CS6200
Information Retrieval

Jesse Anderton
College of Computer and Information Science
Northeastern University

Major Contributors

Gerard Salton
Vector Space Model
Indexing
Relevance Feedback
SMART

Karen Spärck Jones
IDF
Term relevance
Summarization
NLP and IR

Cyril Cleverdon
Cranfield paradigm:
  Test collections
Term-based retrieval
  (instead of keywords)

William S. Cooper
Defining "relevance"
Query formulation
Probabilistic retrieval

Tefko Saracevic
Evaluation methods
Relevance Feedback
Information needs

Stephen Robertson
Term weighting
Combining evidence
Probabilistic retrieval
Bing

W. Bruce Croft
Bayesian inference
  networks
IR language modeling
Galago
UMass Amherst

C. J. van Rijsbergen
Test collections
Document clustering
Terrier
Glasgow

Susan Dumais
Latent Semantic
  Indexing
Question answering
Personalized search

Open Questions in IR

- Which major research topics in IR are we ready to tackle next? SWIRL 2012 picked (out of 27 suggestions from IR research leaders):

  ‣ **Conversational answer retrieval** – asking for clarification

  ‣ **Empowering users to search more actively** – better interfaces and search paradigms

  ‣ **Searching with zero query terms** – anticipating information needs

  ‣ **Mobile Information Retrieval analytics** – toward test collections for mobile search

  ‣ **Beyond document retrieval** – structured data, information extraction…

  ‣ **Understanding people better** – adapting to user interaction

Open Questions in IR

- Today we'll focus on the following topics:

  ‣ **Conversational Search** – asking for clarification

  ‣ **Understanding Users** – collecting better information on user interaction and needs

  ‣ **Test Collections** – how to create test collections for web-scale and mobile IR evaluation

  ‣ **Retrieving Information** – beyond lists of documents

Conversational Search

Conversational Search | Understanding Users
Test Collections | Retrieving Information

Conversational Search

- In the dominant search paradigm, users run a query, look at the results, then refine the query as needed.

- Can we do better?

  ‣ Good idea: Learn from the way the user refines the query throughout a search session

  ‣ Better idea: Recognize when we're doing badly and ask the user for clarification

  ‣ Even better: Create a new interaction paradigm based on a *conversation* with the user

Inspiration

- A major goal for IR throughout its history is to move toward more natural, "human" interactions

- The success and popularity of recent systems that emulate conversational search shows the potential of this approach

- How can we move toward open-domain conversations between people and machines?
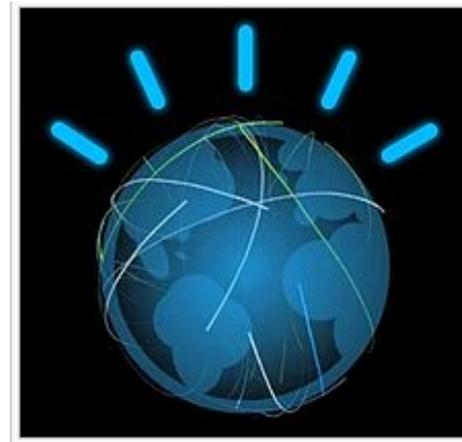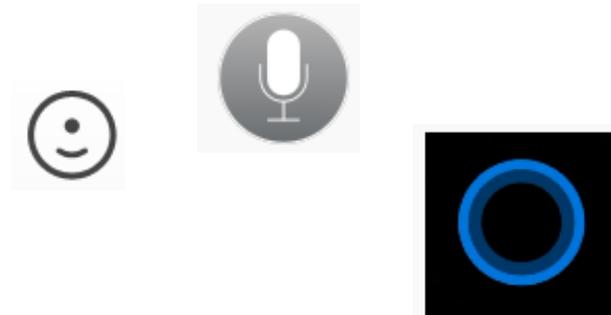
Evi, Siri, Cortana, Watson

Questions

- What does a query look like?

