Ontology Patterns for Tabular Representations of Biomedical Knowledge on Neglected Tropical Diseases

http://www.cin.ufpe.br/~ntdo

Filipe Santana, Daniel Schober, Zulma Medeiros, Fred Freitas, Stefan Schulz

Master Candidate in Computer Science
fss3@cin.ufpe.br

Introduction

Tabular representation of facts in a scientific article

Geographic location	Arthropod (Vector)	Vertebrate (Host)	Protist (Pathogen)	Manifestation (Disease)
Guadeloupe	*Lu.longipalpis*	Human	*L.chagasi*	VL
Mexico	*Lu.longipalpis*	Human	*L.chagasi*	VL
	Lu.olmeca olmeca		*L.mexicana*	CL
			L.sp	ADCL
Paraguay	*Lu.flaviscutellata*	Human	*L.amazonensis*	CL
	L.Longipalpis		*L.chagasi*	ADCL
Peru	*Lu.whitmani*	Human	*L.braziliensis*	CL
	Lu.peruensis		*L.peruviana*	ML
	Lu.verrucarum			

[Sharma & Singh, 2008]

Centro de Informática

Introduction

- **Tabular representations: common format of compact presentation of structured data in scientific publications;**
- **Table cell content denotes:**
  - Numeric content
  - Reference to **Individuals**
  - Reference to **Classes**
- **Table Semantics**
  - hidden assumptions
  - proper interpretation requires in-depth background knowledge

Introduction

- **Hypothesis:**
  - Tables (partly) contain ontological content
  - Use for semi-automated domain ontology enrichment:
  - Extraction of classes, individuals, relations, axioms
- **NTDO (Neglected Tropical Disease Ontology)**
  - Based on **BioTop** (upper-level ontology for biomedicine)
  - Subject Matter: Vector borne diseases, e.g. Filariasis, Leishmaniasis
  - Complex organisms (vectors, parasites) and lifecycles;
  - Geographic locations
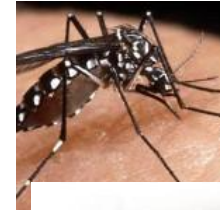 - Use case: querying epidemiological data

Centro de Informática

Geographic location	Arthropod (Vector)	Vertebrate (Host)	Protist (Pathogen)	Manifestation (Disease)
Guadeloupe	Lu.longipalpis	Human	L.chagasi	VL
Mexico	Lu.longipalpis	Human	L.chagasi	VL
	Lu.olmeca		L.mexicana	CL
	olmeca		L.sp	ADCL
Paraguay	Lu.flaviscutellata	Human	L.amazonensis	CL
	L.Longipalpis		L.chagasi	ADCL
Peru	Lu.whitmani	Human	L.braziliensis	CL
	Lu.peruensis		L.peruviana	ML
	Lu.verrucarum			

[Sharma & Singh, 2008]

Methods: Table analysis

- **Columns:**
  - *GeographicRegion*: Individuals;

  - *Vector* (Arthropod): Classes (of insects)

  - Vertebrate (Host): Classes (of patients)

  - *Pathogen* (Protist): Classes (of microorganisms)

  - *PathologicalProcess*: Classes (of diseases)
- **Cells:**
  - one or more entries
- **Rows:**
  - Each row: possible transmission pattern

Centro de Informática U·F·P·E

Geographic location	Arthropod (Vector)	Vertebrate (Host)	Protist (Pathogen)	Manifestation (Disease)
Guadeloupe	*Lu.longipalpis*	Human	*L.chagasi*	VL
Mexico	*Lu.longipalpis*	Human	*L.chagasi*	VL
	Lu.olmeca olmeca		*L.mexicana*	CL
			L.sp	ADCL
Paraguay	*Lu.flaviscutellata*	Human	*L.amazonensis*	CL
	L.Longipalpis		*L.chagasi*	ADCL
Peru	*Lu.whitmani*	Human	*L.braziliensis*	CL
	Lu.peruensis		*L.peruviana*	ML
	Lu.verrucarum			

[Sharma & Singh, 2008]

Methods: Analysis of *Pathogen Transfer Process*

- **Each instance of transmission process:**

  has a location (biotop:**hasLocus**);

  has an agent (biotop:**hasAgent**);

  has a passive participant (biotop:**hasPatient**);

