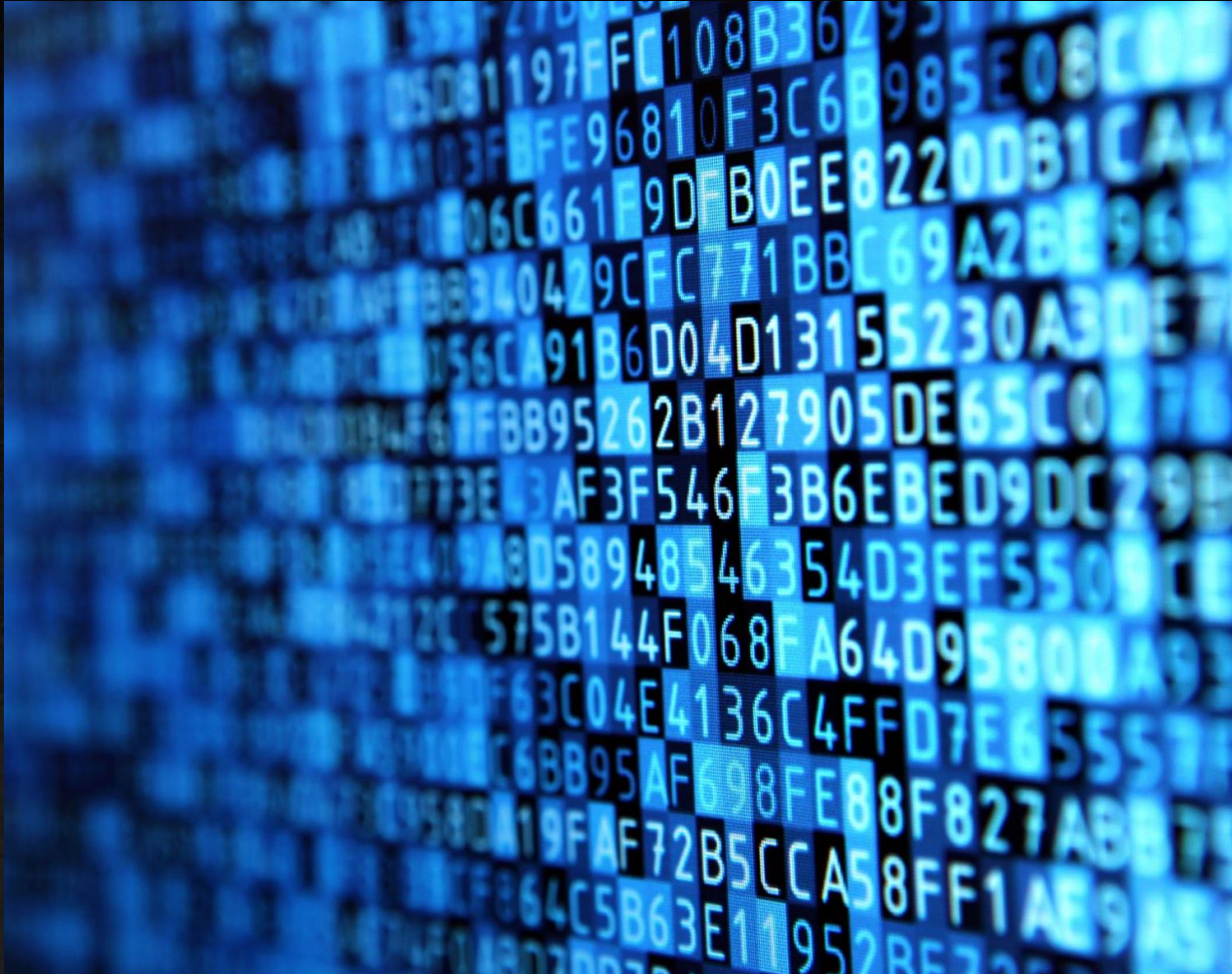


FAQ Maker for Products

Project Presentation



Team:

Madhura Mathkar
Graduate Student
School of Computing
University of Utah

Pranav Shah
Graduate Student
School of Computing
University of Utah

Sreeja Pillai
Graduate Student
School of Computing
University of Utah

Presentation Outline

1. Introduction
 - Problem Statement
 - Motivation
 - Project Scope
2. Dataset
3. Methodology
 - IR Model
 - Pre-Processing
 - Vector Space Model
 - Cosine
 - BM25
4. Evaluation and Applications
 - Results
 - Evaluation Measures
 - Applications
5. References

Introduction

- Customers may ask a few product relevant questions to seller or read FAQ's.
- Hence, it is the seller's responsibility to provide all the necessary information a customer might be looking for.
- Additionally, the seller might find it difficult to survey the online market for similar products and check all user asked questions/suggestions.
- The project aims at creating a model that generates questions related to any specific product for the seller.
- Hence, seller can use these to make product description more informative.
- The seller can also use these questions to get an idea about user's perspective over other products.

Problem Statement:

- Products may be difficult to use, hence a proper product description is necessary.
- Finding a solution to a question may be difficult when there is too much information
- Manufacturers/Sellers need to address such questions efficiently
- In the era of e-commerce, getting a market survey is not easy and comparing thousands of products to get user's perspective is extremely difficult.

