

Riti Gupta
Samanvitha Basole
Neha Gaikwad

Experiments

We chose the following links and made the following queries for 6 topics:

- https://en.wikipedia.org/wiki/Fly_fishing
- <https://www.artofmanliness.com/articles/fly-fishing/>
- <https://www.backcountry.com/explore/fly-fishing-101-an-intro-to-flies>
- <https://1source.basspro.com/index.php/component/k2/67-fly-fishing/3854-a-simple-guide-to-choosing-your-first-fly-fishing-outfit>
- <https://www.takemefishing.org/fly-fishing/>
- https://en.wikipedia.org/wiki/Flying_fish
- <https://www.nationalgeographic.com/animals/fish/group/flying-fish/>
- http://www.softschools.com/facts/animals/flying_fish_facts/318/
- <https://www.nwf.org/Educational-Resources/Wildlife-Guide/Fish/Flying-Fish>
- <http://fishlaboratory.com/fish/flying-fish>

Topic 1 for Fly Fishing: bait

Where to buy fly fishing bait?
Lures and flies for sale?
Fly fishing lures on sale?

Topic 2 for Fly Fishing: intro

What is fly fishing?
Equipment and outfit for fly fishing?
What are the different types of flies for fly fishing?

Topic 3 for Fly Fishing: rod

What are the best fly fishing rods?
Where to buy fly fishing rods
Fly fishing rods for sale

Topic 1 for Flying Fish: intro

What is a flying fish?

california flying fish

What do flying fish eat?

Topic 2 for Flying Fish: facts

flying fish diet?

flying fish habitat?

flying fish adapt?

Topic 3 for Flying Fish: random questions

Is flying fish a food?

how does flying fish survive?

how much is a flying fish?

Then, we made our stemmed corpus and the non-stemmed corpus which contained the content from the 10 links listed above. We used the disjunctive cosine ranking to rank the 10 documents.

Here are the results for the non-stemmed corpus:

Score for query: 1

where to buy fly fishing bait

5 0.0350

2 0.0253

4 0.0234

7 0.0218

10 0.0205

1 0.0180

3 0.0135

6 0.0105

8 0.0000

9 0.0000

Score for query: 2

lures and flies for sale

3 0.0816
5 0.0679
2 0.0330
4 0.0313
1 0.0290
6 0.0077
9 0.0009
10 0.0006
7 0.0005
8 0.0000

Score for query: 3

fly fishing lures on sale

3 0.0552
5 0.0526
4 0.0274
2 0.0233
1 0.0217
6 0.0100
10 0.0062
7 0.0006
8 0.0000
9 0.0000

Score for query: 4

what is fly fishing

5 0.0692
3 0.0455
2 0.0437
4 0.0291
1 0.0208
10 0.0137
7 0.0085
6 0.0074
8 0.0000
9 0.0000

Score for query: 5

equipment and outfit for fly fishing

4 0.1096
1 0.0230
5 0.0123

3 0.0072
2 0.0058
10 0.0029
6 0.0011
7 0.0005
9 0.0004
8 0.0000

Score for query: 6

what are the different types of flies for fly fishing

3 0.0997
5 0.0903
2 0.0696
1 0.0431
4 0.0342
8 0.0186
6 0.0150
10 0.0083
7 0.0051
9 0.0022

Score for query: 7

what are the best fly fishing rods

5 0.1067
4 0.0483
3 0.0458
1 0.0321
7 0.0245
2 0.0245
10 0.0071
6 0.0036
8 0.0013
9 0.0011

Score for query: 8

where to buy fly fishing rods

5 0.0411
4 0.0289
3 0.0267
1 0.0236
2 0.0224
7 0.0211
10 0.0199

6 0.0071
8 0.0000
9 0.0000

Score for query: 9
fly fishing rods for sale

3 0.0734
5 0.0531
4 0.0318
1 0.0280
2 0.0186
6 0.0105
10 0.0068
7 0.0012
9 0.0010

Score for query: 10
what is a flying fish

8 0.0816
10 0.0646
9 0.0585
7 0.0374
6 0.0349
2 0.0136
3 0.0091
5 0.0064
4 0.0053
1 0.0044

Score for query: 11
california flying fish

5 0.0481
7 0.0352
8 0.0301
10 0.0238
9 0.0216
6 0.0120
1 0.0011
2 0.0000
3 0.0000
4 0.0000

