Contributing to the Lapps Grid

Lapps Service Wrapping

Lapps Grid Group

May 26, 2014

Outline

- Introduction
- From Software to Web Service
- From NLP Tool to Lapps Service
    - Java Example
    - Python Example
- Conclusion
- Reference

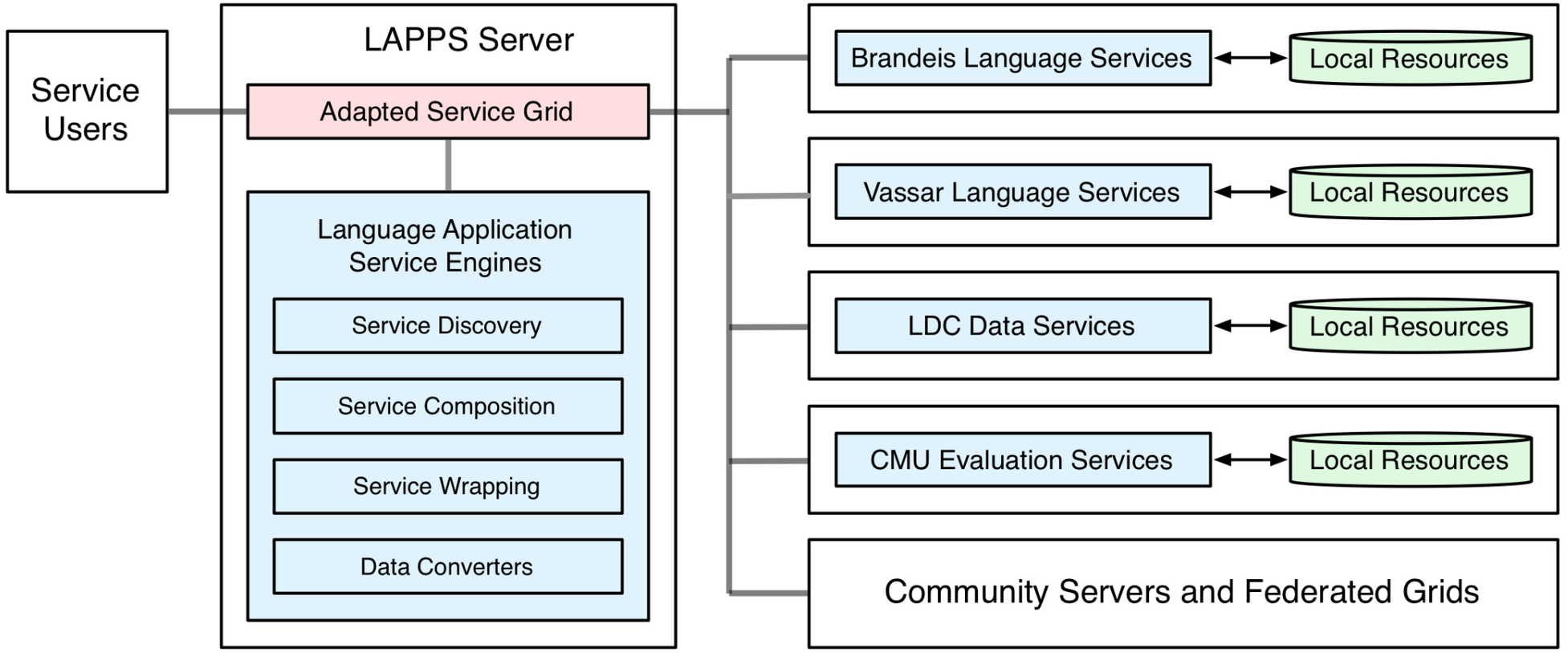
The Language Application Grid

- Availability & Interoperability of NLP Tools
  - Java, Python, tools
    - OpenNLP, Stanford NLP, Gate, NLTK
- Language Application (Lapps) Grid Project
  - Language Service
  - Lapps API Design

Lapps Grid Architecture

Using Composite Lapps Services

Wrapping Atomic Lapps Services



Lapps API Design

- Consistent Interface
- Discriminator
- JSON Format

Consistent Interface (Java)

```java
package org.lappsgrid.api;

import jp.go.nict.langrid.commons.rpc.intf.Service;

@Service(namespace = "lapps:service")
public interface WebService {
   /**
    * Returns the set of data types that must be present in the
    * input to the {@link #execute(Data)} method
    */
   long[] requires();

   /**
    * Returns the set of data types that will be included in the output.
    */
   long[] produces();

   /**
    * Executes a web service on the given input. Returns the output, if any,
    * of the web service in a {@link Data} object.
    */
   Data execute(Data input);

   /**
    * Configures a DataSource.
    * <p/>
    * Returns any errors in a {@link Data} object. Otherwise returns a Data
    * object with the "ok" Discriminator type.
    *
    * @param config
    * @return
    */
   Data configure(Data config);
}
```

Discriminator

long [] requires()
long [] produces()

Discriminator values

Discriminator	Name
Basic data types	
0	error
1	ok
2	meta
3	text
4	xml
5	string-list
Document types	
1024	document
1025	gate
1026	uima
1027	stanford
1028	opennlp
1029	graf
1030	ptb
1031	json
1032	json-ld

JSON


```
{   "@context": "http://vocab.lappsgrid.org/context-1.0.0.jsonld",
    "metadata": {},
    "text": {        "@value": "Hi, how are you today?"    },
    "steps": [
        { "metadata": {
                "contains": {
                    "Token": {
                        "producer": "edu.brandeis.cs.lappsgrid.opennlp.Tokenizer:0.0.4",
                        "type": "tokenizer:opennlp"       }  } },
            "annotations": [
                {  "@type": "Token", "id": "tok0", "start": 0, "end": 2,
                    "features": { "word": "Hi"}    },
                {  "@type": "Token", "id": "tok1", "start": 2, "end": 3,
                    "features": { "word": ","}    },
                {   "@type": "Token", "id": "tok2", "start": 4, "end": 7,
                    "features": { "word": "how"}  },
                {   "@type": "Token", "id": "tok3", "start": 8, "end": 11,
                    "features": { "word": "are" } },
                {   "@type": "Token", "id": "tok4", "start": 12, "end": 15,
                    "features": { "word": "you" } },
                {   "@type": "Token", "id": "tok5", "start": 16, "end": 21,
                    "features": { "word": "today" } },
                {   "@type": "Token", "id": "tok6", "start": 21, "end": 22,
                    "features": { "word": "?" } } ]
        }
    ]
}
```

