A machine-learning based model to identify PhD-level skills in job ads

Lian Chen[1], Hanna Suominen[1-4] and **Inger Mewburn[1]**

[1]The Australian National University (ANU)/ Canberra, ACT, Australia

[2]Data61, Commonwealth Scientific and Industrial Research Organization
(CSIRO/ Canberra, ACT, Australia)

[3]University of Canberra/ Canberra, ACT, Australia

[4]University of Turku/ Turku, Finland

Abstract

Around 60% of doctoral graduates worldwide ended up working in industry rather than academia. There have been calls to more closely align the PhD curriculum with the needs of industry, but an evidence base is lacking to inform these changes. We need to find better ways to understand what industry employers really want from doctoral graduates. One good source of data is job advertisements where employers provide a 'wish list' of skills and expertise. In this paper, a machine learning-natural language processing (ML-NLP) based approach was used to explore and extract skill requirements from research intensive job advertisements, suitable for PhD graduates. The model developed for detecting skill requirements in job ads was driven by SVM. Our preliminary results showed that ML-NLP approach had the potential to replicate manual efforts in understanding job requirements of PhD graduates. Our model offers a new perspective to look at PhD-level job skill requirements.

1   Introduction

Abundant evidence shows that industry employers are often dissatisfied with key aspects of PhD graduates' workplace performance (e.g., Cumming, 2010; G08, 2013; Australian Department of Education, 2014; Hancock, 2019), particularly in relation to professional skills like communications (McCarthy & Wient, 2019). PhD graduates themselves also indicate the education they received during their candidature did not address job market needs outside of academia (Golde & Dore, 2001). Reports indicate that employers were often dissatisfied with doctoral employees' demonstration of soft skills at work (e.g., Cumming, 2010; Cyranoski et al., 2011). Such frustration from stakeholders leads us to question the fitness for purpose of doctoral degrees.

PhD was originally designed to help people into academic careers, but its fitness for purpose has been questioned for over 80 years (Dale, 1930). As the number of PhD graduates increase tremendously (Auriol et al., 2013; Gould, 2015) and the academic job market remains relatively static in scale (Larson et al., 2014), many PhD graduates will be unable to secure academic positions. Despite the fact that most PhD graduates will be working outside academia, universities are still training their candidates based on research competencies desired in academia. One such example of a popular research training framework is the Vitae Researcher Development Framework (RDF) based on research by Bray and Boon (2011). To enhance PhD employability, Mewburn et al. (2018) argue it is not enough to understand only academic workforce requirements: non-academic professions may have different needs.

Some initiatives have been undertaken to understand the so called 'transferrable skills' needed from PhD job seekers. Consequently, many add-on courses based on the long list of skill terms have been put in place at universities (Barnacle & Mewburn, 2010). However, scholars such as Neumann and Tan (2011) and Platow (2012) have expressed concerns about the generic quality of these initiatives. Take professional skills such as teamwork and empathy: these may mean very different things in different workplaces and industry domains. Concerns are reasonable about the ambiguity of previous initiatives suspicious of over-generalisation of skills and neglect of the context in which skills are deployed. However, little has been done so far to empirically test the difference that context makes. We can ask employers to tell us, but relying on retrospective self-reports has an inherent problem of informant inaccuracy (Bernard et al., 1984; Ellison et al., 2020). Other methods such as ethnography are too difficult to scale up considering the large number of industry fields. Therefore, a Machine Learning / Natural Language (ML/NLP) approach is worthy of exploration.

To address the gap described above, we developed a machine-learning-based model to identify employers' expectations of qualified job seekers in job adverts. We first manually labelled 400 job ads across two industries based on Move-Step analysis, an analytic approach widely adopted by applied linguistics researchers to detect contextual difference in written language discourse (Bhatia, 2014). Moves

are coarse-grained categories, and steps are fine-grained categories associated with a particular move. Finer granularity is essential to avoiding abstractness and ambiguity in language use. Hence, a data retrieval interface with fine- and coarse-grained information can provide more accurate results than an interface without differentiation in information granularity (Zhang et al., 2020). After the manual annotation was done, the labelled data were fed into the machine using the SVM algorithm. After the parameters were tuned, the majority of the identified skill categories in the model reached good Area Under the Curve (AUC) performance. Although the model remains to be optimised for certain skill categories, the results showed that natural language processing of job advertisements has the potential to replicate human efforts to provide rich insights into how PhD education can be improved.

This paper seeks to contribute to the literature in three key ways:

1. Workshop design: We outline how we developed a coders' workshop for this project. Our workshop experience can be valuable for future attempts to use natural language processing to inform higher education policy making.

