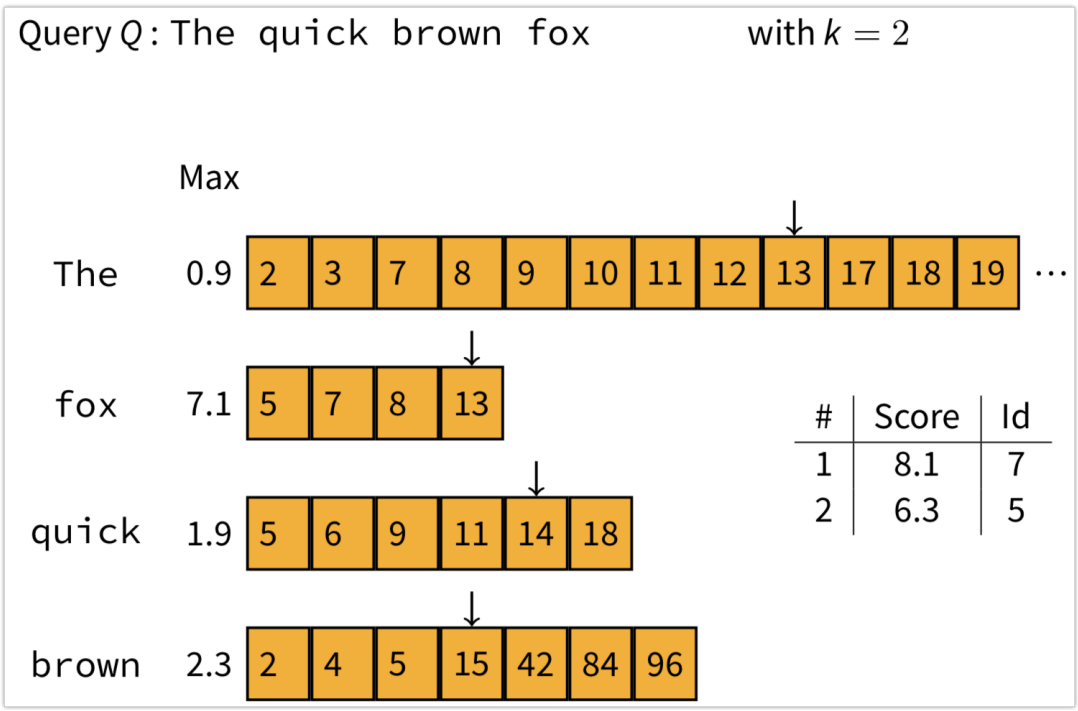
School of Computing and Information Systems
The University of Melbourne
COMP90042 WEB SEARCH AND TEXT ANALYSIS (Semester 1, 2019)

Workshop exercises: Week 3

Discussion

1. Data compression of a **postings list** in an inverted index can help reduce space usage of the index.
   (a) What is the intuition behind compression algorithms used for postings list compression? Why do they work?
   (b) What is **Variable Byte Compression** and how does it compress an integer?
   (c) Determine the values of integers X and Y that were encoded as the byte sequence [52,34,147,42,197] using the Variable Byte algorithm described in the lecture slides 9/10.
2. Algorithms such as WAND help speed up query processing.
   (a) What is the intuition behind WAND? What is the output produced by WAND?
   (b) What extra information is stored for each term to allow algorithms like WAND to skip evaluating documents? How is it computed? What restriction does it place on the query process?
   (c) Assume Document 13 has just been evaluated. In the setting below, what is the next document that will be evaluated?



3. Discuss the concepts of **query expansion** and **relevance feedback** and how they are related.

4. Discuss and give an example of the following forms of relevance feedback.
   (a) User relevance feedback
   (b) Pseudo relevance feedback
   (c) Indirect relevance feedback
5. Is it possible to perform **query expansion** without **relevance feedback** and vice versa? Discuss!

Programming

1. Work on Homework 1! `:-)`

Catch-up

1. Make sure that you have a good understanding of `WSTA_N1B_preprocessing` and `WSTA_N2_information_retrieval` iPython notebooks.