

**Title:** Laboratory 4&5

**Date:** 16/11/21

**Student names:** Vincent Achukwu, Alan Hian Wu

## 4. Experimenting with Query Expansion in IR

### 4.1 Initial Investigations

Using the topic title “*Industrial Espionage*” (ID: 311) with BM25 having parameters  $k=1.2$  and  $b=0.75$ .

Below are the top 5 retrieved documents for the chosen topic.

Document ID	Description
FT944-15440	Article from 18th October 1994 about how industrial espionage is a growing problem in business
FBIS3-58751	An article from 21st January 1994 about how industrial espionage can be dealt with
FT931-3928	Article based from 13th March 1993 on the acquittals in car park spying trial
FBIS4-19393	18th April 1994 article about how Kanther calls upon the industry to take better protective measures against espionage
FT934-7093	22nd November 93 on military espionage

We then wrote a Python script to calculate the  $ow(i)$  and  $rw(i)$  values based on the terms we selected and obtained the following results (including the  $r(i)$  and  $n(i)$  values):

t(i)	r(i) out of R	n(i)	rw(i)	ow(i)
compani	4	158337	1.8677321283786399	7.4709285135145596
industri	5	199735	2.8055933982194605	14.027966991097303
secret	3	12441	4.005083967558967	12.015251902676901
war	1	40885	1.3199297931671095	1.3199297931671095
countri	5	229732	2.5604173582041496	12.802086791020749

We then performed query expansion by using some of these terms to our original query. We obtained the following results by including “compani”, “secret”, and “war” with the original query.

Document ID	Top 5 of original query + “compani”	Top 5 of original query + “secret”	Top 5 of original query + “war”
FT944-15440	FT944-15440	FT944-15440	FBIS3-58751
FBIS3-58751	FT931-3928	FT934-7093	FBIS4-51593
FT931-3928	FBIS3-58751	FBIS3-58751	<b>FBIS4-30090</b>
FBIS4-19393	<b>FBIS3-31719</b>	<b>FT933-8481</b>	<b>FBIS4-30158</b>
FT934-7093	FBIS4-19393	<b>FBIS3-18363</b>	<b>FT933-8481</b>

In conclusion, we can see that appending the term “war” to “Industrial Espionage” resulted in newer results which we have not seen in the top 5 relevant results. Document FT944-15440 stayed consistent with the terms “compani” and “secret” appended to the original query, but then never showed up in the top 5 results when the term “war” was included with the original query. Document FT934-7093 went from originally being last from the top 5 to second when including “secret” with the original query.

## 4.2 Investigations using qrel data

Similar to lab 2, we determined if the resulting documents were relevant to the search queries (including “compani”, “secret”, and “war”) in comparison to what the qrel file has.

“Industrial” “Espionage”	Our Relevance Judgement (BM25)	True relevance feedback (qrel)
FT944-15440	1	1
FBIS3-58751	1	1
FT931-3928	1	1
FBIS4-19393	1	1
FT934-7093	0	0

Results including “compani”, “secret”, and “war”

“Industrial” “Espionage” “company” “secret” “war”	Our Relevance Judgement (BM25)	True relevance feedback (qrel)
FT933-8481	1	1
FT944-15440	1	1
FBIS3-58751	1	1

FT933-8010	1	1
FT933-7932	1	1

We observed that the use of query expansion resulted in more relevant documents being retrieved. The last 2 documents were almost identical, hence why both have the same result. We may assume that expanding queries may result in more relevant results.