Motivation:

- Understand the perspective of the Manufacturers/Sellers
- Manufacturers need to have insight to the user's perception of the product and other similar products in market.
- Reduce redundancy when there are many similar questions from users.
- Generate a set of questions so that users can find relevant information.

Project Scope:

- Help create a question bank for a particular product description using the amazon Q/A and description dataset.

Related Works:

- Every Q/A models involves using Q/A set.
- They suggest questions based on fixed answers.

Our Contribution:

- Providing questions based on real-world data.
- Focusing on questions more and product description than answers. To those questions.

Dataset

Amazon Product Dataset^[2] – It consists of product meta-data.

Amazon Review/Rating Dataset^[2] – It consists of product ratings, review and q/a for each product.

We used the following features from the Datasets:

- ASIN (Product id)
- Question
- Description
- Category
- Title
- Brand

We used Electronics dataset to build our model.



Methodology

- The developed IR model includes the following components:
 - Data Fetch and Pre-processing
 - Descriptions (Description, category, title, brand)
 - Question
 - Query
 - TF – IDF
 - Cosine Similarity and Ranking
 - BM25
- Evaluation of this IR model will be discussed in the later sections.

IR Model:

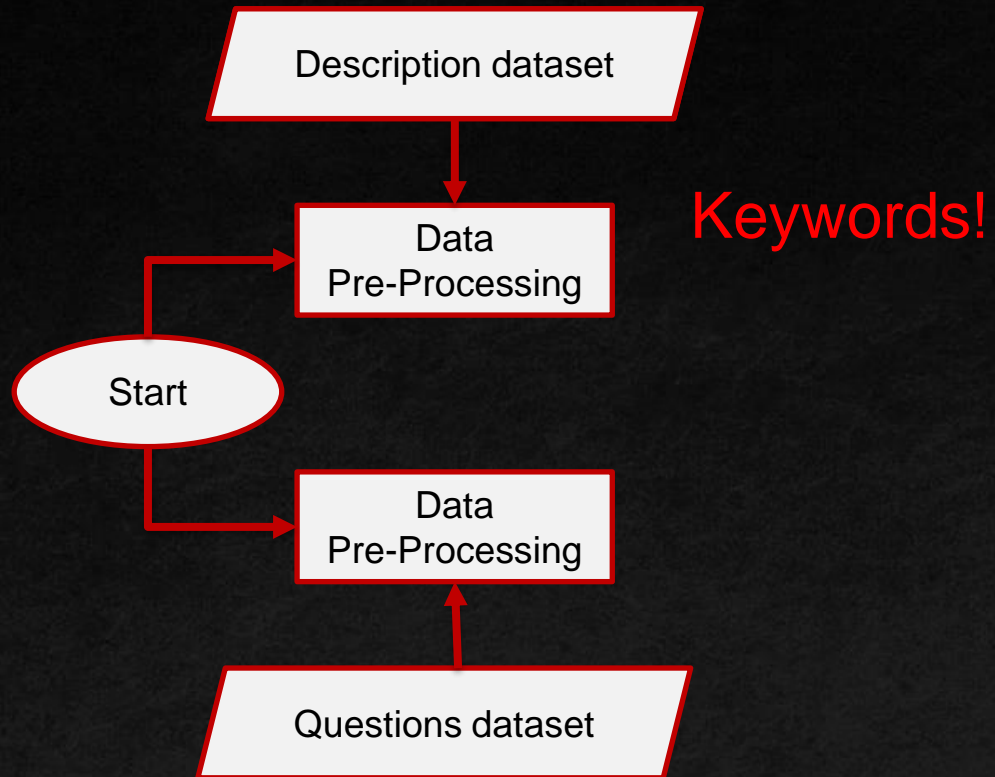


Figure 1: FAQ Maker model

IR Model:

