

Assignment 1

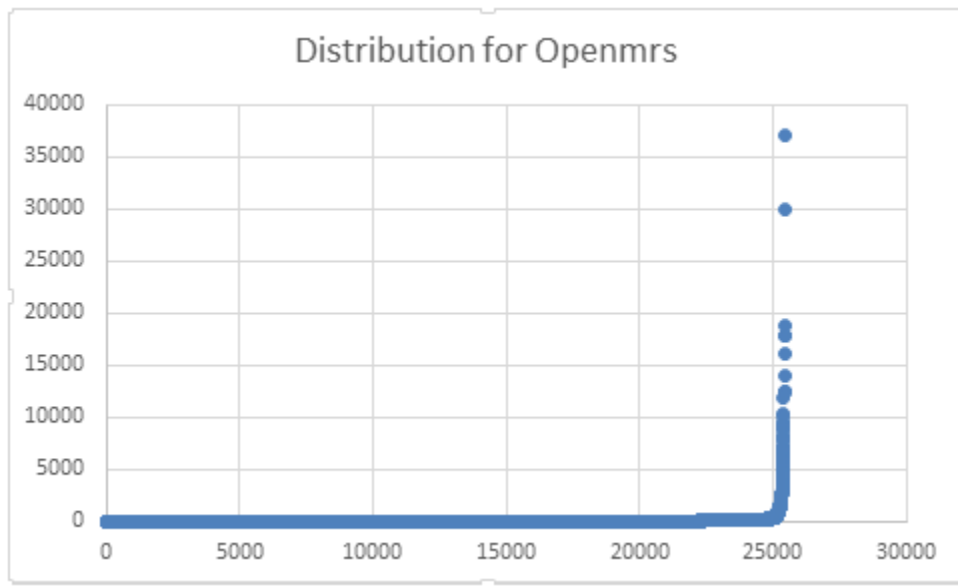
Team Members

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 = 37166
2. the = 30012
3. string = 18745
4. concept = 17894
5. public = 17860
6. set= 16047

7. name = 13950
8. return = 12475
9. null = 12419
10. patient = 12361

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. disclaimer appears in 1531 files:
4. public appears in 1531 files:
5. this appears in 1531 files:
6. and appears in 1531 files:
7. license appears in 1531 files:
8. source appears in 1531 files:
9. not appears in 1531 files:
10. code appears in 1531 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

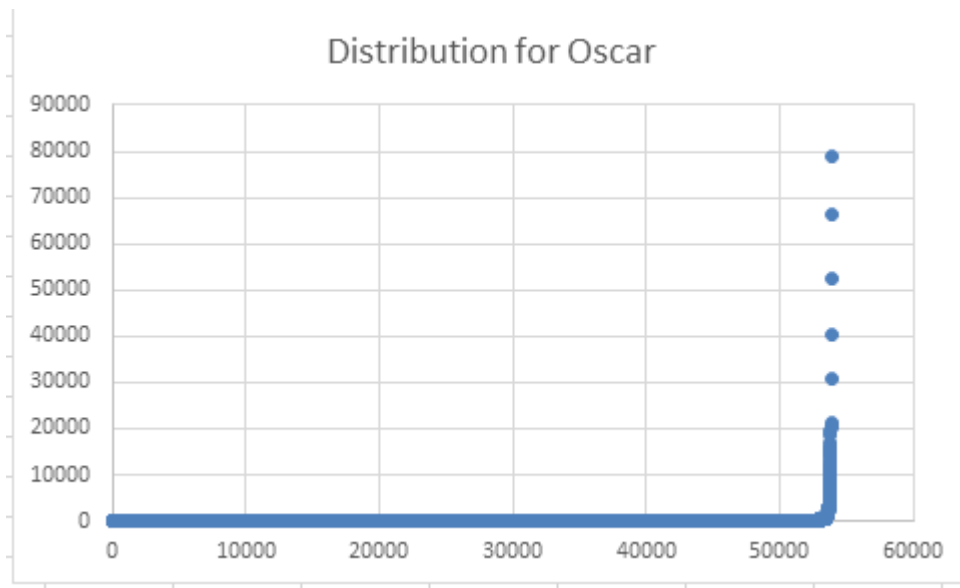
\openmrs-core-1.11.5\api\src\main\java\org\openmrs\logic\op\LogicalOperator.java has 41 unique terms.

\openmrs-core-1.11.5\api\src\main\java\org\openmrs\notification\MessageSender.java has 42 unique terms.\openmrs-core-

1.11.5\api\src\test\java\org\openmrs\annotation\OpenmrsComponent1_8.java has 42 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=78778
2. string=66219
3. the=52351
4. set=40462
5. public=30689
6. date=21398
7. id=20334
8. this=20309
9. value=19343
10. name=18730

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

1. and appears in 3063 files:
2. the appears in 3057 files:
3. public appears in 3052 files:
4. this appears in 3045 files:
5. general appears in 3044 files:
6. version appears in 3041 files:
7. copyright appears in 3041 files:
8. software appears in 3038 files:
9. under appears in 3037 files:
10. license appears in 3037 files:

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

_____ 3 documents with highest unique terms

\oscar\oscar12.1-

61\src\main\java\oscar\oscarDemographic\pageUtil\ImportDemographicDataAction4.java has 1886 unique terms.

\oscar\oscar12.1-61\src\main\java\oscar\ocan\service\OcanDataProcessor.java has 1601 unique terms.

\oscar\oscar12.1-61\src\test\java\org\oscarehr\common\dao\OntarioMDSpec4DataTest.java has 2860 unique terms.

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

_____ 3 documents with lowest unique terms

\oscar\oscar12.1-61\src\main\java\oscar\entities\S21.java has 28 unique terms.

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

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