

Information Retrieval

8 September 2021 – time 45 minutes

Question #1 [rank 5+3]. Given the set of strings $S=\{0001100, 0001111, 00110, 001111\}$.

1. Build a trie data structure over S .
2. Show how it is found the lexicographic position of the string “0010” by searching it in the trie of the previous item.

Question #2 [rank 6]. Decompress the LZ-compressed string $\langle 0,a \rangle \langle 1, 2, b \rangle \langle 3, 4, EOF \rangle$.

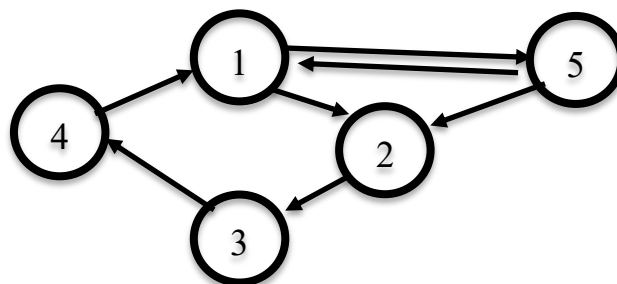
Question #3 [rank 6]. Given the adjacency lists of a Web graph referring to two nodes:

15 \rightarrow {18, 19, 21, 26, 29}

16 \rightarrow {18, 19, 26, 29, 30, 31, 32, 33}

Encode the Adjacency List of node 16 with respect to the one of node 15 via the compression algorithm Web Graph, by returning the so called “copy blocks” and the “extra nodes”.

Question #4 [rank 6+4].



1. Assume that you are given the graph above, and you are interested in computing which is the most related node between 2 and 4 to node 1, without making any assumption on the graph, hence all nodes and edges are equivalent (equiprobable?). How you’d proceed and which is the result?
2. Assume that you know that edges (1,5) and (1,2) have not the same “weight” but the former is twice more important (probable?) than the latter. Recompute the relatedness between nodes 1 and 2.