



# Information Retrieval Exercises

Assignment 1:

**IMDB Spider - Solutions**

Mario Sanger ([saengema@informatik.hu-berlin.de](mailto:saengema@informatik.hu-berlin.de))

# Submission

---

- The vast majority of the groups respected the submission details (e.g. submit executable JARs)
  - But: Source code was missing a few times
- Implementation of the IMDBSpider was correct in general
  - Utilization of JSoup, XPath or CSS selectors to retrieve the information
  - Common issues: Missing entries in character list due to different format styles, handling of gross field
- Implementation of IMDBQueries was more challenging:
  - Missing selections (e.g. restriction to US or SciFi movies)
  - Respect the correct text case (e.g. Q6: Uncreative writer)

# Evaluation setup

---

- Evaluation was performed on a dedicated server of our research group
  - Evaluation set consisted of 10k movies
- Each query of each team was executed five times
  - Average runtime (in ms) was recorded
  - Queries which took longer than 20 minutes were excluded
  - Moreover: Only correctly implemented queries were considered
- For each query a ranking was calculated
  - The team with the highest average rank (concerning only the five best ranks of each team) wins

# Final results: Easy queries (incl. amendment)

Group	Q1		Q2		Q3		Q4		Q5	
	Runtime	Rank	Runtime	Rank	Runtime	Rank	Runtime	Rank	Runtime	Rank
C418	143	5	91	1	-	-	108	18	-	-
Codeworms	232	18	-	-	172	4	32	5	230	6
DominanteWunderWiesel	190	12	-	-	370	10	87	16	130	5
DoppelW	-	-	187	9	173	5	72	12	88	2
EBIR2018	212	17	180	6	-	-	57	10	--	-
Einfallslös	206	16	-	-	-	-	143	20	250	8
Fiedner	98	2	156	3	2122	13	39	6	72	1
FindingMovies	52	1	157	4	198	6	-	-	-	-
Freudon	159	7	-	-	-	-	73	13	-	-
Google	186	11	-	-	-	-	25	2	-	-
GroupNo1	326	20	176	5	-	-	49	9	-	-
KnockKnock	173	10	-	-	269	8	58	11	-	-
KSMR	194	13	-	-	391	11	41	7	-	-
Lejunds	139	4	341	11	-	-	29	4	-	-
Lilaloi	144	6	-	-	-	-	19	1	-	-
Lovelace	313	19	181	7	135	1	-	-	109	3
Nebula	159	7	185	8	161	2	74	14	110	4
Pony-und-kleid	199	14	211	10	-	-	100	17	306	9
RickUndFranz	200	15	-	-	217	7	79	15	-	-
Überwinder	495	21	-	-	163	3	46	8	-	-
Valeria	172	9	-	-	304	9	112	19	-	-
WirSindGroot	131	3	143	2	420	12	28	3	235	7

# Final results: Hard queries (incl. amendment)

Group	Q6		Q7		Q8		Q9		Q10	
	Runtime	Rank	Runtime	Rank	Runtime	Rank	Runtime	Rank	Runtime	Rank
C418	2408	6	380	1	124	6	81	10	6719	5
Codeworms	-	-	680	8	34	1	29	2	-	-
DominanteWunderWiesel	817	3	586	6	51	2	57	7	5214	2
DoppelW	998	4	577	5	-	-	27	1	10799	7
EBIR2018	-	-	-	-	248	10	217	14	-	-
Einfallslos	-	-	-	-	139	7	65	8	-	-
Fiedner	1883	5	1533	10	84	3	48	4	-	-
FindingMovies	-	-	-	-	174	9	119	11	-	-
Freudon	-	-	-	-	617	13	322	17	-	-
Google	-	-	-	-	-	-	-	-	-	-
GroupNo1	-	-	-	-	149	8	149	13	-	-
KnockKnock	-	-	502	3	-	-	-	-	8477	6
KSMR	-	-	1663	11	-	-	69	9	5788	3
Lejunds	-	-	-	-	116	5	46	3	-	-
Lilaloi	-	-	612	7	-	-	-	-	-	-
Lovelace	-	-	471	2	99	4	51	5	3945	1
Nebula	726	1	554	4	1070	14	-	-	-	-
Pony-und-kleid	-	-	-	-	-	-	55	6	-	-
RickUndFranz	797	2	-	-	304	12	297	16	-	-
Überwinder	-	-	-	-	269	11	124	12	-	-
Valeria	-	-	999	9	-	-	-	-	6299	4
WirSindGroot	1067743	7	682959	12	1776	15	285	15	-	-

# Final results: Complete ranking (incl. amendment)

Group	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
C418	5	1	-	18	-	6	1	6	10	5
Codeworms	18	-	4	5	6	-	8	1	2	-
DominanteWunderWiesel	12	-	10	16	5	3	6	2	7	2
DoppelW	-	9	5	12	2	4	5	-	1	7
EBIR2018	17	6	-	10	-	-	-	10	14	-
Einfallslos	16	-	-	20	8	-	-	7	8	-
Fiedner	2	3	13	6	1	5	10	3	4	-
FindingMovies	1	4	6	-	-	-	-	9	11	-
Freudon	7	-	-	13	-	-	-	13	17	-
Google	11	-	-	2	-	-	-	-	-	-
GroupNo1	20	5	-	9	-	-	-	8	13	-
KnockKnock	10	-	8	11	-	-	3	-	-	6
KSMR	13	-	11	7	-	-	11	-	9	3
Lejunds	4	11	-	4	-	-	-	5	3	-
Lilaloi	6	-	-	1	-	-	7	-	-	-
Lovelace	19	7	1	-	3	-	2	4	5	1
Nebula	7	8	2	14	4	1	4	14	-	-
Pony-und-kleid	14	10	-	17	9	-	-	-	6	-
RickUndFranz	15	-	7	15	-	2	-	12	16	-
Überwinder	21	-	3	8	-	-	-	11	12	-
Valeria	9	-	9	19	-	-	9	-	-	4
WirSindGroot	3	2	12	3	7	7	12	15	15	-

