

Semi-Automated Ontology Evaluation Based On Competency Questions and Query Expansion



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

Philipp HOFER^{1a}, Sabrina NEURURER^a, Helga HAUFFE^b,
Thomas INSAM^c, Anette ZEILNER^d and Georg GÖBEL^a

¹*philipp.hofer@i-med.ac.at*

^aDepartment for Medical Statistics, Informatics and Health Economics
Medical University of Innsbruck

^b*Department of Urology, Medical University of Innsbruck*

^c*Department of Obstetrics and Gynecology, Medical University of Innsbruck*

^d*Division of Human Genetics, Medical University of Innsbruck*

Overview



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK



Introduction

Which Information is Stored in a Biobank ?
Relevance of Biomedical Ontologies for Biobanking

Objectives

Semi-Automated Ontology Evaluation Approach

Methods

Query Expansion, Medical Subject Headings & WordNet
Query-Expansion Approach

Results

Competency Questions and Relevant Ontology Concepts
Evaluation Results

Discussion & Outlook

Which Information is Stored in a Biobank ?



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

Introduction

Objectives

Methods

Results

Discussion & Outlook



- Sample related information
 - Collection process, sample processing and storage
- Donor related information
 - Medical history, current disease, medication, family history, therapies
- Information about access control on samples

Relevance of Biomedical Ontologies for Biobanking



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

Introduction

Objectives

Methods

Results

Discussion & Outlook



- Searching for samples across different biobanks needs a common semantic foundation¹
- Harmonization of biobank contents via biomedical ontologies
- It remains unclear which biomedical ontologies are relevant for the biobanking domain
- E.g. anatomical structures and body fluids are represented by several ontologies (FMA, SNOMED, MeSH, etc.)

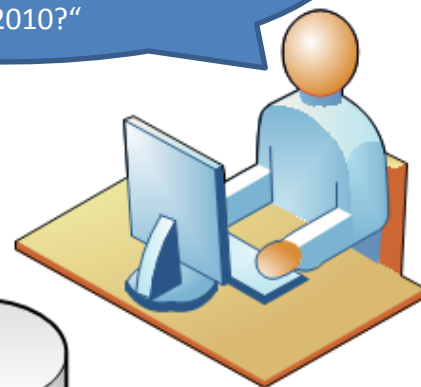
[1] Andrade, A. Q., Kreuzthaler, M., Hastings, J., Krestyaninova, M., & Schulz, S. (2012). Requirements for semantic biobanks. In *Studies in Health Technology and Informatics* (Vol. 180, pp. 569–573)

Semi-Automated Ontology Evaluation Approach

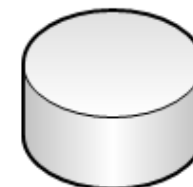
Introduction | Objectives | Methods | Results | Discussion & Outlook

- We want to propose a semi-automated competency evaluation approach to identify candidate ontologies for the biobanking domain
- Competency questions (CQs)
 - Representation of typical user requests to an information system
 - Used to define the scope of ontologies within the ontology engineering process

„Which radical prostatectomies with biopsies, recurrence as well as subsequent radiation and medical therapies existed in the year 2010?“



Ontology for Biobanking



ontology-based
database schema



OMIABIS



Biological Collections Ontology



...



Query Expansion, Medical Subject Headings (MeSH) & WordNet



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

Introduction

Objectives

Methods

Results

Discussion & Outlook



- Query Expansion
 - Extension of the original query by additional search terms to improve retrieval performance
 - Identification of semantic relationships between competency question terms and biobank ontologies by using lexical-semantic relationships from other hierarchical terminologies
- Medical Subject Headings (MeSH)
 - Hierarchical terminology of medical terms and synonyms
- WordNet
 - Database containing words and related synonyms for the English language
- Evaluation of the OMIABIS biobanking ontology

Query-Expansion Approach



Introduction	Objectives	Methods	Results	Discussion & Outlook
--------------	------------	----------------	---------	----------------------

Input:
Competency question CQ
Ontology O

Which radical prostatectomies with biopsies, recurrence as well as subsequent radiation and medical therapies existed in the year 2010?

