MP3—Retrieval Functions

1. Copy and paste your implementation of each ranking algorithm, together with the corresponding final MAP/P@10/MRR/NDCG@10 performance you get from each ranking function. *Use the default parameter settings suggested here* (30pts)

Boolean model:

```
protected float score(BasicStats stats, float termFreq, float docLength) {
        if (termFreq >0)
          return 1;
        return 0;
    }
```

TF-IDF:

```
protected float score(BasicStats stats, float termFreq, float docLength) {
      float df = stats.getDocFreq();
      float tf = (float) (1 + Math.log10(termFreq));
      float idf = (float) Math.log10(( stats.getNumberOfDocuments() + 1)/
df);

      return tf * idf;
    }
```

Okapi BM25:

```
protected float score(BasicStats stats, float termFreq, float docLength) {
      //default value k1=1.5 k2=750 and b=1.0;
      float k1 = 1.5f;
      float k2 = 750.0f;
      float b = 1.0f;
      float df = stats.getDocFreq();
      float cwq = 1.0f;  //assume that the query term frequency is always one
      float n_avg = stats.getAvgFieldLength();
```

```
      float term1 = (float) Math.log((stats.getNumberOfDocuments() - df +
0.5) / (df + 0.5));
      float term2 = (float) (k1 + 1) * termFreq / (k1 * (1 - b + b *
docLength / n_avg) + termFreq);
      float term3 = (float) (k2 + 1) * cwq / (k2 + cwq);

      return term1 * term2 * term3;
    }
```

Pivoted Length Normalization

```
 protected float score(BasicStats stats, float termFreq, float docLength) {
      float s = 0.75f; // default_s = 0.75
      float cwq = 1.0f; //assume that the query term frequency is always one.
      float df = stats.getDocFreq();
      float n_avg = stats.getAvgFieldLength();

      float term1 = (float) (1 + Math.log(1 + Math.log(termFreq))) / (1 - s +
s * docLength / n_avg );
      float term2 = cwq * (float) Math.log((stats.getNumberOfDocuments() + 1)
/ df );

      return term1 * term2;
    }
```

Jelinek-Mercer Smoothing

```
protected float score(BasicStats stats, float termFreq, float docLength) {
      float lamda = 0.1f; //default_lamda = 0.1
      float alpha_d = lamda;
      float Pwd = (float) ((1 - lamda) * termFreq / docLength + lamda *
model.computeProbability(stats));

      return (float) (Math.log10(Pwd / (alpha_d *
model.computeProbability(stats) )) + queryLength * Math.log10(alpha_d));
    }
```

Dirichlet Prior Smoothing

```
protected float score(BasicStats stats, float termFreq, float docLength) {
      float mu = 2500.0f; //default_mu is 2500
      float alpha_d = mu / (mu + docLength);
      float Pwd = (float) (termFreq + mu * model.computeProbability(stats)) /
(docLength + mu);
        return (float) (Math.log10(Pwd / (alpha_d *
model.computeProbability(stats) )) + queryLength * Math.log10(alpha_d));
    }
```

	MAP	P@10	MRR	NDCG
Boolean	0.1765941441732281	0.2881720430107527	0.5943108713270008	0.3498102343135518
TF-IDF	0.22120903844107404	0.3559139784946237	0.6790342284055105	0.4212087244860364
Okapi BM25	0.1867936950148781	0.30752688172043013	0.5933650504215023	0.3681466348357197
Pivoted Length	0.1277415591745546	0.23870967741935492	0.4339758877638715	0.27300760082406833
Jelinek-Mercer	0.2257331643786167	0.34193548387096767	0.6779061335038565	0.4201580145765151
Dirichlet Prior	0.122924162394069	0.1946236559139785	0.46957670675011975	0.24092906541224793

// I use ***defaultNumResults*** = 50 in SearchQuery.java.

2. Please carefully tune the parameters in BM25 and Dirichlet prior smoothed Language Model. Report the best MAP you have achieved and corresponding parameter settings. (20pts)

- BM25:

  k1∈[1.2,2], k2∈(0,1000], b∈[0.75,1.2]

	k1=1.2	k1=1.3	k1=1.4	k1=1.5
b=0.75	**0.23022684043939504**	0.22865027177571 07	0.22643940000879395	0.22554750800092466
b=0.8	0.22616653837356337	0.22458869591306427	0.22131837568832133	0.2247326822947187
b=0.9	0.2176817871892733	0.21302459798005335	0.20872848260687513	0.21930713244907674
b=1.0	0.20098179079433823	0.19622360507078895	0.19006621180414363	0.1867936950148781
b=1.1	0.18587733317251293	0.17786539763406212	0.17203285233412954	0.16650218800795777

MAP_best = 0.23022684043939504 when k1=1.2, b= 0.75. k2 is not sensitive to the result(maybe because our query is short).

