

Boolean Retrieval Baselines

As Indri allows to specify Boolean-style queries using the #band (Boolean AND), #or, and #not operators, we try to transform the Biocaddie "short" queries into Boolean queries using the Indri query language and rerun all the baselines using the Boolean queries.

1. Transform Biocaddie short queries into Boolean style

Using script **boolquery_converter.sh** (https://github.com/nds-org/biocaddie/blob/master/scripts/boolquery_converter.sh)

Output: **boolqueries.combined.short** (<https://github.com/nds-org/biocaddie/blob/master/queries/boolqueries.combined.short>)

2. Run IndriRunQuery on 6 baselines: tldif, okapi, dir, jm, two and rm3.

Results:

Dir, jm, two and rm3 work fine without errors.

```
root@integration-1:~/biocaddie/output/dir/boolean/combined/short# head 500.out
EA1 Q0 704525 1 -4.5747 indri
EA1 Q0 310477 2 -4.59887 indri
EA1 Q0 476233 3 -4.66138 indri
EA1 Q0 22007 4 -4.67851 indri
EA1 Q0 353184 5 -4.71317 indri
EA1 Q0 86285 6 -4.72443 indri
EA1 Q0 750253 7 -4.72995 indri
EA1 Q0 1330 8 -4.7484 indri
EA1 Q0 471708 9 -4.7619 indri
EA1 Q0 307331 10 -4.78789 indri
```

No exception found.

```
root@integration-1:~/biocaddie/output/dir/boolean/combined/short# grep -i exception *
root@integration-1:~/biocaddie/output/dir/boolean/combined/short#
```

```
root@integration-1:~/biocaddie/output/jm/boolean/combined/short# head lambda=0.0
EA1 Q0 733683 1 -2.77259 indri
EA1 Q0 327214 2 -3.46574 indri
EA1 Q0 476233 3 -3.54129 indri
EA1 Q0 149108 4 -3.56953 indri
EA1 Q0 685241 5 -3.61092 indri
EA1 Q0 310477 6 -3.61092 indri
EA1 Q0 121429 7 -3.61092 indri
EA1 Q0 245536 8 -3.61765 indri
EA1 Q0 449101 9 -3.62878 indri
EA1 Q0 155528 10 -3.63099 indri
```

No exception found.

```
root@integration-1:~/biocaddie/output/jm/boolean/combined/short# grep -i exception *
root@integration-1:~/biocaddie/output/jm/boolean/combined/short#
```

```
root@integration-1:~/biocaddie/output/two/boolean/combined/short# head mu=10000:lambda=0.0.out
EA1 Q0 733683 1 -2.77259 indri
EA1 Q0 327214 2 -3.46574 indri
EA1 Q0 476233 3 -3.54129 indri
EA1 Q0 149108 4 -3.56953 indri
EA1 Q0 685241 5 -3.61092 indri
EA1 Q0 310477 6 -3.61092 indri
EA1 Q0 121429 7 -3.61092 indri
EA1 Q0 245536 8 -3.61765 indri
EA1 Q0 449101 9 -3.62878 indri
EA1 Q0 155528 10 -3.63099 indri
```

No exception found.

```
root@integration-1:~/biocaddie/output/two/boolean/combined/short# grep -i exception *
root@integration-1:~/biocaddie/output/two/boolean/combined/short#
```

