ECP: Evaluation Community Portal
A Portal for Evaluation And Collaboration in User Modelling and Personalisation Research

Kevin Koidl, Killian Levacher, Owen Conlan
ADAPT, KDEG,
School of Computer Science & Statistics
Trinity College, Dublin, Ireland
{Kevin.Koidl, Killian.Levacher, Owen.Conlan}@scss.tcd.ie

Ben Steichen
Department of Computer Engineering
Santa Clara University
Santa Clara, CA, USA
bsteichen@scu.edu

ABSTRACT
Researchers conducting evaluations in the fields of User Modelling and Personalisation face the challenge of missing continuing evaluation feedback and collaboration with the overall research community. This missing ability results in limitations such as missing feedback on evaluation approaches, missing insight into other potentially usable evaluation results, and the lack of creating shared evaluation tasks. This paper introduces a community portal *ECP: Evaluation Community Portal*, which is specifically focused on evaluations within the UMAP community (User Modeling, Adaptation, and Personalisation)

CCS Concepts
General and reference~Cross-computing tools and techniques~Evaluations

Keywords
Evaluation;Shared;Evaluation;User Modelling;Research Community Portals;Personalisation

1. INTRODUCTION
Researchers conducting evaluations in the fields of User Modelling and Personalisation face the challenge of missing continuing evaluation feedback and collaboration within the overall research community. This missing ability results in limitations such as missing feedback on evaluation approaches, missing insight into other potentially usable evaluation results, and the lack of shared evaluation tasks to compare different user modeling and personalization approaches.

Other research areas, such as in Information Retrieval (IR) through the TREC and CLEF initiatives, have managed to overcome these barriers by creating evaluation campaigns with shared evaluation tasks, as well as community portals containing shared datasets. Another example of a successful research community portal is the well-known CFP (Call for Paper) wiki, which is an established portal for finding information related to upcoming conferences. Both examples serve as a clear indicator that community portals within and across research communities can serve as a vehicle to overcome limitations and boundaries due to lack of central communication and outreach abilities.

This paper introduces a community portal *ECP: Evaluation Community Portal*, which is specifically focused on evaluations within the UMAP community (User Modelling and Personalisation), aiming to serve as a place for the creation and discussion of shared evaluation tasks from design to results. Furthermore, the portal seeks to provide result data set access to expand on other research and discuss previously conducted work.

The goal of this paper is to spark a discussion on how the proposed portal would assist the UMAP research community and what mechanism would have to be put in place to create and promote such a portal approach.

2. RELATED APPROACHES
Despite a well established User Modelling, Adaptation and Personalisation (UMAP) community, many fundamental evaluation challenges still remain to be solved.

Repeatedly obtaining a sufficiently large number of users to evaluate prototypes is a recurring theme, very familiar across research institutions [1]. In order to overcome this issue, many researchers in the field of Human-Computer Interaction have started using crowdsourcing platforms such as Amazon Mechanical Turk[1] or Crowdflower[2] to perform usability studies. Indeed, the use of such platforms has been shown to be a good substitute for general lab-based usability studies [7][8]. However, the nature of systems and experiments in the field of User Modeling and Personalization typically require prolonged user exposure and interaction with a system in order to i) build accurate user models and ii) truly gauge the effectiveness of personalization techniques, which is often infeasible given the typically short interaction paradigm and setup of crowdsourcing platforms.

Additionally, the ability to assess aggregated research results over time is also hampered by the fact that evaluations are mostly carried out in isolation from each other and are usually not easily reproducible or directly comparable [2, 3], which affects the ability to produce rigorous comparative evaluations between individual systems produced. For example, while there has been substantial work over the last two decades in the development of novel adaptive and personalized e-learning systems, the various research prototypes have generally not been compared to each other through standardised evaluation campaigns.

Within the existing wider research community, two well-established community-based practices are worth pointing out. The first consists of the Call For Papers (CFP) wiki [4], whose

[1] https://www.mturk.com

[2] http://www.crowdflower.com

main purpose is to allow researchers to advertise conference venues, paper submission deadlines, etc. This community-driven platform serves the purpose of both i) *centralising the outreach* needs of the community with respect to a shared unique goal (i.e. attracting as many research submissions as possible), as well as ii) inviting individual researchers to *contribute to the list of venues available* in each field.

