A Brief Review of Current Research in Information Retrieval: Postprint

Authors: Wang Bin, Li Peng

Date: 2017-03-10T00:00:00+00:00

Abstract

In recent years, on the one hand, information retrieval has continuously intersected and integrated with other disciplines while developing on its own; on the other hand, the emergence of new resources and platforms has also promoted its rapid development. Information retrieval research is exhibiting trends toward personalization, collaboration, and socialization. This paper summarizes some new directions in information retrieval research in recent years and analyzes several future development trends.

Full Text

Preamble

Vol. 9 No. 5
Information Technology Letters

A Brief Overview of Current Information Retrieval Research
Wang Bin, Li Peng

Abstract

In recent years, information retrieval has undergone rapid development driven by two parallel forces: continuous cross-disciplinary integration on one hand, and the emergence of new resources and platforms on the other. Research in this field has increasingly exhibited trends toward personalization, collaboration, and socialization. This paper summarizes recent developments in information retrieval research and analyzes several future directions.

Keywords: information retrieval; personalization; collaborative search; social search

1. Introduction

Information Retrieval (IR) is the discipline concerned with how to quickly, accurately, and comprehensively obtain information that users need from large-scale raw data. Originating from the need to locate documents in libraries, IR later expanded into various information processing domains. The advent of the Internet and the acceleration of global digitization have exacerbated the problem of information overload. Finding relevant information from massive datasets has become both an urgent need and a significant challenge, spurring the advancement of IR technologies. In particular, over the past few years, the growing popularity of general-purpose search engines, the emergence of industry-specific search engines, and substantial commercial investment in search technology have created an unprecedented boom in IR research—the core technology underlying search engines. Scholars from fields such as natural language processing, artificial intelligence, machine learning, statistics, cognitive science, databases, and distributed parallel processing have turned their attention to this time-honored yet ever-fresh application domain. Information retrieval has thus become a truly interdisciplinary field. For example, top-tier international conferences that once featured only a handful of IR papers—such as SIGIR—now include numerous IR contributions across diverse venues including SIGKDD, ICDM, WWW, SIGMOD, NIPS, VLDB, ACL, IJCAI, AAAI, EMNLP, and CIKM. At some of these conferences, IR-related papers even constitute the majority of submissions. It is fair to say that IR research has entered an unprecedented era of technological convergence and integration. This paper organizes and summarizes recent research trends in information retrieval to provide a reference for relevant researchers.

Any information retrieval system follows a basic structure: users express their information needs as queries submitted to the system, which then computes similarity between each document in the collection and the query, finally returning a subset of results that may satisfy the user' s needs. Different IR applications may vary in their requirements for queries, document collections, result sets, and similarity computation methods, giving rise to various specialized applications. The most common IR systems include web search engines (typified by World Wide Web search engines), information recommendation systems (such as information subscription systems), and question-answering systems aimed at delivering direct answers.

[Figure 1: see original paper]. Basic structure of an information retrieval system

For convenience of discussion, we categorize IR research into three layers (as shown in Figure 2 [Figure 2: see original paper]), from bottom to top: the resource layer, the technique layer, and the application layer. The resource layer comprises the resources utilized in IR research. In addition to commonly used datasets such as TREC data, WordNet, HowNet, and other publicly available online corpora, we highlight resources and tools that have been widely adopted in recent research, including social network data (e.g., Facebook), microblog data (e.g., Twitter, Sina Weibo), encyclopedia data (e.g., Wikipedia, Hudong Baike), and others. The technique layer encompasses technologies from diverse fields—including Natural Language Processing (NLP), Machine Learning (ML), Data Mining (DM), and Parallel Processing (PP)—that integrate with traditional IR techniques such as modeling, indexing, query expansion, and relevance feedback. The application layer contains numerous IR-based applications, such as web search engines, microblog search, blog search, forum search, and more. Notably, evaluation technology in IR deserves special mention. The proliferation of techniques and applications has made evaluation increasingly important, making it a perennial topic in the IR field.

[Figure 2: see original paper]. Layered architecture of information retrieval research

2. Research on New Applications of Information Retrieval

In recent IR research, studies targeting new applications have occupied a dominant position. On one hand, the proliferation of online information in different formats, forms, and domains has intensified the demand for retrieving such information. On the other hand, the emergence of new information types has introduced distinct requirements and specific research problems. Below we enumerate several representative application areas.

