

HW3 Little Research Report

repository (<https://gitlab.com/krikun/ir-qg>)

0 Structure

```
project/          # little research project
project/docs/     # test document set
project/src/main/scala  # research scala src

project/tasks/    # this reports from .md
```

build with sbt

```
# git clone ssh or https ..
cd project
sbt
...

> run
```

1 Doc set

placed in project/docs/ directory

docId	terms	roughly	tokens
docs/arrays	502	558	1948
docs/mutable-and-immutable-collections	362	396	1048
docs/collection-sets	453	506	1854
docs/collections-trait-traversable	434	487	1633
docs/collections-trait-iterable	278	307	713
docs/maps	447	494	1810
docs/scala-for-java-programmers	962	1055	4983
docs/classes	227	240	430
docs/sets	453	506	1854

Tabel 1. Roughly counted

2.1 Terms

Total terms:

terms : 1709

2.2 Stop words

First 7 by frequency:

the	:	916
of	:	430
a	:	425
is	:	317
to	:	313
in	:	294
and	:	277