RESEARCH PRESENTATION

**PRESENTED BY-
ANUPREET SINGH
11911063
MENTOR-
DR. MUKESH MANN**

TOPIC-INFORMATION RETRIEVAL METHODS:A LITERATURE REVIEW

HOW WE START

This research began by learning about how the web works and what is the basic infrastructure of a web search engine
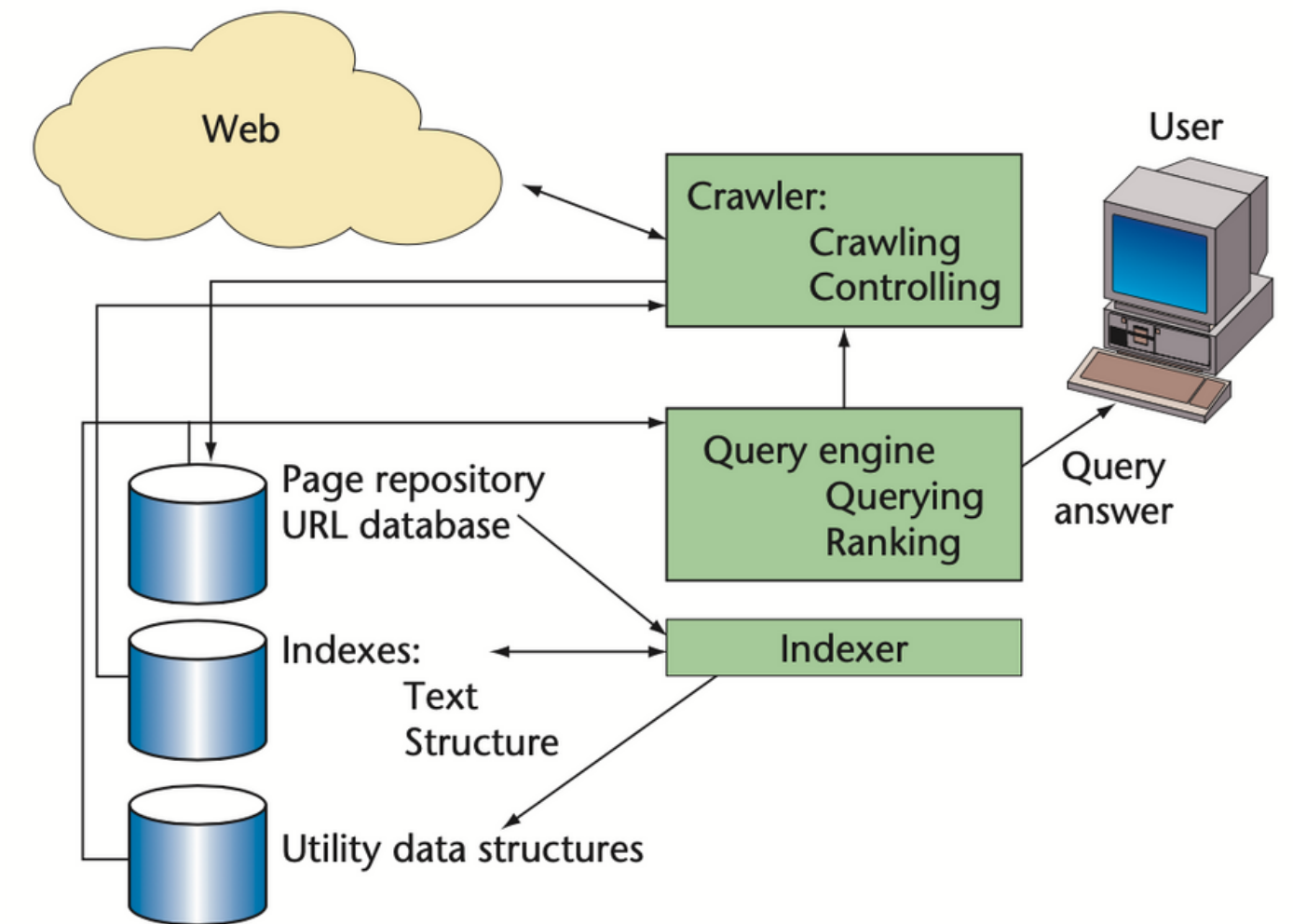

Figure 1. Architecture of a search engine. The modules are the crawler, query engine, and indexer; the data includes the page repository, URL databases, indexes, and utility data structures.

INDEXING

Indexing is the process by which search engines organise information before a search to enable super-fast responses to queries.

