Colouring the Dimensions of Relevance

Ulises Cerviño Beresi[1]*, Yunhyong Kim[1], Mark Baillie[2], Ian Ruthven[2], and Dawei Song[1]

[1] The Robert Gordon University, School of Computing
[2] The Strathclyde University, Department of Computer and Information Sciences

Abstract. In this article we introduce a visualisation technique for analysing relevance and interaction data. It allows the researcher to quickly detect emerging patterns in both interactions and relevance criteria usage. The concept of "relevance criteria profile", which provides a global view of user behaviour in judging the relevance of the retrieved information, is developed. We discuss by example, using data from a live search user study, how these tools support the data analysis.

1 Introduction

In this paper we examine the multi-dimensionality of relevance judgment processes (cf. Borlund's proposed evaluation method[3]): we use Barry and Schamber's relevance criteria classes, Section 2, to encode verbal data gathered from users in a search task, and, define relevance criteria profiles and a session visualisation method (Section 3) which we use to analyse how relevance criteria are used to judge document relevance. We conclude the paper with some final remarks and recommendations for future work (Section 4).

2 Relevance Criteria in (I)IR

Researchers suggest that *"a finite range of [relevance] criteria exists and that these criteria are applied consistently across types of information users, [...]"*[2], defined as an overlap of taxonomies identified within two studies [1, 4]. We adopt the taxonomy and extend it by re-introducing, from [1], three forms of information novelty, users's background knowledge and their ability to understand the information. As a sample, we only list the relevance criteria related to the examples discussed in this paper:

- Depth/Scope/Specificity: related to the range of focus and detail, e.g. how specific it is to the user's needs.
- Currency: whether the information is current or up to date.
- Tangibility: related to tangibility of content, and the inclusion of hard data/facts.
- Affectiveness: related to affective or emotional response to the information aroused in the user.
- Ability to Understand: user's judgement regarding his/her ability to understand the information presented.
- Document novelty: the extent to which the document itself is new to the user.

* This research is funded in part by EPSRC Autoadapt project, grant no: EP/F035705/1.

3 Relevance Criteria Profiles and Session Visualisation

Relevance Criteria Profiles The user data were quantified to produce a user *relevance criteria profile* (RCP). RCPs are defined as a set of counts where each count corresponds to one of the relevance criteria. Counts are defined as the number of utterances made by a user or a group of users that are classified as the corresponding criterion. To *normalise* RCPs, we divide the count of the ith relevance criterion by the total number of user utterances classified as one of the relevance criteria in the encoding schema. That is, the normalised value of relevance criterion i is defined as $rc'_i = rc_i / \sum_{j=0}^{N} rc_j$, where rc_i is the count from the basic RCP, for relevance criterion i, and N is the total number of relevance criteria. Normalising makes RCPs comparable.

Session Visualisation The visualisation proposed in this section is intended to provide a bird's eye view of the relevance judgement process within each search session. The session is visualised as a sequence of *relevance criteria piles*. Each pile (sample in Figure 3) can be viewed as a summarisation, according to relevance criteria, of the utterances taking place between interactions. The pile consists of blocks, each block representing a relevance criterion (each criterion represented by a unique colour). Each block is annotated with the polarity (negativity or positivity) of the utterances, where a negative instance is indicated by a minus sign next to the corresponding block. For example, "this document is too old" might be a negative instance of Currency, while "this document is up-to-date" might be a positive instance. The order of the blocks from the bottom of the pile to the top correspond to the order in which each relevance criteria is first mentioned (subsequent utterances classified as the same criterion and polarity as a previous block are not repeated). The colours are assigned to criteria according to the sequence of colours recommended by Ware's study of effective colour coding [5]. The most frequent relevance criterion is assigned the first colour in the sequence, the second most frequent criterion the second and so on.

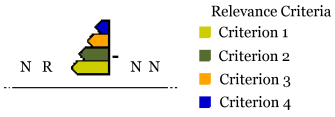

Fig. 1. Example of a relevance criteria pile with four criteria.

By using sequential ordered piles we can analyse whether a user's relevance judgement process exhibits dependencies between relevance criteria. Delimiting processes by interactions will lead to piles not necessarily being aligned with final relevance judgements. However, a finer granularity in interaction encodings in the visualisation may improve alignment.

