Fabrício M. Mendonça
Kátia C.Coelho
André Q. Andrade
Mauricio B. Almeida

School of Information Science
Federal University of Minas Gerais
Brazil

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

ICBO 2012 – Graz, Austria

Topics

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

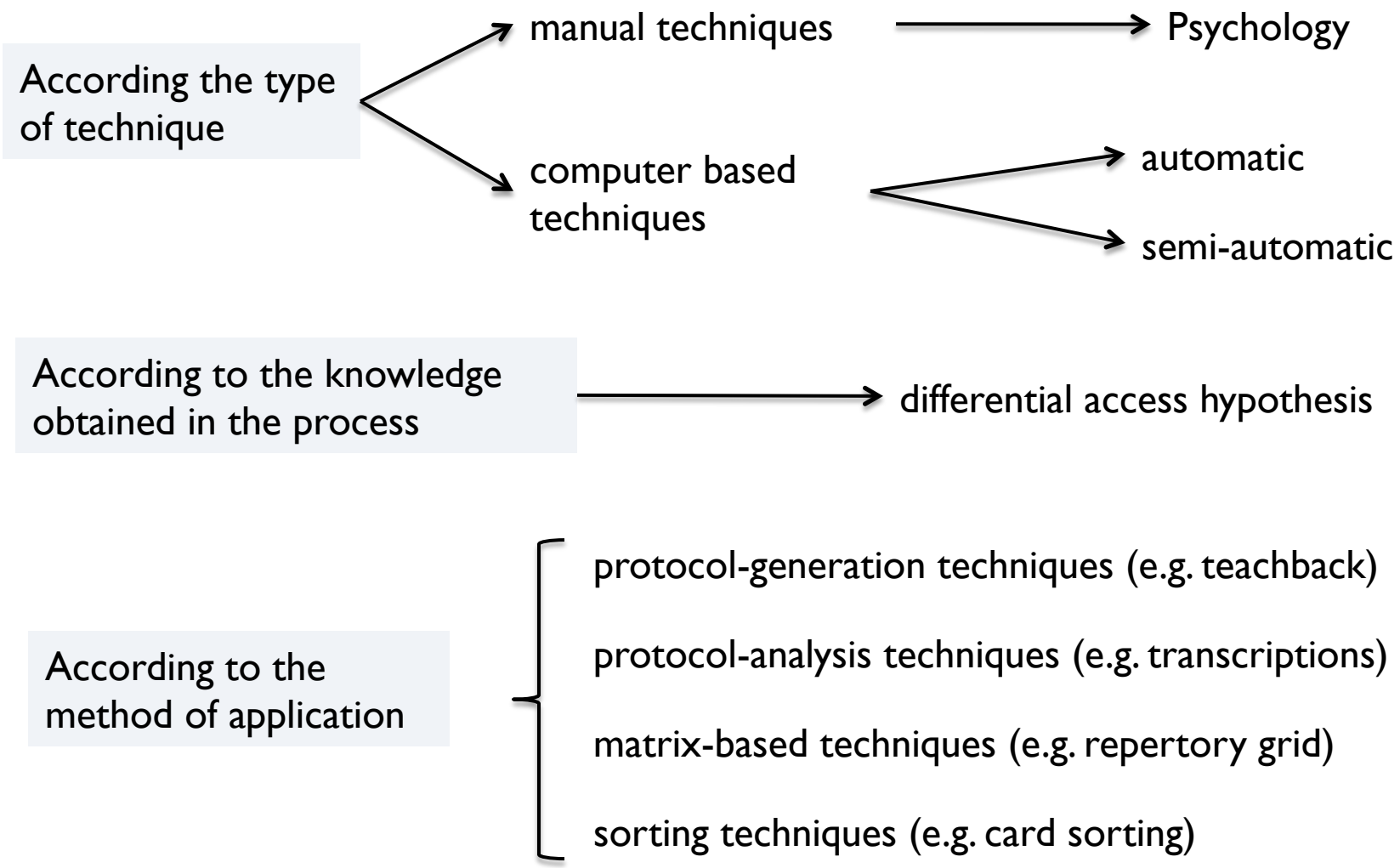
Introduction

- Development of ontologies → 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