2. Curriculum design tools: The coarse and fine granularity in our annotation generates more accurate definitions of skill items, which in turn enables better curriculum design in PhD education.

3. Evidence: Our model can serve as a useful approach to testify the hypothesis that contextual difference exists across industry domains. In addition, it gives evidence of automation being a feasible approach to contribute to human efforts in understanding PhD-level skill requirements in job advertisements by accelerating abilities to analyse text systematically at scale.

2   Background and related work

Job ads contain rich information regarding employers' expectation of qualified job seekers (Walker & Hinojosa, 2014), yet research on the automated identification of skills from job ads is still in its infancy. Most previous studies on skills in job ads were manual content analysis (see for example, Pitt & Mewburn, 2016). The majority of studies on job ads only examined a single industry domain. Studies with manual efforts examined professions such as librarian positions (Clyde. 2002) and accounting (Tan & Laswad, 2018). Nevertheless, studies relying on machine learning techniques seem to exclusively examine computing related job positions (e.g., Aken et al., 2009; Ericsson &

Wingkvist, 2014; Khaouja et al., 2018; Rahhal et al., 2019). In recent years, scholars have started analysing and extracting skills in job ads with machine learning techniques. The effort to automatically retrieve skill requirements from job ads should enable a content analysis approach to job ads to be expanded to other industry domains.

Although there are several ML-based studies on job ads, most authors only took technical skills into consideration. For example, Ericsson and Wingkvist (2014), Khaouja et al. (2018), Rahhal et al. (2019) identified technical skills such as programming languages but ignored non-technical skills such as teamwork, communication skills, user engagement, workplace aesthetics, ethics, etc. Obviously, non-technical skills in job ads remain underexplored by data mining researchers. In order to have a holistic understanding of employers' expectations, we need to examine both technical and non-technical skills listed in job ads.

3   Methods

3.1 Analytic framework

At the manual annotation stage, Move-step analysis was adopted. This analytic framework, first proposed by Swales (1990), has been widely adopted by applied linguistics researchers in exploring conventions in written discourse of communities of practice[1] (Bhatia, 2014; Moreno & Swales, 2018). In Move-step analysis, researchers take the stance that a particulate genre reflects social habitus [2] of a community of practice through events and goals they record in their texts (Bhatia, 2014). When the context of a text genre changes, these components vary in quantity or quality (Connor, 2000; Maswana et al., 2015). A most straightforward example is research articles. The structural components of research articles in different scientific disciplines are not entirely consistent. In a similar vein, job ads across industry domains could differ in textual conventions.

Moves in a genre are the overarching communicative purposes which can be achieved through alternative steps (Swales, 1990). This hierarchical differentiation of information granularity is very useful for users of results of information retrieval tasks (Tange et al., 1998; Fonseca et al., 2002; Zhang et al., 2020). As move step analytic framework is an inductive approach, it allows enough autonomy for annotators to minimise influence from predefined categories which are likely to be vaguely defined (Neumann & Tan, 2011). Such autonomy is particularly important when most categories of skills in existent reports, according to Platow (2012), do not derive from

---

[1] The term 'communities of practice', coined by Wenger (1999), denotes groups of people who share the same goals, interests, and knowledge.

[2] 'Social habitus' refers to habits, dispositions and skills someone has as a result of immersing in a social environment (Bourdieu, 1986).

solid theoretical justification[3].

Below is an example of the move 'Continuous education' and its associated steps identified from the job ad data in this study:

Move: Continuous education

- **Step 1**: Passion & Self-motivation
  Example: *A passion for developing web-based applications.*
- **Step 2**: Participation in training
  Example: *In return we'll offer training, scope for progression and a support buddy to keep you company while you get your feet under the desk.*
- **Step 3**: Sharing of knowledge
  Example: *...recommend corrective actions advice.*
- **Step 4**: Seeking advice
  Example: *Continuously seek(s) feedback and responds proactively.*
- **Step 5**: Self-reflection
  Example: *Demonstrate the ability to identify and support practice improvements and support the implementation of best practice.*

3.2 Data

Research intensive job ads were chosen to explore employability skills required of doctoral graduates. The raw data were purchased from the job market solution company Burning Glass Technologies Inc. Ethics approval was obtained for using the data for our research purpose. Although the purchased data cannot be shared due legal and ethical concerns, future scholarly attempts to verify the results from this study can rely on job ad data from the public domain. One example of using publicly available job ads for analysis is Pitt and Mewburn (2016). Alternatively, scholars interested in accessing the same data could contact Burning Glass Technologies Inc directly. Our intention of purchasing data was for scaling up the analysis across industries in the future.

