Pattern-based Ontology Modeling

Pascal Hitzler
Data Semantics Laboratory (DaSe Lab)
Data Science and Security Cluster (DSSC)
Wright State University
http://www.pascal-hitzler.de
Semantic Web journal

- EiCs: Pascal Hitzler
  Krzysztof Janowicz
- Funded 2010
- 2016 Impact factor of 1.786, top of all journals with “Web” in the title

- We very much welcome contributions at the “rim” of traditional Semantic Web research – e.g., work which is strongly inspired by a different field.
- Non-standard (open & transparent) review process.

- http://www.semantic-web-journal.net/
Data Semantics Laboratory

Two faculty (Michelle Cheatham, Pascal Hitzler)
1 postdoc (Adila Krisnadhi)
9 PhD students
4 Master students

See http://daselab.org/

Data Science and Security Cluster (DSSC):

Combines 5 labs (data semantics, bioinformatics, visualization, cybersecurity, web and complex systems), 7 faculty
Motivation: EarthCube and GeoLink
EarthCube GeoLink Scenario

The NSF EarthCube Program:
Developing a Community-Driven Data and Knowledge Environment for the Geosciences

“concepts and approaches to create integrated data management infrastructures across the Geosciences.”

“EarthCube aims to create a well-connected and facile environment to share data and knowledge in an open, transparent, and inclusive manner, thus accelerating our ability to understand and predict the Earth system.”
GeoLink: An EarthCube “Building Block” project (2014-2017)

How to realize data search across many large-scale geoscience data repositories, such that

- The approach is extendable to new repositories.
- The scope can extend across all of the Geosciences.
- The search capabilities can be made more fine-grained in the future if desired.

Central idea: Use a modular, extendable ontology for the integration of metadata.
The GeoLink Framework

GeoLink Ontology Patterns

- isAssociatedWithCruise, isResultOf
- Publication
  - isAssociatedWithDataset
  - Dataset
    - isFundedBy
    - Digital Object
    - Program
    - Funding Award
  - pAR
  - isPartOf
  - pAR
  - isPerformedBy
    - pAR
  - Property
    - hasProperty
  - Physical Sample
    - derivedUsing
    - hasProperty
  - Sampling Process
    - pAR
  - Event
    - occursAtPlace
    - hasSpatialFootprint
    - Geometry
  - Place
    - isAssociatedWithCruise
  - Cruise
    - isAssociatedWithCruise
    - isUndertakenBy
      - Vessel
      - hasProperty
  - Organization
    - sCO
      - Agent
        - sCO
          - Person
          - isFundedBy
          - pAR
            - providesAgentRole
          - sCO
            - subClassOf
Modeling Patterns
ABoxes as graphs

SesameStreet has Actor JimHenson.
JimHenson hasName “Jim Henson”.

Diagram:

- **SesameStreet** hasActor **JimHenson** hasName **"Jim Henson"**
Problem!

SesameStreet has Actor JimHenson.
MuppetShow has Actor JimHenson.
JimHenson plays Kermit.
JimHenson plays Ernie.
JimHenson hasName “Jim Henson”.
Solution!
Schematically

1. **SesameStreet** → **hasActor** → **JimHenson** → **hasName** → "Jim Henson"

2. **Movie** → **hasActor** → **Person** → **hasName** → “Jim Henson”

3. **Movie** → **hasActor** → **ActorRole** → **performedBy** → **Person**
   - **Character** → **plays** → **xsd:string**
   - **hasName**
An **Ontology Design Pattern** (ODP) is a reusable successful solution to a recurrent ontology modeling problem.

[Gangemi 2005]
Axiomatization

\[ T \subseteq \forall\text{providesAgentRole}.\text{AgentRole} \]
\[ \text{AgentRole} \subseteq \forall\text{performedBy}.\text{Agent} \]
\[ \exists\text{performedBy}.\text{Agent} \subseteq \text{AgentRole} \]
\[ \text{AgentRole} \subseteq \forall\text{startsAtTime}.\text{TimelInstant} \]
\[ \text{AgentRole} \subseteq \forall\text{endsAtTime}.\text{TimelInstant} \]
\[ \text{AgentRole} \subseteq \exists\text{providesAgentRole}^{-}.T \]
\[ \text{AgentRole} \subseteq =1\text{performedBy}.\text{Agent} \]
\[ \text{AgentRole} \subseteq =1\text{startsAtTime}.\text{TimelInstant} \]
\[ \text{AgentRole} \subseteq =1\text{endsAtTime}.\text{TimelInstant} \]
\[ \text{DisjointClasses}(\text{AgentRole}, \text{Agent}, \text{TimelInstant}) \]
A Minimal Event Pattern

