

# Project: Final Report

INFORMATION RETRIEVAL ON SUMMARIZED TEXT DOCUMENTS

ARYA DHORAJIYA

MEET MANSUKHBHAI KATRODIYA

# Section 1

## 1 What is the task, and why is it important to users?

### 1.1 Task:

- The core task involves implementing and evaluating the performance of ad-hoc information retrieval systems which are trained on summaries generated by three state-of-the-art sequence-to-sequence summarization models (Pegasus, Distilbart, Led-Large) of Wikipedia pages.
- Two information retrieval engines, BM25 and Query Likelihood Model, are trained on these generated summaries and tested on manually annotated sample queries to check if it's able to retrieve relevant original documents based on their summaries.
- Given a user query, the system's objective is to retrieve the original documents in a ranked order.

### 1.2 Importance to Users:

- This task is crucial for users who need to quickly access relevant information from extensive datasets.
- Summarization models aid in distilling vast amounts of data into manageable summaries, and the information retrieval engines facilitate the retrieval of pertinent information in response to user queries.
- Summaries can be used as snippets to show users essential content pertinent to their information need.
- This is especially valuable in scenarios where users are looking for swift information retrieval within a large corpus of information.

## 2 In general, what do users' queries look like?

- Users' queries are generally informational queries where they seek information on a specific topic, asking questions or looking for details about a particular topic.
- Users' queries are anticipated to vary widely, encompassing specific inquiries and broad topics. The system needs to handle and respond to this diversity effectively.

Some query examples are given below:

1. Quantum computing algorithms overview
2. Different types of cryptographic algorithms

3. Effects of quantum computing on cryptography
4. Evolution of Social Media Platforms

### 3 What kinds of results would be relevant to these queries? How many relevant results should there be per query?

#### 3.1 Kind of Relevant Results:

- Relevant results would include informative articles, definitions, or explanations related to the query topic.
- Results are ordered based on the document's relevancy score for a query.
- The relevant results consist of the original Wikipedia articles for a query. The system aims to provide users with information that directly addresses or offers insights into their queries.

#### 3.2 The number of Relevant Results:

While the exact number of relevant results can vary, a ranked list containing the top 5 to 10 most relevant summaries/documents is considered an effective approach. This ensures that users are presented with a manageable set of highly pertinent information.

### 4 If relevant to your project, how should the results be organized (ranked list, clusters, summaries, etc.)?

- Results are organized in a ranked list format.
- The organization prioritizes relevance, with the most relevant documents appearing at the top of the list. This structure enables users to quickly identify and access the information they need without going through irrelevant or less important content.

### 5 What evaluation metrics would be appropriate for this task?

#### 5.1 Mean Average Precision (MAP):

- This metric calculates the average precision across all queries, providing a nuanced evaluation of the system's overall performance.
- MAP offers a comprehensive evaluation by considering the precision at different points in the ranked list, giving more weight to higher-ranking relevant results.

- It is well-suited for this task where out of all retrieved results only a limited number of top results are considered, reflecting that only a few of the most relevant documents are suitable.
- We are using MAP@10 for evaluation.

## 5.2 Mean Reciprocal Rank (MRR):

- MRR emphasizes the rank of the first relevant result, indicating the system's efficiency in presenting highly relevant information early in the list.
- It highlights the importance of retrieving the most relevant information quickly, as it focuses on the rank of the first relevant result.
- We are using MRR@10 for evaluation.

## 5.3 Discounted Cumulative Gain (DCG):

- This matrix evaluates the ranking quality of the generated results, considering both relevance and the position of the document in the ranking.
- It highlights the importance of retrieving highly relevant documents are more useful than marginally relevant documents.
- We are using Normalized Discounted Cumulative Gain (NDCG@10) for evaluation.

# 6 A description of your implementation and an analysis of its performance.

## 6.1 Implementation:

### 6.1.1 Generating queries and manual annotations:

- Generated queries covering a variety of topics and added narratives for each query to annotate the documents with relevance scores.
- We annotated relevant and non-relevant Wikipedia articles for all queries on a scale of 0 – 4 with 0 being non-relevant and 4 being most relevant.

