are enrolling in a subject that is unlikely to add much value to their current skill set. While the potential applications of the method presented here are broad, it currently has some limitations. By far the most concerning centers upon the notion of *skillifying*¹ curriculum, occupations, or even a person’s resume or portfolio. We admit that this approach necessarily entails a highly simplified representation of each of these complex entities, and may be critiqued for this reason. However, we feel that the potential utility of the approach in supporting institutions to improve the employability of their graduates justifies this simplification. We believe that RCA can be used for recommending subjects/courses that minimize the skill gaps of each individual with respect to occupation goals that they specify in an interface. Such an interface could also be used to support students in identifying *alternative* occupation goals that they have not yet considered. We reserve these intriguing possibilities for future work.

In conclusion, this paper has provided a proof of concept method for mapping between the skills sought in a local job market, and the skills taught by an institution, an advance that will help to improve student employability in a rapidly changing workforce.

Acknowledgments: We acknowledge the support of Burning Glass in provisioning the API tools and data that were used in this study.

¹See <https://kb.emsidata.com/glossary/skillify/>

References

- [1] M. Al-Yahya, A. Al-Faries, and R. George. Curonto: An ontological model for curriculum representation. In *Proceedings of the 18th ACM conference on Innovation and technology in computer science education*, pages 358–358, 2013.
- [2] A. Alabdulkareem, M. R. Frank, L. Sun, B. AlShebli, C. Hidalgo, and I. Rahwan. Unpacking the polarization of workplace skills. *Science Advances*, 4(7):eaa06030, 2018. doi: 10.1126/sciadv.aao6030.
- [3] B. Balassa. Trade liberalisation and “revealed” comparative advantage 1. *The manchester school*, 33(2): 99–123, 1965.
- [4] N. Brüning and P. Mangeol. What skills do employers seek in graduates?: Using online job posting data to support policy and practice in higher education. 2020.
- [5] T. Chamorro-Premuzic and B. Frankiewicz. Does higher education still prepare people for jobs. *Harvard Business Review*, 7, 2019.
- [6] J. Clemens, L. B. Kahn, and J. Meer. Dropouts need not apply? the minimum wage and skill upgrading. *Journal of Labor Economics*, 39(S1):S107–S149, 2021.
- [7] N. R. Council et al. *Education for life and work: Developing transferable knowledge and skills in the 21st century*. National Academies Press, 2012.
- [8] N. Dawson, M.-A. Rizoïu, and M.-A. Williams. Job transitions in a time of automation and labor market crises. *arXiv e-prints*, pages arXiv–2011, 2020.
- [9] Y. Demchenko, L. Comminiello, and G. Reali. Designing customisable data science curriculum using ontology for data science competences and body of knowledge. In *Proceedings of the 2019 International Conference on Big Data and Education*, pages 124–128, 2019.
- [10] M. Desmarais and R. Baker. A review of recent advances in learner and skill modeling in intelligent learning environments. *User Modeling and User-Adapted Interaction*, 22(1-2):9–38, 2012.
- [11] O. L. U. Enciso, D. S. U. Enciso, and M. d. P. V. Daza. Critical thinking and its importance in education: Some reflections. *Rastros Rostros*, 19(34):78–88, 2017.
- [12] D. Gašević, J. Jovanović, and V. Devedžić. Ontology-based annotation of learning object content. *Interactive Learning Environments*, 15(1):1–26, 2007.
- [13] H. Gasmî and A. Bouras. Ontology-based education/industry collaboration system. *IEEE Access*, 6: 1362–1371, 2017.
- [14] A. Gibson, K. Kitto, and J. Willis. A cognitive processing framework for learning analytics. In *Proceedings of the fourth international conference on learning analytics and knowledge*, pages 212–216, 2014.
- [15] A. Gromov, A. Maslennikov, N. Dawson, K. Musial, and K. Kitto. Curriculum profile: modelling the gaps between curriculum and the job market. In *In Proceedings of The 13th International Conference on Educational Data Mining (EDM 2020)*, pages 610–614, 2020.
- [16] D. Jackson and R. Bridgstock. Evidencing student success in the contemporary world-of-work: Renewing our thinking. *Higher Education Research & Development*, 37(5):984–998, 2018.
- [17] K. Kawintiranon, P. Vateekul, A. Suchato, and P. Punyabukkana. Understanding knowledge areas in curriculum through text mining from course materials. In *2016 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*, pages 161–168. IEEE, 2016.
- [18] K. Kitto, N. Sarathy, A. Gromov, M. Liu, K. Musial, and S. B. Shum. Towards skills-based curriculum analytics: can we automate the recognition of prior learning? In *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, pages 171–180, 2020.
- [19] T. Pham and D. Jackson. The need to develop graduate employability for a globalized world. In *Developing and Utilizing Employability Capitals*, pages 21–40. Routledge, 2020.
- [20] T. Sekiya, Y. Matsuda, and K. Yamaguchi. Curriculum analysis of cs departments based on cs2013 by simplified, supervised lda. In *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge*, pages 330–339, 2015.
- [21] C. Sin, O. Tavares, and A. Amaral. Accepting employability as a purpose of higher education? academics’ perceptions and practices. *Studies in Higher Education*, 44(6):920–931, 2019.
- [22] K. Verbert, J. Klerkx, M. Meire, J. Najjar, and E. Duval. Towards a global component architecture for learning objects: An ontology based approach. In *OTM Confederated International Conferences “On the Move to Meaningful Internet Systems”*, pages 713–722. Springer, 2004.
- [23] L. S. Xun, S. Gottipati, and V. Shankaraman. Text-mining approach for verifying alignment of information systems curriculum with industry skills. In *2015 International Conference on Information Technology Based Higher Education and Training (ITHET)*, pages 1–6. IEEE, 2015.