

Assignment 2

TU Linked Data project (40 pt)

Instructions

Deadline

Make sure to upload **all** your results

before January 24, 2016!

What you should hand in

Please upload your solution to TUWEL by January 24, 2016. This solution should follow the specified naming convention and include:

1. An archive of all artefacts created (cf. expected results)

`SW_2015_TU_project_<group no>.zip (.tar.gz,..)`

2. Your documentation in a pdf document

`SW_2015_TU_project_<group no>.pdf`

Please make sure to upload only a single solution per group. Furthermore, make sure to list your group members in the cover page of your documentation.

Review meetings

We will schedule review meetings at the end of January (registration via TUWEL, at least two alternative dates will be offered).

Questions

Please post general questions in the TUWEL discussion forum of the course. You can also discuss problems and issues you are facing there. We appreciate if you help other students out with general problems or questions regarding the tools used (and may take that into account in case you are short a few points for a better grade). For obvious reasons, however, please do not post any solutions there.

You can also contact me directly (with specific questions) at:

`elmar.kiesling@tuwien.ac.at`

Please use subject line `SW_2015_<your subject>` to minimize the probability that your Email gets lost in my inbox.

Assignment 2

TU Linked Data project (40 pt)

Introduction

In this project assignment, you will obtain data provided by Vienna University of Technology's TISS system, convert it into Linked Data, and create a simple program that programmatically accesses the data to do something interesting with it.

In particular, you will

- Create a domain ontology
- Extract Linked Data from the Web
- Create RDF from structured data
- Store RDF data in a triple store
- Programmatically access the data in an application.

In implementing your project, you will follow an iterative process that involves the following tasks:

1. Create an OWL ontology that models the University domain
2. Extract data from TISS
3. RDFize the extracted data using the vocabulary specified in your ontology
4. Create an RDF graph and load it into a triple store
5. Implement a simple program that does something interesting with the data
6. Suggest an application that could be built on "LinkedTUV" data

You can earn bonus points if you reconcile and link your data externally (e.g., room address with LinkedGeoData, course topics with DBpedia etc.).

Assignment 2

TU Linked Data project (40 pt)

Tasks description

Task 1: Create an OWL ontology that models the University domain

Create an OWL ontology that represents the University domain, including high-level concepts such as *Person*, *Organizational Unit*, *Course*, *Room*, *Project*, as well as both more general and more detailed concepts necessary to represent the domain in later steps of the project. Also model the basic relationships between these concepts. Look at the data available in TISS to inform your modeling process and expect that you may need to revise your ontology once you are in the process of converting the data into rdf later on in the project.

It is not necessary to model all relationships and use complex class constructors to facilitate sophisticated reasoning, but the basic relations necessary to represent the data in a connected graph later on in the project should exist. You can follow the not quite recent, but still very good tutorial on the general process of building an ontology [1]. Also, make sure to reuse existing vocabularies wherever fitting and possible. You can use general Ontology and vocabulary repositories [2][3] and find a list of University-related vocabularies provided by the Linked Universities project [4].

Tools

Protégé: <http://protege.stanford.edu/>

Resources

- [1] http://protege.stanford.edu/publications/ontology_development/ontology101.pdf
- [2] <http://lov.okfn.org/>
- [3] http://www.w3.org/wiki/Ontology_repositories
- [4] <http://linkeduniversities.org/lu/index.php/vocabularies/index.html>

Expected result:

An OWL file of your self-descriptive ontology, including labels and comments.

Task 2: Extract TISS data

You can use some of the following starting points for your data acquisition:

- **Courses** (TISS > Education > Extended search > Faculty ..)
- **Persons** (TISS > Organisation > Address book > Faculties > Institute ..)
- **Rooms** (TISS > Education > Rooms)
- **Org Units** (TISS > Organisation > Address book > Faculties)
- **Projects** (TISS > Research > Research portal > Projects > Faculty)

To extract the data from the web pages, XML versions of the pages or json-API responses, you can use various tools, including:

- Various XPath and XSLT transformation tools for the XML pages (things like the Chrome scrape plugin may come in handy)
- OpenRefine (highly recommended)
- Bulk download of files using URIs that follow a particular naming scheme (e.g., with curl, wget,..)

Assignment 2

TU Linked Data project (40 pt)

- Unfortunately, the API for individual course details, which should be available at <http://tiss.tuwien.ac.at/rest/course/orgUnit/:code> is currently down. The TISS team is working on a fix.

Tools

- OpenRefine
<http://openrefine.org/>
- (wget, xpath, xslt, Chrome Scrape plugin)...

