

Query导向的排序方法评价实验

Data Collection

- 根据数据库内容编写query
 - query的数量
 - query的长度
 - query的质量
 - 包含该query的字段
- 根据query得到hits
 - 对hits根据query进行打分（3分制或4分制或5分制）
 - 3-5人对同一数据集进行打分，不同取多数，无多数取平均
 - 给用户呈现哪些内容
 - 质量和相关性评判标准（打分标准）
 - 是否区分质量分数与相关性分数

- 具体流程：
 - i. 统计content字段、title字段、notes字段中出现次数最多的前100个有效term（去除数字等无意义的term），根据这些term编写query，并测试这些query的hits数量（1000以上？），取不同baseline的前20（30？ 50？）集合。
 - ii. 将得到的hits随机提供给用户（3人以上）进行打分，用户可以查看每个数据集的title、description、url、(content？)、etc.，根据该数据集的质量及与该query的相关程度进行打分。

Baseline

- 单一方法排序
 - TF-IDF
 - BM25
 - FSDM
 - PageRank
 - DING
 - DRank (仅根据数据集的度数排序的native rank)
 - Language Model* (Dirichlet smoothing and Jelinek-Mercer smoothing)

- 混合方法排序
 - Quality(PageRank、DING、DRank) + Relevance(TF-IDF、BM25、FSDM)
*作为自主设计的创新方法
- 多field与单field
 - 全部field
 - 仅content
 - 仅title和description
 - 仅content、title和description

Research Questions

- Q1 不同方法的效果比较（在单field上对比和在多field上对比）
- Q2 不同field对排序效果的影响（相同方法下不同field上搜索的比较）
- Q3 自主设计方法的效果

Evaluation Metrics

- nDCG@k (k=5,10,15,20,50*)
- Precision@k
- Recall@k
- F值 or MAP