

```

solines-mbp:~ solineledesert$ docker ps
CONTAINER ID          IMAGE
COMMAND              CREATED              STATUS              PORTS
NAMES
e512b1bfb907         icij/datashare:1.3
"/entrypoint.sh -w"  2 days ago         Up 2 days
0.0.0.0:8080->8080/tcp datashare_datashare_1
dc1fbc9dda01         docker.elastic.co/elasticsearch/elasticsearch:
6.3.0                "/usr/local/bin/dock... 2 months ago         Up 2 days
0.0.0.0:9200->9200/tcp, 9300/tcp datashare_elasticsearch_1
fbb46d9f2a50         redis:4.0.1-alpine
"docker-entrypoint.s..." 4 months ago         Up 2 days
0.0.0.0:6379->6379/tcp datashare_redis_1
solines-mbp:~ solineledesert$ docker logs e512b1bfb907
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/datashare/lib/datashare-
cli-1.3-jar-with-dependencies.jar!/org/slf4j/impl/
StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/datashare/lib/datashare-nlp-
corenlp-1.3-jar-with-dependencies.jar!/org/slf4j/impl/
StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/datashare/lib/datashare-nlp-
gate-1.3-jar-with-dependencies.jar!/org/slf4j/impl/
StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/datashare/lib/datashare-nlp-
ixapipe-1.3-jar-with-dependencies.jar!/org/slf4j/impl/
StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/datashare/lib/datashare-nlp-
mitie-1.3-jar-with-dependencies.jar!/org/slf4j/impl/
StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/datashare/lib/datashare-nlp-
opennlp-1.3-jar-with-dependencies.jar!/org/slf4j/impl/
StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/datashare/lib/datashare-
web-1.3-jar-with-dependencies.jar!/org/slf4j/impl/
StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an
explanation.
13:23:41,941 |-INFO in ch.qos.logback.classic.LoggerContext[default] -
Could NOT find resource [logback.groovy]
13:23:41,941 |-INFO in ch.qos.logback.classic.LoggerContext[default] -
Could NOT find resource [logback-test.xml]
13:23:41,942 |-INFO in ch.qos.logback.classic.LoggerContext[default] -
Found resource [logback.xml] at [jar:file:/home/datashare/lib/
datashare-cli-1.3-jar-with-dependencies.jar!/logback.xml]
13:23:41,943 |-WARN in ch.qos.logback.classic.LoggerContext[default] -
Resource [logback.xml] occurs multiple times on the classpath.
13:23:41,943 |-WARN in ch.qos.logback.classic.LoggerContext[default] -
Resource [logback.xml] occurs at [jar:file:/home/datashare/lib/
datashare-nlp-gate-1.3-jar-with-dependencies.jar!/logback.xml]

```

13:23:41,943 |-WARN in ch.qos.logback.classic.LoggerContext[default] - Resource [logback.xml] occurs at [jar:file:/home/datashare/lib/datashare-nlp-corenlp-1.3-jar-with-dependencies.jar!/logback.xml]
13:23:41,943 |-WARN in ch.qos.logback.classic.LoggerContext[default] - Resource [logback.xml] occurs at [jar:file:/home/datashare/lib/datashare-nlp-ixapipe-1.3-jar-with-dependencies.jar!/logback.xml]
13:23:41,943 |-WARN in ch.qos.logback.classic.LoggerContext[default] - Resource [logback.xml] occurs at [jar:file:/home/datashare/lib/datashare-nlp-mitie-1.3-jar-with-dependencies.jar!/logback.xml]
13:23:41,943 |-WARN in ch.qos.logback.classic.LoggerContext[default] - Resource [logback.xml] occurs at [jar:file:/home/datashare/lib/datashare-nlp-opennlp-1.3-jar-with-dependencies.jar!/logback.xml]
13:23:41,943 |-WARN in ch.qos.logback.classic.LoggerContext[default] - Resource [logback.xml] occurs at [jar:file:/home/datashare/lib/datashare-cli-1.3-jar-with-dependencies.jar!/logback.xml]
13:23:41,970 |-INFO in ch.qos.logback.core.joran.spi.ConfigurationWatchList@5d5eef3d - URL [jar:file:/home/datashare/lib/datashare-cli-1.3-jar-with-dependencies.jar!/logback.xml] is not of type file
13:23:42,107 |-INFO in ch.qos.logback.classic.joran.action.ConfigurationAction - debug attribute not set
13:23:42,117 |-INFO in ch.qos.logback.core.joran.action.AppenderAction - About to instantiate appender of type [ch.qos.logback.core.ConsoleAppender]
13:23:42,123 |-INFO in ch.qos.logback.core.joran.action.AppenderAction - Naming appender as [STDOUT]
13:23:42,132 |-INFO in ch.qos.logback.core.joran.action.NestedComplexPropertyIA - Assuming default type [ch.qos.logback.classic.encoder.PatternLayoutEncoder] for [encoder] property
13:23:42,228 |-INFO in ch.qos.logback.core.joran.action.AppenderAction - About to instantiate appender of type [ch.qos.logback.core.rolling.RollingFileAppender]
13:23:42,229 |-INFO in ch.qos.logback.core.joran.action.AppenderAction - Naming appender as [FILE]
13:23:42,249 |-INFO in c.q.l.core.rolling.TimeBasedRollingPolicy@1458849419 - No compression will be used
13:23:42,250 |-INFO in c.q.l.core.rolling.TimeBasedRollingPolicy@1458849419 - Will use the pattern ./logs/datashare.%d{yyyy-MM-dd}.log for the active file
13:23:42,254 |-INFO in c.q.l.core.rolling.DefaultTimeBasedFileNamingAndTriggeringPolicy - The date pattern is 'yyyy-MM-dd' from file name pattern './logs/datashare.%d{yyyy-MM-dd}.log'.
13:23:42,254 |-INFO in c.q.l.core.rolling.DefaultTimeBasedFileNamingAndTriggeringPolicy - Roll-over at midnight.
13:23:42,260 |-INFO in

```
c.q.l.core.rolling.DefaultTimeBasedFileNamingAndTriggeringPolicy -
Setting initial period to Fri Jan 11 13:23:42 GMT 2019
13:23:42,261 |-INFO in
ch.qos.logback.core.joran.action.NestedComplexPropertyIA - Assuming
default type [ch.qos.logback.classic.encoder.PatternLayoutEncoder] for
[encoder] property
13:23:42,263 |-INFO in
ch.qos.logback.core.rolling.RollingFileAppender[FILE] - Active log
file name: ./logs/datashare.log
13:23:42,263 |-INFO in
ch.qos.logback.core.rolling.RollingFileAppender[FILE] - File property
is set to [./logs/datashare.log]
13:23:42,264 |-INFO in
ch.qos.logback.classic.joran.action.RootLoggerAction - Setting level
of ROOT logger to INFO
13:23:42,264 |-INFO in
ch.qos.logback.core.joran.action.AppenderRefAction - Attaching
appender named [STDOUT] to Logger[ROOT]
13:23:42,264 |-INFO in
ch.qos.logback.core.joran.action.AppenderRefAction - Attaching
appender named [FILE] to Logger[ROOT]
13:23:42,264 |-INFO in
ch.qos.logback.classic.joran.action.ConfigurationAction - End of
configuration.
13:23:42,265 |-INFO in
ch.qos.logback.classic.joran.JoranConfigurator@6cc4c815 - Registering
current configuration as safe fallback point
```

```
SLF4J: Actual binding is of type
[ch.qos.logback.classic.util.ContextSelectorStaticBinder]
2019-01-11 13:23:42,400 [main] INFO DatashareCli - Running datashare
web server
2019-01-11 13:23:42,404 [main] INFO DatashareCli - with properties:
{nlpParallelism=1, parserParallelism=1, stages=SCAN,INDEX,NLP,
elasticsearchAddress=http://elasticsearch:9200,
messageBusAddress=redis, parallelism=4, web=true, mode=LOCAL,
nlpPipelines=CORENLP,GATENLP,IXAPIPE,MITIE,OPENNLP, cors=no-cors,
redisAddress=redis://redis:6379, clusterName=datashare,
indexName=local-datashare, dataDir=/home/datashare/data}
2019-01-11 13:23:42,410 [main] INFO PropertiesProvider - reading
properties from jar:file:/home/datashare/lib/datashare-web-1.3-jar-
with-dependencies.jar!/datashare.properties
2019-01-11 13:23:43,540 [main] INFO Indexer - indexer defined with
cfg{indexType='doc', indexJoinField='join', docTypeField='type',
shards=1, replicas=1}
2019-01-11 13:23:44,255 [main] INFO Fluent - Production mode
2019-01-11 13:23:44,280 [main] INFO Fluent - Server started on port
8080
2019-01-11 13:24:13,403 [pool-2-thread-1] INFO NlpApp - running
NlpApp for CORENLP pipeline with 1 thread(s)
```

2019-01-11 13:24:13,403 [Worker: RequestDispatcher: Thread-8] INFO TaskResource - waiting for NlpApp CORENLP to listen...

2019-01-11 13:24:13,410 [pool-2-thread-1] INFO NlpForwarder - waiting for messages on host [redis]

2019-01-11 13:24:13,441 [Worker: RequestDispatcher: Thread-8] INFO TaskResource - ...CORENLP is listening

2019-01-11 13:24:13,442 [pool-2-thread-2] INFO ResumeNlpTask - resuming NLP name finding for user local and [CORENLP]

2019-01-11 13:24:14,089 [pool-2-thread-2] INFO ResumeNlpTask - sent EXTRACT_NLP message for 0 files without [CORENLP] pipeline tags

2019-01-11 13:24:14,102 [pool-2-thread-1] INFO RedisSubscriber - Shutdown called. Unsubscribe done.

2019-01-11 13:24:14,102 [pool-2-thread-1] INFO NlpApp - forwarder exited waiting for consumer(s) to finish

2019-01-11 13:24:14,104 [CORENLP-0] INFO NlpConsumer - exiting main loop

2019-01-11 13:24:14,104 [pool-2-thread-1] INFO NlpApp - exiting run

2019-01-11 13:24:14,192 [redisson-netty-1-1] INFO MasterPubSubConnectionPool - 1 connections initialized for redis/172.18.0.2:6379

2019-01-11 13:24:14,193 [redisson-netty-1-4] INFO MasterConnectionPool - 32 connections initialized for redis/172.18.0.2:6379

2019-01-11 13:24:14,208 [pool-2-thread-3] INFO Scanner - Queuing scan of: "/home/datashare/data".

2019-01-11 13:24:14,209 [pool-2-thread-3] INFO ScannerVisitor - Starting scan of: "/home/datashare/data".

2019-01-11 13:24:14,254 [pool-2-thread-3] INFO ScannerVisitor - Entering directory: "/home/datashare/data".

2019-01-11 13:24:14,301 [pool-2-thread-3] INFO ScannerVisitor - Completed scan of: "/home/datashare/data".

2019-01-11 13:24:15,216 [Worker: RequestDispatcher: Thread-12] INFO ElasticsearchSpewer - spewer defined with cfg{indexType='doc', indexJoinField='join', docTypeField='type', shards=1, replicas=1}

2019-01-11 13:24:15,277 [redisson-netty-4-8] INFO MasterPubSubConnectionPool - 1 connections initialized for redis/172.18.0.2:6379

2019-01-11 13:24:15,279 [redisson-netty-4-4] INFO MasterConnectionPool - 32 connections initialized for redis/172.18.0.2:6379

