

Towards Ontological Interpretations

Robert Hoehndorf, Axel Ngonga and Michael Dannemann

Institute for Informatics and Research Group Ontologies in Medicine, Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig

Department of Evolutionary Genetics, Max-Planck-Institute for Evolutionary Anthropology

SMBM 2008

Text Mining and Ontologies

Tasks of Text Mining

- Named Entity Recognition
- Annotation Extraction
- Relationship Extraction
- Ontology Extraction
- Text Classification

Text Mining and Ontologies

Tasks of Ontologies

- Annotation of genes/proteins with categories
- Provision of taxonomic relations
- Provision of definitions for categories and relations
- Provision of axioms

Text Mining and Ontologies

Bridging the gap between ontologies and text mining

- Formalized (axiomatized) ontologies are rich sources of knowledge.
- Knowledge from formalized ontologies is rarely used for text mining.
- Can formal ontologies be used for *verification* of extracted information?
- Can formal ontologies be used for *explanation* and *disambiguation* of extracted information?

Text Mining and Ontologies

Workplan

- Top-level ontological analysis of text mining results
 - Basic conceptualization of results of text mining analyses: Text Mining Structure (TMS)
 - Interpretation of the TMS in a formal ontology: Ontological Interpretation (OI)
- Algorithms to verify and enrich the TMS
 - Knowledge-based algorithms: use logical inferences
 - Reuse existing ontologies and knowledge bases
 - Develop new axiom systems

Verification

Text Mining Structures

A Text Mining Structure (TMS) $\mathcal{TM} = \langle C, I, R, T \rangle$ consists of

- a set of categories C ,
- a set of individuals I ,
- a set of relations R , and
- a set of instances of relations T .

A TMS provides a basic conceptualization for results of text mining analyses.

Verification

TMS Example

Our finding that ASC interacts with pro-caspase-1 prompted us to examine whether PYPAF7 and ASC also play a role in caspase activation.

- recognize entity names
- recognize relationshipnames
- recognize relation-instances (triple)

Verification

TMS Example

Our finding that **ASC** interacts with **pro-caspase-1** prompted us to examine whether **PYPAF7** and **ASC** also play a role in **caspase activation**.

- recognize entity names
- recognize relationshipnames
- recognize relation-instances (triple)

Verification

Ontological Interpretations

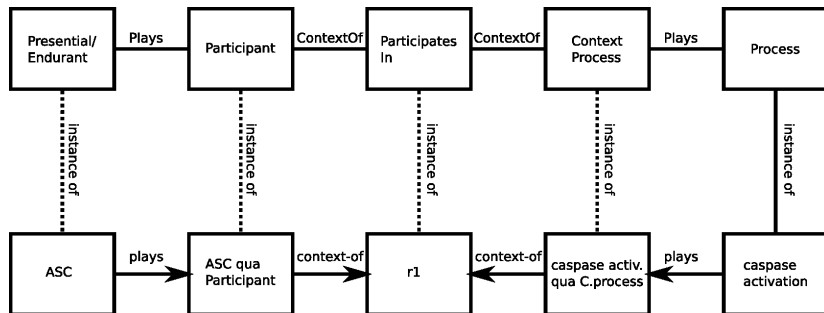
An ontological interpretation (OI) \mathcal{I} of a TMS $\mathcal{TM} = \langle C, I, R, T \rangle$ with respect to ontology $\mathcal{O} = \langle C', R', ::, isa, Ax \rangle$ satisfies:

- for each $c \in C$, $c^{\mathcal{I}} = c'$ such that $c' \in C'$ and either $c :: c'$ or $isa(c, c')$,
- for each $i \in I$, $i^{\mathcal{I}} = i'$ such that there exists a $c' \in C'$ and $i :: c'$,
- for each $r \in R$, $r^{\mathcal{I}} = r'$ such that $r' \in R'$ and $isa(r, r')$,
- for each $t \in T$, $t^{\mathcal{I}} = t'$ such that there exists a $r' \in R'$ and $t' :: r'$.

Verification

Requirements for Ontology of Roles

Our finding that ASC interacts with pro-caspase-1 prompted us to examine whether PYPAF7 and ASC also play a role in caspase activation.