\[ T \sqsubseteq \forall \text{hasSpatioTemporalExtent}.\text{SpatioTemporalExtent} \]

\[ T \sqsubseteq \forall \text{providesAgentRole}.\text{AgentRole} \]

\[ \text{Event} \sqsubseteq \exists \text{hasSpatioTemporalExtent}.\text{SpatioTemporalExtent} \]

\[ \text{Event} \sqsubseteq \forall \text{subEventOf}.\text{Event} \]

\[ \exists \text{subEventOf}.\text{Event} \sqsubseteq \text{Event} \]

\[ \text{subEventOf} \circ \text{subEventOf} \sqsubseteq \text{subEventOf} \]

\[ \text{DisjointClasses} (\text{Event, AgentRole, SpatioTemporalExtent}) \]
Worked Example: Chess
Worked Example: Chess

• Establish a searchable repository for chess data.

• Starting point are PGN files.
• Should be extendable with other information from
  – Chess websites
  – Wikipedia
  – Geographic data
  – News
  – Etc.

• Use an ontology for information integration.
GeoVoCamps modeling approach

- Collaborative modeling, group ideally has
  - More than one domain experts.
  - People familiar with the base data.
  - People understanding possible target use cases.
  - An ontology engineer familiar with the modeling approach.
  - Somebody who understands formal semantics of OWL.

- Domain experts are queried as to the main notions for the application domain.
  - E.g. for chess, these would include
    - Chess game; move; opening; tournament; players; commentary
GeoVoCamps modeling approach

- From available data and from application use cases, devise competency questions, i.e. questions which should be convertible into queries, which in turn should be answerable using the data.

1. Who played against Kasparov in the round 1994 Lineares tournament? Did (s)he play as a white or black player?
2. What is the first move taken by the black player in the Sicilian Defense opening?
3. Find all games in which Bobby Fischer, playing black, lost in the poisoned pawn variation of the Sicilian Defence opening.
4. Are there any recorded games using the Grünfeld Defence from before the 20th century?
5. What did Kasparov say about his opponent’s first two moves in his commentary about his game against Topalov in the 1999 Tournament in Wijk ann Zee?
6. Who was the first non-Russian world champion after Fischer?
7. Did Bobby Fischer ever play against a grandmaster in Germany?
8. List all world championship games won by forfeit.
Then prioritize which notions to model first. In the chess case, e.g.

- chess game
- move/half-move
- players
- opening
- tournaments
- commentary
GeoVoCamps modeling approach

- Understand the nature of the things you are modeling.

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Chess game</td>
<td>...</td>
<td>An Event</td>
</tr>
<tr>
<td>Half-move</td>
<td>...</td>
<td>A Subevent of a chess game</td>
</tr>
<tr>
<td>Player</td>
<td>...</td>
<td>The Role of an Agent</td>
</tr>
<tr>
<td>Opening</td>
<td>...</td>
<td>this is probably complex</td>
</tr>
<tr>
<td>tournaments</td>
<td>...</td>
<td>Events</td>
</tr>
<tr>
<td>commentary</td>
<td>...</td>
<td>this is again more complex</td>
</tr>
</tbody>
</table>
Player as AgentRole

owl:Thing → AgentRole
  providesAgentRole

AgentRole → TimeInstant
  performedBy
  startsAtTime, endsAtTime

BlackPlayerRole rdfs:subClassOf AgentRole

WhitePlayerRole rdfs:subClassOf AgentRole

ChessGame providesAgentRole

AgentRole performedBy

Agent
ChessGame as Event

- TemporalExtent
  - atTime
- Event
  - subEventOf
  - atTime
  - atPlace
- Place
- AgentRole
  - providesAgentRole
  - performedBy
- Agent

- TemporalExtent
  - atTime
- ChessGame
  - providesAgentRole
  - atPlace
- Place
Half-moves
Opening, game result, etc.

We call these “stubs”.

I.e. we’re aware that more fine-grained modeling will be needed for some use cases.

But currently there’s no reason to do it (not in use case, no data), so we only provide “hooks” for future development of the ontology.
Putting things together
Adding commentaries
Adequacy check

• Triplify sample data using the ontology. Does it work?

• Check if competency questions can be answered.

• Add axioms as appropriate (the graph is only for intuition, the OWL axioms are the actual ontology).