The raw data were further filtered into research skill intensive job ads via the algorithm developed by the PostAc® team (Mewburn et al., 2018; Xu et al., 2019) at the Australian National University. The PostAc® filter reached good performance of above 80% accuracy.

Overall, 400 high research skill intensive job ads posted from 2015 to 2016 were randomly chosen for manual annotation. The total word counts of the dataset were 147,089. Of the 400 job ads, 200 were targeting healthcare job seekers, and another 200 were targeting computing job seekers. The reason we chose healthcare and computing professions is because these two domains have the greatest potential to grow in the next five years (Australian Government, 2019). The healthcare dataset consisted of 74,179 of word frequencies. The computing dataset comprised 72,910 of word frequencies.

Several steps were taken to pre-process the data in preparation for the machine learning experiments. We first segmented the data into sentences as units for annotation. This process helped us see if a skill was mentioned several times in one job ad, which possibly indicates the importance of the skill for the employer. If we had treated a whole job ad as an analytic unit, a skill requirement could have been only counted once even if it is mentioned several times. We also removed stop words such as articles and conjunctions from the machine-readable dataset via the stop-word list in the NLTK v3.5 corpus. For the training, testing and validation purposes, the labelled dataset was separated into 70%, 15% and 15% of the overall dataset accordingly.

3.3 The chosen algorithm

The chosen algorithm for running the experiment in the study is Support Vector Machine (SVM) (Cortes & Vapnik, 1995) with the linear kernel. An analytic unit could contain multiple skill requirements. Hence our task is a multilabel text classification task. SVM is an algorithm widely adopted to deal with multilabel text classification tasks (Qin & Wang, 2009; Yang et al., 2009; Wang & Chiang, 2011). Another reason for choosing linear SVM is because it is a computationally cost-effective algorithm which at the same time guarantees good prediction outcome for text classification tasks (Vijayan et al., 2017). It is worth mentioning that we also piloted using Naive Bayes and Logistic Regression classifiers potentially suitable for multi-label text classification tasks with the default parameters. SVM is the one that obviously outperformed these piloted baselines on our dataset. In the experiment, SVM parameters were tuned for optimization using the GridSearchCV [4] tool. The parameters tuned are listed as follows:

- Max-iteration,
- Loss,
- Tolerance,
- Fit intercept, and
- Intercept scaling.

The classifiers' performance was evaluated using the Area Under the Curve (AUC). The AUC measure can avoid the problematically 'too good' results derived from the situation where the accuracy score is high, but the evaluation is biased by class imbalance (Suominen et a., 2009; Narkhede, 2018).

3.5 Coders' workshop

One of the authors and a research assistant were involved in our Coders' workshop. Both annotators hold masters' degrees and were experienced in annotation tasks.

In the coders' workshop, we agreed that several aspects listed as follows are essential to ensuring the quality of the annotation.

1. It is important to have more than one annotator to do the annotation independently. This is to ensure the inter-coder reliability. Our intercoder reliability at step level reached 0.76 measured by Cohen's Kappa[5]. As achieving a good intercoder reliability at the step level is a direct indication that move-level units were labelled reliably, we only calculated Kappa at the step level.

2. Compared to one-off pilot annotation, it is more reasonable to hold several rounds of discussion on improvements among annotators in between annotation efforts. In other words, an iterative process of continuous improvement would be better.

3. There should be a mechanism to resolve interpersonal conflict when disagreement occurs between annotators. In our case, we marked down dubious items in our notes during the discussion before we continue in the next round of annotation with an eye for evidence and justification for our opinions.

4 Results and discussion

Overall, 12 skills at move level and their associated steps were identified from our manual analysis (see table 1). The fine-grained steps in our model serve as a tool to unpack the meaning of coarse-grained 'umbrella terms', which in the past were considered by scholars as having little information regarding their contextual interpretation (Barnacle & Mewburn, 2010). The results and discussion comprise two parts. The first part is the report on machine learning performance. The second part is the report on Chi-square test for skill categories with good ML performance.

4.1 Machine learning performance

The machine learning experiment results in table 1 showed that the AUC scores of many step categories on training, test and validation sets are close, which indicates the model is very likely to produce similar results on unseen data.

It can also be seen from table 1 that 28 out of 61 step-level categories reached AUC scores above 0.8 on all training, validation, and test sets. These categories account for 46% of the overall step-level categories.

