

Please cite the Published Version

Johnson, Frances (2019) Evaluating personalised information retrieval: a perception of trust. In: Second Annual Workshop on the Evaluation of Personalisation in Information Retrieval (WEPIR 2019), 09 March 2019 - 14 March 2019, Glasgow, Scotland.

Publisher: ACM

Version: Published Version

Downloaded from: <https://e-space.mmu.ac.uk/623075/>

Enquiries:

If you have questions about this document, contact openresearch@mmu.ac.uk. Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from <https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines>)

Evaluating personalised information retrieval: a perception of trust

Frances Johnson

iSchool, Department of Languages, Information & Communication
Manchester Metropolitan University
Manchester, UK
f.johnson@mmu.ac.uk

ABSTRACT

This position paper sets out a case for the evaluation of personalised information retrieval from the user perspective, and which focuses on the effect of personalisation on the searcher's perception of a better information retrieval interaction. The proposed approach, complementing system performance evaluation, is about real users, real tasks and their interactions of query, search and assessment of the retrieved results, as well as their intentions driving the interaction. The components of user evaluation are discussed, and the searcher's perception of usefulness is identified to specifically evaluate the effect of personalisation. The proposition is that a perception of 'support to do a task' is formed during use and which constitutes a major part of the user's evaluation or judgment of trust. As such, it is suggested that, a perception of trust and its criteria provide a framework for evaluation. Further consideration is given to the construct of trust as an evaluative instrument with the goal of advancing personalisation in information interactions.

CCS CONCEPTS

- Users and interactive retrieval

KEYWORDS

Evaluation, User behaviour, User perception, Relevance judgments, Trust, Personalised information retrieval

1 Personalisation and its Evaluation

Ghorab et al's [4] survey identifies three types of Personalised Information Retrieval (PIR), individualised, community based and aggregate level systems. The approach for PIR, in general, is to take and use information about users to adapt the query, and/or modify the search (for example, the document ranking algorithm), and/or tailor the presentation of the results returned. Application of any one approach, or some synergy, aims to improve IR performance

in returning relevant results in response to the query. Digging deeper, and of particular interest with respect to the design of evaluation methods for research and development is that, not only is there expected to be an improvement in the relevancy (to the query) of the results returned, but that also they *appear* to be more relevant to the searcher. Specifically, and in the context of information seeking behaviour, the system and the retrieved results appear as more relevant to (or subjectively, we could say that they appear to better satisfy) the user's information need, this being the context and driver for the query submitted. Ultimately, it follows that a holistic goal of PIR is to appear to the user to address the searcher's perceived information need and retrieve documents that are relevant with regards to the searcher's interpretation of the task, referred to as cognitive and situational types or manifestations of relevance by Saracevic back in 1996 [8].

This technology has emerged from decades of research and in 2012 Steichen et al [9] stated that users can expect the next generation of web information retrieval systems to adapt to individual user properties and contexts of use. Current search engines, both commercial and in development will already have some degree of personalisation based on acquired implicit user data. In the last two years, there has been exponential growth in the availability and use of personal data, giving an increased interest in user modelling and personalisation, as well as raising questions on how best data may be used and bringing evaluation into focus. Evaluation of PIR, therefore, from a user perspective, must enable exploration of questions such as, a) how much do we expect, or assume, personalisation in the systems we use; and, b) how do we perceive its influence on the technology we use to help us to find information. As such user evaluation is called for to complement performance or system evaluation that use a data driven approach with test collections (for example, as described in Dou [2]). Crucially, evaluation from the user perspective may allow us to gauge improved search results, and attempt to assess the extent to which, and the way in which, a user's search experience is enhanced. Specifically, this provides the basis for this position paper focusing on PIR that its evaluation should concern the ways in which the user perceives the search and results to be (in their view) better. In PIR, the system may have adapted the query or adapted the search to improve the retrieved results in the user context. For the purposes of evaluation, it maybe suggested that, as a result, the user perceives the retrieval engine to be easier to

query. For example, we might speculate that the searcher forms an assessment that less effort seemed to be necessary to obtain relevant results. Furthermore, if the results presentation (e.g., their ranking) is personalised, the user could perceive this as better supporting them in their subtask of assessing relevancy and/or in recognising results as ‘situationally’ relevant to the information need.

