



Information Retrieval

Results for Assignment 4:

Synonym Expansion with Lucene and WordNet

Patrick Schäfer (patrick.schaefer@hu-berlin.de)

Marc Bux (buxmarcn@informatik.hu-berlin.de)

Approach, Effort Invested (Monday)

Group	Index	Effort (h)
BigGoogle		
Devin&Johannes	query expansion	20
getSchwifty()	index expansion	8
Hendrik&Sinje	query expansion	10
ilGruppo		
jp&bl	query expansion	8
Julian		
Ic&mk		
NewDimensions	query expansion	15
Oceanic2		
Roland&Elias	query expansion	14
TeamJA	query expansion	10

Approach, Effort Invested (Tuesday)

Group	Index	Effort (h)
Bertram&Abdalla		
BeSharps		
Bruno&Christin		
Darko&Mario		
Dinh&Feyco	Query Expansion	1h – 4h
HouseOfMojo		
InfoRetImWinter	Query Expansion	10h
jp&bl		
Julian&Dennis		
lc&mk		
lucy@molly		
Oceanic2		
Paul&Peter		
Vincent&Max		

Student Presentations (Monday)

- Devin&Johannes
- TeamJA

- One team can present their [Lucene WordNet Indexer](#).
- One team can present their [Lucene Query Expansion](#).

Student Presentations (Tuesday)

- ?
- One team can present their [Lucene WordNet Indexer](#).
- One team can present their [Lucene Query Expansion](#).

Evaluation Setup

- One run:
 - build index (discard submission if index not built after one hour),
 - Correctness: run 8 train queries (discard submission if unsuccessful),
 - Evaluation: run 8 train + 2 evaluation queries, and
 - report runtimes for building index and running each evaluation query.
- **Eleven runs** per submission.
- Determine **median runtimes**.
- Determine ranks (across all submissions) for building the index and running each evaluation query.
- Compute **average ranks** across all 11 ranks.

The 10 Evaluation Queries

1. plot:Elvis AND plot:is AND plot:dead AND type:movie
2. (title:Star AND title:Trek) OR (title:Star AND title:Wars) AND year:2016
3. plot:donald AND plot:trump AND plot:president AND type:movie AND year:2016
4. plot:supercomputer AND type:movie
5. title:flesh AND title:fable
6. title:Voyager AND plot:daydream AND plot:doc
7. plot:awful AND type:videogame AND **title:two**
8. plot:omnibus AND plot:driver AND plot:poetry
9. **plot:good AND plot:better AND plot:leisure AND type:video**
10. **gameplot:overflown AND plot:overflow AND plot:zombie AND type:movie**

- query:

plot:good AND plot:better AND plot:leisure AND type:videogame

- +Synonym(plot:adept plot:beneficial plot:commodity plot:dear plot:dependable plot:effective plot:estimable plot:expert plot:full plot:good plot:goodness plot:honest plot:honorable plot:just plot:near plot:practiced plot:proficient plot:respectable plot:right plot:ripe plot:safe plot:salutary plot:secure plot:serious plot:skilful plot:skillful plot:sound plot:soundly plot:thoroughly plot:unspoiled plot:unspoilt plot:upright plot:well)

+Synonym(plot:adept plot:advantageously plot:ameliorate plot:amend plot:beneficial plot:best plot:better plot:bettor plot:break plot:comfortably plot:considerably plot:dear plot:dependable plot:easily plot:effective plot:estimable plot:expert plot:full plot:good plot:honest plot:honorable plot:improve plot:intimately plot:just plot:meliorate plot:near plot:practiced plot:proficient plot:punter plot:respectable plot:right plot:ripe plot:safe plot:salutary plot:secure plot:serious plot:skilful plot:skillful plot:sound plot:substantially plot:unspoiled plot:unspoilt plot:upright plot:wagerer plot:well)

+plot:leisure

+type:videogame

- expected result (6): [MV: Gunlok (2000) (VG), MV: Leisure Suit Larry 1: In the Land of the Lounge Lizards (1991) (VG), MV: Leisure Suit Larry 6: Shape Up or Slip Out! (1993) (VG), MV: Leisure Suit Larry 7: Love for Sail! (1996) (VG), MV: Leisure Suit Larry Goes Looking for Love (In Several Wrong Places) (1988) (VG), MV: Leisure Suit Larry in the Land of the Lounge Lizards (1987) (VG)]

