U.S. DEPARTMENT OF ENERGY

PNNL-19041

Prepared for the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Hierarchical Analysis of the Omega Ontology

C Joslyn
P Paulson

December 2009

Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from the
Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062;
ph: (865) 576-8401
fax: (865) 576-5728
email: reports@adonis.osti.gov

Available to the public from the National Technical Information Service,
U.S. Department of Commerce, 5285 Port Royal Rd., Springfield, VA 22161
ph: (800) 553-6847
fax: (703) 605-6900
email: orders@ntis.fedworld.gov
online ordering: http://www.ntis.gov/ordering.htm

This document was printed on recycled paper.
(9/2003)

PNNL-19041

Hierarchical Analysis of the Omega Ontology

C Joslyn
P Paulson

December 2009

Prepared for the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99352

Hierarchical Analysis of the Omega Ontology*

Cliff Joslyn† and Patrick Paulson‡

December, 2009

Contents

Abstract

We provide an analysis of the hierarchical structure of a version of the Omega Ontology currently in use within the US Government. After providing an initial statistical analysis of the distribution of all link types in the ontology, we then provide a detailed order theoretical analysis of each of the four main hierarchical links present. This order theoretical analysis includes the distribution of components and their properties, their parent/child and multiple inheritance structure, and the distribution of their vertical ranks.

1 Data Preparation

We received the initial data dump of the Omega ontology [4] currently in use, and converted it to OWL. In our current version of the transformation, we assume that all subclasses of the `PROPERTY` should be expressed as object properties. Any Omega concept that is not an object property is

*Initial report for the Ontology Analytics (Ontolytics) project. PNNL Technical Report PNNL-19041.

†Chief Scientist for Knowledge Sciences, National Security Directorate, Pacific Northwest National Laboratory, PNNL/Battelle Suite 400, 1100 Dexter Ave. N, Seattle, WA 98109 `cjoslyn@pnl.gov`, 206-528-3042

‡Senior Research Scientist, Knowledge Systems Group, Pacific Northwest National Laboratory, PO Box 999 MS K7-28, Richland, WA 99352 USA, `patrick.paulson@pnl.gov`, 509-375-3926

transformed into an `owl:Class`. This approach will be revised as we incorporate special knowledge about the `SUBJECT-MATTER` slot and other characteristics of the Omega ontology.

There are several problems with this approach – sometimes the same name is used for both a concept and a property, such as `LOCATION`. To address such problems, we are creating a *meta*-ontology describing the contents of the Omega ontology. In this ontology, all classes and properties are transformed into *instances* of classes that Stand in for OWL's classes, object properties, and datatype properties. The resulting knowledge base can be examined using OWL-aware tools to determine what aspects will not result in a valid OWL-DL ontology.

2 Transitivity Analysis

In consultation with the sponsor, we performed an initial analysis of the node and link types. We identified 660,541 links and 121,658 node declarations (classes and properties). There are also 5 `SUBJECT` concepts that do not appear in the class hierarchy. The links include `DESCRIPTION` records from the omega ontology.

We considered link types together with their inverses, as shown in Table 1. Link types with inverses are shown over two sets of columns, and those without over one. In many cases, the number of forward and reverse links for inverse paris are not the same, and this is measured showing the count range and the % of that range.

After consultion with the sponsor we gained a better understanding of Omega's structure. After understanding that the subject hierarchy was distinct from the class hierarchy, we retained it, bu excluded the `DEFINITION`, `DOMAIN`, `RANGE`, `INVERSE` and `SOURCE` links.

We combined link types with inverses together. For example, `DIRECT-SUPERCLASS` and `SUBCLASSES` are combined into just `SUBCLASSES` (selected arbitrarily), and recorded with the mean of the forward and reverse link counts.

The distribution of the resulting 276,858 combined links is shown in Table 2. This reveals that by far the biggest link types are `SUBCLASSES` and `SUBJECT` (as expected). Together with the next two, `DIRECT-HAS-MEMBER` and `DIRECT-PART-OF`, they comprise 93.2% of all the links. Additionally, each of these links is transitive, that is, hierarchical.

3 Order Theoretical Analysis

We now outline aspects of our order theoretical analysis.

We have four candidate semantic hierarchies: `SUBCLASSES`, `SUBJECT`, `DIRECT-HAS-MEMBER` and `DIRECT-PART-OF`. Each breaks into a number of connected components, and each component is represented as a distinct finite ordered set $\langle P, \leq \rangle$ [1], for example as shown in Fig. 1. In this context, we can initially measure the number of parents and children of each node, revealing the amount of multiple inheritance present.

Further, we are motivated by the our conception of the proper vertical positioning of nodes [2, 3]. Note that in some sense, all children of the root (L, B, X, K) are the same "distance" from the root node 1. But they are not all the same. For example, K is also a leaf, and is similar to e.g. Q in that it is the same distance "from the bottom".