# Final results: Best rankings (incl. amendment)

Group	Best Rank	Second Best Rank	Third Best Rank	Fourth Best Rank	Fifth Best Rank	Average	Rank
C418	1	1	5	5	6	3.60	4
Codeworms	1	2	4	5	6	3.60	4
DominanteWunderWiesel	2	2	3	5	6	3.60	4
DoppelW	1	2	4	5	5	3.40	3
EBIR2018	6	10	10	14	17	11.40	18
Einfallslos	7	8	8	16	20	11.80	19
Fiedner	1	2	3	3	4	2.60	2
FindingMovies	1	4	6	9	11	6.20	10
Freudon	7	13	13	17	-		
Google	2	11	-	-	-		
GroupNo1	5	8	9	13	20	11.00	15
KnockKnock	3	6	8	10	11	7.60	11
KSMR	3	7	9	11	11	8.20	12
Lejunds	3	4	4	5	11	5.40	9
Lilaloi	1	6	7	-	-		
Lovelace	1	1	2	3	4	2.20	1
Nebula	1	2	4	4	7	3.60	4
Pony-und-kleid	6	9	10	14	17	11.20	17
RickUndFranz	2	7	12	15	15	10.20	14
Überwinder	3	8	11	12	21	11.00	15
Valeria	4	9	9	9	19	10.00	13
WirSindGroot	2	3	3	7	7	4.40	8

# Competition: Easy queries

---

- Only teams with at least 5 correctly implemented queries (at the submission deadline) participate in the competition

Group	Q1		Q2		Q3		Q4		Q5	
	Runtime	Rank	Runtime	Rank	Runtime	Rank	Runtime	Rank	Runtime	Rank
C418	143	2	91	1	-	-	108	7	-	-
Codeworms	232	8	-	-	-	-	32	2	230	4
DominanteWunderWiesel	190	5	-	-	-	-	87	6	-	-
DoppelW	-	-	187	5	173	3	72	5	88	1
Einfallslös	206	7	-	-	-	-	143	8	250	6
KnockKnock	173	4	-	-	269	4	58	4	-	-
KSMR	194	6	-	-	391	5	41	3	-	-
Lovelace	313	9	181	3	135	1	-	-	109	2
Nebula	159	3	185	4	161	2	-	-	110	3
WirSindGroot	131	1	143	2	420	6	28	1	235	5



# Competition: Hard queries

---

- Only teams with at least 5 correctly implemented queries (at the submission deadline) participate in the competition

Group	Q6		Q7		Q8		Q9		Q10	
	Runtime	Rank	Runtime	Rank	Runtime	Rank	Runtime	Rank	Runtime	Rank
C418	2408	2	380	1	124	4	81	6	6719	3
Codeworms	-	-	680	7	34	1	-	-	-	-
DominanteWunderWiesel	-	-	586	6	51	2	57	3	-	-
DoppelW	998	1	577	5	-	-	27	1	10799	5
Einfallslos	-	-	-	-	139	5	65	4	-	-
KnockKnock	-	-	502	3	-	-	-	-	8477	4
KSMR	-	-	1663	8	-	-	69	5	5788	2
Lovelace	-	-	471	2	99	3	51	2	3945	1
Nebula	-	-	554	4	-	-	-	-	-	-
WirSindGroot	1067743	3	682959	9	1776	6	285	7	-	-

# Competition: Complete ranking

---

- Only teams with at least 5 correctly implemented queries (at the submission deadline) participate in the competition

Group	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
C418	2	1	-	7	-	2	1	4	6	3
Codeworms	8	-	-	2	4	-	7	1	-	-
DominanteWunderWiesel	5	-	-	6	-	-	6	2	3	-
DoppelW	-	5	3	5	1	1	5	-	1	5
Einfalllos	7	-	-	8	6	-	-	5	4	-
KnockKnock	4	-	4	4	-	-	3	-	-	4
KSMR	6	-	5	3	-	-	8	-	5	2
Lovelace	9	3	1	-	2	-	2	3	2	1
Nebula	3	4	2	-	3	-	4	-	-	-
WirSindGroot	1	2	6	1	5	3	9	6	7	-

# Competition: Best rankings

---

- Only teams with at least 5 correctly implemented queries (at the submission deadline) participate in the competition

Group	Best Rank	Second Best Rank	Third Best Rank	Fourth Best Rank	Fifth Best Rank	Average	Rank
C418	1	1	2	2	3	1.80	2
Codeworms	1	2	4	7	8	4.40	8
DominanteWunderWiesel	2	3	5	6	6	6.20	10
DoppelW	1	1	1	3	5	2.20	3
Einfallslös	4	5	6	7	8	6.00	9
KnockKnock	3	4	4	4	4	3.80	6
KSMR	2	3	5	5	6	4.20	7
Lovelace	1	1	2	2	2	1.60	1
Nebula	2	3	3	4	4	3.60	5
WirSindGroot	1	1	2	3	5	2.40	4

# Competition

---

- Points:
  - 5 Points: Lovelace
  - 3 Points: C418
  - 2 Points: DoppelW
  - 1 Point: Nebula
  - 1 Point: WirSindGroot
- Extra points for custom queries:
  - 1 Point: Pony-&-Kleid

---

# Questions?