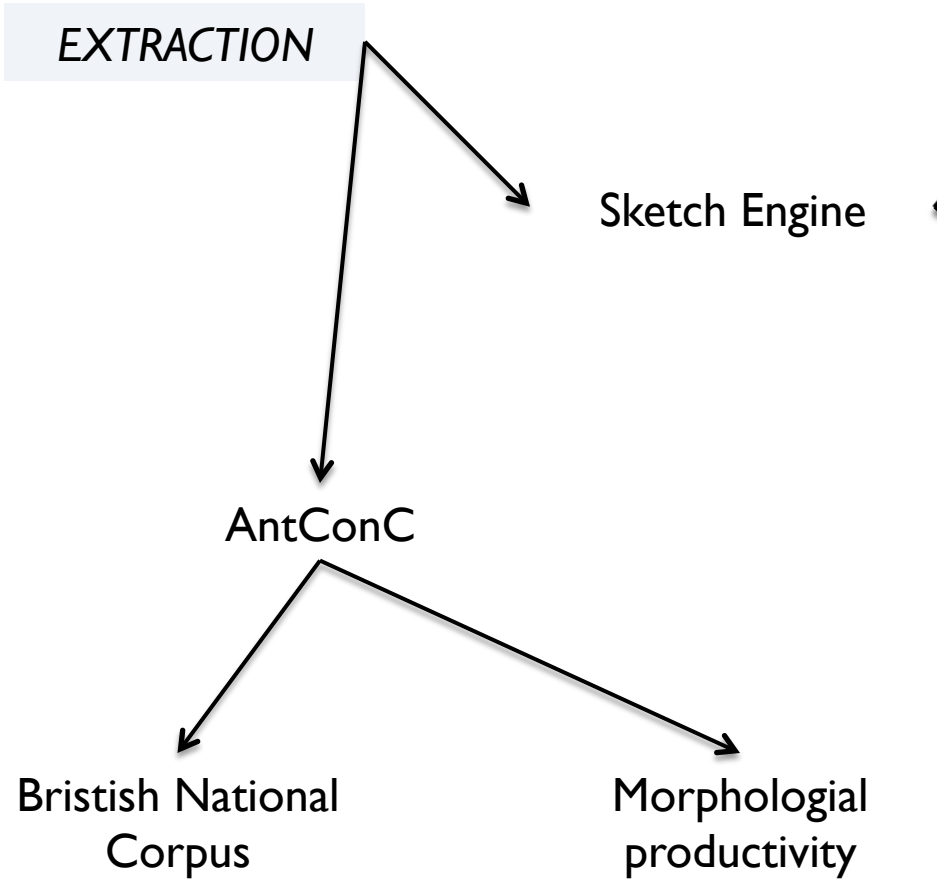


Case Study – The context

- 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



Fragmentation of the selected text into morphemes

text = set of manuals of the AABB

Information retrieval
-apheresis
-desis
-ectomy
-...

Term	Frequency
apheresis	124
phlebotomy	32
cytometry	20
cordocentesis	16
plasmapheresis	15

ELICITATION

Terms from the extraction phase

Interviews, sorting and matrix techniques

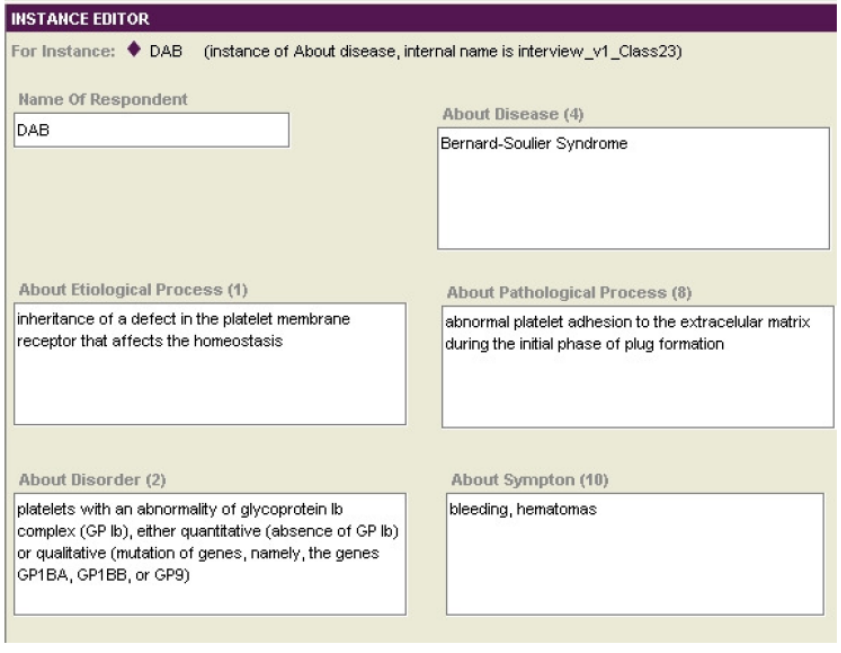
disease as disposition approach
(Scheuermann, Ceusters & Smith, 2009)

ethiological process

course of the disease

therapeutic response

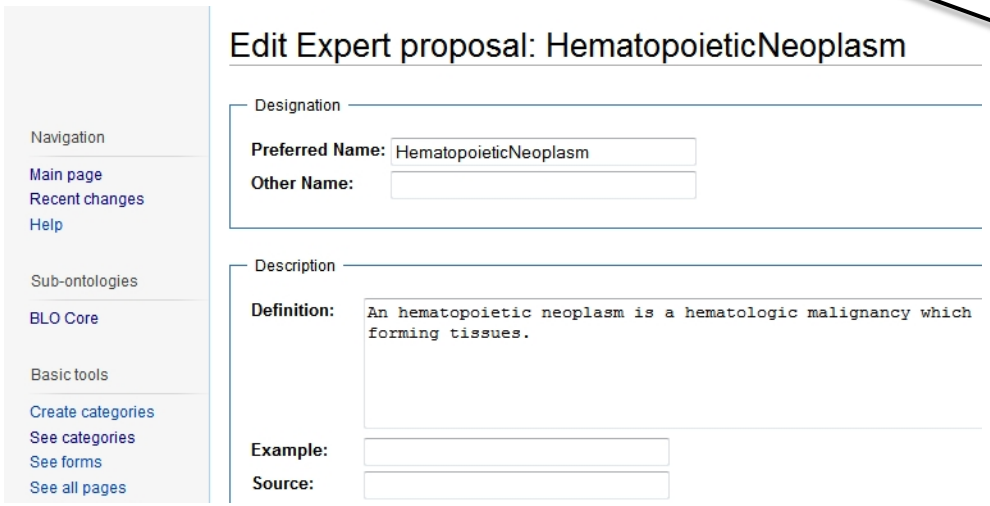
Protege frames template I



VALIDATION

Wiki science tools

collaborative validation



REFINEMENT

Protege frames template II

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

Results - List of methodological steps

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

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