Name	Count	Inverse	Inverse Count	Count Mean	Count Range	% Range
DIRECT-SUPERCLASS	125696	SUBCLASSES	125696	125696	0	0.00%
DEFINITION	115083			115083	0	0.00%
SUBJECT	111450	SUBJECT-MATTER	111450	111450	0	0.00%
DIRECT-HAS-MEMBER	12113	DIRECT-MEMBER-OF	12113	12113	0	0.00%
DIRECT-PART-OF	8638	DIRECT-HAS-PART	8611	8624	27	0.31%
ANTONYM	7605			7605	0	0.00%
IS-PERTAINED-TO-BY	3497	PERTAINS-TO	3497	3497	0	0.00%
SHISHKABOB	1896			1896	0	0.00%
DIRECT-HAS-SUBSTANCE	1505	DIRECT-SUBSTANCE-OF	1502	1503	3	0.20%
INSTRUMENT-OF	918	INSTRUMENT	915	916	3	0.33%
HAS-SPORTS-ACTIVITY	710	IN-DISCIPLINE	710	710	0	0.00%
THEME-OF	486	THEME	489	487	3	0.62%
AGENT-OF	474	AGENT	495	484	21	4.43%
DOMAIN-OF	410	DOMAIN	414	412	4	0.98%
RANGE	389			389	0	0.00%
LOCATION-OF	364	LOCATION	364	364	0	0.00%
INVERSE	288			288	0	0.00%
PRODUCER-OF	183	PRODUCED-BY	184	183	1	0.55%
SOURCE-OF	112	SOURCE	116	114	4	3.57%
BENEFICIARY-OF	73	BENEFICIARY	75	74	2	2.74%
ROLE-FOR-AREA	69	AREA-OF-INTEREST	69	69	0	0.00%
EMPLOYER-OF	63	EMPLOYED-BY	63	63	0	0.00%
MERCHANDISE-OF	50	HAS-MERCHANDISE	50	50	0	0.00%
CUSTOMER-OF	47	HAS-CUSTOMER	48	47	1	2.13%
DESTINATION-OF	47	DESTINATION	47	47	0	0.00%
ACCOMPANIER-OF	41			41	0	0.00%
OBJECT-INVOLVED	40			40	0	0.00%
OWNED-BY	36	OWNER-OF	36	36	0	0.00%
WORK-EQUIPMENT-OF	35	HAS-WORK-EQUIPMENT	35	35	0	0.00%
MEASURED-BY	35	MEASURING-DEVICE-FOR	35	35	0	0.00%
HEADED-BY	33	HEAD-OF	25	29	8	24.24%
CAUSED-BY	30	EFFECT	29	29	1	3.33%
DIRECT-DISJOINT	26			26	0	0.00%
CONTAINED-IN	26	CONTAINS	25	25	1	3.85%
PURPOSE-OF	26	PURPOSE	29	27	3	11.54%
PRECONDITION-OF	24	PRECONDITION	24	24	0	0.00%
AREA-OF-ACTIVITY	23	ACTIVITY-IN-AREA	23	23	0	0.00%
BORDERS-ON	20			20	0	0.00%
EXPERIENCER-OF	20	EXPERIENCER	20	20	0	0.00%
HAS-PRODUCT-TYPE	19	PRODUCT-TYPE-OF	19	19	0	0.00%
REPRESENTS	18	REPRESENTED-BY	18	18	0	0.00%
ROLE-FOR-ACTIVITY	17			17	0	0.00%
SUBSTRATE-OF	15	SUBSTRATE	18	16	3	20.00%
HAS-STANDARD-MEASURE	15	STANDARD-MEASURE-FOR	15	15	0	0.00%
RELIGION-OF	14	HAS-RELIGION	14	14	0	0.00%
AREA-OF-BUSINESS-ACTIVITY	14	ORGANIZATION-IN-BUSINESS-AREA	14	14	0	0.00%
SERVICES	14	SERVICES-OF	14	14	0	0.00%
OPERATOR-OF	11	OPERATED-BY	15	13	4	36.36%
MEASURED-IN	11	MEASURING-UNIT-FOR	11	11	0	0.00%
CONNECTS	9	CONNECTED-TO	9	9	0	0.00%
CONTROLLED-BY	7	CONTROLS	7	7	0	0.00%
AREA-STUDIED-IN	5	OBJECT-STUDIED-IN-AREA	5	5	0	0.00%
UPPER-LIMIT	4	LOWER-LIMIT	4	4	0	0.00%
HAS-AUDIENCE	4	AUDIENCE-OF	4	4	0	0.00%
CO-DOMAIN	4			4	0	0.00%
ESTABLISHED-BY	3	ESTABLISHER-OF	3	3	0	0.00%
HAS-NAME	3	NAME-OF	3	3	0	0.00%
ORIGIN	3	ORIGIN-OF	3	3	0	0.00%
LANGUAGE-OF	3			3	0	0.00%
LOCATION-WITHIN-DOCUMENT	2			2	0	0.00%
AUTHOR-OF	2	AUTHORED-BY	2	2	0	0.00%
HAS-NATIONALITY	2	NATIONALITY-OF	2	2	0	0.00%
HAS-CORPORATE-DIVISION	2	PARENT-CORPORATION-OF	2	2	0	0.00%
ELEMENT-OF	2	HAS-ELEMENT	2	2	0	0.00%
LESS-THAN	2	GREATER-THAN	2	2	0	0.00%
TEXTUAL-RELATION	2	COGNITIVE-RELATION	2	2	0	0.00%
OUTSIDE-OF	1			1	0	0.00%
HAS-REPRESENTATIVE	1	REPRESENTATIVE-OF	1	1	0	0.00%
HAS-CURRENCY	1	CURRENCY-OF	1	1	0	0.00%
SPORTS-CLUB-OF	1	HAS-SPORTS-CLUB	1	1	0	0.00%
HAS-COACH	1	COACH-OF	1	1	0	0.00%
INSIDE-OF	1			1	0	0.00%
HAS-PHONE-NUMBER	1	PHONE-NUMBER-OF	1	1	0	0.00%
HAS-HEADQUARTERS	1	HEADQUARTERS-OF	1	1	0	0.00%
HAS-LABEL	1	LABEL-OF	1	1	0	0.00%
LANGUAGE-REPRESENTED-IN	1			1	0	0.00%
PARTNER-OF	1	HAS-PARTNER	1	1	0	0.00%
OUTCOME	1			1	0	0.00%

