

PWP Ontology

Harry Plantinga

April 9, 2014

Introduction

In order to build a semantic search engine that can “understand” queries and help the user search through related concepts to quickly find resources of the sort that are needed, we need a map of “named entities” (theological concepts, events, biblical people and places, sources and authors, church seasons, etc.) and their relationships—an ontology of bible and theology.

The primary purpose of the ontology is to have information on concepts that can aid search and presentation of information. This will occur in several ways:

1. **Named entity treatment pages.** We may have an entity named Christmas. We should have a page of information to offer about Christmas that is appropriate to the site—in the case of the Preaching and Worship Portal, preaching and worship resources for Christmas sermons. We will also want other views of named entities. For example, we might have a pop-up view of the author of a resource that displays when mousing over the author’s name.
2. **Resource tagging.** We will also use the Christmas entity as a way of restricting other searches. That is, a resource may be tagged as pertaining to “Christmas.” Later, if someone searches for “Christmas videos”, we can restrict our search to those labeled “Christmas”. We will use a subset of 100-200 entities for hand-tagging, but when the site is running we will be able to supplement connections through data mining.
3. **Search synonyms.** Users may type various terms intending the same named entity. We will maintain a list of keywords that users type along with probabilities that the keyword refers to the entity. For example, we may have an entity named “Nativity”, but there may also be a probability that a user typing “Nativity” into the search engine is using it as a synonym for “Christmas”. Therefore we will maintain probabilities that search terms refer to entities.
4. **Search disambiguation.** We can also use the named entities to disambiguate queries. A search for “john” is fairly common, but it may refer to John the apostle, John the book of the Bible, the scripture passage consisting of the book of John, or John Calvin, among many other possible meanings. We can show users the most likely interpretations of the query and let them choose one.

Our goal is that we will eventually have an ontology with a wide enough range of named entities that it covers most of the searches users perform, with enough information about named entities to create the views we need with an appropriate amount of information. Of course, there will be diminishing returns with increasing the size of the ontology. Creating an ontology that allows us to intelligently handle half of the searches might be a reasonable initial goal. A good Bible dictionary or encyclopedia may have 10,000 articles; we’ll probably need at least that number of named entities.

The bigger task will be to add enough information to the database about each entity that we will be able to display a useful information page about each. For about 100-200 important entities we may be able to create “full-treatment” pages of information for preachers preaching on the topic. For less important entities, can gather information from sources such as Bible dictionaries and

encyclopedias and DBpedia, which has the information from Wikipedia. However, this will still require some work, looking up the entity on Wikipedia for example.

Technically, an ontology is a specification for the concepts used in creating a body of knowledge, with the facts themselves being part of a knowledge base. However, these terms are used variously. For simplicity, I will refer to the conceptualization and facts collectively as an ontology.

Ontology Format

Ontologies are generally represented as a collection of named entities (e.g. `John_the_Apostle` and `Book_of_John`) and relationships between them (e.g. “`authored`”). So we might represent an assertion about our entities like this:

```
John_the_Apostle    authored    Book_of_John
```

Each triple has a subject, relationship, and object. Thus, the collection of knowledge can be thought of as a directed graph, and these facts may be stored in a specialized database called a “triple store.” These triples may be stored in various formats, including RDF, Turtle, and Manchester syntax and queried with a query language called SPARQL.¹ There are reasoning engines that make deductions from facts. For example, if we have the fact that Jacob is the son of Isaac and that Isaac is the son of Abraham, the reasoning engine would be able to search for grandchildren of Abraham and find Jacob.

One distinguishing characteristic of triplestores is that there are no data schemas. You can add whatever facts you want: there is no list of fields for a particular type of entity. This is both a blessing and a curse. It makes adding disparate types of entities and facts about them easy, but it makes finding information about a particular type of entity harder. In fact, reasoning systems working on triple stores are notoriously slow. Therefore we will use an ontology and semantic technologies to represent entities and to find information about those entities on the semantic web, but we will create structured, SQL-style tables for each entity type and keep the information in the MySQL database for production use.