(1) Multimedia Retrieval Research

Traditional IR research has primarily focused on textual objects. However, the increasing volume of multimedia documents has intensified the demand for multimedia retrieval. Depending on the media type, multimedia retrieval can be categorized into image retrieval, video retrieval, speech retrieval, music retrieval, and other types, each forming its own research community. For instance, the renowned international conference CIVR (ACM Conference on Image and Video Retrieval) has been held annually since 2002, convening ten times to date. The TRECVID evaluation conference (http://www-nlpir.nist.gov/projects/trecvid/), focused on video retrieval, has also been held for multiple editions. Researchers in music retrieval have formed the International Society for Music Information Retrieval (http://www.ismir.net/) and have organized twelve annual conferences since 2000.

Multimedia retrieval can be viewed as an integrated application of media processing and traditional IR techniques. Through media analysis and processing, low-level features (such as color, shape, texture, and volume) can be extracted. However, user queries typically involve high-level semantic concepts (e.g., "sunrise," "Five-star Red Flag"). The gap between low-level media features and high-level semantic concepts is known as the "semantic gap." Bridging this semantic gap represents the ultimate goal of multimedia retrieval. Recent studies suggest that leveraging multimodal information (such as simultaneous analysis of audio streams, video streams, and text streams in videos) to improve semantic mapping accuracy appears to be a viable approach. Essentially, these studies establish associations between different media representations, leading to the emergence of cross-media IR research. This area can be understood from two perspectives: first, enabling queries in one media format to retrieve documents across different media formats; second, integrating multiple media or modal information to improve retrieval result quality.

Currently, commercial multimedia search systems remain predominantly text-annotation-based. Developing mature applications that truly integrate media processing technologies into retrieval systems will require considerable additional effort.

(2) Domain-Specific and Scenario-Based Information Retrieval Research

Systems that differ from general-purpose search are often termed "vertical search" systems. The term "vertical" refers to information retrieval tailored to specific domains or scenarios. Unlike traditional general search research, vertical search typically demands higher data quality. Consequently, information extraction—the technology for isolating required information from documents—often becomes a key enabling technique. For example, people search requires extracting various attribute information about individuals (e.g., gender, age, occupation) from documents, while product search requires obtaining product attributes (e.g., model, price, origin). When target documents are highly structured, rule-based templates or machine learning methods can be employed for classification. For less structured documents, text understanding techniques are often necessary. When information is scattered across different sources, information integration techniques become essential. The quality of information extraction often determines the effectiveness of vertical search systems. In practice, hybrid human-computer processing methods are commonly employed.

In recent years, domain-specific retrieval research has emerged for medical literature, biological literature, patent documents, legal documents, and more. For instance, the TREC conference added a genomics literature retrieval track in 2003 (some consider this part of "bioinformatics") and introduced legal document retrieval in 2006. Another important evaluation conference, NTCIR (http://research.nii.ac.jp/ntcir/), added patent literature retrieval in 2002. These studies focus on domain-specific requirements, such as incorporating relationships between drugs and symptoms in medical literature retrieval.

Additionally, new web phenomena stimulate IR research. Typical examples include blog retrieval from previous years and microblog search in recent years. As a product of the Web 2.0 era, blogs have attracted widespread attention, creating strong demand for blog analysis and retrieval. Unlike traditional retrieval objects, blogs possess unique characteristics: they comprise bloggers, blog posts, comments, links, trackbacks, and other elements—essentially aggregations of diverse information types. Research on this novel entity, including spam blog identification and filtering, blog community discovery and analysis, blog content mining, and trend prediction, has been quite active in recent years.

Notably, due to the abundance of subjective comments on blogs, sentiment analysis of blog comments has gained increasing attention. This technology clearly has broad applications in product reviews and significant commercial value.

In essence, sentiment analysis examines whether text contains subjective opinions, what objects are being commented upon, whether the sentiment is positive or negative, and the intensity of such sentiment. Some researchers also refer to it as emotion analysis or opinion analysis. TREC introduced a blog retrieval subtask in 2006 that required returned documents to consider both relevance and sentiment presence. Since sentiment analysis involves text processing techniques, it has also attracted widespread interest from natural language processing researchers, with numerous domestic and international studies addressing this problem. The Chinese Information Processing Society of China has even organized Chinese sentiment analysis evaluations. Current results indicate that most work remains at the lexical level (sentiment words), and significant effort is still required before this technology becomes truly practical.