Table 1: Link type distribution, with inverses.

Link	Count	%	Cumulative %
SUBCLASSES	125696	45.4%	45.4%
SUBJECT	111450	40.3%	85.6%
DIRECT-HAS-MEMBER	12119	4.4%	90.0%
DIRECT-PART-OF	8698	3.1%	93.2%
ANTONYM	7605	2.7%	95.9%
PERTAINS-TO	3497	1.3%	97.2%
SHISHKABOB	1896	0.7%	97.9%
DIRECT-SUBSTANCE-OF	1529	0.6%	98.4%
INSTRUMENT	936	0.3%	98.7%
IN-DISCIPLINE	710	0.3%	99.0%
AGENT	527	0.2%	99.2%
THEME	509	0.2%	99.4%
LOCATION	382	0.1%	99.5%
PRODUCED-BY	186	0.1%	99.6%
BENEFICIARY	77	0.0%	99.6%
AREA-OF-INTEREST	69	0.0%	99.6%
EMPLOYER-OF	66	0.0%	99.7%
HAS-CUSTOMER	53	0.0%	99.7%
HAS-MERCHANDISE	50	0.0%	99.7%
DESTINATION	50	0.0%	99.7%
HEADED-BY	42	0.0%	99.7%
ACCOMPANIER-OF	41	0.0%	99.7%
OBJECT-INVOLVED	40	0.0%	99.8%
OWNER-OF	36	0.0%	99.8%
MEASURED-BY	35	0.0%	99.8%
HAS-WORK-EQUIPMENT	35	0.0%	99.8%
CAUSED-BY	32	0.0%	99.8%
PURPOSE	30	0.0%	99.8%
PRECONDITION	26	0.0%	99.8%
DIRECT-DISJOINT	26	0.0%	99.8%
HEAD-OF	25	0.0%	99.9%
CONTAINS	25	0.0%	99.9%
AREA-OF-ACTIVITY	23	0.0%	99.9%
HAS-PRODUCT-TYPE	22	0.0%	99.9%
EXPERIENCER	21	0.0%	99.9%
BORDERS-ON	20	0.0%	99.9%
REPRESENTS	19	0.0%	99.9%
SUBSTRATE	18	0.0%	99.9%
ROLE-FOR-ACTIVITY	17	0.0%	99.9%
OPERATED-BY	16	0.0%	99.9%
AREA-OF-BUSINESS-ACTIVITY	16	0.0%	99.9%
STANDARD-MEASURE-FOR	15	0.0%	99.9%
HAS-STANDARD-MEASURE	15	0.0%	99.9%
SERVICES-OF	14	0.0%	99.9%
SERVICES	14	0.0%	99.9%
RELIGION-OF	14	0.0%	99.9%
ORGANIZATION-IN-BUSINESS-AREA	14	0.0%	100.0%
HAS-RELIGION	14	0.0%	100.0%
MEASURING-UNIT-FOR	11	0.0%	100.0%
MEASURED-IN	11	0.0%	100.0%
CONNECTS	9	0.0%	100.0%
CONNECTED-TO	9	0.0%	100.0%
CONTROLS	7	0.0%	100.0%
CONTROLLED-BY	7	0.0%	100.0%
OBJECT-STUDIED-IN-AREA	5	0.0%	100.0%
AREA-STUDIED-IN	5	0.0%	100.0%
UPPER-LIMIT	4	0.0%	100.0%
LOWER-LIMIT	4	0.0%	100.0%
HAS-AUDIENCE	4	0.0%	100.0%
CO-DOMAIN	4	0.0%	100.0%
ORIGIN	3	0.0%	100.0%
LANGUAGE-OF	3	0.0%	100.0%
HAS-NAME	3	0.0%	100.0%
ESTABLISHED-BY	3	0.0%	100.0%
TEXTUAL-RELATION	2	0.0%	100.0%
PARENT-CORPORATION-OF	2	0.0%	100.0%
NATIONALITY-OF	2	0.0%	100.0%
LOCATION-WITHIN-DOCUMENT	2	0.0%	100.0%
LESS-THAN	2	0.0%	100.0%
HAS-ELEMENT	2	0.0%	100.0%
AUTHORED-BY	2	0.0%	100.0%
SPORTS-CLUB-OF	1	0.0%	100.0%
REPRESENTATIVE-OF	1	0.0%	100.0%
PHONE-NUMBER-OF	1	0.0%	100.0%
PARTNER-OF	1	0.0%	100.0%
OUTSIDE-OF	1	0.0%	100.0%
OUTCOME	1	0.0%	100.0%
LANGUAGE-REPRESENTED-IN	1	0.0%	100.0%
LABEL-OF	1	0.0%	100.0%
INSIDE-OF	1	0.0%	100.0%
HEADQUARTERS-OF	1	0.0%	100.0%
HAS-CURRENCY	1	0.0%	100.0%
HAS-COACH	1	0.0%	100.0%