- Dirichlet prior smoothed Language Model

mu value	**mu= 400**	**mu=500**	**mu=600**	**mu=625**

MAP	0.1442278132241201	0.1429112559100837	**0.14757270211755583**	0.1469277331730575
mu value	**mu=650**	**mu=675**	**mu=700**	**mu=725**
MAP	0.14695158128620875	0.14706909729081027	0.14634030590631675	0.14579466994471588
mu value	**mu= 750**	**mu= 775**	**mu= 800**	**mu=1500**
MAP	0.14636562774293535	0.14561553455183524	0.1454786304051633	0.13041415375838503

MAP_best = 0.14757270211755583 when mu =600

3. With the default document analyzer, choose one or two queries, where Pivoted Length Normalization model performed significantly better than BM25 model in average precision, and analyze what is the major reason for such improvement? Perform the same analysis for Pivoted Length Normalization v.s. Dirichlet Prior smoothed Language Model, and BM25 v.s. Dirichlet Prior smoothed Language Model, and report your corresponding analysis (using your best parameters for BM25 and Dirichlet Prior smoothed Language Model). (20pts)

   All results are top 5 search output, the *defaultNumResults* = 50

3.1 Pivoted Length Normalization v.s. BM25

-PL

Query: methods of producing minimal nets given a logical function in canonical form

1. ... an algorithm for determining minimal representations of a logic function ...
2. ... the minimality of rectifier nets with multiple outputs incompletely specified ...
3. ... notes on the structure of logic nets an elementary approach to a theory of the role of structure in logic nets ...
4. X minimizing mismatch loss simple formulae and curves are given ...
5. minimization of components in electronic switching circuits the details are given of a method using boolean functions for designing switching circuits using a minimum number of if diodes or transistors ...

Pivoted Length Normalization Average Precision: 0.4533333333333333

-BM25

Query: methods of producing minimal nets given a logical function in canonical form

1. an algorithm for determining ***minimal*** representations of a ***logic function***
2. X the **minimality** of rectifier **nets** with multiple outputs incompletely specified
3. notes on the structure of ***logic nets*** an elementary approach to a theory of the role of structure in *logic nets*
4. X **minimizing** mismatch loss simple formulae and curves are ***given***
5. X theory of logical nets two values logic is applied to the study of digital computer circuits a logical net is an array constructed from a stroke element representing circuit components performing logical

functions and a delay element representing storage components various types of net are defined their properties are correlated with those of the associated set of equations and a study is made ...

BM25 Average Precision: 0.43333333333333333

My Reason: I notice that PL will punish long sentence harder than BM25, for there is no long sentence in the PL but there is one very long sentence in the BM25. And PL could suffer repeat words like there are two "logic nets" in the third output. So PL will perform better than BM25 only when the right result contains repeat words and are short sentence. Most cases show that BM25 is better than PL.

3.2 Pivoted Length Normalization v.s. Dirichlet Prior smoothed Language Model

Query: characteristics of the single electrode discharge in the rare gases at low pressures

-Pivoted Length Normalization

1. ... the **single electrode discharge** at **pressures** from a few millimetres of mercury to atmospheric pressure at the frequency of experiments were conducted with discharges in air n and ar to investigate the characteristics of the torch discharge ...
2. X Bdependence of the discharge frequency on the gas pressure and electrode separation ...
3. X ... the low pressure plane symmetric discharge ...
4. X ... the high pressure glow discharge in air report of observations on the glow discharge in air at a pressure of about mm hg between cu and w electrodes characteristics are given for discharge lengths of mm and currents of a ...
5. X ... experimental evidence for beam plasma interaction in a low pressure argon discharge ...

Pivoted Length Normalization Average Precision: 0.5

-Dirichlet Prior smoothed Language Model

1. X ... study of the induced electric discharge in rare gases experiments were performed using a tube of a few centimetres diameter and a frequency of aspects investigated included the conditions for starting the discharge and the radiation from the discharge theory is developed giving a quantitative explanation of the experimental results ...
2. X ... analysis of traces of impurities in rare gases by ultra high frequency excitation of optical radiation technique is discussed in which a cell filled at low pressure with the gas to be analysed is arranged in the field of a continuous microwave oscillation generated by a magnetron ...
3. X 499rare hiss earth current and micropulsations on november brief report of hiss and micropulsations associated with a sudden commencement at alaska ...
4. X ... vertical extent of auroral arcs and bands measurements on photographs taken from three stations indicate that quiet arcs and bands are confined to a narrow layer just about the level the thickness of this layer is most often between and and rarely from the ratio of the number of arcs and bands to the total number of observed forms it is inferred that the percentage of auroral time ...
5. X ... recombination between electrons and positive ions a bibliography of reference is critically

examined nearly always recombination takes place by an interaction involving at least two atoms direct recombination of an electron and a positive ion in a plasma involving the radiation of excess energy is relatively rare ...

Dirichlet Prior smoothed Language Model Average Precision: 0.07142857142857142

My Reason: It is obvious that DP doesn't penalize doc length, so it will prefer long sentence for the higer occurace rate of matched words. However, high term frequency doesn't mean better ranking. There are too much repeat words in long sentences. PL will penalize the doc length and popular term frequency.

