

# INLS 509 Information Retrieval Literature Review

## Objective

The objective of the literature review is for you to gain an in-depth understanding of a particular area of Information Retrieval (IR) that is of interest to **you**.

## Overview

You should form groups of 2-3 people and choose an area of Information Retrieval that you would like to explore in-depth (see examples below). This is a literature review. The goal will be to survey the state of the art in a particular area of IR. You do not need to be familiar with the area. You should pick an area that you are interested in and want to learn more about. This is an opportunity to learn something new! The literature review will be 30% of your grade. You will be responsible for three deliverables:

- Proposal (5%)
- Paper (15%)
- Presentation (10%)

Check the course website for due dates.

## Proposal (5%)

Submit a 2-3 page proposal (one proposal per group). Each proposal must include:

1. The names of every member of the group (2-3 person groups).
2. The IR-related topic you will focus on (see examples below).
3. A preliminary list of research papers you will review (no less than 10, no more than 15 research papers per group). For each paper, provide a 2-3 sentence description for why you chose the paper. Why do you think it is interesting or important?

The goal of the proposal is for me to give you feedback. I will try to point out important work you may have missed and will try to ensure that your chosen area is not too broad or too specific, or that it does not overlap significantly with another group. I will also try to suggest resources (e.g., particular conferences, journals, workshops) where you might find related information.

## Paper (15%)

Your paper should be about 10-15 pages single-spaced and 20-30 pages double-spaced. Please stay with in these page limits. Your paper should include the following information.

1. A description of the problem. What is the problem? Why is it important? Why is it difficult?

2. A survey of how others have attempted to solve the problem. Organize and present the reviewed literature in a way that shows you understand the different approaches to the problem. How are different solutions similar? How are they different? Do different solutions make different assumptions? What are those assumptions? How have solutions evolved with time? How have different approaches built on each other?
3. A survey of evaluation. How are solutions to the problem typically evaluated? Are there agreed upon metrics that are suitable to the task? What assumptions do these metrics make? Are there any specific conferences or workshops that specialize on the task?
4. What do you think? How do you see this area of IR progressing? What are the key problems that remain to be solved? How do you think this technology will change people's lives? Your own view of the problem is very important. Make sure to allocate about 20% of your paper and presentation to this.

## Presentation (10%)

Each group will present their literature review to the class. This is mainly so that the other students can learn you! To accommodate everyone, each presentation should last no more than 10 minutes + 2 minutes of questions from the audience (1 or 2 questions per presentation). You should prepare slides and the expectation is that every group member will contribute to the content of the presentation. You are free to divide the oral presentation between the group members however you like. The presentations will take place during the last two days of class (Monday, November 28 and Wednesday, November 30).

## Example Topics

Some potential areas you might consider include the following. These are only examples. You do not have to choose a topic from this list.

- **Personalized IR:** How can IR systems use information about a user's interests to improve their search experience?
- **Session-based IR:** How can IR systems leverage information from a user's previous retrievals (i.e., same-session queries) to improve their search experience?
- **Clustering of Search Results:** Should systems display results organized by topic or some other dimension rather than a one-dimensional list of results? Are there particular search tasks for which this presentation of results is more appropriate?
- **Book Search:** How do IR systems perform book search? Books are associated with lots of text.
- **Multimedia Search:** How do IR systems perform search for multimedia items that are not inherently associated with text (e.g., images, video) given a textual query?
- **Social Media Data for Forecasting:** Can twitter data be used to predict events in the real world (e.g., the stock market, whether a new drug is causing a certain side-effect, etc.)?
- **Faceted Search:** Many documents have metadata (e.g., publication date, publisher, topic). Can a search system predict when to allow users to filter results based on certain metadata fields?

- **Mobile Search:** How is search from mobile devices different than search from a desktop? Certainly, the interaction is different. Are the information needs different? And, if so, what are systems doing to address this?

## Where to look

Each chapter in the Croft, Metlzer, and Strohan book concludes with a section entitled “References and Further Reading”, which lists research papers related to the chapter that may not have been covered in detail. This is a good place to start. Other books on information retrieval have similar sections. In particular, you may also want to look at [Introduction to Information Retrieval](#), which is freely available online.

In addition, you will want to review the work published at relevant conferences. Here are some conference you might consider.

- Information Retrieval: SIGIR (Information Retrieval), CIKM (Information and Knowledge Management), WWW (World Wide Web), WSDM (Web Search and Data Mining), TREC (Text Retrieval), INEX (XML Retrieval)
- Digital Libraries and Information Science: JCDL (Digital Libraries), ASIST (Information Science and Technology)
- Natural Language Processing: ACL (Computational Linguistics), NAACL (Computational Linguistics), HLT (Human Language Technologies), TAC (Text Analysis)
- Human-Computer Interaction: CHI (Computer-Human Interaction), Ubicomp (Ubiquitous Computing)
- Computer-Supported Collaboration and Learning: CSCW (Computer-Supported Collaborative Work), CSCL (Computer-Supported Collaborative Learning)
- Social-Media: ICWSM (Weblogs and Social Media)

The proceedings for these conferences (with the exception of TREC, INEX and TAC) are available through the [Association of Computing Machinery \(ACM\) Digital Library](#). You should have access to the ACM DL from within the UNC network.

## Tips

- Be scientific. Think broadly about your problem and try to make connections with other areas of IR that may be called something different. Connect the dots! And, there are many dots.
- Be constructive. Every research makes simplifying assumptions (whether or not they are stated explicitly or are implicit). Assumptions (even if at first glance seem unrealistic) are oftentimes necessary and afford analyses that are useful. Being “destructive” is easy (and not impressive). Focus on the pros and cons.
- Be cautious. Be clear about what you know and do not know. Avoid a controversial claim, unless you can support it with a reference.

- When you review a paper, focus on the content and not on the writing. Our goal is not to critique a paper's writing style or organization. The focus should be on different solutions to the problem and how they were evaluated.
- Have fun.