- *plays-role-in is-a participates-in*
- Relational roles in DOLCE and GFO

Knowledge-based post-processing

Inconsistent Interpretations

IsA(Arsenic, Poison)

PlaysRole(Arsenic, Poison)

HasFunction(Arsenic, Poison)

- Inconsistent use of *Poison*: Substance, Role, Function
- Solution: generalization to *IsA(Poison, Entity)*
- Wanted: *Poison* \sqsubseteq (*Substance* \sqcup *Role* \sqcup *Function*)

Knowledge-based post-processing

Generating Explanations for Incomplete Interpretations

- Many possible OIs for each TMS
- Wanted:
 - *Justification* for each fact
 - *Minimality* of justifications
- Generation using additional axioms and abduction
- Abduction: $\{B, A \rightarrow B\} \vdash A$

$$isa(Poison, C_1) \vee \dots \vee isa(Poison, C_n) \rightarrow isa(Poison, Entity)$$

Knowledge-based post-processing

Generating Explanations for Incomplete Interpretations

IsA(Arsenic, Poison)

PlaysRole(Arsenic, Poison)

HasFunction(Arsenic, Poison)

isa(Poison, Entity)

isa(Poison, C₁) ∨ ... ∨ isa(Poison, C_n) → isa(Poison, Entity)

Poison \sqsubseteq (*Substance* \sqcup *Role* \sqcup *Function*)

Recapitulation

- TMS as top-level conceptualization of text mining results
- Ontological Interpretation of a TMS for verification of text mining results
- Use of abductive reasoning to generate Ontological Interpretations
- Choice of ontology determines granularity of verification and generation of knowledge

Acknowledgements



Michael Dannemann



Janet Kelso



Axel Ngonga



Heinrich Herre

Knowledge-based post-processing

Abductive Reasoning

Abduction: $\{B, A \rightarrow B\} \vdash A$

Let Γ be a knowledge base.

- Consistency: $\Gamma \cup A \not\vdash \perp$
- Minimality: A is a *minimal* explanation for B
- Relevance: $A \not\vdash B$
- Explanatoriness: $\Gamma \not\vdash B$

Verification

TMS Example

Our finding that **ASC** interacts with **pro-caspase-1** prompted us to examine whether **PYPAF7** and **ASC** also play a role in **caspase activation**.

Categories:

- ASC (UniProt: Q9ULZ3)
- pro-caspase-1 (UniProt: Q8WXC3)
- PYPAF7 (UniProt: P59046)
- caspase activation (GO:0006919)

Verification

TMS Example

Our finding that **ASC** interacts with **pro-caspase-1** prompted us to examine whether **PYPAF7** and **ASC** also play a role in **caspase activation**.

Relations:

- *interacts-with*
- *plays-role*

Verification

TMS Example

Our finding that **ASC** interacts with **pro-caspase-1** prompted us to examine whether **PYPAF7** and **ASC** also play a role in **caspase activation**.

Relation instances:

- **ASC(actor)** *interacts-with* **pro-caspase-1(actor)**
- **PYPAF7(player)** *plays-role* **caspase activation(role)**
- **ASC(player)** *plays-role* **caspase activation(role)**

Verification

OI Example

Our finding that **ASC** interacts with **pro-caspase-1** prompted us to examine whether **PYPAF7** and **ASC** also play a role in **caspase activation**.

Categories:

- *ASC is-a Protein*
- *ASC instance-of Category_of_protein*
- *pro-caspase-1 is-a Protein*
- *PYPAF7 is-a Protein*
- *caspase_activation is-a Biological_process*

Verification

OI Example

Our finding that **ASC** interacts with **pro-caspase-1** prompted us to examine whether **PYPAF7** and **ASC** also play a role in **caspase activation**.

Relations:

- *interacts-with* equivalent-to *ont:interacts-with*
- *plays-role-in is-a* *participates-in*

Verification

OI Example

Our finding that **ASC interacts** with **pro-caspase-1** prompted us to examine whether **PYPAF7** and **ASC** also **play a role** in **caspase activation**.

Relation instances (triple):

- *ASC interacts-with*
pro-caspase-1= r_1
- $r_1 :: ont:interacts-with$