2 Evaluation from the user perspective

Traditionally, assessment of usability, an evaluation from user perspective, and in specific task contexts, draws on quality components identified by Ben Schneiderman several decades ago. These are described in Preece et al.[7] broadly as follows: a) effectiveness, how good is the product at doing what is it supposed to do; b) efficiency, concerning the way it supports the user in carrying out their task; c) learnability/memorability, how easy it is to learn to use; d) rate of errors, concerning how the user is protected from undesirable situations; and, e) replacing user satisfaction, utility, concerning functionality so that the user can do what they need or want to do.

In this way, usability is evaluated with respect to what the user is doing when using an IR system and for the evaluation of personalisation, Belkin et al [1] bring close attention to the searchers’ series of interactions involved in information retrieval, and moreover that each is motivated by an information seeking intention, such as to learn what is in a database. They suggest that, as such, evaluation must focus on how useful the system has been in supporting these sequences of information seeking intentions. On this, they recommend approaching evaluation from the user assessment of how well the system supported them in their interactions and intentions. This then makes a case for usefulness as the criterion for user evaluation of PIR and implemented during the interactive session by asking the searcher to assess the usefulness of the system (function, features or some quality) in supporting what they want or need to do. This is one approach, and a different approach to evaluation would be to deploy a questionnaire, such as user satisfaction or engagement scale based on an assumption that a desirable aspect or goal of the system and its interface is the user experience, how using it makes you feel. Although it may be reasonably argued that the goal of interactive information retrieval with personalisation is not focused on the user experience, and its hedonistic metrics, a desirable effect of personalisation could nevertheless relate to how the searcher perceives the system to be helpful and supportive during use. A subjective perception formed by the user might be that, or of, the results as relevant to their particular information need or situational context. A perception may be formed of the system effectiveness and ease of use impacted on by the signals of personalisation in the query, search and results presentation or composition. Therefore, rather than asking how useful was the system in supporting the searcher and in helping them to find information, we are interested in how the searcher perceives the system to be useful and which may be impacted on by personalisation.

3 Proposed perspective of trust

The suggestion put forward in this paper is that the searcher would perceive effective (or the effect of) personalisation as supportive in their information seeking intentions and motivating their interactions. The searcher’s task in looking for information has subtasks requiring him or her to formulate and submit a query, the system to perform a search, present the results back and for the searcher to assess these for relevance and in doing so is likely to become involved in other uses of the information, such as learning. Therefore, the effect of personalisation could be that the searcher perceives their query to be better understood by the engine and that its search is more effective, and in turn that the results seem to have greater cognitive or situational relevance. This follows the case made in a previous position paper [1] that evaluation focuses on how useful the system has been in supporting the searchers’ intentions in search interaction. In this position paper, it is suggested that the evaluation is based on how the searcher perceives the support in helping them do what they need or want to do. As we wish to evaluate the effect of personalisation which aims to appear to understand and interpret the searcher’s query and return results that are relevant for the searchers particular context, it seems appropriate to obtain the searcher’s sense of personalisation and feeling of being supported in their intentions.

To put this into an evaluative framework, we may consider the user evaluation of perceived support to be one criterion in the formation of a trust judgment in the system to help find information (and the intentions involved in the session). In the previous section the validity of hedonistic measures of the system goal, for example rating enjoyment, was questioned and so it may seem that an assessment of trust as a metric may be a step too far. Yet recent research and development of ‘trust’ as evaluative construct and framework suggests that this proposition could be worth further investigation.

