

Jonathan D. Warren
Indiana University, SLIS
DSRF abstract, final submission
2005-09-16

“On Similarity”

This presentation and its forthcoming paper suggest the need for a fresh conceptual look at system-centered theories of relevance. Addressing work by the core authors currently writing on relevance from both the user- and system-centered perspectives, the paper argues that, though a vast literature exists on relevance with respect to user perceptions and system performance, a great void of theoretical attention persists around what IR system developers mean when they say that results returned by a search engine are ‘relevant’. In actuality, they mean a great many things: all use one of many similarity metrics as the mechanism handling retrieval and ranking, which in turn takes for granted one of many ways of representing and manipulating resource objects inside an IR system.

The paper begins with an ironic double-criticism of approaches to relevance theory from both user- and system-centered perspectives, with a particular eye to how the perspectives discourse with each other. Going beneath the rhetoric, the paper identifies several difficult paradigmatic and operational issues that hinder both perspectives from realizing a unified vision of relevance theory and method. Key among those hindrances, the paper argues, is fragmentation among system theories of object-comparison (“similarity”). As a first attempt at combatting that fragmentation, the body of the paper proceeds through five categories of the most popular similarity measures, explains in non-technical language how they work along with their assumptions and limitations, and offers both connections between the system- and user-centered notions of relevance as well as criticisms highlighting where user-system miscommunications about the meaning of relevance often occur. The paper ends with a commentarial discussion of four themes that arose in the paper's analysis of system relevance theory, which seem to be primary contributors to user-system miscommunications, if not challenges facing IR as a whole.