### 6.1.2 Fetching required documents and creating a corpus:

- Fetched 100 articles from Wikipedia that cover all the topics we decided on.
- Created a pipeline to pre-process all the fetched documents and a corpus of all documents.

### 6.1.3 Generating document summaries and creating a corpus of summarized documents:

- Created a pipeline to retrieve those documents and split them into chunks of data to fit each model's token size input limits.
- Created Summaries using State-Of-The-Art Seq2Seq models like DistilBart, Pegasus, and Led Large.

#### 6.1.4 Creating and Evaluating BM25 retrieval model:

- Created a BM25 model to rank results for a query based on a corpus of summarized documents.
- Create three BM25 models, one for each summarization model to evaluate performance.
- Evaluated the performance of models using MAP@10, MRR@10, and NDCG@10.

#### 6.1.5 Creating and Evaluating Query Likelihood retrieval model:

- Implemented a Query Likelihood Model from scratch to rank results for a query based on a corpus of summarized documents.
- Create three QLM models, one for each summarization model to evaluate performance.
- Evaluated the performance of models using MAP@10, MRR@10, and NDCG@10.

#### 6.1.6 Tuning hyperparameters for BM25 and Query Likelihood models:

- Created a set of values for  $k_1$  and  $b$  for BM25, and tuned the models based on a combination of these values.
- Tuned the Query Likelihood models based on different values of  $\mu$  for Dirichlet smoothing.

## 6.2 Performance Analysis:

1. **Model Comparison:** Evaluated the summarization models' performance in generating informative and concise summaries.
2. **Engine Performance:** Assessed the information retrieval engines' effectiveness in retrieving relevant results based on the trained models.
3. **Evaluation Metrics:** Analyzed MAP, MRR, and DCG scores to gain insights into the overall performance of the system.

## Size of Summarized Corporuses:

- Average number of words in LED Large corpus: 366.32 words
- Average number of words in Pegasus corpus: 128.09 words
- Average number of words in DistilBART corpus: 3085.03 words

## BM25 Model Performance:

Models	MAP	MRR	NDCG
BM25-DistilBart (k1 = 1.5, b= 0.75)	0.8946	0.9800	<b>0.8469</b>
BM25-DistilBart (k1 = 1.4, b = 0.55)	0.9019	0.9800	0.8466
BM25-Pegasus (k1 = 1.5, b = 0.75)	0.8049	0.8667	0.7068
BM25-Pegasus (k1 = 1.4, b = 0.1)	0.8081	0.8600	0.7067
BM25-Led_Large (k1 = 1.5, b = 0.75)	0.8996	0.9600	0.8105
BM25-Led_Large (k1 = 1.4, b = 0.5)	<b>0.9231</b>	<b>0.9800</b>	0.8108

### Query Likelihood Model Performance:

Models	MAP	MRR	NDCG
QLM-DistilBart (mu = 2000)	<b>0.9267</b>	<b>0.9800</b>	<b>0.8626</b>
QLM-DistilBart (mu = 1700)	0.9267	0.9800	0.8623
QLM-Pegasus (mu = 2000)	0.7704	0.8267	0.6976
QLM-Pegasus (mu = 500)	0.7867	0.8333	0.6995
QLM-Led_Large (mu = 2000)	0.8979	0.9800	0.8004
QLM-Led_Large (mu = 700)	0.9084	0.9800	0.8135

## 7 What milestones did you achieve in your grade contract?

We have achieved all the milestones in our project based on the outlined grade contract. Here are the milestones we completed successfully:

### Milestones for Grade B (Completed):

Annotated 25 queries, scraped and preprocessed documents, generated summaries with a transformer model (DistilBart), integrated BM25 for document retrieval, and evaluated model performance using Mean Average Precision, demonstrating a robust query-based document analysis and retrieval pipeline.