Resources:

- TISS API documentation
- OpenRefine Documentation for Users
<https://github.com/OpenRefine/OpenRefine/wiki/Documentation-For-Users>
- Many OpenRefine tutorials, videos.. etc. available
- Verborgh, R., De Wile, M.: Using OpenRefine, Packt publishing 2013
- Schoolofdata: Scraping multiple Pages using the Scraper Extension and Refine
<http://schoolofdata.org/handbook/recipes/scraping-multiple-pages-with-refine-and-scraper/#sthash.SNHymzwE.dpuf>
- Linked Universities: Tools
<http://linkeduniversities.org/lu/index.php/tools/index.html>

Expected results

- csv files of the extracted data, which should include data on at least three high-level concepts such as Person, Course, Room, Organizational unit, Project etc.
- You don't need to extract all the data available, just a subset that can serve as a proof of concept (e.g., the Persons of a single faculty, the Courses of one of the departments, the rooms of one building etc.)
- Brief documentation of how you did it
- Scripts, XPath expressions, XSLT, GoogleRefine code etc.

Task 3: RDFize your extracted data

The next step consists in lifting your raw data that you obtained and prepared in OpenRefine to RDF, using the concepts and properties you defined in your ontology in Task 1. To this end, you can use the RDFRefine [1] extension to do this step within OpenRefine or export the data in csv format and use mapping tools and languages such as RML [2] to convert it into rdf. Please follow the W3C best practices for publishing linked data [3] where appropriate (e.g., good URIs, standard vocabularies,..).

Tools

- [1] RDFRefine extension: <http://refine.deri.ie>
[2] RML <http://semweb.mmlab.be/rml/spec.html>

Assignment 2

TU Linked Data project (40 pt)

Resources:

[3] W3 Best Practices for Publishing Linked Data <http://www.w3.org/TR/ld-bp/>

Expected result

- RDF file in turtle notation that contains your graph and that can be loaded into a triple store.

Task 4: Load your RDF graph(s) into a triple store

In this step, you should load your graph into a local triple store such as Jena TDB.

Expected result

- Documentation to replicate your setup (which triple store you use, how you organize your data into named graphs – if so - etc.)

Task 5: Implement a simple application

In this final part of the implementation of your project, you should create a simple program that accesses your data programmatically and presents it to the user. It is sufficient to create a simple (e.g., CLI) application that does something simple, but it should make use of the semantics of your data somehow. You can earn bonus points if you create a fancy UI (web app, native app, mobile app).

Expected result

- Simple (e.g., CLI) Java program
- May use Jena Core and/or SPARQL API or Sesame

Task 6: Suggest a TU Linked Data Application

Finally, provide a suggestion for a semantic application that could be built on top of a complete (and possibly extended and externally interlinked) version of a TU Linked Data set. This could be a web or mobile app or some service created on top.

Expected result:

A description of your idea in prose. If applicable, also include an image of a UI mockup of your idea.

Assignment 2 TU Linked Data project (40 pt)

Appendix

TISS Public APIs (<https://tiss.tuwien.ac.at/api/dokumentation>)

Adressbuch

Administration & Services Orgeinheiten der TU Wien

`/adressbuch/adressbuch/administration(/:id(.:format))`

Detailldaten einer Fakultät auf Basis der TISS-ID

`/adressbuch/adressbuch/fakultaet(/:id(.:format))`

Detailldaten einer Fakultät auf Basis des Whitepages-Codes

`/adressbuch/adressbuch/fakultaet_via_code(/:id(.:format))`

Detailldaten einer Orgeinheit auf Basis der TISS-ID

`/adressbuch/adressbuch/orgeinheit(/:id(.:format))`

Detailldaten einer Orgeinheit auf Basis des Whitepages-Kurzzeichens

`/adressbuch/adressbuch/orgeinheit_via_code(/:id(.:format))`

Detailldaten einer Orgeinheit auf Basis der Orgeinheiten-Nummer

`/adressbuch/adressbuch/orgeinheit_via_nummer(/:id(.:format))`

Detailldaten einer Orgeinheit auf Basis der OID

`/adressbuch/adressbuch/orgeinheit_via_oid(/:id(.:format))`

Detailldaten einer Person auf Basis der TISS-ID

`/adressbuch/adressbuch/person(/:id(.:format))`

Detailldaten eines Studenten auf Basis der Matrikelnummer

`/adressbuch/adressbuch/person_via_matrikelnummer(/:id(.:format))`

Detailldaten einer Person auf Basis der OID

`/adressbuch/adressbuch/person_via_oid(/:id(.:format))`

Detailldaten einer Orgeinheit auf Basis des Whitepages-Kurzzeichens

`/api/org_unit/code/:id`

Detailldaten einer Orgeinheit auf Basis der TISS-ID

`/api/org_unit/id/:id`

Detailldaten einer Orgeinheit auf Basis der Orgeinheiten-Nummer

`/api/org_unit/number/:id`

Detailldaten einer Orgeinheit auf Basis der OID

`/api/org_unit/oid/:id`

Detailldaten einer Person auf Basis der TISS-ID

`/api/person/id/:id`

Detailldaten eines Studenten auf Basis der Matrikelnummer

`/api/person/mnr/:id`

Assignment 2 TU Linked Data project (40 pt)

Detailldaten einer Person auf Basis der OID
`/api/person/oid/:id`

Detailldaten einer Person auf Basis der TISS-ID
`/person/:id(.:format)`