SemanticBible Documentation

1. About SemanticBible

1.1. SemanticBible is ...

- The New Testament Hyper-Concordance: an experiment in word-based navigation.
- The Composite Gospel Index: the four Gospel accounts of the life of Jesus in a unified XML-based reference.
- New Testament Names: a semantic knowledge base classifying and describing properties for about 600 names from the New Testament.
- Link to material outside the SemanticBIble site
  Blogos: a weblog with personal reflections on God's Word | our words | meaning, communication, & technology | following Jesus, the Word made flesh.

Everything here is freely available for you to use, download, and share (see License for additional details).

RDF meta-data for SemanticBible.

RDF

1.2. Contribute

SemanticBible is a volunteer effort in the spirit of the open source software development community. Everything here is freely available for you to use, download, and share (see License for additional details).

There's plenty to do if you'd like to contribute your efforts! Send an email describing what kind of help you can offer.

2. Hyper-concordance

2.1. New Testament Hyper-Concordance

The Hyper-Concordance is an experiment in word-based navigation through the New Testament. Each content word is hyperlinked to a page displaying all the verses for that word, preserving context and tightly connecting the content.

Outside link

Jump in through the word index

Outside link

Detailed description

2.2. Hyper-Concordance Code

You can download the code and build your own hyper-concordance, either for local use or to make available to others on the Web.

Space requirements

The program generates static HTML pages for each word. The resulting files require approximately 30Mb of disk space.

The Hyper-Concordance code includes two Perl files (one program and one module), and two data files. You need to supply a New Testament encoded according to the

Outside link

OSIS 1.5 standard. The version available here uses the

Outside link

Revised Standard Version available from the Bible Technologies Group and Crosswire.org. Other versions are available as well.

OSIS version

I have not tried this code with newer versions of the OSIS standard, though i expect to do so as more OSIS-encoded texts become available.

Language

The code and data files assume an English translation. However, there's nothing about the general approach that is language-specific, and i plan to release other versions in the future.

Outside link

Download a .zip file with all the code.

Outside link

List of "stopwords" that are omited from the Hyper-concordance for space purposes.

Outside link

List that maps inflected forms back to their baseforms.

2.3. Frequently Asked Questions

2.3.1. Questions

1. **About the Hyper-Concordance**
   - What Bible translation is used, and why?
   - Can I start from a passage, rather than from a word?
2. **Bugs and Suspicions**
   - Why isn't there an entry for "heaven"/"voice"/"light"/some other word?
   - When I clicked on the word "loved", I got the page for "love". Is this a bug?
   - How many chapters are in the book of James?? You have a reference to James 13:16 ...

2.3.2. Answers

2.3.2.1. 1. About the Hyper-Concordance

1.1. What Bible translation is used, and why?

The Revised Standard Version (RSV) met several key requirements. It is relatively modern and readable (unlike the KJV), but has been made available free of copyright restrictions. Also, the RSV was available in OSIS format, which is likely to become a new standard for formatting Bible texts.

1.2. Can I start from a passage, rather than from a word?

Right now all the navigation is by words. I might add this feature in a later version.

2.3.2.2. 2. Bugs and Suspicions

2.1. Why isn't there an entry for "heaven"/"voice"/"light"/some other word?

The most frequent words have been omitted, particularly function words with little semantic content like pronouns and prepositions. There's a tradeoff between completeness and the space required for the Hyper-concordance: i felt these words wouldn't be missed, and it saves some space.

The complete list of omitted words is in the file
Outside link
stopwords.txt, which is included in the code.

2.2. When I clicked on the word "loved", I got the page for "love". Is this a bug?

No, it's a feature. Plural nouns and Inflected verbs are mapped to their root forms, under the assumption that you if you're looking for 'love' you want to find "loves", "loved", and "loving" as well.

