3bij3: A framework for testing effects of recommender systems on news exposure

Löcherbach, F.; Trilling, D.

DOI
10.1109/eScience.2018.00093

Publication date
2018

Document Version
Final published version

Published in
IEEE 14th International Conference on eScience

Citation for published version (APA):
3bij3: A framework for testing effects of recommender systems on news exposure

Felicia Löcherbach
Department of Communication Science
Vrije Universiteit Amsterdam
Amsterdam, Netherlands
f.loecherbach@vu.nl

Damian Trilling
Department of Communication Science
University of Amsterdam
Amsterdam, Netherlands
d.c.trilling@uva.nl

Abstract—We developed 3bij3, a framework that presents a news app to participants, with contents that are displayed based on different recommendation logics. It tracks usage over time and enables large-scale field experiments.

Index Terms—recommender systems, news exposure, communication science, web application

I. INTRODUCTION

The impact of recommender systems on how we consume and perceive news is still understudied. To shed more light on this, we developed one of the first software solutions for such studies, and present 3bij3. 3bij3 means three by three in Dutch, and signifies the most prominent feature of a news application that forms the interface for study participants. It displays a 3x3 grid of the nine most relevant news articles to a user, selected by different mechanisms (e.g., explicit customization vs based on past behaviour vs random). 3bij3 can be used to conduct large-scale field experiments, in which participants’ use of the site can be tracked over extended periods of time.

II. RELATED RESEARCH

Social-scientific experiments [1]–[4], usually based on the creation of a set of fictitious news articles regarding a specific topic that are varied on factors such as political stance, often create a highly artificial situation. They thus often lack external and ecological validity. Input-output analyses [5], [6] aim at uncovering the workings and effects of existing recommendation algorithms – however, they have to treat the actual algorithm as a black box. Lastly, simulations [7]–[10] evaluate recommendations based on existing or simulated datasets using various metrics. While being a good approach for judging the performance of algorithms regarding certain predefined measures, such studies remain in an “evaluation setting where recommendation approaches are compared without user interaction” [11, p.9].

III. 3BIJ3

Compared to previous work, 3bij3 gives researchers control over the recommendation system under study and creates a realistic environment for the participants. Figure 1 gives an overview of the components of 3bij3: (1) Content retrieval, processing, and enrichment; (2) article selection; (3) user interaction. It integrates web scraping/parsing, supervised machine learning for labeling articles, different recommender systems, a web interface for participants, and gamification elements (earning points for interaction).

3bij3 is written in Python and uses Flask for the web frontend. When signing up, participants are randomly assigned to one of the article selection conditions. Articles are stored in ElasticSearch, user data (incl. clickstream) in MySQL.

IV. LIMITATIONS AND FUTURE WORK

3bij3 has been tested on a first group of participants ($N = 25$). We plan to use it in a large-scale study ($N > 500$) in order to test participants’ reactions towards different selection logics.
REFERENCES