

Weaving the New Testament into the Semantic Web

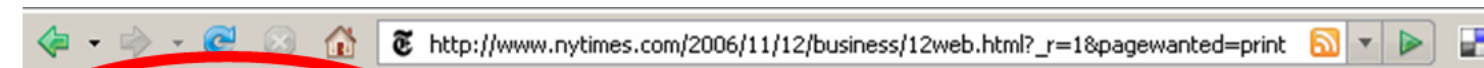
Sean Boisen, <[myfirstname]@semanticbible.org>

Slides at <http://www.semanticbible.org/other/presentations/2006-sbl/Weaving.xhtml>

Society for Biblical Literature Annual Meeting, November 20 2006

Outline

- An introduction to the Semantic Web
- Why Biblical scholars should care
- New Testament Names (NTN): a semantic knowledgebase
- Going further



The New York Times
nytimes.com

November 12, 2006

Entrepreneurs See a Web Guided by Common Sense

By [JOHN MARKOFF](#)

SAN FRANCISCO Nov. 11 — From the billions of documents that form the World Wide Web and the computer scientists and a growing collection of start-up companies are finding new ways to mine hu

Their goal is to add a layer of meaning on top of the existing Web that would make it less of a catalog the foundation for systems that can reason in a human fashion. That level of artificial intelligence, w simply following commands, has eluded researchers for more than half a century.

Referred to as Web 3.0, the effort is in its infancy, and the very idea has given rise to skeptics who b the underlying technologies are rapidly gaining adherents, at big companies like [I.B.M.](#) and [Google](#) a center on simple, practical uses, from producing vacation recommendations to predicting the next h

But in the future, more powerful systems could act as personal advisers in areas as diverse as financ mapping out a retirement plan for a couple, for instance, or educational consulting, with the Web hel right college.

Key Concepts for the Semantic Web

- Principal goal: a *web of data* that allows machines to "connect the dots" and share across application (and organizational) boundaries
 - Other uses: browsing, visualization, inference, statistical analysis ...

-
- Everything has a **URI**
- The RDF triple is a primary data element
- RDF/XML standardizes syntax: OWL standardizes semantics to enable data interoperability
 - Application-specific semantics are not enough



- Just motivating the Semantic Web is a larger topic than i have time for today!
- Application-specific semantics is a common current approach: write specific code to interpret your data. But it means each application must define its own semantics, which defeats reusability, and makes discovery impossible. OWL lets you specify the semantics in a standard fashion.
- Note the focus is not on human readability, but machine access

Web Ontology Language (OWL)

- OWL has been an official W3C recommendation since 2004. It starts from XML and RDF and adds
 - *Ontology specification*
 - *Property characteristics (transitivity, symmetry, functionality)*
 - *Property restrictions (cardinality, value restrictions)*
- An ontology is a representation of a conceptualization (Gruber 1993)
- See the OWL Guide at <http://www.w3.org/TR/owl-guide>
- Beyond OWL: reasoning, logic, ...

The Semantic Web Process

- Model your data (conceptualization)
- Develop an OWL ontology that reflects that model (representation), reusing existing ontologies where they fit
- Populate instance data (preferably by mapping existing data sources)
- Connect to other data
- Deploy "added value" tools like visualization, inference, and validation

Semantic Web and Text Markup

- Ontologies describe what *is*: real things, not documents
 - XML for text markup, OWL for "knowledge markup"
 - *XML can still connect to the Semantic Web through `rdf:ID`*
 - *Topic Maps are perhaps somewhere between XML and OWL*
 - Folksonomic tags (del.icio.us, Technorati) are a different bottom-up alternative
-
- Many similar motivations to OSIS: standards enable reuse, new applications and approaches, and reduce data silos
-

Semantic Web Tools

- Editors: Protege, SemanticWorks, others
- Triple Stores: Sesame, Oracle Spatial 10g, Joseki, Tucana
- Programming libraries for Java (Jena), Perl, Python, etc.
- Longwell is useful for browsing RDF
- SemWeb search through [Swoogle](#): currently > 300M triples
- Also reasoners, validators ...
- W3C has a much more comprehensive list at esw.w3.org/topic/SemanticWebTools

