

School of Computing and Information Systems  
The University of Melbourne  
COMP90049  
Knowledge Technologies (Semester 1, 2018)  
Workshop exercises: Week 5

1. What is the difference between “data retrieval” and “information retrieval”? Why is the latter a knowledge task, but the former is not?
2. (Extension) How many books are there in an average library? How many words are there in an average library? How many documents are there on the World Wide Web? How many words?
3. Identify some different types of “informational needs.”
  - (a) Give examples of queries which might indicate a particular type of informational need. Are some of them ambiguous for the type of need?
  - (b) (Extension) Input some queries of different types of informational needs into a web search engine like Google. Are search engines better at responding to some types of informational need than others? Is there any indication that the search engine is identifying the type of need and tailoring the results toward that?
4. In the context of an **information retrieval** engine, what does it mean for a document to be **returned** for a query? What does it mean for a document to be **relevant** for a query?
5. Identify some differences between **Boolean** querying and **ranked** querying, in an Information Retrieval context.
  - (a) (Extension) Obviously, search engines like Google use ranked querying by default ... or do they? What evidence do we have that there is a “hybrid” Boolean component in Google’s typical behaviour?
6. Identify the two (sometimes three) components of a **TF-IDF model**. Indicate the rationale behind them as in, why would they contribute to a “better” result set?