A. Weisbecker, M. Burmester & A. Schmidt (Hrsg.): Mensch und Computer 2015 Workshopband, Stuttgart: Oldenbourg Wissenschaftsverlag, 2015, S. 255-260.

Unified Information Access for Knowledge Workers via a Federated Recommender System

Heimo Gursch, Hermann Ziak, Roman Kern

Knowledge Discovery Group, Know-Center GmbH

Abstract

The objective of the EEXCESS (Enhancing Europe's eXchange in Cultural Educational and Scientific reSources) project is to develop a system that can automatically recommend helpful and novel content to knowledge workers. The EEXCESS system can be integrated into existing software user interfaces as plugins which will extract topics and suggest the relevant material automatically. This recommendation process simplifies the information gathering of knowledge workers. Recommendations can also be triggered manually via web frontends. EEXCESS hides the potentially large number of knowledge sources by semi or fully automatically providing content suggestions. Hence, users only have to be able to in use the EEXCESS system and not all sources individually. For each user, relevant sources can be set or auto-selected individually. EEXCESS offers open interfaces, making it easy to connect additional sources and user program plugins.

1 Introduction

Due to the vast number and variety of online libraries, reference works, and archives established in the last decades, knowledge workers may experience difficulties in finding the appropriate source of required information. Users have to query every source that can potentially contain useful material for their work. Differences in the user interface designs and query syntax complicate the search process even further. Moreover, formulating the query, i.e., finding descriptive keywords, is an inherent search problem that extends to online databases. (ter Hofstede et al. 1996)

EEXCESS (Enhancing Europe's eXchange in Cultural Educational and Scientific reSources) is an EU-funded research project whose objective is to create an assistant that can automatically suggest novel and relevant content to knowledge workers. Such recommendations should be helpful rather than distracting. In that regard, within the

EEXCESS project components have been developed to be included in the user interface of existing software. Currently, EEXCESS recommendations can be integrated as a Chrome Browser, Google Docs, and WordPress plugin. Once activated, the plugin automatically scans the text in focus and presents recommendations generated by the EEXCESS system to the user. Consequently, the user can access content from various knowledge sources without performing a manual query. Each user can receive recommendations from a different subset of all available sources, which can either be specified manually or automatically selected. The unified EEXCESS interface for querying multiple knowledge sources is especially helpful for workers who rarely query knowledge sources and lack the necessary familiarity with the individual source interfaces.

In order to use the EEXCESS system in an enterprise setting, a connection to the internal knowledge databases has to be established. Databases supporting the OpenSearch API (Application Programming Interface) can be linked directly and require no further actions. For other APIs special connectors have to be developed. Similarly, to integrate the recommendations into other applications, an API is provided to establish a communication between customised clients and the EEXCESS system. Development can be performed by everybody since the EEXCESS source code is available via GitHub[1] under an open-source licence.

2 Related Work

Recommender systems are designed to assist users by providing novel, interesting, and useful information. Nowadays, they are common parts of online selling platforms, suggesting products to potential customers. Recommender systems can be divided into four groups depending on the recommendation mechanism. These four groups are called content based, collaborative filtering, knowledge based, and hybrid recommender systems. Content based systems generate recommendations based on the properties of the recommended items. Collaborative filtering systems recommend items if they are frequently chosen by other users. Knowledge based recommenders rely on recommendation rules specified by domain experts. Hybrid recommenders combine two or all three of the above-mentioned approaches. (Ricci et al. 2011) The EEXCESS system itself can be classified as content based.

Content based recommender systems use many algorithms originally developed for search engines, also termed information retrieval systems (Ricci et al. 2011). Information retrieval systems have a history of federated and distributed approaches. While federated information retrieval systems provide a common interface for multiple knowledge sources, aggregated information retrieval systems go one step further and merge individual result lists into a single one. In other words, while federated information retrieval systems provide one result list per knowledge source, aggregated information retrieval systems provide one result list containing search results from all knowledge sources. (Kopliku et al. 2014)

[1] http://github.com/EEXCESS

Avrahami et al. (2006) showed that federated systems, which rely on small distributed nodes, have advantages over a centralised collection that may quickly become outdated. However, simultaneously querying multiple sources and aggregating the result list creates certain challenges. Although in many cases recommender systems can replace search engines (Guy et al. 2010), very few of them follow the distributed approach. The EEXCESS projects aims at helping popularise federated architectures in the domain of recommender systems.

