https://vvtesh.sarahah.com/

# Information Retrieval

Venkatesh Vinayakarao venkateshv@cmi.ac.in

Chennai Mathematical Institute



What we find changes who we become.

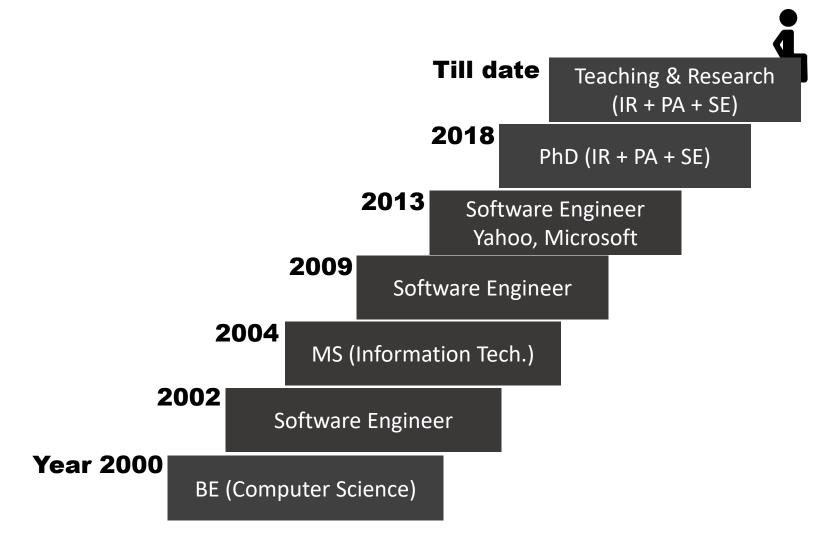
-Peter Morville.



# Agenda

A gentle introduce to "Information Retrieval"

# About Me



### **Acknowledgment**

Some slides are borrowed from the companion website of Manning et al.'s IR book (https://nlp.stanford.edu/IR-book/)

A good teacher can inspire hope, ignite the imagination, and instill a love of learning.

-Brad Henry.





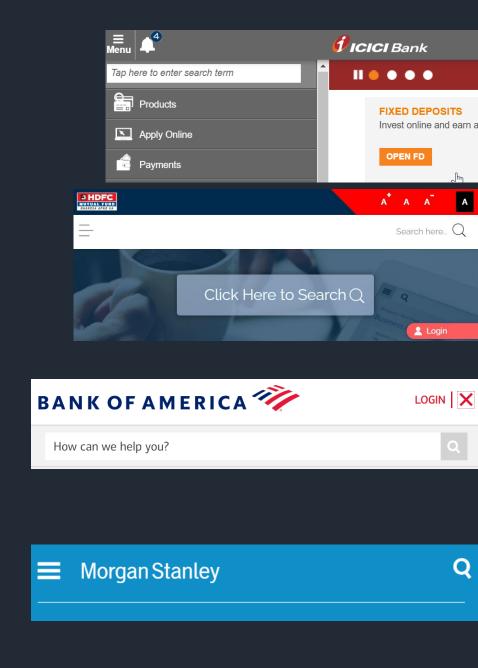


Life without search engines is difficult to imagine!

