Taking Libraries' Cultural Content to the User – Approaches and Experiences from the EEXCESS Project

Timo Borst

ZBW Leibniz Information Centre for Economics, Germany
t.borst@zbw.eu, orcid.org/0000-0002-2481-029X

Maren Lehmann

ZBW Leibniz Information Centre for Economics, Germany
m.lehmann@zbw.eu

Tamara Pianos

ZBW Leibniz Information Centre for Economics, Germany
t.pianos@zbw.eu, orcid.org/0000-0002-4386-519X

Nils Witt

ZBW Leibniz Information Centre for Economics, Germany
n.witt@zbw.eu

Abstract

Before the advent of the internet users had to physically visit libraries and other cultural institutions in order to search, find and access content. With the emergence and massive use of the World Wide Web and associated tools and technologies this situation has changed drastically. If content is to be found and used, it must be present where users do their daily digital work. The aim of the

EU project EEXCESS (http://eexcess.eu/) is to inject the content into the working environments of the users. In this paper, we show some of the approaches, use cases and technical implementations we developed in the project.

Key Words: cultural content; automatic recommendations; EU-project; privacy; Open Source

1. Introduction

Libraries and related cultural institutions used to have a monopoly on providing access to books or journals and to information in general. Users had to physically visit a place in order to find and access library content. Today, the internet and especially technologies like search engines, content hubs like Wikipedia and social media platforms have become means of information and content provision.

A lot of high quality content is freely available on the internet. However, users often cannot access all of the relevant content, because they either do not know it exists, or do not know the specialized search engines or databases they need to access, or do not have the domain knowledge to formulate appropriate search queries. Users very often start their searches either by using the big search engines, mostly Google, or they start their search in Wikipedia (Dzeyk, 2015; Rowlands et al., 2008). Thus, very specific items – like e.g. the image of a special historical garment – may be overlooked as they do not appear on the first two or three pages of the major search engines.

If libraries and other cultural institutions still want their content to be found and used, they must adapt to these new environments in which users expect digital content to be available and accessible directly and at once. Libraries have reacted to the new challenges in a number of different ways, e.g. by making their content more visible in search engines through search engine optimization (SEO), by aligning their online services with user expectations through discovery systems or by offering their services at the point of need via chat or mail.

The EU-funded project EEXCESS,[1] on the other hand, aims to provide another solution for this challenge: Libraries and cultural institutions can deliver their online content directly to the users who will be able to work with their favourite tools and still find information otherwise hidden.

Museums, archives, libraries, etc. possess immeasurable resources of cultural, scientific and educational content. Their holdings comprise of scientific research, historical sound recordings, images of sculptures, films, sheet music and much more. This highly specialized, carefully curated content is still largely invisible to the general public. In the internet context it is therefore often referred to as the long-tail (i.e. a huge body of specialized knowledge existing in the World Wide Web, but hidden from most users).[2]

The vision of EEXCESS is to connect these valuable resources with the main stream content available via internet giants such as Google, Facebook and Wikipedia. EEXCESS is based on the simple principle of taking the content to the user, not the user to the content. It wants to bring the information directly to the users' working environment, on their favourite platforms (Facebook, Twitter etc.) and their preferred devices (tablets, smartphones etc.). Instead of having to navigate through a multitude of libraries, repositories and databases, users should be able to find relevant and specialized information in their habitual environment. The long-tail content is brought to the surface, where it can be put to new and different uses.[3]

The general aim of EEXCESS is to 'inject' digital content into users' everyday work environments like browsers, content management systems, a word processing programme or e-learning environments. Content, thus, is *recommended* by means of an organizational and technical framework of distributed partner recommenders and user profiles.

The unique data sources in conjunction with EEXCESS technologies for recommendation and visualization offer new approaches to discover the best possible results for academic and cultural content.

Relevant academic and cultural content can only be recommended if it is covered by the EEXCESS content partners. Institutions with cultural or academic texts, images etc. can offer their content for inclusion in the EEXCESS recommender in order to make it more visible to potential users.

The EEXCESS project has 10 partners in four countries who are responsible for different aspects like the recommender, privacy issues or visualizations. The project started in February 2013 and ended in July 2016.