  ‣ IR: a keyword list to stem, stop, and expand

  ‣ QA: a question from a limited set of supported types to parse and pattern match

- We want to support questions posed in arbitrary language, which seems like a daunting task

  ‣ Perhaps understanding arbitrary questions is easier than arbitrary sentences in general?

  ‣ A "question" needs a clear working definition: how is a question represented, after processing by the system? Are we constraining the types of possible user input that count as questions somehow?

Dialog

- Given the initial question, the system should provide an answer and/or ask for clarification.

- What does dialog look like?

  ‣ IR: Query suggestion, query expansion, relevance feedback, faceted search

  ‣ QA: Some natural language dialog, mainly resolving ambiguity (e.g. coreferences)

- Our aim is not only to disambiguate terms, but to discriminate between different information needs that can be expressed in the same language.

- We would also like the system to learn about gaps in its understanding through user interaction. Can the user teach the system?

Answers

- Current answers:

  ‣ IR: document lists, snippets, and passages

  ‣ QA: answers extracted from text; usually "factoids"

- Possible answers include the above, but also summaries, images, video, tables and figures (perhaps generated in response to the query). The ideal answer type depends on the question.

- A ranking of other options should be secondary to the primary answer, not the primary search engine output

Research Challenges

- Improved understanding of natural language semantics

- Defining questions and answers for open domain searching

- Techniques for representing questions, dialog, and answers

- Techniques for reasoning about and ranking answers

- Effective dialog actions for improving question understanding

- Effective dialog actions for improving answer quality

- Expectation: this will take >5 years from multiple research teams

Understanding Users

Conversational Search | **Understanding Users**
Test Collections | Retrieving Information

Understanding Users

- There is a surprisingly large gap between the study of how users interact with search engines and the development of IR systems.

- We typically make simplifying assumptions and focus on small portions of the overall system.

- How can we adjust our systems (and research methodology) to better account for user behavior and needs?

User-based Evaluation

- For example, most evaluation measures currently in use make overly-simplistic assumptions about users

  ‣ In most, relevance gained from documents read does not impact the relevance of future documents

  ‣ Users are assumed to scan the list from top to bottom, and to gain all available relevance from each document they observe

- Current research in evaluation is focusing on refining the user gain and discount functions to make this more realistic

User-based Relevance

- In ad hoc web search, we present users with a ranked list of documents. Document relevance should, arguably, depend on:

  ‣ The user's information need (hard to observe)

  ‣ The order in which the user examines documents

  ‣ Relevant information available in documents the user has opened (hard to specify)

  ‣ The amount of time the user spends in documents they open (easy to measure, correlated with information gain)

  ‣ Whether the query has been reformulated, and whether this document was retrieved in a prior version of the query

User-driven Research

- The community would benefit from much more extensive user studies

  ‣ Consider sets of users ranging from individuals, to groups, to entire communities.

  ‣ Consider methods including ethnography, *in situ* observation, controlled observation, experiment, and large-scale logging.

  ‣ In order to provide guidance for the research community, protocols for these research programs should be clearly defined.

Observing User Interactions

- A possible research protocol for controlled observation of people engaged in interactions with information

    ‣ The specific tasks users will engage in

    ‣ Ethnographic details of the participants

    ‣ Instruments for measuring participants' prior **experience** with IR systems, **expectations** of task difficulty, **knowledge** of search topics, **relevance** gained through interactions, level of **satisfaction** after the task is complete, and **aspects** of the IR system which contributed to that.

Large-scale Logging

- A possible research protocol for large-scale logging of search session interactions

  ‣ No particular user tasks; instead, natural search behavior.

  ‣ Logging the **content** of and **clicks** on the search results page, **context** (time of day, location), and **relevance** indicators (clicks, dwell time, returning to the same page next week)

  ‣ Less helpful for personalization, but more helpful for large-scale statistics on information needs and relevance

Research Challenges

- Research community agreement on protocols

- Addressing user anonymity

- Constructing a resource for evaluation and distribution of the resulting datasets in compatible formats

- Dealing adequately with noisy and sparse data

- Cost of data collection

Test Collections

Test Collections

- IR test collections are crucial resources for advancing the state of the art

- There is a growing need for new types of test collections that have proven difficult to gather:

  ‣ Very large test collections for web-scale search

  ‣ Test collections for new interaction modes used on mobile devices

- Here we will focus on the latter

Mobile Test Collections

- Mobile devices are ubiquitous, and used to perform IR tasks across many popular apps and features.