In the context of smart factories information gathering and monitoring becomes an important task of production workers in optimizing the manufacturing processes. Data generated in the production process is usually stored in a number of individual databases an not in a central storage. (Martin et al. 2004). To provide unified and non-distracting access to all involved data stores, user interfaces on mobile devices and augmented reality techniques are adopted for production floors. Federated information retrieval systems can provide a unified access to knowledge sources for user interfaces. (Mooser et al. 2007) Knowledge support is particularly important for decision making, which often requires very specific pieces of information (Zhang & Lu 2007). Since the required material may be obscured by a vast amount of other data that are irrelevant in a particular situation, traditional information retrieval approaches may not be successful (Géczy et al. 2007). Recommender systems that are optimized not only with regard to retrieval precision, but also result diversification and serendipity can help knowledge workers to efficiently and successfully perform their work (Shen et al. 2008).

3 System Architecture

A sketch of the EEXCESS system architecture is shown in Figure 1. Users can choose between the web frontend and client plugins as user interfaces, which differ in terms of how recommendations are triggered. The client plugins automatically infer the current user context and trigger recommendations based on the texts the user works with. The web interface can trigger recommendations based on manually-entered keywords. User privacy is an important aspect of the EEXCESS system's design. To ensure that users cannot be traced, the privacy preservation layer separates the user clients from the recommender components and anonymizes the individual user requests through a variety of strategies. As a result, the users are in control of how much information is disclosed.

The EEXCESS core is the so-called federated recommender component. It is responsible for interpreting the information request, querying one or more sources, and generating a unified result list. Generating recommendations always follows the same procedure. First, sources that should be used have to be identified via the source selection. Depending on the information received from the privacy preservation component, different subsets of all available knowledge sources can be used. After the set of relevant sources has been established, the query reformulation adapts the query specific to each source in order to achieve optimal results. Each source is individually queried via its own source connector. Queries for and replies from the knowledge databases are channelled through the source connectors. They can be viewed as translators between the EEXCESS system and the

knowledge sources. Currently, Brockhaus Wissensservice[2], Deutsche Digitale Bibliothek[3], Deutsche Zentralbibliothek für Wirtschaftswissenschaften (ZBW)[4], Europeana[5], Kooperationsinitiative Museen Baselland (KIM)[6], Mendeley[7], Wikipedia[8], and all knowledge bases with an OpenSearch[9] based interface are supported. Responses from the selected sources are merged into a single result list via the result aggregation mechanism. The final aggregated result list is then sent back to the corresponding user client, again passing through the privacy preservation components.

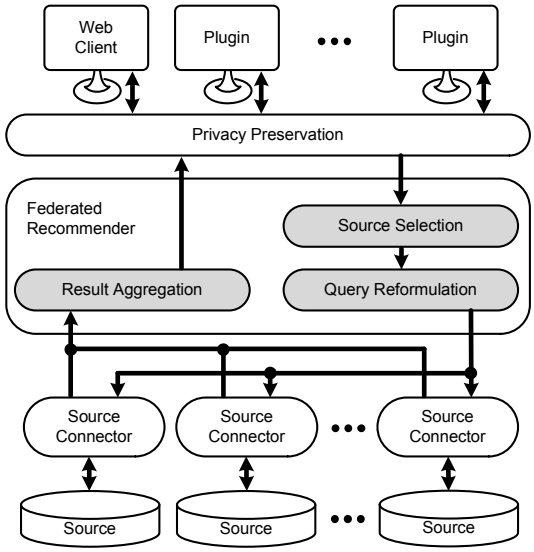

Figure 1: EEXCESS system architecture. User interaction is handled by a web frontend or application plugins. The privacy preservation component prevents the federated recommender from profiling users. The federated recommender is the core of the system and is responsible for generating recommendations based on the material in the knowledge sources.

4 Evaluation and Results

The presented results reflect the current state of the EEXCESS infrastructure. Table 1 provides the response times for various knowledge sources. The average response time of the

2 http://www.brockhaus-wissensservice.com/
3 http://www.deutsche-digitale-bibliothek.de/
4 http://www.zbw.eu/
5 http://www.europeana.eu/
6 http://kgportal.bl.ch/
7 http://www.mendeley.com/
8 http://www.wikipedia.org/
9 http://www.opensearch.org/

EEXCESS system is 576ms. It is significantly affected by the individual sources response times. The federated recommender has to wait for all sources to reply before processing the aggregated results. Source responses are only included in the aggregated result list if the source responds within 2.5s.

Source	Results	Source Response Time	Connector Processing Time
Europeana	82570	404 ms	26 ms
KIM	41	173 ms	19 ms
Mendeley	9	402 ms	14 ms
Brockhaus	52	284 ms	18 ms
ZBW	2334	355 ms	21 ms

Table 1: Source and source connector execution times averaged over 50 queries. The average system response time is 576ms. Depending on the source, about half to three quarters of the EEXCESS system response time relates to the response time of the individual knowledge sources. Note that the average number of results and the average source response time have a Pearson correlation of only 0.47.