In this paper, we show some of the approaches, use cases and technical implementations we developed in EEXCESS. We outline some use cases

which we conceive as representative for our target groups (students and researchers).

Protecting the privacy of the users in connection with personalized recommendations is a major task and effort of the project, so we will briefly cover the implications.

We conclude the paper by depicting some of the approaches towards achieving dissemination and sustainability of the project results.

2. Application Scenarios: Use Cases

The EEXCESS software prototypes work in different ways to support online research (*content consumption*) like looking for information on a particular topic or conducting studies. They also help with writing or editing a Wikipedia article, a blogpost, a Moodle entry or a Google Docs document (*content production*). At the beginning of the project, numerous use cases were identified to cover realistic examples of how students, teachers, researchers or librarians go about their work of searching for information or sorting content etc.

2.1. Content Consumption

As an example of *content consumption*, we take a student researching a topic for a presentation. After a quick search in Google she reads a Wikipedia article in her browser (Figure 1). During the reading process, further information matching the topic of the article – or a section of it – is automatically recommended by EEXCESS.

This example shows a relatively short Wikipedia article on the Ergolz, a river in Switzerland. When users search for Ergolz on Google they find the corresponding Wikipedia entry. But they also find a local hospital, tourist information, a Veterinarian as well as a miniature golf course. The EEXCESS extension on the right side of the screen shows pictures and documents on this topic from specialized cultural and academic databases that would otherwise remain hidden in the long tail of the internet. This example works well in EEXCESS because a museum of this region in Switzerland (Archäologie und

*Fig. 1: Wikipedia and Chrome extension.*
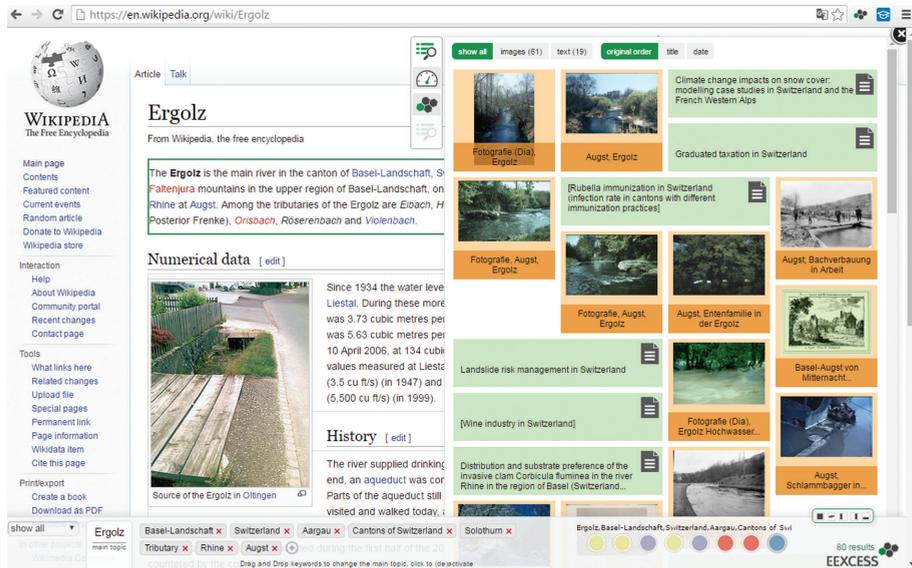
Museum Baselland) is one of the project partners and early content providers. The more content providers have their content included in the EEXCESS recommendation, the more satisfying results will probably be delivered to searches in different contexts.[4]

2.2. Content Production

As an example of *content production*, we consider e.g. the creation and publishing of a blog post, which is supported by automatic recommendations of related documents. After writing the first few lines or paragraphs in the WordPress editor, the EEXCESS plugin can be activated to provide context sensitive recommendations. These documents can easily be inserted and cited in the blogpost using different citation styles. In the same fashion, the plugin also allows users to easily embed images, e.g. for illustrating a blog post (Figure 2).