Table 1. AUC results of the experiment

Moves	Steps	Train	Val	Test
Empathy with	Clients	0.98	0.89	0.90
	Children	0.98	0.89	0.92
	Clients' family members	0.97	0.87	0.83
	Aged group	0.97	0.89	0.83
	Ethnic minorities	0.98	0.87	0.82
	General public	0.88	0.61	0.66
	Novices	0.91	0.79	0.74
	Disabled group	0.98	0.97	0.88
	LGBTIQ+ community[6]	0.98	1.00	1.00
	Women	0.97	0.85	0.50
People skills	Network with peers	0.93	0.77	0.80
	Interpersonal skills	0.96	0.84	0.82
	Multidisciplinary collaboration	0.96	0.87	0.83
	Network with decision makers	0.89	0.63	0.68
	Network with partners	0.93	0.78	0.74
	Network with the public	0.85	0.50	0.50
	Network with public sectors	0.95	0.67	0.65
	Network with private sectors	0.92	0.63	0.50
	Network with research institutions	0.80	0.80	0.79
	Network with project sponsors	0.75	0.75	0.73
	Network with research participants	0.95	0.59	0.69
Continuous education	Participate in training	0.94	0.86	0.84
	Passion & Self-motivation	0.93	0.83	0.80
	Share knowledge	0.92	0.84	0.80
	Seek advice	0.85	0.71	0.68
	Self-reflection	0.98	0.69	0.70
Cognitive abilities	Analytic skills	0.92	0.77	0.76
	Needs extraction	0.95	0.78	0.78
	Understand problems	0.98	0.94	0.88
	Innovation	0.98	0.88	0.90
Professional standards	Ethical conduct	0.67	0.60	0.64
	Policy & Regulation	0.96	0.88	0.82
	Background-check	0.98	0.93	0.96
	Confidentiality	0.95	0.79	0.85
Personal attributes	Personal impact	0.85	0.80	0.79
	Leadership skills	0.95	0.84	0.83
	Result orientation	0.94	0.72	0.75
	High-pressure management	0.91	0.73	0.67

---

[5] Cohen's Kappa equation: $K = \frac{P(a) - P(e)}{1 - P(e)}$ Where P(a) denotes observed percentage of agreement, and P(e) denotes the probability that agreement is met by chance. Cohen's Kappa works for assessing categorical variables (Hallgren, 2013), and hence is suitable for this study.

[6] The step 'Empathy with LGBTIQ+ community' has relatively simpler features. When an analytic unit contains the key words of 'LGBT', 'sexuality' or 'gay', the model would very possibly predict it as this 'Empathy with LGBTIQ+ community'. We therefore agreed that the this is the reason why the AUC results for both Val and Test sets are 1.

	Attention to safety	0.98	0.92	0.88
	Commercial orientation	0.88	0.65	0.73
	Time efficiency	0.94	0.76	0.84
	Independence	0.94	0.77	0.82
	Attention to details	0.93	0.73	0.75
Aesthetics	Maintain workplace	0.96	0.84	0.82
	Manage configuration	0.98	0.83	0.83
	Manage resource	0.95	0.72	0.75
Courage	Work in harsh environment	0.91	0.60	0.65
	Manage conflicts	0.98	0.66	0.77
	Manage risks	0.95	0.83	0.86
	Manage change	0.97	0.80	0.89
	On-call availability	0.92	0.50	0.67
	Driving & Travelling	0.97	0.85	0.80
Proof of qualification	Register in institutions	0.93	0.83	0.81
	Writing skills	0.94	0.79	0.69
	Attain tertiary degree	0.93	0.83	0.88
	General IT skills	0.95	0.70	0.69
	Industry experience	0.96	0.84	0.84
	Oral presentation	0.90	0.80	0.75
	Residency	0.98	0.83	0.96
None category	*The category not containing any skill requirement	0.96	0.89	0.90
Healthcare technical skills		0.88	0.68	0.70
Computing technical skills		0.92	0.82	0.78

Although not all the skill categories reached the rule of thumb gold standard of 0.8 in AUC performance, the results from the experiment so far indicate a likelihood that the manual efforts can be replicated by the machine when further optimization is conducted.

Specifically, step categories such as 'Empathy with the general public' which reached an AUC score above 0.9 on training set but below 0.8 on validation and test sets are likely to have the overfitting problem. The linear SVM is an algorithm that already is less prone to overfitting (Baumes et al., 2006). Hence, future attempts to avoid overfitting for optimization include increasing the number of training samples in order for the machine to capture the key features of the category.