Geographic location	Arthropod (Vector)	Vertebrate (Host)	Protist (Pathogen)	Manifestation (Disease)
$G_{a1}G_{a2}\ldots G_{ak}$	$V_{a1}V_{a2}\ldots V_{aj}$	$H_{a1}H_{a2}\ldots H_{am}$	$P_{a1}P_{a2}\ldots P_{al}$	$D_{a1}D_{a2}\ldots D_{an}$
$G_{b1}G_{b2}\ldots G_{bk}$	$V_{b1}V_{b2}\ldots V_{bj}$	$H_{b1}H_{b2}\ldots H_{bm}$	$P_{b1}P_{b2}\ldots P_{bl}$	$D_{b1}D_{b2}\ldots D_{bn}$
…	…	…	…	…
$G_{n1}G_{n2}\ldots G_{nk}$	$V_{n1}V_{n2}\ldots V_{nj}$	$H_{n1}H_{n2}\ldots H_{nm}$	$P_{n1}P_{n2}\ldots P_{nl}$	$D_{n1}D_{n2}\ldots D_{nn}$
…	…	…	…	…
$G_{z1}G_{z2}\ldots G_{zk}$	$V_{z1}V_{z2}\ldots V_{zj}$	$H_{z1}H_{z2}\ldots H_{zm}$	$P_{z1}P_{z2}..P_{zl}$	$D_{z1}D_{z2}\ldots D_{zn}$

General Pathogen Transfer Pattern

PathogenTransferByVector equivalentTo
  PathogenTranferByVector$_a$ or
  PathogenTranferByVector$_b$ or ... or
  PathogenTranferByVector$_n$ or ... or
  PathogenTranferByVector$_z$

PathogenTransferByVector$_n$ equivalentTo *Transfer* and
  (**hasAgent** some (V_{n1} *or* V_{n2} *or...or* V_{nj})) and
  (**hasLocus** some (G_loc_{n1} or G_loc_{n2}
         or ...or G_loc_{nk})) and
  (**hasPatient** some ((P_{n1} or P_{n2} or...or P_{nl}) and
    (**hasLocus** some (H_{n1} or H_{n2} or...or H_{nm})

PathogenTransferByVector$_n$ subClassOf *Transfer* and
  (**hasAgent** only (V_{n1} or V_{n2} or...V_{nj})) and
  (**hasLocus** only ((not *GeographicLocation*)
    or G_loc_{n1} or G_loc_{n2} or... G_loc_{nk})) and
  (**hasPatient** only ((P_{n1} or P_{n2} or...or P_{nl}) and
    (**causes** only (D_{n1} or D_{n2} or...or D_{nr}))))

Centro
de Informática

Results
Extraction and Ontology Construction

- **Transfer tabular content from publication into Excel spreadsheet**

Geo	Vector	Host	Pathogen	Manifestation
Guadeloupe	LutzomyiaLongipalpis	Human	LeishmaniaChagasi	VisceralLeishmaniasis
Mexico	LutzomyiaLongipalpis LutzomyiaOlmecaOlmeca	Human	LeishmaniaChagasi LeishmaniaMexicana LeishmaniaSp	VisceralLeishmaniasis CutaneousLeishmaniasis AcuteDiffuseCutaneousLeishmaniasis
Paraguay	LutzomyiaFlaviscutellata LutzomyiaLongipalpis	Human	LeishmaniaAmazonensis LeishmaniaChagasi	CutaneousLeishmaniasis AcuteDiffuseCutaneousLeishmaniasis VisceralLeishmaniasis
Peru	LutzomyiaWhitmani LutzomyiaPeruensis LutzomyiaVerrucarum	Human	LeishmaniaBraziliensis LeishmaniaPeruviana	CutaneousLeishmaniasis MucocutaneousLeishmaniasis

- **Creating OWL code using VBA scripting**

Results
Additional Axioms (manually inserted)

- **Reified Geographic Locations (G_Loc_i):**

    G_Loc_i equivalentTo

             GeographicLocation and **hasLocus** value G_i

- **Dispositions and Manifestations:**

    D_disp_i equivalentTo *PathologicalDisposition and*

             (**hasRealization** only D_i)

