

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

- ▶ **Knowledge 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.

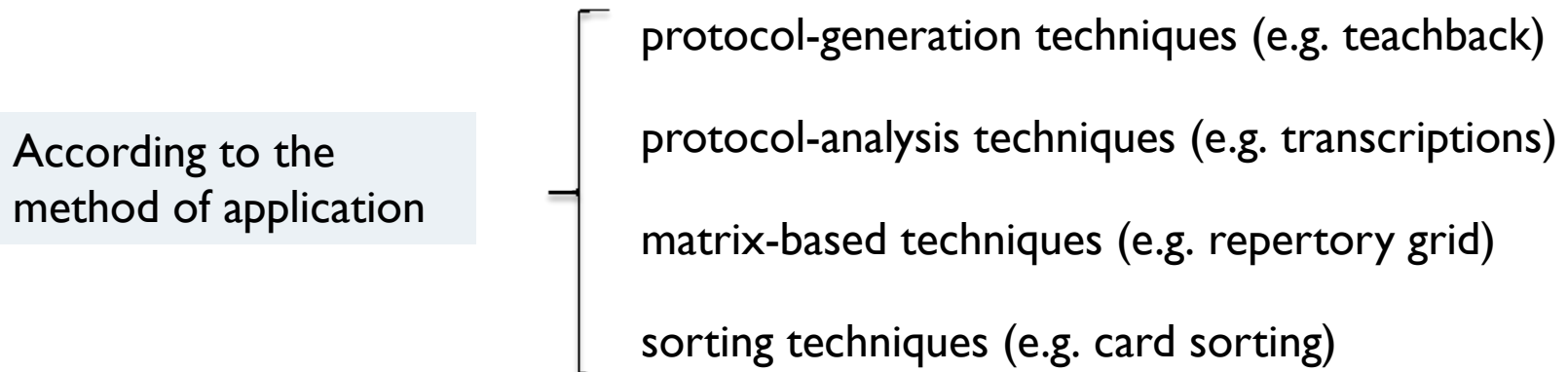
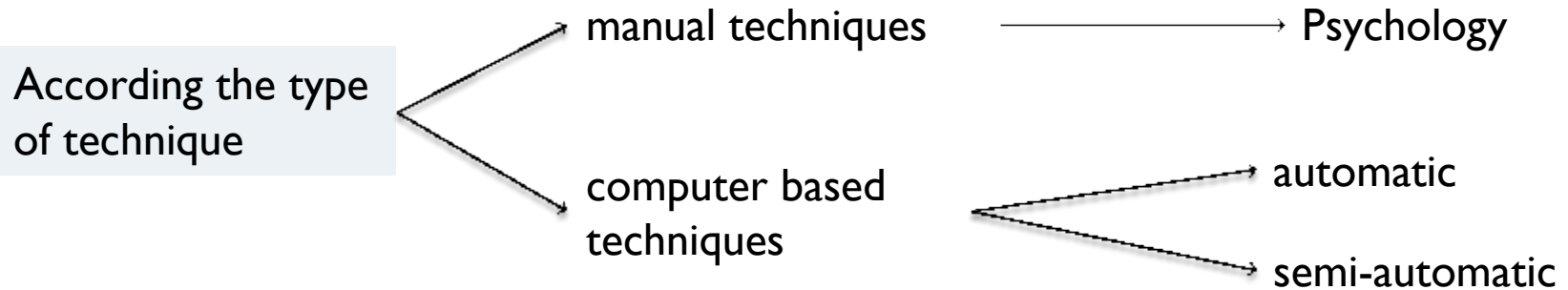