Table 2: Combined link type distribution, classes and properties only.

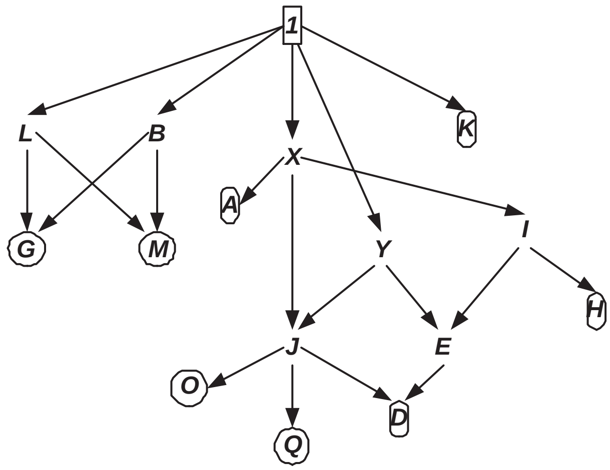

Figure 1: An example ordered set.

Thus we need to consider vertical position dually from the top and bottom. We do this by first providing a global bottom node 0 below all of the leaves (G, M, O, Q, D, H, K). Then for two nodes $a, b \in P$, we let $h^*(a, b)$ be the maximum path length between them, and the **height** to be the ax chain length from the top to the bottom: $\mathcal{H} := h^*(0, 1)$. We than have the following quantities:

Top Rank: Max chain length a to top: $r^t(a) := h^*(a, 1)$

Bottom Rank: Height minus max chain length from the bottom to a: $r^b(a) := \mathcal{H} - h^*(0, b)$

Interval Rank: $R(a) := [r^t(a), r^b(a)]$

Rank Width: $W(a) := \|R(a)\| = r^b(a) - r^t(a)$

Fig. 2 shows the resulting structure (also showing some additional features [3]). For example, we have $\mathcal{H} = 5$, and $R(K) = [1, 4]$, so that K has a top rank of 1, a bottom rank of 4, and a rank width of 3. This is maximal, because K is both "one down from the top" and also "one up from the bottom". This is contrasted with $R(I) = [2, 2]$, so that it is unequivocally at rank 2, being 2 down from the top and 3 up from the bottom.

4 Subclass

We next consider a hierarchical analysis specifically of the `SUBCLASS` hierarchy.

A connection analysis reveals that Omega is essentially "is-a complete": there is one giant component of 121,655 nodes, and eight components of size one due to erroneous roots (see below).

Table 3 shows a portion of the distribution of the number of parents and children. This shows a largely, but not completely, tree-like structure, with 3.0% of the nodes having more than one parent, and seven nodes having six.

Table 4 and Fig. 3 show the rank distribution.

We note a few things.

