# Information Retrieval eXperience (IRX):
# Towards a Human-Centered Personalized Model of Relevance

Frans van der Sluis
*Human Media Interaction (HMI)*
*University of Twente*
*Enschede, The Netherlands*
*f.vandersluis@utwente.nl*

Egon L. van den Broek
*Human-Centered Computing Consultancy*
*http://www.human-centeredcomputing.com/*
*Vienna, Austria*
*vandenbroek@acm.org*

Betsy van Dijk
*Human Media Interaction (HMI)*
*University of Twente*
*Enschede, The Netherlands*
*bvdijk@ewi.utwente.nl*

*Abstract*—**We approach Information Retrieval (IR) from a User eXperience (UX) perspective. Through introducing a model for Information Retrieval eXperience (IRX), this paper operationalizes a perspective on IR that reaches beyond topicality. Based on a document's topicality, complexity, and emotional value, a model of relevance is proposed to influence user's IRX and, consequently, the synthesis and use of the retrieved information. Additionally, methods are discussed to assess UX through interaction and feedback mechanisms. As such, the proposed multi-dimensional IRX model is highly user-dependent and determines document's relevance from a non-traditional human-centered, personalized perspective on IR.**

*Keywords*-**User eXperience (UX), Information Retrieval (IR), Personalization, Human-Centered, Emotion, Relevance**

## I. INTRODUCTION

The goal of an Information Retrieval (IR) system is to solve the information need of its user. Research on how this goal can best be achieved has mainly been dominated by the Cranfield paradigm. This paradigm uses a system-based evaluation, defining precision and recall as evaluation measures for retrieval systems, respectively measuring the number of relevant documents and proportion of retrieved documents that are relevant. Key to such evaluations is the definition of relevance. The decision of relevance is generally performed by domain experts on the basis of topical similarity; i.e., topicality.

Several studies have shown that there is more to relevance than topicality. For example, [12] showed that numerous criteria can be reduced to a core set of five that indicate relevance: topicality, novelty, reliability, understandability, and scope. [1] showed a more lengthy reduction: scope, validity, clarity, currency, tangibility, quality, accessibility, availability, verification, and affectiveness. [7] summarizes relevance in a stratified model, among which a cognitive layer (correspondence between the information and a user's knowledge) and an affective layer (e.g., motivation, intent). If we compare these notions of relevance to topicality, it is clear that topicality is, although the most important, not the only important aspect of relevance.

The multi-faceted notion of relevance pleas for a human-centered approach. A process oriented view on IR further illustrates this. The process of searching information generally has a start and an end. The start is a (more or less clear) information need. The end is some form of synthesis or use of the information. In going from the start to the end, the user is in a continuous conversation with the system: from searching, scanning, judging, to processing Information Objects (IOs). Throughout this iterative process, the user refines her information need. So, essentially an IR system should be human-centered, as it solves the information need of its user, with its user.

To operationalize the pivotal role of the user in solving the information need, we adopt a framework of User eXperience (UX). UX is a fuzzy concept, often defined as technology use beyond its instrumental value (e.g., topicality for IR). Several aspects of UX have been identified; e.g., usability, beauty, hedonic, emotions, temporality, situatedness, enjoyment, motivation, and challenge. Together, these aspects explain part of the UX [3] and are intrinsically related to persistence and effort in information problem solving. Hence, we hypothesize that solving an information need is fostered with an enhanced UX.

A framework of UX would allow to explain how characteristics of IOs influence the UX and how the UX influences the goal of solving an information need. Hence, it can be used to structure what aspects of relevance are of prime importance for a fruitful search experience. One of the rare attempts to operationalize the concept UX can be found in [3], which divides UX in three (partly overlapping) factors:

1) Aesthetic and hedonic factors (e.g., beauty, enjoyment, and extending one's knowledge), regarded as complementary to the instrumental values of a product. Hedonic and aesthetic factors may be a primary reason to search and, thus, be the information need.

2) Emotional factors, addressing the antecedents and consequences of, ideally, positive emotions. Although overlapping with the first group, these factors are not seen as a goal on their own; however, they can aid in solving an information need.

3) Experiential factors, combining all contextual and related factors, including usability. The user states

IEEE computer society

(e.g., mood, expectations, and active goals) interact with the situation and time in creating the experience. Consequently, IR experience is situated and temporal.

In the remainder of this paper we will focus on the second group: emotional factors. This group contains the most influential factors for the instrumental value of IR, as emotional relevance is a core aspect of relevance, related to and overlapping all other types of relevance [2].

