Task Sheet 4
Signatures & Vector Space Model

Preliminary Notes:

All notes and requirements of PR02 & PR03 still apply!

Task 1 – Signatures

1. Implement the usage of signatures as another search mode for the Boolean retrieval! Choose a suitable hash function and use appropriate data structures for the signatures! For the creation of the signatures you may use the parameters $F = 64, D = 4$. You should determine an optimal value for m yourself.
2. Make it also possible for the signature search to use the conjunction and disjunction of terms! Negation may be disregarded. You can assume that users only enter valid queries.

Task 2 – Vector Space Model

1. Enable the usage of the Vector Space Model with inverted lists by implementing the associated methods and classes, particularly `models.VectorSpaceModel` and `buckley_lewit_search`.
2. Use *tf.idf* for the generation of term weights! The generation of the query vector should be done according to Salton/Buckley (1988).
3. Use the base algorithm with inverted lists as it was presented in the lecture!

Hints:

- Since the VSM does not support Boolean junctors, you do not have to keep them in mind when processing a query with this model. You can assume that the user will only insert valid queries, which are terms separated by one space character each.
- You do NOT have to implement both a linear and inverted-list variant of this method. You can start immediately with inverted lists.

Task 3 – Fuzzy Set Model

1. Enable the usage of the Fuzzy Set Model by implementing the associated methods and classes, in particular `models.FuzzySetModel`.
2. Evaluation of Boolean junctors in the query should be done using the Zadeh operations.

Make sure that your solution considers all requirements listed in this file and upload it on Moodle until the specified deadline!