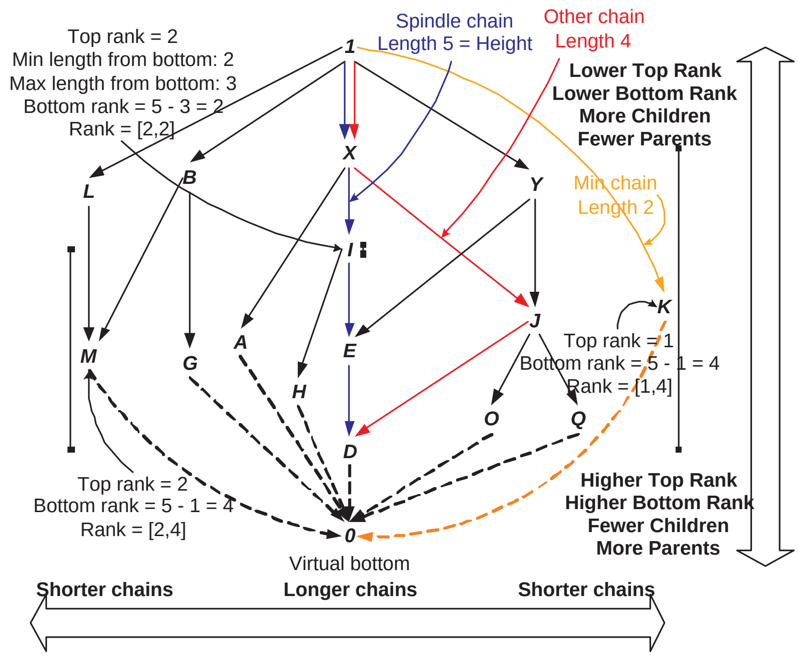

Figure 2: Rank layout.

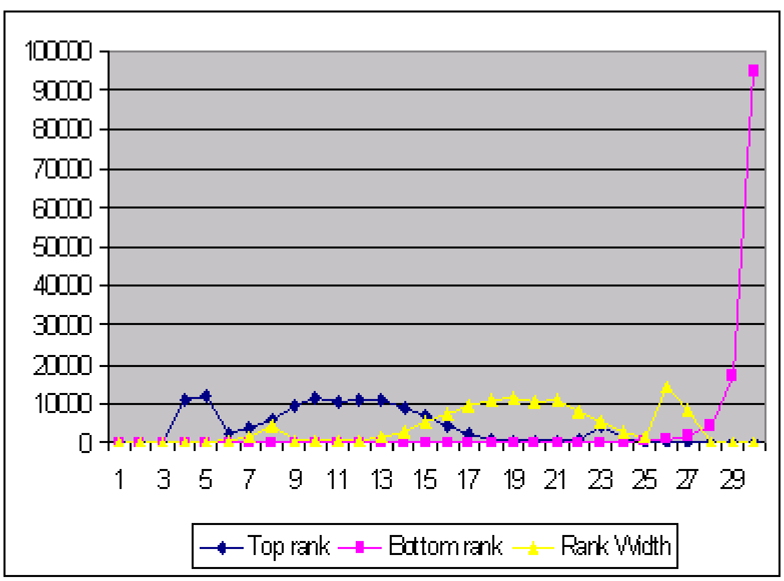

Figure 3: Subclass rank distribution.

# parents/children	Count children	Count parents
0	95913	10
1	9489	118016
2	5202	3309
3	3094	271
4	1976	43
5	1266	7
6	914	7
7	704	
8	494	
9	379	
10	297	
11	280	
12	205	
13	134	
14	136	
15	110	
16	108	
17	88	
18	69	
19	98	
20	59	
21	55	
22	47	
23	39	
24	30	
25	40	
26	20	
27	29	
28	28	
29	20	
30	16	
31	9	
32	16	
33	16	
34	11	
35	14	
36	12	
37	8	
38	10	
39	10	
40	13	
41	11	
42	7	
43	7	
44	4	
45	7	
46	7	
47	2	
48	3	
49	5	
50	4	
51	5	
52	4	
53	4	
54	6	
56	2	
57	3	
58	5	
60	5	
61	4	
62	2	
63	3	
64	1	
67	1	
68	2	
69	4	
70	2	
71	1	
72	1	
73	3	
74	3	
76	3	
77	2	
78	2	
79	1	
80	2	
82	2	
84	1	
85	1	
86	2	
87	3	
88	2	
89	2	
90	1	
91	4	
94	2	
96	1	
104	1	
105	1	
107	1	
108	1	
110	1	
...	...	...
553	1	
620	1	
10977	1	

Table 3: Subclass parent/children node distribution.

	Top Rank	Bottom Rank	Width
0	10	1	74
1	9	1	111
2	39	2	124
3	11124	3	100
4	12176	4	144
5	2272	2	324
6	4172	3	1636
7	6167	2	4387
8	9475	3	734
9	11384	3	355
10	10675	2	315
11	10851	2	686
12	10790	3	1675
13	8788	5	3227
14	6759	6	5570
15	4689	8	7671
16	2505	7	9453
17	1192	16	11108
18	465	22	11311
19	308	24	10473
20	373	48	11107
21	834	66	8156
22	4362	123	5398
23	1580	236	3218
24	219	433	1451
25	121	881	14483
26	105	1882	8338
27	126	4617	25
28	78	17345	
29	15	95913	8

Table 4: Subclass rank distributions.

- The height $\mathcal{H} = 29$.
- There are actually 10 roots (top rank = 0). These are undoubdtedly errors, and are shown in Table 5. We have developed the following opinions:

  AdministrativeDivision: might be mistyped `factbook:AdministrativeDivision`, which occurs in tree

  Building: Might be referring to `BUILDING`?