Manchester syntax seems to be the most concise and human-readable, so we will use that syntax for editing ontology files, known as owl files. There is an open source owl file editor called Protégé, but it is not very efficient or easy to use, so it may be easier to edit the OWL files with a text editor and validate them after editing, writing custom programs to load the information into the database.

Rich ontologies have many upper-level entries. So, for example, a class `Author` may be a subclass of `Human`, which is a subclass of `CognitiveAgent`, which is a subclass of `SentientAgent`, which is in turn subclass of `Agent`, `Object`, `Physical`, and `Entity`. The upper levels of the ontology are less important for our purposes. We will need types such as `Topic`, `Tag`, `Event`, `Nation`, `Place`, `Nationality`, `Person`, `Author`, etc. However, it remains to be seen how rich an upper ontology we need. Perhaps we can start with a fairly shallow one and add to it as needed. It is also not yet known what facts about entities and relationships between them we will need to represent. This will be driven by the information we wish to display in the various entity views we create.

¹ The book *Learning SPARQL* by Bob DuCharme has a nice introduction to semantic technologies.

Treated Terms

For each named entity we need enough information to be able to display a number of views of that entity. For books, we may want a full-page display that has rich information and a one-inch display that may be used in a search result list, with enough context that users can decide whether the book meets their needs. We will also want a sidebar view for a search results page.

The types of views needed and the information that should be displayed in each view will vary according to the entity type. Therefore we will need to develop database tables (schemas) with all the information needed for all views, as the need for views becomes apparent. We will then have to fill out the information as entities are added to the ontology.



We will have different levels of treatment for entities depending on their importance. We will call the richest level of information a “full treatment,” and we will create full treatments for the 100-200 most important entities for preachers. For many other entities, we will look up the entity in online resources such as DBpedia and enter an ID, with the system looking for needed information from the various online sources. We may have to fill out some of the information by hand.

EntityIDs and Names

Each entity will need a unique EntityID and name. The EntityIDs and names should be human readable as much as possible. So, for example, we might use Augustine_of_Hippo rather than 931246 or even Augustine37. Ideally, the EntityID and name would be the same, but names often have spaces and, for various reasons, EntityIDs should not have spaces. Perhaps we can make them the same, with spaces converted to underscore characters for the EntityIDs.

It would also be convenient if the EntityID could be a valid identifier in Manchester syntax. Unfortunately, Manchester syntax is very restrictive for identifiers, and we will need to be more liberal. We may have characters with accents or even words in other alphabets. Therefore we will need to support most Unicode characters. This also allows us to use apostrophes, quotes, and parentheses, for identifiers such as Pilgrim’s_Progress or Confessions_(Augustine). Perhaps we can eliminate whitespace in identifiers and underscore characters in names.

EntityIDs and names will be chosen by humans in many cases. They should be descriptive and in general not abbreviated. Wikipedia IDs are the example to follow, and in fact we should use the same identifier as Wikipedia if they have a page on the entity. In some cases entityIDs may be computer generated, and they may get quite long. For example, the longest current entityID is a section of one of the volumes of the Early Church Fathers series edited by Schaff. It contains the title of the book, the author’s name, and the section identifier. This is the automatically-generated ID:

fathers_of_the_third_and_fourth_centuries_lactantius_venantius_asterius_victorinus_dionysius_a
postolic_teaching_and_constitutions_homily_(schaff)_(iii.ii.viii.xxxviii)

The identifier is 167 characters long. (It should be capitalized properly.) However, sections of books are not entities that will have treatments or frequently be clicked on.

Entity Types

Entities have types, which determine the information we store about the entity. Types also determine when and how to display the information about the entity. Thus, the type chosen should be one from the types we know how to handle, as adding a type means creating a database table and views for the type. Here is the current list of types in the initial ontology:

```
book
author
section
site
page
topic
resource_type
liturgic
tag
holiday
season
person
biblebook
place
```

Initial Ontology

Steve Derosé contributed a number of ontology files he had created for biblical topics, including biblical people, biblical places, bible books, idols, languages, bible manuscripts, and bible translations. The biblical people ontology is the largest of the group, and it has many facts about each person. We can use these as a starting point for an ontology.

