

Introduction to Search and Recommender Systems

Chirag Shah

Introduction and a Brief History of IR

Reading List:

- Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, Introduction to Information Retrieval, Cambridge University Press. 2008. <https://nlp.stanford.edu/IR-book/>
- Culpepper, J. S., Diaz, F., & Smucker, M. D. (2018, August). Research frontiers in information retrieval: Report from the third strategic workshop on information retrieval in lorne (swirl 2018). In ACM SIGIR Forum (Vol. 52, No. 1, pp. 34-90). New York, NY, USA: ACM.
- Harman, D. (2011). Information retrieval evaluation. Synthesis Lectures on Information Concepts, Retrieval, and Services, 3(2), 1-119.
- Mark Sanderson and W. Bruce Croft. (2012). The History of Information Retrieval Research. Proceedings of the IEEE, 100, p1444-1451.
<https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6182576>

A brief timeline

Before being broadly used in search engines, information retrieval systems were discovered in commercial and intelligence applications. During the 20th century, there was a consistency in doubling the amount of digital storage capacity. The development of information retrieval systems reflects the increasing demand for information indexing and searching techniques. At that time, IR systems that were capable of searching and querying databases rapidly became ubiquitous when people found that traditional techniques did not work for large quantities of unstructured information.

The 1960s

Gerard Salton: leader and founder of a IR group - “the father of information retrieval”
He developed the SMART (System for the Mechanical Analysis and Retrieval of Text) information retrieval system with his team at Cornell. His one of the most famous approaches was to view queries and documents as vectors within N-dimensional space where N is the number of unique terms in the corpus. Another significant renovation was the relevance feedback so the previously retrieved documents would be marked as relevant to the search query.

The biggest award for contributions in IR is named after Salton and is given out every three years at ACM SIGIR conference. 2021 is the year of the Salton Award!

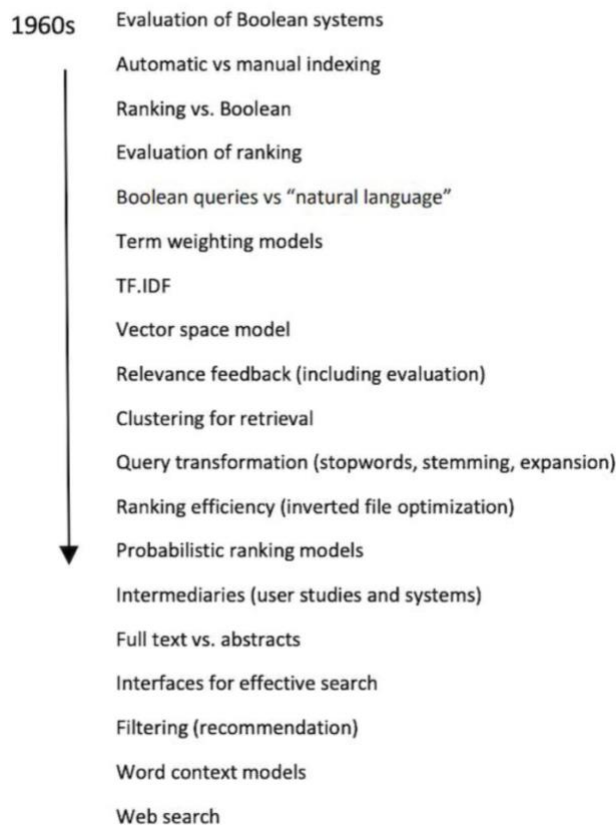
The 1970s

Luhn’s term frequency (tf) weights and inverse document frequency (idf) were developed during this period. Karen Spärck Jones, a british computer scientist who developed the idea of idf states that lower the frequency of a word’s appearance in a document is, higher the importance of that word is during the retrieval process.