3.3 BM25 v.s. Dirichlet Prior smoothed Language Model

Query: transistor phase splitting circuits

BM25

Query: transistor phase splitting circuits

1. ... the phase bistable transistor circuit corrections to paper noted in of ...
2. ... theory of two phase networks phase splitting networks as used in ssb modulation systems are discussed and calculations by approximation methods are described ...
3. X ... a new asymmetric push pull amplifier with extremely low internal impedance an output amplifier circuit is described in which the phase splitting triode is incorporated in one of the push pull arms ...
4. X ... transistorized three phase power supplies a description of the circuits required to convert a dc supply to a phase supply ...
5. X transistorized rc phase shift power oscillator data on the performance of a phase shift circuit operating at frequencies around are given ...

BM25 Average Precision: 0.35299145299145296

DP

Query: transistor phase splitting circuits

1. X... triple splitting of the f echoes polarization measurements of triple split echoes at banaras agree with measurements at high latitudes the process concerned in producing triple splitting at low latitudes is discussed ...
2. X... analysis of the split load stage transistorized expressions are derived for the current and voltage gain output impedance and output return loss of a single stage negative feedback circuit results are incorporated in an analysis of a three stage feedback amplifier with split load output ...
3. X... the early morning e layer and some evidence of pre sunrise f layer splitting observations at haringhata india show that cusps and ridges are regular sunrise phenomena at that location with marked seasonal variations in character and frequency of occurrence a splitting of the f layer during early morning in winter is also observed and the possible bearing of this on the e layer phenomena ...
4. X... f region triple splitting measurements of the direction of arrival of z echoes have been made at

hobart tasmania the results indicate that f region triple splitting is caused by back scattering from a rough layer the directions observed are consistent with the assumption tht reflection at the z level occurs when the angle of incidence is such that the wave normal becomes parallel ...

5. X... study of layer splitting in the ionosphere an equation is derived expressing the conditions for one or more maxima to occur in the height distributions of temperature and recombination coefficient a numerical treatment is given for splitting with uniform temperature distribution would imply a highly improbable value for the recombination coefficient on the other hand the presence of a warm ...

Dirichlet Prior smoothed Language Model Average Precision: 0.07777777777777778

My Reason: Same reason as PL better than BM25. BM25 will penalize the doc length and popular term frequency.

---

4. Pick one of the previously implemented scoring functions out of
   - Okapi BM25
   - Pivoted Length Normalization
   - Language Model with Dirichlet Smoothing

to analyze under what circumstance the chosen scoring function will mistakenly favor some less relevant document (*i.e.*, ranks a less relevant document at a higher position than a more relevant one). Please correspond your analysis with what you have found in Problem 3.

After reading the paper An Exploration of Axiomatic Approaches to Information Retrieval, 1) can you briefly summarize the major contribution of this paper? 2) how do you think you can fix the problem you have identified in the ranking result analysis? Please relate your solution and corresponding implementation in the report. Also report the resulting ranking performance of your revised ranking algorithm. (30pts)

1) briefly summarize the major contribution of this paper

This paper purpose a new axiomatic retrieval model, which decompose the original method into three new constraint component. The first one is the EXP weighting function, which penalizes popular terms in the document(similiar to IDF). The second is the query growth function. There are several constraints in the third component, the Document growth function. The first constaint is each query term should contribute to the weight function if adding new one; the second constraint is punihing the appearance of not-target words. There are two constrains not used in this paper, but could be future work.

2) how to fix problem

Based on the formula (8) in part 3.3.1 in the paper. The new PL is:

$$S(Q,D) = \sum_{t \in D \cap Q} TF(C_t^D) \cdot C_t^Q \cdot weight(t) \frac{avdl + s}{avdl + |D| \cdot s} \qquad (8)$$

where $0 \leq s \leq 1$ and $TF(x) = 1 + ln(1 + ln(x))$.

My implement is:

```
protected float score(BasicStats stats, float termFreq, float docLength) {
      float s = 0.25f; // default_s = 0.75
      float cwq = 1.0f; //assume that the query term frequency is always one.
      float df = stats.getDocFreq();
      float n_avg = stats.getAvgFieldLength();
      double k = 0.5;

      float term1 = (float) (1 + Math.log(1 + Math.log(termFreq)));
      float term2 = cwq * (float) Math.log((stats.getNumberOfDocuments() + 1) /
df );
      float term3 = (n_avg + s)/(n_avg + docLength * s);
      return term1 * term2 * term3;
    }
```

3) resulting ranking performance

In the question 3.1, I compare the PL withBM25. However, there is only one query satisfy the PL is better than BM25. The AP is 0.4533333333333333. After changed, the AP is 0.57. Now it is significantly better than BM25. This imporvemnet is in not punishing long sentence so hard.

The total performance is behind, which is better than before:

MAP: 0.15577671105065125

P@10: 0.3473118279569892

MRR: 0.7023553507424478

NDCG: 0.4332160864672218