Information Retrieval Evaluation in the Field of 10m Air Rifle Shooting

CE205 Assignment 2 2015-16

Evaldas Senavaitis (1402039)

1. Topic and Queries

Questions(Queries):

1. What is shooting?
2. Is 10m air rifle shooting most popular form of shooting?
3. Is it true that air rifle has only one position?
4. Is 10.9 score is highest in shooting competitions?
5. Is shooting Olympic games competition?
6. What guns can be used for shooting competitions?
7. When was shooting first introduced into Olympic games?
8. What positions does rifle shooting has?
9. How many total shots need to be fired for men?
10. How many shots are fired for finals?
11. Is it true that maximum weight of a rifle must be 5.5kg?
12. Maximum score for men competition at 10m air rifle?
13. Is electronic targets are used for shooting competitions?
14. Is Wiesbaden was first ever world championship city for shooting?
15. Is typical 4.5mm 10m air rifle match bullet called pellet?
16. Is air rifle competitions are supervised by ISSF rules?
17. Is AR60 used for men 10m air rifle competition?
18. Is AR40 used for women 10m air rifle competition?
19. Has Johann Riederer ever won any medals in 10m air rifle Olympics?
20. Is Philippe Heberlé Olympic champion?

I chose 10m Air rifle shooting topic, because I used to compete in these kind of competitions, I had trained for about 4 years in this sport. I achieved great results in that time, and won many medals at my age group. I managed to achieve a stage of masterie that people couldn't reach in 6-8 years time. This topic has a special place in my heart, because people that I trained with became like family to me and I always recommend this sport if people want to find bodie and mind balance control.

To devise queries was pretty simple, because I have personal experience about this topic. Most of the queries I given I know the answers, but I used Wikipedia to lookup for some more specific questions about former Olympic games champions. Transforming my queries to seach queries is fairly simple as I focus on finding pieces of my questions in Wikipedia mostly using AND operator because I don't want to exclude any information as it is all important to the research.

2. Indexing the Documents

Indexing was the same as it was in any other previous lab, it took longer than usual to index whole document groups, but it did not encounter any problems.

3. BM25 Performance

3.1 Method

After indexing I used Whoosh library function for BM25 to help me get first 20 results from index. After that I looked at each one finding if it answers my querie and marking it if it is relevant or not. At the end, having all this data it helped me to summarise the numerical results in the table.

3.2 Results

Include the following table, duly completed with your results. Numbers to two decimal places exactly as shown in the table below. The last line is for the averages - examples are shown.

Num	**Query**	**P (n=5)**	**P (n=10)**	**R (n=5)**	**R (n=10)**
1	"Shooting"	0.20	0.10	0.50	0.50
2	"Shooting AND popular"	0.20	0.10	1.00	1.00
3	"Air rifle AND position"	0.40	0.30	0.40	0.60
4	"10.9 AND highest AND shooting"	0.00	0.00	0.00	0.00
5	"Shooting AND Olympic"	1.00	1.00	0.25	0.50
6	"Guns AND shooting"	0.40	0.20	1.00	1.00
7	"First Olympic games AND shooting"	0.20	0.10	0.33	0.33
8	"Rifle AND positions"	0.60	0.40	0.43	0.57
9	"Shots AND men"	0.60	0.40	0.60	0.80
10	"Shots AND finals"	0.20	0.10	1.00	1.00
11	"Maximum AND weight AND rifle"	0.00	0.00	0.00	0.00
12	"Maximum AND score AND men AND rifle"	0.20	0.10	1.00	1.00
13	"Electronics AND shooting"	0.00	0.00	0.00	0.00
14	"Championships AND shooting"	0.00	0.00	0.00	0.00
15	"Pellet AND rifle"	0.40	0.20	1.00	1.00
16	"Rifle AND ISSF"	0.00	0.00	0.00	0.00
17	"AR60 AND men"	0.00	0.00	0.00	0.00
18	"AR40 AND women"	0.00	0.00	0.00	0.00
19	"Johann Riederer AND Olympics"	0.00	0.00	0.00	0.00
20	"Philippe Heberlé AND Olympic"	0.00	0.00	0.00	0.00
Avg		**0.22**	**0.15**	**0.38**	**0.42**

3.3 Discussion

Results show that BM25 gives some relevant data at first few result output. My my queries it was not really specific for what to look for, maybe that is why it have a lot of irelavant data for me to go through. What really caugh my eye is querie 12 and querie 15 for some reason they only returned 2 and 8 total results perspectivly. I took it to my calculations and any rest data would be irrelevant even if I would get more total results. When I look at querie 12 I do not know why but it cant find a specific result even when I have seen Wikipedia page with the right answer. For querie 15 I just guess that there is not really many pages for pellets in Wikipedia. I got total 7 queries totally with none relevant results, or it simply outputted 0 pages. Last four queries got my attention because I don't know why it would not find anything about those queries, I guess maybe because it is old Wikipedia and it should be updated. What really bugged me that I could not find 2 famous Olympic medalist when it is clearly stated in up to date Wikipedia that they exits in their library.

4. TF*IDF Performance (not compulsory, see marking scheme)

4.1 Method

For TF*IDF I did the same as for BM25, I indexed, looked through all the results, took out relevant data, made calculations and make my assumptions.

4.2 Results

Include the following, table duly completed.

