**Meeting with Prof. Diana Inkpen and Pavan Balaji Kumar**

**October 19, 2021**

**Background**

Curricula vitae (CV) provide comprehensive summaries of researchers’ scientific careers, including publications, presentations, grants, committees, *etc*.(Cañibano & Bozeman, 2009; Dietz, Chompalov, Bozeman, Lane, & Park, 2000; Gaughan & Bozeman, 2002; Laurance et al., 2013). Additionally, the scientific CV is moderately standardized, often following a similar format and including similar information across researchers. However, there are also several challenges associated with CV analysis: they can be overly positive/truncated, dated, and tedious to analyze reliably. The time it takes to analyze CVs has been cited as the most prevalent issue in the use of the CV as data. Despite these challenges, CVs have been used alongside other methods of data collection (*e.g.*, demographic surveys, h-indices) to investigate contributions to research success(Cañibano & Bozeman, 2009; Dietz et al., 2000; Gaughan & Bozeman, 2002; Laurance et al., 2013). In this work, we will be analyzing researchers’ CVs alongside survey data to determine what contributes to success in scientific research. We will incorporate lexical analysis and the researchers’ Canadian Commons CVs to limit the time associated with CV analysis.

Our main question: In what ways are measures of chemistry research success associated with English language proficiency, including:

1. Traditional measures of success, such as publication rate, presentations, grant success rate, and research-based awards.
2. Non-traditional measures of success, such as service work (committees), teaching, and non-research-based awards.

* Future work: model could be used to provide feedback on CV structure

**Additional reading**

* Resume parsing

1. Sanyal, S.; Ghosh, N.; Hazra, S.; Adhikary, S. Resume Parser with Natural Language Processing. *Int. J. Eng. Sci. Comput.* **2017**, *7* (2), 4484–4489.
2. Kopparapu, S. K. Automatic Extraction of Usable Information from Unstructured Resumes to Aid Search. *Proc. 2010 IEEE Int. Conf. Prog. Informatics Comput. PIC 2010* **2010**, *1*, 99–103.

* CV analysis

1. Laurance, W. F.; Useche, D. C.; Laurance, S. G.; Bradshaw, C. J. A. Predicting Publication Success for Biologists. *Bioscience* **2013**, *63* (10), 817–823.
2. Gaughan, M.; Bozeman, B. Using Curriculum Vitae to Compare Some Impacts of NSF Research Grants with Research Center Funding. *Res. Eval.* **2002**, *11* (1), 17–26.

**Questions for Pavan and Dr. Inkpen**

* What do you think is the best way to move forward on this? Timeline?
  + Feasible?
    - Yes. Not something new.
    - Challenge will be unstandardized structure
  + Test set—compare against manual coding
  + Pavan aims to complete ~January/February
    - Final report due April
  + Jacky help with
    - Building dataset (collecting CVs)
* Possible to automate online CV retrieval?
  + LinkedIn
  + Can use existing datasets with CVs (more the better)
  + Can manually check for relevance (can automate by locating keywords “chemistry”, BERT?)
  + Can create a general model
    - Retrieve information from general CVs and then apply the general model to our specific research goals
  + Need to transform CVs into text (e.g. PDF to text)
  + Non-English CVs
    - Translation issues could make model building more difficult
* How many needed to develop initial model? Or are there existing models we can use?
* What is feasible in terms of data organization? e.g. number of presentations vs. number local presentations, national presentations, international, etc.?
* For publications, possible to incorporate journal impact factors when the model registers each publication? What about authorship (first, second, third)?
* Differentiating between different types of awards (research vs teaching vs service)
* Online CV’s that could be included in pilot to develop/refine model
* <https://mysite.science.uottawa.ca/organ/content/curriculum-vitae>
* <https://mysite.science.uottawa.ca/dfogg/cv.htm>
* <https://static1.squarespace.com/static/5a4bde8bcd39c3d9f4210aeb/t/5c86b41bf9619a45342beabf/1552331804154/LUMB_CV_Website.pdf>
* <http://www.flynnresearchgroup.com/alison-cv>