

# Assignment 1

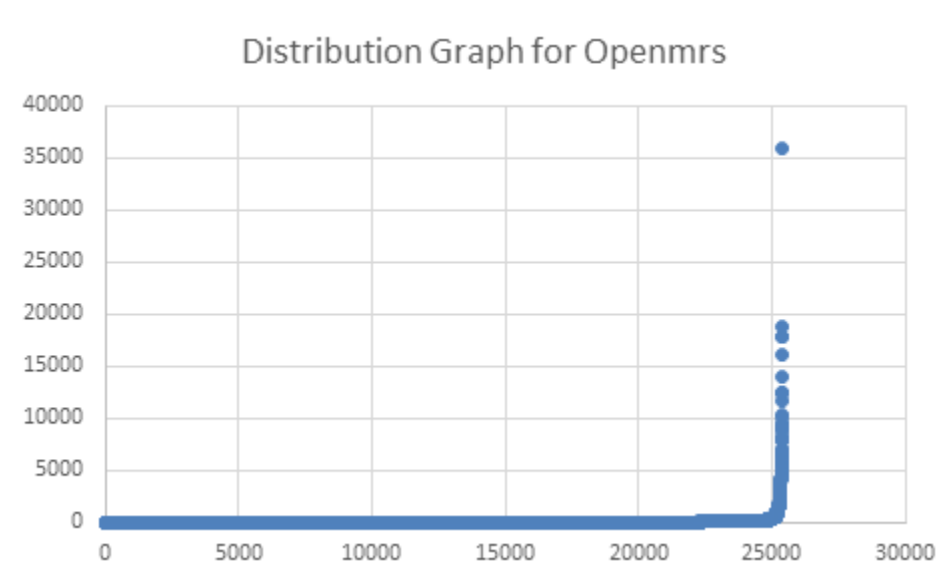
## Team Members

Bennilyn Quek – bxq120430

Kunal Kapoor – Kxk140330

## Results for Openmrs-core-1.11.5

### **Distribution of terms for OPENMRS-**



### **List of the 10 most frequent terms in the corpus (report their tf).**

1. get=35963
2. string=18745
3. concept=17894
4. public=17860
5. set=16047
6. name=13950

7. return=12475
8. null=12419
9. patient=12361
10. exception=11646

### **List of the 10 terms with the highest document frequency (report their df)**

1. form appears in 1531 files
2. copyright appears in 1531 files
3. this appears in 1531 files
4. disclaimer appears in 1531 files
5. public appears in 1531 files
6. license appears in 1531 files
7. source appears in 1531 files
8. code appears in 1531 files
9. http appears in 1530 files
10. open appears in 1530 files

### **List of the 3 documents with the highest number of unique terms.**

\_\_\_\_\_ 3 documents with highest unique terms

\openmrs-core-1.11.5\api\src\test\java\org\openmrs\util\SecurityTest.java has 1654 unique terms.

\openmrs-core-1.11.5\api\src\main\java\org\openmrs\util\OpenmrsConstants.java has 1545 unique terms.

\openmrs-core-1.11.5\api\src\main\java\org\openmrs\util\OpenmrsUtil.java has 1210 unique terms.

### **List of the 3 documents with the lowest number of unique terms.**

\_\_\_\_\_ 3 documents with lowest unique terms

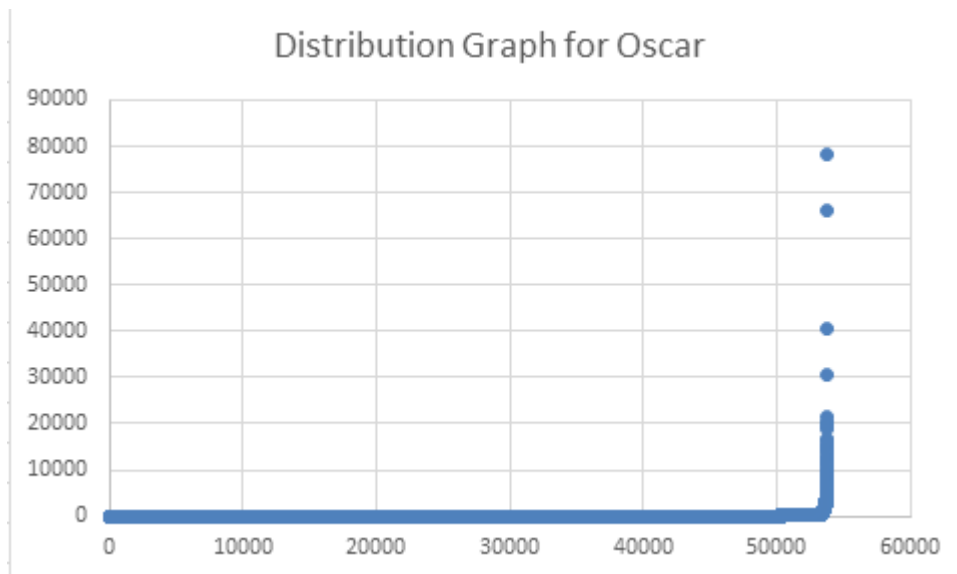
\openmrs-core-1.11.5\api\src\test\java\org\openmrs\util\SecurityTest.java has 1644 unique terms.

\openmrs-core-1.11.5\api\src\main\java\org\openmrs\util\OpenmrsConstants.java has 1495 unique terms.

\openmrs-core-1.11.5\api\src\main\java\org\openmrs\util\OpenmrsUtil.java has 1159 unique terms.

## Results for Oscar 12.1-61

### **Distribution of terms for Oscar**



### **List of the 10 most frequent terms in the corpus (report their tf).**

1. get=78164
2. string=66219
3. set=40462
4. public=30689
5. date=21398
6. id=20334
7. this=20309
8. value=19343
9. name=18730
10. return=16928

### **List of the 10 terms with the highest document frequency (report their df)**

1. public appears in 3052 files
2. this appears in 3045 files

3. general appears in 3044 files
4. version appears in 3041 files
5. copyright appears in 3041 files
6. software appears in 3038 files
7. license appears in 3037 files
8. inc appears in 3037 files
9. terms appears in 3036 files
10. details appears in 3036 files

### **List of the 3 documents with the highest number of unique terms.**

\_\_\_\_\_3 documents with highest unique terms

\oscar12.1-

61\src\main\java\oscar\oscarDemographic\pageUtil\ImportDemographicDataAction4.java has 1866 unique terms.

\oscar12.1-61\src\main\java\oscar\ocan\service\OcanDataProcessor.java has 1593 unique terms.

\oscar12.1-61\src\test\java\org\oscarehr\common\dao\OntarioMDSpec4DataTest.java has 2827 unique terms.

### **List of the 3 documents with the lowest number of unique terms.**

\_\_\_\_\_3 documents with lowest unique terms

\oscar12.1-61\src\main\java\oscar\oscarBilling\ca\bc\MSP\ServiceCodeValidator.java has 11 unique terms.

\oscar12.1-61\src\main\java\oscar\entities\S22.java has 26 unique terms.

\oscar12.1-61\src\main\java\com\indivica\olis\segments\Segment.java has 27 unique terms.