**Your Name: Jiachen Li**

**Your Andrew ID: jiachenl**

**Homework 3**

# Statement of Assurance

I certify that all of the materials I submit are original works that were done by myself.

# Experiment 1: Baselines

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Ranked**  **Boolean** | **BM25**  **BOW** | **Indri**  **BOW** |
| **P@10** | 0.1500 | 0.2900 | 0.2400 |
| **P@20** | 0.1800 | 0.3050 | 0.2750 |
| **P@30** | 0.1667 | 0.3267 | 0.2967 |
| **MAP** | 0.0566 | 0.1325 | 0.1275 |
| **Time** | 00:09 | 00:09 | 00:09 |

# Experiment 2: Different representations

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Indri**  **BOW**  **(body)** | **0.20 url**  **0.20 keywords**  **0.20 title**  **0.20 body**  **0.20 inlink** | **0.05 url**  **0.10 keywords**  **0.20 title**  **0.60 body**  **0.05 inlink** | **0.05 url**  **0.20 keywords**  **0.10 title**  **0.60 body**  **0.05 inlink** | **0.05 url**  **0.10 keywords**  **0.05 title**  **0.60 body**  **0.20 inlink** | **0.20 url**  **0.05 keywords**  **0.10 title**  **0.60 body**  **0.05 inlink** | **0.05 url**  **0.05 keywords**  **0.05 title**  **0.80 body**  **0.05 inlink** |
| **P@10** | 0.2400 | 0. 1100 | 0.2100 | 0.2100 | 0.1800 | 0.2200 | 0.2300 |
| **P@20** | 0.2750 | 0.1600 | 0.2650 | 0.2600 | 0.1950 | 0.2650 | 0.2700 |
| **P@30** | 0.2967 | 0.1633 | 0.2533 | 0.2500 | 0.1933 | 0.2533 | 0.2633 |
| **MAP** | 0.1275 | 0.0853 | 0.1103 | 0.1113 | 0.0934 | 0.1125 | 0.1159 |
| **Time** | 00:09 | 00:09 | 00:09 | 00:09 | 00:09 | 00:09 | 00:09 |

Describe your strategy for setting the weights on the different representations.

Discuss any trends that you observe; whether the different representations behaved as you expected; the Precision and Recall characteristics of each representation; how the differences in accuracy (if any) relate to different computational cost; and any other observations that you may have.

# Experiment 3: Sequential dependency models

**Example Query:** Provide your structured query for query “fickle creek farm”.

#WAND(

0.3 #AND(fickle creek farm)

0.5 #AND(#NEAR/1(fickle creek) #NEAR/1(creek farm))

0.2 #AND(#WINDOW/8(fickle creek) #WINDOW/8(creek farm)) )

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Indri**  **BOW**  **(body)** | **0.50 AND**  **0.30 NEAR**  **0.20 WINDOW** | **0.50 AND**  **0.40 NEAR**  **0.10 WINDOW** | **0.60 AND**  **0.30 NEAR**  **0.10 WINDOW** | **0.45 AND**  **0.45 NEAR**  **0.10 WINDOW** | **0.30 AND**  **0.50 NEAR**  **0.20 WINDOW** |
| **P@10** | 0.2400 | 0.3700 | 0.3700 | 0.3500 | 0.3700 | 0.3700 |
| **P@20** | 0.2750 | 0.3600 | 0.3700 | 0.3600 | 0.3700 | 0.3750 |
| **P@30** | 0.2967 | 0.3633 | 0.3567 | 0.3500 | 0.3567 | 0.3667 |
| **MAP** | 0.1275 | 0.1842 | 0.1852 | 0.1807 | 0.1873 | 0.1909 |
| **Time** | 00:09 | 00:42 | 00:42 | 00:42 | 00:42 | 00:43 |

Describe how you set the weights for the different components of the sequential dependency model.

Discuss any trends that you observe; whether the more complex query behaved as you expected; whether the improvement in accuracy (if any) is worth the increased computational cost; and any other observations that you may have.

# Experiment 4: Multiple representations + SDMs

**Example Query:** Provide your structured query for query “fickle creek farm”.

#WAND(

0.1 #AND(

#WSUM(0.05 fickle.url 0.05 fickle.keywords 0.05 fickle.title 0.8 fickle.body 0.05 fickle.inlink) #WSUM(0.05 creek.url 0.05 creek.keywords 0.05 creek.title 0.8 creek.body 0.05 creek.inlink) #WSUM(0.05 farm.url 0.05 farm.keywords 0.05 farm.title 0.8 farm.body 0.05 farm.inlink))

0.9 #WAND(

0.3 #AND(fickle creek farm)

0.5 #AND(#NEAR/1(fickle creek) #NEAR/1(creek farm))

0.2 #AND(#WINDOW/8(fickle creek) #WINDOW/8(creek farm))) )

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Indri**  **BOW**  **(body)** | **w=1.0**  **(Exp 2)** | **w1=0.9** | **w2=0.7** | **w3=0.5** | **w4=0.3** | **w5=0.1** | **w=0.0**  **(Exp 3)** |
| **P@10** | 0.2400 | 0.2300 | 0.2500 | 0.2900 | 0.3400 | 0.3600 | 0.3700 | 0.3700 |
| **P@20** | 0.2750 | 0.2700 | 0.3000 | 0.3200 | 0.3400 | 0.3650 | 0.3750 | 0.3750 |
| **P@30** | 0.2967 | 0.2633 | 0.3033 | 0.3233 | 0.3333 | 0.3567 | 0.3600 | 0.3667 |
| **MAP** | 0.1275 | 0.1159 | 0.1292 | 0.1502 | 0.1635 | 0.1841 | 0.1883 | 0.1909 |
| **Time** | 00:09 | 00:21 | 00:22 | 00:22 | 00:22 | 00:21 | 00:22 | 00:52 |

Discuss any trends that you observe; whether the more complex query behaved as you expected; whether the improvement in accuracy (if any) is worth the increased computational cost; and any other observations that you may have.