Tfidf, **okapi** got errors *"indri query language operators are not allowed"*

```
root@integration-1:~/biocaddie/output/tfidf/boolean/combined/short# cat k1=0:b=0.0.out
# EXCEPTION in query EA1: IndriRunQuery.cpp(645): Can't run baseline on this query:
    #or(
        #band(TGF- signaling)
        #band(TGF- pathway)
        #band(signaling pathway)
    )
indri query language operators are not allowed.
# EXCEPTION in query EA2: IndriRunQuery.cpp(645): Can't run baseline on this query:
    #or(
        #band(synaptic growth)
        #band(synaptic remodeling)
        #band(synaptic glycolysis)
        #band(synaptic human)
        #band(synaptic brain)
        #band(growth remodeling)
        #band(growth glycolysis)
        #band(growth human)
        #band(growth brain)
        #band(remodeling glycolysis)
        #band(remodeling human)
        #band(remodeling brain)
        #band(glycolysis human)
        #band(glycolysis brain)
        #band(human brain)
    )
indri query language operators are not allowed.
```

```
root@integration-1:~/biocaddie/output/tfidf/boolean/combined/short# grep -i exception * | head
k1=0.2:b=0.0.out:# EXCEPTION in query EA1: IndriRunQuery.cpp(645): Can't run baseline on this query:
k1=0.2:b=0.0.out:# EXCEPTION in query EA2: IndriRunQuery.cpp(645): Can't run baseline on this query:
k1=0.2:b=0.0.out:# EXCEPTION in query EA3: IndriRunQuery.cpp(645): Can't run baseline on this query:
k1=0.2:b=0.0.out:# EXCEPTION in query EA4: IndriRunQuery.cpp(645): Can't run baseline on this query:
k1=0.2:b=0.0.out:# EXCEPTION in query EA5: IndriRunQuery.cpp(645): Can't run baseline on this query:
k1=0.2:b=0.0.out:# EXCEPTION in query EA6: IndriRunQuery.cpp(645): Can't run baseline on this query:
k1=0.2:b=0.0.out:# EXCEPTION in query T1: IndriRunQuery.cpp(645): Can't run baseline on this query:
k1=0.2:b=0.0.out:# EXCEPTION in query T2: IndriRunQuery.cpp(645): Can't run baseline on this query:
k1=0.2:b=0.0.out:# EXCEPTION in query T3: IndriRunQuery.cpp(645): Can't run baseline on this query:
k1=0.2:b=0.0.out:# EXCEPTION in query T4: IndriRunQuery.cpp(645): Can't run baseline on this query:
```

```
root@integration-1:~/biocaddie/output/okapi/boolean/combined/short# head k1=1.0:k3=1.0:b=0.0.out
# EXCEPTION in query EA1: IndriRunQuery.cpp(645): Can't run baseline on this query:
    #or(
        #band(TGF- signaling)
        #band(TGF- pathway)
        #band(signaling pathway)
    )
indri query language operators are not allowed.
```

```

root@integration-1:~/biocaddie/output/okapi/boolean/combined/short# grep -i exception * | head
k1=1.0:k3=1.0:b=0.0.out:# EXCEPTION in query EA1: IndriRunQuery.cpp(645): Can't run baseline on this query:
k1=1.0:k3=1.0:b=0.0.out:# EXCEPTION in query EA2: IndriRunQuery.cpp(645): Can't run baseline on this query:
k1=1.0:k3=1.0:b=0.0.out:# EXCEPTION in query EA3: IndriRunQuery.cpp(645): Can't run baseline on this query:
k1=1.0:k3=1.0:b=0.0.out:# EXCEPTION in query EA4: IndriRunQuery.cpp(645): Can't run baseline on this query:
k1=1.0:k3=1.0:b=0.0.out:# EXCEPTION in query EA5: IndriRunQuery.cpp(645): Can't run baseline on this query:
k1=1.0:k3=1.0:b=0.0.out:# EXCEPTION in query EA6: IndriRunQuery.cpp(645): Can't run baseline on this query:
k1=1.0:k3=1.0:b=0.0.out:# EXCEPTION in query T1: IndriRunQuery.cpp(645): Can't run baseline on this query:
k1=1.0:k3=1.0:b=0.0.out:# EXCEPTION in query T2: IndriRunQuery.cpp(645): Can't run baseline on this query:
k1=1.0:k3=1.0:b=0.0.out:# EXCEPTION in query T3: IndriRunQuery.cpp(645): Can't run baseline on this query:
k1=1.0:k3=1.0:b=0.0.out:# EXCEPTION in query T4: IndriRunQuery.cpp(645): Can't run baseline on this query:

```

3. Run trec_eval (mkeval.sh) on dir, jm, two and RM3

Results:

Model	MAP	NDCG	P@20	NDCG@20	P@100	NDCG@100	Notes	Date
QL (JM)	0.0964	0.3063	0.2976	0.2444	0.1633	0.1964	Sweep lambda	06/02/17
QL (Dir)	0.1314	0.3826	0.3952	0.3501	0.231	0.2792	Sweep mu	06/02/17
QL (TS)	0.0964	0.3063	0.2976	0.2444	0.1633	0.1964	Sweep mu and lambda	06/02/17
RM3	0.1376	0.4149	0.3857	0.2962	0.2676	0.3101	Sweep mu, fbDocs, fbTerms, and lambda	06/04/17

```

root@integration-1:~/biocaddie# scripts/mkeval2.sh dir short combined
map      0.13139523809523812
ndcg     0.3826428571428571
ndcg_cut_5      0.3992285714285714
ndcg_cut_10     0.38498571428571426
ndcg_cut_20     0.35014761904761904
ndcg_cut_100    0.27921428571428575
P_5           0.49523809523809514
P_10          0.48095238095238085
P_20          0.39523809523809517
P_100         0.23095238095238094

```

```

root@integration-1:~/biocaddie# scripts/mkeval2.sh jm short combined
map      0.09635238095238095
ndcg     0.30631904761904766
ndcg_cut_5      0.28415714285714283
ndcg_cut_10     0.28073333333333333
ndcg_cut_20     0.24439047619047624
ndcg_cut_100    0.1964238095238095
P_5         0.4
P_10        0.36190476190476195
P_20        0.29761904761904756
P_100       0.16333333333333336

```

```

root@integration-1:~/biocaddie# scripts/mkeval2.sh two short combined
map      0.09635238095238095
ndcg     0.30631904761904766
ndcg_cut_5      0.28415714285714283
ndcg_cut_10     0.28073333333333333
ndcg_cut_20     0.24439047619047624
ndcg_cut_100    0.1964238095238095
P_5         0.4
P_10        0.36190476190476195
P_20        0.29761904761904756
P_100       0.16333333333333336

```

```

root@integration-1:~/biocaddie# scripts/mkeval2.sh rm3 short combined
map 0.13756190476190475
ndcg 0.41485714285714287
ndcg_cut_5 0.38475238095238096
ndcg_cut_10 0.30699047619047615
ndcg_cut_20 0.2962476190476191
ndcg_cut_100 0.3100904761904762
P_5 0.4476190476190475
P_10 0.5238095238095238
P_20 0.3857142857142857
P_100 0.26761904761904765

```

*****Note:**

jm and **two** results are similar. This is due to *no changes in results when holding lambda but sweeping mu values* for two stage baseline.