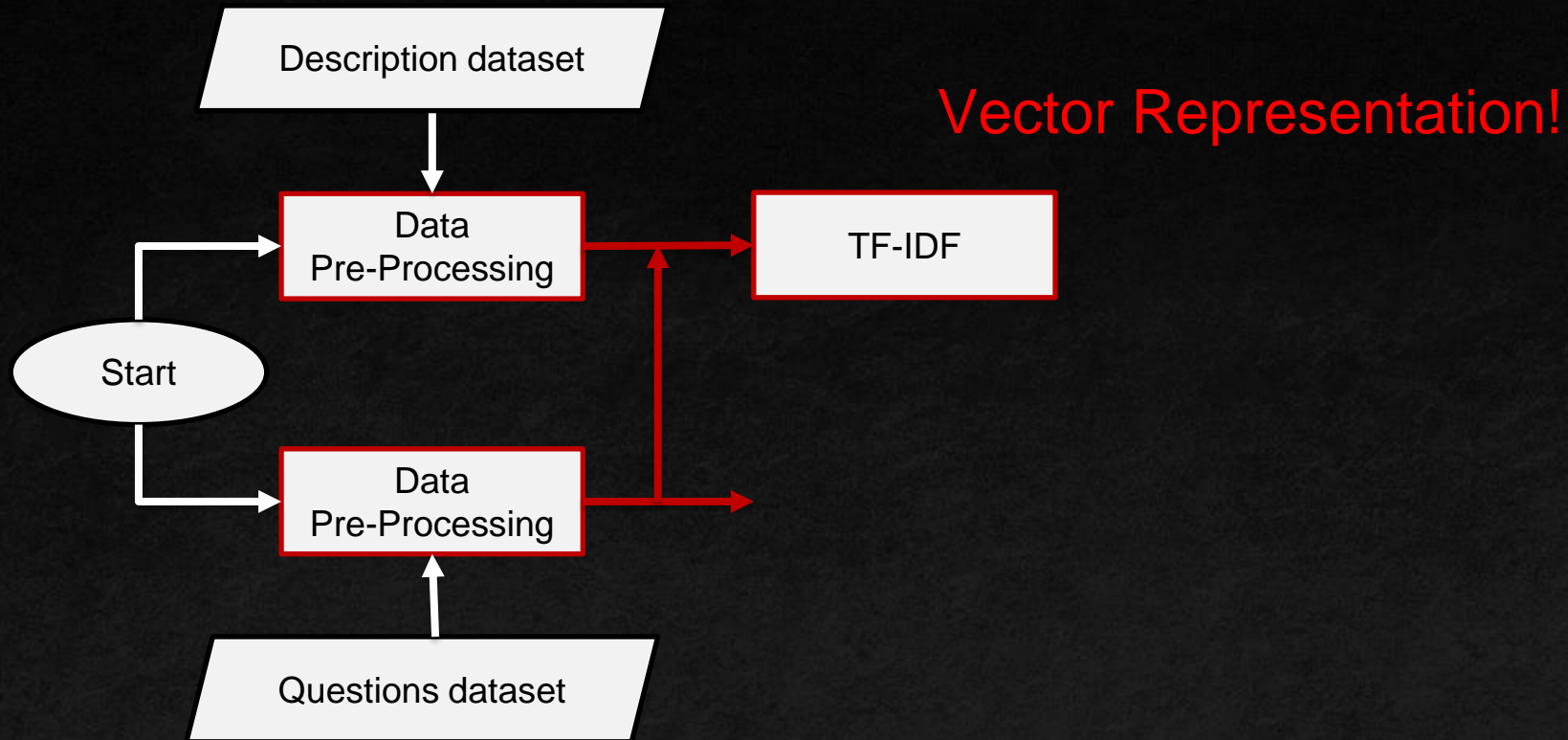


Figure 1: FAQ Maker model

IR Model:

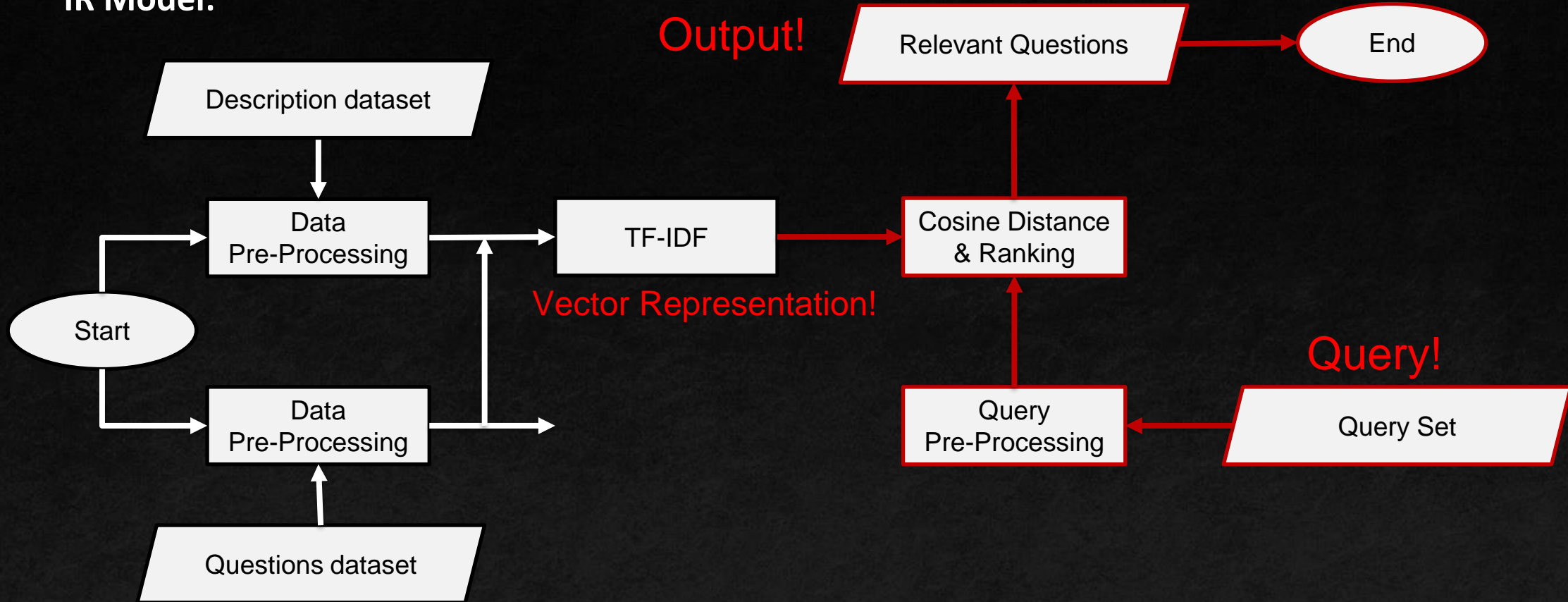


Figure 1: FAQ Maker model

IR Model:

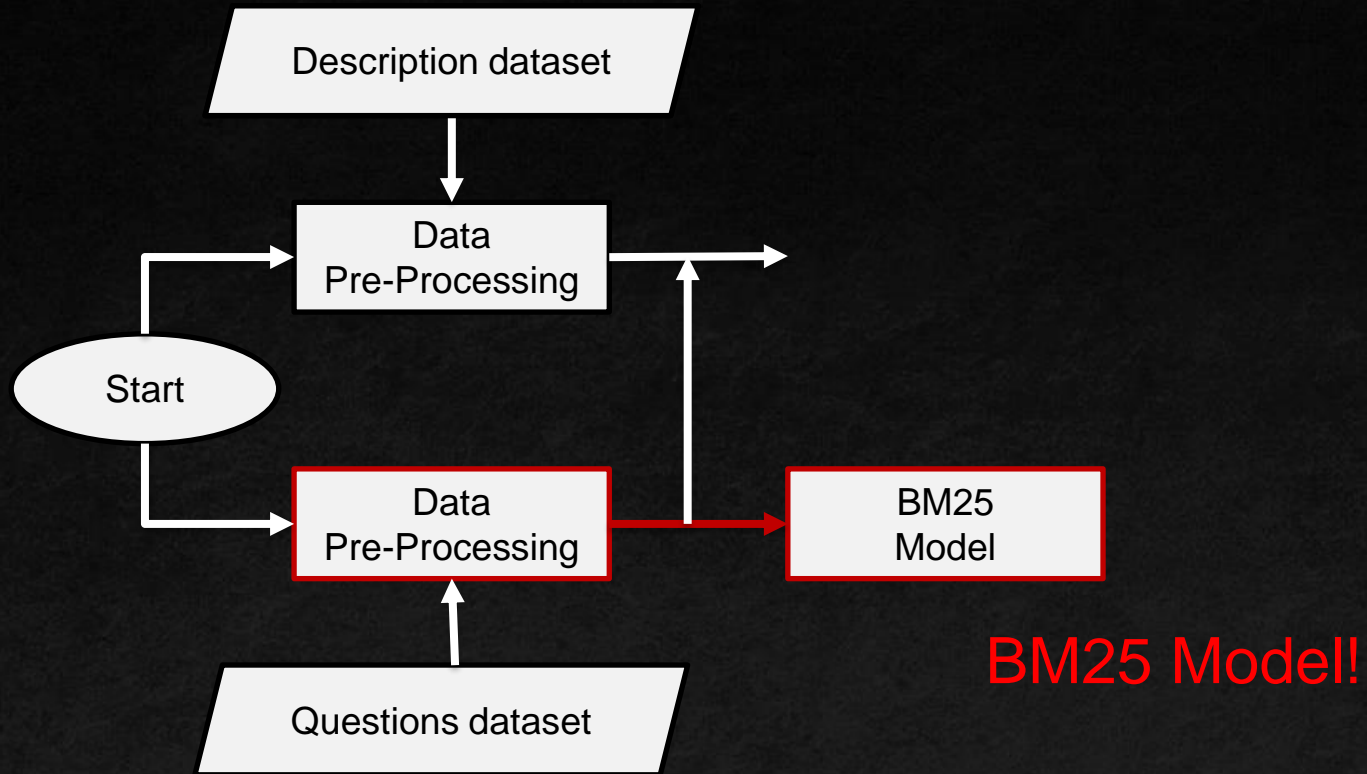


Figure 1: FAQ Maker model

IR Model:

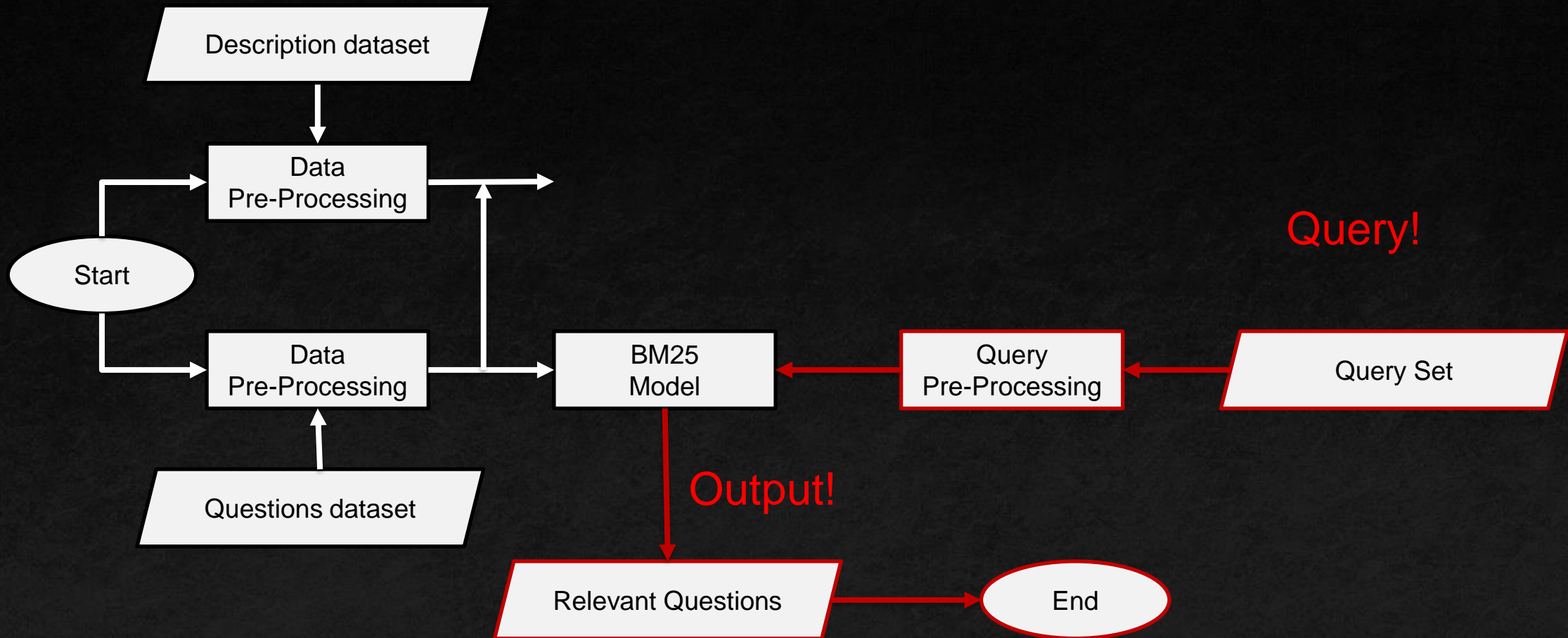


Figure 1: FAQ Maker model

IR Model:

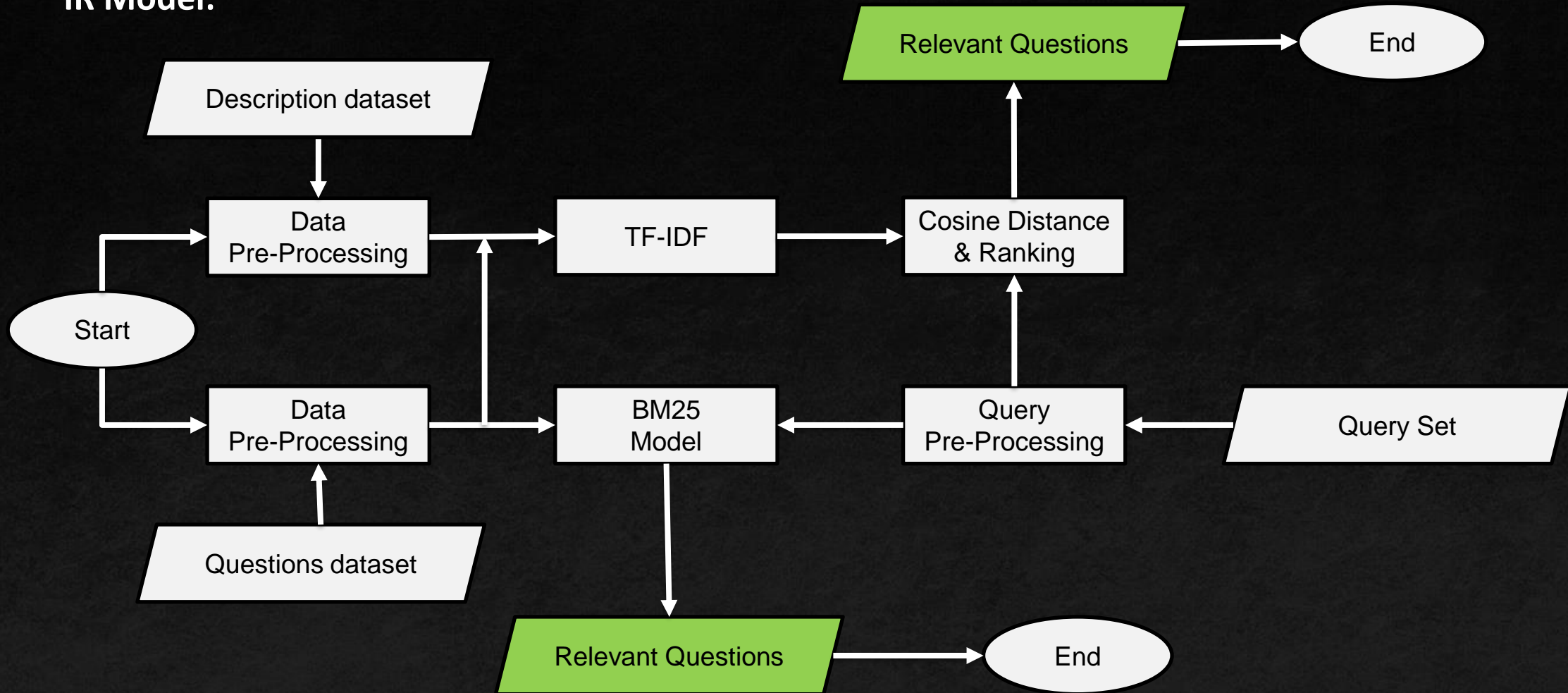


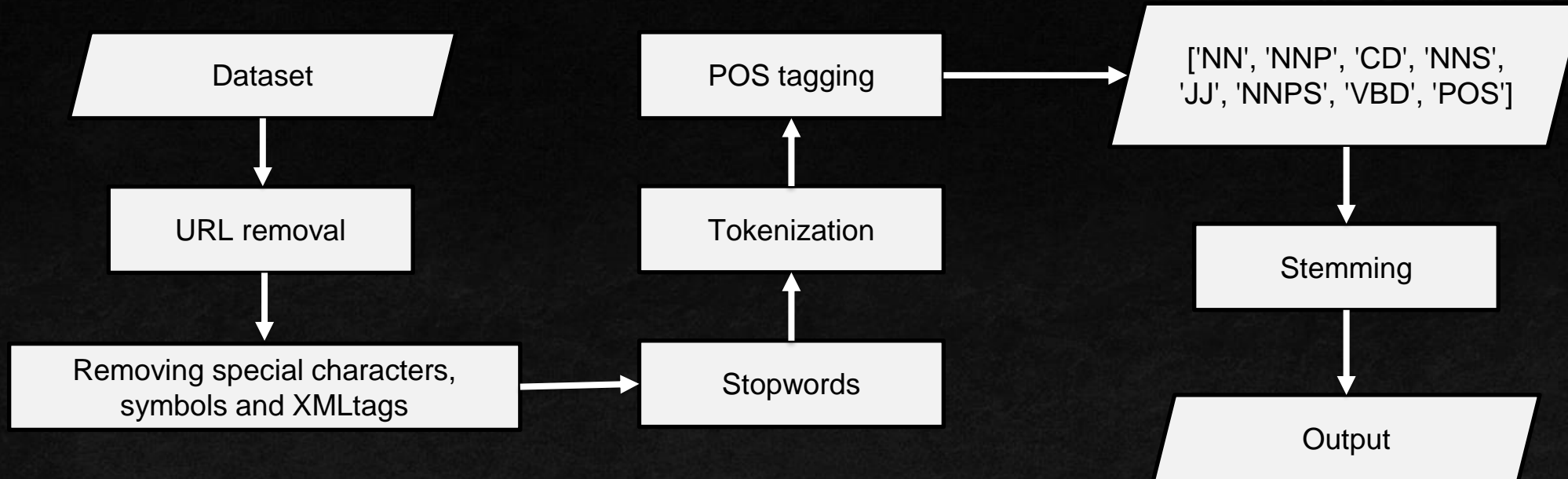
Figure 1: FAQ Maker model

Preprocessing

- Pre-processing means to bring the text into a form that is analyzable for our task^[1].
- Pre-processing also helps removes noise from the data.
- The steps in Pre-processing are:
 - Tokenization: Split a text into individual tokens
 - Stemming: Reducing inflected word to their base or root form.
 - Stop-words: Removing frequent unusual words, called stopwords.
 - Special characters and URL removal: Using Regex
 - POS tagging



Pre-processing:

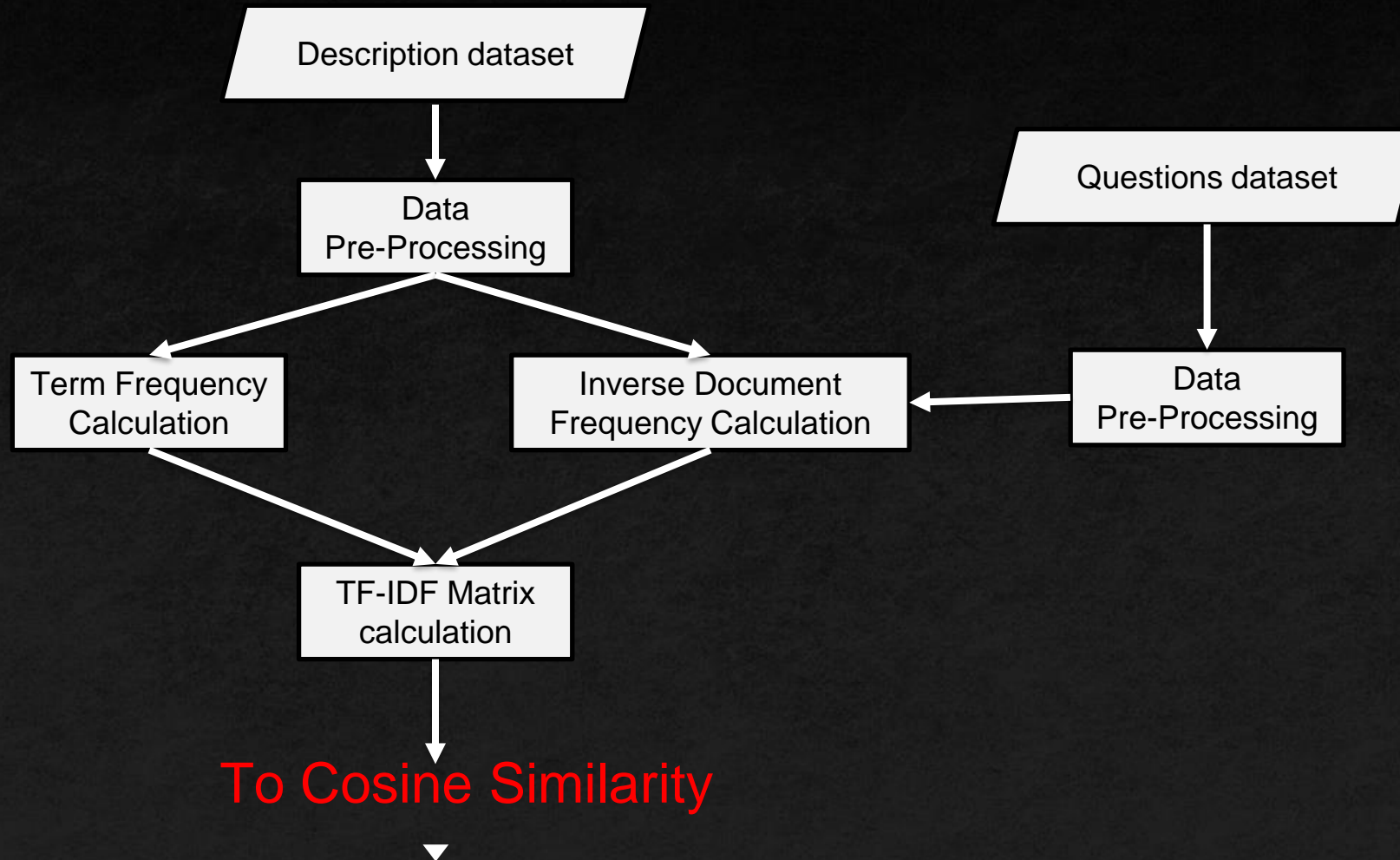


- Example input:

```
['CISCO 2509 CISCO Router', '<div style="text-align:center;width:100%;margin:22px 0;height:1px;border-top:1px solid #DDDDDD"></div> <center><div class="transparency-container aplus-content-container"> <a href="/b?node=12691228011"><h3></h3></a></center></div> <div style="text-align:center;width:100%;margin:22px 0;height:1px;border-top:1px solid #DDDDDD"></div>']
```

- Example Output: ['cisco', '2509', 'cisco', 'route']

Vector Space Model



Cosine

$$\text{Cosine Similarity} = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}}$$

- Where:
 - x: Query
 - y: A row from TF-IDF vector



BM25

BM25 Model:

- BM25 is probabilistic model that is developed by Stephen E. Robertson, Karen Sparck Jones, and others.
- BM stands for 'best match'.
- BM25 model doesn't use a single function, it use set of functions.

$$BM25 = \sum_{t \in q \cap d} \left(\frac{tf}{tf + k_1 \cdot nb} \cdot \log\left(\frac{N - dft + 0.5}{dft + 0.5}\right) \cdot qft \right)$$

- Where:
 - tf: Frequency of term occurrences
 - N: Total number of documents in the collection
 - dft: Number of documents containing the term
 - qtf : frequency of occurrences of a term t in the query,
 - k1: parameters influencing the frequency of terms that is adjusted to 1.2 by default
 - nb: Normalization factor = $(1-b) + b \cdot (tl/tlavg)$
 - tl: Number of terms in the document (document length)
 - tlavg: Average number of terms in the document

Results

- Query: 'Features 22nm Intel Celeron N2807 to deliver to the most intuitive and integrated operating systems in the world Supports 2.5" thickness 7.0/9.5mm Hard Drives (1 x 3Gbps SATA2) Ultra compact PC design - 0.69L(56.1x 107.6 x 114.4mm) 1x SO-DIMM DDR3L Slots (1333 MHz) Preinstall IEEE 802.11 b/g/n Wi-Fi / Bluetooth 4.0 Mini-PCIe card Supports dual displays via a VGA and a HDMI port Gigabit LAN Audio jack (Headphone/MIC) VESA mounting bracket (75 x 75mm + 100 x 100mm)'

Original	Cosine	BM25
Possible using 220v AC power?	I have PCIe 2.0 slot, will this card fit?	Can this shell be used with 18650 batteries (Dimensions: 2.56 x 0.71" / 6.5 x 1.8 cm (L x Dia))?
Will it run Linux (Debian 7, Ubuntu 14.04 or Mint 17)?	Is this Bluetooth 4.0?	can I use it on a TV with VESA 100 x 100?
Does this support 1600 MT/s 204-Pin DDR3 Notebook Memory?	Will this case fit my ASUS ROG G750JS (16.1 x 12.5 x 0.7 ~2.0)	I have PCIe 2.0 slot, will this card fit?
Does it come with power supply and cord?	Does this keyboard have USB 2.0 or 1.0 ports?	who's the exact length of the lens? the description says "82.0 x 100.0 x 3.2 inches," but obviously not correct. what is the length specifically?
does it have any fans?	can I use it on a TV with VESA 100 x 100?	Does this have a barrow size of 2.1 mm x 5.5 mm. (Just double checking one of the reviews that said 2.1 x 5.0)?
Is it possible to boot from a usb thumb drive instead of a SSD?	I found the same product on Home Depot and the dimensions say .5 x 1.5 x 1.25. Amazon says 1 x 4.75 x 9 inches. Which is correct?	Will this case fit my ASUS ROG G750JS (16.1 x 12.5 x 0.7 ~2.0)
Will this boot ChromeOS? ie: run as a Chromebox?	Will it work with OS X Yosemite?	I found the same product on Home Depot and the dimensions say .5 x 1.5 x 1.25. Amazon says 1 x 4.75 x 9 inches. Which is correct?
included power supply? 120/240 Vac at 50/60 Hz.	Can this shell be used with 18650 batteries (Dimensions: 2.56 x 0.71" / 6.5 x 1.8 cm (L x Dia))?	Does this keyboard have USB 2.0 or 1.0 ports?