2.3. How many chapters are in the book of James?? You have a reference to James 13:16 ...

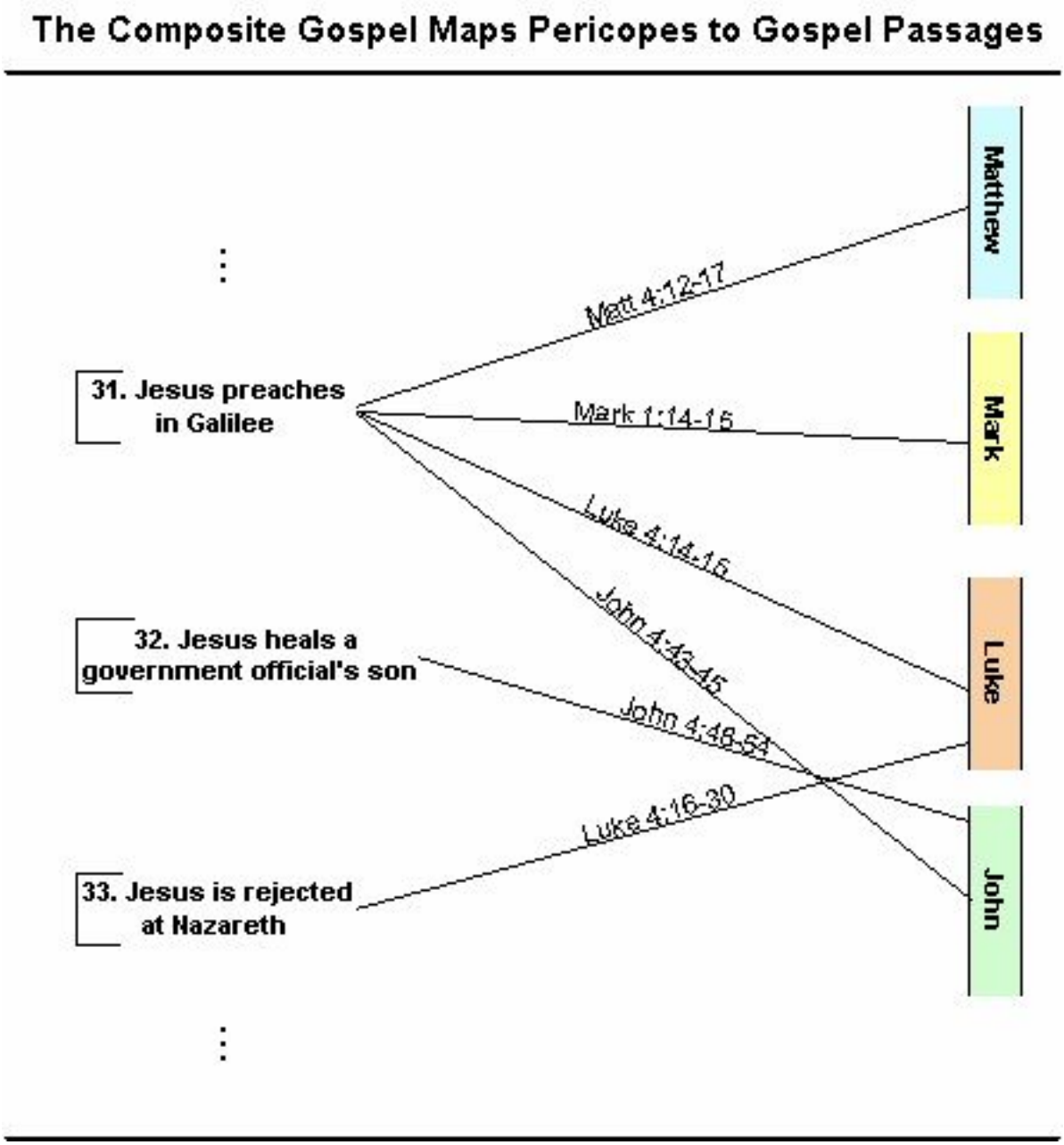
This is a bug in the text i started with which will be corrected in v1.2. There are also missing links for about a dozen words on some specific pages: "have", "go", "do", "eat", "heaven", "will", "one", "voice", "their", "feet", "our", "being", "took". The next version wil also correct this bug. But any word in the master index should linked correctly.