Contributing to Lapps Grid

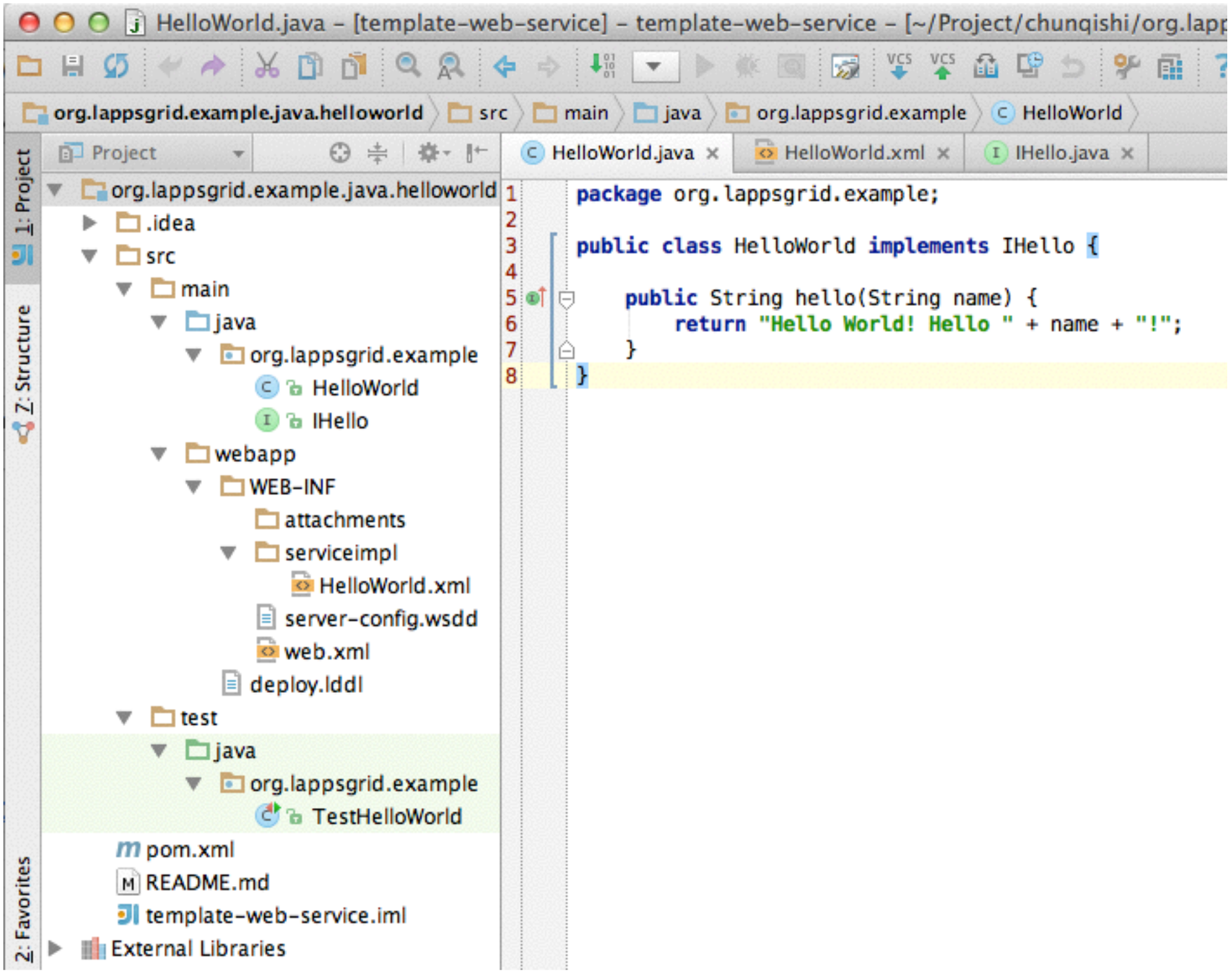
- Wrapping Lapps Service
    - NLP tools + Lapps API to atomic Lapps service
- Registering to Service Manager
    - Atomic Lapps services become available for searching and compositing

Service Wrapping Tutorial

- Web Service: “Hello World!”
  - “Hello World” Program (Java) —> WSDL
- Lapps Service: “Stanford Tagger”
  - Stanford Tagger (Java) + Lapps API —> WSDL
- Lapps Service: “NLTK Tagger”
  - NLTK Tagger (Python) + Lapps API —>WSDL

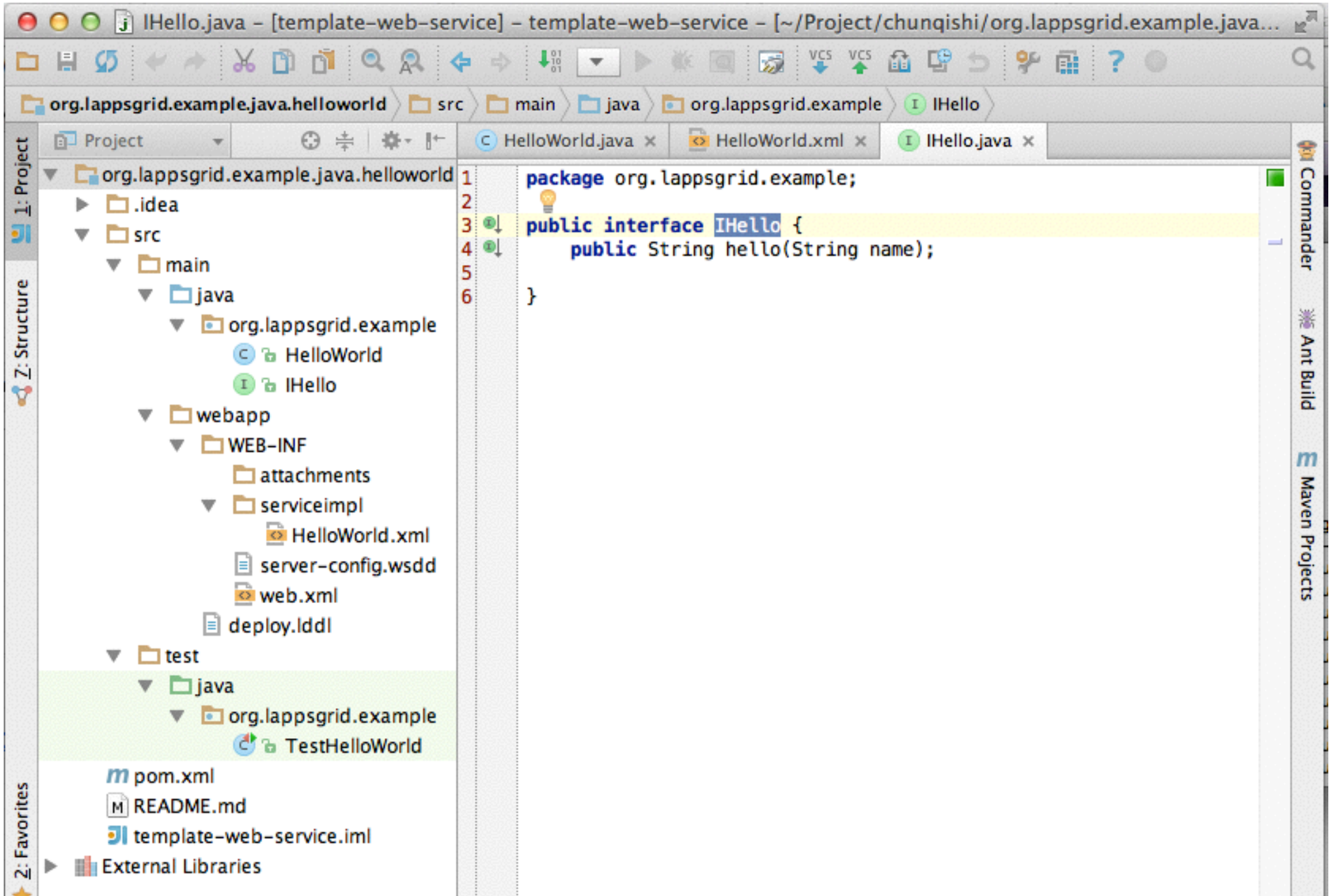
Web Service Wrapping

Hello World (Java)