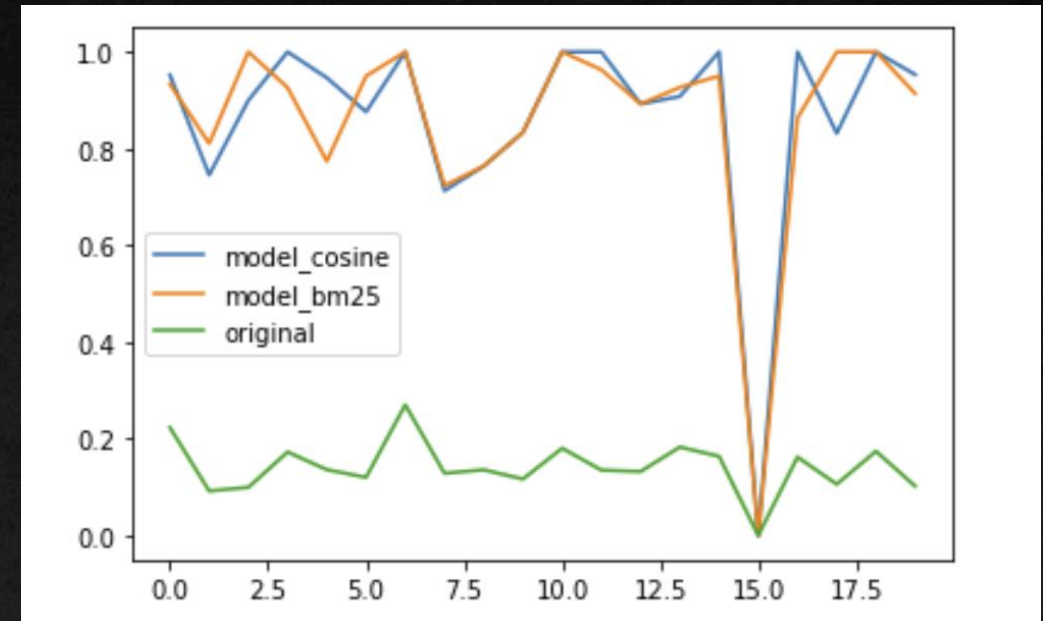
Evaluation

Evaluation:

- We trained the model with 28795 questions and 2610 unique descriptions.
- We tested the model for 20 unseen Queries from the dataset.

Evaluation Technique:

- We ran evaluation by calculating semantic scores.
- We calculated semantic similarity between:
 - Query and Cosine Generated Questions
 - Query and BM25 Generated Questions
 - Query and Original Questions
- The Scores for these evaluations are as follows:
 - Query – Cosine = 0.865760
 - Query – BM25 = 0.861166
 - Query – Original = 0.141647
- Similarity score between Generated and Original Questions:
 - Cosine: 0.251
 - BM25: 0.252



Applications

Applications:

- This model can be highly useful to manufacturers and sellers.
- Generates many Questions which are like the users' questions on similar products.
- Manufacturers can get an idea about user's perspective of different products in the market.
- Manufacturers can create their FAQ to be useful for marketing the product.
- It can also act as a question filter, which can help rank questions and show only relevant information.
- Relevant questions generated can be used to build an informative product description.



Future Scope

- Involving Deep Learning or Machine Learning techniques and algorithms to improvise the model.
- Combining more features from dataset to build a better model.
- Testing the model on bigger machines to include maximum training data.
- Using SPACY or NLTK tools to improvise the model and data pre-processing.



References:

1. <https://towardsdatascience.com/all-you-need-to-know-about-text-preprocessing-for-nlp-and-machine-learning-bc1c5765ff67>
2. <http://jmcauley.ucsd.edu/data/amazon/qa/>
3. A. Tamrakar and S. K. Vishwakarma, "Analysis of Probabilistic Model for Document Retrieval in Information Retrieval," 2015 International Conference on Computational Intelligence and Communication Networks (CICN), Jabalpur, 2015, pp. 760-765, doi: 10.1109/CICN.2015.155.
4. <https://pypi.org/project/rank-bm25/>
5. <https://www.nltk.org/howto/wordnet.html>
6. <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>
7. <https://www.tutorialexample.com/best-practice-to-extract-and-remove-urls-from-python-string-python-tutorial/>
8. <https://towardsdatascience.com/natural-language-processing-feature-engineering-using-tf-idf-e8b9d00e7e76>



Questions?

Thank You 😊