

Team 7

ASSIGNMENT 1

Rajat Sarin (rxs149230)

Sayantani Goswami (sxx145230)

Siddarth Udayakumar (sxu140530)

OpenMRS

10 most frequent term in corpus

Concept = 13927

String = 13152

Set = 12868

Name = 10754

Null = 9552

Patient = 9042

Order = 8503

Type = 8361

Service = 7341

Person = 7189

10 term with highest document frequency

License appears in 1066 files

Copyright appears in 1066 files

Code appears in 1066 files

Http appears in 1065 files

Org appears in 1065 files

Copy appears in 1065 files

Distributed appears in 1065 files

Obtain appears in 1065 files

File appears in 1065 files

Form appears in 1064 files

3 documents with highest unique terms

ArdenBaseParser.java has 1080 unique terms.

OpenmrsConstants.java has 1170 unique terms.

SecurityTest.java has 1654 unique terms.

3 documents with lowest unique terms

LogicalOperator.java has 41 unique terms.

MessageSender.java has 42 unique terms.

OpenmrsComponent1_8.java has 42 unique terms.

JabRef

10 most frequent term in corpus

String = 10845

Entry = 7633

Set = 5470

File = 5331

List = 3927

Field = 3826

Name = 3739

Action = 3681

Bib = 3639

Null = 3019

10 term with highest document frequency

String appears in 731 files

Jab appears in 632 files

Ref appears in 632 files

Version appears in 591 files

Software appears in 582 files

Copy appears in 582 files

Modify appears in 581 files

Useful appears in 581 files

Option appears in 581 files

General appears in 581 files

3 documents with highest unique terms

HTMLConverter.java has 1745 unique terms.

JabRefPreferences.java has 1228 unique terms.

BasePanel.java has 1284 unique terms.

3 documents with lowest unique terms

LocalizationBundle.java has 4 unique terms.

MessageHandler.java has 4 unique terms.

Defaults.java has 5 unique terms.

Eclipse JDT/Core

10 most frequent term in corpus

Type = 160834
String = 120084
Java = 101539
Null = 74579
Node = 62118
Test = 60097
Name = 59507
Binding = 57425
Method = 56168
Source = 54904

10 term with highest document frequency

String appears in 2950 files
Return appears in 2804 files
Eclipse appears in 2752 files
Org appears in 2745 files
Reserved appears in 2724 files
Rights appears in 2720 files
Copyright appears in 2720 files
Distribution appears in 2716 files
Materials appears in 2716 files
Available appears in 2703 files

3 documents with highest unique terms

GenericTypeTest.java has 6250 unique terms.
A_out.java has 6651 unique terms.
XLargeTest.java has 9381 unique terms.

3 documents with lowest unique terms

A.java has 0 unique terms.
X.java has 0 unique terms.
PX.java has 0 unique terms.