• (there are more post-hoc details to be taken care of, but let’s leave it at that)
Axioms in this case are mostly straightforward:

- Inherited from Event or AgentRole
- Scoped domain/range restrictions, possibly with some cardinalities
- Basic existentials
- Non-cyclicity of half-move sequence

What about adding, e.g., the following?

\[ \text{ChessGame} \sqsubseteq \geq 0\text{subEventOf.CheesTournament} \]

If one of the roles of axiomatization is to improve human understanding of the ontology, then such axioms are helpful!
Shortcuts and Views
Shortcuts

\[
\text{ChessGame}(x) \land \text{pAR}(x, y) \land \text{WhitePlayerRole}(y) \land \text{performedBy}(y, z) \\
\land \text{Agent}(z) \land \text{hasName}(z, s) \rightarrow \text{hasWhitePlayer}(x, s) \\
\text{ChessGame}(x) \land \text{pAR}(x, y) \land \text{BlackPlayerRole}(y) \land \text{performedBy}(y, z) \\
\land \text{Agent}(z) \land \text{hasName}(z, s) \rightarrow \text{hasBlackPlayer}(x, s)
\]
Translating the rules

$$\begin{align*}
\text{ChessGame}(x) \land pAR(x, y) \land \text{WhitePlayerRole}(y) \land \text{performedBy}(y, z) \\
\land \text{Agent}(z) \land \text{hasName}(z, s) \rightarrow \text{hasWhitePlayer}(x, s)
\end{align*}$$

$$\begin{align*}
\text{ChessGame}(x) \land pAR(x, y) \land \text{BlackPlayerRole}(y) \land \text{performedBy}(y, z) \\
\land \text{Agent}(z) \land \text{hasName}(z, s) \rightarrow \text{hasBlackPlayer}(x, s)
\end{align*}$$

However note that the introduction of additional role chains may cause violations of regularity restrictions.
We used rules (axioms) to express the mapping from the ontology to the view.

The reverse direction is much more tricky.

\[
\text{ClassA}(x) \land \text{ClassB}(y) \land C_1(x_1) \land \cdots \land C_n(x_n) \land R_1(y_1, y_2) \land \cdots \land R_k(y_k, y_{k+1}) \\
\rightarrow \text{shortcut}(x, y).
\]

\[
\text{shortcut}(x, y) \rightarrow \text{ClassA}(x) \land \text{ClassB}(y) \land \exists x_1 \ldots \exists x_n \exists y_1 \ldots \exists y_n (C_1(x_1) \land \ldots \\
\cdots \land C_n(x_n) \land R_1(y_1, y_2) \land \cdots \land R_k(y_k, y_{k+1}))
\]
Mapping from views

Existential rules may be suitable in principle.

\[ \text{shortcut}(x, y) \rightarrow \text{ClassA}(x) \land \text{ClassB}(y) \land \exists x_1 \ldots \exists x_n \exists y_1 \ldots \exists y_n (C_1(x_1) \land \ldots \land C_n(x_n) \land R_1(y_1, y_2) \land \ldots \land R_k(y_k, y_{k+1})) \]

However automated reasoning with the potentially rather complex rule heads requires investigations, in particular if it is to be integrated with ontology reasoning.

A specific case are right-hand-side role chains:

\[ R \sqsubseteq R_1 \circ \cdots \circ R_n, \]
The GeoLink Modular Ontology (GMO)
The GeoLink Modular Ontology

High-level overview of the GeoLink Modular Ontology (GMO). Each box stands for a module, which has been modeled in its own right.
Example Module: Cruise

Cruise reused e.g. the generic patterns
- AgentRole
- Trajectory
and conceptually cruises are understood to be events.
An (preliminary) interactive demonstration of the integrated GeoLink data is available at

http://demo.geolink.org

At http://www.geolink.org/ there are links to the complete schema, a SPARQL Endpoint, publications, etc.
Several W3C recommendations are relevant in our context:

- The Web Ontology Language OWL for expressing ontologies and sharing them on the Web.

- The Resource Description Framework RDF for expressing e.g. ABox graphs and aligning them to an ontology.

- The SPARQL RDF query language.

(See the references for pointers.)
Thanks!


References


Adila Krisnadhi, Ontology Pattern-Based Data Integration. Dissertation, Department of Computer Science and Engineering, Wright State University, 2015.


Steve Harris, Andy Seaborne, SPARQL 1.1 Query Language. W3C Recommendation 21 March 2013.