**Final Project Proposal#11 Snippet generation algorithms**

**People:** Ziming Song

**Dataset**: same as Assignment 2

**Method**:

Use disjunctive query to get top K result. Limit each snippet length in range(m,n). Here are some possible methods. Each word in a snippet is complete, e.g. no “dinos”, it should be “dinosaur”.

1. Find snippets where one of the query words first occurs (by position).
2. Define the most important query word *key* in a result. Traverse document, find snippets where the query words *key* occurs (by position).
3. Traverse document, combine text where query words occur (by position).
4. Efficient Index-Based Snippet Generation (<https://ad-publications.informatik.uni-freiburg.de/TOIS_snippets_BC_2014.pdf>)
5. Chatgpt

**Evaluations:**

1. time spent on top K results (K as big as possible, 50, 100, 500..)
2. quality

**Ways to evaluate quality:**

1. questionnaire (>=20 respondents)
2. evaluate metrics (e.g. https://github.com/nju-websoft/BANDAR)

**Coding:**

Based on Assignment 2&3, using C++