

In questo report si presentano i risultati delle sperimentazioni condotte su diversi modelli di ripperimento dell'informazione. Si è utilizzato il software [Terrier](#) (vers 4.4). La collezione usata per la sperimentazione consiste di 528155 documenti in lingua inglese (TIPSTER disk 4&5 minus CR). Per la valutazione, si sono usati i TREC-7 topics 351-400 e i relativi *qrels* (due gradi di rilevanza: R e NR) sulla collezione. Per le sperimentazioni, sono state utilizzate diverse indicizzazioni e due modelli: BM25 e TF*IDF. La tabella sotto riporta le quattro configurazioni modello-indicizzazione testate:

Configurazione	Modello	Stoplist	Stemmer
BM25(SL+STEM)	BM25	Sì	Porter
BM25(NO-SL+STEM)	BM25	No	Porter
TF*IDF(SL+STEM)	TF*IDF	Sì	Porter
TF*IDF(NO-SL+NO-STEM)	TF*IDF	No	No

La Stoplist utilizzata è quella predefinita di Terrier. Sono state compiute **due prove** di reperimento e valutazione. Nella *prima prova* è stato utilizzato per il reperimento il campo TITLE dei TREC topics, mentre per la *seconda prova* è stato utilizzato il campo DESC. Per la valutazione, sono state computate le statistiche AP, P@10 e Rprec per ogni topic. Per confrontare le varie configurazioni, per ognuna di queste statistiche è stato compiuto il test 1-way ANOVA con post-hoc Tukey HSD. I risultati delle prove e maggiori dettagli sulla sperimentazione sono disponibili nella repository:

<https://github.com/yokola95/ir-homework1>

Nella prima prova, si osserva che le quattro configurazioni ottengono risultati comparabili per tutte e tre le statistiche. Per ognuna di queste statistiche, il test statistico 1-way ANOVA con post-hoc Tukey HSD non rivela differenze statisticamente significative tra le quattro configurazioni. Per brevità, si riportano nella tabella sotto solo i valori medi di AP, P@10 e Rprec su tutti i Topic per ogni configurazione.

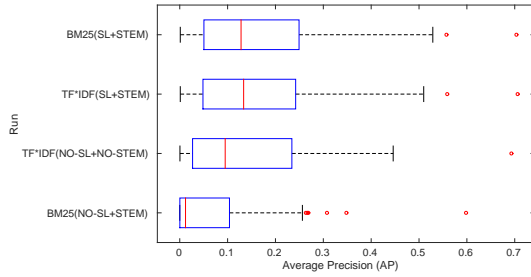
Configurazione	MAP	P@10	Rprec
BM25(SL+STEM)	0.1828	0.4180	0.2391
BM25(NO-SL+STEM)	0.1854	0.4300	0.2406
TF*IDF(SL+STEM)	0.1821	0.4200	0.2391
TF*IDF(NO-SL+NO-STEM)	0.1693	0.4060	0.2290

Tabella 1: Risultati - prima prova (campo TITLE)

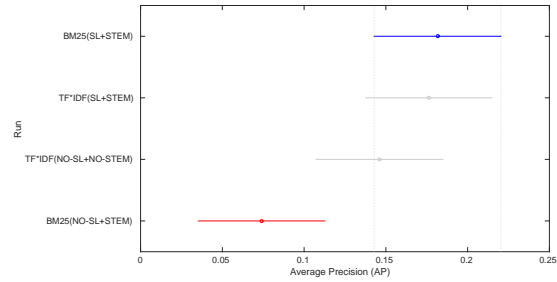
I risultati della seconda prova sono riportati nei grafici di Figura 1. Si osserva che le configurazioni prive di Stoplist ottengono risultati peggiori per tutte e tre le statistiche. Questo è un risultato aspettato, in quanto i campi DESC contengono spesso *stopwords* di cui è tenuto conto nel ranking. Per tutte e tre le statistiche di AP, P@10 e Rprec il test

1-way ANOVA con post-hoc Tukey HSD mostra che la configurazione BM25 senza Stoplist non appartiene al Top Group, avendo una performance significativamente peggiore delle altre tre configurazioni. Il modello BM25 infatti è molto sensibile alla presenza di *stopwords*: nella definizione più comune del modello, una parola della query che compare su più di metà della collezione porta a un contributo negativo nel calcolo dello score.

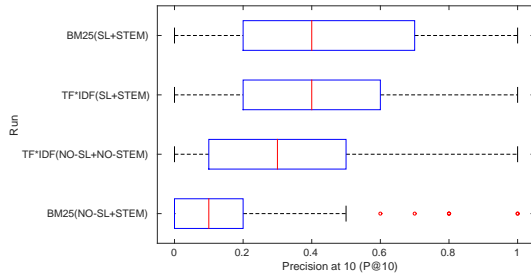
Confrontando la prima e la seconda prova, si può osservare che l'utilizzo del campo TITLE per le query porta ai risultati migliori: i risultati per le configurazioni con stoplist tra le due prove sono comparabili, mentre le configurazioni senza stoplist hanno una performance peggiore utilizzando il campo DESC per le query.



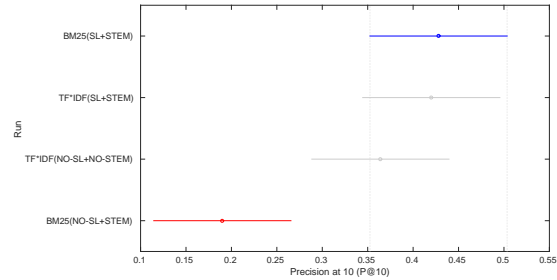
(a) Boxplot - AP



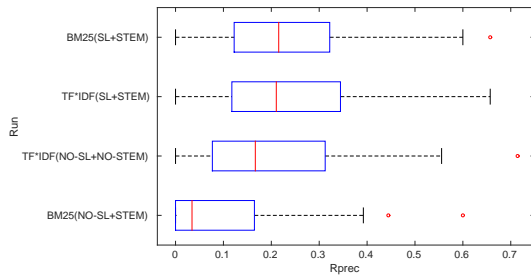
(b) Tukey HSD - AP



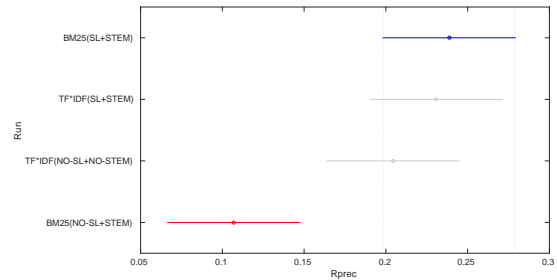
(c) Boxplot - P@10



(d) Tukey HSD - P@10



(e) Boxplot - Rprec



(f) Tukey HSD - Rprec

Figura 1: Risultati - seconda prova (campo DESC)