Ontology engineering
- The basic process

Reminder: What is an Ontology?

- An ontology is an artifact that can be used as a part of a software system and/or published together with some data on the Web

- It is a **model** of something – whether the real world or some imaginary world it is **never** an exact reproduction of it

- In other words it is:
  - a descriptive specification of a set of contextual assumptions about a domain of interest

- While it is usually not:
  - a prescriptive specification of the inner structure of 'true reality'

Reminder: Why develop an ontology?

- To develop a shared common understanding of the structure and meaning of information
    - among people
    - among software agents
    - between people and software
- To enable standardisation and/or reuse of domain knowledge
    - to avoid "re-inventing the wheel"
    - to introduce standards to allow interoperability
- To make domain assumptions explicit
    - easier to change domain assumptions
    - easier to understand and update legacy data
- To separate domain knowledge from operational knowledge (i.e. separate data from operations in a system)
    - re-use domain and operational knowledge separately (e.g., configuration based on constraints)
- …

What is Ontology Engineering?

- Ontologies are artifacts

  - Have a structure (linguistic and logical)

  - Their function is to "encode" a description of the world (actual, possible, counterfactual, impossible, desired, etc.) for some **purpose**

- Ontologies must match both **domain** and **task**

  - Allow the description of the entities ("domain") and their attributes and relations, *e.g. cars and their characteristics*

  - Serve a purpose ("task"), e.g. *finding cars that match some customer criteria*

- Ontologies have a lifecycle

  - Created, evaluated, fixed, and exploited just like any artifact, e.g., like software

  - Their lifecycle has some special characteristics regarding:

    - *Data, processes, argumentation, design patterns…*

Two main kinds of ontologies

- Coverage-oriented ontologies

  – They cover the terminology/metadata/textual corpora/folksonomies … that fit a specific domain

- Task-oriented ontologies

  – They are able to give a structure to a knowledge base that can be used to answer competency questions and/or do reasoning

What is needed for designing ontologies

- Resources ("raw" material) - from domain experts
  - *Reengineering* is key, e.g. from Thesauri, lexicon, tags, code lists, DB schemas etc.
  - Texts, interview transcripts etc.
- Formal languages, e.g. RDF(S) and OWL
- Solutions (target configurations for the raw material)
  - Reusable/standard ontologies and design patterns
- Methods (production from raw material)
  - *Collaboration workflows*, search, evaluation, selection, reengineering procedures, pattern matching, modelling and composition
- Tools
  - Ontology engineering tools (TopBraid Composer, Protégé, WebProtége ...)
  - Ontology evaluation and debugging tools (OOPS! http://oops.linkeddata.es/index.jsp & more on Friday)
  - Management and versioning (github, w3id, Ontoology, ...)
  - API:s and frameworks for using and applying ontologies (Jena, OWL API, various triple stores...)

The overall process

- Project scoping and initialisation
  - Figure out what to do, when, how and why and with what resources
- Project realisation
  - Elicit requirements, formalise the ontology, evaluate and test
- Deployment and maintenance
  - Apply the ontology to your data, and/or in your software system, maintain the ontology over time

LINKÖPING UNIVERSITY

The overall process

- Project scoping and initialisation
  - Figure out what to do, when, how and why and with what resources
- **Project realisation**
  - **Elicit requirements, formalise the ontology, evaluate and test**
- Deployment and maintenance
  - Apply the ontology to your data, and/or in your software system, maintain the ontology over time

LINKÖPING UNIVERSITY

Requirements engineering

What are "requirements"?

- Viewing an ontology as a black box… what should that box provide?

- Functional requirements
  - Query results?
  - Inferences?
  - Error checking?
  - …

  Internal structure, and content

- Non-functional requirements
  - Coverage
  - Efficiency
  - Documentation
  - Changeability – extendibility
  - …

  Overall structure, acceptance -> Guidelines and rules for development

Non-functional Requirements

- Coverage
  - How important is the coverage of the domain? How will the ontology be updated?
- Efficiency
  - What OWL profile to use?
  - Reasoning off-line or online?
  - Query re-writing or reasoner, e.g. not requiring inferences
- Documentation
  - Labels and comments?
  - Naming conventions
- Changeability – extendibility
  - Should future extensions be prepared for?
  - Alignment to online ontologies, standards?

LINKÖPING UNIVERSITY

Functional requirements: Competency Questions (CQs)

- What do we want to ask the knowledge base/knowledge graph/dataset?
  - Typical questions/queries

- Can be used as requirements and as the basis for "unit tests"

- Example
  - Example fact to be represented:
    "Anders works at SAAB training systems"
  - Generalised knowledge, 'instance-free sentence':
    "People work at companies"
  - Potential competency questions:
    - For what company does a certain person work?
    - What persons work at a certain company?

