

# Open Information Systems - Project Specification

## 2018-2019

Jan Hidders, Jonathan Riggio, VUB WISE

### 1 Introduction

The goal of the Open Information Systems project is to design and implement an ontology-based Web application. The assignment will be carried out in teams of 4 people. Additionally, a report describing the design and choices will have to be produced.

### 2 Topic

The topic of this year is:

- **A Web Application for Medical Diagnostics.**

We ask you to implement a web application where a user can register and provide personal information about their health condition. According to the symptoms provided by the user, the system will provide a list of corresponding diseases. In the context of this project, a symptom is defined as a physical or mental feature which is regarded as indicating a condition or a disease. In order to distinguish between different classes of diseases, the symptoms are hierarchically classified. To begin, we ask you to consider the following classes (and subclasses) of symptoms:

- muscular symptom (motor weakness, muscular weakness, muscular cramp)
- abdominal symptom (abdominal rigidity, abdominal mass disorder)
- cardiovascular symptom (arrhythmia, shock)
- skin tissue symptom (edema, lesion, sweat, rash)
- neurological symptom (fatigue, confusion, anxiety, dizziness)
- digestive system disorder (vomiting, gastroenteritis)
- reproductive system symptom (prostate dysfunction, menorrhagia)
- nutritional disorder (weight loss or gain, alteration of appetite)

Some classes of symptoms are related to other classes such as one being a *common consequence* of another one. As examples of this, consider the following relations:

- a nutritional disorder is often a consequence of a digestive system disorder
- neurological symptoms are often the consequence of an instance of one of the other classes of symptoms
- a muscular symptom such as muscular cramp is sometimes a consequence of an abdominal symptom such as abdominal rigidity

Additionally, a user must be able to indicate the degree of severity of a symptom. For example, it might be that they suffer from extreme fatigue but relatively low muscular weakness. It is up to you to find the best way to represent precisely the level of severity for a specific symptom.

Finally, for each disease, the application should be able to provide its name, a description, and a list of associated symptoms.

## 3 Organization

### 3.1 Groups

Students will be assigned to a team of four members. Information about team members can be found on Canvas. We expect a fair partition of the workload between the members of the group. It will not be accepted that a student lets the team work for them without being involved in the project. Every problem with a team member should be reported as soon as possible.

### 3.2 Conceptual schema

By **November 9th** you should provide the conceptual schema of your database modeled using the Entity-Relationship Model. We ask you to use the ER notation provided on the VUB's database course (link [here](#)). At this stage, the conceptual schema should be complete enough to capture every object that needs to be stored by your application. You should therefore provide use cases to describe the usage of the system and verify that the conceptual schema indeed covers all required objects. You should pay attention to the completeness and correctness of the model, and specifically the following aspects:

1. Weak entities should be correctly identified as such.
2. Upper and lower bounds for cardinalities should be correctly identified.
3. All entity types should have at least one natural key, i.e., a key that consists of one or more attributes that have meaning for the users (so not an artificial ID attribute that has no meaning for the users).

### 3.3 Ontologies and rules

We ask you to design the ontologies and rules of your system. You will need to proceed in three steps:

1. Design your own ontologies and provide them as OWL files.

2. Adapt your conceptual schema and/or ontologies in order to ensure consistency between your ontologies and database schema.
3. Design rules that will allow you to infer relevant information from your data.

In order for your application to use real-life data, we require that your system takes into account the following standard ontologies:

- [Human Disease Ontology](#), an ontology for describing the classification of human diseases organized by etiology.
- [Symptom Ontology](#), an ontology of disease symptoms.

You will therefore have to modify your conceptual schema and ontologies in order to ensure consistency with the standard ontologies.

The deliverable of this phase is a small report describing the ontologies including a [WebVOWL](#) visualization of your ontologies and a description of the rules you designed as well as a justification of why those rules are relevant in your system. The deadline for this deliverable is 23rd November.

## 4 Intermediate presentation

The intermediate presentation will take place on the **19th November from 9.00 to 13.00**. This presentation will be the occasion for your group to present your application. Each presentation should last **maximum 12 minutes**. During that time you should be able to

- Give a small overview of the functionalities offered by your system.
- Present and explain the design choices you made for your ontologies.
- Present one scenario showing how your system could be used through the SPARQL endpoint in order to access information. Show which part of the conceptual schema and the ontologies are involved. It is sufficient to present a scenario, no demonstration of a fully working SPARQL endpoint is required at this stage.

## 5 SPARQL endpoint

You will have to provide a SPARQL endpoint accessible within the application that allows external users to query your system. You can either do it using *Ontop's Protégé Plugin* or any other libraries of your choice, as long as you justified its usage. Your final report should detail and explain the mapping between your relational database and the ontologies. During the final presentation, you will have to execute relevant queries on your SPARQL endpoint in order to demonstrate how you used ontologies and rules in order to infer useful information from your information system. We expect your system to be able to at least return the corresponding list of diseases for a list of symptoms.

## 6 Additional functionalities

In addition to the required basic functionality, the team is also asked to design and build additional functionality and use additional ontologies when relevant. For example, the diseases and conditions might be sorted in probability, medication might be advised, possible causes of the disease might be indicated, et cetera.

## 7 Final report

The final report should at least go through the following points:

1. How did you divide and manage the workload inside your team.
2. Motivate the design of your conceptual schema for the database.
3. Motivate the design of your ontology.
4. Explain how you modified your conceptual schema and ontologies in order to adapt your system to the ontology standards and ensure consistency.
5. What are the rules to infer additional information from your database? Why are they relevant for your information system?
6. Present and explain the final ontologies, rules and specification of the mapping including the extensions that were required by the additional functionality.

Don't forget to provide the number of your team, the names, email addresses and enrollment numbers of all members of your team as well as the current academic year in the title page.

The deadlines for the final report is **21st December**.

## 8 Final presentation

The final presentation is scheduled on the **17th December** and will be a private defense of your project. We expect you to make good use of the feedback you received from the intermediate presentation. The final application should be complete enough to illustrate different use case scenarios of your system, allowing you to justify the different design choices for ontologies and rules as well as demonstrate the database mapping. Don't forget to include a working SPARQL endpoint in your application.

## 9 Administrative Requirements

Every deliverable has to be sent by mail to the assistant `jonathan.riggio@vub.be` with the following subject:

- *OIS-2018-Subject-student1-student2-student3-student4*

Here *student1* is the last name of the first team member and *Subject* is the subject of the deliverable (see Section 10). Mails that are not sent according to this header will be ignored! Make sure to send only one mail per team. For each deliverable and report, we expect a report worthy of a MSc student. Reports should be well structured (title, abstract, introduction – middle – conclusion, references, etc.). Plagiarism is not tolerated and we expect that resources are properly cited/attributed.

## 10 Deadlines Recap

The following gives an overview of all the relevant deliverables and dates within the project:

- **November 9th 23:59** - Conceptual schema (by email with subject *OIS-2018-SCHEMA-student1-student2-student3-student4*)
- **November 19th** - Intermediate presentation (public)
- **November 23rd 23:59** - Ontologies and rules (by email with subject *OIS-2018-ONTO-student1-student2-student3-student4*)
- **December 17th** - Final presentation (private, details on organization will be communicated during the semester)
- **December 21st 23:59** - Final Report (by email with subject *OIS-2018-REPORT-student1-student2-student3-student4*)

## 11 Contact Information

If you have any questions, please feel free to contact the assistant `jonathan.riggio@vub.be` or the teacher `jan.hidders@vub.be`.