When lambda=0.0 (IndriRunQuery results)

```

root@integration-1:~/biocaddie/output/two/boolean/combined/short# ls -lrt *lambda=0.0*
-rw-r--r-- 1 root root 652090 Jun  2 16:16 mu=50:lambda=0.0.out
-rw-r--r-- 1 root root 652085 Jun  2 16:16 mu=250:lambda=0.0.out
-rw-r--r-- 1 root root 652083 Jun  2 16:17 mu=1000:lambda=0.0.out
-rw-r--r-- 1 root root 652085 Jun  2 16:17 mu=500:lambda=0.0.out
-rw-r--r-- 1 root root 652080 Jun  2 16:17 mu=2500:lambda=0.0.out
-rw-r--r-- 1 root root 652064 Jun  2 16:17 mu=5000:lambda=0.0.out
-rw-r--r-- 1 root root 652064 Jun  2 16:18 mu=10000:lambda=0.0.out
root@integration-1:~/biocaddie/output/two/boolean/combined/short# head mu=50:lambda=0.0.out
EA1 Q0 733683 1 -2.77259 indri
EA1 Q0 327214 2 -3.46574 indri
EA1 Q0 476233 3 -3.54129 indri
EA1 Q0 149108 4 -3.56953 indri
EA1 Q0 685241 5 -3.61092 indri
EA1 Q0 310477 6 -3.61092 indri
EA1 Q0 121429 7 -3.61092 indri
EA1 Q0 245536 8 -3.61765 indri
EA1 Q0 449101 9 -3.62878 indri
EA1 Q0 155528 10 -3.63099 indri
root@integration-1:~/biocaddie/output/two/boolean/combined/short# head mu=2500:lambda=0.0.out
EA1 Q0 733683 1 -2.77259 indri
EA1 Q0 327214 2 -3.46574 indri
EA1 Q0 476233 3 -3.54129 indri
EA1 Q0 149108 4 -3.56953 indri
EA1 Q0 685241 5 -3.61092 indri
EA1 Q0 310477 6 -3.61092 indri
EA1 Q0 121429 7 -3.61092 indri
EA1 Q0 245536 8 -3.61765 indri
EA1 Q0 449101 9 -3.62878 indri
EA1 Q0 155528 10 -3.63099 indri

```

When lambda=0.7

```
root@integration-1:~/biocaddie/output/two/boolean/combined/short# ls -lrt *lambda=0.7*
-rw-r--r-- 1 root root 652262 Jun  2 16:16 mu=250:lambda=0.7.out
-rw-r--r-- 1 root root 652258 Jun  2 16:16 mu=50:lambda=0.7.out
-rw-r--r-- 1 root root 652271 Jun  2 16:17 mu=500:lambda=0.7.out
-rw-r--r-- 1 root root 652272 Jun  2 16:17 mu=1000:lambda=0.7.out
-rw-r--r-- 1 root root 652272 Jun  2 16:17 mu=2500:lambda=0.7.out
-rw-r--r-- 1 root root 652271 Jun  2 16:18 mu=5000:lambda=0.7.out
-rw-r--r-- 1 root root 652271 Jun  2 16:18 mu=10000:lambda=0.7.out
root@integration-1:~/biocaddie/output/two/boolean/combined/short# head mu=500:lambda=0.7.out
EA1 Q0 733683 1 -3.9758 indri
EA1 Q0 327214 2 -4.66815 indri
EA1 Q0 476233 3 -4.73901 indri
EA1 Q0 149108 4 -4.77178 indri
EA1 Q0 685241 5 -4.81309 indri
EA1 Q0 121429 6 -4.81309 indri
EA1 Q0 310477 7 -4.8131 indri
EA1 Q0 245536 8 -4.81981 indri
EA1 Q0 449101 9 -4.83091 indri
EA1 Q0 155528 10 -4.83312 indri
root@integration-1:~/biocaddie/output/two/boolean/combined/short# head mu=10000:lambda=0.7.out
EA1 Q0 733683 1 -3.9758 indri
EA1 Q0 327214 2 -4.66815 indri
EA1 Q0 476233 3 -4.73901 indri
EA1 Q0 149108 4 -4.77178 indri
EA1 Q0 685241 5 -4.81309 indri
EA1 Q0 121429 6 -4.81309 indri
EA1 Q0 310477 7 -4.8131 indri
EA1 Q0 245536 8 -4.81981 indri
EA1 Q0 449101 9 -4.83091 indri
EA1 Q0 155528 10 -4.83312 indri
```