  COMPUTER_MEDIATED_COMMUNICATION_APPLICATION-SUBJECT: Misspelling of `COMPUTER_MEDIATED_COMMUNICATION_APPLICATIONS-SUBJECT` (note extra 'S')

  fiber-optics: Misspelling of `fiber-optic`

  owl:Thing: In the Omega ontology, unclear why its not under Summum Genus

  JointVentureEvent, CellularPhone, TELLECOMMUNICATION-SUBJECT, TELECOMMUNICATIONS: All are instances of SUBJECT, but some SUBJECTS are also treated as classes.

- There are 95,913 leaves (bottom rank = 29), or 78.8% of the structure.
- The behvaior of the top and bottom ranks are generally as expected. There's a broad fan-out down the structure, tapering towards the bottom, indicated by the generally unimodal top rank distribution and the increasing bottom rank distribution. The trailing spike in bottom rank is the appearance of the leaves.
- Some distinct discrepancies are apparent:
  - An initial spike in top rank at 3 and 4 indicates a distinct collection of nodes at that level.

fiber-optics
JointVentureEvent
TELECOMMUNICATION-SUBJECT
Summum_Genus
Building
NOTHING*
TELECOMMUNICATIONS
COMPUTER_MEDIATED_COMMUNICATION_APPLICATION-SUBJECT
AdministrativeDivision
CellularPhone
PHYSICAL-OBJECT
owl:Thing

Table 5: Root nodes in `SUBCLASSES`.

—FREE_TIME-SUBJECT—
—APPLIED_SCIENCE-SUBJECT—
—FACTOTUM-SUBJECT—
—SOCIAL_SCIENCE-SUBJECT—
—INGESTIBLE$NOUN—
—DOCTRINES-SUBJECT—
—TELECOMMUNICATION-SUBJECT—
—TELECOMMUNICATIONS—
—PURE_SCIENCE-SUBJECT—

Table 6: Roots of the large `SUBJECT` component.

- A corresponding spike in rank width at 25 and 26 indicates that the bulk of those nodes are, in fact, leaves.
- Another bump in rank width at 6 and 7 correlates to a spike in top rank around 22 and 23. This needs more explanation.

5 Subject

An analysis of `SUBJECT` reveals the following:

- A single, large component of 100,313 nodes with the nine roots shown in Table 6.
- A component of size four with two roots: —ARTISTIC-ACTIVITY— and —POLL—, and two children, —EVENT— and —OBJECT—, which both mutually multiply inherit.
- A component of size two with the single root —COMPUTER_MEDIATED_COMMUNICATION_APPLICATION-SUBJECT— and child "bbs".
- 21,344 lone nodes, comprising 17.5% of the structure.

The parents/children distribution for the large component is shown in Table 7. We see a shallow hierarchy (height $\mathcal{H} = 5$) with vast fanout, many many nodes having large numbers of children. The 25 nodes with more than 1000 children is shown in Table 8. Note that one, —FACTOTUM-SUBJECT—, is also a root. But, the amount of multiple inheritance is also moderately high at 9.7% of the structure. There is thus a complex structure here, which will require substantially more analysis to fully understand.

Table 9 shows the rank distribution for the large component, while Fig. 4 shows this graphically. There are the nine roots we saw before, and then a very shallow structure, with the vast bulk of

# parents/children	Count children	Count parents
0	100097	9
1	4	90410
2	5	8786
3	3	974
4	1	129
5	5	5
6	6	
7	1	
8	3	
9	3	
10	4	
11	2	
12	3	
13	3	
15	3	
16	1	
17	2	
18	2	
19	2	
20	2	
21	1	
22	1	
23	1	
24	5	
25	1	
26	3	
28	2	
29	2	
32	4	
33	4	
34	2	
35	2	
37	1	
38	2	
39	2	
41	1	
42	2	
43	1	
44	2	
45	2	
46	1	
48	1	
51	1	
52	1	
53	3	
54	1	
57	1	
59	2	
61	1	
62	1	
63	1	
64	2	
66	1	
67	1	
68	3	
69	1	
71	1	
77	1	
79	1	
87	1	
88	1	
93	1	
97	1	
99	1	
100	1	
102	1	
103	1	
105	1	
107	1	
109	1	
115	1	
124	1	
137	1	
139	1	
146	1	
147	2	
149	2	
153	1	
155	1	
156	1	
159	1	
163	1	
168	1	
186	1	
197	1	
...	...	...
2264	1	
2280	1	
2383	1	
2510	1	
2538	1	
3158	1	
3342	1	
6555	1	
7086	1	
31958	1	

Table 7: Subject parent/children node distribution.