3. Composite Gospel

3.1. The Composite Gospel Index

3.1.1. A New Look at the Story of Jesus

The Composite Gospel Index (CGI) combines the four Gospel accounts of the life of Jesus into a single unified view. Instead of the traditional book/chapter/verse organization, it divides the texts into about 350 *pericopes* (puh-RIH-kuh-pee), each of which describe an event, a teaching, a parable, an interaction, or some other cohesive piece of text. Each pericope has a unique numeric identifier along with a brief descriptive label (e.g. #237, "Jesus speaks to a rich young ruler"), and each is indexed to one or more Gospel passages that it represents, as illustrated here. Every verse of every Gospel is represented by one and only one pericope: no verses are omitted or duplicated.



The Composite Gospel Index maps pericopes to Gospel passages

Attempts to harmonize or unify the Gospels into a single account go back at least as far as Tatian's Diatessaron in the 2nd century. The goal of the Composite Gospel Index is primarily

to bring together accounts which are common to one or more Gospels, and provide a natural arrangement of the meaningful units of the events, settings, and teachings of Jesus. It does not particularly aim to reconcile the different accounts of the Gospel authors, or provide a more chronological order of events.

3.1.2. An XML Index

An additional distinguishing feature is that the CGI is expressed in XML, using OSIS identifiers. That means the CGI supports automated processing: each pericope can be mapped onto its source verses (given OSIS-encoded texts). For example, i've written an example program that retrieves the source texts for a pericope from the ESV web service (here's some background information) and presents them in an HTML table. in the 2nd century. The goal of the Composite Gospel Index is primarily to bring together accounts which are common to one or more Gospels, and provide a natural arrangement of the meaningful units of the events, settings, and teachings of Jesus. It does not particularly aim to reconcile the different accounts of the Gospel authors, or provide a more chronological order of events.

The Composite Gospel Index is intended to be a reusable *abstraction layer* on the Gospels. Because it's independent from the text of any particular translation, you can use it with whatever version you prefer. The pericopes can be organized into higher-level structures by theme (parables, miracles, teachings), topics (faith, the Kingdom of God, eternal life), by chronological order, or other arrangements. From this perspective, it is a reference system, rather than an attempt to construct a single narrative out of four parts.

3.1.3. Looking With Fresh Eyes

It's my hope that the Composite Gospel Index will provide a useful new perspective on the whole story of Jesus which is more integrated and topical, rather than structural and linear. It is no substitute for viewing the material in each Gospel in its own larger and distinctive context. But it may provide fundamental units which better match the way we think about and remember things: as stories, not chapter and verse references.

A fuller description of the CGI with background, justification, and comparison to similar approaches can be found in this longer story.

3.2. The Composite Gospel Index

Outside link

Browse the Composite Gospel Index. Click on the gray areas to the left to show and hide the source texts from the Revised Standard Version. This large file (800k) may take a few moments to open on slower connections. *(Thanks to Jesse at Waileia for fixing my Javascript*

to make this cross-browser compatible!)

Download Note
This is a single HTML file that you can download to your local system and view offline if you wish.

Outside link

View an inverted index that maps references from the Gospels to their corresponding pericope.

XML document

The Composite Gospel Index (version 1.0).

XML document

View the XML Schema Definition (.xsd file).

3.3. Frequently Asked Questions

3.3.1. Questions

1. **The Data**
   - I looked at the CGI but it's all strangely formatted. Isn't it supposed to be readable?
   - What good is the CGI if it isn't human-readable?
2. **Tools**
   - I clicked on the gray area in the Pericope Browser but it didn't display the text.

3.3.2. Answers

3.3.2.1. 1. The Data

1.1. I looked at the CGI but it's all strangely formatted. Isn't it supposed to be readable?

Actually, no: it's meant to be computer-readable, not human-readable. It's expressed in an World WIde Web Consortium (W3C) standard for structuring data called XML.

Outside link

Cover Pages is a good place to start learning about it.