# Search in Banking and Finance

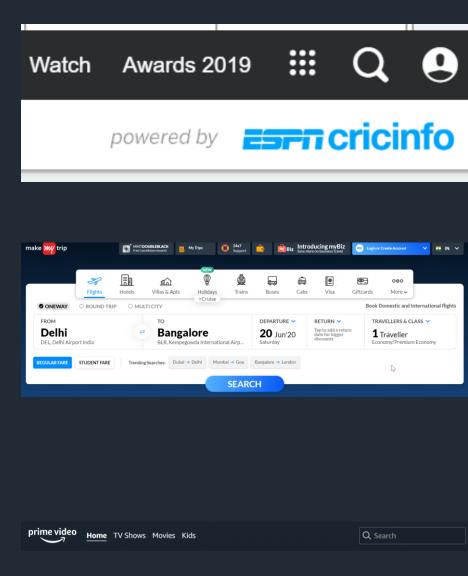
Lots of products to sell

Reach a part of documentation faster



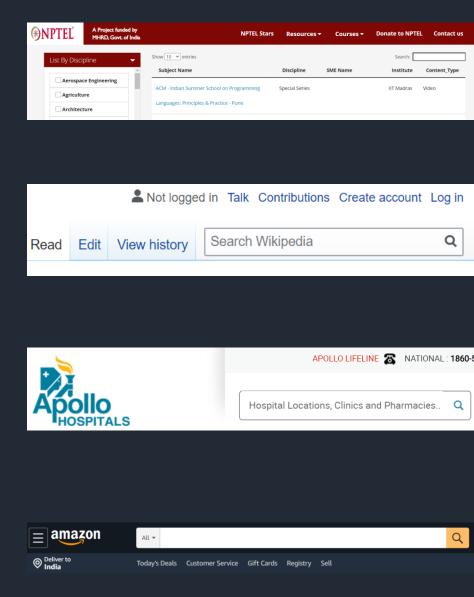
## Search in Sports, Travel & Entertainment

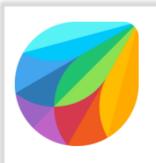
Search events, programs, and schedules



# Search in Education, Ecommerce and Healthcare

Search courses, articles, symptoms, books, etc.





#### Search Relevance Engineer

Freshworks · Hyderabad, IN

Posted 1 month ago · 301 views



## Search COE Engineer

Morgan Stanley · Mumbai, IN

Posted 2 days ago · 32 views



## Technical Architect / Sr SDE -Amazon Search Engine Technologies

Amazon · Bangalore, IN

Posted 3 weeks ago  $\cdot$  271 views



#### Search Engineer

Kubric · Bangalore, IN

Posted 3 months ago · 56 views



## Engineering Manager - Solr Search

Tata CLiQ  $\cdot$  Bangalore, IN

Posted 4 months ago · 174 views



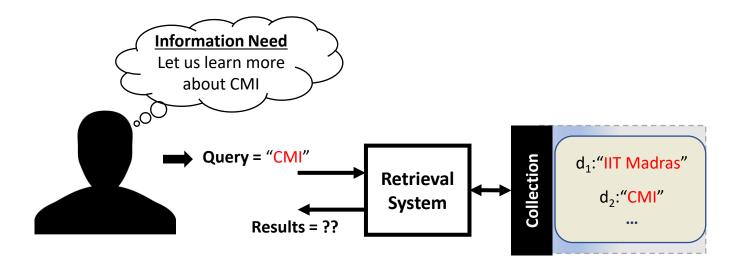
#### Senior Product Manager -Search

Walmart Labs India · Bengaluru,

Karnataka, India

Posted 5 days ago · 2,912 views

# What is Information Retrieval?



Information Retrieval (IR) is finding material (usually documents) of an unstructured nature (usually text) that satisfies an information need from within large collections.

From the Manning et al. IR Book.

## Information

Shannon's Definition, Fisher Information, Neumann Entropy, ...



Information is any entity or form that provides the answer to a question of some kind or resolves uncertainty. - Wikipedia.

# Role of Information

- If only you knew
  - Which stock to invest in?
  - Which faculty to work with?
  - How to get into a top college?
  - Which course to register for?
  - What to study?
  - How to prepare for job interviews?
  - ...
- If only you had the information, you could rule this world!
- What happens when all the information is deprived from you?