Jan 11, 2019 1:24:15 PM
org.apache.tika.config.InitializableProblemHandler\$3
handleInitializableProblem
WARNING: Tesseract OCR is installed and will be automatically applied to image files unless you've excluded the TesseractOCRParser from the default parser. Tesseract may dramatically slow down content extraction (TIKA-2359). As of Tika 1.15 (and prior versions), Tesseract is automatically called.
In future versions of Tika, users may need to turn the

```
TesseractOCRParser on via TikaConfig.
2019-01-11 13:24:15,724 [pool-2-thread-4] INFO IndexTask - Processing
up to 4 file(s) in parallel
2019-01-11 13:24:15,732 [pool-8-thread-1] INFO DocumentQueueDrainer -
Draining to consumer until stopped or interrupted.
2019-01-11 13:24:15,735 [pool-8-thread-1] INFO DocumentConsumer -
Sending to thread pool; will queue if full: "/home/datashare/data/
Trump Multinationales Arte.rtf".
2019-01-11 13:24:15,736 [pool-7-thread-1] INFO DocumentConsumer -
Beginning extraction: "/home/datashare/data/Trump Multinationales
Arte.rtf".
2019-01-11 13:24:15,739 [pool-8-thread-1] INFO DocumentQueueDrainer -
Continuous draining stopped.
2019-01-11 13:24:15,739 [pool-2-thread-4] INFO DocumentQueueDrainer -
Shutting down.
2019-01-11 13:24:15,739 [pool-2-thread-4] INFO DocumentQueueDrainer -
Awaiting completion up to 10 SECONDS.
2019-01-11 13:24:15,739 [pool-2-thread-4] INFO DocumentQueueDrainer -
Terminated.
2019-01-11 13:24:15,740 [pool-2-thread-4] INFO IndexTask - drained 1
documents. Waiting for consumer to shutdown
2019-01-11 13:24:15,741 [pool-2-thread-4] INFO DocumentConsumer -
Shutting down.
2019-01-11 13:24:15,742 [pool-2-thread-4] INFO DocumentConsumer -
Awaiting completion up to 30 MINUTES.
2019-01-11 13:24:15,943 [pool-7-thread-1] ERROR DocumentConsumer -
Exception while consuming file: "/home/datashare/data/Trump
Multinationales Arte.rtf".
java.io.IOException:
    at org.apache.tika.parser.ParsingReader.read(ParsingReader.java:
274)
    at java.io.Reader.read(Reader.java:140)
    at org.icij.spewer.Spewer.copy(Spewer.java:104)
    at org.icij.spewer.Spewer.toString(Spewer.java:114)
    at
org.icij.datashare.text.indexing.elasticsearch.ElasticsearchSpewer.get
Map(ElasticsearchSpewer.java:115)
    at
org.icij.datashare.text.indexing.elasticsearch.ElasticsearchSpewer.pre
pareRequest(ElasticsearchSpewer.java:81)
    at
org.icij.datashare.text.indexing.elasticsearch.ElasticsearchSpewer.ind
exDocument(ElasticsearchSpewer.java:134)
    at
org.icij.datashare.text.indexing.elasticsearch.ElasticsearchSpewer.wri
te(ElasticsearchSpewer.java:72)
    at org.icij.extract.extractor.Extractor.extract(Extractor.java:
272)
    at
org.icij.extract.extractor.DocumentConsumer.lambda$accept$0(DocumentCo
```

```
nsumer.java:125)
    at
java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.j
ava:1149)
    at
java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.
java:624)
    at java.lang.Thread.run(Thread.java:748)
Caused by: java.util.ServiceConfigurationError:
java.nio.charset.spi.CharsetProvider: Provider
org.apache.pdfbox.encoding.PDFBoxCharsetProvider not found
    at java.util.ServiceLoader.fail(ServiceLoader.java:239)
    at java.util.ServiceLoader.access$300(ServiceLoader.java:185)
    at
java.util.ServiceLoader$LazyIterator.nextService(ServiceLoader.java:
372)
    at java.util.ServiceLoader$LazyIterator.next(ServiceLoader.java:
404)
    at java.util.ServiceLoader$1.next(ServiceLoader.java:480)
    at java.nio.charset.Charset$1.getNext(Charset.java:352)
    at java.nio.charset.Charset$1.hasNext(Charset.java:365)
    at java.nio.charset.Charset$2.run(Charset.java:410)
    at java.nio.charset.Charset$2.run(Charset.java:407)
    at java.security.AccessController.doPrivileged(Native Method)
    at java.nio.charset.Charset.lookupViaProviders(Charset.java:406)
    at java.nio.charset.Charset.lookup2(Charset.java:477)
    at java.nio.charset.Charset.lookup(Charset.java:464)
    at java.nio.charset.Charset.forName(Charset.java:528)
    at org.apache.tika.utils.CharsetUtils.forName(CharsetUtils.java:
202)
    at
org.apache.tika.parser.rtf.TextExtractor.getCharset(TextExtractor.java
:317)
    at
org.apache.tika.parser.rtf.TextExtractor.<clinit>(TextExtractor.java:
62)
    at org.apache.tika.parser.rtf.RTFParser.parse(RTFParser.java:97)
    at
org.apache.tika.parser.CompositeParser.parse(CompositeParser.java:280)
    at
org.apache.tika.parser.CompositeParser.parse(CompositeParser.java:280)
    at
org.apache.tika.parser.AutoDetectParser.parse(AutoDetectParser.java:
143)
    at
org.apache.tika.parser.ParserDecorator.parse(ParserDecorator.java:188)
    at
org.apache.tika.parser.DigestingParser.parse(DigestingParser.java:84)
    at
org.apache.tika.parser.ParsingReader$ParsingTask.run(ParsingReader.jav
```

a:235)
... 1 common frames omitted
2019-01-11 13:24:15,944 [pool-2-thread-4] INFO DocumentConsumer - Terminated.
2019-01-11 13:24:15,958 [pool-2-thread-4] INFO IndexTask - exiting
2019-01-11 13:24:51,412 [redisson-netty-6-4] INFO MasterPubSubConnectionPool - 1 connections initialized for redis/172.18.0.2:6379
2019-01-11 13:24:51,413 [redisson-netty-6-4] INFO MasterConnectionPool - 32 connections initialized for redis/172.18.0.2:6379
2019-01-11 13:24:51,417 [pool-2-thread-2] INFO Scanner - Queuing scan of: "/home/datashare/data".
2019-01-11 13:24:51,417 [pool-2-thread-2] INFO ScannerVisitor - Starting scan of: "/home/datashare/data".
2019-01-11 13:24:51,419 [pool-2-thread-2] INFO ScannerVisitor - Entering directory: "/home/datashare/data".
2019-01-11 13:24:51,424 [pool-2-thread-2] INFO ScannerVisitor - Completed scan of: "/home/datashare/data".
2019-01-11 13:24:51,909 [Worker: RequestDispatcher: Thread-8] INFO ElasticsearchSpewer - spewer defined with cfg{indexType='doc', indexJoinField='join', docTypeField='type', shards=1, replicas=1}
2019-01-11 13:24:51,967 [redisson-netty-8-6] INFO MasterPubSubConnectionPool - 1 connections initialized for redis/172.18.0.2:6379
2019-01-11 13:24:51,974 [redisson-netty-8-4] INFO MasterConnectionPool - 32 connections initialized for redis/172.18.0.2:6379
2019-01-11 13:24:51,997 [pool-2-thread-1] INFO IndexTask - Processing up to 4 file(s) in parallel
2019-01-11 13:24:52,011 [pool-13-thread-1] INFO DocumentQueueDrainer - Draining to consumer until stopped or interrupted.
2019-01-11 13:24:52,014 [pool-13-thread-1] INFO DocumentConsumer - Sending to thread pool; will queue if full: "/home/datashare/data/Trump Multinationales Arte.rtf".
2019-01-11 13:24:52,015 [pool-12-thread-1] INFO DocumentConsumer - Beginning extraction: "/home/datashare/data/Trump Multinationales Arte.rtf".
2019-01-11 13:24:52,016 [pool-13-thread-1] INFO DocumentQueueDrainer - Continuous draining stopped.
2019-01-11 13:24:52,017 [pool-2-thread-1] INFO DocumentQueueDrainer - Shutting down.
2019-01-11 13:24:52,017 [pool-2-thread-1] INFO DocumentQueueDrainer - Awaiting completion up to 10 SECONDS.
2019-01-11 13:24:52,018 [pool-2-thread-1] INFO DocumentQueueDrainer - Terminated.
2019-01-11 13:24:52,018 [pool-2-thread-1] INFO IndexTask - drained 1 documents. Waiting for consumer to shutdown
2019-01-11 13:24:52,021 [pool-2-thread-1] INFO DocumentConsumer - Shutting down.

```
2019-01-11 13:24:52,021 [pool-2-thread-1] INFO DocumentConsumer -
Awaiting completion up to 30 MINUTES.
2019-01-11 13:24:52,039 [pool-12-thread-1] ERROR DocumentConsumer -
Exception while consuming file: "/home/datashare/data/Trump
Multinationales Arte.rtf".
java.io.IOException:
    at org.apache.tika.parser.ParsingReader.read(ParsingReader.java:
274)
    at java.io.Reader.read(Reader.java:140)
    at org.icij.spewer.Spewer.copy(Spewer.java:104)
    at org.icij.spewer.Spewer.toString(Spewer.java:114)
    at
org.icij.datashare.text.indexing.elasticsearch.ElasticsearchSpewer.get
Map(ElasticsearchSpewer.java:115)
    at
org.icij.datashare.text.indexing.elasticsearch.ElasticsearchSpewer.pre
pareRequest(ElasticsearchSpewer.java:81)
    at
org.icij.datashare.text.indexing.elasticsearch.ElasticsearchSpewer.ind
exDocument(ElasticsearchSpewer.java:134)
    at
org.icij.datashare.text.indexing.elasticsearch.ElasticsearchSpewer.wri
te(ElasticsearchSpewer.java:72)
    at org.icij.extract.extractor.Extractor.extract(Extractor.java:
272)
    at
org.icij.extract.extractor.DocumentConsumer.lambda$accept$0(DocumentCo
nsumer.java:125)
    at
java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.j
ava:1149)
    at
java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.
java:624)
    at java.lang.Thread.run(Thread.java:748)
Caused by: java.lang.NoClassDefFoundError: Could not initialize class
org.apache.tika.parser.rtf.TextExtractor
    at org.apache.tika.parser.rtf.RTFParser.parse(RTFParser.java:97)
    at
org.apache.tika.parser.CompositeParser.parse(CompositeParser.java:280)
    at
org.apache.tika.parser.CompositeParser.parse(CompositeParser.java:280)
    at
org.apache.tika.parser.AutoDetectParser.parse(AutoDetectParser.java:
143)
    at
org.apache.tika.parser.ParserDecorator.parse(ParserDecorator.java:188)
    at
org.apache.tika.parser.DigestingParser.parse(DigestingParser.java:84)
    at
```


org.apache.tika.parser.ParsingReader\$ParsingTask.run(ParsingReader.java:235)

... 1 common frames omitted

2019-01-11 13:24:52,040 [pool-2-thread-1] INFO DocumentConsumer - Terminated.

2019-01-11 13:24:52,057 [pool-2-thread-1] INFO IndexTask - exiting

2019-01-11 15:57:22,032 [Worker: RequestDispatcher: Thread-15] ERROR ResourceLeakDetector - LEAK: HashedWheelTimer.release() was not called before it's garbage-collected. See <http://netty.io/wiki/reference-counted-objects.html> for more information.

Recent access records:

Created at:

io.netty.util.HashedWheelTimer.<init>(HashedWheelTimer.java:272)

io.netty.util.HashedWheelTimer.<init>(HashedWheelTimer.java:216)

io.netty.util.HashedWheelTimer.<init>(HashedWheelTimer.java:195)

org.redisson.connection.MasterSlaveConnectionManager.initTimer(MasterSlaveConnectionManager.java:318)

org.redisson.connection.MasterSlaveConnectionManager.<init>(MasterSlaveConnectionManager.java:161)

org.redisson.connection.SingleConnectionManager.<init>(SingleConnectionManager.java:34)

org.redisson.config.ConfigSupport.createConnectionManager(ConfigSupport.java:192)

org.redisson.Redisson.<init>(Redisson.java:121)

org.redisson.Redisson.create(Redisson.java:158)

org.icij.extract.redis.RedissonClientFactory.create(RedissonClientFactory.java:70)

org.icij.extract.redis.RedisDocumentQueue.<init>(RedisDocumentQueue.java:43)

org.icij.datashare.extract.RedisUserDocumentQueue.<init>(RedisUserDocumentQueue.java:16)

org.icij.datashare.tasks.IndexTask.<init>(IndexTask.java:54)

org.icij.datashare.tasks.IndexTask\$\$FastClassByGuice\$31c77555.newInstance(<generated>)

com.google.inject.internal.DefaultConstructionProxyFactory\$FastClassProxy.newInstance(DefaultConstructionProxyFactory.java:89)

com.google.inject.internal.ConstructorInjector.provision(ConstructorInjector.java:111)

com.google.inject.internal.ConstructorInjector.construct(ConstructorInjector.java:90)

```
com.google.inject.internal.ConstructorBindingImpl$Factory.get(ConstructorBindingImpl.java:268)
com.google.inject.internal.InjectorImpl$2$1.call(InjectorImpl.java:1019)
com.google.inject.internal.InjectorImpl.callInContext(InjectorImpl.java:1085)
    com.google.inject.internal.InjectorImpl$2.get(InjectorImpl.java:1015)
com.google.inject.assistedinject.FactoryProvider2.invoke(FactoryProvider2.java:776)
    com.sun.proxy.$Proxy10.createIndexTask(Unknown Source)
    org.icij.datashare.TaskResource.indexFile(TaskResource.java:69)
    org.icij.datashare.TaskResource.indexDefault(TaskResource.java:62)
    sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
    java.lang.reflect.Method.invoke(Method.java:498)
net.codestory.http.routes.ReflectionRoute.invoke(ReflectionRoute.java:83)
net.codestory.http.routes.ReflectionRoute.lambda$body$0(ReflectionRoute.java:45)
net.codestory.http.annotations.MethodAnnotations.apply(MethodAnnotations.java:48)
net.codestory.http.routes.ReflectionRoute.body(ReflectionRoute.java:40)
net.codestory.http.routes.RouteWithPattern.body(RouteWithPattern.java:56)
    net.codestory.http.routes.Route.apply(Route.java:25)
net.codestory.http.routes.RouteCollection.lambda$createContextToPayload$98caf044$1(RouteCollection.java:577)
net.codestory.http.routes.RouteCollection.lambda$createContextToPayload$51719a14$1(RouteCollection.java:595)
net.codestory.http.routes.RouteCollection.lambda$null$2339cd96$1(RouteCollection.java:593)
```

org.icij.datashare.session.LocalUserFilter.otherUri(LocalUserFilter.java:27)

net.codestory.http.filters.auth.CookieAuthFilter.apply(CookieAuthFilter.java:78)

net.codestory.http.routes.RouteCollection.lambda\$createContextToPayload\$51719a14\$1(RouteCollection.java:593)

net.codestory.http.routes.RouteCollection.apply(RouteCollection.java:567)

net.codestory.http.AbstractWebServer.handleHttp(AbstractWebServer.java:152)

net.codestory.http.internal.SimpleServerWrapper.handle(SimpleServerWrapper.java:71)

org.simpleframework.http.socket.service.RouterContainer.handle(RouterContainer.java:106)

org.simpleframework.http.core.RequestDispatcher.dispatch(RequestDispatcher.java:121)

org.simpleframework.http.core.RequestDispatcher.run(RequestDispatcher.java:103)

java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1149)

java.util.concurrent.ThreadPoolExecutor\$Worker.run(ThreadPoolExecutor.java:624)

java.lang.Thread.run(Thread.java:748)

2019-01-11 15:57:22,035 [Worker: RequestDispatcher: Thread-15] ERROR ResourceLeakDetector - LEAK: HashedWheelTimer.release() was not called before it's garbage-collected. See <http://netty.io/wiki/reference-counted-objects.html> for more information.