https://github.com/chunqishi/org.lappsgrid.example.java.helloworld

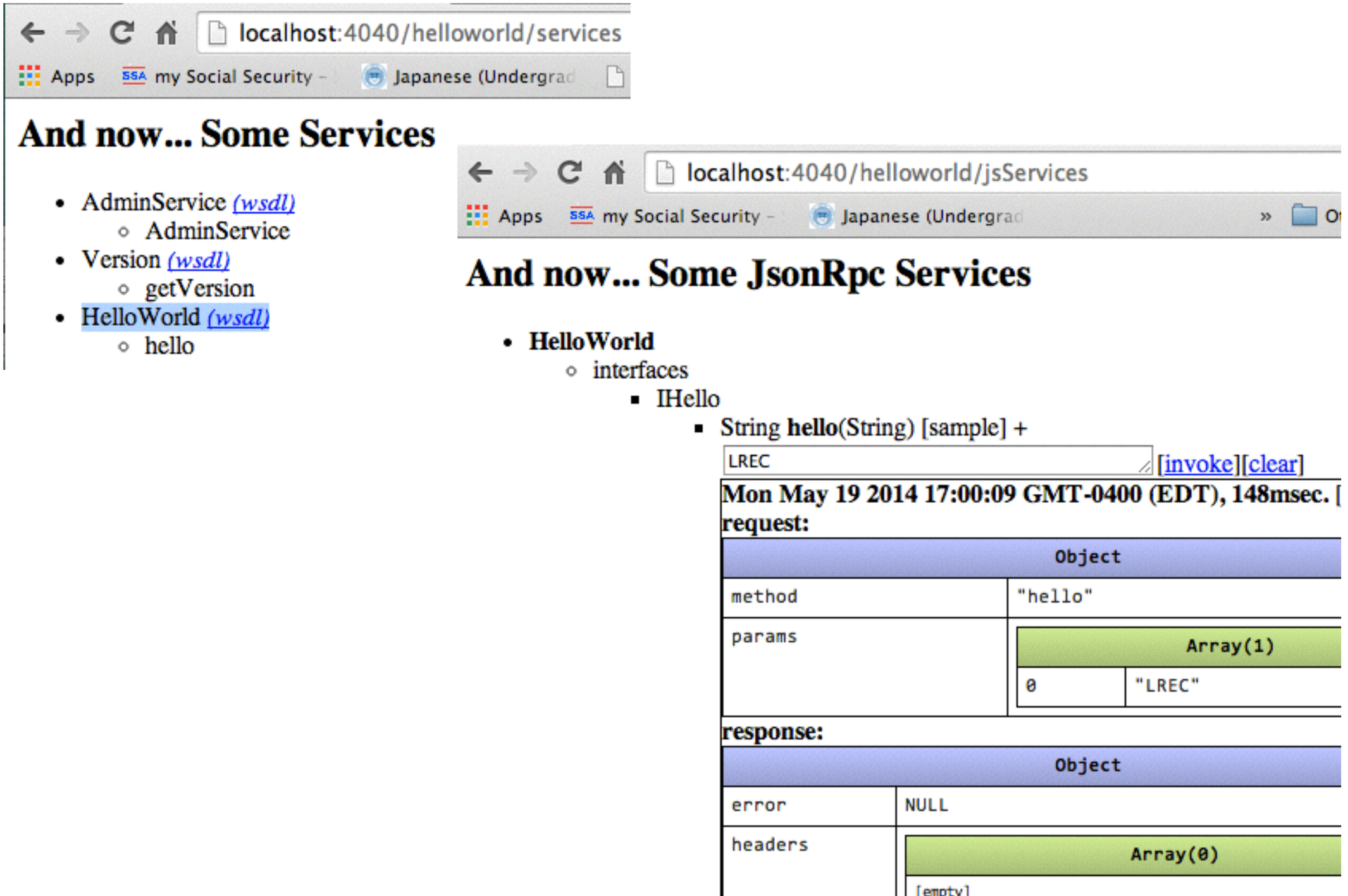
Interface Design



Developing Template

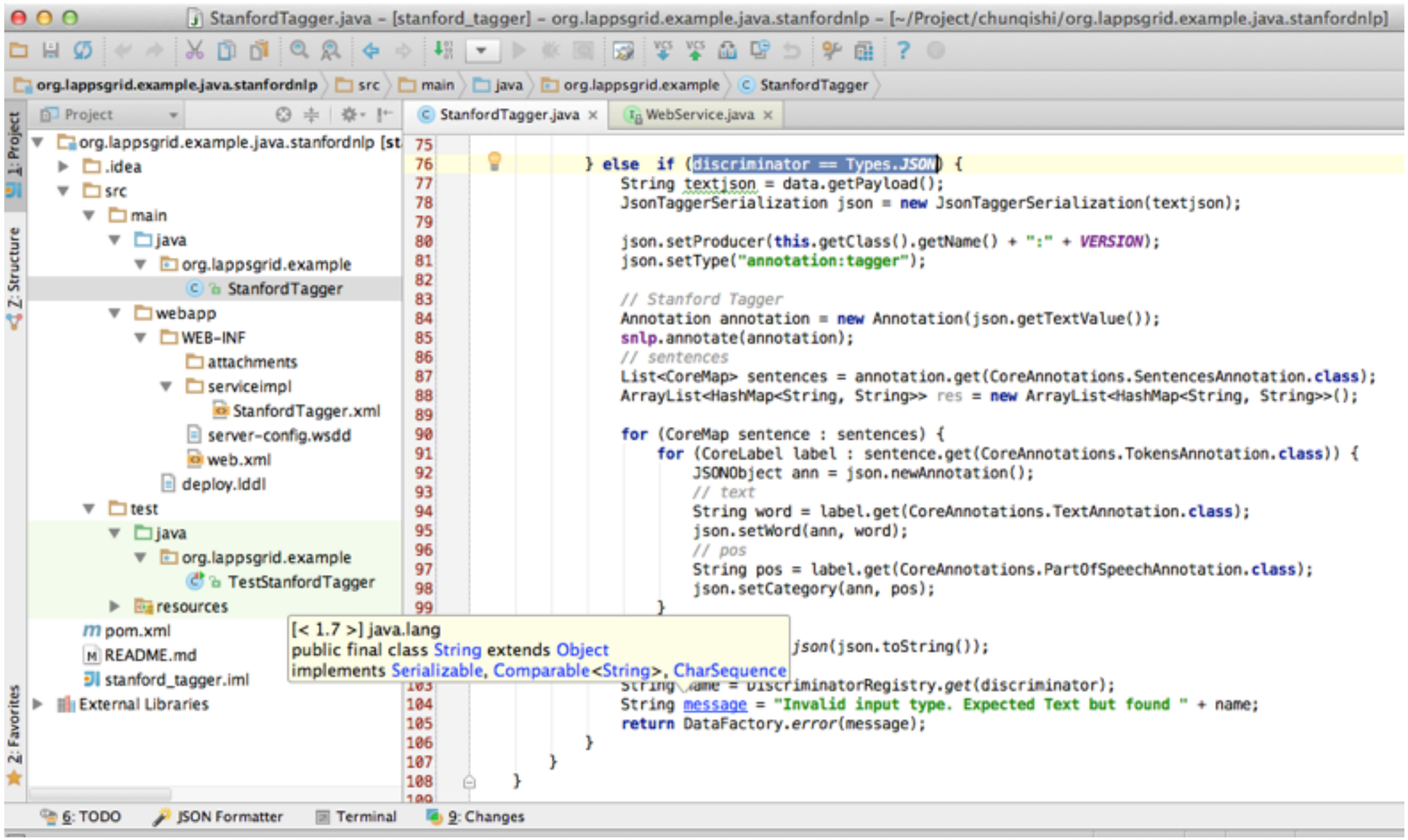
- Developing Template
  - Maven for Dependency Library Management
  - Github Repository
    - https://github.com/chunqishi/org.lappsgrid.example.java.helloworld
- Local Test
  - Maven Compile/Package & Jetty Server based Testing
  - Command: *mvn clean package jetty:run*