Microblogs have flourished in recent years due to their close integration with mobile Internet. Characterized by short messages, rapid transmission, and real-time nature, microblogs contain user information, follow relationships between users, and information forwarding paths. Research indicates that microblog search differs from traditional search in terms of query intent, data distribution, and other aspects, representing a promising research direction.

(3) Mobile Search Research

The high penetration rate of mobile devices (particularly smartphones), extensive mobile network coverage and development prospects, close binding between phones and users, and the enormous user base collectively make mobile information search commercially significant. Top-tier conferences such as WWW and SIGIR have repeatedly held workshops on mobile search. Like conventional search, mobile search must acquire, organize, and provide access to information. However, current mobile search research fundamentally aims to overcome the limitations of mobile output (primarily screen display) and input constraints. Due to limited screen size, mobile search demands more precise results to minimize irrelevant information while requiring more rational result layouts that are concise, highlight key information, and facilitate further browsing. This necessitates integrating various technologies including ranking algorithms, information filtering, text analysis, summarization, and human-computer interaction. Given input constraints on mobile devices, retrieval interaction often employs pinyin-to-character conversion, query recommendation, and query completion to reduce user input burden. Beyond text search, multimedia search constitutes an important component of mobile search, with research focusing on media annotation, display, and transmission. Additionally, mobile devices enable the incorporation of contextual factors such as user profiles and geographic location to optimize search results (e.g., personalized search or local search).

Currently, mobile search research remains in its infancy. Due to various limitations (network speed, data costs, device capabilities, etc.), mobile search adoption remains far below that of Internet search engines. However, mobile search is poised for broader prospects in the near future.

(4) Retrieval-Based Advertising Technology

Most commercial search engines today feature online advertisements. When users input queries, advertisements potentially related to their queries appear alongside or within search results. Online advertising represents a major revenue source for search engine companies and provides significant benefits to advertisers, attracting considerable attention from both industry and academia and even spawning a new term: "Computational Advertising." Recent SIGIR and WWW conferences have made advertising in IR a key topic. Inputs can be search queries—often called paid search—with matched advertisements termed sponsored ads or keyword-driven ads. Alternatively, inputs can be web pages, images, or videos being browsed, with resulting ads called context-driven ads or contextual ads. The system then pushes appropriate advertisements to users based on inferred intent from their input. This process closely resembles information retrieval: matching user queries against an advertisement repository and pushing the most likely ads to users. IR techniques clearly have tremendous potential in computational advertising, which explains why this area has garnered significant attention in the IR community.

Unlike general IR, matching in computational advertising is not simply based on traditional relevance concepts but rather on mining users' commercial intent. For example, when a user queries "Beijing to Shanghai train tickets," the system can infer the user likely plans to travel to Shanghai and consequently push hotel advertisements for Shanghai. This goes beyond simple relevance matching. If no commercial intent can be analyzed from user input, the system may choose not to push any ads. Computational advertising must balance matching accuracy with positive user experience. Ranking in this context is extremely complex, as it must simultaneously consider user experience, advertiser interests, and commercial constraints. Consequently, ranking algorithms often incorporate economic revenue models alongside traditional approaches. Additionally, advertisement texts are typically short and lack the contextual linguistic information available in conventional documents, requiring modifications to traditional IR models. In summary, computational advertising is an emerging and challenging research domain. However, entering this field remains difficult, with data scarcity being a major bottleneck. Advertising data involves commercial interests and is not publicly available. Although Microsoft once claimed it would provide data to support research in this area, this has not yet materialized. Evaluation presents another major challenge, as computational advertising aims to maximize benefits for all stakeholders—optimal user experience and maximum profits for search engines and advertisers—making it difficult to simulate these objectives in research settings. Therefore, developing effective evaluation methodologies for research results remains an important open problem.

