



# Information Retrieval Exercises

Assignment 4:

## **Synonym Expansion with Lucene - Solutions**

Mario Sängner ([saengema@informatik.hu-berlin.de](mailto:saengema@informatik.hu-berlin.de))

# Evaluation setup

---

- Evaluation was performed on a dedicated server of our research group
  - Evaluation set consisted of 9 queries
  - Time for indexing was measured separately
- Each query was executed 10 times for each team
  - Best and worst runtime was excluded for each query
  - Average runtime was recorded
- For each query a ranking was calculated
  - The team which the highest average rank wins

# Evaluation data set

---

- Basically comprised of all validation queries
- One extra query: plot:Kafkaesque AND plot:incubus
  - Extended query: +plot:kafkaesque +Synonym(plot:incubus plot:nightmare)
  - Result:
    - MV: Aaron Cohen's Debt (1999)
    - MV: Das Erwachen des Herrn Schmidt (2010)
    - MV: Deja-moo (2013)
    - MV: Korespondence (2013)
    - MV: Le procès (1962)
    - MV: The Humble Assessment (????)

# Evaluation: Query runtimes

\* indexing times in s, query times in ms

Group	Index	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
C418	66.7	220.3	80.5	50.6	50.0	46.4	42.7	53.1	38.7	31.5
Codeworms	52.5	-	83.3	58.2	51.0	46.3	46.6	58.1	46.5	40.0
DominanteWW	31.5	99.1	25.8	13.0	3.6	5.5	6.2	9.9	4.8	3.4
DoppelW	220.8	-	48.6	31.9	-	-	-	-	29.4	-
EBIR2018	72.9	-	39.3	25.6	25.3	21.8	17.2	34.5	17.3	12.1
Einfallslos	41.2	178.9	64.2	44.4	38.2	41.7	36.2	46.2	36.6	33.4
Fiedner	349.7	164.8	64.7	48.7	54.1	46.0	41.9	42.3	39.5	34.9
FindingMovies	49.6	238.6	78.7	58.9	41.0	48.3	47.1	52.3	43.4	36.8
Freudon	63.5	-	66.6	37.1	35.6	34.0	34.4	43.3	32.9	28.5
Google	215.0	-	57.2	37.8	47.5	32.4	35.2	39.2	30.3	30.3
GroupNo1	700.2	212.8	70.6	50.1	37.9	37.8	35.3	41.3	38.5	31.6
KSMR	48.0	230.9	62.5	32.9	31.9	30.0	25.4	39.1	30.3	19.6
KnockKnock	1657.6	205.3	64.5	58.6	56.3	46.1	47.6	46.8	41.8	51.3
Lejunds	1226.1	185.6	80.5	62.6	62.9	48.7	51.8	50.9	46.6	41.9
Lilaloi	42.4	205.9	75.7	50.3	45.7	43.6	39.2	51.0	37.6	31.8
Lovelace	57.3	248.1	76.7	49.0	45.9	40.8	42.7	57.7	41.1	30.8
Nebula	20.7	168.5	41.2	27.8	18.5	24.7	23.2	30.1	22.1	13.8
PonyUndKleid	581.4	-	70.1	57.7	63.5	54.5	50.7	50.4	39.0	36.0
RickundFranz	60.5	190.5	64.0	38.8	38.7	33.2	29.9	41.7	30.9	28.2
Uberwinder	48.7	231.7	104.0	50.2	51.9	43.3	42.4	50.7	37.5	34.8
Valeria	60.8	-	82.3	57.0	57.8	-	-	-	-	-
WirSindGroot	394.5	160.7	59.0	47.3	38.8	41.5	34.5	36.3	34.9	34.5

# Evaluation: Ranking

Group	Index	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
C418	13	11	18	16	14	17	14	18	14	9
Codeworms	8	-	21	19	15	16	16	20	20	18
DominanteWW	2	1	1	1	1	1	1	1	1	1
DoppelW	16	-	4	4	-	-	-	-	4	-
EBIR2018	14	-	2	2	3	2	2	3	2	2
Einfallslos	3	5	9	9	7	11	10	11	10	12
Fiedner	17	3	11	11	17	14	12	9	16	15
FindingMovies	7	14	17	21	10	18	17	17	19	17
Freudon	12	-	12	6	5	7	6	10	8	6
Google	15	-	5	7	13	5	8	6	5	7
GroupNo1	20	10	14	13	6	8	9	7	13	10
KSMR	5	12	7	5	4	4	4	5	5	4
KnockKnock	22	8	10	20	18	15	18	12	18	20
Lejunds	21	6	18	22	20	19	20	15	21	19
Lilaloi	4	9	15	15	11	13	11	16	12	11
Lovelace	9	15	16	12	12	9	14	19	17	8
Nebula	1	4	3	3	2	3	3	2	3	3
PonyUndKleid	19	-	13	18	21	20	19	13	15	16
RickundFranz	10	7	8	8	8	6	5	8	7	5
Uberwinder	6	13	22	14	16	12	13	14	11	14
Valeria	11	-	20	17	19	-	-	-	-	-
WirSindGroot	18	2	6	10	9	10	7	4	9	13

# Evaluation: Competition

Group	Average Rank	Rank	Position	Points
C418	14.4	12		
Codeworms	-	-		
DominanteWW	1.1	1	1	5
DoppelW	-	-		
EBIR2018	-	-		
Einfallslös	8.7	5		
Fiedner	12.5	9		
FindingMovies	15.7	13		
Freudon	-	-		
Google	-	-		
GroupNo1	11	7		
KSMR	5.5	3	3	2
KnockKnock	16.1	14		
Lejunds	18.1	15		
Lilaloi	11.7	8		
Lovelace	13.1	10		
Nebula	2.7	2	2	3
PonyUndKleid	-	-		
RickundFranz	7.2	4	4	1
Überwinder	13.5	11		
Valeria	-	-		
WirSindGroot	8.8	6	5	1

---

# Questions?