For categories (e.g., the step 'Network with project sponsors') which reached an AUC score below 0.8 on training, validation and test sets, it is very likely that the quality of the manual annotation is not good enough for the machine to learn well. Such underfitting problem can be avoided when future attempt to improve the manual coding procedure is conducted.

Due to the necessity of further optimizing the model, it is still not feasible for our model to be used immediately in another study. Nevertheless, our model can be used to identify those 28 step categories which reached AUC score above 0.8 on all training, test, and validation sets in PhD-level healthcare and computing job ads.

4.2 Chi-square test for step categories with good AUC scores

Previously, there was little empirical evidence to testify the assumption about contextual difference in skill requirements. We therefore did chi-square test for the 27 step-level skill categories (excluding 'None category') with good AUC performance based on our manually labelled data set. The results suggest that there is a significant difference between the two industry domains in 22 of these 27 step-level categories. The Chi-square test results for these categories are listed in table 2 below.

Table 2. Chi-square test results for the 27 steps with good ML performance

Steps	*Comp*	*Health*	*P value*	X^2
Empathy with clients	625	602	> .5	0.4
Empathy with children	8	226	< .0001	203.1
Empathy with clients' family members	5	175	< .0001	160.6
Empathy with aged group	5	143	< .0001	128.7
Empathy with ethnic minorities	101	159	< .001	12.9
Empathy with disabled group	2	124	< .0001	118.1
Empathy with LGBTIQ+ community	7	30	< .001	14.3
Interpersonal skills	367	383	> .5	0.3
Multidisciplinar-y collaboration	106	179	< .0001	18.7
Participate in training	159	221	< .01	10.1
Passion & Self-motivation	474	431	< .5	2
Share knowledge	365	391	< .5	0.9
Understand problems	311	292	< .5	0.6
Innovation	208	149	< .01	9.8
Policy & Regulation	163	452	< .0001	135.8
Background-check	115	213	< .0001	29.3
Leadership skills	199	143	< .01	9.2
Attention to safety	67	170	< .0001	44.8
Maintain workplace	16	145	< .0001	103.4
Manage configuration	82	2	< .0001	76.2
Manage risks	164	99	< .0001	16.1
Manage change	188	123	< .001	13.6
Driving & Travelling	61	70	> .5	0.6

Register in institutions	4	279	$< .0001$	267.2
Attain tertiary degree	112	181	$< .0001$	16.3
Industry experience	157	102	$< .001$	11.7
Residency	70	80	$< .5$	0.7

There is no significant difference in the occurrences of 'Empathy with clients' and 'Interpersonal skills' between computing and healthcare industries. Both industries required qualified job seekers to maintain positive relationships with people at work. Nevertheless, job ads in healthcare industry required job seekers to empathize with a wider range of communities than in computing industry. Healthcare job seekers need to have stronger capacity of emphasizing with children, clients' family members, the aged group, ethnic minorities, people with disability, and the LGBTIQ+ community. Such difference in 'Empathy' might indicate greater subtlety and complexity of healthcare professionals' workplace interpersonal relationships.

'Multidisciplinary collaboration' in healthcare job ads was mentioned more frequently in computing job ads. This difference might be because the healthcare professionals often need to deal with complex health problems beyond one's specialization (Vissers et al., 2013). In comparison, computing professionals receive everyday tasks whose technical scope are already pinpointed. Possibly it is because of a weaker multidisciplinary orientation that computing professionals were required more often to understand problems by one's own (Gardner, 2010), as indicated by more frequent mentioning of the step 'Understand problems' in computing (shown in table 2).

There was only slight difference between the two industries in the steps of 'Participate in training', 'Passion & Self-motivation', 'Share knowledge', 'Innovation', and 'Leadership skills'. Whereas computing job seekers were required more often to have 'Passion & Self-motivation', 'Leadership skills' and 'Innovation', healthcare job seekers were asked more often to 'Share knowledge' and 'Participate in training'. Since computing industry frequently experiences innovation and drives change in technology (as indicated by the step 'Manage change' and 'Innovation' in table 2), professionals in this field naturally need to have stronger passion and self-motivation to keep pace. Also, problems in the field are sometimes ill-structured (Brown, 2008), and hence the ability to lead and manage change in clients' needs and technology plays a significant role in computing industry.