Another example for content production would be the writing or editing of a Wikipedia article. The Wikipedia Reference Butler (WRB), another

Fig. 2: WordPress Plugin: A selected image is inserted in the text.
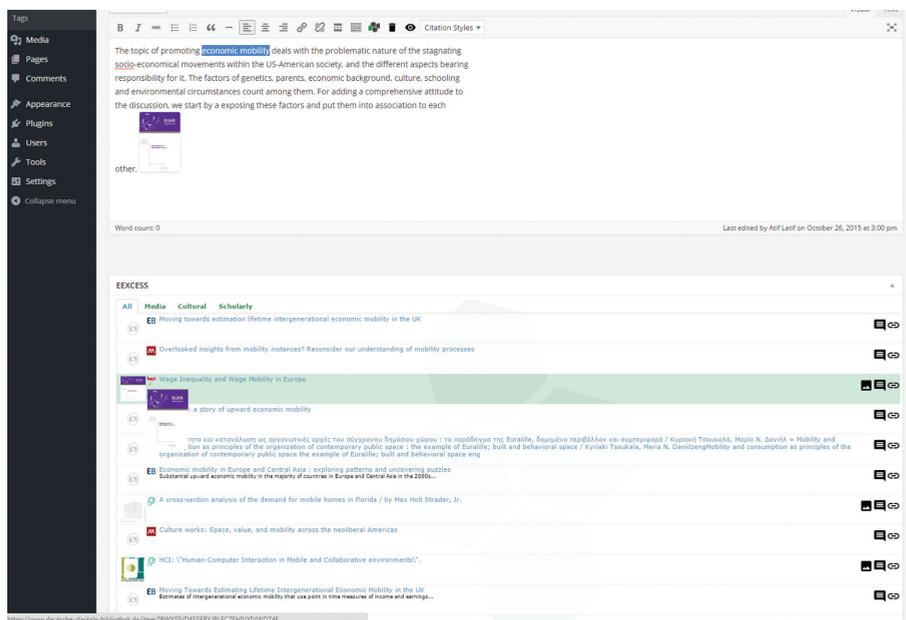
EEXCESS prototype supports editors by showing suitable media-files from WikiCommons (media-container of Wikimedia) to be added to articles and it shows suitable references to scientific literature matching the context of the article.[5]

3. Toolbox: EEXCESS Prototypes, Content, and Partner Wizard

All applications developed in the project are available through the EEXCESS website and can be used freely. The code for the prototypes is freely available under the terms of the Apache 2.0 Licence.

3.1. EEXCESS Prototypes

A wide range of prototypes was developed in the project in order to cover a couple of content consumption and production scenarios.[6] The prototypes and modules are:

- Chrome Browser Extension
- WordPress Plugin
- Google Docs Plugin
- Moodle Plugin
- Wikipedia Reference Butler
- Recommendation Dashboard

The most elaborate EEXCESS prototype is a Chrome browser extension. An unobtrusive icon can be used to trigger the recommendations and change the settings etc. The window showing the recommendations can be resized according to screen size and personal preferences. Multiple options for filtering the results are available to find specific (like e.g. images or publications from a specific era) or serendipitous results.

The Chrome extension was optimized for the use in Wikipedia and works best with English language results, but it can also be used with any other website opened in the Chrome browser.[7]

Depending on the software component installed on either a user's local machine or an application server, the list of recommendations is displayed in different ways: from a classical, text-oriented list, to a visualization of metadata records (Figure 3). Different visualizations can be used to break

Fig. 3: One display option in the Chrome extension Visualization Dashboard.
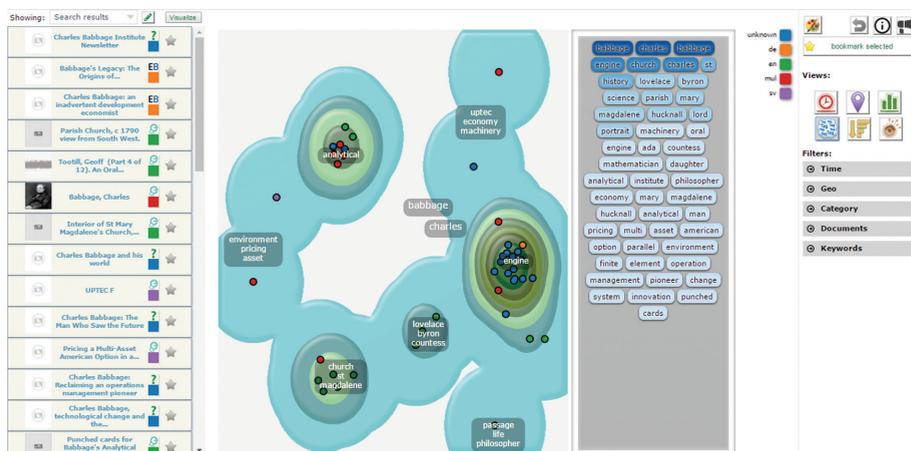
down the number of recommendations (Veas, Mutlu, di Sciascio, Tschinkel, & Sabol, 2015).

