

Unfolding cultural, educational and scientific long-tail content in the Web

Michael Granitzer¹, Christin Seifert¹, Silvia Russegger², and Klaus Tochtermann³

¹University of Passau, Germany

²Joanneum Research, Graz, Austria

³German National Library for Economics, Kiel/Hamburg, Germany

Abstract. This project poster introduces the recently started EEXCESS project, which aims at Enhancing Europes eXchange in Cultural Educational and Scientific Resource. Europe has digitised vast amounts of cultural, scientific and educational content like for example scientific research, historical sound recordings, images of sculptures, films and sheet music. However, since content dissemination on the Web is driven by a small number of large central hubs like social networks or search engines, this cultural and scientific treasures has hardly been recognized by the general public or utilized in scientific and educational processes. EEXCESS aims to develop personalized and contextualized recommendation technologies to augment existing content dissemination channels (e.g. social media) and content creation process (e.g. blogging) for distributing high-quality educational, scientific and cultural content. In this project poster we present the underlying idea and related work with focus on user modeling and personalized, context-aware recommendation.

Keywords: educational and scientific resources, cultural content, personalized recommendation, long-tail content

1 Motivation

In the past decade, Europe conducted tremendous effort for making cultural, educational and scientific resources publicly available. Based on national aggregators like Collections Trusts Culture Grid, initiatives like Europeana nowadays provide a plethora of cultural resources for people worldwide. Concurrently, the semantic web, particularly Linked Open Data, has been growing exponentially providing semantic-enhanced access to and interchange of interesting scientific and cultural resources. Similarly, young start-ups like Mendeley re-shaped the management of scientific resources in a web-centric manner and several educational platforms started to emerge. Although such massive amounts of culturally rich, educating content are available, the potential of its use for educational and scientific purposes remains largely untapped. The primary reason can be seen in todays Web content distribution mechanisms: content dissemination is dominated by a small number of large central hubs like major search engines (e.g. Google), social networks (e.g. Facebook) or online encyclopedias (e.g.

Wikipedia). However, much valuable content is only available in the long-tail (i.e. a theory arguing that in internet-based markets niche content adds up to a huge body of knowledge, but is hidden from most users). In the long-tail content is maintained and curated by a large number of small to medium-sized professional organizations such as memory organizations (e.g. archives and museums), digital libraries and open educational repositories. However, the few large web hubs hardly support disseminating this long-tail content.

In order to reshape content dissemination mechanisms for highly specialized long-tail content EEXCESS relies on augmenting existing web channels with high-qualitative content through personalized, contextualized and privacy preserving recommendations, as discussed in the following.

2 Approach

In our approach, which is presented in an overview in fig. 1, we differentiate between two content related processes in the web: **content consumption** like reading web pages and surfing the web and **content creation** like authoring web pages or social media content. We aim to inject high-quality, long-tail content into those processes through personalized recommendation techniques by so-called augmentation interfaces. Those interfaces unobtrusively inject recommendation results into existing web pages or browsers. For example, when reading a Wikipedia page or a blog post on a certain topic, users should be given additional background material depending on their level of expertise and task. Similarly, during content creation processes, we aim to support the authors in creating the entry by recommending background material from digital libraries like the ZBW¹ or Europeana², digital cultural aggregators like Collection Trust³ or scientific databases like Mendeley⁴.

As in research on knowledge services or just-in-time retrieval [3], personalization along with unobtrusive interfaces will become a corner stone for high user acceptance. Hence, thoughtful user interface design jointly with highly related recommendation will be one major research challenge.

High quality recommendations require personalization and contextualization of the recommendation engines. While machine learning combined with semantic technologies [2] have been successfully applied to this task, privacy considerations remain a crucial task in terms of user acceptance. Hence, EEXCESS aims to retain full user privacy and user control through estimating context mostly on the client and submitting only minimal necessary information to the recommender system. The trade-off between privacy and recommendation accuracy will be the second major research challenge.

Our final research challenge lies in the recommender system itself. Since large-scale recommendation will become cost-intensive, it is unlikely that one

¹ <http://www.zbw.eu/>

² <http://www.europeana.eu/>

³ <http://www.collectionstrust.org.uk/>

⁴ <https://www.mendeley.com/>

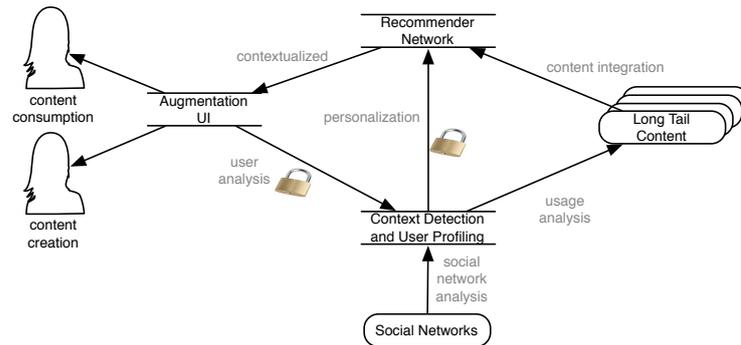


Fig. 1. Overview over EEXCESS envisioned content distribution process

institution will be able to run a single recommender. Therefore we aim to create a recommender network, in which every recommender is specialized on particular content and a subset of a user group. As discussed in the recommender community, like for example in [1], aggregating recommendation result over heterogeneous sources will become challenging in terms of accuracy and timeliness.

3 Impact

Although a large number of challenges have to be solved, very high impact can be expected by achieving the goals of EEXCESS. Particularly, content distribution process will become more open and less driven by big players or commercial interest. Moreover, through improved content creation we also expect to increase the scientific, educational and cultural quality of user generated content.⁵

References

1. Ivan Cantador, Peter Brusilovsky, and Tsvi Kuflik. Second Workshop on Information Heterogeneity and Fusion in Recommender Systems (HetRec2011). In *Proceedings of the fifth ACM Conference on Recommender systems, RecSys '11*, pages 387–388, New York, NY, USA, 2011. ACM.
2. Michael Granitzer, Mark Kröll, Christin Seifert, Andreas S. Rath, Nicolas Weber, Olivia Dietzel, and Stefanie Lindstaedt. Analysis of machine learning techniques for context extraction. In *Digital Information Management, 2008. ICDIM 2008. Third International Conference on*, pages 233–240. IEEE, 2008.
3. Andreas S. Rath, Nicolas Weber, Mark Kröll, Michael Granitzer, Olivia Dietzel, and Stefanie N Lindstaedt. Context-aware knowledge services. *Personal Information Management: PIM*, pages 5–6, 2008.

⁵ **Acknowledgments.** The presented work was developed within the EEXCESS project funded by the European Union Seventh Framework Programme FP7/2007-2013 under grant agreement number 600601.