Your Name

Your Andrew ID

Homework 3

Collaboration and Originality

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.
   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)?
   No
   If you answered Yes, provide the name(s) of anyone that you helped, and describe the type of help that you provided.

3. 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.
   Yes.
   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.

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

Your Name

Leonardo Neves

Your Andrew ID

lneves

Homework 3

1 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

Document the parameter settings that were used to obtain these results.

2 Experiment 2: Different representations

Example Query: Provide your structured query for query "sherwood regional library".

	Indri BOW (body)	0.10 url 0.10 keywords 0.50 title 0.20 body 0.10 inlink	0.20 url 0.10 keywords 0.20 title 0.40 body 0.10 inlink	0.15 url 0.10 keywords 0.15 title 0.35 body 0.25 inlink	0.05 url 0.00 keywords 0.05 title 0.90 body 0.00 inlink	0.01 url 0.01 keywords 0.01 title 0.96 body 0.01 inlink
P@10	0.4000	0.3700	0.3900	0.3800	0.4000	0.4000
P@20	0.4700	0.3800	0.4100	0.3600	0.4600	0.4650
P@30	0.4233	0.3667	0.4167	0.3600	0.4200	0.4233
MAP	0.2057	0.1391	0.1756	0.1566	0.2026	0.2055

Describe your strategy for setting the weights on the different representations. Describe how you expected the different weight combinations to perform (before the experiment), and why.

Discuss the trends and stability that you observed in Experiment 2; whether the different weight combinations behaved as you expected; how Precision and Recall tended to vary as the weights were varied; how the differences in accuracy (if any) relate to different computational cost; and your other observations, interpretations, or conclusions from running this experiment.

For the first set of weights, I have given more weight to the title and body as I thought an important website talking about a specific topic would have important information on content based fields, looking for improving the precision. This was not true, since precision got worse. For the second part, I tried making the title weight smaller and making the url more relevant to see if we were able to find websites were the main focus would be on the url and the content would be related, like something as www.sherwoodlibrary.com, improving the recall value. The results for MAP were slightly better and the low rank precision improved, but the overall result was still much lower than the baseline. The third experiment looked for relevance on the inlink field. The results were worse, so probably this field is not as interesting for the given information needs as expected since the overall accuracy is low. Because our baseline only used the body field, I have then tried to use almost all my weight on the body and just give a small relevance weight to title and url. This made the baseline MAP and p@n worse, meaning those fields can only hurt for the given query set. I have tried to extend this reasoning, now using 0.96 for the body and uniformly dividing the rest among the other fields. Again, our results got worse than the baseline. For this query, it is clear that using different representation and weights per field only makes the baseline worse and the computational cost higher, since we now have to deal with different fields. No additional observations could be made, once no distribution of field weights seems to outperform the baseline of only considering the body field.

3 Experiment 3: Sequential dependency models

Example Query: Provide your structured query for query "sherwood regional library".

	Indri BOW (body)	**0.70 AND 0.20 NEAR 0.10 WINDOW**	**0.40 AND 0.30 NEAR 0.30 WINDOW**	**0.20 AND 0.70 NEAR 0.10 WINDOW**	**0.00 AND 0.00 NEAR 0.00 WINDOW**	**0.20 AND 0.40 NEAR 0.40 WINDOW**
P@10	0.4000	0.4400	0.4700	0.5000	0.5100	0.4900
P@20	0.4700	0.5500	0.5550	0.5350	0.5600	0.5650
P@30	0.4233	0.5467	0.5533	0.5300	0.5167	0.5500
MAP	0.2057	0.2895	0.3012	0.3001	0.2944	0.3022

Describe your strategy for setting the weights for the different components of the sequential dependency model. Describe how you expected the different weight combinations to perform (before the experiment), and why.

Discuss the trends and stability that you observed in Experiment 3; how Precision and Recall tended to vary as the weights were varied; whether the more complex query behaved as you expected; whether the improvement in accuracy (if any) is worth the increased computational cost; and your other observations, interpretations, or conclusions from running this experiment.

The first experiment gave a higher weight for the AND operator in order to achieve a higher recall. We can see that, not only this goal was achieved since we increased MAP and recall based metrics a lot, but also a higher precision at all levels, what shows that this combination (and this model) perform much better than the baseline for accuracy too. In order to assess how important the bow part of the SDM is, the AND weight, we made all of the weights to be closer to each other. From this, a higher precision at lower

ranks and a smaller recall and accuracy was expected. The improvements on P@10 and MAP were as expected, but the recall metrics were not as low as one might have thought, showing that, at least for the given information needs and queries, bigram and context weights are extremely beneficial. In order to evaluate the weights for bigram and context parts, the next two experiments give a higher weight for them. The first gives higher weight for the bigram part, giving 0.7 for this weight. This made the precision at the first ranks to increase, but making the precision on the higher ranks to decrease and the overall MAP to decrease too. Also, the recall based metrics got worse, showing this part is more useful for finding the top documents but it might miss some relevant ones. The context weights showed the same behavior. The difference was more on having a smaller MAP but having a higher recall and better metrics at top ranks like P@10 and P@20. Since both experiments showed this and we know that the BOW weight is more relevant for recall, the last experiment aimed on getting the better precision at top levels possible, like a web search engine would try. We give a higher weight to the bigram and context weights than to the bow weight. This made the best MAP results and a more stable value for the precisions over the top 30 documents. The P@5 was the highest but, for some reason, precision at 10 didn't follow this pattern. The overall accuracy also increased and the computational cost was similar for all the experiments. As an additional comment, I would definitely use this method over the first one for any precision based problem that involves the present information needs.