## Your Name: ADITYA GAUTAM

**Your Andrew ID: agautam1@andrew.cmu.edu**

**Homework 2**

# Collaboration and Originality

Your report must include answers to the following questions:

1. Did you receive help of any kind from anyone in developing your software for this assignment (Yes or No)? It is not necessary to describe discussions with the instructor or TAs.

**Ans) No**

If you answered Yes, provide the name(s) of anyone who provided help, and describe the type of help that you received.

1. Did you give help of any kind to anyone in developing their software for this assignment (Yes or No)?

**Ans) No**

If you answered Yes, provide the name(s) of anyone that you helped, and describe the type of help that you provided.

1. Are you the author of every line of source code submitted for this assignment (Yes or No)? It is not necessary to mention software provided by the instructor.

**Ans) Yes**

If you answered No:

* 1. identify the software that you did not write,
  2. explain where it came from, and
  3. explain why you used it.

1. Are you the author of every word of your report (Yes or No)?

**Ans) Yes**

If you answered No:

* 1. identify the text that you did not write,
  2. explain where it came from, and
  3. explain why you used it.

## Your Name: Aditya Gautam

**Your Andrew ID: agautam1@andrew.cmu.edu**

**Homework 2**

# Experiment 1: Baselines

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Ranked**  **Boolean** | **BM25**  **BOW** | **Indri**  **BOW** |
| **P@10** | 0.1700 | 0.4200 | 0.4000 |
| **P@20** | 0.2800 | 0.3500 | 0.4700 |
| **P@30** | 0.3367 | 0.3667 | 0.4233 |
| **MAP** | 0.1071 | 0.1985 | 0.2057 |

# Experiment 2: Queries with Synonyms and Phrases

## Queries

**Ranked Boolean Query Set :**

69:#OR(#NEAR/3(sewing instruction) #NEAR/3(sewing guidelines) #OR(sewing.title) #OR(sewing.url)

79:#OR(voyager.url voyager.title voyager.keywords voyager.body)

84:#OR(#NEAR/3(continental plates) #NEAR/2(plate continental)) #OR(continental.url)

89:#OR(#NEAR/1(Obsessive-compulsive disorder)) #OR(ocd.url ocd.title)

108:#NEAR/1(ralph owen brewster) #AND(#OR(brewster.url) #OR(owen.url))

141:#AND(va #NEAR/5(dmv registration)) #AND(va #NEAR/5(registration dmv)) #AND(virginia #NEAR/6(dmv registration)) #NEAR/6(virginia motor registration)

146:#OR(#NEAR/1(sherwood regional library)) #OR(#NEAR/2(regional library sherwood))

153:#OR(pocono.url pocono.title pocono.body)

171:#OR(#NEAR/1(ron howard) #NEAR/1(ron.title howard.title)) #OR(RonHoward.url #AND(Ron.url Howard.url))

197:#OR(#NEAR/40(idaho state flower))

**BM25 Query Set :**

69:#NEAR/3(sewing #SYN(instruction guidelines)) #NEAR/6(#SYN(instruction guidelines) sewing)

79:voyager.body voyager.title

84:#NEAR/1(#SYN(tectonic continental) plates) #NEAR/3(plates #SYN(tectonic continental))

89:#NEAR/30(Obsessive compulsive disorder) ocd

108:#NEAR/1(ralph owen brewster) #NEAR/3(ralph brewster) #NEAR/1(owen brewster)

141:#NEAR/6(#SYN(va virginia) #SYN(dmv driving motor vehicle) registration) #NEAR/6(#SYN(va virginia) motor registration)

146:#NEAR/1(sherwood regional library)

153:pocono.body #NEAR/1(pocono.title mountains.title) #NEAR/1(pocono mountains)

171:#NEAR/1(ron howard) #NEAR/1(ron.title howard.title)

197:#NEAR/40(idaho state flower) #NEAR/1(Philadelphus lewisii) #NEAR/1(Philadelphus.title lewisii.title) Syringa.body Syringa.title #NEAR/1(idaho.title state.title flower.title)

**Indri Query Set :**

69:#NEAR/3(sewing #SYN(instruction guidelines)) #NEAR/6(#SYN(instruction guidelines) sewing)79:voyager.body voyager.url

84:#NEAR/3(#SYN(tectonic continental) plates) #NEAR/3(plates #SYN(tectonic continental))

89:#NEAR/120(Obsessive compulsive disorder) ocd.body

108:#NEAR/1(ralph owen brewster) #NEAR/3(ralph brewster) #NEAR/1(owen brewster)

141:#NEAR/40(#SYN(va virginia) #SYN(dmv driving motor vehicle) registration) #NEAR/35(virginia motor registration) #NEAR/35(va motor registration) #NEAR/35(virginia dvm registration) #NEAR/30(va dvm registration)

146:#NEAR/1(sherwood regional library)

153:pocono.body #NEAR/1(pocono.title mountains.title) #NEAR/1(pocono mountains)

171:#NEAR/1(ron howard) #NEAR/1(ron.title howard.title)

197:#NEAR/40(idaho state flower) #NEAR/1(Philadelphus lewisii) #NEAR/1(Philadelphus.title lewisii.title) Syringa.body Syringa.title #NEAR/1(idaho.title state.title flower.title)

## Query descriptions

For each query, provide a brief (1-2 sentences) description that identifies which strategy was used for that query, any important deviations from your default strategies, and your intent, i.e., why you thought that particular structure was a good choice.

