Erasmus Mundus
master course in
Data Mining and
Knowledge Management
a european master

Ontology and Semantic Web: Laboratory Work

Maíra Machado Ladeira

Lecturer: Rémi LEHN

April 20, 2015

Abstract

Report of the laboratory work for the lecture 'Ontology and Semantic Web' of the master's program DMKM. This report presents the developed activities for the creation and reasoner of an ontology that defines the package system for debian packages. It also gives a guide of how to use the developed scripts.

Contents

1 Introduction

The propose of this lab work is to develop and populate an ontology for debian packages. After developing the ontology, the pellet[1] reasoner is used like described on Sirin et al. (2007) for detecting the relations between the packages. The reasoner is used in the command line, aiming to detect:

- The Debian Community packages,
- The Window Manager packages,
- Conflicts of a package,
- Packages that require a given package,
- Suggestions of a package,
- Recommendations of a package,
- Packages provided by a package
- Conflict packages given a list of packages to install

In order to achieve this features, two scripts were developed using Python[2] technology:

- **ontology_generator.py**: receives a Packages' file and the type of the file to be generated (RDF or n3) to generate the populated ontology on the correct format.
- **reasoner.py**: receives the ontology file to reasoner and the requested reasoner action to generate an output file with the results of the reasoner.

For simplification of the work, in order to avoid the file manipulation tasks, the ontology created on Protégé[3] is also generated by the Python script.

2 Ontology

In the first step, an ontology was created to represent the packages on Protégé. The ontology defines the classes architecture, maintainer, genericPackage, debianPackage, debianCommunity and windowManager. Architecture, maintainer and genericPackage are declared disjoint between each other. The structure of the package can be visualized in Fig. 1 In the second step, the object properties and data

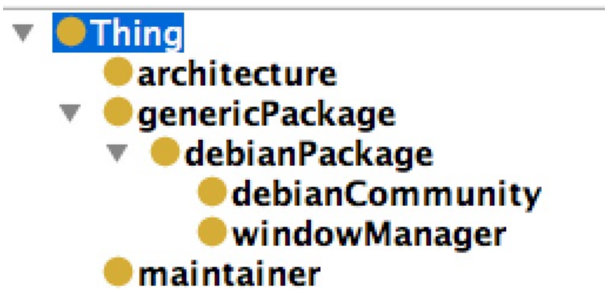

Figure 1: Classes structure for the packages ontology

type properties are set according to the required functionality where a package can have conflicts, dependencies, recommendations and suggestions and can provide another package. The Fig. 2 shows the object and data type properties structure for the ontology. The properties conflicts and depends are treated as transitive and the properties hasArchitecture, hasMaintainer, version and description are treated as functional.
With the help of the ontology created with Protégé, the next step is to develop the python script that generates and populates the ontology, explained on the section 2.1.

[1]https://github.com/Complexible/pellet
[2]https://www.python.org/
[3]http://protege.stanford.edu/

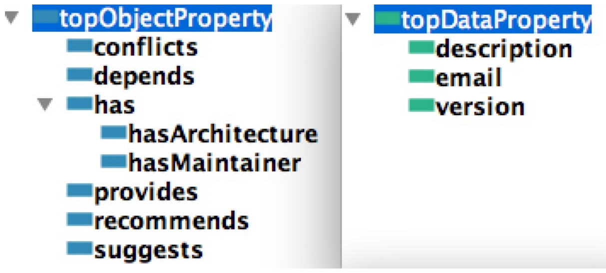

Figure 2: Object and data type properties

2.1 Ontology generator script

The script for generate and populate the ontology is developed in the programming language Python, version 3.4. The script was developed and tested on the operational system OS X Yosemite[4] but should work on Windows and other unix systems as well (without guarantees). The script has two main functionalities: create the package's ontology generated on Protégé in RDF or Notation 3, according to LEHN (2011), and populate this ontology with the debian packages, described on Debian (2014), from a file provided by the user. The program was tested using the packages from the stable distribution of the main section with architecture amd64[5].
The command that automatically executes the script is described below:

```
python ontology_generator.py -p <Packages_file> -t <rdf_or_n3> -o <Output_file>
```