Web Service WSDL



Lapps Service Wrapping (Java)

Developing Template



https://github.com/chunqishi/org.lappsgrid.example.java.stanfordnlp

Stanford Tagger Wrapping

- Java Wrapping

```java
// Stanford Tagger
Annotation annotation = new Annotation(json.getTextValue());
snlp.annotate(annotation);
// sentences
List<CoreMap> sentences = annotation.get(CoreAnnotations.SentencesAnnotation.class);
ArrayList<HashMap<String, String>> res = new ArrayList<HashMap<String, String>>();

for (CoreMap sentence : sentences) {
    for (CoreLabel label : sentence.get(CoreAnnotations.TokensAnnotation.class)) {
        JSONObject ann = json.newAnnotation();
        // text
        String word = label.get(CoreAnnotations.TextAnnotation.class);
        json.setWord(ann, word);
        // pos
        String pos = label.get(CoreAnnotations.PartOfSpeechAnnotation.class);
        json.setCategory(ann, pos);
    }
}
```

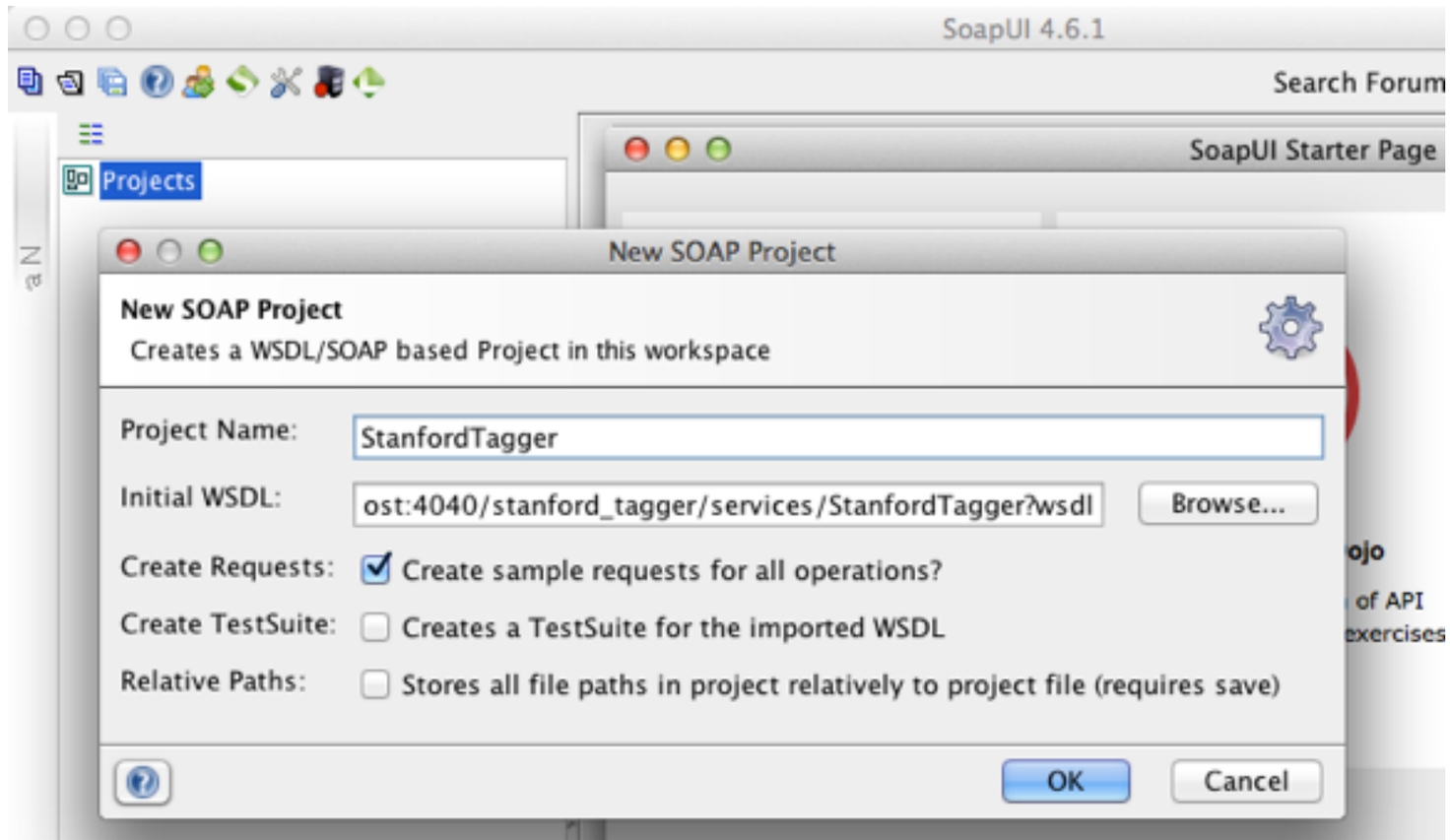
- Jetty Running

```
shis-MacBook-Air:org.lappsgrid.example.java.stanfordnlp shi$
shis-MacBook-Air:org.lappsgrid.example.java.stanfordnlp shi$ export MAVEN_OPTS="-Xmx1024M"
shis-MacBook-Air:org.lappsgrid.example.java.stanfordnlp shi$ mvn jetty:run
[INFO] Scanning for projects...
[INFO]
[INFO] ------------------------------------------------------------------------
[INFO] Building Java Stanford NLP Tagger Example 0.0.1-SNAPSHOT
[INFO] ------------------------------------------------------------------------
```

