2021 IEEE 9th International Conference on Healthcare Informatics (ICHI) | 978-1-6654-0132-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICHI52183.2021.00075

Designing a core competency framework for U.K. clinical informaticians

Alan Davies
Faculty of Biology Medicine & Health
University of Manchester
Manchester, UK
alan.davies-2@manchester.ac.uk

Julia Mueller
MRC Epidemiology Unit
University of Cambridge
Cambridge, UK
julia.mueller@mrc-epid.cam.ac.uk

Alan Hassey
Faculty of Clinical Informatics (FCI)
London, UK
alanhassey@fastmail.com

Georgina Moulton
Faculty of Biology Medicine & Health
University of Manchester
Manchester, UK
Health Data Research UK
London, UK
georgina.moulton@manchester.ac.uk

***Abstract*—This paper outlines the design of a national Core Competency Framework for Clinical Informaticians in the United Kingdom. We provide a brief overview of the iterative process used to create the framework, focusing on the natural language processing methods used to automate the screening of job posts that fed into the information used to generate the framework.**

***Index Terms*—Core competency framework, informatics professionals, job listings, NLP, textmining**

I. Introduction

This work built on earlier work carried out by the UK Faculty of Clinical Informatics (FCI) to define Clinical Informatics [1]. An iterative approach was used to develop and refine the framework. This involved the creation of a draft framework by combining data from multiple sources. A systematic literature review [2] was carried out to identify the core competencies applicable to the different informatics domains (e.g. nursing informatics, pharmacy informatics, medical informatics, etc.). The review identified a large number of papers (n=10,044) with 82 papers included in the final review. This information was combined with analysis of over 50 informatics-related job posts from the NHS jobs website to provide an industry perspective. Once the draft framework was developed, it was shown to informatics experts (n=15) and adjusted based on their feedback to produce a second draft. This was then distributed via a digital survey completed by 87 respondents. An overview of the entire process can be seen in Figure 1.

Information about the details of the survey and interviews are published elsewhere [3]. This paper focuses on the job listing analysis carried out to provide data from the industry perspective, supplementing the academic information derived form the systematic review.

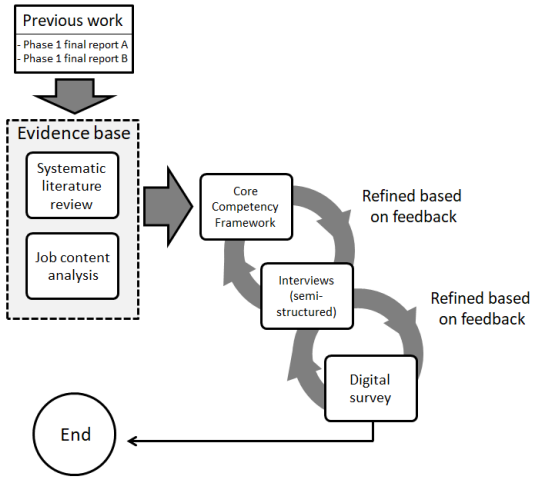

Fig. 1. Overview of framework development process

II. Methods

Job listings were manually downloaded from the NHS jobs website. Details from the job specifications were pasted into separate text files consisting of the following categories:

- Duties and responsibilities
- Abilities
- Qualities
- Experience and knowledge (essential/desirable)
- Qualifications (essential/desirable)
- Skills (essential/desirable)

An automated analysis was then applied to these files using Natural Language Processing (NLP) methods, including word frequency, bigram and trigram analysis. These techniques allow us to explore large volumes of textual data for any

978-1-6654-0132-6/21/$31.00 ©2021 IEEE
DOI 10.1109/ICHI52183.2021.00075

existing patterns/trends. Prior to analysis, a text corpus was created using the R package *tm* [4]. The corpus was then pre-processed by converting textual data to lower case and stripping out numbers, punctuation and stop-words (commonly occurring words such as 'the', 'and' etc.). The documents were then converted into a document-term matrix where documents are represented in columns and terms in rows. These could then be plotted as histograms to show the frequency of words in the documents. This allowed us to visualize the most mentioned skills, abilities, qualifications etc.

In addition to extracting the word frequency, we also carried out n-gram analysis. This allowed us to explore how terms co-occur with one another and how they are related to other terms. We applied this to adjacent words to determine the frequency of the co-occurrence of terms (e.g. 'information' and 'governance'). A pair of tokens (words) is known as a bigram (n=2), whereas a triplet of tokens is referred to as a trigram (n=3) [5].

