1. Features that work well for the given example:
   a. Conjunctions
      1st sentence: like
      2nd sentence: and
   b. Similarity of subject and object in phrase chunks separated by conjunctions
      1st sentence: similarity(man, woman), similarity(fish, bicycle)
      2nd sentence: sim(man, bicycle), sim(woman, bicycle), sim(man, fish), sim(woman, fish)

   The intuition is to see the conjunction ‘like’ and expect the similarity scores to be roughly equal for non-sarcastic sentences and significantly different for sarcastic sentences.
   In the second sentence, conjunction ‘and’ doesn’t imply any such relation.

   Identifying reasonable features - 5 pts
   Explaining why the features work - 3 pts

   No, bag-of-words model doesn’t work well on the given example (3 pts)
   Because, the sentences have roughly similar vocabulary. Sarcasm arises from the way the sentence is structured rather than from the words present in the sentence (3 pts)

   (Total capped to 10 pts)

2. Potential issues that arise from such an approach:
   (5 pts for each well-conveyed issue (total capped to 10 pts))
   a. Validity of the information changes suddenly with events but the frequency of the candidate answers only changes slowly. We might be wrong for many questions whose answer changes frequently.
   b. Synonyms and word order are not being taken care of. E.g: head-of-state etc also refer to the president.

3. The selection of QA system could vary. Grading is based on the following rubric.

   [Grading Rubric]
   -2 No selection of QA system
   -2 No evaluation metric (or not appropriate use)
   -2 No result (or not correct) given the proposed metric
   -2 No appropriate analysis or discussion on the result
Assume that “consumer protection” is more relevant to Doc1, and “Digital technologies for the arts” is related to Doc 3. The query “consumer protection” is relatively easy to translate without the context, so it would be likely to be search similarity when we use tf-based or Boolean query model according to Table 1 and 2. For the query, “Digital technologies for the arts”, the term “art” could be translated differently when we use the context in the document, but Google translate for “query” works well in this case, and it does not make some difference for both case.

In practice, translating all the document computationally expensive, and the quality of translation and context sharing could be not consistent across sentences. As long as the query is not confusing without any other contextual information, translating the query and then searching is safer way.

[Grading Rubric]
-2 Does not provide correct analysis
-2 Does not provide actual evidence or clues about arguments (Thus, too much high-level)
5. Describe how relevance feedback can be used in content based filtering systems:
The idea behind relevance feedback is to take the results that are initially returned from a given
query, to gather user feedback, and to use information about whether or not those results are
relevant to perform a new query.
For example, if a user is interested in watching thrill and action movies, he/she initially would
receive a random list of movies based on other’s ratings of the movies. Once he/she interacts with
the system by clicking on some movies to watch, we can filter based on these initial
recommendations to provide more movies with thrill and action.

Comparison: Relevance feedback is much more effective in content based filtering than it is with
ad-hoc retrieval.
Reasoning: In content-based filtering system, each user would have a user profile. We can use
relevance feedback to finetune user profile in order to give better recommendations.
In ad-hoc retrieval model, it does not track specific user data, so when a user issues a query, it will
only return what most people think what is relevant to the query, but it might not be what this specific
user is looking for.

[Grading Rubric]
5 Describe how relevance feedback can be used in content filtering
3 Comparison correct
2 Reasoning for comparison correct