(5) Personal Information Management (PIM) and Desktop Search

Research

While many researchers focus on massive Internet resources, the decreasing cost of hard drives has led to users accumulating increasingly large personal datasets, overwhelming individual machines. Personal Information Management has become an important and urgent need, gradually gaining widespread attention from the research community. Since 2005, an annual international workshop on PIM has been convened, and conferences such as SIGIR and SIGCHI (human-computer interaction) have listed PIM as a major topic in recent years. The top-tier journal ACM Transactions on Information Systems (TOIS) organized a special issue on PIM at the end of 2008. PIM primarily studies the acquisition, organization, management, maintenance, and retrieval of personal information and represents core technology for the contested "future desktop." Desktop search, which we are familiar with, is a key component of PIM. In recent years, desktop search technology (searching user data on hard drives) has become a focus for search engine companies, many of which have developed their own desktop search engines. However, commercial desktop search tools still have room for improvement in terms of user satisfaction.

Compared to web search, desktop search has access to much less information. Web search can leverage not only keywords but also link relationships between web pages and user logs. Initially, desktop search only had access to text information and file access times. Meanwhile, users demand higher accuracy from desktop search than from web search, typically expecting to locate desired files directly. Therefore, obtaining more information and leveraging it for result ranking presents a difficult challenge in desktop search. Desktop search does have certain advantages, such as close integration with the user, enabling the utilization of user information. We believe that incorporating user behavioral information represents a promising and feasible research direction for desktop search. Existing literature has utilized logs of user file accesses and queries, applying machine learning methods to develop personalized ranking algorithms. Others have used access patterns to establish relationships between files and applied algorithms similar to PageRank for ranking.

(6) Social Mining and Search Research

In recent years, the Internet has increasingly exhibited clear socialization trends. Social tagging sites such as Delicious have accumulated vast amounts of data, social networks like Facebook have gained widespread participation, and new information sharing and dissemination mechanisms represented by Twitter and Sina Weibo have shown vigorous vitality. These emerging phenomena have not only enabled new search applications such as Twitter search and microblog search but also provided valuable data for other applications due to their rich social information (user profiles, relationship information, behavioral information, and information associations). Researchers are mining the underlying patterns and profound implications of this data to further improve IR effectiveness.

3. Research on Information Retrieval Techniques

This section introduces research on IR-related techniques. Traditional IR techniques primarily include IR models, relevance feedback, query expansion, and indexing technologies. In recent years, technologies from machine learning, natural language processing, statistics, and other fields have been more extensively applied to IR. Key developments are summarized below.

(1) Research on Information Retrieval Models

The essence of retrieval models is to model the relevance between user needs and documents, primarily comprising representation techniques for queries and documents and relevance ranking techniques. Early models—including the Boolean model, vector space model, probabilistic model, and the statistical language modeling approach introduced in 1998—continue to be improved and applied. Additionally, several new models have emerged, with improvements focusing on two main aspects: first, incorporating relationships between terms to overcome the traditional term independence assumption; second, moving beyond traditional factors such as term frequency (TF), inverse document frequency (IDF), and document length to improve retrieval models.

The Markov Random Field (MRF) model represents an outcome of the first approach. It comprehensively considers various combinations of term relationships and integrates them into a retrieval model through different weights. Experimental results demonstrate that MRF models can outperform traditional retrieval models, with more pronounced effects on noisy data (such as web pages). Traditional IR models—whether the decades-old vector space and probabilistic models or the language modeling approach from the past decade—have only incorporated three factors: term frequency, inverse document frequency, and document length. Incorporating additional factors has long been a research goal. Some researchers have proposed introducing term proximity relationships into IR models and conducted preliminary experiments.

A particularly interesting study on IR models comes from the University of Illinois at Urbana-Champaign (UIUC). The researchers proposed that a good IR model must satisfy certain basic constraints and proved that existing models either partially satisfy or fail to fully meet these conditions (depending on parameters). Experimental results on traditional models with varying parameters confirmed their theoretical analysis. Building on this foundation, they proposed a framework and methodology for constructing new retrieval models, with the newly proposed models demonstrating certain advantages.

(2) Research on Machine Learning-Based Information Retrieval Models

In recent years, research on machine learning-based IR models has surged in popularity. The topic of "Learning to Rank" has occupied a significant portion of SIGIR papers; machine learning conferences such as NIPS have also embraced this topic; Microsoft Research Asia has even established a dedicated website (http://research.microsoft.com/en-us/um/beijing/projects/letor/index.html) providing relevant papers, datasets, evaluation standards, and tools for researchers. Unlike existing heuristic ranking functions, these studies assume that ranking functions take a certain form and then determine model parameters through training on labeled datasets to obtain the final ranking function. Learning to rank elegantly transforms the ranking problem into a machine learning problem, attracting particular attention from machine learning researchers.

