Research-Centres Centred Living Labs

Liadh Kelly

ADAPT Centre, Trinity College, Dublin, Ireland `Liadh.Kelly@tcd.ie`

Abstract. Developing means to conduct shared evaluation in the user modelling, adaptation and personalization (UMAP) space is inherently difficult. Not least because of privacy concerns and individual differences in behaviours between users of systems. In this paper we propose a *research-centre* centred living labs approach as one potential way to overcome these difficulties and to allow for shared task generation in the UMAP domain.

1 Introduction

Existing shared evaluation tasks using the living labs methodology, specifically *living labs for information retrieval evaluation* LL4IR [1] and CLEF NEWSREEL [2] , take an API-centred approach. Challenge participants plug their developed approaches into the challenge provided API. Live commercial systems can then communicate through the API to use, and hence test, challenge participants' algorithms (or techniques), in place of, or in conjunction with, the commercial systems algorithms (or approaches). Figure 1(a) depicts the LL4IR instantiation of this general architecture. However, living labs can also be interpreted in different ways.

2 Research-Centre Centred Living Labs Approach

A tool for creating a living lab that centres on research centres providing data and users for shared evaluation is presented in [3] (see Figure 1(b)). The paper focuses on a living lab for evaluation of retrieval techniques for personal desktop collections. In this approach, researchers wishing to evaluate their technologies would participate in a collaborative evaluation effort. Whereby required protocols and technology to gather data for, and to conduct the evaluation, would be distributed to the participating research centres. The retrieval algorithms/techniques developed by each participating research centre would also be distributed to the research centres for evaluation. Individual research centres would then recruit experiment subjects locally, who install and run the provided tool on their personal computer (PC). The tool indexes the items on the PCs. Using the provided protocols and tool, experiment subjects issue personal queries and conduct relevance assessment. Participating research centres' IR algorithms are evaluated locally on subjects' PCs using the generated index, queries and relevance assessments, with only performance measures returned to investigators thus preserving privacy.

2.1 Proposal

The high-level concept of this *research-centres* centred living labs approach could be generalized to allow for shared task evaluation in the UMAP space. Whereby evaluation goal specific tools and protocols, and challenge participants algorithms/software
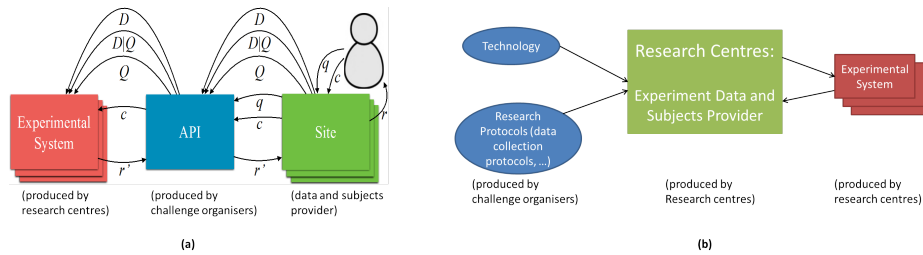
Fig. 1. (a) Schematic representation of the LL4IR API (adapted from [1]). Where, Q = frequent queries; (D—Q) = candidate documents for each query; c = user interactions with ranking r' for query q∈Q. **(b)** Schematic representation of user centric Living Labs API.

are distributed to individual research centres. These are then either used (as shown in Figure 1(b)) to: generate static collections for evaluations as described above; run controlled experiments with the participants' software locally in each research centre; or ideally run the participants' software live, for evaluation purposes, in place of individuals' typical software as they go about their normal activities. Or indeed a hybrid of this *research-centres* centred and the earlier *API-centred* approach might prove most useful in the UMAP space, depending on the precise scenario to be evaluated. This general evaluation paradigm has potential for evaluation of any tool or algorithm supporting individuals interacting with digital data on both mobile and stationary devices.

Realising such living labs requires addressing several challenges associated with living labs architecture and design, hosting, maintenance, security, privacy, participant recruiting, and scenarios and tasks for use development. Lessons can be learned here from the experiences of the IR and recommender systems living labs shared tasks [1, 2].

3 Conclusions

Current instantiations of living labs in IR and recommender system shared challenges focus on an API-centred living labs methodology. Other interpretations of living labs are also possible. In this paper we put forward the use of a research-centres centred living labs methodology for shared UMAP challenges. This research-centres centred living labs methodology could also have application in evaluation of other research spaces.

Bibliography

[1] K. Balog, L. Kelly, and A. Schuth. Head first: Living labs for ad-hoc search evaluation. In *CIKM'14*, 2014.

[2] F. Hopfgartner, B. Kille, A. Lommatzsch, T. Plumbaum, T. Brodt, and T. Heintz. Benchmarking news recommendations in a living lab. In *CLEF'14*, 2014.

[3] L. Kelly, P. Bunbury, and G. J. F. Jones. Evaluating personal information retrieval. In *Proc. of ECIR '12*, 2012.