Recent access records:

Created at:

io.netty.util.HashedWheelTimer.<init>(HashedWheelTimer.java:272)

io.netty.util.HashedWheelTimer.<init>(HashedWheelTimer.java:216)

io.netty.util.HashedWheelTimer.<init>(HashedWheelTimer.java:195)

org.redisson.connection.MasterSlaveConnectionManager.initTimer(MasterSlaveConnectionManager.java:318)

org.redisson.connection.MasterSlaveConnectionManager.<init>(MasterSlaveConnectionManager.java:161)

```
org.redisson.connection.SingleConnectionManager.<init>(SingleConnectionManager.java:34)

org.redisson.config.ConfigSupport.createConnectionManager(ConfigSupport.java:192)
    org.redisson.Redisson.<init>(Redisson.java:121)
    org.redisson.Redisson.create(Redisson.java:158)

org.icij.extract.redis.RedissonClientFactory.create(RedissonClientFactory.java:70)

org.icij.extract.redis.RedisDocumentQueue.<init>(RedisDocumentQueue.java:43)

org.icij.datashare.extract.RedisUserDocumentQueue.<init>(RedisUserDocumentQueue.java:16)
    org.icij.datashare.tasks.ScanTask.<init>(ScanTask.java:30)
    org.icij.datashare.tasks.ScanTask$$FastClassByGuice$de989a04.newInstance(<generated>)

com.google.inject.internal.DefaultConstructionProxyFactory$FastClassProxy.newInstance(DefaultConstructionProxyFactory.java:89)

com.google.inject.internal.ConstructorInjector.provision(ConstructorInjector.java:111)

com.google.inject.internal.ConstructorInjector.construct(ConstructorInjector.java:90)

com.google.inject.internal.ConstructorBindingImpl$Factory.get(ConstructorBindingImpl.java:268)

com.google.inject.internal.InjectorImpl$2$1.call(InjectorImpl.java:1019)

com.google.inject.internal.InjectorImpl.callInContext(InjectorImpl.java:1085)
    com.google.inject.internal.InjectorImpl$2.get(InjectorImpl.java:1015)

com.google.inject.assistedinject.FactoryProvider2.invoke(FactoryProvider2.java:776)
    com.sun.proxy.$Proxy10.createScanTask(Unknown Source)
    org.icij.datashare.TaskResource.scanFile(TaskResource.java:76)
    org.icij.datashare.TaskResource.indexFile(TaskResource.java:67)
    org.icij.datashare.TaskResource.indexDefault(TaskResource.java:62)
    sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)

sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
```

```
sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccess
orImpl.java:43)
    java.lang.reflect.Method.invoke(Method.java:498)

net.codestory.http.routes.ReflectionRoute.invoke(ReflectionRoute.java:
83)

net.codestory.http.routes.ReflectionRoute.lambda$body$0(ReflectionRout
e.java:45)

net.codestory.http.annotations.MethodAnnotations.apply(MethodAnnotatio
ns.java:48)

net.codestory.http.routes.ReflectionRoute.body(ReflectionRoute.java:
40)

net.codestory.http.routes.RouteWithPattern.body(RouteWithPattern.java:
56)
    net.codestory.http.routes.Route.apply(Route.java:25)

net.codestory.http.routes.RouteCollection.lambda$createContextToPayloa
d$98caf044$1(RouteCollection.java:577)

net.codestory.http.routes.RouteCollection.lambda$createContextToPayloa
d$51719a14$1(RouteCollection.java:595)

net.codestory.http.routes.RouteCollection.lambda$null$2339cd96$1(Route
Collection.java:593)

org.icij.datashare.session.LocalUserFilter.otherUri(LocalUserFilter.ja
va:27)

net.codestory.http.filters.auth.CookieAuthFilter.apply(CookieAuthFilt
er.java:78)

net.codestory.http.routes.RouteCollection.lambda$createContextToPayloa
d$51719a14$1(RouteCollection.java:593)

net.codestory.http.routes.RouteCollection.apply(RouteCollection.java:
567)

net.codestory.http.AbstractWebServer.handleHttp(AbstractWebServer.java
:152)

net.codestory.http.internal.SimpleServerWrapper.handle(SimpleServerWra
pper.java:71)

org.simpleframework.http.socket.service.RouterContainer.handle(RouterC
ontainer.java:106)
```

org.simpleframework.http.core.RequestDispatcher.dispatch(RequestDispatcher.java:121)

org.simpleframework.http.core.RequestDispatcher.run(RequestDispatcher.java:103)

java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1149)

java.util.concurrent.ThreadPoolExecutor\$Worker.run(ThreadPoolExecutor.java:624)

java.lang.Thread.run(Thread.java:748)

2019-01-11 15:57:22,070 [redisson-netty-10-5] INFO MasterPubSubConnectionPool - 1 connections initialized for redis/172.18.0.2:6379
2019-01-11 15:57:22,072 [redisson-netty-10-4] INFO MasterConnectionPool - 32 connections initialized for redis/172.18.0.2:6379
2019-01-11 15:57:22,075 [pool-2-thread-3] INFO Scanner - Queuing scan of: "/home/datashare/data".
2019-01-11 15:57:22,075 [pool-2-thread-3] INFO ScannerVisitor - Starting scan of: "/home/datashare/data".
2019-01-11 15:57:22,083 [pool-2-thread-3] INFO ScannerVisitor - Entering directory: "/home/datashare/data".
2019-01-11 15:57:22,095 [pool-2-thread-3] INFO ScannerVisitor - Completed scan of: "/home/datashare/data".
2019-01-11 15:57:22,958 [Worker: RequestDispatcher: Thread-15] INFO ElasticsearchSpewer - spewer defined with cfg{indexType='doc', indexJoinField='join', docTypeField='type', shards=1, replicas=1}
2019-01-11 15:57:22,997 [redisson-netty-12-2] INFO MasterPubSubConnectionPool - 1 connections initialized for redis/172.18.0.2:6379
2019-01-11 15:57:22,999 [redisson-netty-12-4] INFO MasterConnectionPool - 32 connections initialized for redis/172.18.0.2:6379
2019-01-11 15:57:23,025 [pool-2-thread-4] INFO IndexTask - Processing up to 4 file(s) in parallel
2019-01-11 15:57:23,027 [pool-18-thread-1] INFO DocumentQueueDrainer - Draining to consumer until stopped or interrupted.
2019-01-11 15:57:23,030 [pool-18-thread-1] INFO DocumentConsumer - Sending to thread pool; will queue if full: "/home/datashare/data/Trump Multinationales Arte.rtf".
2019-01-11 15:57:23,031 [pool-17-thread-1] INFO DocumentConsumer - Beginning extraction: "/home/datashare/data/Trump Multinationales Arte.rtf".
2019-01-11 15:57:23,032 [pool-18-thread-1] INFO DocumentQueueDrainer - Continuous draining stopped.
2019-01-11 15:57:23,033 [pool-2-thread-4] INFO DocumentQueueDrainer - Shutting down.

```
2019-01-11 15:57:23,033 [pool-2-thread-4] INFO DocumentQueueDrainer -
Awaiting completion up to 10 SECONDS.
2019-01-11 15:57:23,033 [pool-2-thread-4] INFO DocumentQueueDrainer -
Terminated.
2019-01-11 15:57:23,033 [pool-2-thread-4] INFO IndexTask - drained 1
documents. Waiting for consumer to shutdown
2019-01-11 15:57:23,035 [pool-2-thread-4] INFO DocumentConsumer -
Shutting down.
2019-01-11 15:57:23,036 [pool-2-thread-4] INFO DocumentConsumer -
Awaiting completion up to 30 MINUTES.
2019-01-11 15:57:23,081 [pool-17-thread-1] ERROR DocumentConsumer -
Exception while consuming file: "/home/datashare/data/Trump
Multinationales Arte.rtf".
java.io.IOException:
  at org.apache.tika.parser.ParsingReader.read(ParsingReader.java:
274)
  at java.io.Reader.read(Reader.java:140)
  at org.icij.spewer.Spewer.copy(Spewer.java:104)
  at org.icij.spewer.Spewer.toString(Spewer.java:114)
  at
org.icij.datashare.text.indexing.elasticsearch.ElasticsearchSpewer.get
Map(ElasticsearchSpewer.java:115)
  at
org.icij.datashare.text.indexing.elasticsearch.ElasticsearchSpewer.pre
pareRequest(ElasticsearchSpewer.java:81)
  at
org.icij.datashare.text.indexing.elasticsearch.ElasticsearchSpewer.ind
exDocument(ElasticsearchSpewer.java:134)
  at
org.icij.datashare.text.indexing.elasticsearch.ElasticsearchSpewer.wri
te(ElasticsearchSpewer.java:72)
  at org.icij.extract.extractor.Extractor.extract(Extractor.java:
272)
  at
org.icij.extract.extractor.DocumentConsumer.lambda$accept$0(DocumentCo
nsumer.java:125)
  at
java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.j
ava:1149)
  at
java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.
java:624)
  at java.lang.Thread.run(Thread.java:748)
Caused by: java.lang.NoClassDefFoundError: Could not initialize class
org.apache.tika.parser.rtf.TextExtractor
  at org.apache.tika.parser.rtf.RTFParser.parse(RTFParser.java:97)
  at
org.apache.tika.parser.CompositeParser.parse(CompositeParser.java:280)
  at
org.apache.tika.parser.CompositeParser.parse(CompositeParser.java:280)
```

```
    at
org.apache.tika.parser.AutoDetectParser.parse(AutoDetectParser.java:
143)
    at
org.apache.tika.parser.ParserDecorator.parse(ParserDecorator.java:188)
    at
org.apache.tika.parser.DigestingParser.parse(DigestingParser.java:84)
    at
org.apache.tika.parser.ParsingReader$ParsingTask.run(ParsingReader.jav
a:235)
    ... 1 common frames omitted
2019-01-11 15:57:23,082 [pool-2-thread-4] INFO DocumentConsumer -
Terminated.
2019-01-11 15:57:23,092 [pool-2-thread-4] INFO IndexTask - exiting
2019-01-13 22:23:30,623 [Worker: RequestDispatcher: Thread-26] INFO
TaskResource - waiting for NlpApp CORENLP to listen...
2019-01-13 22:23:30,633 [pool-2-thread-2] INFO NlpApp - running
NlpApp for CORENLP pipeline with 1 thread(s)
2019-01-13 22:23:30,680 [pool-2-thread-2] INFO NlpForwarder - waiting
for messages on host [redis]
2019-01-13 22:23:30,689 [Worker: RequestDispatcher: Thread-26] INFO
TaskResource - ...CORENLP is listening
2019-01-13 22:23:30,697 [pool-2-thread-1] INFO ResumeNlpTask -
resuming NLP name finding for user local and [CORENLP]
2019-01-13 22:23:30,768 [pool-2-thread-1] INFO ResumeNlpTask - sent
EXTRACT_NLP message for 0 files without [CORENLP] pipeline tags
2019-01-13 22:23:30,787 [redisson-netty-14-6] INFO
MasterConnectionPool - 32 connections initialized for redis/
172.18.0.2:6379
2019-01-13 22:23:30,788 [redisson-netty-14-4] INFO
MasterPubSubConnectionPool - 1 connections initialized for redis/
172.18.0.2:6379
2019-01-13 22:23:30,791 [pool-2-thread-3] INFO Scanner - Queuing scan
of: "/home/datashare/data".
2019-01-13 22:23:30,791 [pool-2-thread-3] INFO ScannerVisitor -
Starting scan of: "/home/datashare/data".
2019-01-13 22:23:30,813 [pool-2-thread-3] INFO ScannerVisitor -
Entering directory: "/home/datashare/data".
2019-01-13 22:23:30,836 [pool-2-thread-2] INFO RedisSubscriber -
Shutdown called. Unsubscribe done.
2019-01-13 22:23:30,836 [pool-2-thread-2] INFO NlpApp - forwarder
exited waiting for consumer(s) to finish
2019-01-13 22:23:30,837 [pool-2-thread-3] INFO ScannerVisitor -
Completed scan of: "/home/datashare/data".
2019-01-13 22:23:30,838 [CORENLP-0] INFO NlpConsumer - exiting main
loop
2019-01-13 22:23:30,838 [pool-2-thread-2] INFO NlpApp - exiting run
2019-01-13 22:23:31,452 [Worker: RequestDispatcher: Thread-19] INFO
ElasticsearchSpewer - spewer defined with cfg{indexType='doc',
indexJoinField='join', docTypeField='type', shards=1, replicas=1}
```


