

HCSNet Next Generation Search Workshop Talk

Baden Hughes

The University of Melbourne

baden.hughes@unimelb.edu.au

Areas of Interest

Research

- Research areas which overlap the NGS agenda
 - Digital libraries, particularly search services for distributed data environments and metadata standards
 - Web data mining, large scale semi-structured data acquisition and information retrieval engines
 - The intersection of generic search technologies and domain-specific highly structured, such as linguistic data on the web
 - Application of broad coverage classification techniques (eg for language identification of documents) to web data as a catalyst for higher order domain specific search applications.

Solutions Delivery

- Looking at next generation search from a different perspective – Information Services at UniMelb
 - Stronger focus on engineering and solution delivery rather than research in shorter term
 - How can we more effectively use the empirical data we have about users and their search behaviour ?
- Over 5 year timeframe, considering how “enterprise” search at UniMelb may change:
 - New information discovery paradigms
 - New information delivery mechanisms
 - Emergence of new problems which may drive a research/engineering cycle
 - Tractable solutions to open problems

Areas of Intersection

Ranked Lists or Something Else ?

- Decades of research in information retrieval and “relevance”
- Emergence of more generally accessible interfaces for information discovery such as web search engines
- (Web) search landscape still dominated a reasonably standard output display : ranked lists
 - Assumption that ranking is the best practical method for instantiating relevance of a document to a query
- Wide variety of alternative display types for engaging in information discovery tasks are available
 - But these are not actively or widely deployed
- Cognitive science research shows that different modes of information engagement result in different information outcomes
 - Comparatively little research in how knowledge about human communication preferences can be applied in the web search context, particular in results display and user interaction alternatives

Binary or Graduated Relevance ?

- Fine grained semantic distinctions can have profound effects on interpretation of communicative intent
 - A single utterance or piece of text can have many different interpretations depending on the context it is presented in, and the perspective from which the consumer approaches it
- Such semantic ambiguity contrasts markedly with the granularities typically adopted in information retrieval
 - Considering typical IR judgement tasks as "Is this document relevant to a given query ? Yes or No". (ie the Cranfield model permeates)
- Some research has been conducted into graduated assessments which allow for greater human interpretation of the concept of 'relevance' between a query and a document set
 - A large number of open questions, in both science and engineering terms, as to how to effectively provision for semantic ambiguity in determining relevance