# Solitary Confinement is Cruel



# Information

#### Several retrieval

systems: Lycos, Altavista, MSN, B idu, Yahoo!, Ask. etc.,

Google

**Digital** Libraries

1970's

**Universal Digital** Library,

**Project** Gutenberg, etc.

30 Trillion documents

1998

> 130 Trillion in 2016

nationale de France

300 BC.

**Royal Library of** Alexandria

Bibliothèque

British

Library

1970's

170+ Million

Collection

1463

Information Retrieval – Road Ahead Crawling Content Processing Documents Relevance Query Understanding and Ranking Query Index Retrieval **Processed Content System Results** Index Compression Query **Evaluation Results Human Judges Techniques** 

# Technologies & Frameworks







Apache



Apache



Univ. of Glasgow

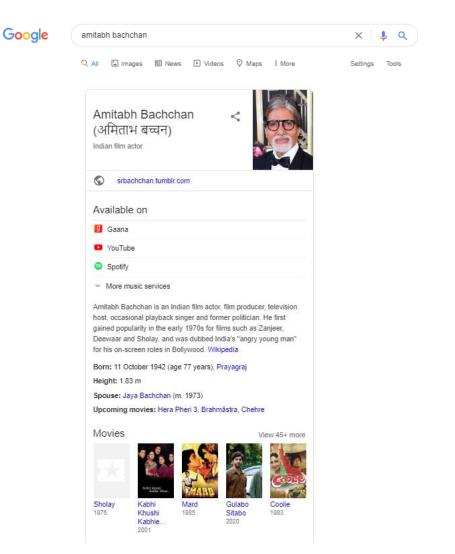


**UMass & CMU** 

There are many more....

Thanks to these... We can now focus on more complex problems.

# **Entity Search**





#### amitabh bachchan



#### Amitabh Bachchan

Indian Film Actor

Amitabh Bachchan is an Indian film actor, film producer, television host, occasional playback singer and former politician. He first gained popularity in the early 1970s for films such as Zanjeer, Deewaar and Sholay, and was dubbed India's "angry young man" for his on-screen roles in Bollywood. Referred to as the Shahenshah of Bollywood, Sadi ka Mahanayak, Sta... +











wikipedia

Twitter

acebook

Official site Instagram

Born: 11 Oct 1942 (age 77) · Allahabad, India

Height: 6' 0"

Spouse: Jaya Bachchan (m. 1973)

Children: Abhishek Bachchan (Son) · Shweta Bachchan-Nanda (Daughter)

Upcoming films: Chehre · Wisdom for Heroes · Brahmastra

Parents: Harivansh Rai Bachchan (Father) · Teji Bachchan (Mother)

#### Measuring Recommendation Explanation Quality: The Conflicting Goals of Explanations

Krisztian Balog Google London, UK krisztianb@google.com

Filip Radlinski Google London, UK Iftah Gamzu

Amazon Tel-Aviv, Israel

Operationalizing the Legal Principle of Data Minimization for Personalization

Marina Haikin Amazon

**Query Rewriting for Voice Shopping Null Queries** 

Tel-Aviv, Israel

Nissim Halabi Amazon Tel-Aviv, Israel nissimh@amazon.com

Research contributions from leading corporates in SIGIR 2020

Asia J. Biega Microsoft Research Montréal

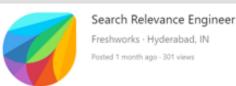
Peter Potash Microsoft Research Montréal Hal Daumé III Microsoft Research NYC University of Maryland

Fernando Diaz Microsoft Research Montréal Michèle Finck
Max Planck Institute for Innovation
and Competition

In spite of all these developments, "search"ing effectively has been an art.

(This is why) The academia, research labs, and the software industry needs you.

Let us strive to build better search experiences.





Search COE Engineer

Morgan Stanley · Mumbai, IN

Posted 2 days ago - 32 views

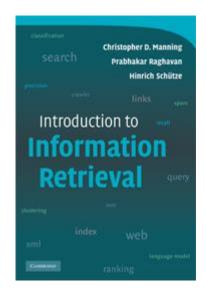




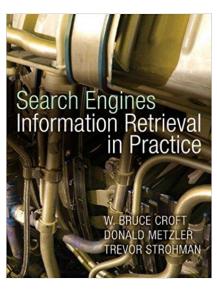
Search Engi Kubric - Banga Posted 3 months as

Amazon - Bangalore, IN
Posted 3 weeks ago - 271 views

# Resources



**Course Text** 



Reference

# Thank You