Various machine learning methods have been introduced into IR, including supervised learning and semi-supervised learning, as well as generative and discriminative approaches. Traditional machine learning methods have been "retrofitted" for learning to rank, with methods such as Ranking SVM, RankBoost, and RankNet emerging in recent years. Despite ongoing controversies and theoretical imperfections, learning to rank is undeniably gaining widespread attention. More importantly, it enables machine learning researchers to quickly engage in IR research, injecting new vitality into the field.

(3) Research on Query Analysis Techniques

In a typical IR system, users express their information needs as queries submitted to the system, which then returns results ranked by relevance. Queries serve as the interactive "language" between users and computers, playing a pivotal role: they must accurately reflect user information needs while being understandable and processable by the retrieval system. However, for various reasons (such as differences in user background and experience), user-inputted queries may not accurately reflect their true information needs, and initial queries may not return satisfactory results. Traditional IR often relies on keyword matching and cannot understand the deeper intent behind queries (for example, a user querying "word processing shareware download" likely wants a website for quickly downloading relevant software, yet many search engines only provide keyword-matched irrelevant results). Therefore, deep analysis and understanding of user queries are necessary to reconstruct original queries or conduct targeted retrieval based on user intent, thereby improving IR precision. The ultimate goal is to reduce the time users need to reach target documents.

In recent years, research on query analysis has proliferated, occupying significant space at major conferences. Query intent classification holds an important position in query analysis. Various classification schemes exist: for instance, some categorize queries into informational, navigational, and transactional types. Informational queries seek various types of information related to the query (e.g., "Chinese history"). Navigational queries target specific website entry points or personal homepages (e.g., "Sina homepage"). Transactional queries typically require subsequent interactive operations after obtaining results (e.g., downloading software after searching for "word processing shareware download"). Once classified, targeted retrieval methods can be applied. For example, research has shown that anchor text is highly effective for navigational queries, so once such a query is identified, anchor text weight can be increased to improve results.

Beyond this classification scheme, queries can also be semantically categorized by domain (e.g., "computer science," "physics," "chemistry"), by presence of commercial intent, by ambiguity, or by multimedia requirements. Since queries are typically short and provide limited information, query intent classification often requires external resources, such as using query logs for training to derive classification rules.

Another query analysis technique is query difficulty prediction, which typically judges result quality based on the query and returned results. Queries yielding good results are classified as "easy," while those yielding poor results are "difficult." Distinguishing between "difficult" and "easy" queries has many applications —for example, when a user inputs a "difficult" query, we can recommend related "easy" queries to obtain better results. Query difficulty prediction research has been very active in recent years, with numerous related papers published at SIGIR and CIKM. These studies primarily judge result quality based on characteristics of result distributions.

Additionally, some research analyzes other query characteristics for targeted retrieval, such as whether queries require personalization or can be localized.

4. Utilization of New Resources in Information Retrieval

This section introduces the use of new resources in IR. The rise of new resources not only supplements traditional resources but also provides unique characteristics leveraged by IR research.

(1) Wikipedia

Wikipedia is an open online encyclopedia. Its free, collaborative, and open-content encyclopedia project has attracted participants worldwide and has grown into the world' s largest online encyclopedia. As of January 2008, the English version of Wikipedia contained over 6,000,000 entries, with continued growth expected. In recent years, Wikipedia has emerged as a research hotspot in IR and natural language processing due to its large scale, high quality, and collaborative editing features. In Wikipedia, each entry corresponds to an article, with the entry serving as a summary or topic description of the article' s subject. These topics are mostly composed of phrases. Since Wikipedia allows free editing within certain guidelines, these topics better reflect commonly used semantic concepts in real life. As Wikipedia continues to be edited and expanded, we can expect its collection of commonly used phrases to become more comprehensive and timely reflect changes in language usage.