#### **Milestones for Grade B+ (Completed):**

Integrated an additional summarization model (Pegasus) to generate document summaries, conducted retrieval on the new corpus, and rigorously evaluated and compared overall model performance between the baseline and the additional summarization model. This expanded approach enhances the analysis of document retrieval effectiveness.

#### **Milestones for Grade A- (Completed):**

Integrated a third summarization model (Led\_Large) for document summaries, executed retrieval on the updated corpus, and conducted a thorough evaluation and comparison of overall model performance against the baseline and the additional summarization model. Additionally, fine-tuned BM25 hyperparameters to enhance retrieval efficacy, providing a comprehensive refinement to the document analysis and retrieval system.

#### **Milestones for Grade A (Completed):**

- 1. Generated summaries of documents with three summarization models:**  
We created a corpus of summarized documents from our original dataset (Scraped from Wikipedia) with three models: Led Large, Distilbart, and Pegasus.
- 2. Incorporated another retrieval model alongside BM25:**  
We successfully integrated Query Likelihood retrieval model alongside BM25, enhancing the capabilities of our system. Implemented these models on corpus of summaries of documents.
- 3. Tuned new retrieval model's hyperparameters:**  
We fine-tuned hyperparameters for both given retrieval models, optimizing their performance.
- 4. Compared the overall performance of all implemented models:**  
We conducted a comparison through evaluation metrics like MAP@10, MRR@10, and NDCG@10 to consider the overall performance of all implemented models, providing a holistic assessment.
- 5. Analyzed the strengths and drawbacks of the project:**  
We presented a detailed analysis of the strengths and drawbacks of the project, showcasing our understanding of the system's performance and limitations.

#### **7.1 Additional Achievements:**

##### **Successfully addressed potential barriers:**

1. Integrated transformer-based summarization models with BM25 and additional summarization models.
2. Tuned hyperparameters for BM25 and additional retrieval models, finding an optimal balance between time complexity and improved performance.
3. Managed computational resources efficiently, considering the time cost of generating summaries and training Information Retrieval Models.

## 7.2 Overall Assessment:

We demonstrated a systematic and progressive approach to the project, effectively overcoming potential barriers and consistently meeting or exceeding the outlined milestones. The completed milestones reflect our commendable level of proficiency in implementing, tuning, and evaluating complex models for information retrieval and summarization. Our detailed analysis of strengths and drawbacks further indicates a comprehensive understanding of the project's outcomes. We are proud of the successful completion of our project!

## 8 For group projects, what did each team member contribute?

For generating manual annotations, we both created a set of suitable queries along with narratives. We then exchanged the queries, annotated documents for each query by searching Wikipedia articles, and added a relevance score on a scale of 0 to 4 for each document.

### 8.1 Arya's Contributions:

Arya played a crucial role in the summarization phase of the project. He employed three advanced summarization models (DistilBart, Pegasus and Led Large) to generate concise and informative summaries for the collection of documents. Arya also took responsibility for annotating 30 percent of the documents, ensuring that the system had relevant and labeled data for training and evaluation. Additionally, he implemented a function for Mean Reciprocal Rank (MRR), contributing to the evaluation metrics of the project, and tuned the hyperparameters for BM25.

### 8.2 Meet's Contributions:

Meet focused on the information retrieval aspect of the project. He worked extensively on the BM25 and Query Likelihood Estimation (QLE) models, contributing to the document retrieval process. Meet's efforts included annotating a substantial 70 percent of the documents, providing a comprehensive dataset for training and testing the retrieval models. He also took charge of encapsulating functions and ensuring a proper paradigm in the codebase. Furthermore, Meet created functions for Mean Average Precision (MAP) and Normalized



Discounted Cumulative Gain (NDCG), contributing to the overall evaluation framework of the project and tuned the hyperparameter for Query Likelihood model.

Both team members collaborated effectively to cover both the summarization and information retrieval components of the project. Arya's work ensured the availability of quality summaries, while Meet's contributions were pivotal in developing and evaluating robust information retrieval models. Their combined efforts led to the successful integration of summarization and retrieval components, ultimately achieving the project goals.

