

# A Review of User-centred Information Retrieval Tasks

Ali Hosseinzadeh vahid  
ali.h.vahid2010@gmail.com

Roghayeh Gachpaz hamed  
ramisa.hamed@adaptcentre.ie

Kevin Koidl  
kevin.koidl@scss.tcd.ie

ADAPT Centre  
Trinity College Dublin  
Dubin- Ireland

## ABSTRACT

Although, most of the recent studies within the IR domain tend to target how users behave while addressing their information needs. However, even though the collection of document sets and user profiling is a top research problem, it holds inherent difficulties for the establishment of a comparative task to evaluate various approaches. Also finding a comprehensive metrics to evaluate different affects of various aspects, that play a significant role in satisfaction of users of personalized IR systems, seems to be another noticeable issue. With the review of related tasks in well-known IR evaluation communities, this paper will discuss the way of gathering users profiles and objects of their interest in last 5 years of IR evaluation campaigns.

## CCS Concepts

•Information systems → Personalization; Information retrieval; Retrieval tasks and goals;

## Keywords

Information Retrieval, User-centred evaluation tasks

## 1. INTRODUCTION

Evaluation tasks are well-known and innovative ways of providing the infrastructure necessary for stimulating, demonstrating and evaluating substantial improvements of information retrieval methodologies. “Future information retrieval systems must anticipate user needs and respond with information appropriate to the current context.”[1] These systems are challenged in three stages: (1) How to model the information about the user, task, and context; (2) How to find and acquire “objects of interest” and (3) how to exploit this information in order to retrieve the most relevant results, which satisfy the users information needs. In recent years different IR evaluation campaigns have focused on the development of a variety of task for the exchange of research ideas on user-centred IR systems. This paper

investigates and compares the objectives, approaches and impacts of such tasks with the hope to find more appropriate approaches to evaluate improved user-dependent and context-aware IR systems.

## 2. USER-CENTRED IR TASKS

Considering the daily-growing tendency to user-centred IR systems, evaluation campaigns are promoted to the exploration of new evaluation methodologies for such systems. Nevertheless, due to the complexity of the evaluation of personalized IR systems and the involved potential costs of evaluation, some IR evaluation forums such as NII Testbeds and Community for Information access Research (NTCIR)<sup>1</sup> and Forum for Information Retrieval Evaluation (FIRE)<sup>2</sup> have not focused on user-centred tasks yet. So, this section describes a user-oriented and context-based task approach which has been provided in other IR evaluation communities.

### 2.1 Contextual Suggestion Task of TREC

Starting in 1992, the Text REtrieval Conference (TREC)<sup>3</sup>, co-sponsored by the National Institute of Standards and Technology (NIST) and U.S. Department of Defense, is the most popular IR evaluation campaign which has also provided the first large-scale evaluations of cross-lingual and multilingual document retrieval tasks. TREC has also introduced evaluations for open-domain question answering, content-based retrieval of digital video and retrieval of recordings of speech.

The goal of the contextual suggestion track [3] is to evaluate the search techniques for complex information needs of users with respect to context and their point of interest. Introduced in TREC 2012, this track investigates to develop a system that is able to make suggestions of sites with the goal to explore an unknown city based upon the users personal interests in the users home city. A set of user preferences, example suggestions and a set of contexts are given to participants as inputs: Constant number of manually gathered suggestions consist of a title, a short description and a website URL of different attractions within specific, predefined regions have been recommended to a user as something they find interesting. Profiles are built by conducting a survey advertised to crowdsourcing workers to indicate their preferences to the set of above mentioned example suggestions.

<sup>1</sup><http://research.nii.ac.jp/ntcir/index-en.html>

<sup>2</sup><http://fire.irsi.res.in/>

<sup>3</sup><http://trec.nist.gov/>

These assessors asked to give two ratings for each attraction: 1) How interesting the suggested attraction seemed to them based on its description and 2) based on its website, respectively. Contexts describe which city a user is currently located in. There were 50 cities chosen randomly from the list of primary cities in metropolitan areas in the United States from Wikipedia. Each submitted run consists of up to 50 ranked suggestions for each profile-context pair, with formatting similar to that of the sample suggestions. Participants have been able to gather suggestions from either the open web, ClueWeb12<sup>4</sup>, or fixed set of documents. Precision at Rank 5 (P@5), Mean Reciprocal Rank (MRR) and a modified version of Time-Biased Gain (TBG)[5] are used to rank runs of participants.