The more frequent requirement in healthcare industry of 'Policy & Regulation', 'Background-check', 'Attention to safety', 'Maintaining workplace' and 'Managing risks' (table 2) could be associated with the patient-related uncertainties in healthcare context. As

healthcare professionals' practices are directly relevant to patients' physical and mental wellbeing, familiarization with policies, regulations, safety guidelines, risk prevention is crucial in order to cope with potential hazards and disputes.

The step 'Manage configuration' seemed to be a requirement specific to the computing industry, as shown in table 2. Configuration is domain-specific resource that computing practitioners need to manage in their daily work. According to Stevenson (2010), configuration in computing refers to the set-up of software and hardware components of a product. Hence, configuration is a special resource in computing that is different from other resource such as those in the healthcare industry.

Through Chi-square test, we pinpointed difference in the number of skills required in computing and healthcare industries. These difference challenges the view that doctoral graduates' identity is monolithic, especially under the current circumstance where more than 60% PhDs ended up working outside of academia (Larson et al., 2014). As suggested by Gardner, (2010), doctoral students' identity formation, skill development and socialization should be linked with specific industrial and professional contexts when devising support in their programs. In this study, we illustrate how it would be problematic to offer add-on courses and de-contextualized interpretation of skills to PhD students. Analysis of job ads in different industries, and machine-learning-enabled automation of the analysis are potential methods to enable better decision making in PhD education.

5   Conclusion

In this study, we developed a model to automatically identify some PhD-level skill requirements in job ads. The move-step analysis we adopted in our manual annotation procedure allows for the model to pinpoint both coarse-grained skill items and their associated contextual interpretation. The moves and steps we identified can be used in curriculum design to enhance PhD candidates' awareness of skill requirements in different industry domains. Our human coder's workshop experience can be useful for scholars who also intend to conduct ML driven analysis of textual data for enabling better decision making in higher education. In addition, the ability of our model to quantify skills provides evidence that contextual difference exists in the number of skills required of qualified PhD job seekers. Our finding challenges the problematic view that we can set aside contextual factors in PhD training.

Our model has limitations. In this study, we pinpointed areas for further optimizing the model. Before the model can be used to automatically identify all the skill categories, the overfitting and underfitting problems need to be solved. Additionally, we only experimented with limited number of data in healthcare

and computing domains. Therefore, extra effort to manually label more data in these two and other industries is necessary before the model can be used to identify skills across different industry contexts.

Acknowledgement

References

Aken, Andrew., Litecky, Chuck., Ahmad, Altaf., & Nelson, Jim. (2009). Mining for computing jobs. *IEEE software*, *27*(1), 78-85.

Auriol, Laudeline., Misu, Max., & Freeman, R. Ann. (2013, March). Careers of doctorate holders: Analysis of labour market and mobility indicators. Retrieved from https://www.oecd-ilibrary.org/docserver/5k43nxgs289w-en.pdf

Australian Department of Education (2014, June). *Initiatives to enhance the professional development of research students.* Retrieved from https://docs.education.gov.au/system/files/doc/other/initiatives_to_enhance_the_prof_development_of_research_students_0.pdf

The Australian Government. (2019). *Future Outlook*.

Barnacle, Robyn., & Mewburn, Inger. (2010). Learning networks and the journey of 'becoming doctor'. *Studies in Higher education*, *35*(4), 433-444.

Baumes, L. A., Serra, J. M., Serna, P., & Corma, A. (2006). Support vector machines for predictive modeling in heterogeneous catalysis: a comprehensive introduction and overfitting investigation based on two real applications. *Journal of combinatorial chemistry, 8*(4), 583-596.

Bernard, H. Russell., Killworth, Peter., Kronenfeld, David., & Sailer, Lee. (1984). The problem of informant accuracy: The validity of retrospective data. *Annual review of anthropology*, *13*(1), 495-517.

Bhatia, Vijay. (2014). *Analysing genre: Language use in professional settings*. London: Routledge.

Bourdieu, Pierre. (1986) 'The Forms of Capital', in J.G. Richardson (ed.) *Handbook of Theory and Research for the Sociology of Education* (pp. 241-59). New York: Greenwordpress.

Bray, Robert., & Boon, Stuart. (2011). Towards a framework for research career development. *International Journal for Researcher Development.*

Brown, T. (2008). Design thinking. *Harvard business review*, *86*(6), 84.

Clyde, Laurel. A. (2002). An instructional role for librarians: an overview and content analysis of job advertisements. *Australian academic & research libraries*, *33*(3), 150-167.

Connor, Ulla. (2000). Variation in rhetorical moves in grant proposals of US humanists and scientists. Retrieved from https://scholarworks.iupui.edu/handle/1805/2663

Cortes, Corinna., & Vapnik, Vladimir. (1995). Support-vector networks. *Machine learning, 20*(3), 273-297.