Where $< Packages\_file >$ should be replaced by the file containing the debian's packages downloaded from the previous URL (the file should be on the same folder as the python script), $< rdf\_or\_n3 >$ should be replaced by the strings rdf or n3, representing the desired format of the ontology to be generated and $< Output\_file >$ represents the name of the output file to be generated, **without** extension. The output file should be without extension because the script makes sure that the generated extension is the appropriated one to be used on the pellet reasoner. Example of valid command to use the script:

```
python ontology_generator.py -p Packages -t rdf -o rdf_ontology
```

The program also supports the calling without parameters and in this case, a command line interface guides the user to provide the necessary information during the execution type. This interface is visualized on Fig. 3.

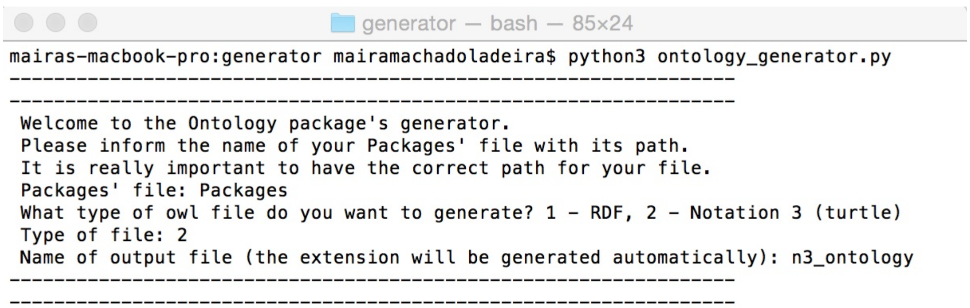

Figure 3: Ontology generator command line interface

The name of the packages, their descriptions, versions and maintainers are changed in order to follow

[4] http://en.wikipedia.org/wiki/OS_X_Yosemite
[5] Available for download on: http://ftp.fr.debian.org/debian/dists/stable/main/amd64/Packages.gz

the requisites of the ontology formats. The list of changes is presented below:

- information between () or ¡¿ are removed.
- / is replaced by _or_
- & is replaced by _and_
- The character + is replaced by _
- the characters ”, ’ and are removed.
- the characters . and space are replaced by _

The results files have the format .owl for RDF/XML ontologies and .ttf for Notation 3 (turtle) ontologies. For simplification reasons, the version of dependencies on the packages are not taken in consideration and only the important fields of the packages descriptions are translated into the ontology.
The execution time of the script, varies between 6 and 10 seconds, depending on the parallel activities performed by the computer. Feedback messages of the performed tasks are given during the whole process of execution of the script.

3 Reasoner

During the reasoner part of the project, the generated ontology is used to infer logical consequences on the packages ontology. This functionality can be very useful for the analysis of the content on the packages. With reasoning it is possible to obtain, for instance, all the packages that are recommended by a package or all the packages that are required by a package. Inconsistencies and unsatisfiabilities on the ontology are also detected by the reasoner which helped on the debugging process of the ontology generator script. Query languages like sparql and N3QL are auxiliary to the reasoner process and allow more structured data manipulation during the reasoning process. For the lab work, the Pellet reasoner in version 2.3.1 and the query language: sparql, according to LEHN (2015) and Sparql (2013) were used. A script that integrates the reasoner in order to automatically generate queries was developed in Python and will be defined on section 3.1

3.1 Reasoner script

A python script was developed in order to automatize the sparql query generation based on actions requested by the user. The script was developed in Python version 3.4 and tested on the operational system OS X Yosemite. As the system access and executes pellet on command line, this script won’t work on Windows operational system, but it should work on other unix systems.
The script can perform the following reasoning:

- get the packages with debian community type
- get packages with window manager type
- get the conflicts of a package
- get the packages that depends on a given package
- get the suggestions of a package
- get the recommendations of a package
- get the packages provided by a package
- get the list of packages that conflict with a given list of packages.