At least two clear lines of research on emotion in IR can be identified. One strain of research shows the effect of difficulty (or challenge) compared to the skills of the user. Namely, experienced difficulty leads to negative emotions. Moreover, numerous studies have shown that understandability, comprehension, and complexity are among the core relevance criteria users apply to documents [2], [12]. A second line of research is occupied with reading the emotional value of a text, image, or video. The emotional value of an IO can be considered the most direct antecedent to emotional experience. Please note that these two lines, complexity and affect, do not cover all the emotional antecedents. Moreover, for the total UX of IR, many other aspects of relevance are likely to be influential as well.

In the next section, the salient parts of UX in relation to solving users' information need are identified. This is followed by the introduction of a three-factor model of relevance that influences UX. We finish with a brief discussion in Section III.

## II. Human-Centered Model of Information Retrieval eXperience (IRX)

In this section a model is presented, which will show the feasibility of incorporating different features into a coherent model of relevance, aimed at the Information Retrieval eXperience (IRX). Figure 1 illustrates the general outline of this model: a series of features of the IOs are analyzed
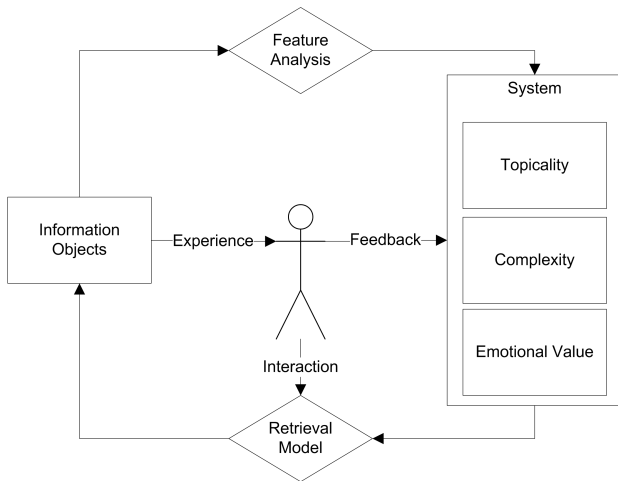


Figure 1. The proposed multi-dimensional, human-centered, personalized model of Information Retrieval eXperience (IRX).

and incorporated into a retrieval model, which supplies IOs directly targeted at the whole UX, contrary to only topicality.

### A. Features

A retrieval system can act by either excluding or including an IO from the search results or by changing the ranking of an IO within the search results. Both possibilities are dependent upon the features that can be derived from the IOs. We will now review the features relevant to the IRX; see also Figure 1 and Section I.

*Topicality*. Key to retrieval models is the weights assigned to query terms. Various weighting methods have been proposed, consisting of three main factors: 1) term frequency $tf$: the number of times a term is present in a document; 2) document frequency $df$: the total number of documents containing the term; and 3) document length $dl$, to normalize for higher frequencies in longer documents. A classic weighting method is tf-idf:

$$w = tf \times log\frac{N}{df}, \qquad (1)$$

Combining the weights ($w$) for different terms can be done with the following basic retrieval model:

$$T(\mathbf{d}, \mathbf{q}) = \sum_{\mathbf{t_i} \in \mathbf{Q}} w_{\mathbf{t_i}}, \qquad (2)$$

where $\mathbf{d}$ and $\mathbf{q}$ are the term-vectors of respectively the document and the query, $t_i$ denotes term $i$, and $\mathbf{Q}$ is the set of terms $t_i$. Over the years, more sophisticated weighting methods (e.g., Okapi MB25) and retrieval models (e.g., vector space models, probabilistic models) have become the standard to estimate topicality [8].

*Readability*. The common approach to estimating the difficulty of a text is by readability measures. These are rough measures, based on textual characteristics such as words per sentence, syllables per word, ratio of polysyllabic words per word, and characters per word. A popular implementation is the Flesch-Kincaid readability formula [4]. It indicates the reading level ($C_1$) of a text, from grade 5 to college level:

$$C_1(d) = 0.39wps(d) + 11.80spw(d) - 15.59, \qquad (3)$$

with $wps$ being words per sentence and $spw$ being syllables per word.

*Entropy*. The amount of information tells something about the resources needed to process that information. Entropy is a measure for the amount of information and has been shown to be indicative for the complexity of information [10]. It indicates the predictability of a next symbol (e.g., a word), based on the occurrence of (a sequence of) previous words. The entropy for a sequence of symbols is defined as:

$$C_{2,n}(d) = - \sum_{B \in A^n, s \in A} p(B,s) \log_b \frac{p(B,s)}{p(B)}, \qquad (4)$$

where $A$ is the set of all possible symbols, $s$ is a symbol, $A^n$ is the collection of all sequences of length $n$, $p(B)$ the probability of sequence $B$, $p(B,s)$ is the probability of sequence $B$ followed by symbol $s$, and $b$ is the logarithmic scale (usually 2; i.e., bits). Larger values of $n$ give a more precise measure of the information content of a signal.