Functional requirements: Competency Questions (CQs)

- Requirements of an ontology = CQs + additional constraints/restrictions and reasoning requirements
  - Additional restrictions (axioms) to be defined on the model, usually restrictions over data
  - Reasoning requirements specify the facts that have to be inferred before they can be retrieved through a query, i.e. they are not explicitly stored in/input to the knowledge base

- Example constraints/restrictions
  - Each flower shop sell at least 2 kinds of flowers.
  - Every flower shop sells some roses.
- Example reasoning requirement
  - The class of "Popular flowers" is the flowers that have been sold more than twice the past week

Trade-off: Software vs. Ontology

- What functionality is going to be put into the software and what is going to be part of the ontology?
  - An OWL reasoner is nothing more than general-purpose code for processing data – why not use specific code in our system instead?
- Ontology pro:s
  - The ontology makes assumptions explicit
    - Ontologies can be published and shared together with the data
  - The ontology can be changed at runtime without changing the code (or with minimal changes)
  - The reasoning procedures are sound and well-defined, and they are reused for all inferences
- Software pro:s
  - More efficient?
  - Expressivity

Hands-on: CQs

- **Context**

  - Tourist information about cities: A company providing a tourism web portal would like to semantically represent all their information about cities in Europe. Following is a typical description of the kind of information that is present in the portal, and that should be contained in (and queried from) the knowledge base.

- **Example of information that the system should manage**

  - Rome is the capital of Italy, it is located in the Lazio region. Rome has two airports. Fiumincino airport is served by Alitalia flights, while Ciampino is served by Ryanair and Wizzair. Rome has several train stations, the main station is Termini located in the center of Rome, but there are also the Trastevere station in the west part of Rome (the Trastevere district), and Tiburtina in the south east. When getting around in Rome by public transport there are buses, trams and a metro with two lines (A and B). The metro lines A and B run until 00.30, and resume at 5.30 in the morning. In Rome you can stay in a hotel. Hotels have different ratings (from five to one star), and rates within a price range. Price ranges of the hotels differ depending on the time of the year that you decide to visit Rome.

    Rome is full of sights, such as churches, monuments and museums. Popular churches are St Peters and the Pantheon, and a monument is the Colosseum. To enter the Colosseum there is a fee of 12 EUR for adults, a fee of 7,50 EUR for EU citizens below the age of 24 and free entrance for EU citizens below 18 years old. There are organised tours to see the sights of Rome, either on foot starting from the Pantheon or by bus leaving from the Termini station, with different starting times on different weekdays. Rome also has numerous museums. Museums are generally closed on Mondays, but the Vatican Museums are instead closed on Sundays.

    While in Rome a visitor should take the opportunity to visit one of the many events taking place there during different periods, such as theater and opera performances, and concerts. During September and October 2009 the Teatro Roma offers two theater shows each evening. The play in this period is the "Merchant of Venice", written by Shakespeare in 1598, given through an ensemble of well-known actors. During the second and third week of September the role of Shylock is played by Arnold Schwarzenegger as a special guest actor.

    There are a number of travel guides available for Rome, many of them are available in several formats, for example as a guidebook or on a CD. Maps of Rome can also be found in paper format or in an electronic format.

- **Problem divided into sub-parts, "requirement stories"**
  - Rome is the capital of Italy, it is located in the Lazio region.

  - Rome has two airports. Fiumincino airport is served by Alitalia flights, while Ciampino is served by Ryanair and Wizzair. Rome has several train stations, the main station is Termini located in the center of Rome, but there are also the Trastevere station in the west part of Rome (the Trastevere district), and Tiburtina in the south east.

  - When getting around in Rome by public transport there are buses, trams and a metro with two lines (A and B). The metro lines A and B run until 00.30, and resume at 5.30 in the morning.

  - In Rome you can stay in a hotel. Hotels have different ratings (from five to one star), and rates within a price range. Price ranges of the hotels differ depending on the time of the year that you decide to visit Rome.

  - Rome is full of sights, such as churches, monuments and museums. Popular churches are St Peters and the Pantheon, and a monument is the Colosseum. To enter the Colosseum there is a fee of 12 EUR for adults, a fee of 7,50 EUR for EU citizens below the age of 24 and free entrance for EU citizens below 18 years old.

  - There are organised tours to see the sights of Rome, either on foot starting from the Pantheon or by bus leaving from the Termini station, with different starting times on different weekdays.

  - Rome also has numerous museums. Museums are generally closed on Mondays, but the Vatican Museums are instead closed on Sundays.

  - While in Rome a visitor should take the opportunity to visit one of the many events taking place there during different periods, such as theater and opera performances, and concerts. During September and October 2009 the Teatro Roma offers two theater shows each evening.

  - The play in this period is the "Merchant of Venice", written by Shakespeare in 1598, given through an ensemble of well-known actors. During the second and third week of September the role of Shylock is played by Arnold Schwarzenegger as a special guest actor.