    D_i equivalentTo *PathologicalProcess* and

             (**realizationOf** some D_disp_i)

    D_i subClassOf **realizationOf** only D_disp_i

Results
Competency Questions

- **What pathogen can be transmitted by a given vector in a geographic location?**

Query:

Query (class expression)

Protist **and not** (patientIn **some** (PathogenTransferByVector **and** hasLocus **some** (GuadeloupeLocation **and** hasAgent **only** LutzomyiaLongipalpis)))

Execute   Add to ontology

Query results

Sub classes (5)

- LeishmaniaAmazonensis
- LeishmaniaBraziliensis
- LeishmaniaMexicana
- LeishmaniaPeruviana
- LeishmaniaSp

☐ Super classes
☐ Ancestor classes
☐ Equivalent classes
☑ Subclasses
☐ Descendant classes
☐ Individuals

Query:

Query (class expression)

Protist **and not** (LeishmaniaAmazonensis **or** LeishmaniaBraziliensis **or** LeishmaniaMexicana **or** LeishmaniaPeruviana **or** LeishmaniaSp)

Execute   Add to ontology

Query results

Sub classes (1)

- LeishmaniaChagasi

☐ Super classes
☐ Ancestor classes
☐ Equivalent classes
☑ Subclasses
☐ Descendant classes
☐ Individuals

Results
Competency Questions

- **Can disease X be transmitted by some vector in a given geographic location Z?**

Query:

Query (class expression)

```
PathogenTransferByVector and
(hasLocus some MexicoLocation) and
(hasPatient some (Protist and
causes some MucocutaneousLeishmaniasis))
```

[Execute] [Add to ontology]

Query results

Equivalent class

● Nothing   (?)

Sub classes (0)

☐ Super classes
☐ Ancestor classes
☑ Equivalent classes
☑ Subclasses
☐ Descendant classes
☐ Individuals

Centro de Informática

Results
Competency Questions

■ **What kind of disease can be transmitted in a given geographic location ?**

Query:

Query (class expression)

PathologicalProcess and not (causedBy some (Protist and (patientIn some (PathogenTransferByVector and hasLocus some PeruLocation))))

[Execute] [Add to ontology]

Query results

Sub classes (2)

● **AcuteDiffuseCutaneousLeishmaniasis**

● **VisceralLeishmaniasis**

☐ Super classes
☐ Ancestor classes
☐ Equivalent classes
☑ Subclasses

Query:

Query (class expression)

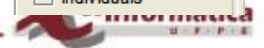PathologicalProcess and not (AcuteDiffuseCutaneousLeishmaniasis or VisceralLeishmaniasis)

[Execute] [Add to ontology]

Query results

Sub classes (2)

● **CutaneousLeishmaniasis**

● **MucocutaneousLeishmaniasis**

☐ Super classes
☐ Ancestor classes
☐ Equivalent classes
☑ Subclasses
☐ Descendant classes
☐ Individuals

Conclusions

- **Legacy information extracted from tables can be transformed in a formal ontology**

- **OWL-DL sufficient to represent moderately complex biological situations;**

- **Dependency on principled ontological foundations:**
  - rigid set of object properties from *BioTop*;
  - closed-world interpretation of the table content

- **DL queries instead of RDF-based SPARQL queries:**
  - using simple DL Manchester syntax
  - supports reasoning about taxonomies and transitive relations

Centro
de Informática
U · F · P · E

Conclusions

- **Querying:**
  - Subclass retrieval and satisfiability testing;
  - Broad use of OWL *DL* constructors, i.e. disjunctions, negations, value restrictions and complex class definitions;
- **Limitations:**
  - Numeric content (rudimentary represented by OWL Data Properties), probabilistic associations; default expressions
  - Scalability of ontologies in expressive DL
  - Tools for Complex DL Queries: Supported by Protégé, but could be improved for end users;

Centro de Informática

Acknowledgements

- **This work was supported by:**

Thanks
for your attention!!

Filipe Santana
Bsc Biology
Msc Candidate in Computer Science
fss3@cin.ufpe.br

For additional information,
http://www.cin.ufpe.br/~ntdo