## 9 Strengths and Drawbacks:

### 9.1 Strengths:

- Summarized documents streamline information retrieval, reducing search time and enhancing system efficiency.
- Users benefit from concise, focused results, minimizing information overload and improving understanding.
- Retrieval on summarized documents is resource-efficient, suitable for scenarios with limited bandwidth or storage.

### 9.2 Drawbacks:

- Summarized documents may omit nuanced details, potentially impacting the depth of information.
- System effectiveness relies on the quality of underlying summarization algorithms.
- Summarized documents may struggle with ambiguous, detail-specific queries or those requiring context.

## 10 Conclusion:

The experimentation with different Information Retrieval Systems has revealed intriguing insights into their performance. Notably, a hybrid approach utilizing the BM25 algorithm alongside the Led\_Large Summarization model yielded a commendable Mean Average Precision (MAP) of **0.9231**. This combination surpassed all other State-of-the-Art (SOTA) models examined in our project.

Equally impressive was the synergy between the QLM algorithm and the DistilBart Summarization model, which emerged as the top-performing model in our comprehensive study. This combination achieved an outstanding MAP of **0.9267**, marking it as the most effective model in our project.

It's noteworthy that despite variations in document size among the summaries—with DistilBart presenting the largest size—all models showcased commendable results. This outcome underscores the robustness and versatility of the selected models, reflecting a significant achievement in our research endeavors.

GitHub link: [https://github.com/arya1234/CS\\_6200](https://github.com/arya1234/CS_6200)

## Section 2

### **Queries, Narratives, and Annotated Documents with relevance score:**

Documents are represented as the title of the corresponding Wikipedia page.

#### 1. Quantum computing algorithms overview

- Narrative : Relevant documents must contain information about various algorithms used in quantum computing. These algorithms can run on a quantum computer. Documents on that do not mention any of the quantum algorithms like Deutsch–Jozsa algorithm, Bernstein-Vazirani algorithm, Simon’s algorithm, Quantum phase estimation algorithm, Shor’s algorithm, Grover’s algorithm, etc. should be considered irrelevant.
- Relevant Judgements – Relevance Score:
  - Quantum algorithm - 4
  - Algorithm - 0
  - Shor’s algorithm - 3
  - Grover’s algorithm - 3
  - Deutsch–Jozsa algorithm - 3
  - Quantum computing - 2
  - Quantum mechanics – 0

#### 2. Shor’s algorithm in quantum computing

- Narrative: The user is interested in gaining a comprehensive understanding of Shor’s algorithm, a key component of quantum computing. The focus should be on the principles behind Shor’s algorithm, its applications, and implications in the field of quantum computing. Relevant information may include its significance in integer factorization, its potential impact on cryptography, and any recent advancements or variations.
- Relevant Judgements – Relevance Score:
  - Shor’s algorithm - 4
  - Quantum algorithm - 3
  - Quantum computing - 2
  - Quantum mechanics - 0

- Quantum cryptography - 0
- Peter Shore - 2
- Quantum algorithm - 3
- Grover's algorithm – 1

### 3. Symmetric Key Cryptography

- Narrative: Relevant documents should include comprehensive information on symmetric key cryptography. The focus should encompass the principles behind symmetric key algorithms, their applications, and potential vulnerabilities. Relevant details may include popular symmetric key algorithms, key management strategies, and real-world use cases.
- Relevant Judgements – Relevance Score:
  - Symmetric-key algorithm - 4
  - Advanced Encryption Algorithm - 4
  - Data Encryption Standard - 4
  - Cryptography - 2
  - Public-key cryptography - 1
  - Quantum cryptography - 3
  - Quantum algorithm - 0
  - Quantum computing - 2