I have supplemented this starting point from three sources. I added the “HoseeHundred” list of a little more than 100 recommended terms for cataloging sermon preparation material, from Scott Hosee. I added some events, church seasons, and resource types previously identified as needed for cataloguing resources for the PWP. I added entities for all the CCEL books and authors. Finally, I looked through the top 1000 search queries entered at the CCEL and added all of the ones that seemed as though they ought to have ontology entries. These supplements resulted in nearly 400 additional entities. This resulted in a total of 5,510 entities (not including sections of books) of which 3,762 are bible people and 1,308 are CCEL books or authors. However, there are as yet no facts about these entities. See the appendix for the additional 440 entities.

There is an entity table in the semantics database on Julian containing these entities and the keywords linked to them. It was populated from simplified entity files, not the richer Manchester syntax. The files containing this data may be found on `julian:/home/hplantin/work/ontology`.

Continuing Development

Probably the current list of entities will only need a few additions in the first year or so. Most of work will be in gathering needed facts about these entities and relations between them. We will need a database table for each entity type with a suitable list of fields. We will need a way to enter

information from full treatments, and we will need a way to enter keys from other databases such as DBPedia and other information for minor treatments. We will have to look up each entity in various online sources and add the key to the database..

One preliminary step should be to work on creating views of entity types. As a view is defined, the facts we need about an entity type will be determined. We can also immediately start adding keys for other information resources for entities. At a minimum, we will want a VIAF ID if available, since it contains keys to other information sources such as DBPedia. If a VIAF ID is not available, we will want a DBPedia key if available. We will also want keys for the dictionaries and encyclopedias at CCEL, most notably the Schaff-Herzog and Catholic encyclopedias.

Once the needed facts and relationships are known, we can start filling out the facts and relationships needed. Some data will be added by hand, such as VIAF IDs and disambiguation text. We can also work on programs to gather information from the various sources on the semantic web to populate the database tables.

When the current list of entities is treated and the search engine is in place and running, we can start adding additional entities to handle more queries intelligently. In order to identify the most important ones to add, we can find the most frequent queries that are not being handled intelligently. After a couple years' work on the ontology, we will be able to make a better estimate of the appropriate level of effort for continued work on the ontology as a percentage of total project funding.

Appendix: Basic Named Entities

Not including biblical people, bible books, books, or authors

Draft of April 9, 2014

```
#
# File:  basicterms.ent: basic entities for the initial PWP ontology
#
# Based on Scott Hosee's keywords for tagging resources for preachers. Topics have
# been added from the top 1000 CCEL searches, cataloging needs, other denominations, etc.
#
# Disambiguation strings, urls, and search keywords are not shown.
#
# Columns are separated by one or more tabs, entries within a column by commas
#
# entityID          name                      type
# =====          =====
#
# Resource type
sermon_starter_type      Sermon starter          resource_type
sermon_outline_type     Sermon outline          resource_type
sermon_type              Sermon                  resource_type
commentary_type          Commentary/exegesis     resource_type
essay_type               Blog post/essay         resource_type
reflection_type          Reflection               resource_type
book_citation_type      Book citation           resource_type
movie_citation_type     Movie citation           resource_type
topic_treatment_type    Topic treatment         resource_type
map_type                 Map                     resource_type
image_type               Image                   resource_type
video_type               Video                   resource_type
song_type                Hymn/song               resource_type
drama_type               Drama                   resource_type
activity_type            Activity                 resource_type
childrens_type           Children's resource     resource_type
liturgy_type             Complete liturgy        resource_type
liturgic_type            Liturgical element      resource_type

# Element of worship
agnus_dei_liturgic      Agnus Dei (liturgical element)  liturgic
amen_liturgic           Amen (Liturgical element)       liturgic
baptism_liturgic        Baptism                          liturgic
benediction_liturgic    Benediction (liturgical element) liturgic
call_to_worship_liturgic Call to Worship (liturgical element) liturgic
closing_liturgic        Closing(liturgical element)     liturgic
communion_liturgic      Communion/Eucharist (liturgical element) liturgic
confession_liturgic     Confession (liturgical element) liturgic
credo_liturgic          Credo/creed (liturgical element) liturgic
gloria_liturgic         Gloria (liturgical element)     liturgic
gospel_acclamation_liturgic Gospel acclamation (liturgical element) liturgic
illumination_liturgic  Illumination (liturgical element) liturgic
kyrie_liturgic          Kyrie (liturgical element)      liturgic
lords_prayer_liturgic   Lord's Prayer (liturgical element) liturgic
lords_supper_liturgic   Lord's Supper                   liturgic
memorial_acclamation_liturgic Memorial Acclamation (liturgical element) liturgic
offertory_liturgic      Offertory (liturgical element)  liturgic
praise_liturgic         Praise (liturgical element)     liturgic
prayer_liturgic         Prayer (liturgical element)      liturgic
response_liturgic       Response (liturgical element)    liturgic
sanctus_liturgic        Sanctus (liturgical element)     liturgic

# Attitude
irony                    Irony                            tag
humorous                 Humorous                         tag

# Life Events
baptism                  Baptism                          topic
birth                    Birth                             topic
```

