* Crawler

HTTP request

Save, parse, extract, and store (in dictionaries ??) the data (which data ??)\*

Follow links and loop through

Create the database

* Data preprocess

Remove tags, punctuation, and stopwords

Stemming

* Vector space model

Document-Term matrix

Calculate TF, IDF, and TF-IDF

…

* Embeddings

Choose a model (Word2Vec, FastText, or BERT)

Train on our dataset OR use pretrained

…

* Boolean model

…

* Query

autocomplete/autosuggest

Spell checking

Preprocess

* Retrieval and presentation

Snippet Generation

* Feedback and personalization

Create profiles ??

* + GUI ??