1.2. What good is the CGI if it isn't human-readable?

Being computer-readable, it provides a foundation for computer programs to produce a variety of other results, including human-readable ones. For example, the

Outside link

Pericope Browser is produced automatically from a combination of the CGI and the Gospel texts (formatted according to the OSIS format).

3.3.2.2. 2. Tools

2.1. I clicked on the gray area in the Pericope Browser but it didn't display the text.

This behavior is only available with Internet Explorer v.5 or higher. Other browser like Netscape or Opera don't work correctly yet.

(If you're a javascript expert and you can tell me how to make this code more browser agnostic, please let me know!)

4. NT Names

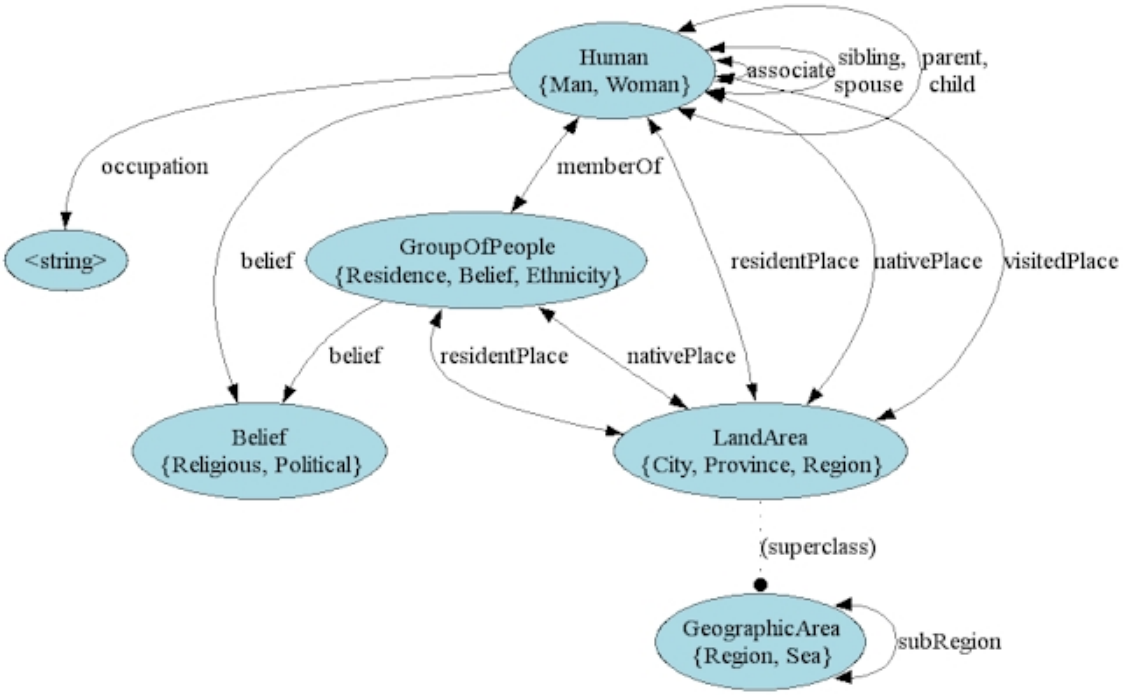
4.1. New Testament Names: a Semantic Knowledge Base

4.1.1. Overview

NTN (for New Testament Names) is a semantic knowledge base describing each named thing in the New Testament, about 600 names in all. Each named thing (an *entity*) is categorized according to its *class*, including God, Jesus, individual men and women, groups of people, and locations. These entities are related to each other by *properties* that interconnect the entities into a web of information, all represented in a standardized language with formal semantics, and shared on the Web with URI's for others to use and extend. You can download it from this page.

NTN is *the first semantic Bible information available on the Web*, as far as i can tell (and i've looked). See below for what this claim really means. Of course, there are many wonderful websites with Bible texts, search tools, reference materials, maps, and the like, all designed around natural language for human readers. But they don't provide any standardization, which is essential to machine processing of the information. Some Bible providers (and blogs) are now using RDF as part of an RSS feed to provide e.g. a daily Bible verse: however, these are still primarily web services and text containers, with little semantic information or meta-data included. NTN, on the other hand, provides a reusable Web-based

Bible resource for the Digital Age, a first step toward incorporating content from the New Testament into the Semantic Web.



