Information Retrieval

Assignment 4 (Bonus)

Due: Sunday, 18th May. (You can submit by Email to the lab instructor "Dr. Nour")

---

1. **(**2 Points**) In the permuterm index, each permuterm vocabulary term points to the original vocabulary term(s) from which it was derived. How many original vocabulary terms can there be in the postings list of a permuterm vocabulary term?**

2. **(**2 Points**) Write down the entries in the permuterm index dictionary that are generated by the term "world".**

3. **(**2 Points**) If you wanted to search for "r*al" in a permuterm wildcard index, what key(s) would one do the lookup on?**.

4. **(**2 Points**) In the k-grams index, the vocabulary terms in the posting lists are lexicographically ordered. Why is this ordering useful?**

5. **(**2 Points**) Consider the query fi*mo*er. What Boolean query on a bigram index would be generated for this query?**

6. **(**3 Points**) If |S| denotes the length of string S, show that the edit distance between s1 and s2 is never more than max{|s1|, |s2|}.**

7. **(**5 Points**) Write down a Java (or C) program to apply the edit distance algorithm? What this algorithm used for?**

8. **(**2 Points**) Compute the edit distance between layer and royal. Write down the 5 × 5 array of distances between all prefixes as computed by the edit distance algorithm.**

9. **(**2 Points**) Compute the Jaccard coefficients between the query "bord" and each of the terms {border, lord, morbid, sordid}.**