  - There are a number of travel guides available for Rome, many of them are available in several formats, for example as a guidebook or on a CD. Maps of Rome can also be found in paper format or in an electronic format.

LINKÖPING UNIVERSITY

Possible solutions

- Rome is the capital of Italy, it is located in the Lazio region.

  - A city is the capital of a country. A city is located in a region.

    - CQs: What city is the capital of a certain country? In what region is a certain city located? In what country is a certain city or region located?

- Rome has two airports. Fiumincino airport is served by Alitalia flights, while Ciampino is served by Ryanair and Wizzair. Rome has several train stations, the main station is Termini located in the center of Rome, but there are also the Trastevere station in the west part of Rome (the Trastevere district), and Tiburtina in the south east.

  - A city has a number of airports. An airport is served by a number of airlines. A city has a number of train stations. One station is the main station. A station is located in a district.

    - CQs: In what city is a certain airport located? What airlines serve a specific airport? In what city is a certain train station located? Which is the main station of a certain city? In what district is a certain train station located? What are the districts of a certain city?

- When getting around in Rome by public transport there are buses, trams and a metro with two lines (A and B). The metro lines A and B run until 00.30, and resume at 5.30 in the morning.

  - When getting around in a city by public transport there are a number of means of public transport with a number of lines. Lines run until a certain time and resume a certain time in the morning.

    - CQs: What are the means of public transport in a certain city? What lines of a certain means of public transport exist in a certain city? What are the hours of operation for a certain line? What lines are operating at a certain time in a certain city?

- In Rome you can stay in a hotel. Hotels have different ratings (from five to one star), and rates within a price range. Price ranges of the hotels differ depending on the time of the year that you decide to visit Rome.

  - In a city you can stay in a hotel. A hotel has a rating and a price range. The price range differs depending on the time of the year.
    - CQs: In what city is a certain hotel located? What is the star rating of a certain hotel? What is the price range of a certain hotel at a certain time? What are the hotels within a certain price range at a certain time?

Possible solutions (cont.)

- Rome is full of sights, such as churches, monuments and museums. Popular churches are St Peters and the Pantheon, and a popular monument is the Colosseum. To enter the Colosseum there is a fee of 12 EUR for adults, a fee of 7,50 EUR for EU citizens below the age of 24 and free entrance for EU citizens below 18 years old.

    – A city has sights of different categories. To enter a sight there is a fee of a certain amount for a certain category and a certain age range.

        • CQs: What are the sights of a certain city? What sights of a certain category are located in a certain city? What is the entrance fee to a certain sight for a person of a certain category within a certain age range?

- There are organised tours to see the sights of Rome, either on foot starting from the Pantheon or by bus leaving from the Termini station, with different starting times on different weekdays.

    – There are organized tours to see the sights of a city, by different means of transport and starting from different points in the city, at different starting times on different weekdays.

        • CQs: What are the guided tours available in a certain city? What is the means of transport of a certain tour? What is the starting time of a certain tour on a certain day? What is the starting point of a certain tour?

- Rome also has numerous museums. Museums are generally closed on Mondays, but the Vatican Museums are instead closed on Sundays.

    – A city has museums. Museums are generally closed on a specific weekday. Specific museums are closed on a specific weekday.

        • CQs: What are the museums of a certain city? What day are museums generally closed in a certain city? What day is a specific museum closed?

- While in Rome a visitor should take the opportunity to visit one of the many events taking place there during different periods, such as theater and opera performances, and concerts. During September and October 2009 the Teatro Roma offers two theater shows each evening.

    – While in a city a visitor should take the opportunity to visit events that take place there during different periods. Events are of different categories. During a time period a certain facility offers a specific event at certain times.

        • CQs: What are the events taking place in a certain time interval in a certain city? What is the category of a specific event? What is the time period when a specific event is available? At what days and times is a certain event available? In what facility is a certain event taking place?

- The play in this period is the "Merchant of Venice", written by Shakespeare in 1598, given through an ensemble of well-known actors. During the second and third week of September the role of Shylock is played by Arnold Schwarzenegger as a special guest actor.

    – A certain theater event is the realization of a certain work, created by a certain author in a certain year and performed by a set of people. During a certain period of time a certain role in a theater production is played by a certain person.

        • CQs: What work is presented in a certain theater event? Who created this work? In what year was this work created? Who are the persons participating in this event? What role does a certain person play in a certain production on a certain day and time?

- There are a number of travel guides available for Rome, many of them are available in several formats, for example as a guidebook or on a CD. Maps of Rome can also be found in paper format or in an electronic format.

    – There are a number of travel guides available for a certain city. They are available in certain formats. Maps of a city are also available in certain formats.

        • CQs: What are the travel guides available for a certain city? In what editions are a certain travel guide available? What maps are available for a certain city? In what editions are a certain map available?

www.liu.se