

AI Notetaking Tool

S pomočjo obstoječih knjižnic želimo integrirati sledeče funkcionalnosti:

Keyword extraction – input PDF

Summary generation – input PDF, output kratka obnova teksta.

Prepoznavna strokovnih izrazov, ki jih podkrepimo z definicijami.

Testiranje na prosojnicah

KEYWORD EXTRACTION

BERT Embedding

<https://www.analyticsvidhya.com/blog/2022/03/keyword-extraction-methods-from-documents-in-nlp/>

Testiramo kodo za keyword extraction pri čemer uporabljamo BERT embeddingse. Potrebne so večje GPU enote. Iščemo alternativo s hitrejšim computational časom.

Alternative: ALBERT,

Uporabili bomo že pripravljene knjižnice, kjer je potrebno sprva izvoziti tekst nato šele obnoviti sprva pridobljen tekst.

Knjižnica za pridobitev teksta iz PDF-ja <https://github.com/py-pdf/pypdf>

Knjižnica, ki uporablja BERT za povzemanje besedila <https://github.com/dmmiller612/bert-extractive-summarizer>