

# Is Relevance the Right Criterion for Evaluating Interactive Information Retrieval?

Nicholas J. Belkin, Michael Cole, Ralf Bierig  
SCILS, Rutgers University

[nick@belkin.rutgers.edu](mailto:nick@belkin.rutgers.edu), [mcole@scils.rutgers.edu](mailto:mcole@scils.rutgers.edu),  
[bierig@gmx.net](mailto:bierig@gmx.net)

RUTGERS



## ABSTRACT

The standard criterion on which to evaluate information retrieval systems has long been the relevance of document(s) to information problem. However there is also a long history of alternatives to relevance as a criterion, and there is reason to believe that relevance is an inappropriate criterion for *interactive* IR. We suggest that: a) *usefulness* is a more appropriate evaluation criterion than relevance for interactive information retrieval (IIR); b) IIR requires evaluation of the entire information-seeking episode; and c) there are unlikely to be universally applicable measures based on usefulness, and therefore measures for evaluating IIR need to be tailored not only to the task which led to information seeking behavior, but also to the searching tasks or information seeking strategies in which a searcher is engaged at any point in an information-seeking episode.

# Way Beyond Binary Relevance

- This position statement tries to take the most general statement of the goal of IR seriously, in terms of evaluation of IR system performance
- This leads us to question whether relevance is a criterion according to which we really should be evaluating IR system performance.

# Goals, Criteria, and Measures

- The *goal* of a system determines the *criteria* against which its performance should be evaluated
- The criteria, and the *context*, then determine *measures* which allow evaluation of the system's performance

# How did Relevance Come to be the IR Criterion?

- Pre-History of IR
  - Documentation, subject bibliography
  - Characterize the topics of documents, identify documents treating a topic
  - Static, one time event: statement of need, response
- Evaluation of methods for representation, identification
  - Treating, or being about a topic, was identified as a document's being *relevant* to a topic
  - Success interpreted as extent to which goal of subject bibliography is achieved, i.e., that a bibliography was both complete and correct

# IR versus Documentation

- Shift from fairly general topics of long-standing interest to more specific, immediate topics
- Shift from human to mechanical identification of topicality
- *But*, remains a static, one-time event, and
- Relevance, as treating the topic, remains the criterion for evaluation

# What has been Construed as the Goal of Information Retrieval?

- Helping a person to find information useful in accomplishing a task or achieving a goal
- Helping a person to find information appropriate to his/her situation
- Helping a person to find information appropriate to a problem
- Helping a person to find information about a topic

# What has been Meant by *Relevance*?

- Topically relevant
  - About a topic
- Communicationally relevant
  - Appropriate to a specific discourse or problem
- Situationally relevant
  - In congruence with a person's circumstances
- Pragmatically relevant
  - Useful in achieving some goal or accomplishing some task



# Alternatives to Relevance

- Utility
- Use
- Satisfaction
- Task-specific criteria, e.g.
  - Question-answering
  - Topic distillation
  - Aspectual recall

# What's Wrong with this View of IR?

- Static, one-time event
  - IR is inherently an interactive process; previous view based on historical and technical constraints
- Goal is construed as having found information
  - Goal of IR should be resolution of problem, accomplishment of task, achievement of user's goal
- Criterion is topical relevance
  - Criterion should be *usefulness* in accomplishment of goal

# Why does IR Happen?

- IR is initiated by a person's real-life *task* or *goal*
- Achievement of goal requires resolution of an *information problem*
- Person engages in IR system, in order to resolve information problem and thereby achieve goal (accomplish task)

# IR as an Interactive Process

- IR construed as an *information-seeking episode*, rather than one event, or a sequence of unrelated events
- An episode consists of a sequence of interactions between person and information object(s), related to one-another with respect to initiating task and related information problem

# Information Seeking Strategies and Situations

- Each interaction in an information-seeking episode can be construed as a specific *ISS*
- Selection of an ISS is conditioned by the current *situation*
- Each ISS has its immediate goal, and its goal with respect to accomplishment of the initiating task, resolution of the general information problem

# Evaluation of Interactive IR

- Evaluation of the information-seeking episode as a whole, with respect to accomplishment of task, resolution of problem
- Evaluation of each interaction, with respect to its specific goal, and to its contribution to accomplishment of the overall goal

# Criterion for Evaluation of IIR

- How *useful* was the information-seeking episode in accomplishing the goal
- How *useful* was each interaction in helping to achieve the overall goal
- How *useful* was the system in supporting each interaction

# Measures ...

- For the information-seeking episode as a whole
  - Task specific, e.g. perceived usefulness
  - General, e.g. task accomplishment, cumulated effort
- For each interaction
  - ISS specific, e.g. amount learned,
  - General, e.g. effort
  - WRT overall goal, e.g. perception of usefulness



# ... and Methods

- Formal
  - Fuhr suggests cumulating cost of decisions over sequence of situations
- Interactional, situational
  - Setting specific tasks in experimental settings
  - Direct elicitation of usefulness, and related, judgments

# Acknowledgements

- This research supported by a grant from the Institute of Museum and Library Studies
- Related work on this topic at <http://scils.rutgers.edu/imls/poodle>

## Some References

- Fuhr (2008) A probability ranking principle for interactive information retrieval. *Information Retrieval*, 11: 251-265
- Yuan, X.-J. & Belkin, N.J. (2007) Supporting multiple information-seeking strategies in a single system framework. In *SIGIR 2007* (pp. 247-254). New York: ACM.
- Belkin, N.J. (1996) Intelligent information retrieval: whose intelligence? In *ISI '96* (pp. 25-31). Konstanz: Universitätsverlag Konstanz.
- Belkin, N.J., Cool, C., Stein, A. & Thiel, U (1995) Cases, scripts and information seeking strategies: on the design of interactive information retrieval systems. *Expert Systems with Applications*, vol. 9 (1995): 379-395.