Example: Geodata Integration

- [Google Earth data for Bible Places](#) (John Philip and others) provides latitude, longitude, and altitude in Keyhole Markup Language (KML)

```
<Placemark>
<name>Antipatris</name> ...
<LookAt>
  <longitude>34.954...</longitude>
  <latitude>32.1068...</latitude>
```

Google Earth Visualization



- "... the next day we touched at Samos; and the day after that we went to Miletus. For Paul had decided to sail past Ephesus ...". (Ac 20:15-16, ESV)

The Integration Problem

- Variation in the <name> element
 - “Ashtaroth, Astaroth”
 - “Akkad (Accad)”
 - “Acco / Ptolemais”
 - “Antioch” (Syria), “Antioch” (Pisidia)
 - “Ar / Ar of Moab”
 - “Bethany”, “Bethany Beyond the Jordan”
- Type variation: “Sea of Galilee”, “Mount of Olives”, “Solomon's Porch”
- How can we form semantic connections to this data?

A Semantic Web Solution

- Give resources a URI like

```
http://semanticbible.org/ns/2006/NTNames.owl#Antioch
```

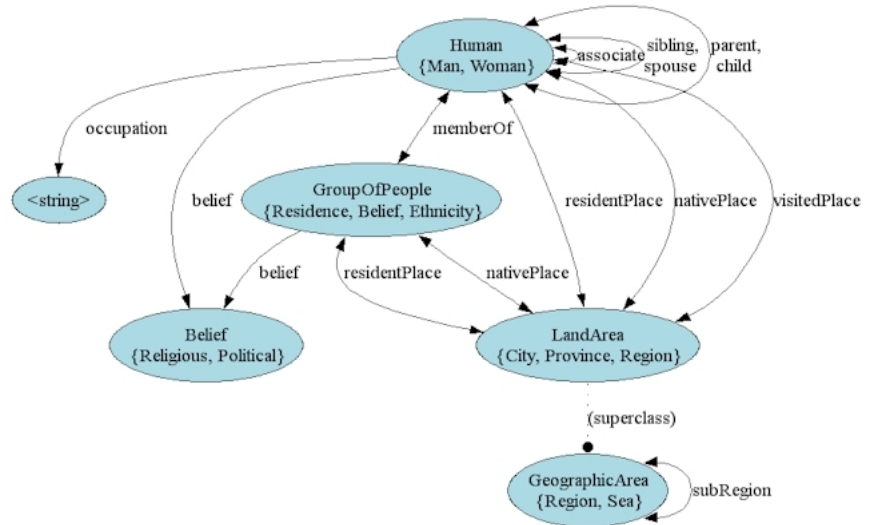
- Every resource has a *type*:

```
<http://semanticbible.org/ns/2006/NTNames#Antioch>  
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>  
<http://semanticbible.org/ns/2006/NTNames#City> .
```

- Use `rdfs:label` to capture alternate names and spellings
 - Use `owl:sameAs` to connect resources across ontologies
-

New Testament Names: a Semantic Knowledgebase

- An instance for each unique entity named in the New Testament
 - Provides unambiguous identifiers
- Provides an initial reference ontology with instances
 - 48 classes, 38 properties
 - 647 instances, 4200 RDF triples
- The content is in the identifiers and their relationships, rather than prose
- This is not anything like a complete ontology of objects in the NT (an enormous undertaking!): initially limited to named objects, to keep the effort manageable and grounded in concrete, unambiguous things (e.g. not theology).

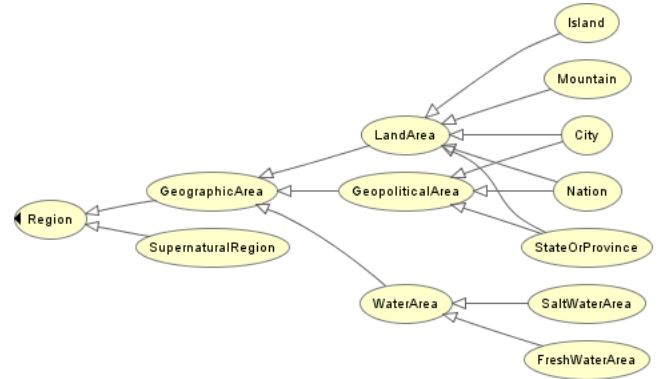


New Testament Names (2)