INVERTED INDEXING

Forward Index

Inverted Index

Search engines use inverted indexing as a system wherein a database of text elements is compiled along with pointers to the documents which contain those elements.
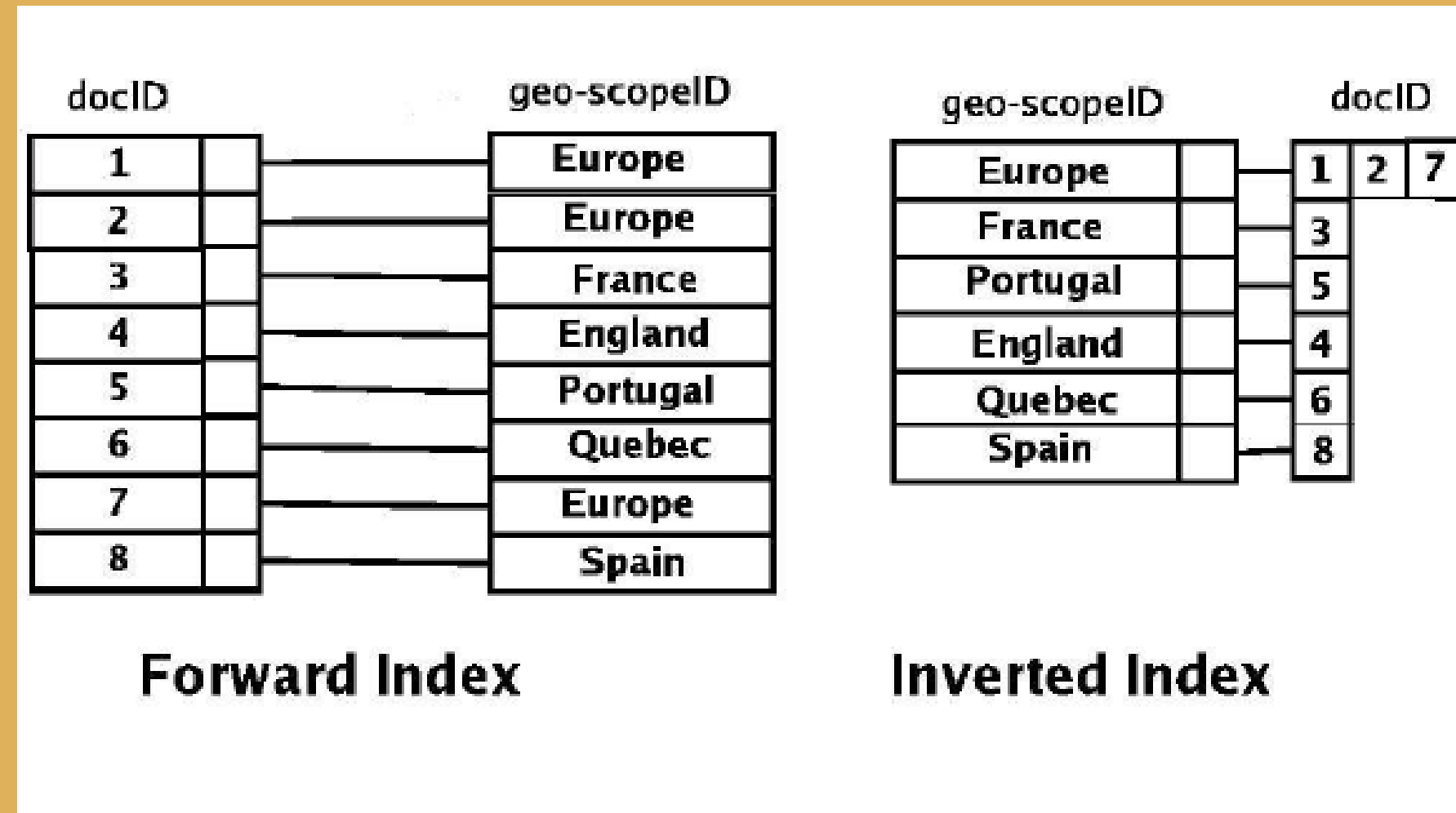
PAGE RANK ALGORITHM

- Page Rank Algorithm was google's first and most famous query result ranking algorithm

- It ranks the results by measuring the importance of website pages

- PageRank works by counting the number and quality of links to a page to determine a rough estimate of how important the webpage is.

$$PageRank\ of\ site = \sum \frac{PageRank\ of\ inbound\ link}{Number\ of\ links\ on\ that\ page}$$

