Ontology for Literary Text

Andrew U. Frank

Department of Geodesy and Geoinformation
TU Wien

March 20, 2018

Outline

1. My Starting Point
   - My past research
   - Why and when are Ontology Projects Successful
2. Ontology for GIS
3. Computational Reading
   - My current interest
   - Layers of Natural Language Processing
4. Representation of Meaning
   - AMR Abstract Meaning Representation
   - Extracting Ontology from Text

Outline

1. My Starting Point
   - My past research
   - Why and when are Ontology Projects Successful
2. Ontology for GIS
3. Computational Reading
   - My current interest
   - Layers of Natural Language Processing
4. Representation of Meaning
   - AMR Abstract Meaning Representation
   - Extracting Ontology from Text

GIS Research

- Professor at TU Wien for Geoinformation for 25 years.
  NCGIA (University of Maine branch) before.
- Spatial relations, especially spatial reasoning with distances and directions.
- "Semantics of spatial data" (part of the title of my Ph.D. thesis in 1982).
  Connection between
  - datastructures for databases
  - object oriented programming
    using the functional, second order, programming language Haskell
  - visualization of data
  - economic value of data and data quality
- Extend ontologies beyond taxonomy:
  add actions and operations (verbs) to the ontology

Outline

1. My Starting Point
   - My past research
   - Why and when are Ontology Projects Successful
2. Ontology for GIS
3. Computational Reading
   - My current interest
   - Layers of Natural Language Processing
4. Representation of Meaning
   - AMR Abstract Meaning Representation
   - Extracting Ontology from Text

Example Successful Projects

- Bio- or Medical and similar:
  of interest to pharma and health industry
- Manufacturing: description of parts for order processing
- Accounting and banking
- Law: managing the documents relevant for large cases.

Factors for Success

- Substantive commercial interest
- An authority can fix the ontology ("fiat" ontologies)
- Seldom: factual or scientific justification
- Issue: Reistance to the structure imposed by the ontology
  a modern form of imperialism

Example

The political discussion on bank accounting rules (US vs. Europe)

Ontologies for GIS cannot be "fiat"

- Classification of real world (geographic) objects depend on the point of vie,i.e. the intended use of the information
- Efforts to describe object properties "objectively" (Werner Kuhn and his students)

Tiers of ontology

Group data on the "objectiv" to "subjective" spectrum
i.e. the amount of expected intersubject agreement

Tier 0: human independent reality (unknown)
Tier 1: observation of physical world (quantitative, Partial Differential Equations)
Tier 2: physical objects with properties
Tier 3: social reality
Tier 4: subjective knowledge

Outline

1. My Starting Point
   - My past research
   - Why and when are Ontology Projects Successful
2. Ontology for GIS
3. Computational Reading
   - My current interest
   - Layers of Natural Language Processing
4. Representation of Meaning
   - AMR Abstract Meaning Representation
   - Extracting Ontology from Text

Computational Comparative Literature

- Computational analysis of literary text
- Intelligent search in a literary text:

Example Task: Find all sentences in a collection of text where Animals think or talk.
E.g. fables and fairy tales.

- Analysis with Natural Language Processing and Semantic Tools required

Outline

1. My Starting Point
   - My past research
   - Why and when are Ontology Projects Successful
2. Ontology for GIS
3. Computational Reading
   - My current interest
   - Layers of Natural Language Processing
4. Representation of Meaning
   - AMR Abstract Meaning Representation
   - Extracting Ontology from Text

NLP analysis adds gradually information to a text

tokenization breaking text in words
Information added: the readers knowledge of word formation and word boundary markers

lemmatization identification of base word forms
Information added: the readers knowledge of the lexicon

POS tagging determination of Part of Speech
Information added: the readers knowledge of grammar

NLP analysis adds gradually information to a text

dependencies grouping words into meaningful units
Information added: knowledge of grammar

co-references resolve references
(e.g. what is "it" or "they" referring to?)
Information added: the readers knowledge of semantics

action patterns marked
Information added: knowledge of semantics and real world processes

Gradual analysis comparable to Tiers of Ontology?

- results of initially analysis steps are comparable between subjects
- research focus on: "Information contributed by the reader to the reading process"
- what is the "objective meaning" of a text?