This helped to provide some context in terms of which words appeared with other words. To manage the volume of words and help us to rapidly find associations, we plotted the results of the n-gram analysis using network plot visualisations.

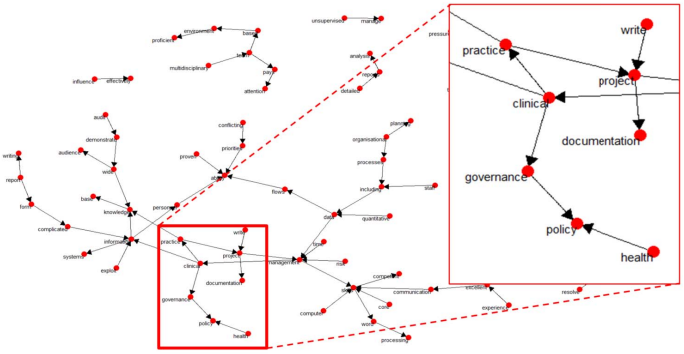

Fig. 2. Example bi-gram visualisation (skills)

III. Results

Figure 2 shows an example of the bigram analysis on the 'job skills' document. Here we see terms such as 'clinical' associated with 'practice' and 'governance', 'governance' with 'policy', and 'policy' with 'health'. This aids in the understanding of the sometimes complex relationships between terms. This also reveals the hard and soft skills that are required and their commonality across multiple job adverts.

This can be visualised as histograms as shown in Figure 3 depicting the word frequency for desirable qualifications such as Agile, management experience and the use of methods such as PRINCE project management.

IV. Discussion

The systematic literature review gave us an overview of the academic perspective on informatics. To overcome the theory-practice gap, this was combined with terms discovered by

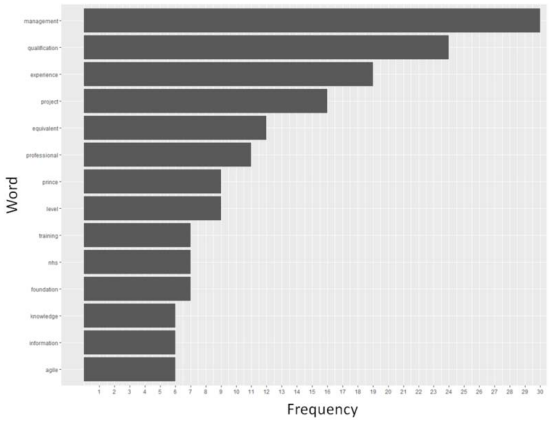

Fig. 3. Desirable qualifications word frequencies

the job listings analysis to capture any differences between industry and academia and provide a more complete picture. A combination of automated and manual methods produced the best results. Many jobs are structured in a similar format with a focus on qualifications, skills, experience and knowledge etc. This made it easier to extract these components and amalgamate them across multiple jobs. The automated analysis allowed us to rapidly extract terms from a large quantity of text. This was further enhanced through use of visualisation techniques to aid in analysis and further strengthened by combining the results with the systematic literature review, interviews, survey and iterative approach to create the final framework (which can be viewed on the FCI's website [6]).

V. Conclusion

The final framework was developed robustly utilising a mixed methods approach to framework generation using a mixture of qualitative and quantitative methods with an iterative approach to capture informatics requirements.

Acknowledgment

We wish to thank the Faculty of Clinical Informatics for funding this work

References

[1] N. Quinn, A. Hassey, L. Jidkov, "Development of Core Competencies for Clinical Informaticians in the United Kingdom: Final Report a (V1.1)", 2019

[2] A. Davies, J. Mueller, G. Moulton, "Core competencies for clinical informaticians: A systematic review" International Journal of Medical Informatics 141, 2020

[3] A. Davies, J. Mueller, A. Hassey, G. Moulton, "Development of a Core Competency Framework for Clinical Informatics" BMJ Health and Care Informatics, 2021 [IN PRESS]

[4] I. Feinerer, K. Hornik, "tm: Text Mining Package" R package version 0.7-8, 2020

[5] A. Müller, S. Guido, "Introduction to Machine Learning with Python: A Guide for Data Scientists" O'Reilly, 2016

[6] Faculty of Clinical Informatics, "Competency framework for Clinical Infotmatics" https://facultyofclinicalinformatics.org.uk/core-competency-framework, Accessed: 21-12-2020, 2020