Optimising E-portfolio's through the means of xAPI and Entity Extraction of Job Advertisements.

Alan Mark Berg
ICTS, University of Amsterdam
a.m.berg@uva.nl

Gábor Kismihók
Education and Labour Analytics Research group , TIB Hannover
Gabor.Kismihok@tib.eu

ABSTRACT: Within this hackathon theme, we explore the optimisation of e-portfolio's: Firstly, how entity extraction of authentic tasks from Vacancy data eases the pain of adoption and improves the learner's context. Secondly, how the standardising of xAPI e-portfolio profiles enable us to obtain scalable datasets for predictive models.

Keywords: E-Portfolio, Job Market Intelligence, Barriers to Adoption, Vacancy mining, Big Data, Entity Extraction

1   INTRODUCTION

Highly scalable e-portfolio systems exist including the open source systems such as Mahara (Gerbic & Maher, 2008), the next generation Open Portfolio System (Cambridge et al., 2008) and it's successor Karatu[1]. However, studies have consistently reported a negative perception of learners to e-portfolios (Rahayu & Sensuse, 2015). Issues include the difficulty of use for students, teachers and mentors, the quality of mentorship and the stability and complexity of the underlying system. In this hackathon, we explore how to make the student, teachers and mentors life easier through the application of Learning Analytics (LA) and Job Market Intelligence (JMI).

E-portfolio's have a potential to intersect with Learning Analytics, for example, Aguiar et al. (2014) showed how to use the activity of learners in e-portfolios to significantly improve the prediction of student dropout. Enabling e-Portfolio's through the xAPI protocol is a means of standardising the capture of the student digital trace and through this means eases the comparison between predictive models across organisations and encapsulation of learning moments, etc. Beckers, Dolmans, and Van Merriënboer (2016) suggested four central e-portfolio related themes: 1) Mandated as a dossier, which details achievements that an employer can hold against a standard. 2) The reflective e-portfolio, which is a voluntary version of the dossier and involves self-assessment on how the learner views themselves. This type of e-portfolio is often used when arguing for promotion. 3) The training e-portfolio, which aims at keeping track of learning, especially for employers and 4) the voluntary

---

[1] http://karutaproject.org

version is the personal development e-portfolio. We ask: *How do we capture these different usages through xAPI profiles and reuse the traces captured within predictive models?*

The authors acknowledge that E-portfolios are deployed for many different purposes and that scenario's 1) Detailing achievement and 3) keeping track of learning is probably the most straightforward situation's to define within xAPI statements/profiles. However, to aid in expanding the range we intend a) to provide a real E-portfolio system so the audience can experiment with the realistically structuring usage of portfolio's and b) offer a simple synthetic data generator (Berg, Mol, Kismihok, & Sclater, 2016) that is easily configurable to add new statements and then populate xAPI compliant applications. Through these means, we can quickly plug and play with the broader infrastructure provided for example by JISC. Another scenario is to plug and play with already existing recipes supplied for Goal setting (Berg, Scheffel, Drachsler, Ternier, & Specht, 2016). The audience can then experiment with streamlining interaction with Job Market Intelligence by encapsulating new events in the Job a market, such as newly discovered skills in a set of job adverts, again through the means of the xAPI standard.

In text mining, entity and feature extraction is a mature field with many decades of published research (Ramya et al., 2017; Nadeau & Sekine, 2007). Kobayashi et al. (2017a), provides a tutorial in the text analysis of job adverts, including code and data on how to perform extraction within the context of organisational research with specific reference to gathering the skills necessary for Nursing. Karakatsanis et al., 2017, conducted a broad scan of the job market for skill shortages based on the O-net database and Latent Semantic Indexing (LSI) and verified via crowdsourcing the accuracy of their approach and identified occupation clusters around skills, which are highly relevant within the collection of e-portfolios. Rahayu and colleagues (2017) reviewed recommendation systems that positively influence e-portfolio personalisation and concluded that researchers had shown the value of Hybrid and Collaborative Filtering, choosing the most relevant tasks together through online voting or tracking choice. One can easily imagine the content of Job adverts as a source of authentic tasks which feed the options for students as part of the e-portfolio recommendation process.