Node	# Parents	# Children
—FACTOTUM-SUBJECT—	0	31958
—ZOOLOGY-SUBJECT—	1	7086
—BOTANY-SUBJECT—	1	6555
—BIOLOGY-SUBJECT—	1	3342
—GEOGRAPHY-SUBJECT—	1	3158
—GASTRONOMY-SUBJECT—	1	2538
—MEDICINE-SUBJECT—	1	2510
—CHEMISTRY-SUBJECT—	1	2383
—QUALITY-SUBJECT—	1	2280
—ANATOMY-SUBJECT—	1	2264
—ADMINISTRATION-SUBJECT—	1	2232
—PERSON-SUBJECT—	1	1985
—BUILDING_INDUSTRY-SUBJECT—	1	1715
—RELIGION-SUBJECT—	1	1601
—MILITARY-SUBJECT—	1	1517
—LINGUISTICS-SUBJECT—	1	1486
—LAW-SUBJECT—	1	1436
—PSYCHOLOGY-SUBJECT—	1	1366
—METROLOGY-SUBJECT—	1	1350
—ECONOMY-SUBJECT—	1	1276
—TRANSPORT-SUBJECT—	1	1143
—PHYSICS-SUBJECT—	1	1068
—POLITICS-SUBJECT—	1	1013
—INDUSTRY-SUBJECT—	1	1009
—MUSIC-SUBJECT—	1	1005

Table 8: Subject nodes with > 1000 children.

	Top Rank	Bottom Rank	Width
0	9	1	11
1	32213	4	1947
2	32297	7	33990
3	33898	40	32199
4	1890	164	32166
5	6	100097	

Table 9: Rank distributions for Subject large component.

the nodes living between top ranks 1 and 3, tracking the large number of single-node components. There are far more leaves here, 90.1% of the structure.

6 Member

An analysis of `DIRECT-HAS-MEMBER` reveals the distribution of component sizes shown in Table 10. We note:

- There are 109,059 lone nodes comprising 89.6% of the structure; or, in other words, only 10.4% of the structure is actually included within the `MEMBER` hierarchy.
- The one largest component with 5,291 members has 4.4% of the structure.
- The second largest component with 4,688 members as 3.9% of the structure.

The number of roots and leaves of the components with five or more elements is shown in Table 11, with the roots shown for any with one or two roots.

We can observe the following about the two largest components:

ID 241: 5291 nodes, height $\mathcal{H} = 11$. Its 13 roots are show in Table 12, its parents/children distribution in Table 13, and its rank distribution in Table 14. There is very little multiple inheritance, with only 0.4% of the structure having more than one parent. 53.7% are leaves.

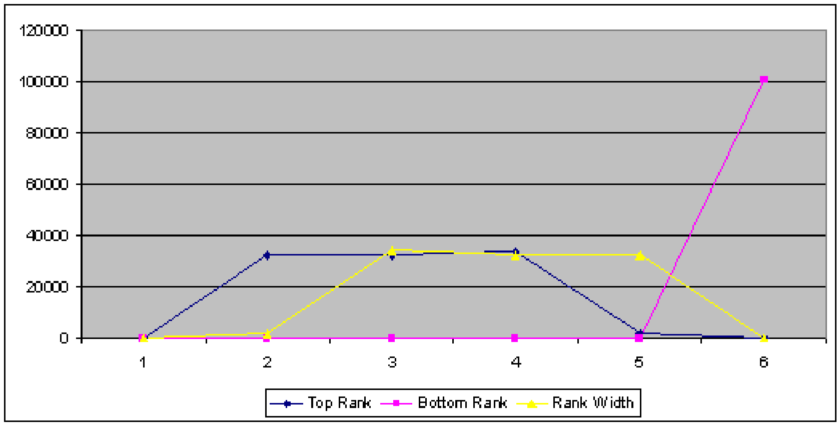

Figure 4: Rank distribution of Subject large component.

Size	# Components
5291	1
4688	1
515	1
177	1
150	1
124	1
66	1
26	1
25	1
21	1
20	1
17	3
15	1
14	2
13	2
12	1
11	2
10	5
9	2
8	3
7	6
6	2
5	19
4	25
3	72
2	395
1	109059

Table 10: Component distribution for `DIRECT-HAS-MEMBER`.