### 4. Different types of cryptographic algorithms

- Narrative: The user is interested in gaining knowledge about various types of cryptographic algorithms. The results should encompass a range of cryptographic algorithms, including symmetric and asymmetric key algorithms, hash functions, and digital signatures. Relevant documents should cover the principles, applications, and strengths of each type of algorithm.
- Relevant Judgements – Relevance Score:
  - Cryptography - 4
  - Quantum cryptography - 1
  - Symmetric-key algorithm - 4
  - Public-key cryptography - 4
  - Cryptanalysis - 2
  - Hash function - 4
  - Quantum algorithm - 0
  - Quantum computing – 0

### 5. Effects of quantum computing on cryptography

- Narrative: Relevant documents should include the effects of quantum computing on traditional cryptographic methods. The narrative should cover the potential vulnerabilities introduced by quantum algorithms, such as Shor's algorithm, to commonly used encryption schemes. Additionally, documents may show how to withstand quantum attacks.
- Relevant Judgements – Relevance Score:

- Quantum algorithm - 0
- Quantum computing - 2
- Quantum cryptography - 4
- Cryptography - 3
- Symmetric-key algorithm - 0
- Public-key cryptography - 0
- Hash function - 0
- Cryptanalysis - 2
- Post-quantum cryptography - 4
- Quantum key distribution – 4
- Shor’s algorithm – 3

#### 6. Cars 2011 Movie

- Narrative: The user is specifically searching for information about the 2011 installment in the Cars movie franchise, exploring details about the characters, plot, and unique elements introduced in this particular release.
- Relevant Judgements – Relevance Score:
  - Cars 2 - 4
  - Cars (film) - 2
  - Jaguar Cars - 0
  - Cars (franchise) - 3
  - Jaguar Cars - 0
  - Cars - 0
  - Cars 3 – 2

#### 7. Solar power efficiency

- Narrative: The user is keen on delving into the intricacies of solar power efficiency, seeking in-depth insights into the latest technological advancements, breakthroughs, and emerging trends in maximizing the conversion of sunlight into electricity.
- Relevant Judgements – Relevance Score:
  - Solar Cell Efficiency - 4
  - Shockley Queisser Limit - 2
  - Power (social and political) - 0
  - Solar cell - 3
  - Solar eclipse - 0
  - Solar energy - 2
  - Solar power - 3
  - Power (Physics) – 0

#### 8. Energy storage solutions

- Narrative: The user is actively seeking comprehensive information on modern energy storage solutions, with a particular interest in understanding the

efficiency, sustainability, and diverse applications of advanced technologies such as batteries, pumped hydro storage, and other innovative methods.

- Relevant Judgements – Relevance Score:

- Energy - 0
- Energy storage - 3
- Grid energy storage - 4
- Storage room - 0
- Energy storage as a service - 1
- Solution (Chemistry) - 0
- Storage room - 0
- Thermal energy storage – 3

#### 9. Lists of Presidents in US

- Narrative: The user is looking for a straightforward and easily accessible list of U.S. presidents, intending to gain a quick overview of the individuals who have held the Presidential Position in the country, including their respective terms in office.

- Relevant Judgements – Relevance Score:

- List of presidents of the United States - 4
- President (corporate title) - 0
- List of Presidents of the United States by education - 1
- List of states and territories of the United States - 0
- List of Presidents of the United States by Home State - 1
- United States - 0
- List of Presidents of the United States by Other Offices Held – 3

#### 10. First American President

- Narrative: The user is interested in a detailed exploration of the first American president, George Washington, including insights into his leadership during pivotal moments such as the Revolutionary War, and the enduring impact of his presidency on the foundations of the United States.

- Relevant Judgements – Relevance Score:

- List of presidents of the United States - 2
- George Washington - 4
- United States - 0
- List of Presidents of the United States Presidential firsts - 3
- First President - 0
- Citizenship of the United States - 0
- Franklin D Roosevelt – 2

#### 11. Impact of T20 cricket on other formats

- Narrative: The user is interested in gaining a comprehensive understanding of the impact of T20 cricket on other formats of the game. The focus should be on the principles behind T20 cricket, its history, and implications in the field of cricket.