*Semantic coherence.* Coherence is related to the complexity of a text, such that incoherent texts are often perceived as more complex [5]. The following measure quantifies the degree of connectivity across sentences, based on the idea that coherent texts contain a high number of semantically related words. In general the following definition is used:

$$C_3(d) = \frac{\sum_{i=1}^{n-1} \text{sim}(S_i, S_{i+1})}{n-1} \qquad (5)$$

where $d$ is a text, and $\text{sim}(S_i, S_{i+1})$ is a function for the similarity between two sentences; e.g., for word-overlap, the number of words co-occurring in both sentences [5].

*Emotional keyword spotting.* This is the most basic approach available to detect the emotional value of a text, merely counting the occurrence of unambiguous emotion words (e.g., happy and sad), which are often grouped into emotion categories [6]. The count of occurrences of each (unambiguous emotional) word $w$ of lexicon $L$, emotion category $L_c$ in IO $d$ is defined as:

$$E_1(d) = \#\{w \in d | w \in L_c\} \qquad (6)$$

*Lexical affinity.* This measure is similar to keyword spotting but includes ambiguous emotion words as well. Every word $w$ in the lexicon gets a probability $p_c(w)$ assigned, indicating its affinity with a particular emotion $c$ [6].

$$E_2(d) = \frac{\sum_{w \in d} p_c(w)}{\#\{w \in d\}} \qquad (7)$$

Together these measures result in a set of features: $\{T, C_{1...3}, E_{1,2}\}$. Note that more features can be calculated than mentioned in this overview. This overview merely served to illustrate the possibility of deriving features from IOs that are relevant to the IRX.

*B. Model*

To process the set of features, we need a multi-dimensional model of relevance. One of the possibilities is to use a Linear Regression Model (LRM). A LRM is an optimal linear model of the relationship between one dependent variable (e.g., relevance) and several independent variables (e.g., topicality $T$, understandability $U$, and emotional value $E$). A LRM typically takes the following form:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \varepsilon, \qquad (8)$$

where $\varepsilon$ represents unobserved random noise, and $p$ represents the number of predictors (i.e., independent variables

Table I
HUMAN-CENTERED MEASURES

Short overview of measures to operationalize the presented concepts.

**Interaction**. The most common interaction paradigm for IR is query-based, capturing the user's information need in a query. The interaction can easily be extended with the possibility to indicate a need for IOs that are easy to understand or elicit positive emotions.

**Explicit feedback**. Besides the interaction, a user can indicate several values about the retrieved IOs through explicit feedback mechanisms. Such are commonly used; e.g., indicating if a search result is liked (social tagging).

**Implicit feedback**. A broad plethora of objective user measures exists: physiological measures, movement analysis, computer vision techniques, and speech processing. All measures have shown to be useful in human-computer interaction. Moreover, for IR, it is common to use click-through data (or related measures) as an indication of topical relevance.

$x$ and regression coefficients $\beta$). With this scheme, the following models can be based on the methods to act:

$$R \to \{T, U, E\} \qquad \text{Relevance;} \qquad (9)$$
$$U \to \{C_{1...3}\} \qquad \text{Understandability;} \qquad (10)$$
$$E \to \{E_{1,2}\} \qquad \text{Emotional value.} \qquad (11)$$

Please note that the user plays a role in each of the three models: for topicality, this is through the query; for emotional value, the user's preferences will influence the final experienced emotion, and understandability is not only dependent upon the complexity of the IO but also on the skills and knowledge of the user.

The regression coefficients ($\beta$) are commonly derived by a linear regression analysis, trying to solve multiple equations of the type of Equation 8. Each equation represents an (empirical) observation of the dependent and independent variables. Consequently, observations are of key importance for creating the model: the user really has to be incorporated in the model; Figure 1 and Section II-C illustrate how.

*C. Personalization*

This section shows three methods to incorporate the user into the model: interaction, implicit, and explicit feedback. Table I gives an overview of each. Some of the methods in Table I are more preferable than others. Interaction and explicit feedback bring substantial costs to the user, concerning the time and effort needed for these measures. Implicit feedback mechanisms have low costs, provided that they are unobtrusive and reliable; hence, these mechanisms are preferable. The usefulness of each measure, and in particular for each factor of the model (understandability, emotion, and relevance), will be reviewed next.