Schematic of NTN classes and relationships

The main classes and properties that relate them to each other are illustrated above (omiting many details). The most important classes include:

- `Human`, as well as other kinds of `Agent` (God and other supernatural beings, whether real and malevolent (Satan) or supposed (Zeus, Artemis))
- `GeographicArea`, with sub-classes of `City`, `StateOrProvince`, `Island`, bodies of water, and a few other types
- `GroupOfPeople`, either defined by beliefs (Jews), ethnicity (Israelites, Samaritans), residence in a particular area (Galileans), or membership in political or religious organizations (Pharisees, Sadduccees)

Properties express the relationships between classes:

- `associate`, `spouseOf`, and `parentOf`/`childOf` are the main properties between `Human`s
- `member` expresses the relationship between a `Human` and a `GroupOfPeople`
- a `GeographicArea` can have a `subregion` (or be a `subregionOf`), and can be a `nativePlaceOf`, `residentPlaceOf`, or `visitedPlaceOf` a `Human`

- less-constrained properties like `occupation` and `description` are simple strings

Several other classes are either included to help the class organization, to maintain consistency with other ontologies, or to represent less frequent name classes: for example, `Character` (Alpha, Omega) and `Language` (Greek, Aramaic, Hebrew).

4.1.2. Why Do We Need This?

The purpose of NTN is to provide a standardized and sharable representation of *data*. So it's not intended for direct consumption by end-users, and it's not itself an application: instead, it provides a resuable knowledge base of Bible information that can provide a foundation for many different applications.

The W3C's Semantic Web home page is a great jumping-off point for information about the endeavor of bringing the Web up to the next level of usefulness by adding semantic meta-data. Ultimately, this work is just one tiny step in the direction of a complete conceptual annotation of the Scriptures, which is the grand vision of the Semantically-Annotated New Testament (SemANT). That's still a distant dream, but every bit of progress helps clarify the rest of the journey.

4.1.2.1. OWL, the Semantic Web Language

NTN is represented in OWL, the Ontology Web Language which was recently released as a W3C recommendation. OWL in turn is built on RDF, and expressed in XML. If you look at the NTN data with a text viewer, you'll see a dismaying number of angle brackets, pound signs, and other notation. While the file is technically readable text, it's not very understandable unless you know about OWL, RDF, and XML: it's really meant for processing by program. Most people will want to view it using special editors like Protege that understand OWL and provide a more friendly interface (be sure to get the OWL plug-in) . Here's a screenshot of (most of) the classes, and some of the instances of the Woman class, displayed in Protege. This (large) diagram shows the whole class structure.

4.1.3. What's It Good For?

Even though NTN isn't an end-user application, it's hard to see its value without some concrete examples. I soon hope to develop at least a few applications to show what you can do with it in the areas of:

Search:
There's no easy way to search for the names of Timothy's mother and grandmother, unless you already know them (though you can find them in a good Bible dictionary). There's no way whatsoever to easily understand who worked alongside Barnabas, or what cities they visited, other than undertaking an

exhaustive survey. But linking information together conceptually makes these kinds of searches possible.

Visualization:

Geographic data is particularly appropriate for visualizations: where were the people Paul addressed, and what routes did his co-workers travel?

Social Networks (who knew who):

Graphing social networks can reveal patterns of relationships that are otherwise obscured in textual descriptions

Beyond software applications, the process of making knowledge explicit provides an important benefit to our overall understanding. Often it's only when we attempt to formalize our knowledge that we discover the areas where our understanding is still implicit or incomplete.

4.2. View and Download NT Names

Download the most current released version (2004/09) of the New Testament names in OWL.This release has a complete set of names for the Woman class (48 in all), with relationship data for native/resident/visitedPlace, spouse/childOf, and religiousBelief.

