RELATIONAL SUPPORT FOR PROTEGE

Raimundo Lozano
Felipe Geva
Xavier Pastor
CSC
INTRODUCTION

- SCOPE: “Structuring Concepts for Online Publishing Environment” (Project 22016Y1C1DMAL2)
  - **Goal**: Structuring scientific information in an ontology
  - **Medical domain**: Gastroenterology and Hepatology (G&H)
  - **Hypothesis**: Users able to search and retrieve information with a higher level of abstraction than with actual keyword-based systems
  - **Implementation**: Integrated tool for building and maintaining medical ontologies
OBJECTIVES

- Related functionality
  - Implement semantic search of contents
  - Knowledge representation
  - Multilingual support

- Related interface
  - Friendly interface
  - Structured presentation of results

- Related technology
  - Use of standards
  - Open source contribution
METHODOLOGY

- Concepts
  - Medical terminology framework of reference: UMLS (NLM)
    - Metathesaurus: + 800.000 concepts
    - Multilingual support
    - Relational database developed
  
  Knowledge representation system
  - RDF: Relational database developed
  - Protégé 2000: Extended
    - Articles categorization
    - RDF models storage
    - UMLS search capabilities

- Retrieve system: On the Web
IMPLEMENTATION

- UMLS Relational DB
- RDF Relational DB
- Developed Plugins for Protégé
  - UMLS
  - Categorization
  - RDF DB
- Web searching system
Is organized by concept or meaning; its purpose is to link alternative names and views of the same concept together and to identify useful relationships between different concepts.
UMLS - accessed from Protégé
UML - Functionality

INPUT QUERY

liver

english liver
german leber
spanish hígado
french foie
italian fegato

OUTPUT QUERY

C0023895 Disease of liver
C0023908 Liver transplant
C0085605 Liver function failure
C0023899 Liver Extract
C0019204 Carcinoma of liver cell
RDF “statements” consist of

- **resources** (= nodes)
  - which have **properties**
    - which have **values** (= nodes, strings)

```
  resource  property  value
            = subject = predicate = object
```

The sentence “http://www.w3.org/Home/Lasilla has creator Ora Lasilla” would thus be diagrammed as:

```
http://www.w3.org/Home/Lasilla  Creator  Ora Lasilla
```

*From W3C RDF Model and Syntax Specification*
RDFS

- Collection of RDF resources that can be used to describe other resources
- Provide a mechanism to define vocabularies

From W3C RDF Schema Specification
RDF STORAGE - Requirements

- Wide scope, not limited to SCOPE project needs
- Conceptual representation. Not attached to any specific format
- Portable between different DBMS.
  - Sybase Adaptive Server Anywhere
  - Sybase Adaptive Server Enterprise
  - Oracle 8i
- Efficiency retrieving concepts
RDF STORAGE

- No good models proposed
  - very simple
  - not efficient

- Solution
  - to design a new storage model
  - taking advantage of relational capabilities
    - making explicit all RDF components defined in the RDFS specification: classes, properties, literals, etc.
RDF – DB design (1)
RDF – DB design (2)
RDF – DB design (3)
RDF – DB design (4)
Classes organized in a tree with indexes
very fast searches of subclasses
RDF STORAGE – Multiple inheritance

- disease
  - ulcer
    - gastric
      - acute
    - duodenal
      - Chronic ulcer
    - Gastroduodenal chronic ulcer
  - Gastroduodenal chronic ulcer

- disease
  - ulcer
    - gastric
      - acute
    - duodenal
      - Chronic ulcer
      - Gastroduodenal chronic ulcer
    - Gastroduodenal chronic ulcer

- disease
  - ulcer
    - gastric
      - acute
    - duodenal
      - Chronic ulcer
      - Gastroduodenal chronic ulcer
    - Gastroduodenal chronic ulcer
RDF STORAGE - Interface

- Basic element: the Statement

- Stored procedures
PLUGINS – Common features

- Each plugin is implemented by a class derived from AbstractTabWidget.
  - Access to Protégé classes
    - KnowledgeBase
    - Class management ->Cls
    - Properties management -> Slot
    - Tree interface -> ClsesPanel
  - Tab presentation
  - Easy configuration
- Database access using jdbc:odbc.
  - It is allowed to choose the database
    - Plugin RDF.
    - Plugin Categorization.
PLUGINS - UMLS

jdbc:odbc connection
PLUGINS - UMLS

- Concept search
  - Variable number of terms allowed
  - Ordered result list with the most similar concept highlighted

- Adding a concept to the ontology
  - Multiple parents selection allowed
  - Automatic addition of:
    - UMLS code
    - UMLS semantic type
    - Semantic description
PLUGINS - Categorization

jdbc:odbc connection
PLUGINS - Categorization

- Show the list of articles
  - Title, volume, issue, abstract...

- Categorisation
  - Selecting an article
  - Article class automatically created
  - Article identifier automatically added
  - Volume and issue parents automatically stated
  - Allow selecting other parents
PLUGINS - RDF

RDF-XML file

Model comparison and statements extraction

Stored procedures on the database

JENA
Java API for RDF

RDF
RDF STORAGE – Integrity

- Valid model needed ⇐ Protégé
- Not ordered statements in RDF: validity assumed
  - Automatic creation of needed resources

  e.g.: (Gastritis, type, Disease)
  If not exists Disease ⇒ class Disease is created
PROBLEMS – Name modification

- The common identifier between the database and Protégé is the resource name
- The user is allowed to modify the name
- Changes on Protégé needed
  - List of modified elements in DefaultKnowledgeBase
  - New attributes and functions in DefaultFrame
PROBLEMS – Type definition

- Problems with abbreviated format of type definition
  - Protégé read as a literal
  - RDFFrameWalker.getDirectType(Resource resource) modified to create the class
AKNOWLEDGEMENTS

- SCOPE partners
  - Universitat Pompeu Fabra: the coordinating institution for SCOPE
  - DOYMA: a branch of Havas-MediMedia
  - OrbiTeam Software GmbH: a spin-off company of GMD, the German National Research Center for Information Technology
  - SESI group of the University of Wales, Bangor

- Other institutions
  - Stanford Medical Informatics
  - National Library of Medicine