

# TikaAndMITIE

## MIT Information Extraction (MITIE) with Tika

MIT Information Extraction provides free state-of-the-art information extraction tools. The current release includes tools for performing named entity extraction and binary relation detection as well as tools for training custom extractors and relation detectors.

Support for MITIE is provided as a runtime binding in Tika `org.apache.tika.parser.ner.mitie.MITIENERrecogniser` class

- [MIT Information Extraction \(MITIE\) with Tika](#)
  - [Installation](#)
  - [Running MITIE with Tika-App](#)
  - [Running MITIE with Tika-Server](#)

## Installation

1. **Simple by downloading [mitie-resources](#)** : Use following commands to set up your mitie-resources.

*MAC OS Requirement:* Download and install [Homebrew](#).

*Linux/Windows:* No pre-requisite.

```
git clone https://github.com/manalishah/mitie-resources
cd mitie-resources
# absolute path to mitie-resources folder
export NER_RES=$PWD
chmod a+x install.sh
./install.sh
```

## Running MITIE with Tika-App

For running MITIE, the following steps are essential:

- Supply the `java.library.path` as absolute path to jni shared library obtained from building MITIE (required for MAC OS only)
- Supply the `javamitie.jar` in classpath
- Supply the complete model path to `ner_model` obtained from building MITIE
- Supply the ner implementation class as `MITIENERrecogniser`

1. \* For Mac OS\*

```
export TIKA_APP={your/path/to/tika-app}/target/tika-app-1.13-SNAPSHOT.jar

java -Djava.library.path=$NER_RES/MITIE/mitielib -Dner.mitie.model=$NER_RES/MITIE/MITIE-models/english/ner_model.dat -Dner.impl.class=org.apache.tika.parser.ner.mitie.MITIENERrecogniser -classpath $NER_RES/MITIE/mitielib/javamitie.jar:$TIKA_APP org.apache.tika.cli.TikaCLI --config=$NER_RES/tika-config.xml -m $NER_RES/sample.txt
```

2. \* For LINUX/Windows\*

```
export TIKA_APP={your/path/to/tika-app}/target/tika-app-1.13-SNAPSHOT.jar

java -Dner.mitie.model=$NER_RES/MITIE/MITIE-models/english/ner_model.dat -Dner.impl.class=org.apache.tika.parser.ner.mitie.MITIENERrecogniser -classpath $NER_RES/MITIE/mitielib/javamitie.jar:$TIKA_APP org.apache.tika.cli.TikaCLI --config=$NER_RES/tika-config.xml -m $NER_RES/sample.txt
```

This will output metadata keys along with named entities extracted using mitie:

```
Content-Length: 63
Content-Type: text/plain
NER_LOCATION: Los Angeles
NER_LOCATION: California
X-Parsed-By: org.apache.tika.parser.CompositeParser
X-Parsed-By: org.apache.tika.parser.ner.NamedEntityParser
resourceName: sample.txt
```

## Running MITIE with Tika-Server

### 1. \* For Mac OS\*

```
export TIKA_SERVER={your/path/to/tika-server}/target/tika-server-1.13-SNAPSHOT.jar

java -Djava.library.path=$NER_RES/MITIE/mitielib -Dner.mitie.model=$NER_RES/MITIE/MITIE-models/english/ner_model.dat -Dner.impl.class=org.apache.tika.parser.ner.mitie.MITIENERecogniser -classpath $NER_RES/MITIE/mitielib/javamitie.jar:$TIKA_SERVER org.apache.tika.server.TikaServerCli --config=$NER_RES/tika-config.xml -p 9998
```

### 2. \* For LINUX/Windows\*

```
export TIKA_SERVER={your/path/to/tika-server}/target/tika-server-1.13-SNAPSHOT.jar

java -Dner.mitie.model=$NER_RES/MITIE/MITIE-models/english/ner_model.dat -Dner.impl.class=org.apache.tika.parser.ner.mitie.MITIENERecogniser -classpath $NER_RES/MITIE/mitielib/javamitie.jar:$TIKA_SERVER org.apache.tika.server.TikaServerCli --config=$NER_RES/tika-config.xml -p 9998
```

This will start the Tika-Server enabled with MITIE Named Entity Parser at <http://localhost:9998>  
To test the server try the sample.txt file provided in the mitie-resources folder

```
curl -T $NER_RES/sample.txt http://localhost:9998/meta -H "Accept: application/json"
```

This should return metadata keys in a JSON format:

```
{
  "Content-Type": "text/plain",
  "NER_LOCATION": [ "Los Angeles", "California" ],
  "X-Parsed-By": [ "org.apache.tika.parser.CompositeParser", "org.apache.tika.parser.ner.NamedEntityParser" ],
  "language": "sl"
}
```