Score for query: 12

what do flying fish eat

10 0.0672

9 0.0655

8 0.0417

7 0.0297

2 0.0263

6 0.0203

5 0.0100

3 0.0094

4 0.0070

1 0.0067

Score for query: 13

flying fish diet

9 0.0712

10 0.0606

8 0.0591

7 0.0311

6 0.0300

1 0.0060

2 0.0000

3 0.0000

4 0.0000

5 0.0000

Score for query: 14

flying fish habitat

9 0.0554

8 0.0485

6 0.0431

10 0.0384

4 0.0240

7 0.0185

1 0.0073

2 0.0000

3 0.0000

5 0.0000

Score for query: 15

flying fish adapt

8 0.0389

7 0.0319

10 0.0308

6 0.0283
9 0.0278
1 0.0091
2 0.0000
3 0.0000
4 0.0000
5 0.0000

Score for query: 16
is flying fish a food

8 0.0816
10 0.0697
9 0.0637
7 0.0374
6 0.0349
1 0.0080
4 0.0041
2 0.0027
3 0.0000
5 0.0000

Score for query: 17
how does flying fish survive

10 0.0362
2 0.0360
8 0.0242
7 0.0229
6 0.0183
9 0.0173
1 0.0099
5 0.0076
3 0.0045
4 0.0024

Score for query: 18
how much is a flying fish

8 0.0459
3 0.0423
2 0.0396
10 0.0363
9 0.0329
1 0.0261
7 0.0234

6 0.0196
5 0.0144
4 0.0121

Here are the results for the stemmed corpus:

Score for query: 1
where to buy fly fish bait

2 0.0348
4 0.0233
5 0.0218
10 0.0211
1 0.0164
7 0.0156
6 0.0119
3 0.0042
8 0.0000
9 0.0000

Score for query: 2
lure and fly for sale

5 0.0835
4 0.0295
3 0.0258
1 0.0175
6 0.0132
2 0.0086
9 0.0005
10 0.0004
7 0.0003
8 0.0000

Score for query: 3
fly fish lure on sale

1 0.0199
6 0.0175
2 0.0095
4 0.0083
3 0.0065
7 0.0044
5 0.0000
8 0.0000
9 0.0000

10 0.0000

Score for query: 4

what is fly fishing

5 0.1086

4 0.0414

3 0.0352

2 0.0330

1 0.0270

7 0.0022

9 0.0018

6 0.0008

8 0.0000

10 0.0000

Score for query: 5

equipment and outfit for fly fishing

4 0.1152

5 0.0271

1 0.0134

3 0.0082

2 0.0077

9 0.0004

10 0.0003

7 0.0002

6 0.0002

8 0.0000

Score for query: 6

what ar the differ type of fli for fly fishing

5 0.1098

3 0.0873

4 0.0479

2 0.0460

1 0.0437

8 0.0118

6 0.0041

9 0.0036

7 0.0034

10 0.0022

Score for query: 7

what ar the best fly fish rods

5 0.0953
4 0.0394
3 0.0333
1 0.0221
7 0.0220
2 0.0022
9 0.0020
8 0.0014
10 0.0012
6 0.0009

Score for query: 8
where to buy fly fish rods

5 0.0418
4 0.0307
2 0.0259
3 0.0238
1 0.0207
10 0.0188
7 0.0138
6 0.0068
8 0.0000
9 0.0000

Score for query: 9
fly fish rod for sale

1 0.0252
3 0.0221
5 0.0182
6 0.0179
4 0.0166
2 0.0084
9 0.0007
10 0.0005
7 0.0004
8 0.0000

Score for query: 10
what is a fly fish

3 0.0136
2 0.0104
5 0.0072
7 0.0071

9 0.0059
4 0.0058
6 0.0026
1 0.0016
8 0.0000
10 0.0000

Score for query: 11
california fly fish

5 0.0519
7 0.0256
1 0.0000
2 0.0000
3 0.0000
4 0.0000
6 0.0000
8 0.0000
9 0.0000
10 0.0000

Score for query: 12
what do fly fish eat

10 0.0352
9 0.0300
2 0.0274
8 0.0197
5 0.0113
7 0.0104
3 0.0103
4 0.0074
1 0.0069
6 0.0041

