

Data_Retrieval	
PK	<u>retrieval_id int NOT NULL</u>
	company_name char(50) NOT NULL
	extra_name char(50)
	all_headers char(2000)
	all_pdfs_texts char(30000)

Keyword_Extraction	
PK	<u>extraction_id int NOT NULL</u>
FK1	retrieval_id int NOT NULL
	extracted_keywords_from_headers(1000)
	extracted_keywords_from_pdfs(1000)

Use_retrieved_data

Use_extracted_keywords

Text_Generation	
PK	<u>generation_id int NOT NULL</u>
FK1	extraction_id int NOT NULL
	generated_mail char(1000)
	generated_newsletter char(1000)

- Giving a company name on the GUI kicks off the script to retrieve data of e.g. the first 10 search pages
- > The results of this retrieval is saved in a specific line of the Data_Retrieval entity

- If in addition there should be a keyword extraction with the retrieved texts, the data of a specific line from Data_Retrieval is used and the NLP-script for the keyword extraction is started
- > the result of this script is saved in a specific line of Keyword_Extraction

- This can be used for generating phishing-Texts
- > The data in a specific line of Keyword_Extraction is used to generate a phishing Mail oder a phishing Newsletter
- > These are then stored in a new entry in the entity Text_Generation