

Advanced Programming Techniques: Spring 2023

CMPN306

Search Engine Project

Team 1: Thursday group

Hazem Montasser	2200003
Mahmoud Khaled Mahmoud	1190141
Ahmed Mohamed Saad	1190184

Crawler algorithms

- Breadth first search with Threading
- URI normalization & CompactString Concept

Indexer algorithms

- Word Stemming(PortStemmer algorithm)
- Jsoup(to Download HTML Content)

Ranker algorithms

- TF-IDF
- PageRank
- Cosine-Similarity
- Harmonic mean