Your Name:

Your Andrew ID:

Homework 1

0 Introduction

0.1 Collaboration and Originality

1. Did you receive help of any kind from anyone (other than the instructor or TAs) in developing your software for this assignment (Yes or No)? If you answered Yes, provide the name(s) of anyone who provided help, and describe the type of help that you received.

2. Did you give help of any kind to anyone in developing their software for this assignment (Yes or No)? If you answered Yes, provide the name(s) of anyone that you helped, and describe the type of help that you provided.

3. Did you examine anyone else's software for this assignment (Yes or No)? Do not describe software provided by the instructor.

4. Are you (or the course instructor) the author of every line of source code submitted for this assignment (Yes or No)? If you answered No:
    a. identify the software that you did not write,
    b. explain where it came from, and
    c. explain why you used it.

5. Are you the author of every word of your report (Yes or No)? If you answered No:
    a. identify the text that you did not write,
    b. explain where it came from, and
    c. explain why you used it.

0.2 Format

Instructions are shown in this red italic bold font. Do not include instructions in your report. For example, delete this subsection, and in the next section, delete the instruction paragraphs.

Leave the page breaks between sections, as shown in this template. For example, Sections 1.1 and 1.2 must be on different pages.

There is a 2 point deduction for not following format instructions because it creates extra work during grading.

1 Structure in Boolean Queries

1.1 Structure in Boolean Queries: Queries

Show the structured queries used in your experiments. Briefly describe the thinking behind each structured query, i.e., what it was intended to do or why you thought it would work well. Hint: Probably a sentence or two is all you need for each query.

1.2 Structure in Boolean Queries: Experimental Results

Present your experimental results for each query set, in the format shown below.

Your .zip / .tgz file must include files named HW1-Exp-1.1a.qry, HW1-Exp-1.1a.param, etc., in the QryEval directory. The experimental results shown below must be reproducible by these files.

	Unranked Boolean		
	BOW #AND (Exp-1.1a)	**BOW #NEAR/3 (Exp-1.1b)**	**Structured (Exp-1.1c)**
MRR	0.0000	0.0000	0.0000
P@10	0.0000	0.0000	0.0000
P@20	0.0000	0.0000	0.0000
P@30	0.0000	0.0000	0.0000
MAP	0.0000	0.0000	0.0000
Run Time	mm:ss	mm:ss	mm:ss

	Ranked Boolean		
	BOW #AND (Exp-1.2a)	**BOW #NEAR/3 (Exp-1.2b)**	**Structured (Exp-1.2c)**
MRR	0.0000	0.0000	0.0000
P@10	0.0000	0.0000	0.0000
P@20	0.0000	0.0000	0.0000
P@30	0.0000	0.0000	0.0000
MAP	0.0000	0.0000	0.0000
Run Time	mm:ss	mm:ss	mm:ss

1.3 Structure in Boolean Queries: Analysis

Discuss the value of structured queries for the two Boolean ranking algorithms. There are a wide range of issues that you might discuss. You may discuss what seems most appropriate or interesting from your experiments. Some example issues are interactions between type of query with type of algorithm; accuracy vs. running time; accuracy vs. time spent forming queries; the behavior / usefulness of the different query operators and fields, and your success and failure at using them in structured queries; and which operators or fields are more or less useful. You have the freedom to discuss what you find interesting. These are just some examples to help your thinking.

Do not just summarize the results – we can read the tables. Instead, explain what conclusions you can reach based on the experiment. This is your chance to show what you learned from the experiments. We are primarily interested in your observations about general trends, not quirky queries. Usually a good analysis addresses several issues. This advice applies to all of the analysis sections.

Recommended length: Less than ¾ of a page.

2 Document Representations

2.1 Document Representations: Parameters

Document the parameters used to obtain these results (ranking lengths, BM25 parameters, etc).

2.2 Document Representations: Experimental Results

	Ranked Boolean				
	body (Exp-1.2a)	**title (Exp-2.1b)**	**url (Exp-2.1c)**	**keywords (Exp-2.1d)**	**inlink (Exp-2.1e)**
MRR	0.0000	0.0000	0.0000	0.0000	0.0000
P@10	0.0000	0.0000	0.0000	0.0000	0.0000
P@20	0.0000	0.0000	0.0000	0.0000	0.0000
P@30	0.0000	0.0000	0.0000	0.0000	0.0000
MAP	0.0000	0.0000	0.0000	0.0000	0.0000
Run Time	mm:ss	mm:ss	mm:ss	mm:ss	mm:ss

	BM25				
	body (Exp-2.2a)	**title (Exp-2.2b)**	**url (Exp-2.2c)**	**keywords (Exp-2.2d)**	**inlink (Exp-2.2e)**
MRR	0.0000	0.0000	0.0000	0.0000	0.0000
P@10	0.0000	0.0000	0.0000	0.0000	0.0000
P@20	0.0000	0.0000	0.0000	0.0000	0.0000
P@30	0.0000	0.0000	0.0000	0.0000	0.0000
MAP	0.0000	0.0000	0.0000	0.0000	0.0000
Run Time	mm:ss	mm:ss	mm:ss	mm:ss	mm:ss

The Exp-1.2a result from Section 1.2 is used as a baseline for comparison.

2.3 Document Representations: Analysis

Discuss the behavior / usefulness of the different fields. You may find it useful to compare the behavior of different queries and/or documents to develop a better understanding of what the different fields tend to contain and how effectively the two algorithms use them.

3 Multifield BM25

3.1 Multifield BM25: Parameters

Document the parameters used to obtain these results (ranking lengths, BM25 parameters, field weights, etc). Briefly explain the thinking behind your field weights. Hint: Parameter sweeps don't demonstrate much thought.

3.2 Multifield BM25: Experimental Results

	body (Exp-2.2a)	**0.00 inlink 0.00 url 0.00 keywords 0.00 title 0.00 body** (Exp-3.1b)	**0.00 inlink 0.00 url 0.00 keywords 0.00 title 0.00 body** (Exp-3.1c)	**0.00 inlink 0.00 url 0.00 keywords 0.00 title 0.00 body** (Exp-3.1d)	**0.00 inlink 0.00 url 0.00 keywords 0.00 title 0.00 body** (Exp-3.1e)
P@10	0.0000	0.0000	0.0000	0.0000	0.0000
P@20	0.0000	0.0000	0.0000	0.0000	0.0000
P@30	0.0000	0.0000	0.0000	0.0000	0.0000
MAP	0.0000	0.0000	0.0000	0.0000	0.0000
R@100	0.0000	0.0000	0.0000	0.0000	0.0000
R@1000	0.0000	0.0000	0.0000	0.0000	0.0000

The Exp-2.2a result from Section 2.2 is used as a baseline for comparison.

3.3 Multifield BM25: Analysis

Discuss the value of different combinations of document representations. You may find it useful to compare the behavior of different queries and/or documents to develop a better understanding of how your combinations behaved.