

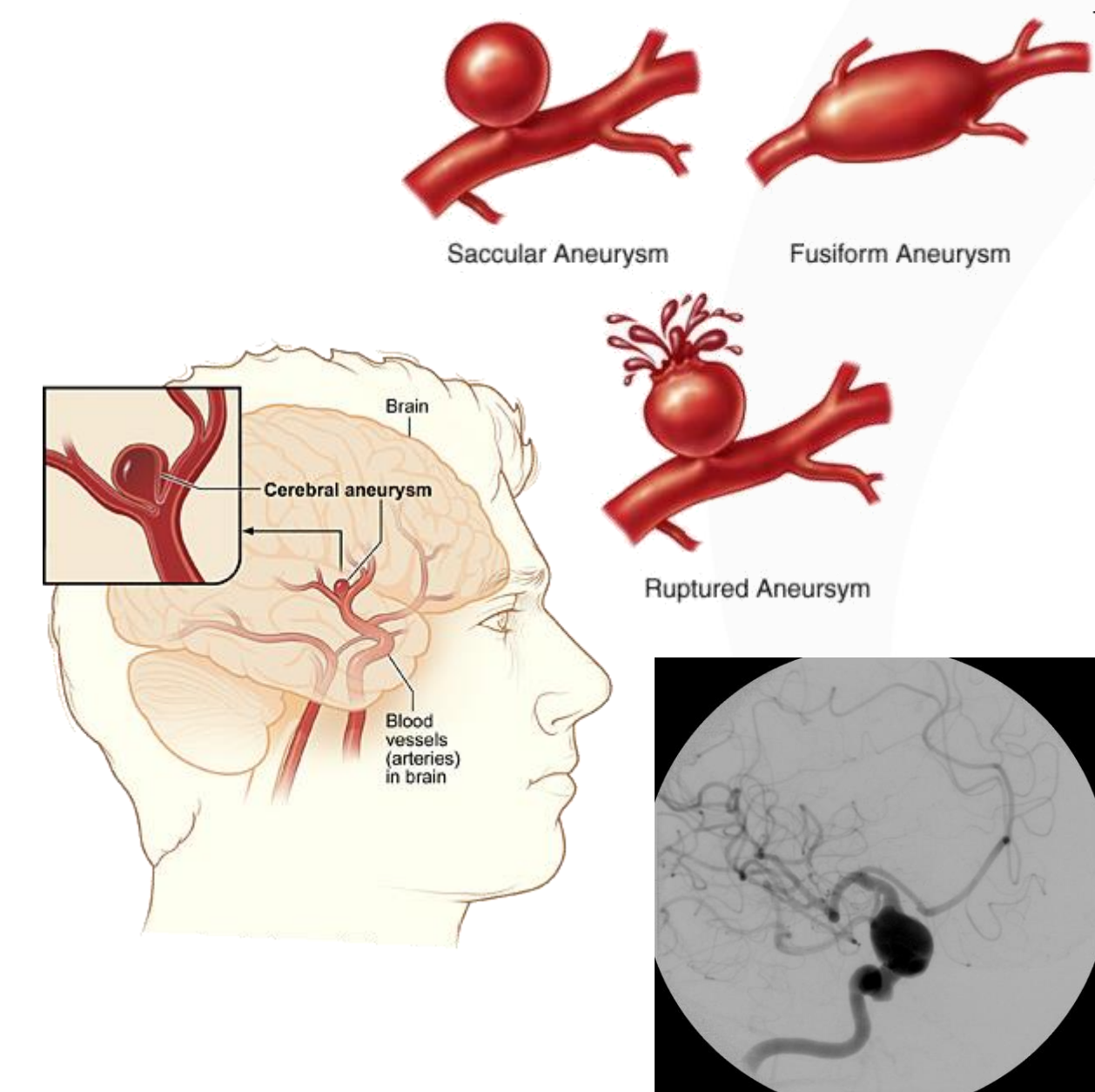
DEVELOPING AN OWL-DL ONTOLOGY FOR RESEARCH AND CARE OF INTRACRANIAL ANEURYSMS – CHALLENGES AND LIMITATIONS

Holger Stenzhorn, Martin Boeker, Stefan Schulz, Susanne Hanser

Institute for Medical Biometry and Medical Informatics, Freiburg University Medical Center, Freiburg, Germany

INTRODUCTION

The European project @neurIST aims to provide an integrated information infrastructure related to intracranial aneurysms and subarachnoid bleedings. Its benefits for clinicians and scientists include improved support of diagnosis and treatment planning and an easier access to domain knowledge.



The @neurIST ontology integrates several disease description levels, e.g. clinical, genetic, epidemiologic views from various information sources, e.g. clinical databases, literature and terminologies.

SOURCES

- Clinical databases and information models
- Literature abstracts
- UMLS Metathesaurus
- Domain experts
- Open biomedical databases

AVAILABILITY

The ontology and related material can be downloaded at <http://ontology.aneurist.org>

Upper-Level Distinctions

“Is a disease an event or a state ?”

