



2nd International Conference on Computer Science and Computational Intelligence 2017, ICCSCI  
2017, 13-14 October 2017, Bali, Indonesia

## Interest Indicators in Structured Scientific Articles

Saadia Malik<sup>a,b</sup>, Muhammad Saleem<sup>a,b,\*</sup>

<sup>a</sup>King Abdulaziz University, Rabigh Campus, KSA

<sup>b</sup>University of Duisburg - Essen, 47057 Duisburg, Germany

---

### Abstract

Interaction with the systems has been focus of study since many years to devise the interest of user. This information can be used to implement implicit relevance feedback and to provide personalized suggestion in recommender systems. The aim of this article is to investigate the interaction patterns that can be predictors of users' interest in the context of structured documents. The investigated predictors include time spent on a page, clicks to navigate within the document, query and result presentation overlap. Descriptive statistical and machine learning techniques are used to find the relationship between searcher's interest and their explicit given feedback. The results indicate that reading time is predictive of user interest at document level.

© 2017 The Authors. Published by Elsevier B.V.

Peer-review under responsibility of the scientific committee of the 2nd International Conference on Computer Science and Computational Intelligence 2017.

*Keywords:* Interest predictors, relevance indicators, structured documents, INEX iTrack 2005

---

### 1. Introduction

Relevance feedback is a very effective retrieval technique and its main goal is to generate a query that is as close as possible to the searcher's information need. The most common and obvious method for applying relevance feedback is to ask for the explicit rating of the retrieved items, where users tell the system what they think about some object or piece of information. However, forcing the user to decide about the relevance can alter the normal pattern of reading and browsing<sup>4</sup>. The users may stop providing the ratings if they perceive that there is no benefit from it<sup>5</sup>. Hence the user continues to read the information and provides no relevance at all. With the GroupLens system<sup>15</sup>, it was found that users were reading much more information than they were rating. There might be a significant difference between

---

\* Corresponding author. Tel.: +966-58-112 1173.

E-mail address: [msaleim1@kau.edu.sa](mailto:msaleim1@kau.edu.sa)

a user's real interest level and the user's explicit rating since users sometimes have difficulties expressing their interest explicitly on a single numeric scale<sup>16</sup>.

Hence explicit ratings may not be as reliable and especially not as complete as is often presumed. Systems can rely on other sources for getting the relevant/irrelevant information. The possible alternative is to obtain the rating unobtrusively by examining the searchers' interaction with the system and estimating the level of interest based on this data. Though these estimates are not as accurate as the explicit rating, the underlying data can be captured for free, and the combination with the explicit ratings can help in finding out the implicit interest indicators. The need for methods that can estimate the interests has been identified by<sup>12, 14</sup>.

*We believe an ideal solution is to improve the user interface to acquire implicit ratings by watching user behavior. Implicit ratings include measures of interest such as whether the user reads an article and, if so, how much time the user spent reading it<sup>14</sup>.*

*More tools that allow for the accurate and reliable collection of data, such as the browser developed by Claypool, et al. need to be developed, tested and shared, and further research should be done into how the collection process can encourage implicit feedback to closely match the user's underlying intent<sup>12</sup>.*

Another possible way to obtain explicit accounts of why information was assessed at a certain relevance level is through the use of more sophisticated equipment and experimental techniques. For example, it is possible to use eye tracking equipment to monitor the users' eye movements while reading the contents. By analyzing fixation periods and saccades, it is possible to make inferences about the users' perception of importance of the various information. Granka et al.<sup>6</sup> and Guo et al.<sup>7</sup> investigated how users interact with the result page of a WWW search engine using eye-tracking.

A substantial research effort has been put into XML retrieval, with the Initiative for the Evaluation of XML Retrieval (INEX) as the main driving force. Noteworthy advances have been made in the investigation of the possible benefits of document structure in Information Retrieval (IR) (see, e.g.,<sup>8</sup>). This study is performed when searchers performed their tasks by using the XML based retrieval system. Several possible interest indicators are investigated. These include time spent on reading, navigating, and browsing patterns within the document. The selection of result list items to view the details could be due to overlap with the query terms. This overlap can also be considered as a possible evidence.

This paper is structured as follows: Section 2 presents a brief review of related work. Research questions are listed in section 3. Section 4 describes the experimental settings. The implicit indicators studied and discussion on them is given in section 5. In section 6, interest indicators as relevance predictors are evaluated. The conclusion of the study is summarized in the last section.

## 2. Related Work

There are number of behaviors that have been described in the literature as potential relevance feedback indicators. The relevance can be inferred from these observable behaviors to perform implicit relevance feedback retrieval. These techniques obtain the implicit relevance information by watching the users' natural interaction with the system. Such measures are generally thought to be less accurate than explicit measures<sup>17</sup> but, as large quantities of implicit data can be gathered at no extra cost for the user, they are an attractive alternative. Oard et al<sup>18</sup> studied how implicit indicators can be used in place of explicit ratings for a recommender system. The study focused on three broad categories of retention, examination and reference as useful criteria for making predictions. They also found that reading time is a good indicator for measuring relevance. This assertion was affirmed by<sup>11, 12, 23</sup>.

Behaviors such as time spent on reading<sup>16, 15</sup>, dwelling time<sup>1, 2, 9</sup>, mouse activity<sup>5, 22</sup>, scrolling behavior<sup>4</sup>, items bookmarked<sup>19</sup> and interactions with a document<sup>14</sup> have been examined as implicit measures of user interest. These behaviors can be used to indicate interest for a variety of systems such as recommender systems, information filtering systems, etc.

Reading time has been found to be a good indicator of interest for news reading<sup>15, 16</sup> and web browsing<sup>4, 19</sup> but contradictory results have been found for IR tasks<sup>12, 20</sup>. Browser approach to data (key, mouse and dwell time) collection provides an efficient alternative for accurately predicting relevant web documents<sup>8</sup>. Joachim's study<sup>10</sup> focused on the use of clickthrough as an implicit indicator for measuring relevance. He argues that clickthrough by itself is not a significant indicator to measure user interest.

Download English Version:

<https://daneshyari.com/en/article/4960439>

Download Persian Version:

<https://daneshyari.com/article/4960439>

[Daneshyari.com](https://daneshyari.com)