Relevant information may include its significance in the game, its potential impact on other formats, and any recent advancements or variations. Documents that do not mention anything about T20 cricket should be considered irrelevant.

- Relevant Judgements – Relevance Score:
  - Twenty20 - 4
  - Cricket - 3
  - Cricket (insect) - 0
  - Forms of Cricket - 2
  - Limited overs cricket - 1

## 12. History of cricket

- Narrative: The user is interested in gaining a comprehensive understanding of the history of cricket. Relevant information should cover the inception of the game, iconic moments, legendary players, and the evolution of the game over the years. Documents that do not include any information about the game of cricket should be considered irrelevant.
- Relevant Judgements – Relevance Score:
  - Twenty20 - 0
  - FIFA World Cup - 4
  - Stadium - 0
  - List of FIFA World Cup finals - 3
  - History of the FIFA World Cup - 4
  - Association football – 2

## 13. History of Football World Cup

- Narrative: Explore documents providing a detailed historical account of the Football World Cup. Relevant information should cover the inception of the tournament, iconic moments, legendary players, and the evolution of the World Cup over the years. Documents that do not include any information about the football World Cup should be considered irrelevant.
- Relevant Judgements – Relevance Score:
  - Laws of the Game (association football) - 4
  - American football rules - 4
  - Association football - 2
  - History of the FIFA World Cup - 0
  - Regulation of sport – 1

## 14. Rules and regulations of football

- Narrative: The user is looking to find documents explaining the basic rules and regulations of football. Relevant information should cover aspects such as scoring, fouls, and offside rules. Documents that just contain information about the football and do not mention anything about the rules and regulations should be considered irrelevant.
- Relevant Judgements – Relevance Score:

- Laws of the Game (association football) - 4
- American football rules - 4
- Association football - 2
- History of the FIFA World Cup - 0
- Regulation of sport – 1

#### 15. Basics of Kinetic Energy

- Narrative: The user is interested in gaining a comprehensive understanding of the basics of kinetic energy. Relevant information should cover the foundational aspects of kinetic energy, understanding its definition, calculation, and its role in physics. Documents that do not include any information about kinetic energy should be considered irrelevant.
- Relevant Judgements – Relevance Score:
  - Energy – 1
  - Kinetic Energy – 4
  - Classical mechanics – 3
  - Solar energy – 0
  - Energy storage – 0

#### 16. Kinetic Energy vs. Potential Energy

- Narrative: The user is interested in gaining a comprehensive understanding of the difference between kinetic energy and potential energy. Relevant information should cover the foundational aspects of kinetic energy and potential energy, understanding their definitions, calculations, and their role in physics. Documents that do not include any information about kinetic energy and potential energy should be considered irrelevant.
- Relevant Judgements – Relevance Score:
  - Energy – 1
  - Kinetic Energy – 4
  - Potential Energy – 4
  - Classical mechanics – 3
  - Solar energy – 0
  - Chemical energy – 0
  - Energy storage – 0

#### 17. History of Cinema

- Narrative: The user is interested in gaining a comprehensive understanding of the history of cinema. Relevant information should cover the inception of cinema, iconic moments, legendary actors, and the evolution of cinema over the years. Documents that do not include any information about cinema should be considered irrelevant.
- Relevant Judgements – Relevance Score:
  - History of film - 4
  - Visual arts - 0

- History of film technology - 3
- Filmmaking - 1
- Film - 2
- Cars (film) – 0

#### 18. Impact of Streaming Services on the Film Industry

- Narrative: The user is interested in gaining a information of the impact of streaming services on the film industry. Relevant documents should cover the principles behind streaming services, its evolution, and implications in the field of film industry. Relevant information may include its significance in the film industry, its potential impact on the film industry, and any recent advancements or variations. Documents that do not mention anything about streaming services and film industry should be considered irrelevant.
- Relevant Judgements – Relevance Score:
  - History of film – 0
  - Visual arts – 0
  - Streaming television – 3
  - Film – 2
  - Streaming media – 4

