

# Assignment 3

## Information Retrieval and Web Search

Ahmed Rekik - 790063 & Tim Henning - 789242

November 15, 2017

### 2. Boolean Retrieval

#### 2.a. One of the drawbacks of the Boolean retrieval model lies in the size of the returned result set. Why is the size typically difficult to control?

The fact that the IR with a Boolean retrieval model returns an out of hand result set is due to its absolute Boolean property. While the result depends on query terms, these could occur in relevant as non relevant documents, thereby the size of the result will be hard to control and unpredictable.

#### 2.b. According to query log statistics, most queries do not contain search operators. Why are Boolean operators nevertheless necessary in professional search? Are consumer and professional search precision or recall oriented?

Although they might not be frequently employed, search operators represent an efficient tool to express queries in a more expressive way in order to increase chance to hit relevant documents. Since professional search might generally intend very specific documents, those could be not easy to reach with the regular queries and without specifying certain predicates.

A very specific document aimed for a professional use may contain frequent word that occur in other documents often, therefore by querying without employing several predicates irrelevant documents will submerge the relevant one. Thus, by using operator ("AND", "OR") for providing very specific key-word and eliminating the undesirable documents (using "NOT" operator) the professional could reach the precise document faster.

As expressed above, the professional use in information retrieval tend more to the effective and pertinent result. Since they look at (very) specific information their search should rather precision-oriented. Otherwise, the consumer in general is more tolerant in his search and can consider multiple sources as long as they still relevant, thus the consumer use would be more recall-oriented.

word	Query				Document			
	tf	wf	df	q_i = wf-idf	tf	wf	d_i = Normalized wf	q_i . d_i
digital	1	3	10000	3	1	3		
video	0	2	100000	0	1	2		
cameras	1	2.3	50000	2.3	2	2.3		

### 3. Boolean Retrieval in Practice

#### 3.a. Evaluate the query: $q_1 = (t_1 \text{OR} t_5) \text{NOT} t_2$ .

$t_1 \text{OR} t_5 = D_1, D_2, D_4, D_5, D_6$ ;

$t_2 = D_2, D_6$

$q_1 = D_1, D_4, D_5$

#### 3.b. $q_2 = (t_1 \text{AND} t_5) \text{OR} (t_3 \text{AND} t_2)$

$t_1 \text{AND} t_5 = D_1, D_2$ ;

$t_2 \text{AND} t_3 = \emptyset$

$q_2 = D_1, D_2$

#### 3.c. Vector Space Model

Compute the vector space similarity between the query "digital cameras" and the document "digital cameras and video cameras" by filling out the empty columns in the Table 1. Assume  $N=10,000,000$ , logarithmic term weighting (wf columns) for query and document, idf weighting for the query only and cosine normalization for the document only. Treat and as a stop word. Enter term counts in the tf columns. What is the final similarity score?

### 4. (Programming) Boolean Queries

#### 4.a. Print a list of the comments (comment text only; no other attributes) in your dataset that match the queries:

Searching: October

Found 48 results in 0.21ms.

But sadly most of people's referendum are declared as "not binding" from the beginning – as are the coming one's in Italy for Lombardy and Venetia on the 22nd of October. The Red Army was created after the success of the October Revolution in 1917 led by V.I. Ulyanov (Lenin). Thirty five years later the Red Army smashed the fascist armies of Germany, Finland, Hungary, Romania, Italy, Croatia and Slovakia. The bodies of tens of thousands rascist/anti-

semitic fascists fertilized Soviet territory helping the recovery of Soviet lands.\nYou stand with all that is immoral.

Yep. Between 6–8 million between April 1945 and October 1949. Yeah but they're always wet as October. That's always a plus. May called Swamp on October 10th on JCPOA. It was decided then. Brits are ahead of Trump in the lineup for Tel Aviv.

Searching: jobs

Found 2847 results in 2.29ms.

You name three countries brought to their knees by islamic invasions and terrorism ,sjw terrorist infiltration of govt jobs but its soon going to end. Camps for the terrorists and their support base.. just wait n watch history repeat. consequences are that his political career is over. This situation clearly demonstrate that EU and PACE is not about solutions. That is not their job. Their job is to execute orders and this guy see the full consequences of thinking independently and toward peace.

"form a film and production crew for the purposes of producing a video documentary based on its research associated with Mr . Guelen." so the job is similar to other USA 'consulting firm', like film about Maidan, White Helmet, ISIS beheading series ( wonder when the season finale is.)

Agree—and he is doing the same job, the British elite did 80 years or more before...

So true.\n\nAcademically questioning holocaust will make you lost job in University and land you in a jail.

Searching: Trump

Found 7455 results in 6.08ms.

Who believes you Trump? You are the biggest self-loving opportunistic liar in politics. You arm Saudi Arabia and call Iran terrorists. You want peace in the middle east but support the settlement building of Israel in Palastine. You claim that the actions against Qatar are in support of your vision on Iran, yet you have a military base in Qatar from where you bomb Syria. Go away Trump!

Just months ago, Donald Trump was thanking Wikileaks and asking them to expose more about Hillary Clinton. Now he wants to arrest and jail the founder?

Trump lies about crime numbers, terrorist attacks, his taxes, ...

I think she has the goods on Trump.

No need. Trump is just a wee little froggie in a great, big swamp.

Searching: hate

Found 2247 results in 1.73ms.

I am a bosnian turk. I support Russia Putin and RT and support BRICS. Hate NWO and USA and NATO.\nMost of the Turks are just sheeps with no awareness no understanding what is going on earth.\nBut not all Turks my Russian friend.

Do you blame Pinochet, or the people that made it possible for him to commit those atrocities? \nOh and yes Russian names, though mostly western educated jews. As much as i hate jewbashing.

You don't hate to think about it actually

Symbols and words are hate crimes, but bullets and explosions are not

True tru and true again. 90's were disgucting disaster. While those oligarch liberals were pocketing on USSR wealth, and flushign it into foreign banks, regular working Russian people were starving. And of course those liberals miss those days when they could easily rob, and wish to get back to those times. They hate Russia, they only love money they can steal from the people, and those people who support them are so gullible and dumb.

### Implementation Choices

- Stopword List: as there are no special stopwords in our use-case for RT.com, we are using nltk.corpus.stopwords (Porter et al 2,400 stopwords for 11 languages)
- Stemmer: using nltk.stem.porter.PorterStemmer (widely regarded as the most universal stemmer in NLTK)
- Lemmatizer: not used at the moment (search results were sufficient without it)