Knowledge Acquisition in the construction of ontologies: a case study in the domain of hematology
Topics

- Knowledge Acquisition
  - Background
  - Classification of techniques
- Case Study
  - Context
  - Methodological steps
- Results
  - Consolidated methodology
  - Observations on the KA process
Introduction

- Development of ontologies $\rightarrow$ knowledge acquisition (KA);
- Literature mentions difficulties in communication between experts and knowledge engineers;
- We investigate the KA activity within biomedicine;
- Scope of the investigation: a project about human blood;
- Goal of the project: a knowledge base for scientific and educational purposes;
- Contributions of this paper: a methodology for KA; observation of problems during the activity.
Background

- KA includes knowledge collection, analysis, structuring and validating for representation purposes;

- Involves manual and computer-based tasks;

- Diverse definitions; theories and methods underlying KA rely on diverse fields:
  - Computer Science
  - Cognitive Science
  - Linguistics
  - Psychology
Classification of KA techniques

According to the type of technique:
- Manual techniques
- Computer-based techniques
  - Psychology
  - Automatic
  - Semi-automatic

According to the knowledge obtained in the process:
- Differential access hypothesis

According to the method of application:
- Protocol-generation techniques (e.g. teachback)
- Protocol-analysis techniques (e.g. transcriptions)
- Matrix-based techniques (e.g. repertory grid)
- Sorting techniques (e.g. card sorting)
Case Study – The context

- Blood bank responsible for healthcare services for a population of 20 million people
- Project for organization of information which includes construction of ontologies
- Direct observation of the activities of the blood bank
- Interviews with a group of 20 experts in a period of six months
Case Study – Methodological Steps

- Four steps: extraction, elicitation, validation and refinement

**EXTRACTION**

- Sketch Engine
- Information retrieval
- Fragmentation of the selected text into morphemes
  - text = set of manuals of the AABB
  - Information retrieval
    - *apheresis*
    - *desis*
    - *ectomy*
    - ...

**AntConC**

- British National Corpus

**Morphological productivity**

<table>
<thead>
<tr>
<th>Term</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>apheresis</td>
<td>124</td>
</tr>
<tr>
<td>phlebotomy</td>
<td>32</td>
</tr>
<tr>
<td>cytometry</td>
<td>20</td>
</tr>
<tr>
<td>cordocentesis</td>
<td>16</td>
</tr>
<tr>
<td>plasmapheresis</td>
<td>15</td>
</tr>
</tbody>
</table>
ELICITATION

Terms from the extraction phase

Interviews, sorting and matrix techniques

Protege frames template 1

- disease as disposition approach (Scheuermann, Ceusters & Smith, 2009)
- ethiological process
- course of the disease
- therapeutic response
VALIDATION

Wiki science tools

collaborative validation

REFINEMENT

Protege frames template II

Specialized Ontologies
- Anatomy
- Gene
- Proteins
- ...

Edit Expert proposal: HematopoieticNeoplasm

Designation

Preferred Name: HematopoieticNeoplasm
Other Name: 

Description

Definition: An hematopoietic neoplasm is a hematologic malignancy which forming tissues.

Example: 

Source: 

Navigation
Main page
Recent changes
Help

Sub-ontologies
BLO Core

Basic tools
Create categories
See categories
See forms
See all pages
## Results - List of methodological steps

<table>
<thead>
<tr>
<th>Phase</th>
<th>Task</th>
<th>Description</th>
<th>Resources and people involved</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Extraction</td>
<td>1.1 build a corpus</td>
<td>Create a corpus from texts</td>
<td>-Medical texts -K. engineer</td>
</tr>
<tr>
<td></td>
<td>1.2 codification</td>
<td>Automatically fragment texts</td>
<td>-Sketch Engine tool -K. engineer</td>
</tr>
<tr>
<td></td>
<td>1.3 information retrieval</td>
<td>Obtain terms through suffixes</td>
<td>-Sketch Engine tool -K. engineer</td>
</tr>
<tr>
<td>(2) Contact</td>
<td>2.1 obtain knowledge</td>
<td>Hold interviews with experts</td>
<td>-Template Protégé and teachback; -K. engineer, experts</td>
</tr>
<tr>
<td></td>
<td>2.2 know the terminology</td>
<td>Identify experts’ rationale</td>
<td>-Matrix Techniques -K. engineer and expert</td>
</tr>
<tr>
<td></td>
<td>2.3 see ad-hoc organization</td>
<td>Understand how experts sort concepts</td>
<td>-Sorting techniques -Experts</td>
</tr>
<tr>
<td>(3) Validation</td>
<td>3.1 validate knowledge</td>
<td>Obtain approval of terms acquired</td>
<td>-Wiki Page -Expert</td>
</tr>
<tr>
<td></td>
<td>3.2 updating</td>
<td>Update data after each validation</td>
<td>Wiki Page -K. engineer</td>
</tr>
<tr>
<td>(4) Refinement</td>
<td>4.1 integration between granularities</td>
<td>Characterize related genes, proteins, etc</td>
<td>-Template Protégé -K. engineer</td>
</tr>
<tr>
<td></td>
<td>4.2 connection with top-level</td>
<td>Connect data with other ontologies</td>
<td>-Template Protégé -K. engineer</td>
</tr>
</tbody>
</table>
Results

- An OWL ontology with more than 300 classes and 50 properties

Observations:

- Use of list of terms automatically extracted: avoid the KEs’ needing to start interviews from scratch
- Use of ontological disease model for interviews (Scheuermann, Ceusters & Smith, 2009): very useful, experts in general approved this framework
- Identified intervenient factors in process of KA: experst’s lack of time, lack of access to relevant data sources, deficiencies in the organizational structure