Wikipedia provides new resources and methods for solving problems in natural language understanding and IR. Concepts in Wikipedia have multiple relationships, including hypernym-hyponym relations. A direct application is using Wikipedia as a dictionary to compute semantic similarity between concepts by constructing relationship graphs. Some researchers have used Wikipedia for named entity disambiguation, expanding original text fragments and providing contextual information necessary for disambiguation based on Wikipedia content. Many studies have leveraged Wikipedia to improve retrieval results.

(2) Open Directory Project (ODP)

In June 1998, programmer Rich Skrenta became frustrated with outdated and dead links in Yahoo' s search results and issued a call on the Internet, requesting volunteers worldwide to help edit a directory. The initiative quickly gained support from enthusiastic volunteers, giving birth to the revolutionary management approach known as the Open Directory Project (ODP). The most famous ODP website is Dmoz (http://www.dmoz.org), which is generally what people refer to when discussing ODP. Dmoz is maintained by over 80,000 volunteer editors who have categorized more than 4 million websites into over 590,000 detailed categories, with all information freely available to any individual or organization. ODP contains extensive manually edited and professionally curated directory information, which many companies and researchers have used to improve retrieval, classification, and clustering effectiveness. For example, Google considers website information from ODP in its search result ranking. Many researchers have achieved good results using ODP' s directory information, such as Paolo Ferragina et al., who used ODP to cluster search results and improve user search experience. This has attracted researcher attention, as evidenced by the KDD-CUP 2005 task of classifying 800,000 queries into 67 categories.

(3) Folksonomy

The term "Folksonomy" combines "folks" (people) and "taxonomy" (classification), sometimes translated as "social classification." It refers to a collaborative approach where users annotate web resources (web pages, images, etc.), and collective annotations from the crowd are used to classify resources. In this model, users are both consumers and creators of tags. User-applied tags reflect individual cognition of resources, while different users tagging the same resource with different tags reveal various aspects of the resource' s attributes. In the Web 2.0 era, web platforms provide mechanisms for annotating web pages, images, and other content, enabling Internet users to easily tag browsed information. Typical folksonomy systems include http://delicious.com/ (bookmark sharing) and http://flickr.com/ (photo sharing). Through delicious.com, users can save favorite web pages and tag them based on relevance judgments, allowing other users with similar interests to find and browse these pages through tags.

Folksonomy offers many benefits, such as enabling ontology, thesaurus, and classification system development; facilitating search and browsing; discovering new things and communities; and enabling collaborative recommendation. However, it also presents challenges, including mixed-language tags, difficulty controlling singular/plural forms, ambiguity, synonymy, and levels of abstraction, as well as spam tag interference. Some scholars are conducting research using natural language processing and other technologies to address these issues.

Tags play an important role in information organization and retrieval. In text retrieval, tags can be incorporated into vector space models to participate in document representation, or used to represent relationships between documents to assist in ranking. Tags are also significant for organizing and retrieving non-textual information such as images and videos. Additionally, tag information can help users manage knowledge, track developments, discover new resources, and find like-minded individuals.

(4) Search Logs

Search engines record user search behaviors, generating large volumes of logs. Typically, search logs contain user IDs, query terms, result lists, and click information. Much research on search logs involves statistical analysis, such as examining query length, click counts per query, query distribution across domains, average number of queries needed to obtain desired information, query frequency (popularity), and emerging queries. These analyses greatly help understand user search habits, monitor public opinion, and identify social hotspots.

As research has deepened, many researchers have begun using search logs to improve search engines themselves, while others treat them as a new data resource for discovering valuable information. For example, the Sogou Input Method' s lexicon uses new words discovered from user search logs. Increasing research has also emerged on named entity recognition, query understanding, and mining query similarity for query recommendation based on search logs.

5. Evaluation Research in Information Retrieval

Evaluation has always been a fundamental issue in IR, involving all aspects of retrieval, with the most important work focusing on retrieval effectiveness. Retrieval effectiveness evaluates system performance based on the ranking of relevant documents in returned results, involving both evaluation processes and metrics.

(1) Evaluation Process Research

The goal of the evaluation process is to obtain relevance judgments. Traditional relevance judgments are completed through manual annotation. Since retrieval corpora are typically large-scale, manually annotating all relevant documents for a specific query is infeasible. The pooling method provides a solution. Its basic idea is to take the union of top-ranked retrieval results submitted by each system, remove duplicate documents, and constitute the document set to be judged. Documents in this set are considered all relevant documents, while those not included are assumed irrelevant. This method cannot annotate all relevant documents, but analysis of evaluation results shows that its impact on outcomes is minimal, leading to its widespread adoption in TREC.