Understandability and emotion can be measured implicitly, allowing to make Equation 10 and 11 user-centered. For example, the cognitive load can be measured, indicative of the complexity of an IO. Moreover, the user's query history has been proposed as an indication of what is likely understandable for the user [11]. Emotion has been measured

by heart rate and skin conductance response, or other inter-action modalities such as speech [9]. Explicit mechanisms are also common and of prime importance for the evaluation of the models in Equation 10 and 11. Smileys, social tagging (liking), Likert-scales, and semantic differentials can allow the user to supply feedback on emotion or understandability.

Relevance is commonly measured via both explicit and implicit feedback. Using such measures it is unlikely only topical relevance is measured, as opposed to the more general concept of how relevant an IO is for a user in a certain context. Hence, we pose that multi-dimensional relevance can be measured through feedback mechanisms that allow to refine the model in Equation 9 to user's preferences.

Given the available measures, the model as proposed in Section II-B can be made user-centered, using both implicit and explicit feedback mechanisms. Alternatively, a work-around is possible by including the different parameters in the interaction cycle.

## III. DISCUSSION

Although IR aims to solve people's information need, the field is not truly human-centered and seems captured by its own formal methods. In contrast, with this article we pose to approach IR from a UX perspective and coin IRX. As UX is a fuzzy concept, we introduce a three-factor model of relevance that operationalizes UX. This model addresses the emotional experience of users and, consequently, assesses document relevance from a non-traditional perspective.

The model and its features have yet to be benchmarked and personalized, using the methods Section II-C describes. Moreover, the model is not a full reflection of the IRX of the user: some uncertainty will be inevitable. Using novel interaction paradigms, part of this uncertainty can be solved. However, is must be acknowledged that the perfect IRX is still far beyond reach.

The IRX framework presented in this article is founded on the integration of notions that originate from various scientific disciplines. As such, IRX is a true interdisciplinary endeavor. It is shown how the fuzzy concept UX can be utilized in the formally specified field of IR and, as such, illustrates that a human-centered approach can be formal as well. As no experimental validation or testing has been employed so far, IRX still has to prove its use in practice. Notwithstanding, in time, we believe that IRX will bring us human-centered personalized IR.

## ACKNOWLEDGMENT

## REFERENCES

[1] C. L. Barry and L. Schamber. Users' criteria for relevance evaluation: A cross-situational comparison. *Information Processing & Management*, 34(2-3):219 – 236, 1998.

[2] E. Cosijn and P. Ingwersen. Dimensions of relevance. *Information Processing & Management*, 36(4):533 – 550, 2000.

[3] M. Hassenzahl and N. Tractinsky. User experience - a research agenda. *The American Journal of Psychology*, 25(2):91–97, 2006.

[4] J. P. Kincaid and et al. Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. Technical report, National Technical Information Service, Springfield, Virginia, 1975.

[5] M. Lapata and R. Barzilay. Automatic evaluation of text coherence: models and representations. In *IJCAI'05: Proceedings of the 19th international joint conference on Artificial intelligence*, pages 1085–1090, San Francisco, CA, USA, 2005. Morgan Kaufmann Publishers Inc.

[6] H. Liu, H. Lieberman, and T. Selker. A model of textual affect sensing using real-world knowledge. In *IUI '03: Proceedings of the 8th international conference on Intelligent user interfaces*, pages 125–132, New York, NY, USA, 2003. ACM.

[7] T. Saracevic. 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, 2007.

[8] A. Singhal. Modern information retrieval: A brief overview. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*, 24(4)::35–43, 2001.

[9] E. L. Van den Broek, J. H. Janssen, J. H. D. M. Westerink, and J. A. Healey. Prerequisits for Affective Signal Processing (ASP). In P. Encarnação and A. Veloso, editors, *Biosignals 2009: Proceedings of the International Conference on Bio-Inspired Systems and Signal Processing*, pages 426–433, Porto – Portugal, 2009.

[10] F. Van der Sluis and E. L. Van den Broek. Applying ockham's razor to search results: Using complexity measures in information retrieval. In *Information Interaction in Context (IIiX) Symposium*, New York, USA, in press. ACM.

[11] F. Van der Sluis and E. L. Van den Broek. Modeling user knowledge from queries: Introducing a metric for knowledge. In *Proceedings of the 2010 International Conference on Active Media Technology*, Lecture Notes in Computer Science. Springer, in press.

[12] Y. C. Xu and Z. Chen. Relevance judgment: What do information users consider beyond topicality? *J. Am. Soc. Inf. Sci. Technol.*, 57(7):961–973, 2006.