- Predominant classes:
 - *Man* (330 instances)
 - *Woman* (46 instances)
 - *City* (90 instances)
 - NTN is the first and (so far) only publically available Bible data in OWL
 - *First version released 2004-04*
 - The ontology and instance data are freely available
 - *Ontology:* <http://www.semanticbible.org/2006/11/NTNames.owl>
 - *Instances:* <http://www.semanticbible.org/2006/11/NTN-individuals.owl> (the instance data).
-

New Testament Names: Classes

- Classes are defined consistently with existing ontologies where possible
 - *A comment on ontology reuse*
- OWL supports multiple inheritance



New Testament Names: Properties

- Humans have attributes like `description`, `occupation`, `name_en`, `name_gk`
 - Properties between Humans and
 - *Humans*:
 - family relations (parent/child, spouse, sibling)
 - inter-personal relations: collaborates with, antagonist of, enemy of, knows
 - *Locations*: *native place*, *resident place*, *visited place*
 - Religious belief, ethnicity, group membership
 - Locations have latitude/longitude, spatial inclusion
-

New Testament Names: Instances

- URIs disambiguate Alexander (5), Joseph (4), James (3), Mary (6), etc.
- Principles:
 - *Limit to what's explicit in the text*
 - *Conservative grouping*
- Browser demonstration using Longwell: Cities in Macedonia

An Integration Example: Connecting the Composite Gospel and NTN

- Pericope.121: Herod kills John the Baptist
- Add properties for textual mentions of Herod, Herodias, John, Philip (and others who aren't named) using NTN identifiers
- Now NTN data is available to enrich a display or other application

Extending NTN

- Other relevant entities that are either
 - *mentioned in the text without names*
 - *relevant to the context of New Testament study*
 - Add Bible references as properties; annotate textual mentions with NTN URIs
 - Incorporate other reference sources (e.g. [Wikipedia's category for New Testament people](#))
 - Incorporate events and time (support timeline visualization)
 - Connect to specific vocabularies (theological terms, historical/cultural background)
-

Other OWL Applications for Bible scholars

- OSIS is a complementary step toward standardizing text markup
- Modeling the life of Jesus:
 - *Composite Gospel (RDF)* describes relationships between textual units of the Gospels
 - *Events of His life*
 - *Concepts that He taught*
- Authorship information (the Biblical canon, Ante-Nicene Fathers, etc.)
- Lexical semantics (could we integrate [Wordnet](#) and Louw-Nida?)
- Topical markup

What else besides New Testament Names?

- Many similar motivations to OSIS: standards enable reuse, new tools, and new approaches, and reduce data silos

Some Possible Objections (*and Responses*)

- Requires black-and-white judgments about inherently "gray" matters
 - *The ontologist is responsible for the model*
- Requires too much effort and technical savvy
 - *Tools are improving*
 - *Opportunistic beginnings provide important foundations*
- Limited present value
 - *Standards are always forward-looking*
 - *Agreement is easier before competing standards become entrenched*

An Invitation to Broader Participation

- The primary data are open to enrichment and extension
- [Metcalf's Law](#) applies here too
- Biblical scholarship would benefit from greater web-enablement: this requires semantic standardization

References

1. "Why use OWL?" Presentation by Roger L. Costello and David B. Jacobs (MITRE).
2. "Semantic Web: Building on what exists", by Tim Berners-Lee (W3C), <http://www.w3.org/2006/Talks/0404-mit-tbl>
3. "Semantic Web concepts", by Tim Berners-Lee (W3C), <http://www.w3.org/2005/Talks/0517-boit-tbl>
4. "Questions (and Answers) on the Semantic Web", by Ivan Herman (W3C), <http://www.w3.org/2006/Talks/0927-Berlin-IH>
5. "A translation approach to portable ontologies", by T. R. Gruber, *Knowledge Acquisition*, 5(2):199-200, 1993.