1) One type of sentiment analysis that has posed particular challenges is detecting sarcasm. What are good features for a sarcasm detection for the following corpus?
   D1: A woman needs a man like a fish needs a bicycle. (Sarcasm)
   D2: A woman and a man went to the market on their bicycles and bought some fish. (non-sarcasm)

   Explain why the features you suggest work. Give another example of sarcasm where the feature you suggest works well, and an example where it doesn’t.

   Would a classifier trained on bag-of-words features be able to classify the above corpus effectively? Why or why not?

2) Consider the following approach to answer questions in an automated fashion:

   We have a collection of articles in the archive. We have a question and possible candidate answers for it. We augment each candidate answer to the question to form a candidate answer statement. Then we score the answers based on the frequency of candidate answer statements occurring in the corpus.

   For example:
   Q1: What is the capital of Columbia?
   C1) Cartagena
   C2) Cucuta
   C3) Bogota
   C4) Medellin

   We would form following candidate statements:
   Cartagena is the capital of Columbia
   Cucuta is the capital of Columbia
   Bogota is the capital of Columbia
   Medellin is the capital of Columbia

   Presumably C3 would score much higher than the rest in the corpus we have.

   Mention two potential issues that can arise from such an approach. As one possibility, consider how the approach might fail for the question Who is the current president of the USA?
   C1) Donald Trump
   C2) Barack Obama
   C3) George Bush
   C4) Abraham Lincoln
3) Find a question answering system on the web. Devise a (brief) evaluation process for the system and evaluate it.

4) Given the following web pages (in Danish):
https://europa.eu/european-union/topics/consumers_da
https://europa.eu/european-union/topics/culture_da

And the following queries (in English):
Consumer protection
Digital technologies for the arts

Consider two methods of cross-lingual information retrieval: Machine translation applied to the pages and then issuing the original query against the translated documents, and machine translation applied to the query then issuing the query against the original documents. Try both methods (you will need to find a machine translation website), and determine which approach seems more likely to be effective. Explain why.

5) Describe how you might use relevance feedback in a content-based filtering system. Would you expect relevance feedback to be more or less effective in content-based filtering than in ad-hoc retrieval? Explain why.