Panel session: How could we build a community for sharing annotation activities, methods, and outcomes?

Emma L. Tonkin
Email: e.l.tonkin@bristol.ac.uk
Digital Health, School of Electrical and Computer Engineering, The University of Bristol, Bristol, UK
Kristina Yordanova
Email: kristina.yordanova@uni-rostock.de
Institute of Computer Science, University of Rostock

Panel Abstract

Creation and testing of methods, tools, vocabularies, taxonomies, and ontologies for annotation of user data have been reported in many places, including prior editions of this workshop [1]. How can we work together to increase the impact and visibility of these outcomes?

In this panel, we discuss means to develop the sustainability, reusability and ultimately the impact of our work. In so doing, we draw insight from the approaches taken in other domains. For example, the fields of metadata and subject indexing publish and regularly update design artefacts such as taxonomies, ontologies and knowledge schemas. The sustainability of data publication is limited by the long-term availability of the platforms on which they are published and the technologies on which they depend. A prior study has found that many knowledge structures are not formally published, leading to low preservation of these artefacts [2]. Standardisation efforts may significantly increase the likelihood of uptake and ongoing availability of these artefacts, which are expensive [3] and often time-consuming to create. What can we learn from practices in other fields, or gain from making use of the outcomes of prior investigations in this area? Knowledge structures take time and effort to produce, test and validate: are there emerging 'de facto' standards in the field that may benefit from being brought forward toward formal standardisation?

As instance metadata, annotated datasets are highly valuable as resources in pervasive computing, with significant potential to advance the field. However, annotation of real-world user data additionally raises problems that arise less frequently in classical information management, such as ethical constraints and risks to participant privacy. What practices exist that may help us to appropriately balance participant privacy and relevant legislation against the need to make the best use of participant data, maximising the impact of their participation and ensuring that their effort is not wasted?

Tertiary outcomes, including software such as annotation platforms, validators and data visualisation methods or frameworks, are equally of value and are likely to increase both in complexity and in relevance to the full lifecycle of data annotation pipelines in production. For example, there is an increasing expectation in current legislation that decision-making by automated systems be *explainable*, openly communicating system strengths and limitations so that participants may make informed decisions about data processing. There is a corresponding strand of research that shows that the public are cautious users of 'AI'-focused applications in areas such as healthcare [4], whilst the effectiveness of interventions may depend on user acceptance in these application areas [5]. Practical applications require a human-centric design focus. How best can this activity be supported and documented within the community?

References

[1] K. Yordanova, A. Paiement, M. Schröder, E. Tonkin, P. Woznowski, C. M. Olsson, J. Rafferty, and T. Sztyler, "Challenges in annotation of user data for ubiquitous systems: Results from the 1st arduous workshop," arXiv preprint, Tech. Rep. arXiv:1803.05843, March 2018. [Online]. Available: https://arxiv.org/abs/1803.05843

[2] E. L. Tonkin, "Supporting unsupervised context identification using social and physical sensors," Ph.D. dissertation, University of Bristol, Bristol, UK, 2015.

[3] H. Alani and C. Brewster, "Metrics for ranking ontologies," in *Evaluating Ontologies forthe Web Workshop (EON2006), 15th International World Wide Web Conference*, 2006.

[4] T. Nadarzynski, O. Miles, A. Cowie, and D. Ridge, "Acceptability of artificial intelligence (ai)-led chatbot services in healthcare: A mixed-methods study," *Digital health*, vol. 5, p. 2055207619871808, 2019.

[5] A. Naidu, "Factors affecting patient satisfaction and healthcare quality," *International journal of health care quality assurance*, 2009.