Relevance Criteria Profiles and Visualisation in Action In this section, we discuss RCPs and session visualisations from real data collected between January to August of 2008. A total number of 21 research scientists, affiliated to one of three school in the Robert Gordon University (the School of Computing, the Information Management Group and the School of Pharmacy) participated. The participants searched outside their research field for literature related to their own area of research, and verbalised their thoughts throughout the session. This was recorded and processed to produce the RCPs.

The aggregated RCP for the group of all users shows that *tangibility* and *depth/scope/specificity* are the most frequently mentioned criteria (595 and 406 times respectively). This tendency exhibited as an aggregated group is not preserved, however, when we consider user behaviour according to research background. The differences resulting from research background on the usage of relevance criteria is observable in Figure 2, where we present the normalised RCPs across groups of users clustered according to their school affiliation. In the figure we can observe that members form the School of Computing show a preference for hard, tangible data while this preference is not as evident in the case of members of the Information Management Group. Members from the School of Pharmacy prefer properties such as length and depth over tangibility.

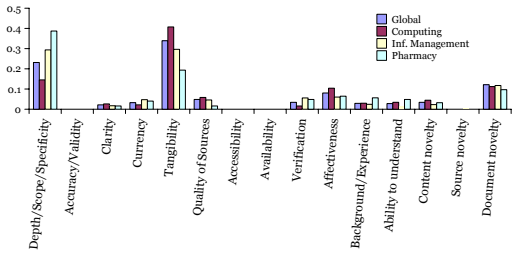

Fig. 2. A comparison of normalised school RCPs and the global RCP

An example of a session visualised is presented in figure 3. The first pile shows that the session begins with a mention of *document novelty*, after which the user navigates away. The second pile ends with a negative mention of *ability to understand* (top most block), again leading the user to navigate away from the document. In the next pile, the user again encounters a known document (black box denoting *document novelty*). The judgement process continues, however, suggesting that the importance of *document novelty* can be overridden, for example, by an affective response to the document. The visualisation also highlights the fact that negative relevance judgements are likely to be preceded by a pattern of the form *negative relevance criteria* $\rightarrow$ *navigation*. Out of the 14 sequences of the form *negative relevance criteria* $\rightarrow$ *navigation*, 12 are indeed implicit negative relevance judgements.

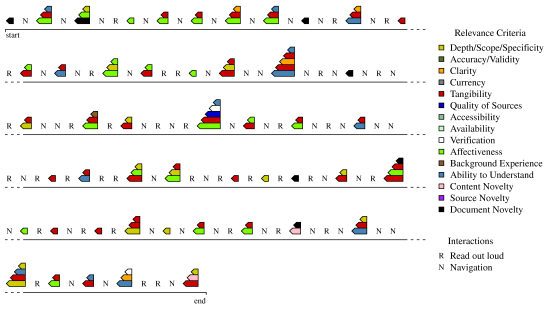

Fig. 3. Example search session visualised: participant 2.

4 Concluding Remarks

In this article, we presented relevance criteria profiles (RCP) and a visualisation technique for relevance judgement processes within a search session. We demonstrated how we might use these to analyse data, and how comparing RCPs can help uncover similarities and differences in relevance criteria usage across users and groups. We also illustrated the potential of visualising relevance judgement processes throughout a session might highlight patterns that lead to final relevance judgements. The next step would be to carry out a fuller comparison of users' relevance criteria, and session visualisations across different groups of users. User interactions might also be further classified to align visualisations with relevance judgements. A comprehensive analysis would likely require aligning relevance criteria, judgement processes, final judgements, and click-through data.

References

1. C.L. Barry. User-defined relevance criteria: an exploratory study. *Journal of the American Society for Information Science*, 45(3):149–159, 1994.
2. C.L. Barry and L. Schamber. Users criteria for relevance evaluation: A cross-situational comparison. *Information Processing and Management*, 34(2-3):219–236, 1998.
3. Pia Borlund. The IIR evaluation model: a framework for evaluation of interactive information retrieval systems. *Information Research*, 8(3):8–3, 2003.
4. L. Schamber. Users'criteria for evaluation in a multimedia environment. *Proceedings of the 54 Annual Meeting of the American Society for Information Science*, 28:126–133, 1991.
5. C. Ware. Color sequences for univariate maps: Theory, experiments and principles. *IEEE Computer Graphics and Applications*, 8(5):41–49, 1988.