3.2. EEXCESS Content

Some of the project partners also act as content providers: Museum Baselland, Mendeley, and EconBiz/ZBW. Another early content provider is Europeana. During the last year of the project a number of other content providers were included: The Digital Public Library of America, The National Archives UK, Swissbib, Deutsche Digitale Bibliothek, Rijksmuseum – The Museum of the Netherlands, and Core.ac.uk, so that EEXCESS can now recommend contents from ten different providers covering a wide range of cultural and academic topics. More than 150 Mio items can be found through EEXCESS in August 2016.

New content can be easily submitted for integration by anyone who owns a repository or database with academic or cultural content accessible through an Application Program Interface (API) supporting search functionalities.

3.3. Submitting Content for Inclusion and Partner Wizard

Institutions are encouraged to make their digital content and metadata discoverable through the EEXCESS recommendation technology. EEXCESS is not a content aggregator or archive, but a broker between cultural, educational and academic content and users. Providers who want to provide suitable content need to have a search API and comply with some metadata standards (Orgel, Höffernig, Bailer, & Russegger, 2015).[8]

EEXCESS accepts metadata about items in partner collections in LIDO or EDM standards. All metadata supplied by content providers to the EEXCESS aggregation infrastructure will be converted, without user intervention, to the EEXCESS metadata scheme [similar to Europeana Data Model (EDM)]. Every digital object needs to be published with a rights label that describes its copyright status.

The partner wizard helps to customize the connection from the new database to the EEXCESS framework, i.e. if the new database delivers best results

Fig. 4: Process of creating personalized recommendations.
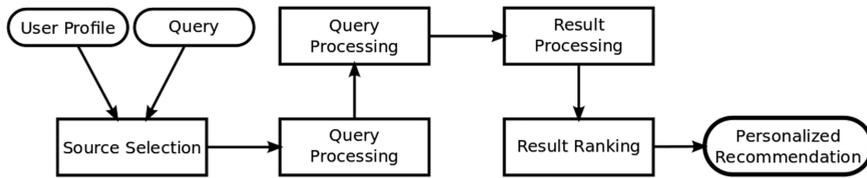
to queries formulated with AND or OR operators, the recommender will adapt this behaviour when querying the new database. New partners can click through result lists generated by prepared queries and thereby optimize the queries, rather than writing program code or metadata mappings themselves. The optimal results will be found automatically with the help of the partner wizard.

Once a content partner has connected to this framework, the results will be listed and merged with the results of the other partners (Figure 4).

4. User tests and Feedback

All prototypes developed during the project were tested by users. This was repeated in different stages of prototype development. As a general outcome, the overall acceptance of EEXCESS content and tools is high, but depends heavily on the relevance of the recommendations and usability of the software components.

The first prototypes were created early in the project. It was quite obvious that they would not be able to meet all user expectations, but user tests were still conducted at early stages in order to get feedback and adapt the application.

During the first part of the development many users said they liked the features and saw the potential of the project. However, the recommendations actually provided – especially by the early prototypes – were not matching the expectations of a number of users due to a couple of (technical) issues. The first recommendation tool did not work well with the different types of

databases responding in many different ways. Also, there were only a small number of content providers hooked up to the central system in the beginning, so that there was only a relatively small pool of recommendations. Over the course of the project the recommendation system was fine-tuned and more content was integrated, so that the matching was improved. A number of experiments were conducted to enhance personalized recommendations (Seifert, Schlötterer, & Granitzer, 2015).