Based on the literature we suggest the following **research questions**:

> **RQ1:** *What is the definition(s) of an authentic task in the context of e-portfolio's?*
> **RQ2:** *How do we apply machine learning techniques to the extraction of authentic tasks?*
> **RQ3:** *How do we populate Karuta an open source e-portfolio system with authentic tasks?*
> **RQ4:** *What are the definitions of xAPI profiles for Job Market Intelligence enriched e-portfolio systems?*
> **RQ5:** *Which variables captured by xAPI profiles describe the most variance in predictive models for student success?*

Logistics: Monsterboard has kindly donated 20,000 job advertisements. The authors have converted the data to Rdata and CSV format. We will also provide links and supporting help documentation for the installation of an opensource e-portfolio system with pointers for data scientists and programmers.

2 IMPACT

By populating a highly visable open source e-portfolio system with authentic tasks we signal to the market that entity extraction of job adverts is a viable approach to personalising the learner e-portfolio context. We expect this will lead to an accelerated market adoption, thus decreasing barriers to the acceptance by students and teachers of e-portfolio systems due to their at times unnecessary complexity and burden of use. Further, we extend the likelihood of further personalisations of e-portfolios.

Through the integration of e-portfolio systems with LA and JMI xAPI standards-based infrastructure, we increase the opportunity to compare and thus improve the accuracy and range of predictive models and the range of supporting services provided.

3 ACKNOWLEDGEMENTS

4 REFERENCES

Aguiar, E., Ambrose, G. A., Chawla, N. V., Goodrich, V., & Brockman, J. (2014). Engagement vs Performance: Using Electronic Portfolios to Predict First Semester Engineering Student Persistence. *Journal of Learning Analytics*, *1*(3), 7–33. https://doi.org/10.18608/jla.2014.13.3

Berg, A., Scheffel, M., Drachsler, H., Ternier, S., & Specht, M. (2016). The dutch xAPI experience. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK '16* (pp. 544–545). New York, New York, USA: ACM Press. https://doi.org/10.1145/2883851.2883968

Berg, A., Mol, S., Kismihok, G., & Sclater, N. (2016). Scaling Learning Analytics: The practical application of synthetic data. In *Proceedings of the 9th European Distance and E-Learning Network Research Workshop* (p264-269).

Beckers, J., Dolmans, D., & Van Merriënboer, J. (2016). e-Portfolios enhancing students' self-directed learning: A systematic review of influencing factors. *Australasian Journal of Educational Technology*. https://doi.org/10.14742/ajet.2528

Cambridge, D., Fernandez, L., Kahn, S., Kirkpatrick, J., & Smith, J. (2008). The impact of the open source portfolio on learning and assessment. *MERLOT Journal of Online Learning and Teaching*, *4*(4), 492-502.

Gerbic, P., & Maher, M. (2008). Collaborative self-study supporting new technology: The Mahara e-portfolio project. *Proceedings of ASCILITE Melbourne*, 320-324.

Karakatsanis, I., AlKhader, W., MacCrory, F., Alibasic, A., Omar, M. A., Aung, Z., & Woon, W. L. (2017). Data mining approach to monitoring the requirements of the job market: A case study. *Information Systems*, *65*, 1–6. https://doi.org/10.1016/j.is.2016.10.009

Kobayashi, V. B., Mol, S. T., Berkers, H. A., Kismihók, G., & Den Hartog, D. N. (2017a). Text Classification for Organizational Researchers. Organizational Research Methods, 1094428117719322. https://doi.org/10.1177/1094428117719322

Kobayashi, V. B., Mol, S. T., Berkers, H. A., Kismihók, G., & Den Hartog, D. N. (2017b). Text Mining in Organizational Research. Organizational Research Methods, 1094428117722619. http://dx.doi.org/10.1177/1094428117722619

Nadeau, D., & Sekine, S. (2007). A survey of named entity recognition and classification. Lingvisticae Investigationes, 30(1), 3-26. . https://doi.org/10.1075/li.30.1.03nad

Rahayu, P., Sensuse, D. I., Purwandari, B., Budi, I., Khalid, F., & Zulkarnaim, N. (2017). A Systematic Review of Recommender System for e-Portfolio Domain. In *Proceedings of the 5th International Conference on Information and Education Technology - ICIET '17* (pp. 21–26). New York, New York, USA: ACM Press. https://doi.org/10.1145/3029387.3029420

Ramya, R. S., Venugopal, K. R., Iyengar, S. S., & Patnaik, L. M. (2017). Feature Extraction and Duplicate Detection for Text Mining: A Survey. *Global Journal of Computer Science and Technology*, *16*(5). Retrieved from https://www.computerresearch.org/index.php/computer/article/view/1459