## 2.2 Social Book Search Task of CLEF

Conference and Labs of the Evaluation Forum, formerly known as Cross-Language Evaluation Forum (CLEF)<sup>5</sup> promotes development of information access systems with an emphasis on multilingual and multimodal information with various levels of structure.

The Social Book Search [4] investigates Evaluation methodologies for book search task using a combination of various aspects of retrieval and recommendations dealing with professional and user-generated meta-data.

As a continuation of the INEX SBS Track that ran from 2011 up to 2014, the task is targeted to returning a list of recommended books in reply to a user request posted on a LibraryThing<sup>6</sup>(LT) discussion forums by matching the user's information need. A set of book requests and a set of user profiles have been assumed as inputs of the task and a submitted ranked list of recommended books has been evaluated as the result of participant's system.

The test collection [2] consists of 2.8 million book records from Amazon, extended with social metadata from LT. Each book record is an XML file with fields like *isbn*, *title*, *author*, *publisher*, *dimensions*, *numberofpages* and *publicationdate*. The social metadata from Amazon and LT is stored in the *tag*, *rating*, and *review* fields. To improve the quality of the meta-data, they are extended with library catalogue records from the Library of Congress (LoC) and the British Library (BL).

The topic set is focused on requests which are provided as a narrative description of the information need of a user and one or more example books to guide the suggestions. Users typically describe what they are looking for, give examples of what they like and do not like, indicate which books they already know and ask other members for recommendations. There are also annotated fields by crowdsourcing workers to indicate whether the example book had been read by requester and to judge his/her attitude about the book.

The books suggested by members, which are directly linked to their corresponding records on Amazon, have been used as initial relevance judgements for evaluation of participated systems in the Suggestion Track.

The rich user profiles of the topic creators and other LT users have been used as valuable resources of User profiles and personal catalogues. These profiles generally contain information on which books they have in their personal catalogue on LT, which ratings and tags they assigned to them

<sup>4</sup><http://lemurproject.org/clueweb12/>

<sup>5</sup><http://www.clef-initiative.eu/web/clef-initiative/home>

<sup>6</sup><https://www.librarything.com/>

and a social network of friendship relations, interesting library relations and group memberships.

The official evaluation measure for this task is nDCG@10. It takes graded relevance values into account and is designed for evaluation based on the top retrieved results. In addition, P@10, MAP and MRR scores will also be reported, with the evaluation results.

## 3. CONCLUSION

Although defining tasks and scenarios for evaluation purposes in IR domain is one of the most common ways for the exploration of new methodologies and innovative ways in using and discussing experimental data, it has to be noted that the comparison of approaches, the exchange of ideas and transfer of knowledge has been considered a valuable contribution to evaluation tasks during last decades. However, the tendency of modern user-centred IR system for taking user preferences and interests into account through information seeking process also has changed the identification, setting and evaluation of shared tasks. This paper reviewed and compared existing user tasks and described their way of collecting task resources, methods and metrics with hope to help improving them with combining/ summarising them to propose new user centered tasks in future.

## 4. ACKNOWLEDGMENTS

"The ADAPT Centre for Digital Content Technology is funded under the SFI Research Centres Programme (Grant 13/RC/2106) and is co-funded under the European Regional Development Fund."

## 5. REFERENCES

- [1] J. Allan, B. Croft, A. Moffat, and M. Sanderson. Frontiers, challenges, and opportunities for information retrieval: Report from swirl 2012 the second strategic workshop on information retrieval in lorne. In *ACM SIGIR Forum*, volume 46, pages 2–32. ACM, 2012.
- [2] T. Beckers, N. Fuhr, N. Pharo, R. Nordlie, and K. N. Fachry. Overview and results of the inex 2009 interactive track. In *Research and Advanced Technology for Digital Libraries*, pages 409–412. Springer, 2010.
- [3] A. Dean-Hall, C. L. Clarke, J. Kamps, P. Thomas, N. Simone, and E. Voorhees. Overview of the trec 2013 contextual suggestion track. Technical report, DTIC Document, 2013.
- [4] M. Koolen, T. Bogers, M. Gäde, M. Hall, H. Huurdeman, J. Kamps, M. Skov, E. Toms, and D. Walsh. Overview of the clef 2015 social book search lab. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction*, pages 545–564. Springer, 2015.
- [5] M. D. Smucker and C. L. Clarke. Time-based calibration of effectiveness measures. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, pages 95–104. ACM, 2012.