A user test was conducted at a computer lab at the University of Passau with 77 individuals. The Chrome Browser with the EEXCESS extension was used for testing. The experiment was focussed on the following questions:

- How accurate is the extraction/detection of paragraphs on a page?
- Are users satisfied with the automatic query, in general?
- What is a satisfying query?
- Does the personalisation work?
- What is the influence of explanations on perceived quality and trust?

The test persons considered the quality of paragraph extraction/detection to be quite high. However, in general the rating of query results still varies a lot. At the time, there were no clear positive effects of personalization. Language filters are required by users to eliminate results in languages that they do not understand.[9]

Towards the end of the project the major findings of user tests and feedback are:

- The different implementations are accepted as useful;
- Easy to understand, simple components are more easily accepted than more complex components;
- The more data providers are included, the better the results,
- Respondents in several test beds stated that a client technology for Microsoft Word would be desirable.

The quantitative evaluation revealed that both, the number of data providers and client technology have increased significantly since the last evaluation. This is a positive sign implying that the take-up by programming communities starts to show effect.

5. Privacy Protection

A key objective of the project is privacy protection. Therefore, a dedicated work package focusing on privacy issues while providing personalized recommendations at the same time was part of the project.

Private, sensitive information is exchanged at multiple levels like for example during recommendation and usage mining. Individuals reveal information about their usage behaviour through e.g. click rates, technical details of the devices used, profile information with respect to topical interests etc. Wherever data about individuals is involved, this raises privacy issues. On the other hand, relevant recommendations are best achieved if they are tailored to the user needs and preferences.

Hence, privacy protection plays an important role in the research goals of EEXCESS. The project partners developed and implemented methods and protocols that guarantee privacy-preserving augmentation, recommendation, and user mining.

All user profiles and context information are stored on the user's device (instead of a central server), submitting only very little information to the recommender system. Data is only collected with prior and explicit permission by the user and it is anonymized. Furthermore, only information that enables better recommendations is stored. (i.e. no names, no exact address information etc.).

A privacy proxy ensures that users can observe all traces, analyse the activity of the proxy and the type and volume of data exchanged etc.

Research on the theoretical and practical implications of privacy protection was also part of the project (Petit, Ben Mokhtar, Brunie, & Kosch, 2014).

6. Sustainability of the Results

To achieve sustainability of the project, dissemination and exploitation of the project results were a core element and a separate work package. Scientific dissemination mainly happened through conferences in the fields of Computer Science, cultural heritage and libraries. For general dissemination, the project

website as a central hub for all information about the project proved very useful.

Exploitation so far included internal exploitation of the technology within the partner institutions on the one hand, and external exploitation on the other hand. Regarding internal exploitation, EEXCESS partner institution Joanneum Research integrated EEXCESS features into their collection management system *imdas pro*. Also, BitMedia and ZBW have integrated EEXCESS software modules into their systems. The servers that host the EEXCESS recommender will run beyond project lifetime at Joanneum Research.

To achieve external exploitation (exploitation of the tools outside the EEXCESS consortium), EEXCESS partners approach cultural institutions as well as private enterprises, demonstrating the tools and exploring the options of integrating them into existing systems (Figure 5).

The software components developed in EEXCESS are made available on platforms like Chrome Web Store, WordPress Plugin repository and Github. Some additional efforts will be needed after the end of the project to cover later releases of software environments like browsers or WordPress versions. Since a number of different institutions were involved in the development of individual components and more stakeholders are interested in the distribution

