

A Twitter Bot: Recommending long-tail content

Sebastian Bayerl and Michael Granitzer

University of Passau, Germany
{forename.surname}@uni-passau.de

Abstract Sources like the German National Library of Economics (ZBW)¹ provide high quality content, but potentially interested user groups are often unaware of these long-tail data. Twitter is an open platform and provides access to its 284 million monthly active users. The purpose of the Twitter Bot is to distribute the previously mentioned content to interested users and thereby connecting the unaware community with the service. A recommender, developed within the EEXCESS project², is used in the ZBW use-case to generate recommendations for a specific tweet. Identifying the relevant subsets of users is hereby a major problem.

The research questions can be formulated as following:

- Is Twitter a viable medium to connect specialized memory institutions like libraries with their unknown world wide distributed user base?
- Respecting API limits and terms of service, how can interested users be identified and approached?

The not trivial problem about implementing the bot is maintaining its functionality. Twitter restricts the number of API calls and expects personal and varying content. Violating these terms leads to a suspension of the account. By tweeting a lot of similar structured content, e.g. URLs with descriptions on a regular schedule, the bot could be considered a spam bot and be suspended. Therefore, a high quality of service with varying content must be maintained.

Tweets can be pre-filtered using a keyword map, e.g. the STW Thesaurus for Economics provided by the ZBW. Recommending resources only to people, which are actually posting within the topic increases the chance that they are interested in the service. Twitter analytics dashboard shows the the keyword-based approach is suitable to identify suitable user-groups. These users are now entry points to the network, because they probably follow and are followed by users interested in the same topics. Hereby, a network of use-case interested persons is build up. Retweeting the tweets of identified experts can enhance the quality of the stream and brings more variety to it.

This work will show if and how long-tail data can be recommended to domain-experts within the Twitter environment in the long-term.

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 <http://www.zbw.eu/>

² <http://eexcess.eu/>