



# Information Retrieval Exercises

Assignment 3:

**Boolean IR with Lucene - Solutions**

Mario Sängler ([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 11 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: Runtimes

Group	Index	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11
C418	52.2	189.1	128.3	56.6	112.2	37.3	41.3	35.3	42.7	38.2	36.0	33.5
Codeworms	52.1	211.2	156.3	66.1	201.6	68.2	52.7	40.8	53.9	43.5	39.7	38.1
DominanteWW	26.2	58.9	16.3	3.4	53.0	4.7	2.8	1.7	10.5	3.5	1.9	1.7
DoppelW	42.0	119.5	98.3	33.6	84.2	28.6	41.5	30.3	34.4	27.2	24.7	22.8
EBIR2018	88.7	143.6	114.7	27.6	90.7	16.4	17.5	14.5	29.1	18.0	12.4	13.3
FindingMovies	45.7	179.2	127.1	48.4	117.1	32.4	35.9	42.2	49.6	70.6	51.1	32.7
Freudon	55.2	-	505.0	542.3	327020.8	330.2	-	851.0	895.7	6413.6	435.3	591.7
Friedner	46.5	201.3	151.1	55.7	122.6	46.0	47.1	45.0	60.8	53.2	49.6	48.0
Google	47.8	190.0	138.4	48.1	127.4	37.1	50.0	33.3	42.9	34.9	31.1	31.4
GruppeNo1	690.2	194.4	122.2	41.9	102.4	29.7	37.4	30.2	40.4	35.5	30.4	30.5
KSMR	38.7	183.3	133.0	47.8	101.5	35.7	44.6	37.6	46.8	37.8	34.5	38.0
KnockKnock	727.8	206.5	118.1	47.8	115.1	39.2	44.3	35.4	55.3	39.0	33.1	33.5
Lejunds	46.3	184.4	140.1	57.5	112.8	40.8	47.2	47.1	53.2	56.0	36.6	42.3
Lilaloi	43.6	162.1	126.3	45.2	101.5	31.0	34.1	31.3	40.4	34.7	29.7	30.2
Lovelace	55.1	208.3	153.0	63.8	122.5	39.8	44.3	38.2	54.4	47.4	38.9	39.1
Nebula	14.3	200.4	128.7	45.3	108.6	32.2	38.2	30.9	47.7	33.4	30.0	33.5
PonyUndKleid	47.9	180.5	126.7	48.1	110.9	37.8	42.3	37.2	47.8	39.0	31.6	35.7
RickUndFranz	51.4	196.7	147.1	56.6	118.6	38.8	48.0	41.1	54.9	47.9	38.6	39.5
Überwinder	67.7	212.5	135.7	60.2	125.6	38.9	44.0	49.8	75.2	74.8	35.7	36.8
Valeria	66.0	217.4	145.1	57.0	122.6	37.6	46.0	44.8	60.7	46.5	39.1	38.1
WirSindGroot	44.5	178.8	128.7	40.2	101.7	32.0	35.1	31.2	40.5	32.0	30.1	31.5

\* indexing times in s, query times in ms

# Evaluation: Ranking

Group	Index	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11
C418	14	10	9	14	10	11	8	9	7	10	13	9
Codeworms	13	18	20	20	20	20	20	14	14	13	18	15
DominanteWW	2	1	1	1	1	1	1	1	1	1	1	1
DoppelW	4	2	2	3	2	3	9	4	3	3	3	3
EBIR2018	19	3	3	2	3	2	2	2	2	2	2	2
FindingMovies	7	6	8	12	13	8	5	16	12	19	20	8
Freudon	16	-	21	21	21	21	-	21	21	21	21	21
Friedner	9	15	18	13	16	19	16	18	19	17	19	20
Google	10	11	14	10	19	10	19	8	8	7	8	6
GruppeNo1	20	12	5	5	7	4	6	3	4	8	7	5
KSMR	3	8	12	8	4	9	14	12	9	9	11	14
KnockKnock	21	16	4	8	12	16	12	10	17	11	10	9
Lejunds	8	9	15	17	11	18	17	19	13	18	14	19
Lilaloi	5	4	6	6	4	5	3	7	4	6	4	4
Lovelace	15	17	19	19	15	17	12	13	15	15	16	17
Nebula	1	14	10	7	8	7	7	5	10	5	5	9
PonyUndKleid	11	7	7	10	9	13	10	11	11	11	9	12
RickUndFranz	12	13	17	14	14	14	18	15	16	16	15	18
Überwinder	18	19	13	18	18	15	11	20	20	20	12	13
Valeria	17	20	16	16	16	12	15	17	18	14	17	15
WirSindGroot	6	5	10	4	6	6	4	6	6	4	6	7

# Evaluation: Competition

Group	Average Rank	Rank	Position	Points
C418	10.2	10		
Codeworms	17.2	19		
DominanteWunderWiesel	1.1	1	1	5
DoppelW	3.5	2	2	3
EBIR2018	4	3		
FindingMovies	10.6	11		
Freudon	-	-		
Friedner	16	17		
Google	11.6	12		
GruppeNo1	7.4	6		
KSMR	8.8	8		
KnockKnock	12.7	13		
Lejunds	14.5	14		
Lilaloi	5	4	3	2
Lovelace	15.7	16		
Nebula	7.4	6	5	1
PonyUndKleid	10	9		
RickUndFranz	14.9	15		
Überwinder	17.2	19		
Valeria	16.1	18		
WirSindGroot	5.7	5	4	1

# Assignment 4

---

- Updated version of test queries and results is now online!
- Keep pre-processing of synonyms in mind!
  - Synonyms have to be pre-processed in the same way as the textual content!
- Example: StandardAnalyzer of Lucene
  - Pre-processing of "Dr." leads to dr
  - However, WordNet refers "Dr." as synonym for "doc", "doctor", "medico", "physician", ....

---

# Questions?