Fig. 5: Achieving Sustainability through involvement of different communities.
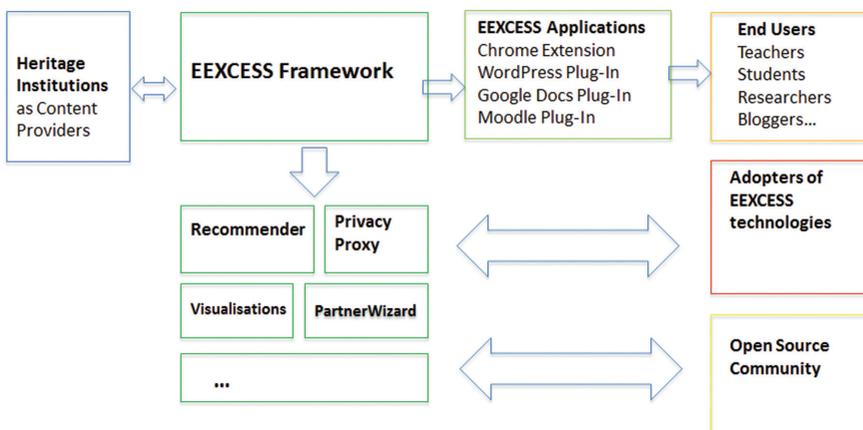
of their content, at least the prototypes most heavily used and needed by the communities will be adapted and maintained.

Other components like the partner wizard for content integration focus on automated workflows which leads to minimized maintenance efforts.

Since all software modules developed in the project are open source, a potentially large community of developers can maintain and enhance the components.

A Hackathon during the project phase was held to get other developers involved in the project. Also, individual project partners worked together with Wikipedians on particular topics. Thus, there is reasonable hope that the institutions, developers and Wikipedians will all contribute to further developments of the project results.

At the German National Library for Economics (ZBW) we also included some of the prototype features into our search portal EconBiz.

## 7. How Libraries and Other Institutions Can Benefit from the Results and Outlook

In addition to the fact that the software developed for the prototypes can be reused by anyone, there are other ways for content providers to get involved. Libraries that have content they would like to share with potential users can disseminate their content through EEXCESS.

Cultural heritage institutions, cultural aggregators, academic libraries and other organizations with valuable digitized content are invited to contribute to EEXCESS.[10]

Components that are useful in many different parts of the project were developed as independent modules, so that they can be easily used and reassembled.

At ZBW, we included two components into our EconBiz Beta Services to make them available for testing. We will evaluate the benefits of including them into the main service.

Fig. 6: uRank Technology as part of the visualization widget.
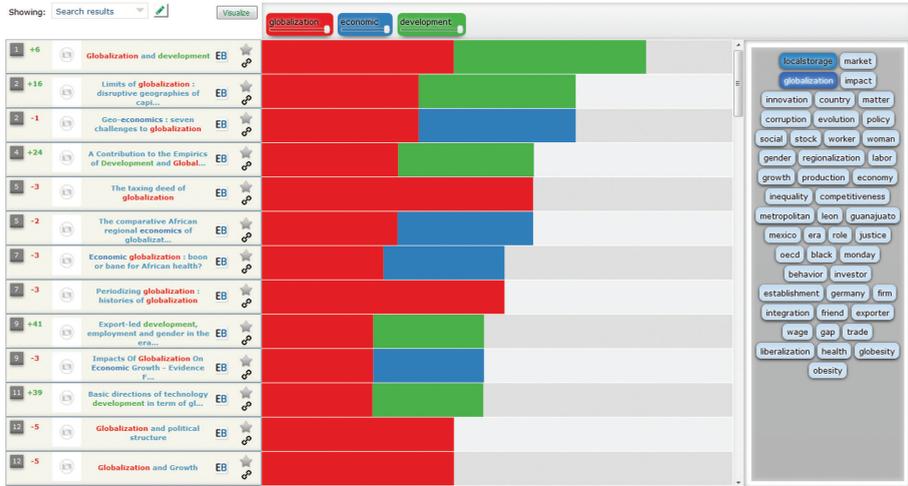
The visualization especially of the uRank technology where search results can be re-ranked may be of interest for a number of libraries and related institutions (Figure 6).

The results change as more keywords are dragged and dropped into a special field. Users can move sliders to give individual keywords more or less weight.

Last but not least, institutions with the abilities and resources to modify the code can easily adapt the prototypes or individual modules to their needs.

Acknowledgement

References

Borst, T., & Witt, N. (2016). *Content recommendation by means of EEXCESS* (Blogpost). http://zbw.eu/labs/de/blog/content-recommendation-by-means-of-eexcess.