death	Death	topic
divorce	Divorce	topic
funeral	Funeral	topic
suicide	Suicide	topic
wedding	Wedding	topic
# Special days		
all_saints_day	All Saints' Day	holiday
ascension_day	Ascension Day	holiday
ash_wednesday	Ash Wednesday	holiday
christmas_day	Christmas	holiday
easter_day	Easter	holiday
epiphany_day	Epiphany	holiday
good_friday	Good Friday	holiday
pentecost_day	Pentecost	holiday
reformation_day	Reformation Day	holiday
transfiguration_day	Transfiguration Day	holiday
# Church seasons		
advent_season	Advent	season
christmas_season	Christmas	season
epiphany_season	Epiphany	season
lent_season	Lent	season
easter_season	Easter	season
pentecost_season	Pentecost	season
# Theme clusters for preaching		
beatitudes	Beatitudes	topic
fruit_of_the_spirit	Fruit of the Spirit	topic
gifts_of_the_spirit	Gifts of the Spirit	topic
lords_prayer	Lord's Prayer	topic
parables	Parables	topic
sermon_on_the_mount	Sermon on the Mount	topic
seven_deadly_sins	Seven deadly sins	topic
ten_commandments	Ten Commandments	topic
# Christian life		
beauty	Beauty	topic
conversion	Conversion	topic
culture	Culture	topic
family	Family	topic
fear	Fear	topic
grief	Grief	topic
guilt	Guilt	topic
healing	Healing	topic
leadership	Leadership	topic
marriage	Marriage	topic
ministry	Ministry	topic
money	Money	topic
penance	Penance	topic
pilgrimage	Pilgrimage	topic
politics	Politics	topic
poverty	Poverty	topic
promises	Promises	topic
religion	Religion	topic
repentance	Repentance	topic
reverence	Reverence	topic
sexuality	Sexuality	topic
spirituality	Spirituality	topic
suffering	Suffering	topic
temptation	Temptation	topic
tithing	Tithing	topic
tongues	Tongues	topic
work	Work	topic
# Spiritual practices		
confession	Confession	topic
discipleship	Discipleship	topic
evangelism	Evangelism	topic
fasting	Fasting	topic
forgiveness	Forgiveness	topic
praise	Praise	topic
prayer	Prayer	topic
rejoicing	Rejoicing	topic