The 1980s

Based on the development of tf-idf and probability theories, the ranking function BM25 was invented and proved to be highly effective so that it is still being used nowadays. Latent semantic indexing (LSI) was also developed during this period, as an advanced version of vector space representation. In late 1980s, Text REtrieval Conference (TREC) was formed by Voorhees and Harman, where researchers built a large magnitude of document collections and discovered that the previous ranking techniques didn't work for all kinds of collections. It was realized that different weighting and ranking systems need to be developed and applied to different collections.

Research Timeline



<https://ciir.cs.umass.edu/downloads/essir2019/doing-ir-research-croft.pdf>

Cranfield experiments

Reading List:

- Ellen M. Voorhees. [The Philosophy of Information Retrieval Evaluation](#)
- William Webber. [The Cranfield tests « Evaluating E-Discovery](#)

The Cranfield experiments contain a series of experiments started in the 1960s. Cranfield 1 evaluated the relative efficiency of indexing systems, while Cranfield 2 tested in detail to examine what elements of an index language device make it more effective.

In Cranfield 1, four indexing experiments were compared and evaluated.

1. Universal Decimal Classification
2. The Alphabetical Subject Catalogue
3. The Faceted Classification Scheme
4. Mortimer Taube's Uniterm system of coordinate indexing

These four systems were tested with the same set of document collections and the same information provided, and they used precision and recall ratio as the metrics to evaluate the effectiveness of their results. The relevance judgement was manually made before the experiments by domain experts.

The results of Cranfield 1 showed that there was little difference between different indexing systems, which motivated them to continue on Cranfield 2 and examine closely each component. The results of Cranfield 2 were also unexpected. It suggested that simple terms actually perform better than controlled languages and simpler devices are better than complicated techniques. This was the main outcome of Cranfield experiments, which was evolutionary but significant, showing that complex and controlled indexing techniques were unnecessary in terms of effectiveness in indexing and ranking. It also suggested that the query terms should be better left in their original forms.

The results of these experiments remained as continual debates for years. However, the influence of Cranfield experiments was profounding. It provided an inspiring methodology for IR evaluation later in the future, although the original Cranfield experiments were developed based on some simplifying assumptions which were not true in reality. The document collection used in the experiments was also widely used for many years as a testbed corpus at that time.

TREC (Text REtrieval Conference)

Reading List:

- [Text REtrieval Conference \(TREC\) Overview](#)
- TREC Experiment and Evaluation in Information Retrieval. Edited by Ellen M. Voorhees and Donna K. Harman. <https://mitpress.mit.edu/books/trec>

Beginning in 1992, TREC was funded by DARPA (US Defense Advanced Research Project) and run by NIST (National Institute of Standards and Technology). The purpose of this conference is to provide a platform for researchers and professionals in the field of IR to evaluate IR methodologies. As provided on their official website, there are four main goals:

1. to encourage research in IR based on large test collections, uniform scoring procedures, and results comparisons;
2. to increase communication among industry, academia, and government by creating an open forum for the exchange of research ideas;
3. to speed the transfer of technology from research labs into commercial products by demonstrating substantial improvements in retrieval methodologies on real-world problems; and
4. to increase the availability of appropriate evaluation techniques for use by industry and academia, including development of new evaluation techniques more applicable to current systems.

As TREC was established, one of its primary motivations was to validate the test collection evaluation paradigm as introduced in the Cranfield experiments. They aimed to develop an evaluation methodology that was systematic and simplified with only a set of queries, a test collection, an IR

engine, and a series of judgements. TREC was successful and pioneering. During the first six years of TREC, the effectiveness of IR systems had approximately doubled, which also accelerated the transfer of research ideas into commercial usages. Many of today's search engines were first developed in TREC. In addition, TREC created the first large-scale evaluation of the retrieval of non-English and multi-language documents.

TREC consists of multiple tracks, each of them focuses on a particular retrieval task. Some of the past tracks are no longer removed from recent TREC.

TREC 2021 Call for Participation <https://trec.nist.gov/pubs/call2021.html>

Available tracks:

- Clinical Trials Track
- Conversational Assistance Track
- Deep Learning Track
- Fair Ranking Track
- Health Misinformation Track
- Incident Streams Track
- New Track
- Podcasts Track