OR

$$PR(u) = (1 - d) + d \times \sum \frac{PR(v)}{N(v)}$$

OTHER PROMINENT RANKING ALGORITHMS

1. Weighted Page Rank
2. Hyperlink Induced Topic Search

IR MODELS

- BOOLEAN MODEL
- VECTOR SPACE MODEL
- PROBABILISTIC MODEL
- INFERENCE NETWORK MODEL

GOOGLE'S RECENT RANKING ALGORITHMS EXAMPLES

- Panda(2011)

- Penguin(2012)

- Humming bird(2013)
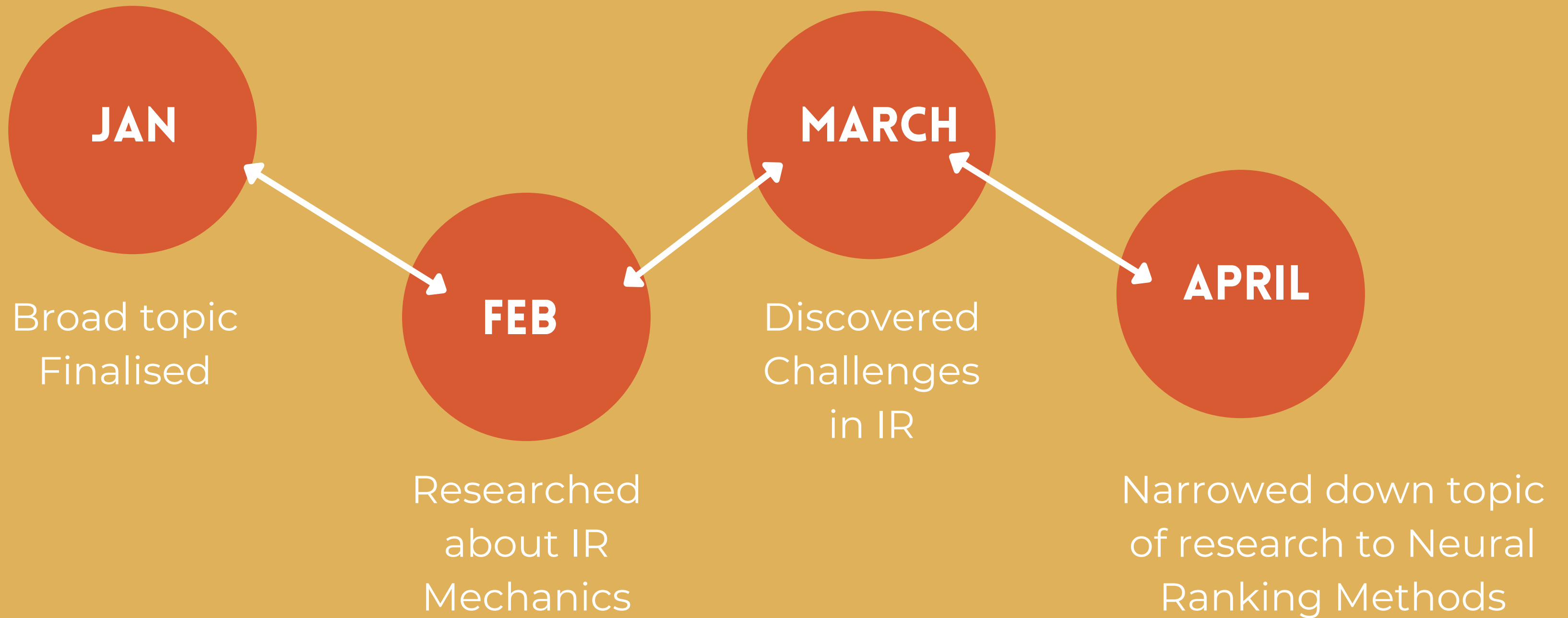
- Rank Brain(2015)

- Fred(2017)

CHALLENGES IN WEB SEARCH ENGINES

- Spam-
1. Text Spam
2. Link Spam
3. Cloaking

- Content Quality

- Quality Evaluation of updates

- Web Conventions(anchor text, hyperlink,Meta Tags)