In recent years, TREC' s evaluation corpora have grown increasingly large. The terabyte track and million-query track use the GOV2 corpus, which is 426GB in size and contains 25,205,197 documents. For each topic, typically the top 10,000 documents (terabyte track) or top 1,000 documents (million-query track) are selected for performance evaluation, with 50 topics (terabyte track) or 1,700 topics (million-query track) used. Simple calculations reveal that even with pooling, the number of documents requiring judgment is substantial, as pooling treats all documents in the pool equally, requiring complete judgment. Can manual annotation be reduced? The answer is yes, thanks to the work of Ben and Javed.

The fundamental goal of evaluation is to obtain a ranking of system retrieval performance. If a system ranking consistent with the original evaluation can be obtained, the evaluation is effective. Since Average Precision (AP) calculation depends on the positions of relevant documents, those ranked higher contribute more to the final metric and produce greater differences in system comparisons, while lower-ranked documents have relatively less impact. Ben proposed ranking documents by their contribution to system comparison and annotating relevance in this order, stopping at a certain point. By reformulating the AP formula and treating document relevance as a random variable, Ben made AP itself a random variable. For judged documents, relevance values are determined; for unjudged documents, Ben designed an order regression-based model to predict document relevance probability. This enables calculation of each system' s expected AP per topic and consequently the expected Mean Average Precision (MAP) across systems. Systems are then ranked by expected MAP. This partial judgment approach can significantly reduce annotation volume, and with a well-designed prediction model, the predicted relevance probabilities can be robust for final evaluation ranking. Javed's method uses sampling to obtain samples for estimating AP values, also substantially reducing annotation volume. Their methods have been applied in the million-query task.

(2) Evaluation Metric Research

In evaluation, an evaluation metric is typically calculated for each topic, with system performance evaluated by averaging across multiple topics. Common metrics include AP, R-precision, b-pref, NDCG, and inferred AP. These metrics measure different aspects of retrieval effectiveness, with b-pref and inferred AP introduced in the terabyte track to ensure evaluation consistency with full judgments when relevance judgments are reduced. Evaluation metrics generally desire good mathematical properties while producing stable, robust comparison results that are not easily affected by changes in relevance judgments.

Due to space and capability limitations, we have only briefly enumerated part of the current state of IR research. It should be noted that while numerous technologies are being introduced into IR, IR techniques themselves have gradually become foundational technologies that researchers introduce into other fields. For example, some have applied the PageRank technique from IR to software engineering with remarkable results.

6. Conclusion

We conclude with a brief summary of current IR research, which exhibits several fundamental characteristics:

1. **User-centric focus with the goal of improving interactive experience.** The ultimate goal of IR applications is to satisfy user needs, which must drive research directions. Modern IR research emphasizes the user' s central role, driving technological development and exhibiting a clear trend toward personalization.

2. **Harnessing the wisdom of crowds.** Whether considering resources such as microblogs, Wikipedia, ODP, or search logs, all concentrate the wisdom of large user populations. From this data, commonalities can be extracted to improve IR result accuracy. From a methodological perspective, collaborative approaches based on user similarity are considered essential techniques in both filtering and retrieval.

3. **New resources and platforms have spawned new research topics.** Examples include microblog mining and search, sentiment analysis of reviews, mobile search, and advertising recommendation. These new research directions offer both significant practical value and considerable challenges, greatly enriching IR research. Social search based on social resources is attracting widespread attention.

4. **Large-scale data analysis and learning as primary means.** Modern IR research emphasizes user-centric approaches, making large-scale data analysis and learning indispensable. Current IR research integrates technologies from various fields and attracts researchers from diverse backgrounds. We can anticipate that future IR research will become even more vibrant and diverse.

Author Biographies:

Wang Bin: Associate Professor and Ph.D. Supervisor, Institute of Computing Technology, Chinese Academy of Sciences. Email: wangbin@ict.ac.cn

Li Peng: Ph.D. Candidate, Institute of Computing Technology, Chinese Academy of Sciences

Note: Figure translations are in progress. See original paper for figures.