Ontology Java Library API

Prerequisite

- JavaSE-1.7
- UTF-8 File Encoding

EHowNet

Load EHowNet Library and EHowNet Ontology

- Add `ontologyAcquisition.jar` to classpath
- Get an instance of the Ontology file `ehownet_ontology.txt`

```
EHowNetTree tree = EHowNetTree.getInstance("./docs/ehownet_ontology.txt");
```

Search

- For example, we search for「開心」

```
List<EHowNetNode> results = tree.searchWord("開心");

EHowNetNode node = results.get(0);
```

- If there's no result, an empty List will be returned

Data within a Node

- `node.getNodeType()` : return `NodeType.WORD` or `NodeType.TAXONOMY`
    - Node with type `NodeType.WORD` has no Hyponym, since it is at the bottom of the Ontology
- For word node:
    - `node.getSid()` : return an integer denoting the id of the word, for example `61549`
    - `node.getNodeName()` : return a string denoting the name of the word, for example `開心`
    - `node.getPos()` : return a string denoting the part-of-speech tag of the word, for example `Nv4,VH21`
    - `node.getEhownet()` : return a string denoting the ehownet's definition of the word, for example `{joyful|喜悅}`
- For taxonomy node:
    - `node.getNodeName()` : return a string denoting the name of the taxonomy, for example `物體`
    - `node.getEhownet()` : return a string denoting the ehownet's definition of the word, for example `object|物體`

Hypernym

- `node.getHypernym()` : return an `EHowNetNode` instance, which is the parent of the node. If the node is at the top of the Ontology, the returned value will be `null`

Hyponym

- `node.getHyponymList()` : return a `List<EHowNetNode>` instance, containing all the children of the node. If the node is at the bottom of the Ontology, an empty List will be returned

CKIP Document Converter

Convert a Text File into CKIP-Tagged Document

- Add `ontologyAcquisition.jar` and `jsoup-1.9.2.jar` to classpath
- Set the input/output files and convert

```
Converter.toCKIP("ckip_input.txt", "ckip_output.txt");
```

- We can also convert the documents online: http://sunlight.iis.sinica.edu.tw/uwextract/demo.htm

Ontology Acquisition

Load the Acquisition Tools

- Add `ontologyAcquisition.jar` and `jxl.jar` to classpath
- Initialize and start with root concept, CKIP-documents and EHowNet

```
OntologyAcquisition oa = new OntologyAcquisition("教育", "./docs/ckip", "./docs/ehownet_ontology.txt");
oa.start();
```

Search for a specific concept

- For example, we search for「會議」

```
OntologyNode node = oa.searchConcept("會議");
```

- If the concept does not exist, `null` will be returned

Data within a Node

- `node.getConcept()` : return a string denoting the name of the concept, for example 會議 and 記錄
- `node.getAttr()` : return a `List<String>` instance, containing all the related concept(but not Hypernym or Hyponym) of the node. If the node has no attributes, an empty List will be returned

Hypernym

- `node.getHypernym()` : return an `OntologyNode` instance, which is the parent of the node. If the node is at the top of the Ontology, the returned value will be `null`

Hyponym

- `node.getCategories()` : return a `List<OntologyNode>` instance, containing all the children of the node. If the node is at the bottom of the Ontology, an empty List will be returned

Term/Document Frequency

- `oa.getTermFreq("教育")` : return an integer, which is the term frequency of 教育
- `oa.getDocFreq("教育")` : return an integer, which is the document frequency of 教育

Compile and Run the Sample Project

- `OntologyDemo` is an Eclipse sameple project of EHowNet, CKIP-Converter and Ontology Acquisition
- For Eclipse:
    - `Properties-JavaBuildPath-Libraries` : add all the JAR files in `libs`
    - `Windows-Perferences-General-Workspace` : set the text file encoding to `UTF-8`
- For Shell:
    - `Makefile` is available
        - `OntologyDemo$ make` to compile, `OntologyDemo$ make run` to run
    - Command to Compile and Run

```
OntologyDemo$ javac -d bin -sourcepath src -encoding utf8 -cp libs/jsoup-1.9.2.jar;libs/jxl.jar;libs/ontologyAcquis
ition.jar src/Main.java

OntologyDemo$ java -Dfile.encoding=UTF-8 -cp bin;libs/jsoup-1.9.2.jar;libs/jxl.jar;libs/ontologyAcquisition.jar Mai
n`
```

Reference

- JExcel
- JSoup
- CKIP Service