2019-01-13 22:23:31,502 [redisson-netty-16-2] INFO
MasterPubSubConnectionPool - 1 connections initialized for redis/
172.18.0.2:6379
2019-01-13 22:23:31,509 [redisson-netty-16-4] INFO
MasterConnectionPool - 32 connections initialized for redis/
172.18.0.2:6379
2019-01-13 22:23:31,541 [pool-2-thread-4] INFO IndexTask - Processing
up to 4 file(s) in parallel
2019-01-13 22:23:31,544 [pool-25-thread-1] INFO DocumentQueueDrainer
- Draining to consumer until stopped or interrupted.
2019-01-13 22:23:31,545 [pool-25-thread-1] INFO DocumentConsumer -
Sending to thread pool; will queue if full: "/home/datashare/data/
Trump Multinationales Arte.rtf".
2019-01-13 22:23:31,547 [pool-25-thread-1] INFO DocumentQueueDrainer
- Continuous draining stopped.
2019-01-13 22:23:31,548 [pool-2-thread-4] INFO DocumentQueueDrainer -
Shutting down.
2019-01-13 22:23:31,548 [pool-2-thread-4] INFO DocumentQueueDrainer -
Awaiting completion up to 10 SECONDS.
2019-01-13 22:23:31,548 [pool-2-thread-4] INFO DocumentQueueDrainer -
Terminated.
2019-01-13 22:23:31,549 [pool-2-thread-4] INFO IndexTask - drained 1
documents. Waiting for consumer to shutdown
2019-01-13 22:23:31,549 [pool-24-thread-1] INFO DocumentConsumer -
Beginning extraction: "/home/datashare/data/Trump Multinationales
Arte.rtf".
2019-01-13 22:23:31,551 [pool-2-thread-4] INFO DocumentConsumer -
Shutting down.
2019-01-13 22:23:31,551 [pool-2-thread-4] INFO DocumentConsumer -
Awaiting completion up to 30 MINUTES.
2019-01-13 22:23:31,571 [pool-24-thread-1] ERROR DocumentConsumer -
Exception while consuming file: "/home/datashare/data/Trump
Multinationales Arte.rtf".
java.io.IOException:
at org.apache.tika.parser.ParsingReader.read(ParsingReader.java:
274)
at java.io.Reader.read(Reader.java:140)
at org.icij.spewer.Spewer.copy(Spewer.java:104)
at org.icij.spewer.Spewer.toString(Spewer.java:114)
at
org.icij.datashare.text.indexing.elasticsearch.ElasticsearchSpewer.get
Map(ElasticsearchSpewer.java:115)
at
org.icij.datashare.text.indexing.elasticsearch.ElasticsearchSpewer.pre
pareRequest(ElasticsearchSpewer.java:81)
at
org.icij.datashare.text.indexing.elasticsearch.ElasticsearchSpewer.ind
exDocument(ElasticsearchSpewer.java:134)
at
org.icij.datashare.text.indexing.elasticsearch.ElasticsearchSpewer.wri

```
te(ElasticsearchSpewer.java:72)
    at org.icij.extract.extractor.Extractor.extract(Extractor.java:
272)
    at
org.icij.extract.extractor.DocumentConsumer.lambda$accept$0(DocumentCo
nsumer.java:125)
    at
java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.j
ava:1149)
    at
java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.
java:624)
    at java.lang.Thread.run(Thread.java:748)
Caused by: java.lang.NoClassDefFoundError: Could not initialize class
org.apache.tika.parser.rtf.TextExtractor
    at org.apache.tika.parser.rtf.RTFParser.parse(RTFParser.java:97)
    at
org.apache.tika.parser.CompositeParser.parse(CompositeParser.java:280)
    at
org.apache.tika.parser.CompositeParser.parse(CompositeParser.java:280)
    at
org.apache.tika.parser.AutoDetectParser.parse(AutoDetectParser.java:
143)
    at
org.apache.tika.parser.ParserDecorator.parse(ParserDecorator.java:188)
    at
org.apache.tika.parser.DigestingParser.parse(DigestingParser.java:84)
    at
org.apache.tika.parser.ParsingReader$ParsingTask.run(ParsingReader.jav
a:235)
```

... 1 common frames omitted

```
2019-01-13 22:23:31,572 [pool-2-thread-4] INFO DocumentConsumer -
Terminated.
2019-01-13 22:23:31,602 [pool-2-thread-4] INFO IndexTask - exiting
2019-01-13 22:27:02,027 [Worker: RequestDispatcher: Thread-12] INFO
TaskResource - waiting for NlpApp CORENLP to listen...
2019-01-13 22:27:02,027 [pool-2-thread-1] INFO NlpApp - running
NlpApp for CORENLP pipeline with 1 thread(s)
2019-01-13 22:27:02,030 [pool-2-thread-1] INFO NlpForwarder - waiting
for messages on host [redis]
2019-01-13 22:27:02,033 [Worker: RequestDispatcher: Thread-12] INFO
TaskResource - ...CORENLP is listening
2019-01-13 22:27:02,037 [pool-2-thread-2] INFO ResumeNlpTask -
resuming NLP name finding for user local and [CORENLP]
2019-01-13 22:27:02,054 [pool-2-thread-2] INFO ResumeNlpTask - sent
EXTRACT_NLP message for 0 files without [CORENLP] pipeline tags
2019-01-13 22:27:02,069 [pool-2-thread-1] INFO RedisSubscriber -
Shutdown called. Unsubscribe done.
2019-01-13 22:27:02,070 [pool-2-thread-1] INFO NlpApp - forwarder
exited waiting for consumer(s) to finish
```

2019-01-13 22:27:02,070 [CORENLP-0] INFO NlpConsumer - exiting main loop
2019-01-13 22:27:02,077 [pool-2-thread-1] INFO NlpApp - exiting run
2019-01-13 22:27:02,082 [redisson-netty-18-3] INFO MasterPubSubConnectionPool - 1 connections initialized for redis/172.18.0.2:6379
2019-01-13 22:27:02,088 [redisson-netty-18-4] INFO MasterConnectionPool - 32 connections initialized for redis/172.18.0.2:6379
2019-01-13 22:27:02,091 [pool-2-thread-3] INFO Scanner - Queuing scan of: "/home/datashare/data".
2019-01-13 22:27:02,091 [pool-2-thread-3] INFO ScannerVisitor - Starting scan of: "/home/datashare/data".
2019-01-13 22:27:02,093 [pool-2-thread-3] INFO ScannerVisitor - Entering directory: "/home/datashare/data".
2019-01-13 22:27:02,097 [pool-2-thread-3] INFO ScannerVisitor - Completed scan of: "/home/datashare/data".
2019-01-13 22:27:02,798 [Worker: RequestDispatcher: Thread-19] INFO ElasticsearchSpewer - spewer defined with cfg{indexType='doc', indexJoinField='join', docTypeField='type', shards=1, replicas=1}
2019-01-13 22:27:02,885 [redisson-netty-20-2] INFO MasterPubSubConnectionPool - 1 connections initialized for redis/172.18.0.2:6379
2019-01-13 22:27:02,886 [redisson-netty-20-8] INFO MasterConnectionPool - 32 connections initialized for redis/172.18.0.2:6379
2019-01-13 22:27:02,917 [pool-2-thread-4] INFO IndexTask - Processing up to 4 file(s) in parallel
2019-01-13 22:27:02,923 [pool-31-thread-1] INFO DocumentQueueDrainer - Draining to consumer until stopped or interrupted.
2019-01-13 22:27:02,925 [pool-31-thread-1] INFO DocumentConsumer - Sending to thread pool; will queue if full: "/home/datashare/data/Trump Multinationales Arte.rtf".
2019-01-13 22:27:02,929 [pool-30-thread-1] INFO DocumentConsumer - Beginning extraction: "/home/datashare/data/Trump Multinationales Arte.rtf".
2019-01-13 22:27:02,930 [pool-31-thread-1] INFO DocumentQueueDrainer - Continuous draining stopped.
2019-01-13 22:27:02,932 [pool-2-thread-4] INFO DocumentQueueDrainer - Shutting down.
2019-01-13 22:27:02,933 [pool-2-thread-4] INFO DocumentQueueDrainer - Awaiting completion up to 10 SECONDS.
2019-01-13 22:27:02,933 [pool-2-thread-4] INFO DocumentQueueDrainer - Terminated.
2019-01-13 22:27:02,934 [pool-2-thread-4] INFO IndexTask - drained 1 documents. Waiting for consumer to shutdown
2019-01-13 22:27:02,937 [pool-2-thread-4] INFO DocumentConsumer - Shutting down.
2019-01-13 22:27:02,937 [pool-2-thread-4] INFO DocumentConsumer - Awaiting completion up to 30 MINUTES.

```
2019-01-13 22:27:02,962 [pool-30-thread-1] ERROR DocumentConsumer -
Exception while consuming file: "/home/datashare/data/Trump
Multinationales Arte.rtf".
java.io.IOException:
    at org.apache.tika.parser.ParsingReader.read(ParsingReader.java:
274)
    at java.io.Reader.read(Reader.java:140)
    at org.icij.spewer.Spewer.copy(Spewer.java:104)
    at org.icij.spewer.Spewer.toString(Spewer.java:114)
    at
org.icij.datashare.text.indexing.elasticsearch.ElasticsearchSpewer.get
Map(ElasticsearchSpewer.java:115)
    at
org.icij.datashare.text.indexing.elasticsearch.ElasticsearchSpewer.pre
pareRequest(ElasticsearchSpewer.java:81)
    at
org.icij.datashare.text.indexing.elasticsearch.ElasticsearchSpewer.ind
exDocument(ElasticsearchSpewer.java:134)
    at
org.icij.datashare.text.indexing.elasticsearch.ElasticsearchSpewer.wri
te(ElasticsearchSpewer.java:72)
    at org.icij.extract.extractor.Extractor.extract(Extractor.java:
272)
    at
org.icij.extract.extractor.DocumentConsumer.lambda$accept$0(DocumentCo
nsumer.java:125)
    at
java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.j
ava:1149)
    at
java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.
java:624)
    at java.lang.Thread.run(Thread.java:748)
Caused by: java.lang.NoClassDefFoundError: Could not initialize class
org.apache.tika.parser.rtf.TextExtractor
    at org.apache.tika.parser.rtf.RTFParser.parse(RTFParser.java:97)
    at
org.apache.tika.parser.CompositeParser.parse(CompositeParser.java:280)
    at
org.apache.tika.parser.CompositeParser.parse(CompositeParser.java:280)
    at
org.apache.tika.parser.AutoDetectParser.parse(AutoDetectParser.java:
143)
    at
org.apache.tika.parser.ParserDecorator.parse(ParserDecorator.java:188)
    at
org.apache.tika.parser.DigestingParser.parse(DigestingParser.java:84)
    at
org.apache.tika.parser.ParsingReader$ParsingTask.run(ParsingReader.jav
a:235)
```

