Stockholms universitet

Building a RAG Chat Assistant for DSV

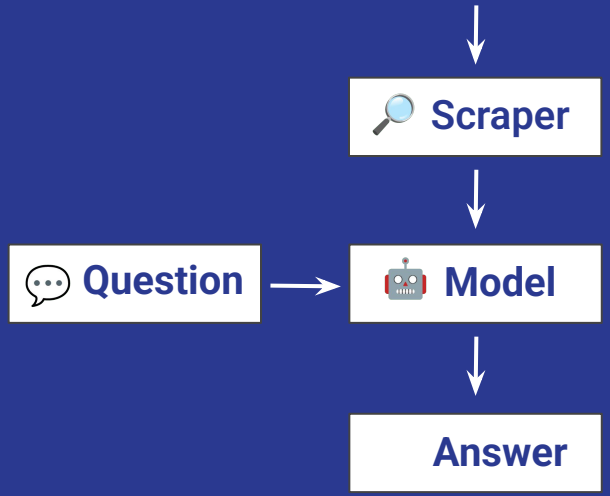
NLP Group Project Presentation - 3rd June 2025

Ippokratis Pantelidis
Fabian Sperber
Ramon Ventura Navarro
Lorenzo Vercilli

A Smarter Way to Ask Questions About DSV

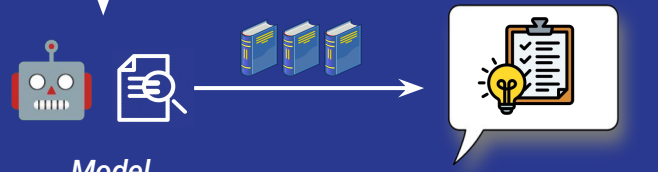
What

Stockholm University | Department of Computer and Systems Sciences



Why

Lots of information!

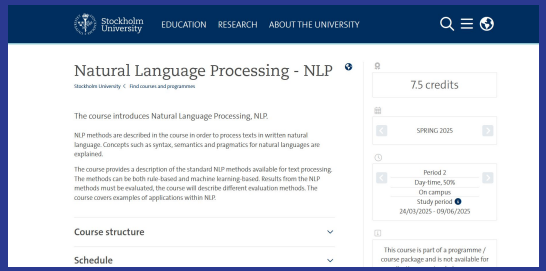


Model

Concise answers

Building the Dataset

- URL Scraping to gather information from the department websites (186 URLs explored)
- Extracted the text content from each URL and saved it in JSON format
- Used a pre-trained Sentence Transformer to generate the word embeddings (paraphrase-mpnet-base-v2 from the Hugging Face's Sentence Transformers library)
- The embeddings are then used to build the Retriever



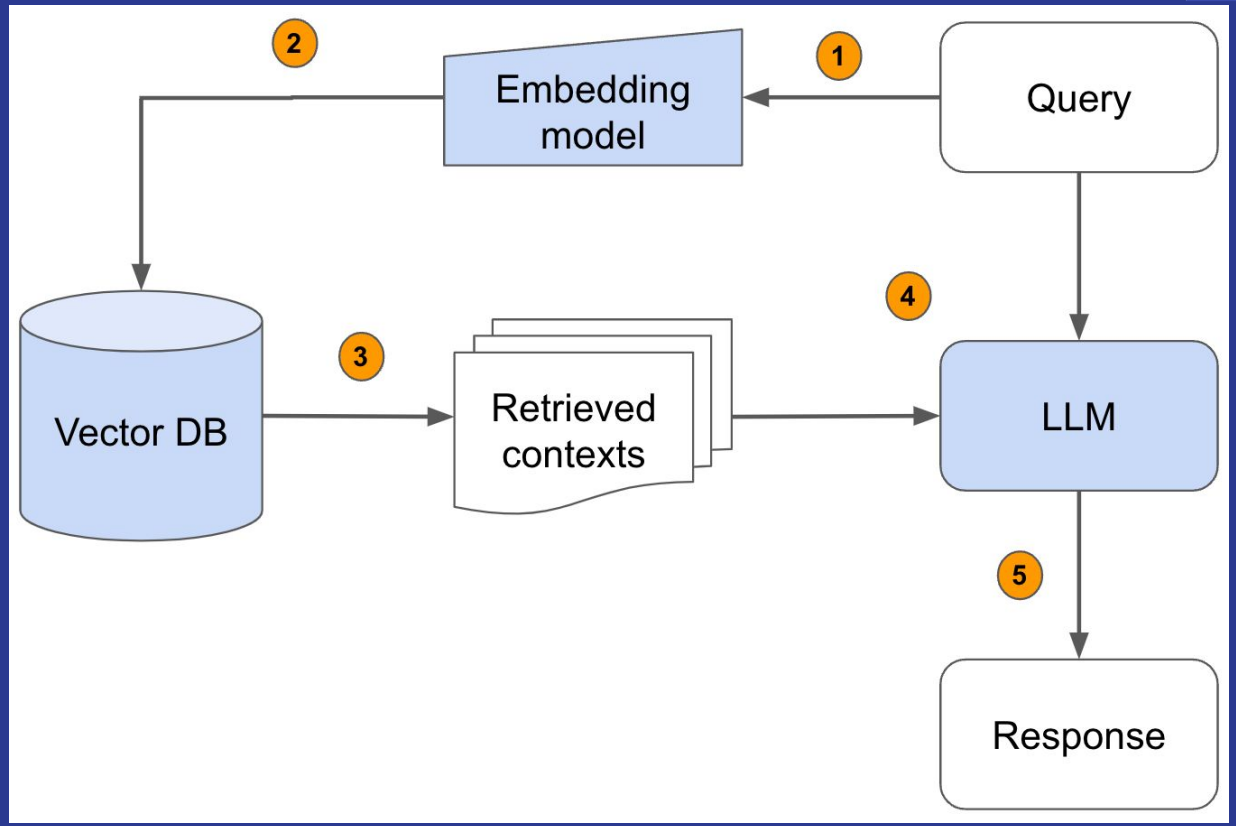
Scraping

{"url":"https://www.su.se/english/search-courses-and-programmes/ml273n-1.581408",
"content": "The course introduces Natural Language Processing, NLP...}

Chunking + Embedding

[0.08765432, 0.03456789, ..., 0.03456789, -0.06789012]

Model Architecture



Model Evaluation

Method

- Extrinsic: Manual sample queries used (top-down evaluation)
- Relevance and accuracy of responses assessed
- Focus on response comprehensibility and accuracy

Insight

- Works well with clearly formulated and specific questions
- Weaknesses with vague queries (too many hits)
- No hallucinating, but answers can get creative

Question 💬
How can I contact the DSV International Office?

→

Response
Response: You can contact the DSV International Office by sending an email to exchange@dsv.su.se (for partner universities), exchange-in@dsv.su.se (for incoming students), or exchange-out@dsv.su.se (for outgoing students). If you have questions about buddies, ...

Thank You

Do you have any questions?