

Measures of Measurements: Robust Evaluation of Search Systems

Justin Zobel

National ICT Australia

justin.zobel@nicta.com.au

A good search system is one that helps a user to find useful documents. When building a new system, we hope, or hypothesise, that it will be more effective than existing alternatives. We apply a measure, which is often a drastic simplification, to establish whether the system is effective. Thus the ability of the system to help users and the measurement of this ability are only weakly connected, by assumptions that the researcher may not even be aware of. But how robust are these assumptions? If they are poor, is the research invalid? Such concerns apply not just to search, but to many other data-processing tasks. In this talk I introduce some of the recent developments in evaluation of search systems, and use these developments to examine some of the assumptions that underlie much of the research in this field.