ABSTRACT

Cultural, scientific and educational resources often remain buried in deep web portals like libraries or museum archives in the so called long-tail of the web. The reason for these long-tail contents largely remaining untapped is twofold: One factor is that most users are not aware of the existence of these contents, as they do not appear high-ranked on general purpose search engines, due to their lower popularity. The second problem is that the contents are spread widely over the Web, residing in a multitude of small repositories, which cover specific domains.

We present the current prototype of a browser extension\(^1\), using a federated just-in-time retrieval service for bringing the content to the user, extending previous work [1]. More concretely, to face the user problem and make people aware of the existence of valuable long-tail resources, our prototype notifies the user about available resources from the long-tail, which are deemed relevant to the task at hand. The proactive search for resources is based on the user’s current context, i.e., the current web page or a text selection. Facing the technical problem of distributed resources, the extension queries a federated retrieval system, which aggregates search results from different content providers, by the time of writing content from ZBW\(^2\), Mendeley\(^3\) and Europeana\(^4\). Generally, further content providers can register themselves easily to the service. Europeana itself is already an aggregator for digital museum artifacts, so the content base of potentially retrievable results within our system is even larger. When relevant content is detected for the current user context, the user is unobtrusively notified. Notification is done by changing the icon of the browser extension, showing the number of newly retrieved results. If the user decides to have a look at those results, she gets (beyond a traditional result list) various visual interfaces, which facilitate the exploration of the retrieved results. Figure 1 shows an example of the user interface showing the result list. We think that our approach is promising for making users aware of long-tail contents, for which they can and do not state an information need, because they do not know of their existence.

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.

1. REFERENCES

Figure 1: Screenshot of the Wikipedia page about Arsène Houssaye and search results for the selection "La Couronne de bluets". The terms in the query box are generated out of the selected text, but may be adapted by the user.