... 1 common frames omitted
2019-01-13 22:27:02,965 [pool-2-thread-4] INFO DocumentConsumer - Terminated.
2019-01-13 22:27:02,991 [pool-2-thread-4] INFO IndexTask - exiting
2019-01-13 23:04:01,592 [Worker: RequestDispatcher: Thread-8] INFO TaskResource - waiting for NlpApp CORENLP to listen...
2019-01-13 23:04:01,592 [pool-2-thread-2] INFO NlpApp - running NlpApp for CORENLP pipeline with 1 thread(s)
2019-01-13 23:04:01,595 [pool-2-thread-2] INFO NlpForwarder - waiting for messages on host [redis]
2019-01-13 23:04:01,598 [Worker: RequestDispatcher: Thread-8] INFO TaskResource - ...CORENLP is listening
2019-01-13 23:04:01,599 [pool-2-thread-1] INFO ResumeNlpTask - resuming NLP name finding for user local and [CORENLP]
2019-01-13 23:04:01,625 [pool-2-thread-1] INFO ResumeNlpTask - sent EXTRACT_NLP message for 0 files without [CORENLP] pipeline tags
2019-01-13 23:04:01,633 [pool-2-thread-2] INFO RedisSubscriber - Shutdown called. Unsubscribe done.
2019-01-13 23:04:01,633 [pool-2-thread-2] INFO NlpApp - forwarder exited waiting for consumer(s) to finish
2019-01-13 23:04:01,637 [CORENLP-0] INFO NlpConsumer - exiting main loop
2019-01-13 23:04:01,637 [pool-2-thread-2] INFO NlpApp - exiting run
2019-01-13 23:04:01,662 [redisson-netty-22-1] INFO MasterPubSubConnectionPool - 1 connections initialized for redis/172.18.0.2:6379
2019-01-13 23:04:01,663 [redisson-netty-22-4] INFO MasterConnectionPool - 32 connections initialized for redis/172.18.0.2:6379
2019-01-13 23:04:01,666 [pool-2-thread-3] INFO Scanner - Queuing scan of: "/home/datashare/data".
2019-01-13 23:04:01,667 [pool-2-thread-3] INFO ScannerVisitor - Starting scan of: "/home/datashare/data".
2019-01-13 23:04:01,676 [pool-2-thread-3] INFO ScannerVisitor - Entering directory: "/home/datashare/data".
2019-01-13 23:04:01,702 [pool-2-thread-3] INFO ScannerVisitor - Completed scan of: "/home/datashare/data".
2019-01-13 23:04:02,126 [Worker: RequestDispatcher: Thread-15] INFO ElasticsearchSpewer - spewer defined with cfg{indexType='doc', indexJoinField='join', docTypeField='type', shards=1, replicas=1}
2019-01-13 23:04:02,190 [redisson-netty-24-8] INFO MasterPubSubConnectionPool - 1 connections initialized for redis/172.18.0.2:6379
2019-01-13 23:04:02,196 [redisson-netty-24-4] INFO MasterConnectionPool - 32 connections initialized for redis/172.18.0.2:6379
2019-01-13 23:04:02,216 [pool-2-thread-4] INFO IndexTask - Processing up to 4 file(s) in parallel
2019-01-13 23:04:02,217 [pool-37-thread-1] INFO DocumentQueueDrainer - Draining to consumer until stopped or interrupted.

```
2019-01-13 23:04:02,221 [pool-37-thread-1] INFO DocumentConsumer -
Sending to thread pool; will queue if full: "/home/datashare/data/
Trump Multinationales Arte.rtf".
2019-01-13 23:04:02,223 [pool-37-thread-1] INFO DocumentConsumer -
Sending to thread pool; will queue if full: "/home/datashare/data/
sample (1).pdf".
2019-01-13 23:04:02,224 [pool-36-thread-2] INFO DocumentConsumer -
Beginning extraction: "/home/datashare/data/sample (1).pdf".
2019-01-13 23:04:02,224 [pool-36-thread-1] INFO DocumentConsumer -
Beginning extraction: "/home/datashare/data/Trump Multinationales
Arte.rtf".
2019-01-13 23:04:02,226 [pool-37-thread-1] INFO DocumentConsumer -
Sending to thread pool; will queue if full: "/home/datashare/data/
SAMPLE-field-guide-to-fake-news.pdf".
2019-01-13 23:04:02,228 [pool-37-thread-1] INFO DocumentConsumer -
Sending to thread pool; will queue if full: "/home/datashare/data/pdf-
sample.pdf".
2019-01-13 23:04:02,230 [pool-36-thread-4] INFO DocumentConsumer -
Beginning extraction: "/home/datashare/data/pdf-sample.pdf".
2019-01-13 23:04:02,230 [pool-37-thread-1] INFO DocumentConsumer -
Sending to thread pool; will queue if full: "/home/datashare/data/
sample.pdf".
2019-01-13 23:04:02,230 [pool-36-thread-3] INFO DocumentConsumer -
Beginning extraction: "/home/datashare/data/SAMPLE-field-guide-to-
fake-news.pdf".
2019-01-13 23:04:02,299 [pool-36-thread-1] ERROR DocumentConsumer -
Exception while consuming file: "/home/datashare/data/Trump
Multinationales Arte.rtf".
java.io.IOException:
    at org.apache.tika.parser.ParsingReader.read(ParsingReader.java:
274)
    at java.io.Reader.read(Reader.java:140)
    at org.icij.spewer.Spewer.copy(Spewer.java:104)
    at org.icij.spewer.Spewer.toString(Spewer.java:114)
    at
org.icij.datashare.text.indexing.elasticsearch.ElasticsearchSpewer.get
Map(ElasticsearchSpewer.java:115)
    at
org.icij.datashare.text.indexing.elasticsearch.ElasticsearchSpewer.pre
pareRequest(ElasticsearchSpewer.java:81)
    at
org.icij.datashare.text.indexing.elasticsearch.ElasticsearchSpewer.ind
exDocument(ElasticsearchSpewer.java:134)
    at
org.icij.datashare.text.indexing.elasticsearch.ElasticsearchSpewer.wri
te(ElasticsearchSpewer.java:72)
    at org.icij.extract.extractor.Extractor.extract(Extractor.java:
272)
    at
org.icij.extract.extractor.DocumentConsumer.lambda$accept$0(DocumentCo
```

```
nsumer.java:125)
    at
java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.j
ava:1149)
    at
java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.
java:624)
    at java.lang.Thread.run(Thread.java:748)
Caused by: java.lang.NoClassDefFoundError: Could not initialize class
org.apache.tika.parser.rtf.TextExtractor
    at org.apache.tika.parser.rtf.RTFParser.parse(RTFParser.java:97)
    at
org.apache.tika.parser.CompositeParser.parse(CompositeParser.java:280)
    at
org.apache.tika.parser.CompositeParser.parse(CompositeParser.java:280)
    at
org.apache.tika.parser.AutoDetectParser.parse(AutoDetectParser.java:
143)
    at
org.apache.tika.parser.ParserDecorator.parse(ParserDecorator.java:188)
    at
org.apache.tika.parser.DigestingParser.parse(DigestingParser.java:84)
    at
org.apache.tika.parser.ParsingReader$ParsingTask.run(ParsingReader.jav
a:235)
    ... 1 common frames omitted
2019-01-13 23:04:02,305 [pool-36-thread-1] INFO DocumentConsumer -
Beginning extraction: "/home/datashare/data/sample.pdf".
2019-01-13 23:04:02,306 [pool-37-thread-1] INFO DocumentConsumer -
Sending to thread pool; will queue if full: "/home/datashare/data/
pdfsample.pdf".
2019-01-13 23:04:02,681 [Apache Tika: sample (1).pdf] INFO
PDDeviceRGB - To get higher rendering speed on JDK8 or later,
2019-01-13 23:04:02,682 [Apache Tika: sample (1).pdf] INFO
PDDeviceRGB - use the option -
Dsun.java2d.cmm=sun.java2d.cmm.kcms.KcmsServiceProvider
2019-01-13 23:04:02,682 [Apache Tika: sample (1).pdf] INFO
PDDeviceRGB - or call System.setProperty("sun.java2d.cmm",
"sun.java2d.cmm.kcms.KcmsServiceProvider")
2019-01-13 23:04:03,766 [Apache Tika: pdf-sample.pdf] WARN
FileSystemFontProvider - New fonts found, font cache will be re-built
2019-01-13 23:04:03,766 [Apache Tika: pdf-sample.pdf] WARN
FileSystemFontProvider - Building on-disk font cache, this may take a
while
2019-01-13 23:04:04,028 [Apache Tika: pdf-sample.pdf] WARN
FileSystemFontProvider - Finished building on-disk font cache, found
91 fonts
2019-01-13 23:04:04,044 [Apache Tika: pdf-sample.pdf] WARN
PDTrueTypeFont - Using fallback font 'LiberationSerif' for
'TimesNewRomanPSMT'
```

2019-01-13 23:04:04,056 [Apache Tika: sample (1).pdf] WARN
PDType1Font - Using fallback font LiberationSans for base font Symbol
2019-01-13 23:04:04,057 [Apache Tika: sample (1).pdf] WARN
PDType1Font - Using fallback font LiberationSans for base font
ZapfDingbats
2019-01-13 23:04:04,315 [Apache Tika: pdf-sample.pdf] WARN
PDType1Font - Using fallback font LiberationSans for Symbol
2019-01-13 23:04:04,338 [Apache Tika: pdf-sample.pdf] WARN
PDTrueTypeFont - Using fallback font 'LiberationSerif-BoldItalic' for
'TimesNewRomanPS-BoldItalicMT'
2019-01-13 23:04:04,877 [Apache Tika: sample (1).pdf] WARN Type1Lexer
- NULL byte in font, skipped
2019-01-13 23:04:04,878 [Apache Tika: sample (1).pdf] WARN Type1Lexer
- NULL byte in font, skipped
2019-01-13 23:04:04,878 [Apache Tika: sample (1).pdf] WARN Type1Lexer
- NULL byte in font, skipped
2019-01-13 23:04:04,879 [Apache Tika: sample (1).pdf] WARN Type1Lexer
- NULL byte in font, skipped
2019-01-13 23:04:04,880 [Apache Tika: sample (1).pdf] WARN Type1Lexer
- NULL byte in font, skipped
2019-01-13 23:04:04,880 [Apache Tika: sample (1).pdf] WARN Type1Lexer
- NULL byte in font, skipped
2019-01-13 23:04:04,881 [Apache Tika: sample (1).pdf] WARN Type1Lexer
- NULL byte in font, skipped
2019-01-13 23:04:04,883 [Apache Tika: sample (1).pdf] WARN Type1Lexer
- NULL byte in font, skipped
2019-01-13 23:04:04,883 [Apache Tika: sample (1).pdf] WARN Type1Lexer
- NULL byte in font, skipped
2019-01-13 23:04:05,494 [Apache Tika: sample (1).pdf] WARN Type1Lexer
- NULL byte in font, skipped
2019-01-13 23:04:05,495 [Apache Tika: sample (1).pdf] WARN Type1Lexer
- NULL byte in font, skipped
2019-01-13 23:04:05,496 [Apache Tika: sample (1).pdf] WARN Type1Lexer
- NULL byte in font, skipped
2019-01-13 23:04:05,496 [Apache Tika: sample (1).pdf] WARN Type1Lexer
- NULL byte in font, skipped
2019-01-13 23:04:05,497 [Apache Tika: sample (1).pdf] WARN Type1Lexer
- NULL byte in font, skipped
2019-01-13 23:04:05,497 [Apache Tika: sample (1).pdf] WARN Type1Lexer
- NULL byte in font, skipped
2019-01-13 23:04:05,499 [Apache Tika: sample (1).pdf] WARN Type1Lexer
- NULL byte in font, skipped
2019-01-13 23:04:05,499 [Apache Tika: sample (1).pdf] WARN Type1Lexer
- NULL byte in font, skipped
2019-01-13 23:04:05,500 [Apache Tika: sample (1).pdf] WARN Type1Lexer
- NULL byte in font, skipped
2019-01-13 23:04:05,500 [Apache Tika: sample (1).pdf] WARN Type1Lexer
- NULL byte in font, skipped
2019-01-13 23:04:05,502 [Apache Tika: sample (1).pdf] WARN Type1Lexer
- NULL byte in font, skipped

2019-01-13 23:04:05,503 [Apache Tika: sample (1).pdf] WARN Type1Lexer
- NULL byte in font, skipped
2019-01-13 23:04:05,506 [Apache Tika: sample (1).pdf] WARN Type1Lexer
- NULL byte in font, skipped
2019-01-13 23:04:05,506 [Apache Tika: sample (1).pdf] WARN Type1Lexer
- NULL byte in font, skipped
2019-01-13 23:04:05,507 [Apache Tika: sample (1).pdf] WARN Type1Lexer
- NULL byte in font, skipped
2019-01-13 23:04:05,507 [Apache Tika: sample (1).pdf] WARN Type1Lexer
- NULL byte in font, skipped
2019-01-13 23:04:05,508 [Apache Tika: sample (1).pdf] WARN Type1Lexer
- NULL byte in font, skipped
2019-01-13 23:04:05,508 [Apache Tika: sample (1).pdf] WARN Type1Lexer
- NULL byte in font, skipped
2019-01-13 23:04:05,509 [Apache Tika: sample (1).pdf] WARN Type1Lexer
- NULL byte in font, skipped
2019-01-13 23:04:05,510 [Apache Tika: sample (1).pdf] WARN Type1Lexer
- NULL byte in font, skipped
2019-01-13 23:04:05,511 [Apache Tika: sample (1).pdf] WARN
PDSimpleFont - No Unicode mapping for negationslash (54) in font
YLJAAA+CMSY10
2019-01-13 23:04:05,511 [Apache Tika: sample (1).pdf] WARN
PDSimpleFont - No Unicode mapping for star (63) in font FWAAAA+CMMI12
2019-01-13 23:04:05,925 [pool-36-thread-2] INFO ElasticsearchSpewer -
Document 4c76...4ddc added to elasticsearch in 323ms: /home/datashare/
data/sample (1).pdf
2019-01-13 23:04:05,926 [pool-36-thread-4] INFO ElasticsearchSpewer -
Document 60bd...0d5b added to elasticsearch in 1425ms: /home/
datashare/data/pdf-sample.pdf
2019-01-13 23:04:05,927 [pool-36-thread-1] INFO ElasticsearchSpewer -
Document 8dec...9e51 added to elasticsearch in 1424ms: /home/
datashare/data/sample.pdf
2019-01-13 23:04:05,931 [pool-36-thread-2] INFO DocumentConsumer -
Beginning extraction: "/home/datashare/data/pdfs/sample.pdf".
2019-01-13 23:04:05,932 [pool-37-thread-1] INFO DocumentQueueDrainer
- Continuous draining stopped.
2019-01-13 23:04:05,935 [pool-2-thread-4] INFO DocumentQueueDrainer -
Shutting down.
2019-01-13 23:04:05,935 [pool-2-thread-4] INFO DocumentQueueDrainer -
Awaiting completion up to 10 SECONDS.
2019-01-13 23:04:05,936 [pool-2-thread-4] INFO DocumentQueueDrainer -
Terminated.
2019-01-13 23:04:05,937 [pool-2-thread-4] INFO IndexTask - drained 6
documents. Waiting for consumer to shutdown
2019-01-13 23:04:05,939 [pool-2-thread-4] INFO DocumentConsumer -
Shutting down.
2019-01-13 23:04:05,939 [pool-2-thread-4] INFO DocumentConsumer -
Awaiting completion up to 30 MINUTES.
2019-01-13 23:04:10,120 [Apache Tika: SAMPLE-field-guide-to-fake-
news.pdf] WARN PDSimpleFont - No Unicode mapping for f_f_i (151) in