An example query generated for the conflicts detection can be seen on Fig. 4
The results are saved in a file selected by the user. Fig. 5 shows an example result file for the packages that conflicts with a given package.
The script can be executed with the following code:

```
PREFIX url: <http://www.semanticweb.org/ontologies/2015/3/n3_ontology_test.owl#>
SELECT ?conflicts
WHERE {
        url:ndoutils-common a url:debianPackage.
        url:ndoutils-common url:conflicts ?conflicts.
}
```

Figure 4: Query generated by the reasoner script for detect the conflicts of the package: ”fonts-sil-andika”

```
Query Results (1 answers):
conflicts
======================
ndoutils-nagios2-mysql
```

Figure 5: Results file of the reasoner script for the package ”nordugrid-arc-doc”

```
python reasoner.py
```

The interface of the program is visualized on Fig. 6 The script automatically parses the name of the

```
mairas-macbook-pro:generator mairamachadoladeira$ python3 reasoner.py
----------------------------------------------------------------------------
 Welcome to the reasoner! We are using the pellet reasoner, version 2.3.1


 IMPORTANT: This project assume that the namespace of the ontology is the file n
ame


 File to reasoner on (must be an RDF(.owl) or Notation 3(.ttl) file): n3_ontolog
y_test.ttl


 Select the desired reasoner:
 1) Get the packages from the DEBIAN COMMUNITY type
 2) Get the packages from the WINDOW MANAGER type
 3) Get the CONFLICTS of a given package
 4) Get the list of packages that DEPENDS of a given package
 5) Get the SUGGESTIONS of a given package
 6) Get the RECOMMENDATIONS of a given package
 7) Get the packages PROVIDED by a given package
 8) Get the list of the conflicts given a list of packages to install
 Option: 3
 Package to extract the Conflicts from: ndoutils-common
 Output file: conflicts_packages
 Executing command: ./pellet.sh query -v -input-format Turtle -q /Users/mairamac
hadoladeira/PycharmProjects/owlpackagesgenerator/generator/query_temp_conflicts
/Users/mairamachadoladeira/PycharmProjects/owlpackagesgenerator/generator/n3_ont
ology_test.ttl > /Users/mairamachadoladeira/PycharmProjects/owlpackagesgenerator
/generator/conflicts_packages
 Be patient this process takes a while.
----------------------------------------------------------------------------
```

Figure 6: Interface of the reasoner script

packages in order to fit the modifications done by the ontology generator script.
For the running the reasoner, a manual set up of 7G of ram for the java virtual machine of Pellet was performed and this configuration should be changed in computers that do not support it.

4 Conclusion

This project gave me the opportunity to better understand the differences between the formats OWL, RDF and Notation 3. It also helped me to work with a practical use of ontology to describe a real big data.

With this project it was also possible to understand and apply reasoner techniques and query language, showing the importance of logical inference to determine, for instance, the dependency of a dependency package.

The tools developed can certainly be improved in a further work, incorporating for instance, the generation of ontology on the OWL format. However, the current status allows the user to have a good comparison between RDF and Notation 3 on properties like file size and time to reasoner on the ontology.

The reasoner script allows the user to easily get the dependencies, conflicts, suggestions and recommendations of a package, being useful for unix users that install a new debian package, allowing full access to the query and result files.

References

Debian 2014
DEBIAN: *Debian Policy Manual Chapter 7 - Declaring relationships between packages.* `https://www.debian.org/doc/debian-policy/ch-relationships.html`. Version: 2014

LEHN 2011
LEHN, Rémi: *RDF, OWL and Semantic Web.* 2011. – DMKM Slides

LEHN 2015
LEHN, Rémi: *SPARQL.* 2015. – DMKM Slides

Sirin et al. 2007
SIRIN, Evren; PARSIA, Bijan; GRAU, Bernardo C.; KALYANPUR, Aditya; KATZ, Yarden: Pellet: A practical owl-dl reasoner. In: *Web Semantics: science, services and agents on the World Wide Web* 5 (2007), Nr. 2, pages 51–53

Sparql 2013
SPARQL: *SPARQL 1.1 Overview.* `http://www.w3.org/TR/sparql11-overview/`. Version: 2013