#### 19. Football vs Soccer

- Narrative: Find documents that explain the difference between football and soccer. Relevant information should cover aspects such as the history of the two sports, the naming, the rules and regulations, or the differences in the gameplay. Documents that do not contain information about football and soccer and do not mention anything about the differences between the two sports should be considered irrelevant.
- Relevant Judgements – Relevance Score:
  - Names for association football – 4
  - Association football – 4
  - Laws of the Game (association football) – 2
  - History of the FIFA World Cup – 0
  - History of cricket – 0

#### 20. Evolution of electric cars

- Narrative: The user is interested in gaining a comprehensive understanding of the evolution of electric cars. Relevant information should cover the inception of electric cars, the evolution of electric cars over the years, and the impact of electric cars on the environment. Documents that do not include any information about electric cars should be considered irrelevant.
- Relevant Judgements – Relevance Score:
  - Electric car – 4
  - Electric vehicle – 3
  - Battery electric vehicle – 4



- Car – 2
- Jaguar Cars – 0

## 21. Evolution of Social Media Platforms

- Narrative: Explore the chronological evolution of social media platforms, highlighting key milestones, innovations, and the impact of these platforms on communication. Relevant information should cover the inception of social media platforms, the evolution of social media platforms over the years, and the impact of social media platforms on communication. Documents that do not include any information about social media platforms should be considered irrelevant.
- Relevant Judgements – Relevance Score:
  - Electric car – 0
  - Social media – 4
  - Virtual community – 1
  - Content creation – 2
  - Social network – 3

## 22. Social media for marketing

- Narrative: The user is interested in gaining knowledge of the use of social media for marketing. Relevant information should cover the way in which social media is used for marketing, the impact of social media on marketing, and the evolution of social media marketing over the years.
- Relevant Judgements – Relevance Score:
  - Social media marketing – 4
  - Social media – 4
  - Virtual community – 0
  - Content creation – 1
  - Social network – 0
  - Digital marketing – 3

## 23. Basics of stock market

- Narrative: The user is looking for basic information such as fundamentals, understanding the stock market, and how it works. Relevant information should cover the basics of the stock market, including concepts like stocks, dividends, and market indices. Documents that do not include any information about the stock market should be considered irrelevant.
- Relevant Judgements – Relevance Score:
  - Stock market – 4
  - Stock – 4
  - Dividend – 3
  - Stock exchange – 2
  - Social network – 0

## 24. Importance of credit score

- Narrative: The user wants to know about the significance of credit scores, factors that influence them, and tips for maintaining a good credit standing. Relevant information should cover the basics of credit scores, including concepts like credit score, credit report, and credit history.
- Relevant Judgements – Relevance Score:
  - Credit score – 4
  - Credit history – 4
  - Credit rating – 2
  - Credit risk – 1
  - Social network – 0

## 25. Personal Finance Tips

- Narrative: The user wants to explore documents that discover practical tips for managing personal finances, budgeting, and building wealth. Relevant information should cover the basics of personal finance, including concepts like budgeting, saving, and investing.
- Relevant Judgements – Relevance Score:
  - Personal finance – 4
  - Budget – 2
  - Financial management – 2
  - Credit risk – 0
  - Social network – 0

## References:

1. Search Engines: Information Retrieval in Practice, Croft, Metzler and Strohman (Addison-Wesley)
2. <https://huggingface.co/google/pegasus-large>
3. <https://huggingface.co/sshleifer/distilbart-cnn-12-6>
4. <https://huggingface.co/pszemraj/led-large-book-summary>
5. [https://en.wikipedia.org/wiki/Main\\_Page](https://en.wikipedia.org/wiki/Main_Page)
6. <https://nlp.stanford.edu/IR-book/html/htmledition/okapi-bm25-a-non-binary-model-1.html>
7. <https://nlp.stanford.edu/IR-book/html/htmledition/using-query-likelihood-language-models-in-ir-1.html>
8. <https://www.kaggle.com/>