font CWBBIH+HKGrotesk-Regular
2019-01-13 23:04:20,193 [Apache Tika: SAMPLE-field-guide-to-fake-news.pdf] WARN PDSimpleFont - No Unicode mapping for f_f_i (151) in font CWBBIH+HKGrotesk-Regular
2019-01-13 23:04:21,025 [Apache Tika: SAMPLE-field-guide-to-fake-news.pdf] WARN PDSimpleFont - No Unicode mapping for .notdef (31) in font JFTRCT+MyriadPro-Regular
2019-01-13 23:04:24,081 [Apache Tika: SAMPLE-field-guide-to-fake-news.pdf] WARN PDSimpleFont - No Unicode mapping for .notdef (31) in font HPZFAX+MyriadPro-Regular
2019-01-13 23:04:29,643 [pool-36-thread-2] INFO ElasticsearchSpewer - Document 5d5f...a7f5 added to elasticsearch in 377ms: /home/datashare/data/pdfs/sample.pdf
2019-01-13 23:04:32,281 [Apache Tika: SAMPLE-field-guide-to-fake-news.pdf] WARN PDSimpleFont - No Unicode mapping for f_f_i (152) in font NQHPGL+HKGrotesk-Bold
2019-01-13 23:04:44,521 [Apache Tika: SAMPLE-field-guide-to-fake-news.pdf] WARN PDSimpleFont - No Unicode mapping for f_f_i (152) in font NQHPGL+HKGrotesk-Bold
2019-01-13 23:05:06,377 [pool-36-thread-3] INFO ElasticsearchSpewer - Document fdca...eeda added to elasticsearch in 198ms: /home/datashare/data/SAMPLE-field-guide-to-fake-news.pdf
2019-01-13 23:05:06,378 [pool-2-thread-4] INFO DocumentConsumer - Terminated.
2019-01-13 23:05:06,402 [pool-2-thread-4] INFO IndexTask - exiting
2019-01-13 23:14:51,263 [Worker: RequestDispatcher: Thread-28] INFO TaskResource - waiting for NlpApp CORENLP to listen...
2019-01-13 23:14:51,263 [pool-2-thread-1] INFO NlpApp - running NlpApp for CORENLP pipeline with 1 thread(s)
2019-01-13 23:14:51,267 [pool-2-thread-1] INFO NlpForwarder - waiting for messages on host [redis]
2019-01-13 23:14:51,270 [Worker: RequestDispatcher: Thread-28] INFO TaskResource - ...CORENLP is listening
2019-01-13 23:14:51,270 [pool-2-thread-2] INFO ResumeNlpTask - resuming NLP name finding for user local and [CORENLP]
2019-01-13 23:14:51,554 [pool-2-thread-2] INFO ResumeNlpTask - sent EXTRACT_NLP message for 5 files without [CORENLP] pipeline tags
2019-01-13 23:14:51,528 [pool-2-thread-1] INFO RedisSubscriber - Shutdown called. Unsubscribe done.
2019-01-13 23:14:51,529 [pool-2-thread-1] INFO NlpApp - forwarder exited waiting for consumer(s) to finish
2019-01-13 23:14:51,554 [CORENLP-0] INFO NlpConsumer - extracting CORENLP entities for document
8decc8571946d4cd70a024949e033a2a2a54377fe9f1c1b944c20f9ee11a9e51
2019-01-13 23:14:51,568 [CORENLP-0] INFO CorenlpPipeline - initializing CORENLP ENGLISH [TOKEN, SENTENCE, POS, LEMMA, NER]
2019-01-13 23:14:51,572 [CORENLP-0] INFO CoreNlpNerModels - sync models : true
2019-01-13 23:14:52,304 [CORENLP-0] INFO CoreNlpNerModels - downloading models for language ENGLISH under dist/models/corenlp/

```
3-8-0/en
2019-01-13 23:14:57,358 [CORENLP-0] WARN NlpConsumer - error in
consumer main loop
java.lang.RuntimeException: Couldn't create parent directories for /
home/datashare/dist/models/corenlp/3-8-0/en/stanford-corenlp-3.8.0-
models-english.jar
    at
com.amazonaws.services.s3.transfer.TransferManager.downloadDirectory(T
ransferManager.java:1275)
    at
com.amazonaws.services.s3.transfer.TransferManager.downloadDirectory(T
ransferManager.java:1157)
    at
com.amazonaws.services.s3.transfer.TransferManager.downloadDirectory(T
ransferManager.java:1148)
    at org.icij.datashare.io.RemoteFiles.download(RemoteFiles.java:65)
    at
org.icij.datashare.text.nlp.AbstractModels.downloadIfNecessary(Abstrac
tModels.java:114)
    at
org.icij.datashare.text.nlp.AbstractModels.load(AbstractModels.java:
66)
    at
org.icij.datashare.text.nlp.AbstractModels.get(AbstractModels.java:55)
    at
org.icij.datashare.text.nlp.corenlp.CorenlpPipeline.initializeNerAnnot
ator(CorenlpPipeline.java:180)
    at
org.icij.datashare.text.nlp.corenlp.CorenlpPipeline.initialize(Corenlp
Pipeline.java:77)
    at
org.icij.datashare.text.nlp.NlpConsumer.findNamedEntities(NlpConsumer.
java:66)
    at org.icij.datashare.text.nlp.NlpConsumer.run(NlpConsumer.java:
40)
    at
java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.j
ava:1149)
    at
java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.
java:624)
    at java.lang.Thread.run(Thread.java:748)
2019-01-13 23:14:57,366 [CORENLP-0] INFO NlpConsumer - extracting
CORENLP entities for document
60bdd13ea4827b8de375c79dc3ff847f83b55bd73b6461523fdf8f843b5a0d5b
2019-01-13 23:14:57,366 [CORENLP-0] INFO CorenlpPipeline -
initializing CORENLP ENGLISH [TOKEN, SENTENCE, POS, LEMMA, NER]
2019-01-13 23:14:57,369 [CORENLP-0] INFO CoreNlpNerModels -
downloading models for language ENGLISH under dist/models/corenlp/
3-8-0/en
```

```
2019-01-13 23:14:57,851 [CORENLP-0] WARN NlpConsumer - error in
consumer main loop
java.lang.RuntimeException: Couldn't create parent directories for /
home/datashare/dist/models/corenlp/3-8-0/en/stanford-corenlp-3.8.0-
models-english.jar
    at
com.amazonaws.services.s3.transfer.TransferManager.downloadDirectory(T
ransferManager.java:1275)
    at
com.amazonaws.services.s3.transfer.TransferManager.downloadDirectory(T
ransferManager.java:1157)
    at
com.amazonaws.services.s3.transfer.TransferManager.downloadDirectory(T
ransferManager.java:1148)
    at org.icij.datashare.io.RemoteFiles.download(RemoteFiles.java:65)
    at
org.icij.datashare.text.nlp.AbstractModels.downloadIfNecessary(Abstrac
tModels.java:114)
    at
org.icij.datashare.text.nlp.AbstractModels.load(AbstractModels.java:
66)
    at
org.icij.datashare.text.nlp.AbstractModels.get(AbstractModels.java:55)
    at
org.icij.datashare.text.nlp.corenlp.CorenlpPipeline.initializeNerAnnot
ator(CorenlpPipeline.java:180)
    at
org.icij.datashare.text.nlp.corenlp.CorenlpPipeline.initialize(Corenlp
Pipeline.java:77)
    at
org.icij.datashare.text.nlp.NlpConsumer.findNamedEntities(NlpConsumer.
java:66)
    at org.icij.datashare.text.nlp.NlpConsumer.run(NlpConsumer.java:
40)
    at
java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.j
ava:1149)
    at
java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.
java:624)
    at java.lang.Thread.run(Thread.java:748)
2019-01-13 23:14:57,935 [CORENLP-0] INFO NlpConsumer - extracting
CORENLP entities for document
fdcafc384c4e5e249f7489727d0105c0033054d6bd8a5e141a6f8442f321eeda
2019-01-13 23:14:57,935 [CORENLP-0] INFO CorenlpPipeline -
initializing CORENLP ENGLISH [TOKEN, SENTENCE, POS, LEMMA, NER]
2019-01-13 23:14:57,940 [CORENLP-0] INFO CoreNlpNerModels -
downloading models for language ENGLISH under dist/models/corenlp/
3-8-0/en
2019-01-13 23:14:58,419 [CORENLP-0] WARN NlpConsumer - error in
```

```
consumer main loop
java.lang.RuntimeException: Couldn't create parent directories for /
home/datashare/dist/models/corenlp/3-8-0/en/stanford-corenlp-3.8.0-
models-english.jar
    at
com.amazonaws.services.s3.transfer.TransferManager.downloadDirectory(T
ransferManager.java:1275)
    at
com.amazonaws.services.s3.transfer.TransferManager.downloadDirectory(T
ransferManager.java:1157)
    at
com.amazonaws.services.s3.transfer.TransferManager.downloadDirectory(T
ransferManager.java:1148)
        at org.icij.datashare.io.RemoteFiles.download(RemoteFiles.java:65)
        at
org.icij.datashare.text.nlp.AbstractModels.downloadIfNecessary(Abstrac
tModels.java:114)
        at
org.icij.datashare.text.nlp.AbstractModels.load(AbstractModels.java:
66)
            at
org.icij.datashare.text.nlp.AbstractModels.get(AbstractModels.java:55)
            at
org.icij.datashare.text.nlp.corenlp.CorenlpPipeline.initializeNerAnnot
ator(CorenlpPipeline.java:180)
            at
org.icij.datashare.text.nlp.corenlp.CorenlpPipeline.initialize(Corenlp
Pipeline.java:77)
            at
org.icij.datashare.text.nlp.NlpConsumer.findNamedEntities(NlpConsumer.
java:66)
                at org.icij.datashare.text.nlp.NlpConsumer.run(NlpConsumer.java:
40)
                at
java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.j
ava:1149)
                    at
java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.
java:624)
                        at java.lang.Thread.run(Thread.java:748)
2019-01-13 23:14:58,428 [CORENLP-0] INFO  NlpConsumer - extracting
CORENLP entities for document
4c76b7a367c810aa717ec49caf5bd8ee3edeefd197241f6bd3698ed5de2c4ddc
2019-01-13 23:14:58,428 [CORENLP-0] INFO  CorenlpPipeline -
initializing CORENLP ENGLISH [TOKEN, SENTENCE, POS, LEMMA, NER]
2019-01-13 23:14:58,441 [CORENLP-0] INFO  CoreNlpNerModels -
downloading models for language ENGLISH under dist/models/corenlp/
3-8-0/en
2019-01-13 23:14:58,922 [CORENLP-0] WARN  NlpConsumer - error in
consumer main loop
```

```
java.lang.RuntimeException: Couldn't create parent directories for /
home/datashare/dist/models/corenlp/3-8-0/en/stanford-corenlp-3.8.0-
models-english.jar
    at
com.amazonaws.services.s3.transfer.TransferManager.downloadDirectory(T
ransferManager.java:1275)
    at
com.amazonaws.services.s3.transfer.TransferManager.downloadDirectory(T
ransferManager.java:1157)
    at
com.amazonaws.services.s3.transfer.TransferManager.downloadDirectory(T
ransferManager.java:1148)
    at org.icij.datashare.io.RemoteFiles.download(RemoteFiles.java:65)
    at
org.icij.datashare.text.nlp.AbstractModels.downloadIfNecessary(Abstrac
tModels.java:114)
    at
org.icij.datashare.text.nlp.AbstractModels.load(AbstractModels.java:
66)
    at
org.icij.datashare.text.nlp.AbstractModels.get(AbstractModels.java:55)
    at
org.icij.datashare.text.nlp.corenlp.CorenlpPipeline.initializeNerAnnot
ator(CorenlpPipeline.java:180)
    at
org.icij.datashare.text.nlp.corenlp.CorenlpPipeline.initialize(Corenlp
Pipeline.java:77)
    at
org.icij.datashare.text.nlp.NlpConsumer.findNamedEntities(NlpConsumer.
java:66)
    at org.icij.datashare.text.nlp.NlpConsumer.run(NlpConsumer.java:
40)
    at
java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.j
ava:1149)
    at
java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.
java:624)
    at java.lang.Thread.run(Thread.java:748)
2019-01-13 23:14:58,930 [CORENLP-0] INFO NlpConsumer - extracting
CORENLP entities for document
5d5fd9b0a661367012b15064bd0feb70d0a7d2a762c704fe0daf2a922cf0a7f5
2019-01-13 23:14:58,931 [CORENLP-0] INFO CorenlpPipeline -
initializing CORENLP ENGLISH [TOKEN, SENTENCE, POS, LEMMA, NER]
2019-01-13 23:14:58,934 [CORENLP-0] INFO CoreNlpNerModels -
downloading models for language ENGLISH under dist/models/corenlp/
3-8-0/en
2019-01-13 23:14:59,414 [CORENLP-0] WARN NlpConsumer - error in
consumer main loop
java.lang.RuntimeException: Couldn't create parent directories for /
```

```
home/datashare/dist/models/corenlp/3-8-0/en/stanford-corenlp-3.8.0-
models-english.jar
    at
com.amazonaws.services.s3.transfer.TransferManager.downloadDirectory(T
ransferManager.java:1275)
    at
com.amazonaws.services.s3.transfer.TransferManager.downloadDirectory(T
ransferManager.java:1157)
    at
com.amazonaws.services.s3.transfer.TransferManager.downloadDirectory(T
ransferManager.java:1148)