Stanford Tagger Testing

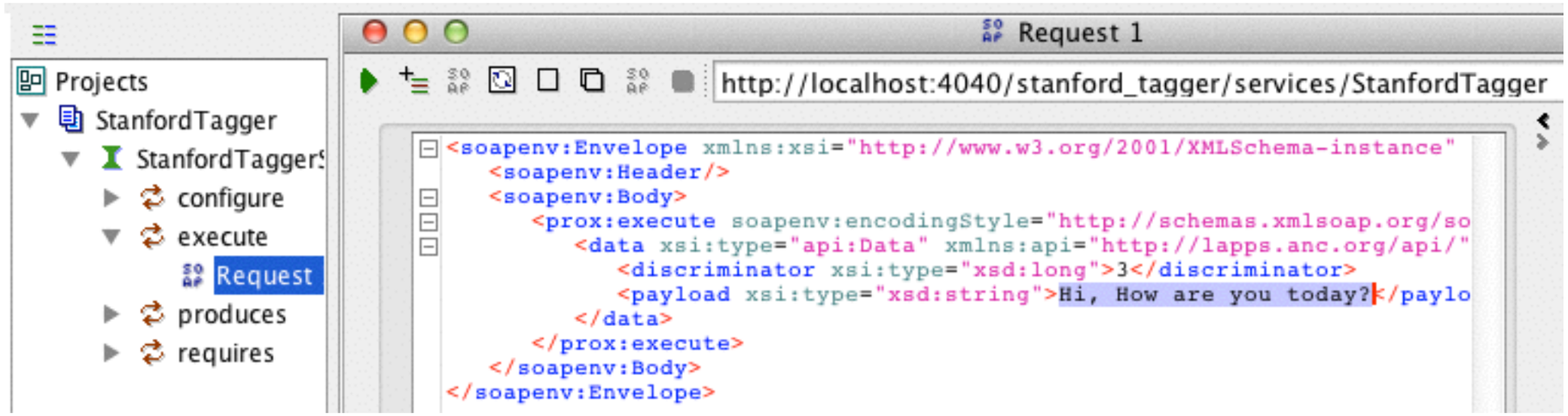
- Local Service

- SoapUI Testing

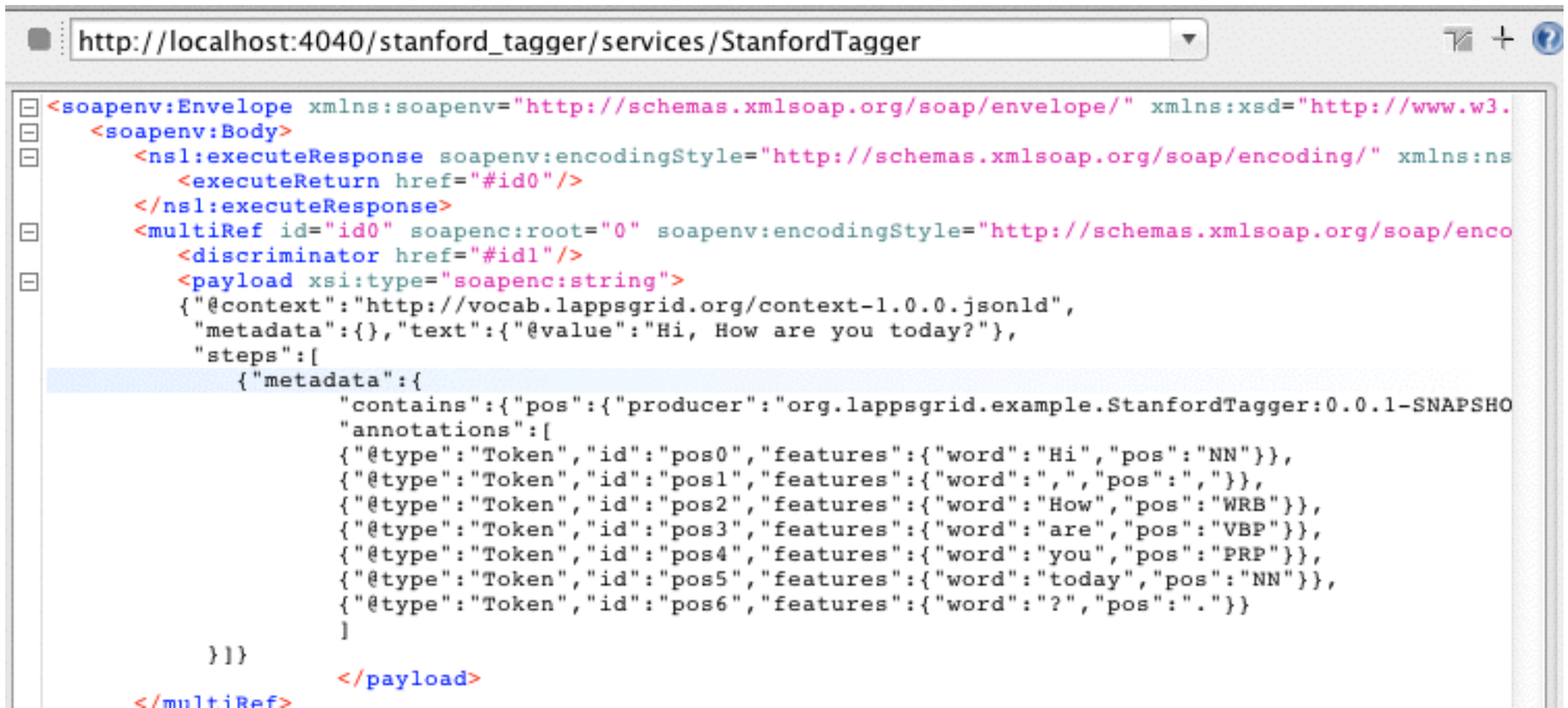


Stanford Tagger Testing Result

- Request

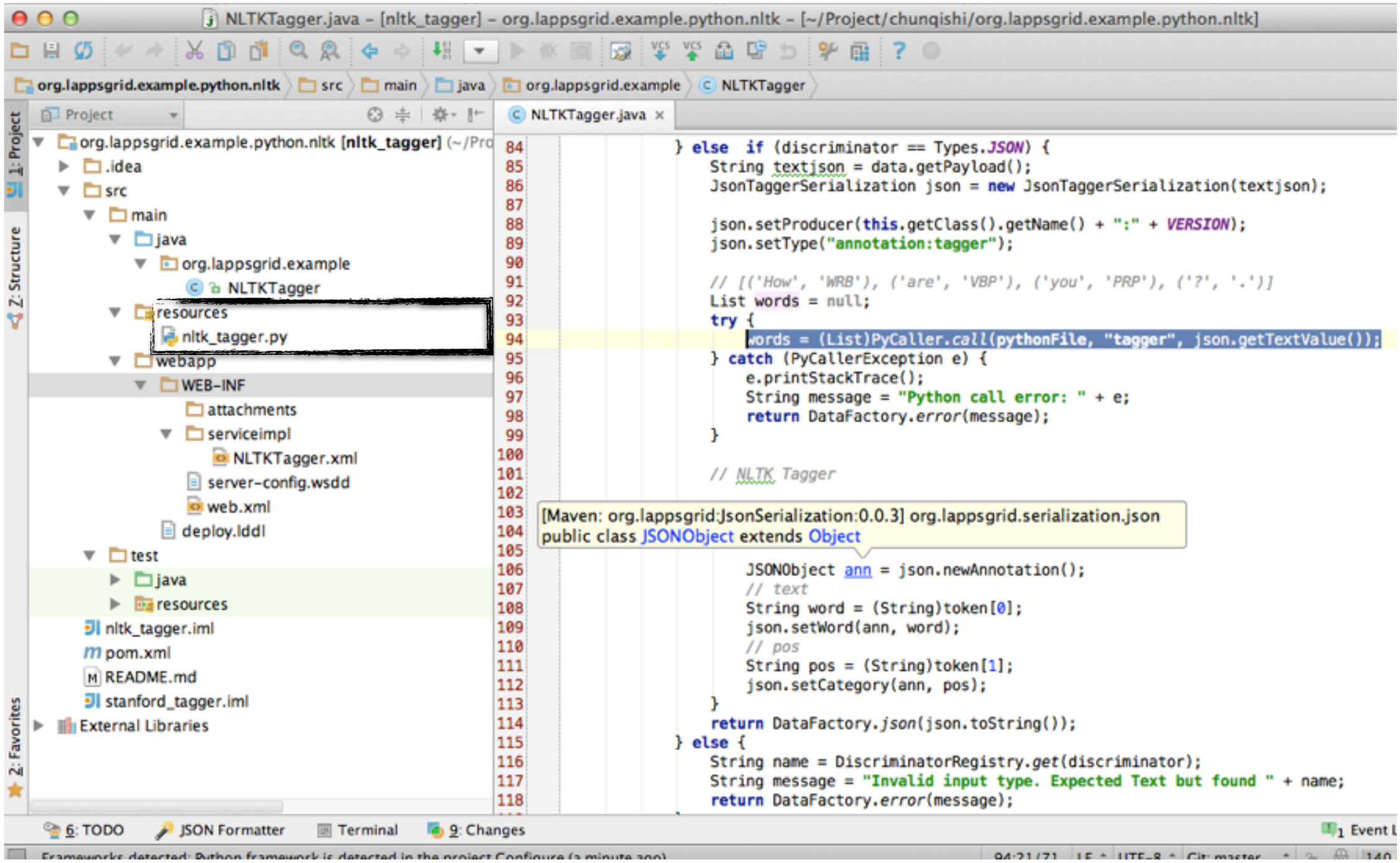


- Response



Lapps Service Wrapping (Python)

Developing Template



NLTK Python

- Python Program

nltk_tagger.py

```
#!/usr/bin/python
import nltk

def tagger(sent):
    text = nltk.word_tokenize(sent)
    return nltk.pos_tag(text)

if __name__ == "__main__":
    import sys
    print tagger(sys.argv[1])
~
```

- Python Result

```