The previous version (2004/04) was the initial release, which is now deprecated.

If you're seriously interested in looking at the data, i strongly recommend the Protege editor (you also need the OWL plug-in).

This HTML file provides an semi-readable overview of the ontology, produced by Mike Dean's dumpont tool. Only the hyperlinks for properties seem to work correctly: other links break something in the dumpont code. The long lists of instances make it a bit unwieldy, and it doesn't include any of the documentation: it's still better than reading raw OWL though.

4.3. NTN Development Roadmap

4.3.1. Phase 1

The initial release includes classes and properties for the ontology, and all the instances, appropriately categorized into class.

Released 2004 Apr 01.

4.3.2. Phase 2

The goal of Phase 2 is to add initial relationship data, generally taken from the immediate

context in which names occur, and to enable parallel development of a search application. Activities:

- add alternative labels (names) for instances of `Human` (e.g. "John Mark" as another name for Mark).
- add properties for other `Human` relationships: `childOf`/`parentOf`, `spouseOf`.
- add a brief `description` for `Human` instances where appropriate
- specify membership in `BeliefGroup`
- add a ministry role where appropriate
- add properties for `nativePlace(Of)` and `residentPlace(Of)`

4.3.3. Phase 3

The goal of Phase 3 is to add Scripture references, exhaustively when feasible, and incorporate more detailed information about Paul's missionary journeys.

- add `visitedPlace(Of)` data

4.3.4. Phase 4

The goal of Phase 4 is to include additional geographic information (latitude and longitude), as well as consulting and referencing additional standard works like Smith's, Easton's, ISBE, etc. Another option is to add variant name spellings from the KJV to support automated or semi-automated integration of other KJV-based resources. This is the final phase envisioned for the NTNames.

4.3.5. NT Entities

The next step beyond NT Names is to incorporate other Humans mentioned (but not named): the Gerasene demoniac, the Samaritan woman at the well, etc.

5. SemANT

5.1. Semantically-Annotated New Testament (SemANT)

- The Vision of a Semantic New Testament (from Blogos)
- Frequently Asked Questions about the SemANT Project

5.2. Frequently Asked Questions

5.2.1. Questions

1. **Approach**

- How can you possibly capture all the subtle meanings of Scripture?

5.2.2. Answers

5.2.2.1. 1. Approach

1.1. How can you possibly capture all the subtle meanings of Scripture?

Translation always involves tradeoffs, and not everything can be translated. But even though translations aren't perfect, we still find them useful: if not, we'd all need to learn to read Koine Greek.

6. Other Resources

6.1. Index of Other Resources

6.1.1. New Testament Names and Frequences

This lists the proper names and their frequencies in the

Outside link

English Standard Version translation of the New Testament. It's a plain text file, tab-delimited. (Thanks to their webmaster for sharing this)

Outside link

View the list

6.1.2. Starting to Read the Bible for Yourself

Slides from a seminar at

Outside link

Cedar Ridge Community Church on helping newer believers get started with personal Bible study. It's not fully self-explanatory, but some parts may be useful apart from the presentation.

Outside link

View the slides

6.2. Other Writings

Outside link

Exercising Humilty: a modernized paraphrase of "Acts of Humility" from Holy Living by

Jeremy Taylor, a 17th century devotional writer.

Outside link

Engaging the Heart through the Tongue: on the relationship between external speech and internal feeling.

Outside link

Haggai's Habits of Highly Effective People: the power of reflection and motivation in accomplishing great works of God.

Outside link

James Strong, IT Heavyweight: an information perspective on Strong's Bible Concordance.

Outside link

The Important of Knowing Divine Truth: a modern paraphrase of a spiritual classic by Jonathan Edwards (unfinished).

6.3. Links

- Many of the materials available here are encoded using the

  Outside link

  Open Scriptural Information Standard (OSIS), an XML schema for marking up scripture and related text that is under active development by the

  Outside link

  Bible Technologies Group.
- There's a longer list of links to other resources and blogs at

  Outside link

  Blogos.