- However, there is little understanding of mobile information access patterns across tasks, interaction modes, and software applications.

- How can we collect this information?

- Once we have it, how can we use it to enable high-quality research?

Data of Interest

- There are several types of data we'd like to include in a hypothetical mobile test collection

  ‣ The information-seeking task the user carries out

  ‣ Whether the resulting information led to some later action (e.g. buying a movie ticket)

  ‣ Contextual information: location, time of day, mobile device type and platform, application used

  ‣ Cross-app interaction patterns: seeking information from several apps, or acting in app B as a result of a query run in app A

Data Collection

- We can develop a data collection toolkit for application developers to include in their software

- There are obvious privacy concerns here, and the methodology has to be carefully developed

- Ideally, we would persuade major search app developers to include the toolkit

- To protect users, data collection should be anonymized and (perhaps) based on periodic opt-in

- Many people don't mind providing anonymized information to promote social good, such as advancing research, if trust is maintained

Given the data…

- Supposing we could readily collect the data, there is still work to be done to ensure it results in quality research

- Standard research task definitions and evaluation metrics need to be developed, e.g. by TREC

- The task definitions will specify exactly what types of data to collect, the format of that data, and how to distribute the data to research teams

Research Challenges

- Persuading thousands of people to allow their personal usage to be tracked

- Developing data collections with sufficient data to be useful, but which are sufficiently anonymized

- Developing a collection methodology that university ethics boards and mobile device and application developers find acceptable

Retrieving Information

Conversational Search | Understanding Users
Test Collections | **Retrieving Information**

Retrieving Information

- The most widely-used IR task is currently retrieving lists of documents in response to a keyword query.

- However, recent products and usage patterns (mobile platforms, social networks) appear to be disrupting that paradigm

- Some systems have been developed to support factoid question answering, and to integrate structured data into search results.

- Can we improve search results by pulling in linked data, information extraction, collaborative editing, and other structured information?

Motivating Examples

- It is easy to find queries where the information need is not most naturally addressed with a document list:

  ‣ Researching a job applicant's employment history – generating a work-centric biography would be better

  ‣ "How to" queries – a reliable list of instructions would be better

  ‣ "Is Myrtle Beach crowded today?" – presenting data on recent and current beach occupancy patterns is better

Structuring Data

- Even plain text documents have some latent semantic structure.

- Users routinely exploit this structure when they scan through documents to find the information they're seeking.

- Can we identify the structure in documents and use that to inform our query results?

- Can we somehow merge this automatically-structured information with explicitly-structured information from information services?

- Can we extract the relevant information from a document, and merge it with information from other documents?

Crowdsourced Search

- Can we include human intelligence as a component of a search system?

  ‣ We could crowdsource the task of identifying semantic structure in a document

  ‣ We could "friend-source" certain queries, e.g. by asking your friends for movie recommendations on your behalf

Research Challenges

- Keyword queries may be the wrong kinds of questions for this data. We will need to define the query language used in this domain.

- Creating good general representations of structured and unstructured information, and storing that information for fast querying, merging, reasoning, and retrieval on free form queries.

- Keeping a notion of information uncertainty, source reliability, and privacy is important.

- Result presentation – How do we create a useful and aesthetic representation of the results?

- Evaluation – How can we measure result quality, especially when result format can vary? Relatedly, how can we create test collections for this new task?

Summary

- Recommended reading:

  ‣ Recommended Reading for IR Research Students, Alistair Moffat, Justin Zobel, David Hawking (eds.), 2004.

  ‣ Frontiers, Challenges, and Opportunities for Information Retrieval: Report from SWIRL 2012, James Allan, Bruce Croft, Alistair Moffat, and Mark Sanderson (eds.), 2012.