

Data processing

Text processing – analyzing, manipulating and generation of electronic text

- stop word removal
- lemmatization

Job recommendation techniques

Topic modeling: discover the abstract topic in a collection of documents

Discovers words and phrases and patterns and clusters word in group and similar expression that best characterize the documents

Natural Language processing (NLP)- LDA

Cosine similarity – measures similarities between vectors by calculating the angle between the two vectors list.

TF-IDF and word2vec

TF-IDF - used to evaluate the importance of a keyword in any content and assign the value to that keyword based on its frequency of appearance in the analyzed document

word2vec – converts text into numerical feature vectors

LSH - It is a technique that collects similar input items into the same “buckets” with high probability. This technique can be used for data clustering and nearest neighbor search.