We typically define trust in terms of our reliance on the ability and integrity of a person or thing, usually to do something for us and our assessment of trustworthiness may be considered an aspect of credibility assessment. Recent and ongoing interest in trust as a construct is evident in various research studies [such as 3, 5, 6]. These aim to understand the judgment made on its core dimensions of ability (or competence), reliability and integrity (or some benevolent intent), as well as the factors that influence its formation (when interacting with the potential ‘trustee’) and, furthermore how this assessment might vary in user and or task contexts. Based on the premise that we evaluate, or at least form an assessment of, the systems we use to help us find information, the position made is that the searcher’s evaluation of the interaction (and the effect of personalisation) has the components of the criteria used in forming a judgment of trust. In the trust framework the searcher may assess the competence and reliability of the engine, in so far as it appears to be working and doing the task expected, and additionally assess its usefulness in helping the searcher in their task. From the perspective of the user and in a search context, this third criterion, of usefulness, may be realized through some assessment of the information provider’s (or retrieval engine’s) intent, as might be made in a trust judgment. In information seeking context we would

expect the information provider to be intent in helping the information seeker in their quest, and that this intention may be signaled through some of the effects of personalisation. It is proposed then that an important effect of personalisation would be that the searcher perceives the interaction and their intention as supported, influencing their judgment of trust formed by an assessment of the retrieval system's effectiveness (or perceived as competent and reliable) and assessment of usefulness (a perception of the engine's intent as helpful and supportive). The evaluation based on trust thus focuses on the perception of usefulness in supporting searcher intentions and interactions through an assessment of the provider's or the engine's intent in working in the searcher's best interests in helping them to accomplish their task.

The case made is for the evaluation of personalisation based on a perception of trust. This, in particular, presents a possible means to evaluate the somewhat intangible impact of personalisation, whereby the search appears to be better for that user. If deemed to be a promising approach, further research is needed for the IR community to build trust evaluation tools for their use across development of PIR technology and interfaces. For example, user studies to elicit how the searcher evaluates their search experience and the criteria used in assessing trustworthiness could be used to develop a validated trust scale for use in A/B testing or comparative evaluation studies. Ultimately, evaluating the goal of PIR to be that whereby interactions and intentions are perceived as trusted.

REFERENCES

- [1] NJ Belkin, D Hienert, P Mayr and C Shah, (2018) Data Requirements for Evaluation of Personalisation of Information Retrieval – A Position Paper. Proceedings for CLEF 2018, Laboratory Evaluation of personalized information retrieval, Avignon, France. <https://arxiv.org/ftp/arxiv/papers/1809/1809.02412.pdf>
- [2] Z Dou, R Song, and JR Wen, (2007) A large-scale evaluation and analysis of personalized search strategies. In WWW '07: Proceedings of the 16th international conference on World Wide Web, 581–590, New York, NY, USA., ACM.
- [3] BJ Fogg, C Soohoo, DR Danielson, L Marable, J Stanford and ER Tauber, (2003). How do users evaluate the credibility of web sites? A study with over 2,500 participants. Proceedings of DUX2003, Designing for User Experiences Conference.
- [4] MR Ghorab, D Zhou, A O'connor and V Wade, (2013). Personalised information retrieval: survey and classification. *User Modeling and User-Adapted Interaction*, 23(4), 381-443.
- [5] F Johnson, J Rowley and L Scaffi, (2014) Modelling trust formation in health information contexts. *Journal of Information Science*, 41(4), 415-429
- [6] T Lucassen and JM Schraagen, (2013). Topic familiarity and information skills in online credibility evaluation. *Journal of the American Society for Information Science and Technology*, 64, 254-264.
- [7] J Preece, Y Rogers and H Sharp, 2015 *Interaction Design: Beyond Human-Computer Interaction*. (4th ed): John Wiley & Sons Chichester
- [8] T Saracevic (2007) Relevance A review of the literature and a framework for thinking on the notion in information science. Part II: Nature and manifestations of relevance, *Journal of the American Society for Information Science and Technology*, 58(13), 1915-1933.
- [9] B Steichen, H Ashman and V Wade, (2012) A comparative survey of personalized information retrieval and adaptive hypermedia techniques, *Information Processing & Management*, 48(4), 698-724.