# Task9

Nawar Saeed

Artificial Intelligence for the Web, VT21
DT506A

16 april 2021

Let's assume that we have a list L of top-k documents and its scores according to the set below:

$$X = \{id_1 : s_1, \ldots, id_k : s_k\}$$

Lets assume that the set is sorted, the highest score appears first in the set in such way:

$$s_1 \geq, \ldots, \geq s_k$$

The task is about to inspect the top-k documents depending on how the set is sorted, namely whether the set should be sorted by document ID or by score. For this, two additional sets will be created to perform a comparison in order to find the most sufficient sorted order.

$$Set_1\{id_i : s_i, \ldots, id_n : s_n\}, for\ i = 1 \ldots n$$

$$Set_2\{id_j : s_j, \ldots, id_m : s_m\}, for\ j = 1 \ldots m$$

Now, if the elements in $Set_1$ and $Set_2$ are sorted after the id, it should be needed to inspect the whole $Set_1$ and $Set_2$ . This should be so due to unknown positions to the document ID with the highest score.
This will require many comparisons between the elements of $Set_1$ and $Set_2$ in order to find the top-k documents. It will require $k * (m + n)$ times before top-k documents are found, which is not so efficient if the sets are incredibly large.
If the elements in $Set_1$ and $Set_2$ are sorted by scores, there will be little difference here. The difference here is that you do not need to inspect the whole sets before finding the top-k documents, because the elements are sorted by scores, the highest is first in the list and the smallest is last in the list. This means that it will require a k-inspections / comparison to find the top-k documents.
In summary, sorting by scores will be a sufficient reason for inspection of each inverted list.