shis-MacBook-Air:resources shi$ python nltk_tagger.py "Hi, how are you today?"
[('Hi', 'NNP'), (',', ','), ('how', 'WRB'), ('are', 'VBP'), ('you', 'PRP'), ('today', 'NN'), ('?', '.')]
shis-MacBook-Air:resources shi$
```

- Java Wrapping

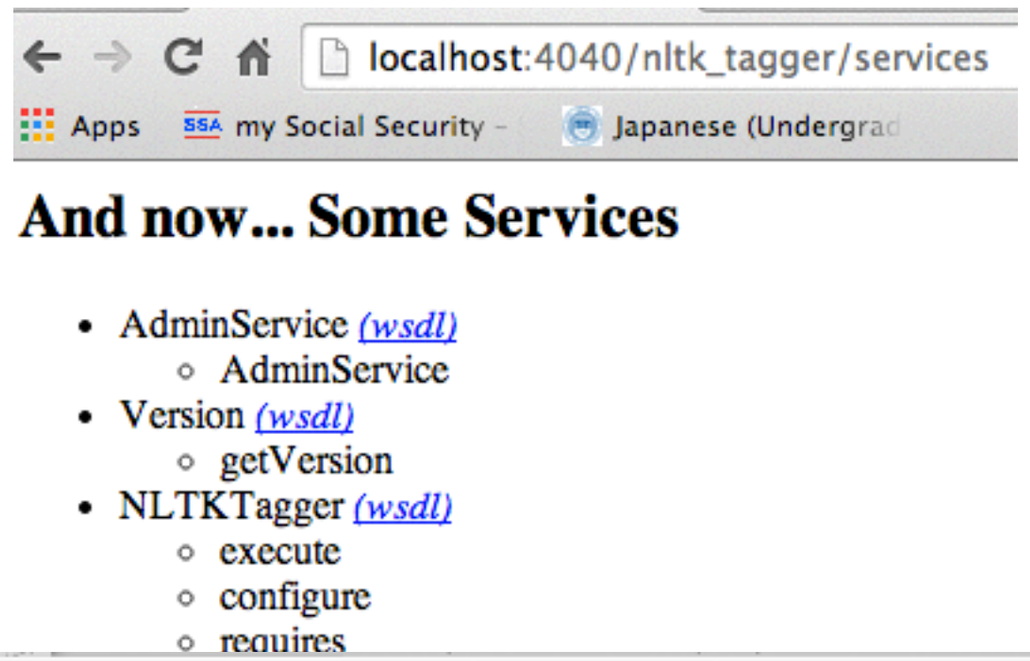
```
// [('Hi', 'NNP'), (',', ','), ('how', 'WRB'), ('are', 'VBP'), ('you', 'PRP'), ('today', 'NN'), ('?', '.')]
List words = null;
try {
    words = (List)PyCaller.call(pythonFile, "tagger", json.getTextValue());
} catch (PyCallerException e) {
    e.printStackTrace();
    String message = "Python call error: " + e;
    return DataFactory.error(message);
}
```