Naïve Approach

Is “BrainNeoplasm \sqsubseteq Event” or is “BrainNeoplasm \sqsubseteq Stative”?

Critique

Ontologically, there are two different entities: 1) the disease itself and 2) the course of the disease. But this distinction is not needed in the given context!

Solution

Introduce the disjunction class “StateOrProcess \sqsubseteq Event \sqcup Stative” and hence “BrainNeoplasm \sqsubseteq StateOrProcess”

Contextual Knowledge

“Smoking is a risk factor for aneurysm rupture”

Naïve Approach

“TobaccoSmoking \sqsubseteq AneurysmRuptureRiskFactor
TobaccoSmoking \sqsubseteq Process”

Critique

A context-dependent statement is represented here but not a generic property of TobaccoSmoking. Still, such classes are considered important for ontology navigation and retrieval.

Solution

Clearly separate 1) the ontology proper from 2) the context ontology
1) “TobaccoSmoking \sqsubseteq Process \sqsubseteq Perdurant \sqsubseteq Particular”
2) “TobaccoSmoking \sqsubseteq AneurysmRuptureRiskFactor \sqsubseteq RiskFactor \sqsubseteq ParticularInContext”

REFERENCES

- S. Hanser, J. Fluck, L. Furlong, C. Friedrich, M. Hofmann-Apitius, H. Stenzhorn, M. Boeker, Knowledge Structuring and Retrieval for Intracranial Aneurysm Research. In Proc. of the HealthGrid Conference, Chicago, USA, 2008.
- A. Gangemi, N. Guarino, N. Masolo, A. Oltramari, L. Schneider, Sweetening ontologies with DOLCE. In Proc. of the Internat. Conf. on Knowledge Engin. and Management (EKAW), Sigüenza, Spain, 2002.
- S. Schulz, H. Stenzhorn, Ten Theses on Clinical Ontologies, In Proc. of the International Council on Medical and Care Computetics Event (ICMCC 2007), Amsterdam, The Netherlands, 2007.

Information Entities

“ABCC5 gene encodes the protein MRP5”

Naïve Approach

“ABCC5 \sqsubseteq \exists encodes.MRP5”

Critique

There may be ABCC5 (sensu nucleotide chain) instances that happen to never encode any instance of the protein MRP5

Solution 1 (preferred)

Use value restrictions “ABCC5 \sqsubseteq \forall encodes.MRP5”

Solution 2

Consider ABCC5 and MRP5 not as classes but as information entity instances “ABCC5 \in InformationEntity
MRP5 \in InformationEntity
<ABCC5, MRP5> \in encodes”

Uncertainty

“Pregnancy is a suspected risk factor for aneurysm rupture”

Naïve Approach

“Pregnancy \sqsubseteq SuspectedAneurysmRuptureRiskFactor”

Critique

The context-dependent statement on pregnancy as a risk factor is modified by a modal expression. This is not an issue to be handled by ontologies!

Objection

The distinction between suspected and proven risk factors is crucial for the use of the ontology (retrieval)

Solution

Maintain the naïve approach but encode it within the context ontology

Dispositions

“Headache is associated with intracranial aneurysm”

Naïve Approach

“Headache \sqsubseteq \exists associatedWith.IntracranialAneurysm
IntracranialAneurysm \sqsubseteq \exists associatedWith.Headache”

Critique

Not every headache is a symptom of intracranial aneurysm and not every intracranial aneurysm produces headache.

Solution

Introduce (anonymous) dispositions
“IntracranialAneurysm \sqsubseteq
 \exists hasDisposition.(\forall hasRealization Headache)”

Residual Classes

“BBS2 is a protein with unknown function”

Naïve Approach

“BBS2 \sqsubseteq ProteinWithUnknownFunction \sqsubseteq Protein”

Critique

Whether a certain function is known or unknown is ontologically irrelevant

Objection

Such classes – typical for statistical classification systems – are important for ontology navigation and housekeeping

Solution 1

Eliminate irrelevant class “BBS2 \sqsubseteq Protein”

Solution 2

Mark such a class as housekeeping or navigational class
“BBS2 \sqsubseteq NAVProteinwithUnknownFunction \sqsubseteq Protein”

ARCHITECTURE

- DOLCE lite top-level ontology
- OWL-DL (SHIN(D))
- Editor: Protégé 4
- Reasoner: Pellet, Fact++
- Web-based Ontology browser

COVERAGE

- About 2800 classes
- Scope:
 - anatomy, surgery, neurology
 - fluid dynamics
 - epidemiology
 - molecular biology
- 98 relations (with 70 inherited form DOLCE)
- Linked to lexicon with about 9000 entries

CURRENT STATE

- Used for text mining
- Linkage to clinical information systems in preparation

FURTHER CHALLENGES

- Create convincing use cases for demonstrating the benefit for ontology in @aneurIST
- Avoid overlap between ontology and information model design
- Communicate the rationale of ontology support for semantic mediation
- Train curators in applying ontology best practices and avoiding systematic modeling errors (see examples)