RANK BASED RETRIEVAL BASED ON TERM WEIGHTS USING IMPLICIT RELEVANCE FEEDBACK

PAWAN KALYAN D

KARTHIK G

SURYA TEJA A

AGENDA

Introduction.
What is rank based retrieval??

Rank based retrieval using term frequency weights and IDF.

Rank based retrieval based on cosine similarity.

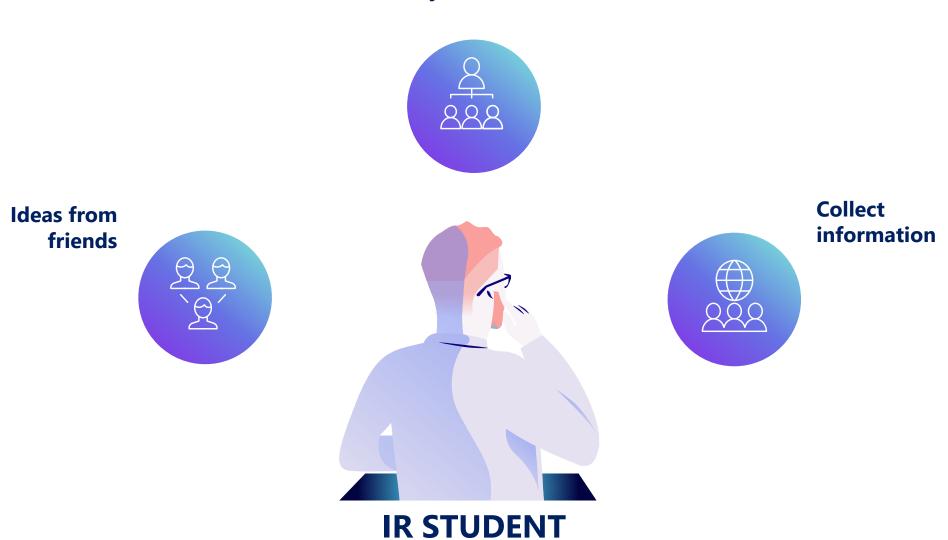*Rank based retrieval with implicit feedback*

After a hectic semester, what does a typical IIITSian do?

o *Search for the best movies which have good IMDB rating, right!*

What does an IR student do?

Retrieve the most relevant idea which could satisfy his/her need

Ideas from friends

Collect information

IR STUDENT

A Survey says that most of the students suggest 'movies'.

Our system is composed of two different retrieval methods, one is based on Tf-Idf weights and other on Cosine Similarity

Rank based retrieval using TF-IDF score

o *The tf–idf is the product of two statistics, term frequency weights and inverse document frequency*

o *Term frequency weight can be obtained using*

$$tf(t,d) = (1 + \log(Tf(t,d)))$$

o *Inverse document frequency can be obtained using.*

$$\text{idf}(t, D) = \log \frac{N}{|\{d \in D : t \in d\}|}$$

o *The Tf-Idf can obtained using*

$$\text{tfidf}(t, d, D) = \text{tf}(t, d) \cdot \text{idf}(t, D)$$

Rank based retrieval using Cosine similarity

o *Cosine similarity is a measure of similarity between two non-zero vectors of an inner product space that measures the cosine of the angle between them.*

o *Two vectors with the same orientation have a cosine similarity of 1, two vectors oriented at 90° relative to each other have a similarity of 0*

o *Cosine similarity can be obtained using*

$$Cosine\ Similarity(d,q)=Dot\ product(d,q)/||d||*||q||$$

Rank based retrieval Integrated with implicit feedback

o *The Tf-Idf system is integrated with implicit relevance feedback to increase the relevance of the retrieved documents.*

o *Click Data for each document is stored in a log file.*

o *Click data from the log file is used to predict preferences between pairs of documents.*

o *Document which is often opened by users will be the famous document.*

o *So integrating ranked documents with relevance makes the retrieval more efficient.*

– click data

$$
\begin{aligned}
&d_1 \\
&d_2 \\
&d_3 \ (\text{clicked}) \\
&d_4
\end{aligned}
$$

– generated preferences

$$
\begin{aligned}
d_3 &> d_2 \\
d_3 &> d_1 \\
d_3 &> d_4
\end{aligned}
$$

Thank You