## Experimental Results

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Ranked**  **Boolean** | **BM25**  **BOW** | **Indri**  **BOW** | **Ranked**  **Boolean**  **Syn/Phr** | **BM25**  **Syn/Phr** | **Indri**  **Syn/Phr** |
| **P@10** | 0.1700 | 0.4200 | 0.4000 | 0.5500 | 0.0000 | 0.0000 |
| **P@20** | 0.2800 | 0.3500 | 0.4700 | 0.6550 | 0.0000 | 0.0000 |
| **P@30** | 0.3367 | 0.3667 | 0.4233 | 0.6333 | 0.0000 | 0.0000 |
| **MAP** | 0.1071 | 0.1985 | 0.2057 | 0.3200 | 0.0000 | 0.0000 |

## Discussion

Discuss any trends that you observe; whether the use of synonyms and phrases behaved as you expected; and any other observations that you may have.

# Experiment 3: BM25 Parameter Adjustment

## k1

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **k1** | | | | | | | |
| 1.2 | 0 | 0.1 | 0.4 | 0.8 | 1 | 1.5 | 5 |
| **P@10** | 0.6100 | 0.4900 | 0.5600 | 0.5600 | 0.5600 | 0.5600 | 0.5700 | 0.5500 |
| **P@20** | 0.5850 | 0.4500 | 0.5700 | 0.5700 | 0.5750 | 0.5700 | 0.5650 | 0.5400 |
| **P@30** | 0.5400 | 0.5333 | 0.5367 | 0.5367 | 0.5333 | 0.5367 | 0.5267 | 0.5233 |
| **MAP** | 0.2964 | 0.2184 | 0.2941 | 0.2949 | 0.2917 | 0.2926 | 0.2933 | 0.2905 |

.

## b

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **b** | | | | | | | |
| 0.75 | 0 | 0.1 | 0.2 | 0.4 | 0.6 | 0.8 | 1.0 |
| **P@10** | 0.6100 | 0.5600 | 0.5500 | 0.5200 | 0.5200 | 0.5500 | 0.5500 | 0.5700 |
| **P@20** | 0.5850 | 0.5950 | 0.5900 | 0.5750 | 0.5400 | 0.5550 | 0.5600 | 0.5550 |
| **P@30** | 0.5400 | 0.5900 | 0.5800 | 0.5700 | 0.5467 | 0.5500 | 0.5300 | 0.5300 |
| **MAP** | 0.2964 | 0.2964 | 0.3048 | 0.3013 | 0.3008 | 0.3003 | 0.2903 | 0.2737 |

## Discussion

Explain your reasons for choosing the values that you tested, and how those reasons are related to how BM25 works. Discuss any changes in retrieval performance that you observed, and the significance of any trends that you observed.

# Indri Parameter Adjustment

## µ

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **µ** | | | | | | | |
| 2500 | 0 | 500 | 1000 | 1500 | 2000 | 3000 | 5000 |
| **P@10** | 0.5400 | 0.5100 | 0.5900 | 0.5600 | 0.5700 | 0.5700 | 0.5600 | 0.5900 |
| **P@20** | 0.6050 | 0.5400 | 0.5450 | 0.5700 | 0.5900 | 0.5900 | 0.6100 | 0.6200 |
| **P@30** | 0.6000 | 0.5200 | 0.5433 | 0.5567 | 0.5800 | 0.5900 | 0.6033 | 0.6067 |
| **MAP** | 0.3206 | 0.2831 | 0.3117 | 0.3165 | 0.3203 | 0.3198 | 0.3187 | 0.3197 |

## 𝜆

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **𝜆** | | | | | | | |
| 0.4 | 0 | 0.1 | 0.2 | 0.5 | 0.6 | 0.8 | 1 |
| **P@10** | 0.5400 | 0.5400 | 0. | 0.5400 | 0.0000 | 0.5400 | 0.5300 | 0.5700 |
| **P@20** | 0.6050 | 0.6050 | 0.0000 | 0.6050 | 0.0000 | 0.6000 | 0.5950 | 0.5200 |
| **P@30** | 0.6000 | 0.6000 | 0.0000 | 0.6000 | 0.0000 | 0.6000 | 0.6000 | 0.4933 |
| **MAP** | 0.3206 | 0.3204 |  | 0.3205 | 0.0000 | 0.3199 | 0.3197 | 0.2455 |

## Discussion

Explain your reasons for choosing the values that you tested, and how those reasons are related to how Indri works. Discuss any changes in retrieval performance that you observed, and the significance of any trends that you observed.