thanksgiving	Thanksgiving	topic
# Fruit of the spirit, virtues		
compassion	Compassion	topic
contentment	Contentment	topic
faith	Faith	topic
faithfulness	Faithfulness	topic
gentleness	Gentleness	topic
goodness	Goodness	topic
hope	Hope	topic
hospitality	Hospitality	topic
humility	Humility	topic
joy	Joy	topic
kindness	Kindness	topic
love	Love	topic
mercy	Mercy	topic
patience	Patience	topic
peace	Peace	topic
self_control	Self control	topic
selfishness	Selfishness	topic
wisdom	Wisdom	topic
# Sins, vices		
addiction	Addiction	topic
anger	Anger	topic
blasphemy	Blasphemy	topic
corruption	Corruption	topic
envy	Envy	topic
folly	Folly	topic
gluttony	Gluttony	topic
gossip	Gossip	topic
greed	Greed	topic
hatred	Hatred	topic
hypocrisy	Hypocrisy	topic
idolatry	Idolatry	topic
lust	Lust	topic
lying	Lying	topic
pride	Pride	topic
selfishness	Selfishness	topic
sloth	Sloth	topic
stealing	Stealing	topic
# Ten Commandments		
first_commandment	First Commandment	topic
second_commandment	Second Commandment	topic
third_commandment	Third Commandment	topic
fourth_commandment	Fourth Commandment	topic
fifth_commandment	Fifth Commandment	topic
sixth_commandment	Sixth Commandment	topic
seventh_commandment	Seventh Commandment	topic
eighth_commandment	Eighth Commandment	topic
ninth_commandment	Ninth Commandment	topic
tenth_commandment	Tenth Commandment	topic
# Church life # (')		
children	Children	topic
church	Church	topic
church_discipline	Church discipline	topic
liturgy	Liturgy	topic
revival	Revival	topic
worship	Worship	topic
# Trinity		
god	God the Father	topic
jehovah	Jehovah	topic
jesus	Jesus Christ	topic
holy_spirit	Holy Spirit	topic
trinity	Trinity	topic
incarnation	Incarnation	topic
nativity	Nativity	topic
crucifixion	Crucifixion	topic
resurrection	Resurrection	topic
# Theological topics		

angels	Angels	topic
covenant	Covenant	topic
creation	Creation	topic
election	Election	topic
epiphany	Epiphany	topic
eschatology	Eschatology	topic
ethics	Ethics	topic
good_works	Good works	topic
glory	Glory	topic
grace	Grace	topic
heaven	Heaven	topic
hell	Hell	topic
image_of_god	Image of God	topic
judgment	Judgment	topic
justice	Justice	topic
kingdom_of_god	Kingdom of God	topic
law	Law	topic
lords_supper	Lord's Supper	topic
morality	Morality	topic
miracles	Miracles	topic
providence	Providence	topic
redemption	Redemption	topic
sabbath	Sabbath	topic
sacrifice	Sacrifice	topic
salvation	Salvation	topic
scripture	Scripture	topic
shalom	Shalom	topic
sin	Sin	topic

The above are used for categorizing preaching resources. Most should probably also have their own full treatments.

Those that follow are from the top 100 CCEL search terms that are not covered by the topics above, scripture passages, book and author names, full text searches, etc. These may not need full treatments.

questions

- # - is interpretation the same as hermeneutics? Should we have multiple "interpretation" entities?
- # - is charity the same as love? Should we have separate entries for agape, caritas, etc?
- # - which of salvation, regeneration, justification, sanctification, remission, forgiveness, etc to use? which are synonyms?
- # - trust, faith, faithfulness: redundant?
- # - should there be separate God and Jehovah entities? How about Yaweh?
- # - How to handle queries like "jews"?
- # - should there be separate entities for rome (the place), and rome as a theological/biblical topic?
- # - Use a different type for physical things like ark of the covenant, temple, etc?
- # - what kind of entities should be used to handle queries like "Women"?
- # - do we handle the fairly frequent query "homosexuality"?