Cumming, Jim. (2010). Contextualised performance: Reframing the skills debate in research education. *Studies in Higher Education*, *35*(4), 405-419.

Cyranoski, David., Gilbert, Natasha., Ledford, Heidi., Nayar, Anjali., & Yahia, Mohammed. (2011). The PhD factory: The world is producing more PhDs than ever before. Is it time to stop?. *Nature*, *472*(7343), 276-279.

Dale, Edgar. (1930). The Training of Ph. D.'s. *The Journal of Higher Education*, *1*(4), 198-202.

Ellison, William. D., Trahan, Alec. C., Pinzon, Joseph. C., Gillespie, Megan. E., Simmons, Logan. M., & King, Kendel. Y. (2020). For whom, and for what, is experience sampling more accurate than retrospective report? *Personality and Individual Differences*, *163*, 10071.

Ericsson, Morgan., & Wingkvist, Anna. (2014, June). *Mining job ads to find what skills are sought after from an employers' perspective on IT graduates*. [Paper presentation]. 2014 conference on Innovation & technology in computer science education, New York, US. https://dl.acm.org/doi/10.1145/2591708.2602670

Fonseca, Frederico., Egenhofer, Max., Davis, Clodoveu., & Câmara, Gilberto. (2002). Semantic granularity in ontology-driven geographic information systems. *Annals of mathematics and artificial intelligence*, *36*(1-2), 121-151.

G08 (2013, March). *The Changing PhD: Discussion paper*. Retrieved from https://go8.edu.au/files/docs/the-changing-phd_final.pdf

Gardner, Susan. K. (2010). Contrasting the socialization experiences of doctoral students in high-and low-completing departments: A qualitative analysis of disciplinary contexts at one institution. *The Journal of Higher Education*, *81*(1), 61-81.

Golde, Chris. M., & Dore, Timothy. M. (2001). At cross purposes: What the experiences of today's doctoral

students reveal about doctoral education. Retrieved from https://files.eric.ed.gov/fulltext/ED450628.pdf

Gould, Julie. (2015). How to build a better PhD. *Nature News*, *528*(7580), 22.

Hallgren, Kevin. A. (2012). Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in quantitative methods for psychology*, *8*(1), 23.

Hancock, Sally. (2019). A future in the knowledge economy? Analysing the career strategies of doctoral scientists through the principles of game theory. *Higher Education*, *78*(1), 33-49.

Khaouja, Imane., Rahhal, Ibrahim., Elouali, Mehdi., Mezzour, Ghita., Kassou, Ismail., & Carley, Kathleen. M. (2018, April). Analyzing the needs of the offshore sector in Morocco by mining job ads. [Paper presentation]. *2018 IEEE Global Engineering Education Conference (EDUCON)*, Tenerife,Spain.
https://ieeexplore.ieee.org/document/8363390

Larson, Richard. C., Ghaffarzadegan, Navid., & Xue, Yi. (2014). Too many PhD graduates or too few academic job openings: the basic reproductive number R0 in academia. *Systems research and behavioral science*, *31*(6), 745-750.

Maswana, Sayako., Kanamaru, Toshiyuki., & Tajino, Akira. (2015). Move analysis of research articles across five engineering fields: What they share and what they do not. *Ampersand*, *2*, 1-11.

McCarthy, Paul., & Wient, Maaike. (2019). *Who are the top PhD employers*: *Advancing Australia's know ledge economy*. Retrieved from https://blogs.deak in.edu.au/bl-research-student-news/wp-content/up loads/sites/227/2019/09/Advancing_Australias_K nowledge_Economy-0021.pdf

Mewburn, Inger., Grant, Will. J., Suominen, Hanna., & Kizimchuk, Stephanie. (2018). A machine learning analysis of the non-academic employment opportunities for Ph. D. graduates in Australia. *Higher Education Policy*, 1-15.

Moreno, Ana. I., & Swales, John. M. (2018). Strengthening move analysis methodology towards bridging the function-form gap. *English for Specific Purposes*, *50*, 40-63.

Narkhede, Sarang. (2018). Understanding AUC-ROC Curve. *Towards Data Science*, *26*.

Neumann, Ruth., & Tan, Kim. K. (2011). From PhD to initial employment: the doctorate in a knowledge economy. *Studies in Higher Education*, *36*(5), 601-614.

Pitt, R., & Mewburn, I. (2016). Academic superheroes? A critical analysis of academic job descriptions. *Journal of Higher Education Policy and Management*, *38*(1), 88-101.