Dzeyk, W. (2015). *Die Nutzung von Social-Media-Diensten in der Wissenschaft-Merkmale und Typologie.* Online-Studie 2015 im Auftrag von Goportis-Leibniz-Bibliotheksverbund Forschungsinformation. Retrieved July 19, 2016, from http://www.goportis.de/fileadmin/user_upload/Bericht_Goportis_Nutzung_von_Social-Media-Diensten_in_der_Wissenschaft_Daten_und_Ergebnisse_2015.pdf.

Orgel, T., Höffernig, M., Bailer, W., & Russegger, S. (2015). A metadata model and mapping approach for facilitating access to heterogeneous cultural heritage assets. *International Journal on Digital Libraries*, *15*(2–4), 189–207. http://dx.doi.org/10.1007/s00799-015-0138-2.

Petit, A., Ben Mokhtar, S., Brunie, L., & Kosch, H. (2014). *Towards efficient and accurate privacy preserving web search*. Presented at the Middleware for Next Generation Internet Computing Workshop of the 15th International Middleware Conference. Retrieved July 19, 2016, from. http://eexcess.eu/wp-content/uploads/2013/03/Towards_Efficient_and_Accurate_Privacy_Preserving_WebSearch.pdf.

Rowlands, I., Nicholas, D., Williams, P., Huntington, P., Fieldhouse, M., Gunter, B., … Tenopir, C. (2008). The Google generation: the information behaviour of the researcher of the future. *Aslib Proceedings*, *60*(4), 290–310. http://dx.doi.org/10.1108/00012530810887953.

Schlötterer, J., Seifert, C., Lutz, W., & Granitzer, M. (2015). *From context-aware to context-based: Mobile just-in-time retrieval of cultural heritage objects*. In A. Hanbury, G. Kazai, A. Rauber, & N. Fuhr (Eds.): Advances in Information Retrieval – 37th European Conference on IR Research, ECIR 2015, Proceedings – Lecture Notes in Computer Science 9022 (pp. 805–808). Berlin: Springer. Retrieved July 19, 2016, from http://eexcess.eu/wp-content/uploads/2013/03/Schloetterer2015a_ecirdemo_eexcess-android-app.pdf.

Seifert, C., Schlötterer, J. & Granitzer, M. (2015). *Towards a feature-trich data set for personalized access to long-tail content*. In SAC '15, Proceedings of the 30th Annual ACM Symposium on Applied Computing (PP. 1031–1038). New York: ACM. http://dx.doi.org/10.1145/2695664.2695671.

Veas, E., Mutlu, B., di Sciascio, C., Tschinkel, G., & Sabol, V. (2015). Visual recommendations for scientific and cultural content. In J. Braz, A. Kerren, & L. Linsen (Eds.), *Proceedings of the 6th International Conference on Information Visualization Theory and Applications* (pp. 256–261). Setúbal, Portugal: Scitepress. Retrieved July 19, 2016, from http://www.scitepress.org/DigitalLibrary/PublicationsDetail.aspx?ID=ZVQcsK976XQ=&t=1. http://dx.doi.org/10.5220/0005352802560261.

Notes

---

[1] http://eexcess.eu/.

[2] The concept of the "long tail" describes a distribution where a few mega-sellers or hubs have a lot of usage whereas specialized content has only very few buyers or users. Since there is a very huge number of special items a graph of this distribution shows a "long tail". See e.g. https://en.wikipedia.org/wiki/Long_tail.

[3] Compare EEXCESS Vision: http://eexcess.eu/vision/.

[4] On the recommendation process see Borst and Witt (2016). On context detection of the users and their circumstances see Schlötterer, Seifert, Lutz and Granitzer (2015).

[5] Wikipedia Reference Butler: http://eexcess.eu/wikipedia-reference-butler/.

[6] All prototypes and visualizations can be found at http://eexcess.eu/results/.

[7] Wikipedia was used because a number of surveys have shown that it is very often the most important source for starting research on a new topic even in academic contexts. Also, the structure with individual paragraphs and headlines supports the automatic detection of section topics.

[8] Details on the requirements can be found at http://eexcess.eu/content-providers/.

[9] All project deliverables are made available to the public on the eexcess website. They can be found on the project website from September 2016: http://eexcess.eu/publications/.

[10] Information on how to become a content provider can be found at http://eexcess.eu/content-providers/.