Score for query: 13
fly fish diet

9 0.0366
10 0.0175
7 0.0110
6 0.0082
1 0.0051
2 0.0000
3 0.0000
4 0.0000

5 0.0000
8 0.0000

Score for query: 14
fly fish habitat

6 0.0281
4 0.0277
9 0.0242
1 0.0067
2 0.0000
3 0.0000
5 0.0000
7 0.0000
8 0.0000
10 0.0000

Score for query: 15
fly fish adapt

7 0.0191
1 0.0177
6 0.0143
2 0.0000
3 0.0000
4 0.0000
5 0.0000
8 0.0000
9 0.0000
10 0.0000

Score for query: 16
i fly fish a food

2 0.0569
10 0.0317
3 0.0272
7 0.0032
9 0.0027
1 0.0027
4 0.0020
6 0.0012
5 0.0000
8 0.0000

Score for query: 17

how doe fly fish survive

7 0.0321

2 0.0249

1 0.0071

5 0.0058

3 0.0035

4 0.0018

6 0.0005

8 0.0000

9 0.0000

10 0.0000

Score for query: 18

how much is a fly fish

3 0.0504

2 0.0347

1 0.0290

5 0.0167

4 0.0138

7 0.0068

6 0.0015

8 0.0000

9 0.0000

10 0.0000

Next, we calculate the precision for the non-stemmed queries/corpus.

Query 1 precision: 5/8

Query 2 precision: 5/9

Query 3 precision: 5/8

Query 4 precision: 5/8

Query 5 precision: 5/9

Query 6 precision: 5/10

Query 7 precision: 5/10

Query 8 precision: 5/8

Query 9 precision: 5/10

Query 10 precision: 5/10

Query 11 precision: 5/7

Query 12 precision: 5/10

Query 13 precision: 5/6

Query 14 precision: 5/7

Query 15 precision: 5/6

Query 16 precision: 5/8

Query 17 precision: 5/10

Query 18 precision: 5/10

Then, we calculate the precision for the stemmed queries/corpus.

Query 1 precision: 5/8

Query 2 precision: 5/9

Query 3 precision: 4/6

Query 4 precision: 5/8

Query 5 precision: 5/9

Query 6 precision: 5/10

Query 7 precision: 5/10

Query 8 precision: 5/8

Query 9 precision: 5/9

Query 10 precision: 3/8

Query 11 precision: 1/2

Query 12 precision: 5/10

Query 13 precision: 4/5

Query 14 precision: 2/4

Query 15 precision: 2/3

Query 16 precision: 4/8

Query 17 precision: 2/7

Query 18 precision: 2/7

Relevant docs for topic 1, 2, 3 - [docs 1, 2, 3, 4, 5]

Relevant docs for topic 4, 5, 6 - [docs 6, 7, 8, 9, 10]

Next, we calculate the average precision for the non-stemmed queries/corpus.