Pitt, Rachael., & Mewburn, Inger. (2016). Academic superheroes? A critical analysis of academic job descriptions. *Journal of Higher Education Policy and Management*, *38*(1), 88-101.

Platow, Michael. J. (2012). PhD experience and subsequent outcomes: A look at self-perceptions of acquired graduate attributes and supervisor support. *Studies in Higher Education*, *37*(1), 103-118.

Qin, Yu. P., & Wang, Xiu. K. (2009, August). *Study on multi-label text classification based on SVM*. [Paper presentation]. 2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery, Tianjin, China. https://ieeexplore.ieee.org/abstract/document/5358597

Rahhal, Ibrahim., Makdoun, Ibtissam., Mezzour, Ghita., Khaouja, Imane., Carley, Kathleen., & Kassou, Ismail. (2019, April). *Analyzing Cybersecurity Job Market Needs in Morocco by Mining Job Ads*. [Paper presentation]. 2019 IEEE Global Engineering Education Conference *(EDUCON)*, Dubai, United ArabEmirates. https://ieeexplore.ieee.org/document/8725033

Stevenson, Angus. (Ed.) (2010). Oxford dictionary of English (3rd ed.). Oxford: Oxford University Press.

Suominen, H., Pyysalo, S., Hiissa, M., Ginter, F., Liu, S., Marghescu, D., Pahikkala, T., Back, B., Karsten, H., Salakoski, T. (2008). Performance evaluation measures for text mining. In M. Song & Y-FB. Wu (Eds.), *Handbook of Research on Text and Web Mining Technologies* (pp. 724–747). Pennsylvania, USA: IGI Global.

Swales, John. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.

Tan, Lin. M., & Laswad, Fawzi. (2018). Professional skills r-equired of accountants: what do job advertisements tell us?. *Accounting Education*, *27*(4), 403-432.

Tange, Huibert. J., Schouten, Harry. C., Kester, Arnold. D., & Hasman, Arie. (1998). The granularity of medical narratives and its effect on the speed and completeness of information retrieval. *Journal of the American Medical Informatics Association*, *5*(6), 571-582.

Vijayan, Vikas. K., Bindu, K. R., & Parameswaran, Lat-ha. (2017, September). *A comprehensive study of text clas-sification algorithms.* [Paper presentation]. 2017 International Conference on Advances in Computing, Communications and Informatics (IC ACCI), Udupi, India. https://ieeexplore.ieee.org/document/8125990

Vissers, Kris. C., van den Brand, Maria. W., Jacobs, Jose., Groot, Marieke., Veldhoven, Carel., Verhagen, Constans., Hasselaar, Jeroen., & Engels, Yvonne. (2013). Palliative medicine update: a multidisciplinary approach. *Pain Practice*, *13*(7), 576-588.

Walker, H. Jack., & Hinojosa, Amanda. S. (2014). Recruitment: The Role of Job Advertisements. In K. Yu & D. M. Cable (Eds.), *The Oxford*

handbook of recruitment (pp. 269-283). New York: Oxford University Press.

Wang, Tai. Y., & Chiang, Huei. M. (2011). Solving multi-label text categorization problem using support vector machine approach with membership function. *Neurocomputing*, *74*(17), 3682-3689.

Wenger, Etienne. (1999). *Communities of practice: Learning, meaning, and identity*. Cambridge: Cambridge university press.

Xu, Chenchen., Mewburn, Inger., Grant, Will. J., & S uominen, Hanna. (2019, July). *PostAc: A Visual Interactive Search, Exploration, and Analysis Pl atform for PhD Intensive Job Postings*. [Paper pr esentation]. The 57th Annual Meeting of the Ass ociation for Computational Linguistics: System Demonstrations, Florence, Italy. https://www.acl web.org/anthology/P19-3008.pdf

Yang, Bishan., Sun, Jian. T., Wang, Tengjiao., & Chen, Zheng. (2009, June). *Effective multi-label active l earning for text classification*. [Paper presentation]. The 15th ACM SIGKDD international conferen ce on Knowledge discovery and data mining, New York, US. https://dl.acm.org/doi/10.1145/15570 19.1557119

Zhang, Min., Li, Feng., Wang, Yang., Zhang, Zequan., Zhou, Yanhai., & Li, Xiaoyu. (2020). Coarse and Fine Granularity Graph Reasoning for Interpretab le Multi-Hop Question Answering. *IEEE Access* , *8*, 56755-56765.