    at org.icij.datashare.io.RemoteFiles.download(RemoteFiles.java:65)
    at
org.icij.datashare.text.nlp.AbstractModels.downloadIfNecessary(Abstrac
tModels.java:114)
    at
org.icij.datashare.text.nlp.AbstractModels.load(AbstractModels.java:
66)
    at
org.icij.datashare.text.nlp.AbstractModels.get(AbstractModels.java:55)
    at
org.icij.datashare.text.nlp.corenlp.CorenlpPipeline.initializeNerAnnot
ator(CorenlpPipeline.java:180)
    at
org.icij.datashare.text.nlp.corenlp.CorenlpPipeline.initialize(Corenlp
Pipeline.java:77)
    at
org.icij.datashare.text.nlp.NlpConsumer.findNamedEntities(NlpConsumer.
java:66)
    at org.icij.datashare.text.nlp.NlpConsumer.run(NlpConsumer.java:
40)
    at
java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.j
ava:1149)
    at
java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.
java:624)
    at java.lang.Thread.run(Thread.java:748)
2019-01-14 07:30:08,894 [Worker: RequestDispatcher: Thread-12] INFO
TaskResource - waiting for NlpApp CORENLP to listen...
2019-01-14 07:30:08,895 [pool-2-thread-3] INFO NlpApp - running
NlpApp for CORENLP pipeline with 1 thread(s)
2019-01-14 07:30:08,900 [pool-2-thread-3] INFO NlpForwarder - waiting
for messages on host [redis]
2019-01-14 07:30:08,903 [Worker: RequestDispatcher: Thread-12] INFO
TaskResource - ...CORENLP is listening
2019-01-14 07:30:08,903 [pool-2-thread-4] INFO ResumeNlpTask -
resuming NLP name finding for user local and [CORENLP]
2019-01-14 07:30:08,934 [pool-2-thread-4] INFO ResumeNlpTask - sent
EXTRACT_NLP message for 5 files without [CORENLP] pipeline tags
```

```
2019-01-14 07:30:08,935 [pool-2-thread-3] INFO RedisSubscriber -
Shutdown called. Unsubscribe done.
2019-01-14 07:30:08,936 [pool-2-thread-3] INFO NlpApp - forwarder
exited waiting for consumer(s) to finish
2019-01-14 07:30:08,956 [CORENLP-0] INFO NlpConsumer - extracting
CORENLP entities for document
8decc8571946d4cd70a024949e033a2a2a54377fe9f1c1b944c20f9ee11a9e51
2019-01-14 07:30:08,957 [CORENLP-0] INFO CorenlpPipeline -
initializing CORENLP ENGLISH [TOKEN, SENTENCE, POS, LEMMA, NER]
2019-01-14 07:30:08,965 [CORENLP-0] INFO CoreNlpNerModels -
downloading models for language ENGLISH under dist/models/corenlp/
3-8-0/en
2019-01-14 07:30:11,592 [CORENLP-0] WARN NlpConsumer - error in
consumer main loop
java.lang.RuntimeException: Couldn't create parent directories for /
home/datashare/dist/models/corenlp/3-8-0/en/stanford-corenlp-3.8.0-
models-english.jar
    at
com.amazonaws.services.s3.transfer.TransferManager.downloadDirectory(T
ransferManager.java:1275)
    at
com.amazonaws.services.s3.transfer.TransferManager.downloadDirectory(T
ransferManager.java:1157)
    at
com.amazonaws.services.s3.transfer.TransferManager.downloadDirectory(T
ransferManager.java:1148)
    at org.icij.datashare.io.RemoteFiles.download(RemoteFiles.java:65)
    at
org.icij.datashare.text.nlp.AbstractModels.downloadIfNecessary(Abstrac
tModels.java:114)
    at
org.icij.datashare.text.nlp.AbstractModels.load(AbstractModels.java:
66)
    at
org.icij.datashare.text.nlp.AbstractModels.get(AbstractModels.java:55)
    at
org.icij.datashare.text.nlp.corenlp.CorenlpPipeline.initializeNerAnnot
ator(CorenlpPipeline.java:180)
    at
org.icij.datashare.text.nlp.corenlp.CorenlpPipeline.initialize(Corenlp
Pipeline.java:77)
    at
org.icij.datashare.text.nlp.NlpConsumer.findNamedEntities(NlpConsumer.
java:66)
    at org.icij.datashare.text.nlp.NlpConsumer.run(NlpConsumer.java:
40)
    at
java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.j
ava:1149)
    at
```



```
java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.  
java:624)  
    at java.lang.Thread.run(Thread.java:748)  
2019-01-14 07:30:11,598 [CORENLP-0] INFO NlpConsumer - extracting  
CORENLP entities for document  
60bdd13ea4827b8de375c79dc3ff847f83b55bd73b6461523fdf8f843b5a0d5b  
2019-01-14 07:30:11,598 [CORENLP-0] INFO CorenlpPipeline -  
initializing CORENLP ENGLISH [TOKEN, SENTENCE, POS, LEMMA, NER]  
2019-01-14 07:30:11,603 [CORENLP-0] INFO CoreNlpNerModels -  
downloading models for language ENGLISH under dist/models/corenlp/  
3-8-0/en  
2019-01-14 07:30:12,150 [CORENLP-0] WARN NlpConsumer - error in  
consumer main loop  
java.lang.RuntimeException: Couldn't create parent directories for /  
home/datashare/dist/models/corenlp/3-8-0/en/stanford-corenlp-3.8.0-  
models-english.jar  
    at  
com.amazonaws.services.s3.transfer.TransferManager.downloadDirectory(T  
ransferManager.java:1275)  
    at  
com.amazonaws.services.s3.transfer.TransferManager.downloadDirectory(T  
ransferManager.java:1157)  
    at  
com.amazonaws.services.s3.transfer.TransferManager.downloadDirectory(T  
ransferManager.java:1148)  
    at org.icij.datashare.io.RemoteFiles.download(RemoteFiles.java:65)  
    at  
org.icij.datashare.text.nlp.AbstractModels.downloadIfNecessary(Abstrac  
tModels.java:114)  
    at  
org.icij.datashare.text.nlp.AbstractModels.load(AbstractModels.java:  
66)  
    at  
org.icij.datashare.text.nlp.AbstractModels.get(AbstractModels.java:55)  
    at  
org.icij.datashare.text.nlp.corenlp.CorenlpPipeline.initializeNerAnnot  
ator(CorenlpPipeline.java:180)  
    at  
org.icij.datashare.text.nlp.corenlp.CorenlpPipeline.initialize(Corenlp  
Pipeline.java:77)  
    at  
org.icij.datashare.text.nlp.NlpConsumer.findNamedEntities(NlpConsumer.  
java:66)  
    at org.icij.datashare.text.nlp.NlpConsumer.run(NlpConsumer.java:  
40)  
    at  
java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.j  
ava:1149)  
    at  
java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.
```

```
java:624)
    at java.lang.Thread.run(Thread.java:748)
2019-01-14 07:30:12,162 [CORENLP-0] INFO NlpConsumer - extracting
CORENLP entities for document
fdcafc384c4e5e249f7489727d0105c0033054d6bd8a5e141a6f8442f321eeda
2019-01-14 07:30:12,162 [CORENLP-0] INFO CorenlpPipeline -
initializing CORENLP ENGLISH [TOKEN, SENTENCE, POS, LEMMA, NER]
2019-01-14 07:30:12,166 [CORENLP-0] INFO CoreNlpNerModels -
downloading models for language ENGLISH under dist/models/corenlp/
3-8-0/en
2019-01-14 07:30:12,667 [CORENLP-0] WARN NlpConsumer - error in
consumer main loop
java.lang.RuntimeException: Couldn't create parent directories for /
home/datashare/dist/models/corenlp/3-8-0/en/stanford-corenlp-3.8.0-
models-english.jar
    at
com.amazonaws.services.s3.transfer.TransferManager.downloadDirectory(T
ransferManager.java:1275)
    at
com.amazonaws.services.s3.transfer.TransferManager.downloadDirectory(T
ransferManager.java:1157)
    at
com.amazonaws.services.s3.transfer.TransferManager.downloadDirectory(T
ransferManager.java:1148)
    at org.icij.datashare.io.RemoteFiles.download(RemoteFiles.java:65)
    at
org.icij.datashare.text.nlp.AbstractModels.downloadIfNecessary(Abstrac
tModels.java:114)
    at
org.icij.datashare.text.nlp.AbstractModels.load(AbstractModels.java:
66)
    at
org.icij.datashare.text.nlp.AbstractModels.get(AbstractModels.java:55)
    at
org.icij.datashare.text.nlp.corenlp.CorenlpPipeline.initializeNerAnnot
ator(CorenlpPipeline.java:180)
    at
org.icij.datashare.text.nlp.corenlp.CorenlpPipeline.initialize(Corenlp
Pipeline.java:77)
    at
org.icij.datashare.text.nlp.NlpConsumer.findNamedEntities(NlpConsumer.
java:66)
    at org.icij.datashare.text.nlp.NlpConsumer.run(NlpConsumer.java:
40)
    at
java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.j
ava:1149)
    at
java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.
java:624)
```

```
    at java.lang.Thread.run(Thread.java:748)
2019-01-14 07:30:12,682 [CORENLP-0] INFO NlpConsumer - extracting
CORENLP entities for document
4c76b7a367c810aa717ec49caf5bd8ee3edeefd197241f6bd3698ed5de2c4ddc
2019-01-14 07:30:12,683 [CORENLP-0] INFO CorenlpPipeline -
initializing CORENLP ENGLISH [TOKEN, SENTENCE, POS, LEMMA, NER]
2019-01-14 07:30:12,687 [CORENLP-0] INFO CoreNlpNerModels -
downloading models for language ENGLISH under dist/models/corenlp/
3-8-0/en
2019-01-14 07:30:13,226 [CORENLP-0] WARN NlpConsumer - error in
consumer main loop
java.lang.RuntimeException: Couldn't create parent directories for /
home/datashare/dist/models/corenlp/3-8-0/en/stanford-corenlp-3.8.0-
models-english.jar
    at
com.amazonaws.services.s3.transfer.TransferManager.downloadDirectory(T
ransferManager.java:1275)
    at
com.amazonaws.services.s3.transfer.TransferManager.downloadDirectory(T
ransferManager.java:1157)
    at
com.amazonaws.services.s3.transfer.TransferManager.downloadDirectory(T
ransferManager.java:1148)
    at org.icij.datashare.io.RemoteFiles.download(RemoteFiles.java:65)
    at
org.icij.datashare.text.nlp.AbstractModels.downloadIfNecessary(Abstrac
tModels.java:114)
    at
org.icij.datashare.text.nlp.AbstractModels.load(AbstractModels.java:
66)
    at
org.icij.datashare.text.nlp.AbstractModels.get(AbstractModels.java:55)
    at
org.icij.datashare.text.nlp.corenlp.CorenlpPipeline.initializeNerAnnot
ator(CorenlpPipeline.java:180)
    at
org.icij.datashare.text.nlp.corenlp.CorenlpPipeline.initialize(Corenlp
Pipeline.java:77)
    at
org.icij.datashare.text.nlp.NlpConsumer.findNamedEntities(NlpConsumer.
java:66)
    at org.icij.datashare.text.nlp.NlpConsumer.run(NlpConsumer.java:
40)
    at
java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.j
ava:1149)
    at
java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.
java:624)
    at java.lang.Thread.run(Thread.java:748)
```

```
2019-01-14 07:30:13,236 [CORENLP-0] INFO NlpConsumer - extracting
CORENLP entities for document
5d5fd9b0a661367012b15064bd0feb70d0a7d2a762c704fe0daf2a922cf0a7f5
2019-01-14 07:30:13,237 [CORENLP-0] INFO CorenlpPipeline -
initializing CORENLP ENGLISH [TOKEN, SENTENCE, POS, LEMMA, NER]
2019-01-14 07:30:13,240 [CORENLP-0] INFO CoreNlpNerModels -
downloading models for language ENGLISH under dist/models/corenlp/
3-8-0/en
2019-01-14 07:30:13,857 [CORENLP-0] WARN NlpConsumer - error in
consumer main loop
java.lang.RuntimeException: Couldn't create parent directories for /
home/datashare/dist/models/corenlp/3-8-0/en/stanford-corenlp-3.8.0-
models-english.jar
    at
com.amazonaws.services.s3.transfer.TransferManager.downloadDirectory(T
ransferManager.java:1275)
    at
com.amazonaws.services.s3.transfer.TransferManager.downloadDirectory(T
ransferManager.java:1157)
    at
com.amazonaws.services.s3.transfer.TransferManager.downloadDirectory(T
ransferManager.java:1148)
    at org.icij.datashare.io.RemoteFiles.download(RemoteFiles.java:65)
    at
org.icij.datashare.text.nlp.AbstractModels.downloadIfNecessary(Abstrac
tModels.java:114)
    at
org.icij.datashare.text.nlp.AbstractModels.load(AbstractModels.java:
66)
    at
org.icij.datashare.text.nlp.AbstractModels.get(AbstractModels.java:55)
    at
org.icij.datashare.text.nlp.corenlp.CorenlpPipeline.initializeNerAnnot
ator(CorenlpPipeline.java:180)
    at
org.icij.datashare.text.nlp.corenlp.CorenlpPipeline.initialize(Corenlp
Pipeline.java:77)
    at
org.icij.datashare.text.nlp.NlpConsumer.findNamedEntities(NlpConsumer.