# isms		
anabaptist	Anabaptist	topic
arianism	Arianism	topic
arminianism	Arminianism	topic
calvinism	Calvinism	topic
catholicism	Catholicism	topic
cistercian	Cistercian	topic
gnosticism	Gnosticism	topic
methodist	Methodist	topic
monasticism	Monasticism	topic
montanism	Montanism	topic
mysticism	Mysticism	topic
orthodoxy	Orthodoxy	topic
puritan	Puritans	topic
catholic_apostolic_church	Catholic Apostolic Church	topic

# people		
aristotle	Aristotle	topic
socrates	Socrates	topic

# biblical topics		
ark_of_the_covenant	Ark Of The Covenant	topic

golden_calf	Golden Calf	topic
nephilim	Nephilim	topic
rosary	Rosary	topic
temple	Temple	topic
tabernacle	Tabernacle	topic
# theological topics		
adoption	Adoption	topic
agape	Agape	topic
antichrist	Antichrist	topic
apocalypse	Apocalypse	topic
apostle	Apostle	topic
apostolic_tradition	Apostolic Tradition	topic
ascension	Ascension	topic
atonement	Atonement	topic
begotten	Begotten	topic
blood	Blood	topic
biblical_canon	Biblical Canon	topic
bishop	Bishop	topic
catechism	Catechism	topic
charity	Charity	topic
contemplation	Contemplation	topic
creation	Creation	topic
demons	Demons	topic
discipline	Discipline	topic
doctrine	Doctrine	topic
dreams	Dreams	topic
early_church	Early Church	topic
exorcism	Exorcism	topic
fellowship	Fellowship	topic
flesh	Flesh	topic
free_will	Free Will	topic
glory	Glory	topic
gospel	Gospel	topic
heart	Heart	topic
heresy	Heresy	topic
inerrancy	Inerrancy	topic
interpretation	Interpretation	topic
jubilees	Jubilees	topic
justification	Justification	topic
light	Light	topic
living_water	Living Water	topic
logos	Logos	topic
martyr	Martyr	topic
meditation	Meditation	topic
messiah	Messiah	topic
missions	Missions	topic
natural_law	Natural Law	topic
original_sin	Original Sin	topic
parables	Parables	topic
passover	Passover	topic
persecution	Persecution	topic
preaching	Preaching	topic
predestination	Predestination	topic
priesthood	Priesthood	topic
prophecy	Prophecy	topic
purgatory	Purgatory	topic
rapture	Rapture	topic
reconciliation	Reconciliation	topic
regeneration	Regeneration	topic
remission	Remission	topic
relics	Relics	topic
sacrament	Sacrament	topic
sanctification	Sanctification	topic
soul	Soul	topic
spiritual_warfare	Spiritual Warfare	topic
symbol	Symbol (creed, catechism)	topic
theology	Theology	topic
world	World	topic
theotokos	Theotokos	topic
transfiguration	Transfiguration	topic
tree_of_life	Tree Of Life	topic
worship	Worship	topic

# virtues		
courage	Courage	topic
integrity	Integrity	topic
obedience	Obedience	topic
perseverance	Perseverance	topic
righteousness	Righteousness	topic
trust	Trust	topic
# events		
council_of_chalcedon	Council of Chalcedon	topic
council_of_nicea	Council Of Nicea	topic
crusades	Crusades	topic
# places		
babylon	Babylon	place
carthage	Carthage	place
israel_place	Israel (place)	place
jerusalem	Jerusalem	place
rome	Rome	place
# other topics		
abortion	Abortion	topic
body	Body	topic
bride	Bride	topic
conscience	Conscience	topic
darkness	Darkness	topic
deacon	Deacon	topic
health	Health	topic
education	Education	topic
eye	Eye	topic
honey	Honey	topic
islam	Islam	topic
israel_nation	Israel (nation)	topic
israel_nationality	Israel (nationality)	topic
jew	Jews	topic
magic	Magic	topic
pope	Pope	topic
reformation	Reformation	topic
saints	Saint, Saints	topic
salt	Salt	topic
shepherd	Shepherd	topic
slavery	Slavery	topic
strength	Strength	topic
thrones	Thrones	topic
time	Time	topic
truth	Truth	topic
virginity	Virginity	topic
virtue	Virtue	topic
war	War	topic
warriors	Warriors	topic
wife	Wife	topic
transfiguration	Transfiguration	topic
women	Women	topic