Num	Query	P (n=5)	P (n=10)	R (n=5)	R (n=10)
1	"Shooting"	0.40	0.30	0.50	0.75
2	"Shooting AND popular"	0.00	0.00	0.00	0.00
3	"Air rifle AND position"	0.00	0.00	0.00	0.00
4	"10.9 AND highest AND shooting"	0.00	0.00	0.00	0.00
5	"Shooting AND Olympic"	1.00	0.70	0.45	0.64
6	"Guns AND shooting"	0.20	0.20	0.33	0.67
7	"First Olympic games AND shooting"	0.40	0.20	1.00	1.00
8	"Rifle AND positions"	0.00	0.00	0.00	0.00
9	"Shots AND men"	0.00	0.00	0.00	0.00
10	"Shots AND finals"	0.00	0.10	0.00	1.00
11	"Maximum AND weight AND rifle"	0.00	0.00	0.00	0.00
12	"Maximum AND score AND men AND rifle"	0.00	0.00	0.00	0.00
13	"Electronics AND shooting"	0.00	0.00	0.00	0.00
14	"Championships AND shooting"	0.00	0.00	0.00	0.00
15	"Pellet AND rifle"	0.20	0.10	1.00	1.00
16	"Rifle AND ISSF"	0.20	0.20	0.50	0.50
17	"AR60 AND men"	0.00	0.00	0.00	0.00
18	"AR40 AND women"	0.00	0.00	0.00	0.00
19	"Johann Riederer AND Olympics"	0.00	0.00	0.00	0.00
20	"Philippe Heberlé AND Olympic"	0.00	0.00	0.00	0.00
Avg		**0.12**	**0.09**	**0.20**	**0.30**

4.3 Discussion

Results showed that TF*IDF performed worst in every aspect. It not only did not find relevant data, but it completely gave irrelevant data that is nothing to do with my topic. Looked at the table you can see that it showed similar results to BM25, but overral picture of TF*IDF does not look good. Maybe if queries would be more specified on looking for several words not just a couple of them, it would have gave me more relevant data, but it is only predictions and I would not count on it. If we look at results after number 10 it most of the time gave me atleast 1 relevant page, but first 10 were completely irrelevant. At the end I would suggest using BM25 as it gives more results that are relevant to the topic.

5. Features Added to Whoosh (not compulsory, see marking scheme)

5.1 First Feature - Snippets (just an example)

Explain how you implemented it. State which parts of your submitted program implement it. Also include screen shots of your code actually working.

5.2 Second Feature - Display of Documents (just an example)

Explain how you implemented it. State which parts of your submitted program implement it. Also include screen shots of your code actually working.

Appendix 1

Include the queries you used for your BM25 evaluation and list the IDs of up to four right answers found in the first ten responses returned by the system, listed in the order in which they are returned. Note that there might be as many as ten right answers found in the first ten, or as few as zero. Generally, only a minority of your queries should have no answers at all in the first ten. If more then four correct answers are returned, just list the first four here. Note that we can use this information to verify that these answers are really returned in response to the query, and that these answers are indeed 'correct' answers.

Num	Query	IDs of Answers
1	"Shooting"	28495
2	"Shooting AND popular"	2452530
3	"Air rifle AND position"	2940587, 2775293, 2452530,
4	"10.9 AND highest AND shooting"	None
5	"Shooting AND Olympic"	2528826, 2771181, 924680, 2529371
6	"Guns AND shooting"	28498, 875653,
7	"First Olympic games AND shooting"	2528826, 2255965
8	"Rifle AND positions"	2451920, 2830613, 960818,
9	"Shots AND men"	2451920, 2452537, 2452530, 2482304
10	"Shots AND finals"	438961
11	"Maximum AND weight AND rifle"	None
12	"Maximum AND score AND men AND rifle"	438961
13	"Electronics AND shooting"	None
14	"Championships AND shooting"	None
15	"Pellet AND rifle"	145263, 1387010,
16	"Rifle AND ISSF"	2529371, 2451943
17	"AR60 AND men"	None
18	"AR40 AND women"	None
19	"Johann Riederer AND Olympics"	None
20	"Philippe Heberlé AND Olympic"	None

Appendix 2

Include the same queries as in Appendix 1 plus the IDs of the right answers found when TF*IDF was used.

Num	Query	IDs of Answers
1	"Shooting"	28498, 875653,
2	"Shooting AND popular"	None
3	"Air rifle AND position"	None
4	"10.9 AND highest AND shooting"	None
5	"Shooting AND Olympic"	22576, 863486, 48644, 869815
6	"Guns AND shooting"	145263, 28498
7	"First Olympic games AND shooting"	22576, 48644
8	"Rifle AND positions"	None
9	"Shots AND men"	None

10	“Shots AND finals”	438961
11	“Maximum AND weight AND rifle”	None
12	“Maximum AND score AND men AND rifle”	None
13	“Electronics AND shooting”	None
14	“Championships AND shooting”	None
15	“Pellet AND rifle”	145263
16	“Rifle AND ISSF”	438961, 869815, 2529371
17	“AR60 AND men”	None
18	“AR40 AND women”	None
19	“Johann Riederer AND Olympics”	None
20	“Philippe Heberlé AND Olympic”	None

Reminder: When you submit your report, make sure you convert to .pdf first! .doc files cannot be accepted. Then place your report with your .py code file in a .zip and submit to Faser, following the instructions exactly.