java:66)
    at org.icij.datashare.text.nlp.NlpConsumer.run(NlpConsumer.java:
40)
    at
java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.j
ava:1149)
    at
java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.
java:624)
    at java.lang.Thread.run(Thread.java:748)
2019-01-14 07:43:20,415 [Worker: RequestDispatcher: Thread-24] INFO
```

```
TaskResource - waiting for NlpApp CORENLP to listen...
2019-01-14 07:43:20,415 [pool-2-thread-2] INFO NlpApp - running
NlpApp for CORENLP pipeline with 1 thread(s)
2019-01-14 07:43:20,416 [pool-2-thread-2] INFO NlpForwarder - waiting
for messages on host [redis]
2019-01-14 07:43:20,418 [Worker: RequestDispatcher: Thread-24] INFO
TaskResource - ...CORENLP is listening
2019-01-14 07:43:20,419 [pool-2-thread-4] INFO ResumeNlpTask -
resuming NLP name finding for user local and [CORENLP]
2019-01-14 07:43:20,462 [pool-2-thread-4] INFO ResumeNlpTask - sent
EXTRACT_NLP message for 5 files without [CORENLP] pipeline tags
2019-01-14 07:43:20,476 [pool-2-thread-2] INFO RedisSubscriber -
Shutdown called. Unsubscribe done.
2019-01-14 07:43:20,476 [pool-2-thread-2] INFO NlpApp - forwarder
exited waiting for consumer(s) to finish
2019-01-14 07:43:20,485 [CORENLP-0] INFO NlpConsumer - extracting
CORENLP entities for document
8decc8571946d4cd70a024949e033a2a2a54377fe9f1c1b944c20f9ee11a9e51
2019-01-14 07:43:20,485 [CORENLP-0] INFO CorenlpPipeline -
initializing CORENLP ENGLISH [TOKEN, SENTENCE, POS, LEMMA, NER]
2019-01-14 07:43:20,499 [CORENLP-0] INFO CoreNlpNerModels -
downloading models for language ENGLISH under dist/models/corenlp/
3-8-0/en
2019-01-14 07:43:25,689 [CORENLP-0] WARN NlpConsumer - error in
consumer main loop
java.lang.RuntimeException: Couldn't create parent directories for /
home/datashare/dist/models/corenlp/3-8-0/en/stanford-corenlp-3.8.0-
models-english.jar
    at
com.amazonaws.services.s3.transfer.TransferManager.downloadDirectory(T
ransferManager.java:1275)
    at
com.amazonaws.services.s3.transfer.TransferManager.downloadDirectory(T
ransferManager.java:1157)
    at
com.amazonaws.services.s3.transfer.TransferManager.downloadDirectory(T
ransferManager.java:1148)
    at org.icij.datashare.io.RemoteFiles.download(RemoteFiles.java:65)
    at
org.icij.datashare.text.nlp.AbstractModels.downloadIfNecessary(Abstrac
tModels.java:114)
    at
org.icij.datashare.text.nlp.AbstractModels.load(AbstractModels.java:
66)
    at
org.icij.datashare.text.nlp.AbstractModels.get(AbstractModels.java:55)
    at
org.icij.datashare.text.nlp.corenlp.CorenlpPipeline.initializeNerAnnot
ator(CorenlpPipeline.java:180)
    at
```

```
org.icij.datashare.text.nlp.corenlp.CorenlpPipeline.initialize(Corenlp
Pipeline.java:77)
    at
org.icij.datashare.text.nlp.NlpConsumer.findNamedEntities(NlpConsumer.
java:66)
    at org.icij.datashare.text.nlp.NlpConsumer.run(NlpConsumer.java:
40)
    at
java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.j
ava:1149)
    at
java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.
java:624)
    at java.lang.Thread.run(Thread.java:748)
2019-01-14 07:43:25,699 [CORENLP-0] INFO  NlpConsumer - extracting
CORENLP entities for document
60bdd13ea4827b8de375c79dc3ff847f83b55bd73b6461523fdf8f843b5a0d5b
2019-01-14 07:43:25,699 [CORENLP-0] INFO  CorenlpPipeline -
initializing CORENLP ENGLISH [TOKEN, SENTENCE, POS, LEMMA, NER]
2019-01-14 07:43:25,703 [CORENLP-0] INFO  CoreNlpNerModels -
downloading models for language ENGLISH under dist/models/corenlp/
3-8-0/en
2019-01-14 07:43:26,176 [CORENLP-0] WARN  NlpConsumer - error in
consumer main loop
java.lang.RuntimeException: Couldn't create parent directories for /
home/datashare/dist/models/corenlp/3-8-0/en/stanford-corenlp-3.8.0-
models-english.jar
    at
com.amazonaws.services.s3.transfer.TransferManager.downloadDirectory(T
ransferManager.java:1275)
    at
com.amazonaws.services.s3.transfer.TransferManager.downloadDirectory(T
ransferManager.java:1157)
    at
com.amazonaws.services.s3.transfer.TransferManager.downloadDirectory(T
ransferManager.java:1148)
    at org.icij.datashare.io.RemoteFiles.download(RemoteFiles.java:65)
    at
org.icij.datashare.text.nlp.AbstractModels.downloadIfNecessary(Abstrac
tModels.java:114)
    at
org.icij.datashare.text.nlp.AbstractModels.load(AbstractModels.java:
66)
    at
org.icij.datashare.text.nlp.AbstractModels.get(AbstractModels.java:55)
    at
org.icij.datashare.text.nlp.corenlp.CorenlpPipeline.initializeNerAnnot
ator(CorenlpPipeline.java:180)
    at
org.icij.datashare.text.nlp.corenlp.CorenlpPipeline.initialize(Corenlp
```

```
Pipeline.java:77)
    at
org.icij.datashare.text.nlp.NlpConsumer.findNamedEntities(NlpConsumer.
java:66)
    at org.icij.datashare.text.nlp.NlpConsumer.run(NlpConsumer.java:
40)
    at
java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.j
ava:1149)
    at
java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.
java:624)
    at java.lang.Thread.run(Thread.java:748)
2019-01-14 07:43:26,186 [CORENLP-0] INFO NlpConsumer - extracting
CORENLP entities for document
fdcafc384c4e5e249f7489727d0105c0033054d6bd8a5e141a6f8442f321eeda
2019-01-14 07:43:26,186 [CORENLP-0] INFO CorenlpPipeline -
initializing CORENLP ENGLISH [TOKEN, SENTENCE, POS, LEMMA, NER]
2019-01-14 07:43:26,190 [CORENLP-0] INFO CoreNlpNerModels -
downloading models for language ENGLISH under dist/models/corenlp/
3-8-0/en
2019-01-14 07:43:26,669 [CORENLP-0] WARN NlpConsumer - error in
consumer main loop
java.lang.RuntimeException: Couldn't create parent directories for /
home/datashare/dist/models/corenlp/3-8-0/en/stanford-corenlp-3.8.0-
models-english.jar
    at
com.amazonaws.services.s3.transfer.TransferManager.downloadDirectory(T
ransferManager.java:1275)
    at
com.amazonaws.services.s3.transfer.TransferManager.downloadDirectory(T
ransferManager.java:1157)
    at
com.amazonaws.services.s3.transfer.TransferManager.downloadDirectory(T
ransferManager.java:1148)
    at org.icij.datashare.io.RemoteFiles.download(RemoteFiles.java:65)
    at
org.icij.datashare.text.nlp.AbstractModels.downloadIfNecessary(Abstrac
tModels.java:114)
    at
org.icij.datashare.text.nlp.AbstractModels.load(AbstractModels.java:
66)
    at
org.icij.datashare.text.nlp.AbstractModels.get(AbstractModels.java:55)
    at
org.icij.datashare.text.nlp.corenlp.CorenlpPipeline.initializeNerAnnot
ator(CorenlpPipeline.java:180)
    at
org.icij.datashare.text.nlp.corenlp.CorenlpPipeline.initialize(Corenlp
Pipeline.java:77)
```

```
    at
org.icij.datashare.text.nlp.NlpConsumer.findNamedEntities(NlpConsumer.
java:66)
    at org.icij.datashare.text.nlp.NlpConsumer.run(NlpConsumer.java:
40)
    at
java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.j
ava:1149)
    at
java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.
java:624)
    at java.lang.Thread.run(Thread.java:748)
2019-01-14 07:43:26,675 [CORENLP-0] INFO  NlpConsumer - extracting
CORENLP entities for document
4c76b7a367c810aa717ec49caf5bd8ee3edeedfd197241f6bd3698ed5de2c4ddc
2019-01-14 07:43:26,675 [CORENLP-0] INFO  CorenlpPipeline -
initializing CORENLP ENGLISH [TOKEN, SENTENCE, POS, LEMMA, NER]
2019-01-14 07:43:26,679 [CORENLP-0] INFO  CoreNlpNerModels -
downloading models for language ENGLISH under dist/models/corenlp/
3-8-0/en
2019-01-14 07:43:27,155 [CORENLP-0] WARN  NlpConsumer - error in
consumer main loop
java.lang.RuntimeException: Couldn't create parent directories for /
home/datashare/dist/models/corenlp/3-8-0/en/stanford-corenlp-3.8.0-
models-english.jar
    at
com.amazonaws.services.s3.transfer.TransferManager.downloadDirectory(T
ransferManager.java:1275)
    at
com.amazonaws.services.s3.transfer.TransferManager.downloadDirectory(T
ransferManager.java:1157)
    at
com.amazonaws.services.s3.transfer.TransferManager.downloadDirectory(T
ransferManager.java:1148)
    at org.icij.datashare.io.RemoteFiles.download(RemoteFiles.java:65)
    at
org.icij.datashare.text.nlp.AbstractModels.downloadIfNecessary(Abstrac
tModels.java:114)
    at
org.icij.datashare.text.nlp.AbstractModels.load(AbstractModels.java:
66)
    at
org.icij.datashare.text.nlp.AbstractModels.get(AbstractModels.java:55)
    at
org.icij.datashare.text.nlp.corenlp.CorenlpPipeline.initializeNerAnnot
ator(CorenlpPipeline.java:180)
    at
org.icij.datashare.text.nlp.corenlp.CorenlpPipeline.initialize(Corenlp
Pipeline.java:77)
    at
```



```
org.icij.datashare.text.nlp.NlpConsumer.findNamedEntities(NlpConsumer.  
java:66)  
    at org.icij.datashare.text.nlp.NlpConsumer.run(NlpConsumer.java:  
40)  
        at  
java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.j  
ava:1149)  
            at  
java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.  
java:624)  
                at java.lang.Thread.run(Thread.java:748)  
2019-01-14 07:43:27,161 [CORENLP-0] INFO NlpConsumer - extracting  
CORENLP entities for document  
5d5fd9b0a661367012b15064bd0feb70d0a7d2a762c704fe0daf2a922cf0a7f5  
2019-01-14 07:43:27,161 [CORENLP-0] INFO CorenlpPipeline -  
initializing CORENLP ENGLISH [TOKEN, SENTENCE, POS, LEMMA, NER]  
2019-01-14 07:43:27,165 [CORENLP-0] INFO CoreNlpNerModels -  
downloading models for language ENGLISH under dist/models/corenlp/  
3-8-0/en  
2019-01-14 07:43:27,635 [CORENLP-0] WARN NlpConsumer - error in  
consumer main loop  
java.lang.RuntimeException: Couldn't create parent directories for /  
home/datashare/dist/models/corenlp/3-8-0/en/stanford-corenlp-3.8.0-  
models-english.jar  
    at  
com.amazonaws.services.s3.transfer.TransferManager.downloadDirectory(T  
ransferManager.java:1275)  
        at  
com.amazonaws.services.s3.transfer.TransferManager.downloadDirectory(T  
ransferManager.java:1157)  
            at  
com.amazonaws.services.s3.transfer.TransferManager.downloadDirectory(T  
ransferManager.java:1148)  
                at org.icij.datashare.io.RemoteFiles.download(RemoteFiles.java:65)  
                    at  
org.icij.datashare.text.nlp.AbstractModels.downloadIfNecessary(Abstrac  
tModels.java:114)  
                        at  
org.icij.datashare.text.nlp.AbstractModels.load(AbstractModels.java:  
66)  
                            at  
org.icij.datashare.text.nlp.AbstractModels.get(AbstractModels.java:55)  
                                at  
org.icij.datashare.text.nlp.corenlp.CorenlpPipeline.initializeNerAnnot  
ator(CorenlpPipeline.java:180)  
                                    at  
org.icij.datashare.text.nlp.corenlp.CorenlpPipeline.initialize(Corenlp  
Pipeline.java:77)  
                                        at  
org.icij.datashare.text.nlp.NlpConsumer.findNamedEntities(NlpConsumer.
```

```
java:66)
  at org.icij.datashare.text.nlp.NlpConsumer.run(NlpConsumer.java:
40)
  at
java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.j
ava:1149)
  at
java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.
java:624)
  at java.lang.Thread.run(Thread.java:748)
solines-mbp:~ solineledesert$
```