Outline

1. My Starting Point
   - My past research
   - Why and when are Ontology Projects Successful
2. Ontology for GIS
3. Computational Reading
   - My current interest
   - Layers of Natural Language Processing
4. Representation of Meaning
   - AMR Abstract Meaning Representation
   - Extracting Ontology from Text

"Not an interlingua, but close" (Xue et al 2014)

- The representation of the meaning of a sentence should be independent of the language used.
- Demonstration: a text and two translations (English, Chinese and Czech) result in the same AMR representation.
- AMR is much influence by English language structures and semantics, but:
- various other research efforts to extract more semantics from text, AMR is probably the most ambitous one

Example: Where is Homer Simpson when you need him?

c. Kde je Homer Simpson, když ho potřebujete?

```
(b / byt_umisten
    :ARG0 (p / person
                :name ( h / name
                          :op1 "Homer"
                          :op2 "Simpson"))
    :ARG1 (a / amr-unknown)
    :time (p2 / potřebovat-1|need-01
                :ARG0 (v / vy|you)
                :ARG1 p))
```

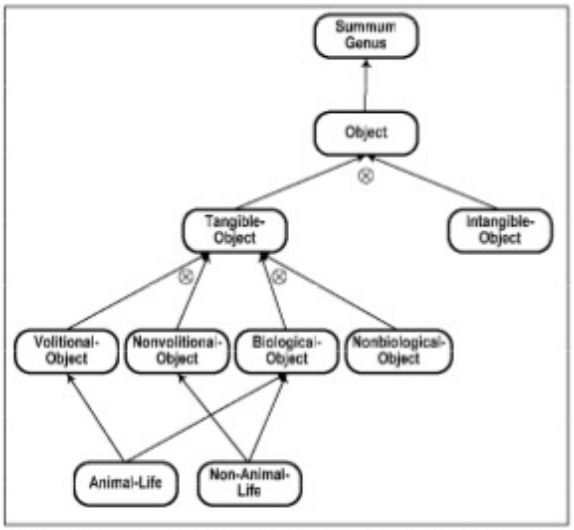
b. 当你需要他时, 霍默　辛普森在哪里?

```
(b / be-located-at
    :ARG0 (p / person
                :name (n / name
                          :op1 " 霍默　辛普森|Homer Simpson"))
    :ARG1 (a / amr-unknown)
    :time (n2 / 需要|need
                :ARG0 (y / 你|you)
                :ARG1 p))
```

(1) a. Where is Homer Simpson when you n

```
(b / be-located-at-91
    :ARG0 (p / person
                :name ( h / name
                          :op1 "Homer"
                          :op2 "Simpso
    :ARG1 (a / amr-unknown)
    :time (n / need-01
                :ARG0 (y / you)
                :ARG1 p))
```

Ontology?

- Semantic annotations for nouns are elements of a taxonomy
- Major component: Wordnet synsets
- Plans to merge with other upper level ontologies
- AMR uses the Omega Ontology (Philpot et al, 2005) a shallow, lexically oriented, term taxonomy
- Construct as a lattice over mutually disjoint concepts
- 

Outline

1. My Starting Point
   - My past research
   - Why and when are Ontology Projects Successful
2. Ontology for GIS
3. Computational Reading
   - My current interest
   - Layers of Natural Language Processing
4. Representation of Meaning
   - AMR Abstract Meaning Representation
   - Extracting Ontology from Text

Probability and Ontology

- NLP is fundamentally probabilistic because human language is polysemous, imprecise and context dependent
- Influence of context is captured with probability
- Construct ontology with probability (Ding and Peng, 2004)

Example

Trivial: Probability 50% of "x is male" if "x is an animal", but not when "x is a cow" or "x is a chicken".

A Literary Text creates a Reality

- The imaginary "reality" constructed by the author is the only reality of a fictional text.
- The ontology of the text may be different from everyday ontology as we experience it.
- The difference in the ontology may signal to the user that this is a work of fiction.
- Research question: how does the ontology represented change in time or between countires?

Large amount of Text

- Large amounts of text available.

Example

About 9000 English novels of the 19th century are in the Project Gutenberg.

- How much of a probabilistic "ontology" can be extracted automatically?

Achieved so far:

- Automatic processing to construct multi-language corpora of analyzed literary text.
- Storage of annotated text in RDF triplestore
- Query with SPARQL
- Challenges:
  - Construction of probabilistic ontology useful for NLP
  - Extraction of ontology from text
  - AMR or similar "interlingua" annotation systems.

Summary

- New research in NLP use ontologies
- NLP methods can be routinely used to process large amounts of text.
- Literary text (e.g. novels) contain much information about cultural and social situation. Computational comparative analysis can reveal important trends in time and difference between societies.
- Ontologies used for NLP and extracted from text are probabilistic (fuzzy, or rough set) based

References:

DING, Zhongli and PENG, Yun. A probabilistic extension to ontology language OWL. 2004

XUE, Nianwen, BOJAR, Ondrej, HAJIC, Jan, PALMER, Martha, URESOVA, Zdenka and ZHANG, Xiuhong. Not an Interlingua, But Close: Comparison of English AMRs to Chinese and Czech.

Banarescu, L., Bonial, C., S. Cai, M., Georgescu, Griffitt, K., Hermjakob, U., Knight, K., Koehn, P., Palmer, M. and Schneider, N. Abstract Meaning Representation for Sembanking Proc. Linguistic Annotation Workshop, 2013