When aggregating the result, the EEXCESS system considers various optimization measures, such as diversity, accuracy, novelty, and serendipity. To achieve good user acceptance, recommendations have to be comprehensible and reproducible. Users can become frustrated if recommendations change although the underlying context does not. This needs to be taken into account as an additional design criterion.

One major issue associated with federated content based recommendation systems is the integration of knowledge sources. Although in the past efforts were made to develop a standardized interface (e.g., OpenSearch), the individual sources cannot be integrated easily. Although a lot of querying operations are supported by all sources (e.g., grouping, faceting, and range queries), they tend to behave differently even at the level of Boolean queries. For example, disjunction queries might be interpreted as *one or the other or both* or strictly as *one or the other*. In addition to these conceptual differences, establishing a common data format is also a challenge. The EEXCESS sources use JSON and XML, and the necessary format conversion has a negative impact on the system's performance.

5 Conclusion

EEXCESS can simplify research tasks by suggesting possible content selected from a variety of knowledge sources. However, it has certain restrictions. Users will only accept federated recommenders if they provide quick responses and comprehensible suggestions. These conditions come on top of other restrictions imposed by the sources having different query and result formats. All of these are important considerations with regard to the success of a recommender system simplifying the access to knowledge sources also for production workers not experienced in search. In the future, work in the EEXCESS project is focused on improving the source selection and the query reformulation so that the EEXCESS system is able to better handle the knowledge source peculiarities.

Acknowledgments

The EEXCESS project is funded by the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 600601. The Know-Center is funded within the Austrian COMET Program—Competence Centers for Excellent Technologies—under the auspices of the Austrian Federal Ministry of Transport, Innovation and Technology, the Austrian Federal Ministry of Economy, Family and Youth and by the State of Styria. COMET is managed by the Austrian Research Promotion Agency FFG.

References

Avrahami, T. T., Yau, L., Si, L. & Callan, J. (2006). The FedLemur Project: Federated Search in the Real World. *Journal of the American Society for Information Science and Technology 57*(3), pp. 347-358.

Géczy, P., Izumi, N., Akaho, S. & Hasida, K. (2007). Long Tails and Analysis of Knowledge Worker Intranet Browsing Behavior. In Abramowicz, W. (Ed.): *Business Information Systems.* Berlin, Germany: Springer, pp. 584-597.

Guy, I., Jaimes, A., Agulló, P., Moore, P., Nandy, P., Nastar, C. & Schinzel, H. (2010). Will Recommenders Kill Search? Recommender Systems - an Industry Perspective. In: *Proceedings of the fourth ACM conference on Recommender systems*. Barcelona, Spain: ACM, pp. 7-12.

Shen, J., Geyer, W., Muller, M., Dugan, C., Brownholtz, B. & Millen, D. (2008). Automatically Finding and Recommending Resources to Support Knowledge Workers' Activities. In: *Proceedings of the 13th International Conference on Intelligent User Interfaces. Gran Canaria*, Spain: ACM, pp. 207-216.

Kopliku, A., Pinel-Sauvagnat, K. & Boughanem, M. (2014). Aggregated Search: A New Information Retrieval Paradigm. *ACM Computing Surveys (CSUR). 46*(3), pp. 1-31.

Martin, B., Lamine, J. & Oliver, S. (2004). Smart Factory - Mobile Computing in Production Environments. In: *Proceedings of the Mobisys Workshop on Applications of Mobile Embedded Systems*. Boston, MA, USA: Association for Computing Machinery, pp. 18-20.

Mooser, J., Wang, L., You, S. & Neumann, U. (2007). An Augmented Reality Interface for Mobile Information Retrieval. In: *Proceedings of the International Conference on Multimedia and Expo. Beijing*, China: IEEE, pp. 2226-2229.

Ricci, F., Rokach, L., Shapira, B. & Paul, K.B. (Ed.) (2011). *Recommender Systems Handbook*. New York: Springer.

ter Hofstede, A. H., Proper, H. A. & van der Weide, T. P. (1996). Query Formulation as an Information Retrieval Problem. In: *The Computer Journal. 39*(4), pp. 255-274.

Zhang, N. & Lu, W.-F. (2007). A Framework for Managing Enterprise Knowledge for Collaborative Decision Support. In: *International Conference on Industrial Informatics*. Vienna, Austria: IEEE, pp. 517-522.

Contact Information

Heimo Gursch, Knowledge Discovery Group, Know-Center – Kompetenzzentrum für wissensbasierte Anwendungen und Systeme Forschungs- und Entwicklungs GmbH, Inffeldgasse 13/6, 8010 Graz, Austria, hgursch@know-center.at, http://www.know-center.at