<http://purl.obofoundry.org/obo/omiabis.owl>

Lexical Processing
Stemming and
stopword removal

E04:Surgical Procedures, Operative
E04.950:Urogenital Surgical Procedures
E04.950.774:Urologic Surgical Procedures
E04.950.774.860:Urologic Surgical Procedures, Male
E04.950.774.860.625:Prostatectomy

{biopsy, radiation, medical, therapy, 2010, existed, prostatectomy, radical, year, recurrence, subsequent}

```
SELECT ?label ?concept
{
  ?concept a owl:Class .
  ?concept rdfs:label ?label .
  FILTER(REGEX(STR(?label), „<CQ term>“))
}
```

Lexical
matching
SPARQL

no direct match
→
{prostatectomy,
biopsy, ...}

Query
expansion



direct match

Output:
Matched
entities

Lexical Processing
Stemming and
stopword removal

prostatectomy → OMIABIS_0001041: surgical procedure,
medical → OMIABIS_0001026: medical record,
...

Competency Questions and Relevant Ontology Concepts



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

Introduction

Objectives

Methods

Results

Discussion & Outlook



- Competency questions were defined in collaboration with medical experts from different departments at the Medical University of Innsbruck
- Manual definition of relevant ontology entities for the evaluation process

CQ Terms	Example: Relevant entity relationships (OMIABIS)
biopsy	primary diagnosis, diagnosis, sample handling, diagnostic process, surgical procedure, specimen surgical removal
radiation	treatment
medical	medical record, sample medical record
therapy	treatment
2010	value specification
existed	exists at, during which exists
prostatectomy	surgical procedure, specimen surgical removal
radical	surgical procedure, specimen surgical removal
year	date of data entry, time unit
recurrence	specimen disease state data, disease state specimen data
subsequent	-

Evaluation Results

Introduction Objectives Methods **Results** Discussion & Outlook

- “Which radical prostatectomies with biopsies, recurrence as well as subsequent radiation and medical therapies existed in the year 2010?”

Evaluation of OMIABIS

535 concepts, 77 object properties (including imports from higher ontologies) (22.2.2015)

Keywords	True positive	False positive	Total matches
biopsy	3	63	66
radiation	0	1	1
medical	2	0	2
therapy	0	0	0
2010	0	0	0
existed	2	0	2
prostatectomy	2	3	5
radical	0	8	8
year	0	1	1
subsequent	0	0	0
recurrence	2	3	5

Discussion & Outlook



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

Introduction

Objectives

Methods

Results

Discussion & Outlook



- Pilot-approach for systematical evaluation of candidate ontology entities based on a query expansion to state the usability of specific ontologies for the biobanking domain
- However there are still limitations to be considered
 - Compound nouns, e.g. “radical prostatectomy”, “medical therapy”
 - We plan to extend and refine the term matching process towards integrating composite terms to reduce false positive matches.
- For next steps we plan to conduct broader study on several different biomedical ontologies using an extended set of competency questions

Thank you for your attention !



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

Philipp Hofer

Department of Medical Statistics, Informatics
and Health Economy
Medical University of Innsbruck

Sabrina Neururer

Department of Medical Statistics, Informatics
and Health Economy
Medical University of Innsbruck

Georg Goebel

Department of Medical Statistics, Informatics
and Health Economy
Medical University of Innsbruck

Helga Hauffe

Department of Urology
Medical University of Innsbruck

Anette Zeilner

Division of Human Genetics
Medical University of Innsbruck

Thomas Insam

Department of Obstetrics and Gynecology
Medical University of Innsbruck

Funding
BBMRI.AT



BBMRI
Biobanking and
Biomolecular
Resources Research
Infrastructure