Information Retrieval

hussein suleman
uct cs honours 2016

Things still to finish…

- Assignment 1b (Thursday 19 May)
- Assignment 2 (Thursday 2 June)

- No Exam

- Lecturer Evaluation (today)

What did we discuss?

- Basic concepts in IR
  - term, document, relevant, query, filtering/ranking
- Text content algorithms
  - boolean, extended boolean, vector, LSI, probabilistic
- Key optimizations
  - inverted files, idf, stemming, stopping, thesauri, relevance feedback, decision trees,...
- Link algorithms
  - PageRank, HITS
- IR evaluation
  - recall, precision, F, MAP, NDCG
- Testbeds, metadata and crowdsourcing
- IR applications
  - recommendation, classification, CBIR, CLIR,...

What did we not discuss?

- How to create your own search engine in practice.

- If you want to create a document-based IR system?
  - Use Apache's SOLR
- If you have a database, many DBMSes will support IR-style "full text" indexing.
  - `SELECT documentID FROM documents WHERE MATCH(content) AGAINST('apples,bananas')`

What are the take-home messages?

- Use IR where appropriate.
- Search is King!
- …

What should be done differently?

- ☐ How to improve the module?
- ☐ Or any other comments?

Where to from here?

- Research in IR at UCT
  - IR in marginalized languages (isiXhosa, isiZulu, Arabic, |Xam, !Kun)
  - IR for Development
  - IR for Education
- Research in IR elsewhere
  - University or Industrial Research Labs
- Applications of IR
  - Honours project!
  - Future projects at work and elsewhere

The End ...

- ❑ well, not really, but lectures in this module anyway :)