- Duplicate Hosts detection

- Vaguely structured data

MILESTONES

JAN

Broad topic
Finalised

FEB

Researched
about IR
Mechanics

MARCH

Discovered
Challenges
in IR

APRIL

Narrowed down topic
of research to Neural
Ranking Methods

NARROWED DOWN TOPIC OF RESEARCH-

Neural Ranking Methods Architecture

BASE PAPER-

Neural Ranking Models for Document Retrieval

link-
https://link.springer.com/content/pdf/10.1007/s10791-021-09398-0.pdf

KEYPOINTS OF NEURAL RANKING MODELS –

- Neural Ranking models use deep learning techniques

- They improve Overtime using implicit raw data

- That is why they overcome the limitations of hand crafted features

- Examples of Deep learning techniques-
1. Convolutional neural network (CNN)
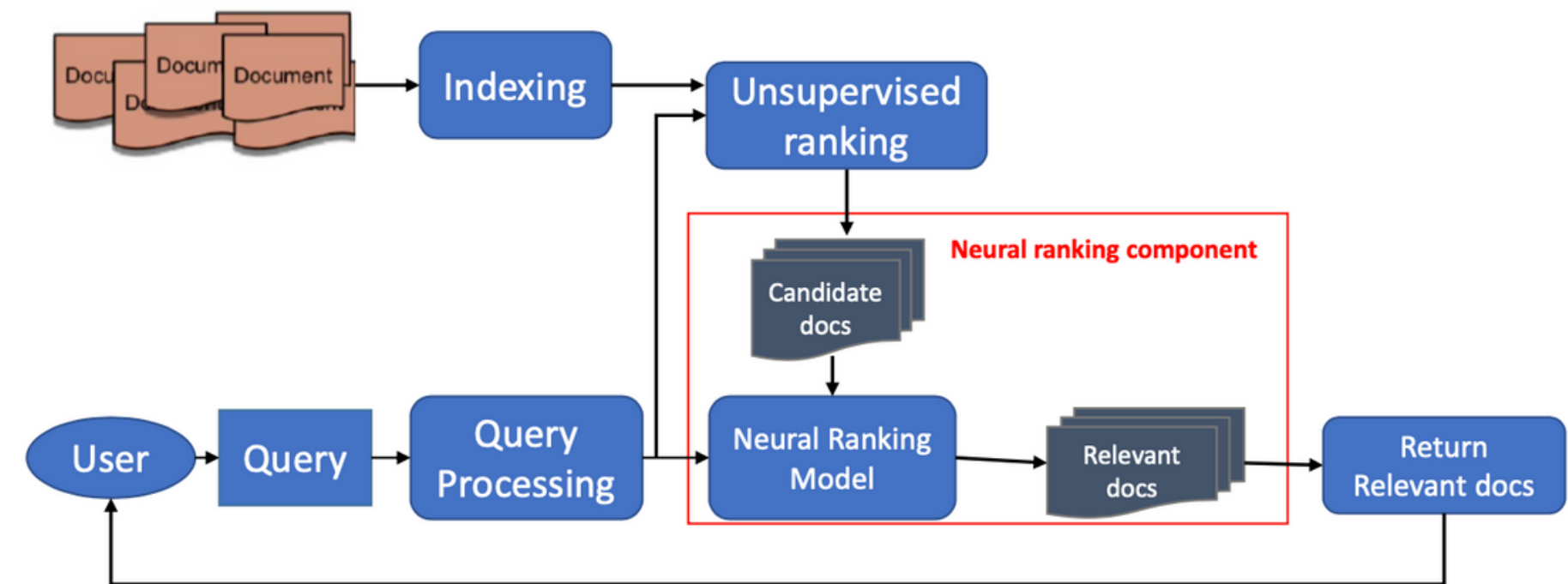2. Long short-term memory (LSTM)
3. Wordembedding

Fig. 1 Overview of the flowchart of the neural ranking based document retrieval. The neural ranking component is highlighted within the red box. The inputs to the neural ranking model are the processed query and the candidate documents that are obtained from the traditional ranking phase. The final output of the neural ranking model is a ranking of relevant documents to the user's query

DEEP LEARNING TECHNIQUES INTRODUCED

- CNN
- RNN
- LSTM
- GRU
- ATTENTION MECHANISM
- WORD EMBEDDING
- DEEP CONTEXTUALIZED LANGUAGE MODELS
- KNOWLEDGE GRAPHS

MAJOR APPLICATION OF NEURAL RANKING MODELS

1. Ad_Hoc Retrieval

2. Question Answering

3. Community Question Answering

4. Automatic Conversation

PLAN FORWARD-

Improve the quality and Add more information about latest neural ranking alorithms

THANK YOU

QUESTIONS

I will be happy to answer your questions