Considering the recurrent need for large number of users in each UMAP evaluation, it is surprising that no equivalent platform exists for the purpose of evaluation within the community. As of today, there is no central location in which to advertise individual UMAP evaluation calls. Evaluation calls are mostly performed through dedicated institution-wide or field specific research mailing lists[3] to which one needs to subscribe. As a result, the wider research community and general public is often unaware of these calls. An equivalent ECP wiki platform would not only centralise and simplify the process of advertising on-going evaluations within each field of personalisation, it could also contribute to the larger evaluation needs and analysis of the community through the a-posteriori publication of datasets, evaluation metrics and results for each experiment.

The second community-based practice of interest consists of the CLEF [5] and TREC [6] shared tasks initiatives. As part of these tasks, separate systems are designed within the context of a common set of evaluation constraints (eg: common scenario, dataset, metrics etc.) and users to compare each approach proposed. In addition to pooling resources, which lets researchers focus their efforts on developing their systems, this approach embeds comparative evaluation as the core evaluation strategy. Again, the UMAP community lacks such shared tasks and therefore similar research prototype systems are typically not compared to each other through a rigorous process. A CFE platform, as proposed above, could be augmented to form the basis for the creation of similar tasks within the personalisation community. Existing evaluation datasets and results published on the platform could organically increase the number of independent evaluations being carried out upon identical datasets, eventually leading to dedicated shared evaluation tasks.

3. PORTAL OVERVIEW

Based on the discussion above we propose a community focused portal, which is inspired by work done within CLEF, and is based on the simplicity of CFP. We propose the following key features as a starting point for this community effort:

- Ability to post calls for participation in evaluations. This feature, which is similar to CFP, requires the linking to surveys and online systems where the evaluation can be conducted.
- Ability to discuss approaches and findings in a forum manner. This may include following evaluations and/or discussions to receive notification on status and outcome.
- Ability to upload and present data that can be shared and used in other evaluations.

ECP will require substantial community-driven effort to ensure it remains useful and impactful. For this reason, the portal has to be designed in an open and simple fashion by using easy to implement and extensible platforms such as Content Management Systems or Wikis. Similar to other community efforts, the portal does not require a central structure or organisation once the basic portal is established. Its growth and success depends mostly on researchers to pick up tasks and extend the portal where needed.

[3]User modelling mailing list: https://www.di.unito.it/listserver/subrequest/um, Adaptive Hypermedia mailing list: http://pegasus.tue.nl/mailman/listinfo/ah

4. CONCLUSION

Based on the challenges of evaluation in User Modelling and Personalisation we propose a community driven portal introduced as ECP (Evaluation Community Portal). We discussed the overall motivation to this topic and related projects successfully applied in other research communities such as Information Retrieval. We furthermore introduce a brief overview of required high level features. We envisage that the main challenges related to ECP will be in bootstrapping the Portal and gaining initial community momentum. Like any community lead approach it requires a certain amount of traction to ensure it is widely used across different research institutes. Furthermore, an initial task force (community champions) leading these efforts needs to be identified which should include more than one research institute across more than one continent.

5. ACKNOWLEDGMENTS

The ADAPT Centre for Digital Content Technology is funded under the SFI Research Centres Programs (Grant 13/RC/2106) and is co-funded under the European Regional Development Fund

6. REFERENCES

[1] Paramythis, A., Weibelzahl, S., & Masthoff, J. (2010). Layered evaluation of interactive adaptive systems: framework and formative methods. *User Modeling and User-Adapted Interaction*, *20*(5), 383–453.

[2] Chen, L., & Pu, P. (2012). Experiments on user experiences with recommender interfaces. *Behaviour & Information Technology*, *33*(January), 1–23

[3] Hernández del Olmo, F., & Gaudioso, E. (2008). Evaluation of recommender systems: A new approach. *Expert Systems with Applications*, *35*(3), 790–804.

[4] Wiki Call For Papers: http://www.wikicfp.com/cfp/ Accessed on: 06/05/16

[5] CLEF 2016: http://clef2016.clef-initiative.eu/ Accessed on: 06/05/16

[6] TREC 2016: http://trec.nist.gov/pubs/call2016.html Accessed on: 06/05/16

[7] Aniket Kittur, Ed H. Chi, and Bongwon Suh. 2008. Crowdsourcing user studies with Mechanical Turk. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2008), pp. 453-456.

[8] Komarov, S., Reinecke, K., & Gajos, K. Z. (2013, April). Crowdsourcing performance evaluations of user interfaces. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2013), pp. 207-21