1. Web Crawling
   a) One important pre-processing aspect for hyperlinks that have been detected by a web crawler is the transformation to a canonical representation. Give a list of different URI examples that have to be transformed to a uniform representation.

   b) Compare the two approaches breadth-first and depth-first search with each other and give a short evaluation for them.

2. TFIDF
   Imagine a set of four text documents that contain one of the following sentences each:
   - “Computer Science is a very important course of studies.”
   - “All science is either computer science or stamps collection.”
   - “Computer Science is no more about computers than astronomy is about telescopes.”
   - “I hate computers.”

   Calculate the overall TFIDF ranking value of the query “computer science” for each of the documents. Which document is ranked on the highest position?

3. PageRank
   a) Have a look at the simple document graph depicted below.

   Calculate the PageRank by an iterative approach based on:
   1. The simple random surfer model without adjustment factor.
   2. The regular random surfer model with adjustment factor (0.15).

   b) What are shortcomings of the regular PageRank approach?