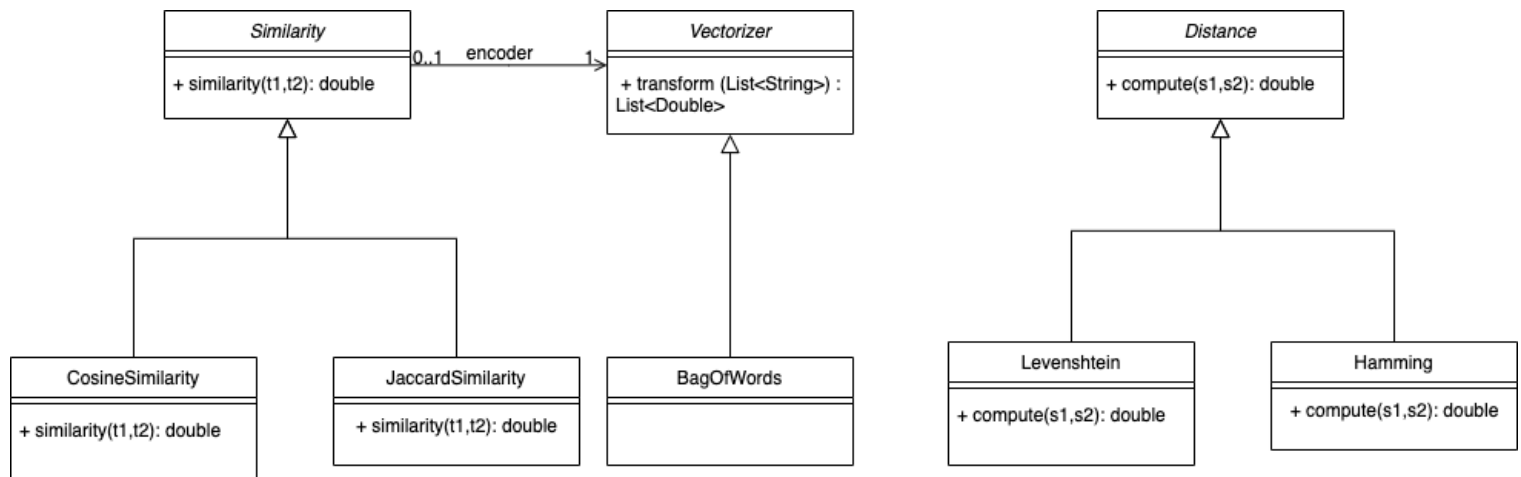


Seminar 5



1. Implementieren Sie auf Basis des obigen Diagramms eine NLP-Bibliothek (Natural Language Processing). In dieser Iteration soll die Bibliothek die folgenden Funktionalitäten bereitstellen:

- das Berechnen der Similarity zwischen zwei Strings und Texts

2. Gegeben sei eine CSV Datei: train_all.csv. Implementieren Sie eine Klasse Distancer mit zwei Methoden:

- `findWord (String word_to_find):` gibt den am besten passenden String in der Datei (Spalte qtext) zurück
- `findText (String text_to_find):` gibt den am besten passenden Text in der Datei (Spalte aText) zurück

3. Links für Algorithmen

- Bag of Words: <https://machinelearningmastery.com/gentle-introduction-bag-words-model/>
- Cosine Similarity: <https://paulminogue.com/index.php/2019/09/29/introduction-to-cosine-similarity/>
- Levenshtein Distance: https://en.wikipedia.org/wiki/Levenshtein_distance
- Hamming Distance: https://en.wikipedia.org/wiki/Hamming_distance
- Jaccard Similarity: <https://deepai.org/machine-learning-glossary-and-terms/jaccard-index>

4. Test Framework

Implementieren Sie ein einfaches Software-Test-Framework mithilfe des Builder-Patterns. Ein möglichen Anwendungsfall ist:

```
TestBuilder.describe("Sum of numbers").expect(result).equal().to(expectedResult);
```

Nutzen Sie das Framework um ein Algorithmus zu testen