Background

- ▶ 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



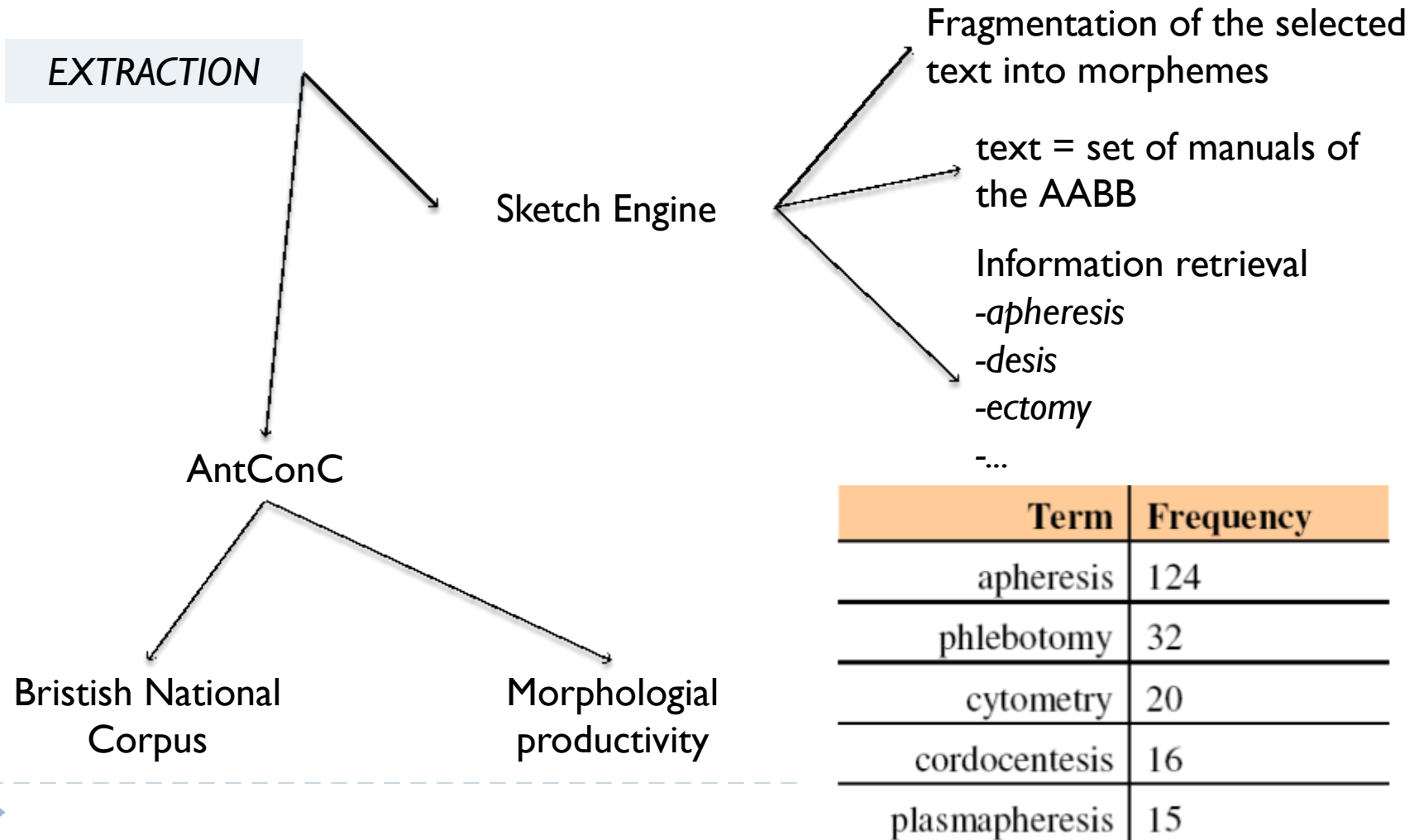
Case Study – The context

- ▶ 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



ELICITATION

Terms from the extraction phase

Interviews, sorting and matrix techniques

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

Protege frames template I

ethiological
process

course of
the disease

therapeutic
response

INSTANCE EDITOR

For Instance: ◆ DAB (instance of About disease, internal name is interview_v1_Class23)

Name Of Respondent DAB	About Disease (4) Bernard-Soulier Syndrome
About Etiological Process (1) inheritance of a defect in the platelet membrane receptor that affects the homeostasis	About Pathological Process (8) abnormal platelet adhesion to the extracellular matrix during the initial phase of plug formation
About Disorder (2) platelets with an abnormality of glycoprotein Ib complex (GP Ib), either quantitative (absence of GP Ib) or qualitative (mutation of genes, namely, the genes GP1BA, GP1BB, or GP9)	About Symptom (10) bleeding, hematomas

VALIDATION

Wiki science tools

collaborative
validation

The screenshot shows a web interface for editing an expert proposal. On the left is a navigation sidebar with links: 'Main page', 'Recent changes', 'Help', 'Sub-ontologies', 'BLO Core', 'Basic tools', 'Create categories', 'See categories', 'See forms', and 'See all pages'. The main content area is titled 'Edit Expert proposal: HematopoieticNeoplasm'. It contains several input fields: 'Designation' (empty), 'Preferred Name' (filled with 'HematopoieticNeoplasm'), 'Other Name' (empty), 'Description' (filled with 'An hematopoietic neoplasm is a hematologic malignancy which forming tissues.'), 'Example' (empty), and 'Source' (empty).

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 <i>ad-hoc</i> 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 : expert's lack of time, lack of access to relevant data sources, deficiencies in the organizational structure