- Jetty Running

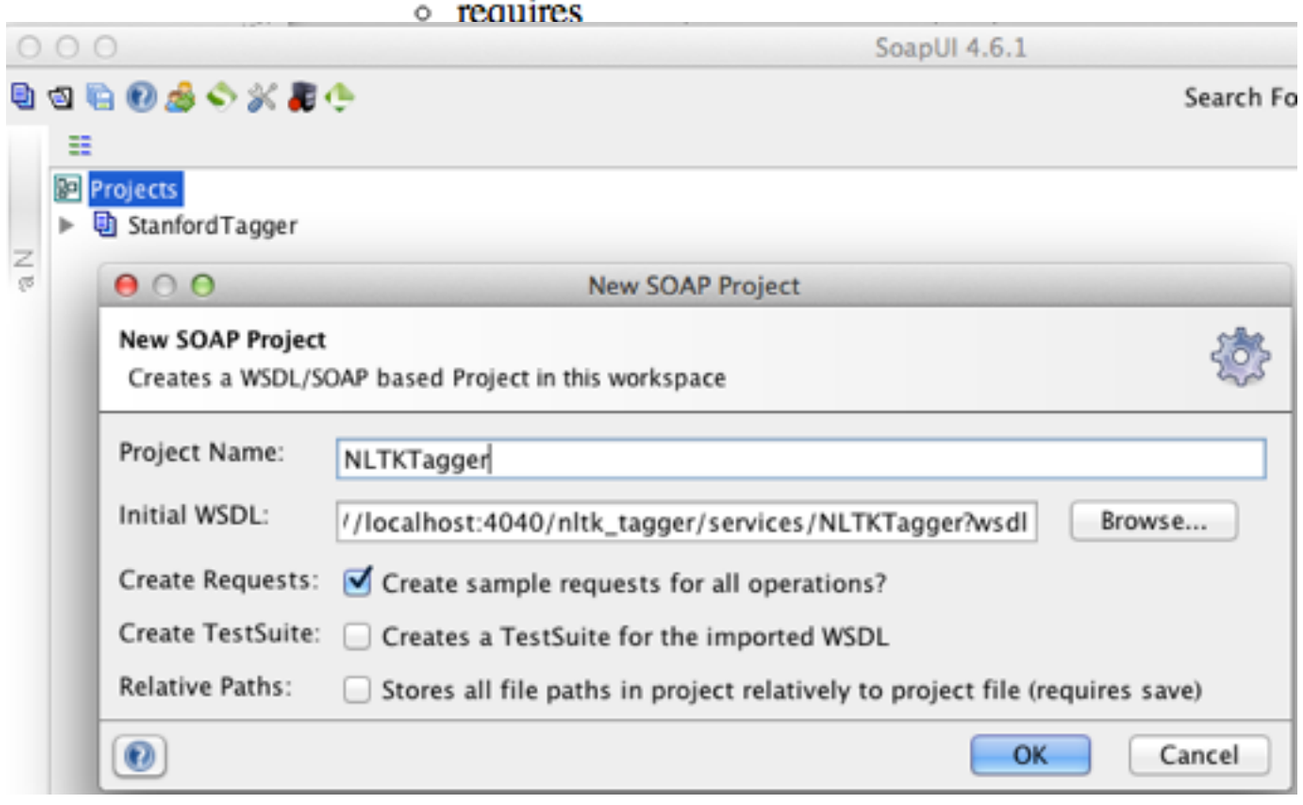
```
shis-MacBook-Air:org.lappsgrid.example.python.nltk shi$
shis-MacBook-Air:org.lappsgrid.example.python.nltk shi$ mvn jetty:run
[INFO] Scanning for projects...
[INFO]
[INFO] ------------------------------------------------------------------------
[INFO] Building NLTK Tagger Example 0.0.1-SNAPSHOT
[INFO] ------------------------------------------------------------------------
```