Query 1 Average precision - $1/5 * (1/1 + 2/2 + 3/3) = 3/5$

Query 2 Average precision - $1/5 * (1/1 + 2/2 + 3/3 + 4/4 + 5/5) = 1.0$

Query 3 Average precision - $1/5 * (1/1 + 2/2 + 3/3 + 4/4 + 5/5) = 1.0$

Query 4 Average precision - $1/5 * (1/1 + 2/2 + 3/3 + 4/4 + 5/5) = 1.0$

Query 5 Average precision - $1/5 * (1/1 + 2/2 + 3/3 + 4/4 + 5/5) = 1.0$

Query 6 Average precision - $1/5 * (1/1 + 2/2 + 3/3 + 4/4 + 5/5) = 1.0$

Query 7 Average precision - $1/5 * (1/1 + 2/2 + 3/3 + 4/4) = 4/5$

Query 8 Average precision - $1/5 * (1/1 + 2/2 + 3/3 + 4/4 + 5/5) = 1.0$

Query 9 Average precision - $1/5 * (1/1 + 2/2 + 3/3 + 4/4 + 5/5) = 1.0$

Query 10 Average precision - $1/5 * (1/2 + 2/3 + 3/4 + 4/5) = 0.543$

Query 11 Average precision - $1/5 * (1/2 + 2/3 + 3/4 + 3/5) = 0.503$

Query 12 Average precision - $1/5 * (1/1 + 2/2 + 3/3 + 4/4) = 0.8$

Query 13 Average precision - $1/5 * (1/1 + 2/2 + 3/3 + 4/4 + 5/5) = 1.0$

Query 14 Average precision - $1/5 * (1/1 + 2/2 + 3/3 + 4/4) = 0.8$
 Query 15 Average precision - $1/5 * (1/1 + 2/2 + 3/3 + 4/4 + 5/5) = 1.0$
 Query 16 Average precision - $1/5 * (1/1 + 2/2 + 3/3 + 4/4 + 5/5) = 1.0$
 Query 17 Average precision - $1/5 * (1/1 + 2/3 + 3/4 + 4/5) = 0.64$
 Query 18 Average precision - $1/5 * (1/1 + 2/4 + 3/5) = 0.42$

Then, we calculate the average precision for each of the topics (non-stemmed):

Topic 1 average precision: 0.86

Topic 2 average precision: 1.0

Topic 3 average precision: 0.93

Topic 4 average precision: 0.615

Topic 5 average precision: 0.93

Topic 6 average precision: 0.68

Then, we calculate the average precision for the stemmed queries/corpus.

Query 1 Average precision: $1/5 * (1/1 + 2/2 + 3/3 + 4/5) = 0.76$
 Query 2 Average precision: $1/5 * (1/1 + 2/2 + 3/3 + 4/4) = 4/5$
 Query 3 Average precision: $1/5 * (1/1 + 2/3 + 3/4 + 4/5) = 0.64$
 Query 4 Average precision: $1/5 * (1/1 + 2/2 + 3/3 + 4/4 + 5/5) = 1.0$
 Query 5 Average precision: $1/5 * (1/1 + 2/2 + 3/3 + 4/4 + 5/5) = 1.0$
 Query 6 Average precision: $1/5 * (1/1 + 2/2 + 3/3 + 4/4 + 5/5) = 1.0$
 Query 7 Average precision: $1/5 * (1/1 + 2/2 + 3/3 + 4/4) = 0.8$
 Query 8 Average precision: $1/5 * (1/1 + 2/2 + 3/3 + 4/4 + 5/5) = 1.0$
 Query 9 Average precision: $1/5 * (1/1 + 2/2 + 3/3 + 4/5) = 0.76$
 Query 10 Average precision: $1/5 * (1/4 + 2/5) = 0.13$
 Query 11 Average precision: $1/5 * (1/2) = 0.1$
 Query 12 Average precision: $1/5 * (1/1 + 2/2 + 3/4) = 0.55$
 Query 13 Average precision: $1/5 * (1/1 + 2/2 + 3/3 + 4/4) = 0.8$
 Query 14 Average precision: $1/5 * (1/1 + 2/3) = 0.33$
 Query 15 Average precision: $1/5 * (1/1 + 2/3) = 0.33$
 Query 16 Average precision: $1/5 * (1/2 + 2/4 + 3/5) = 0.32$
 Query 17 Average precision: $1/5 * (1/1) = 0.2$
 Query 18 Average precision: $1/5 * (0) = 0$

Then, we calculate the average precision for each of the 6 topics (stemmed):

Topic 1 average precision: 0.73

Topic 2 average precision: 1.0

Topic 3 average precision: 0.85

Topic 4 average precision: 0.33

Topic 5 average precision: 0.46

Topic 6 average precision: 0.173

Then, finally, to find the MAP, we add the average precision scores and divide by the number of topics.

In the case of the non-stemmed, we have

$$\text{MAP} = (0.86 + 1.0 + 0.93 + 0.615 + 0.93 + 0.68) / 6 = \mathbf{0.84}$$

In the case of the stemmed, we have

$$\text{MAP} = (0.73 + 1.0 + 0.85 + 0.33 + 0.46 + 0.173) / 6 = \mathbf{0.60}$$

Conclusion:

According to the above MAP results, it looks like stemming actually gave a lower MAP. This was probably because it considered “fly fish” and “fly fish” as the same thing. Stemming removed the ‘ing’ parts in ‘fishing’ and ‘flying’.

So:

fly fish(ing) = fly(ing) fish.

Hence, the lower score since non-relevant docs were showing up in stemmed results.