Evaluation Query 2

- query:
plot:overflown AND plot:overflow AND plot:zombie AND type:movie
- +Synonym(plot:overflow plot:overflown plot:overfly plot:overrun)
+Synonym(plot:flood plot:outpouring plot:overflow plot:overrun
plot:overspill plot:runoff)
+Synonym(plot:automaton plot:zombi plot:zombie)
+Synonym(type:film type:flick type:movie type:pic type:picture)
- expected result (22): [MV: Arise (2009), MV: Bubba's Chili Parlor (2005), MV: Busanhaeng (2016), MV: Cotard (2013), MV: Dawn of the Dead (2004), MV: Dead at the Box Office (2005), MV: Deadworld (2017), MV: Fistful of Brains (2008), MV: Let There Be Zombies (2014), MV: Monster Minis (2010), MV: Not Quite Dead (2007), MV: Raiders of the Damned (2007), MV: ReZistance (2014), MV: Rise of the Zombies (2012), MV: Shaun of the Dead (2004), MV: The Running Dead (2013), MV: Virus (1980), MV: Z's (2011), MV: Zombie Hood (2013), MV: Zombieland (2009), MV: Zombies During Third Period: Again (2011), MV: Zombies i Orten (2014)]

Median Runtimes (10 out of 20 submissions)

	Index / s	Query 1 /ms	Query 2	Query 3	Query 4	Query 5	Query 6	Query 7	Query 8	Query 9	Query 10
BigGoogle	53,90	18,15	12,74	10,43	7,82	9,21	23,14	44,20	11,16	21,62	15,15
bl_jp	46,93	31,90	31,82	28,83	24,20	27,80	26,03	36,32	25,65	-1,00	-1,00
Dinh&Feyco	19,97	2,62	1,14	0,95	0,51	0,79	0,67	1,10	0,94	1,04	0,84
InformationRetrievalImWinter	22,17	19,68	5,61	8,01	3,38	9,19	8,02	10,97	8,49	26,02	5,05
lc+mk	63,29	297,87	355,12	367,29	152,28	165,45	206,92	199,92	243,53	-1,00	-1,00
lucy@molly	25,34	16,73	4,49	6,05	2,93	5,51	4,82	10,01	36,08	13,43	5,21
newDimensions	18,89	24,91	19,45	11,47	14,76	10,40	10,36	17,17	12,50	27,67	25,88
Oceanic2	59,18	168,00	131,19	124,30	77,94	81,46	104,18	90,07	108,62	-1,00	164,39
Roland&Elias	62,59	24,41	16,61	20,07	16,41	16,98	16,38	20,17	16,90	21,39	17,01
TeamJA	64,65	37,37	36,62	29,21	27,49	34,70	31,74	29,41	25,73	36,04	28,57

Average Ranks (Top 10, Final Results)

	Ranks	Points
Dinh&Feyco	1,5	5
lucy@molly	3,3	3
InformationRetrievalImWinter	3,5	2
BigGoogle	5,0	1
newDimensions	5,2	1
bl_jp	5,6	
Roland&Elias	5,8	
Oceanic2	8,2	
TeamJA	8,2	
lc+mk	8,3	

Old Ranking (Assignments 1-3)

Group	Points
Dinh&Feyco	15
lucy@molly	6
Roland&Elias (InfoRet2000)	4
ilGruppo	3
BigGoogle	3
BeSharps	3
NewDimensions	2
TeamJA	2
Julian&Dennis	1
jp&bl	1
HendrikundSinje	1

Current Ranking (Assignments 1-4)

Group	Points
Dinh&Feyco	20
lucy@molly	9
Roland&Elias (InfoRet2000)	4
BigGoogle	4
ilGruppo	3
BeSharps	3
NewDimensions	3
TeamJA	2
InformationRetrievalImWinter	2
Julian&Dennis	1
jp&bl	1
HendrikundSinje	1