NLTK Tagger Testing

- Local Service

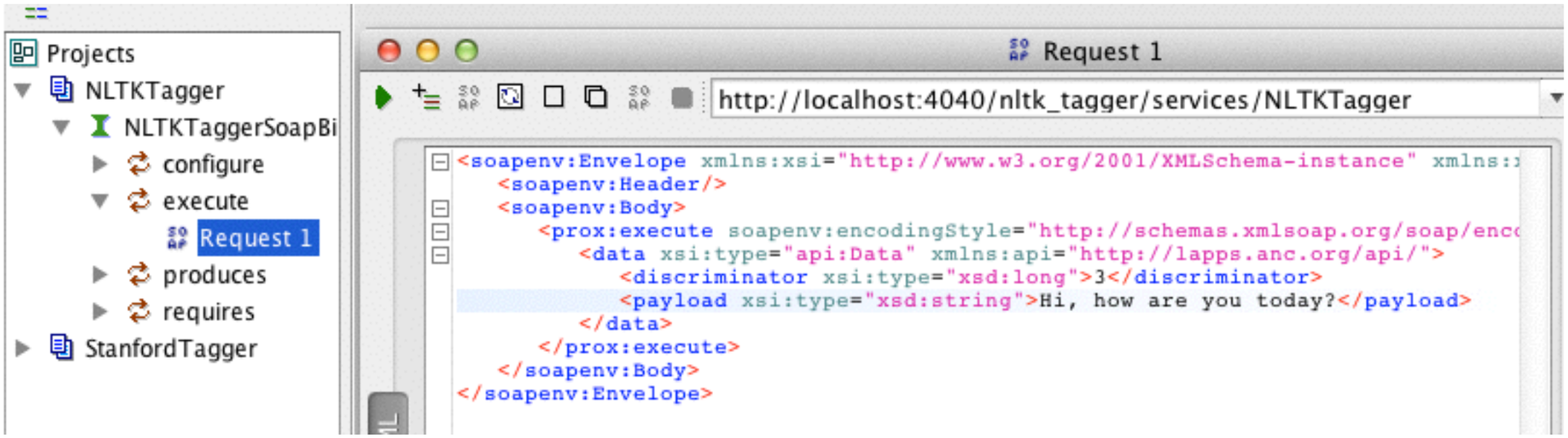


- SoapUI Testing

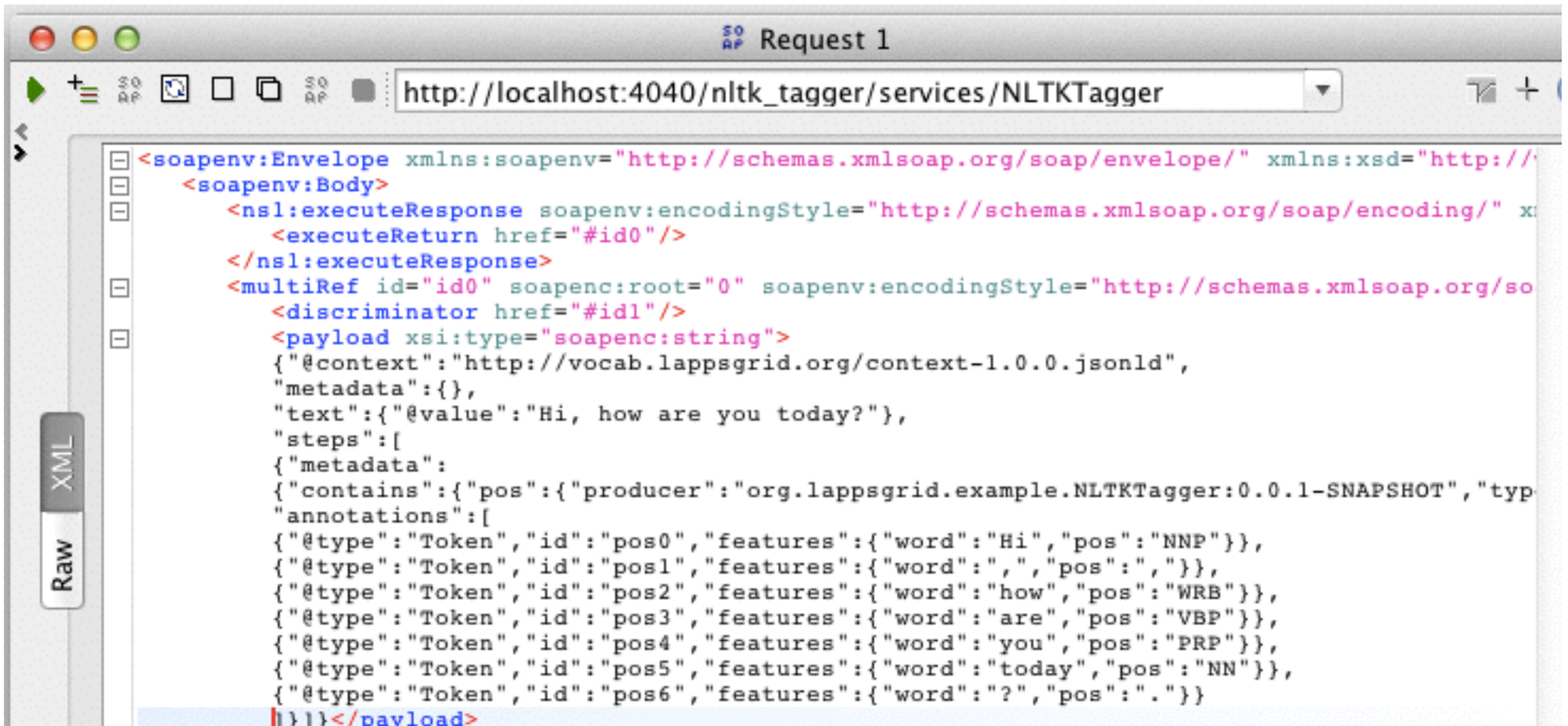


NLTK Tagger Testing Result

- Request

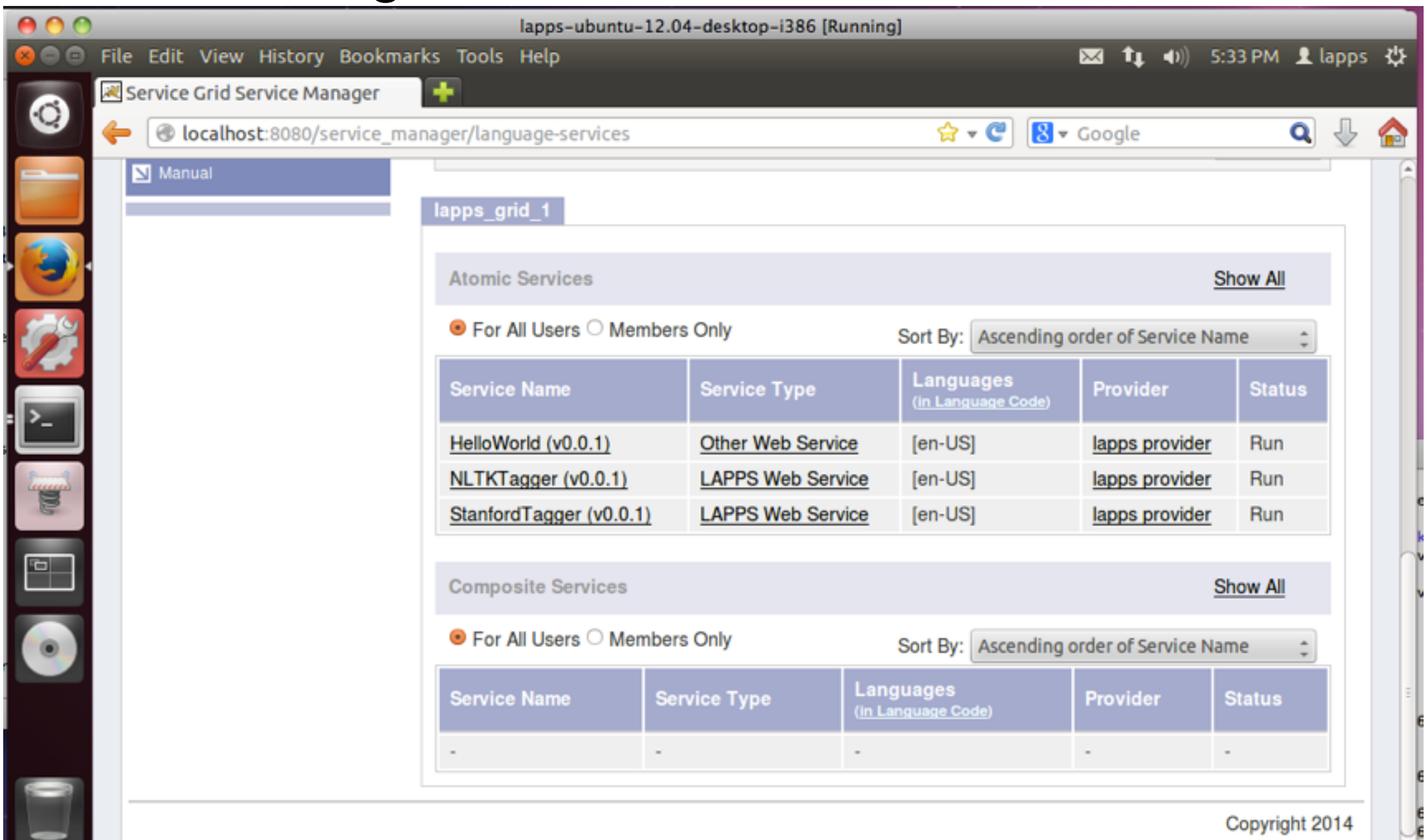


- Response



Service Register

- Service Manager



Conclusion

- Contributing to the Lapps Grid
    - Wrapping Lapps Service
        - Java / Python Wrapping
        - Templates from Github Repository
- Registering into Service Manager
    - Service Manager Installation Script
- Developing Environment
    - VirtualBox Image: Ubuntu

Reference

- API Docs: http://www.anc.org/projects/lapps/api/project-info.html
- Service Templates:
  - https://github.com/chunqishi/org.lappsgrid.example.java.helloworld
  - https://github.com/chunqishi/org.lappsgrid.example.java.stanfordnlp
  - https://github.com/chunqishi/org.lappsgrid.example.python.nltk
- Service Managers:
  - http://eldrad.cs-i.brandeis.edu/service_manager/language-services
  - http://grid.anc.org:8080/service_manager/language-services
- VirtualBox Image:
  - http://eldrad.cs-i.brandeis.edu/download/lapps-ubuntu-12.04-desktop-i386.tar.gz