Compid	Size	Height	# Roots	# Leaves	Root(s)
241	5291	11	13	2843	
252	4688	9	1	2779	—Plantae—
2902	515	8	1	305	—Fungi—
2738	177	7	1	87	—Protoctista—
231	150	2	8	80	
3227	124	6	1	59	—Monera—
8529	66	3	1	40	—Ericaceae—
7592	26	1	1	25	—Roman_alphabet—
8474	25	1	1	24	—Greek_alphabet—
4558	21	1	5	16	
7636	20	5	1	11	—Phaeophyta—
7568	17	1	1	16	—Hebrew_alphabet—
633	17	2	1	12	—CIS—
18645	17	3	1	12	—football_league—
17187	15	1	1	14	—US_Cabinet—
4545	14	2	4	9	
1627	14	7	1	7	—kingdom¿Fungi—
575	13	1	2	11	—Dixie—,—Deep_South—
12365	13	1	1	12	—solar_system—
875	12	2	4	7	
26389	11	1	1	10	—fishing_gear—
46588	11	2	1	6	—Scorpaenidae—
14095	10	5	2	2	—army—,—regiment—
16358	10	1	1	9	—baseball_team—
22704	10	1	1	9	—Hanseatic_League—
5272	10	3	1	5	—U.S.—
28295	10	2	1	5	—Hinduism¿Brahminism—
6102	9	1	1	8	—UN—
29338	9	1	1	8	—British_Cabinet—
17039	8	1	1	7	—wedding—
27490	8	1	1	7	—Soviet_Russia—
3221	8	1	1	7	—amphibole_group—
5815	7	1	3	4	
2154	7	1	3	4	
58812	7	1	2	5	—Windsor—,—Saxe-Coburg-Gotha—
58805	7	1	1	6	—Hohenzollern—
14574	7	1	1	6	—Bloomsbury_Group—
6645	7	1	1	6	—royalty<house—
38627	6	1	1	5	—Cyperus—
6894	6	3	1	3	—military—
13497	5	1	3	2	
907	5	1	2	3	—nurse-patient_relation—,—doctor-patient_relation—
8497	5	2	2	2	—cosmos—,—constellation¿Ara—
58817	5	1	1	4	—Tudor<dynasty—
48335	5	1	1	4	—tea_set—
21872	5	1	1	4	—Beatles—
32704	5	1	1	4	—Home_Counties—
23465	5	1	1	4	—Centaurus—
37782	5	1	1	4	—Marx_Brothers—
70551	5	1	1	4	—Siberia—
6808	5	2	1	3	—navy—
6839	5	2	1	3	—ship's_company—
44823	5	2	1	3	—congeries—
26639	5	2	1	3	—electorate—
7522	5	2	1	3	—Taurus<constellation—
21011	5	2	1	3	—basketball_league—
12147	5	3	1	2	—underworld<class—
70957	5	2	1	2	—Giraffidae—
76288	5	2	1	2	—parliament—

Table 11: Component details for MEMBER components with more five or more elements; roots shown for those with one or two roots.

—gaggle<flock—
—swarm—
—school<<group—
—pod—
—herd¿remuda—
—covey<flock—
—Animalia—
—pack¿wolf_pack—
—flock<<group—
—pride—
—clowder—
—flock¿bevy—
—covert—

Table 12: Roots for the MEMBER largest component.

# parents/children	Count children	Count parents
0	2843	13
1	1561	5255
2	360	21
3	199	2
4	100	
5	63	
6	43	
7	21	
8	24	
9	25	
10	13	
11	7	
12	3	
13	3	
14	1	
15	2	
16	3	
18	5	
19	1	
20	3	
22	2	
23	1	
25	1	
26	1	
28	1	
33	1	
35	2	
40	1	
56	1	

Table 13: Parent/children node distribution for member component # 241.

ID 252, Plantae: 4688 nodes, height $\mathcal{H} = 9$. Its parent/child and rank distributions are shown in Tables 15 and 16. This is a pure tree (no multiple inheritance), with 52.5% of the nodes being leaves.

7 Part Of

An analysis of `DIRECT-HAS-PART` reveals the distribution of component sizes shown in Table 17. This is even less connected, with 111,968 lone nodes, and one largest component with only 5,915 members, or 4.5% of the structure. We note:

- The bulk of the ontology is not included in `PART-OF`, with 111,960 lone nodes comprising 92.0% of the structure; or, in other words, only 8.0% of the structure is actually included within the `PART-OF` hierarchy.
- In general the distribution is much flatter than `MEMBER`, with far more smaller components.
- The one largest component with 2,564 members has 2.1% of the structure.

	Top Rank	Bottom Rank	Width
0	13	1	28
1	34	1	176
2	73	1	1341
3	168	3	1256
4	338	4	1145
5	624	9	680
6	906	18	386
7	1140	50	145
8	1192	154	71
9	719	586	38
10	80	1621	25
11	4	2843	

Table 14: Rank distributions for member component # 241.

# parents/children	Count children	Count parents
0	2779	1
1	1278	4687
2	269	
3	126	
4	52	
5	51	
6	23	
7	18	
8	15	
9	8	
10	10	
11	5	
12	5	
13	6	
14	3	
15	7	
16	5	
17	4	
18	1	
19	3	
20	2	
22	1	
24	1	
25	4	
28	2	
29	1	
30	1	
31	1	
50	1	
51	1	
54	1	
68	1	
77	1	
103	1	
176	1	

Table 15: Parent/children node distribution for member component # 252.

	Top Rank	Bottom Rank	Width
0	1	1	388
1	12	1	2954
2	18	1	923
3	54	2	199
4	123	7	145
5	234	18	46
6	501	55	17
7	1688	250	8
8	1826	1574	8
9	231	2779	

Table 16: Rank distributions for member component # 252.

Size	# Components
2564	1
967	1
786	1
96	1
89	1
88	1
36	1
31	1
28	1
27	1
25	1
24	1
23	2
20	3
19	2
18	1
17	1
16	2
15	2
14	1
13	7
12	4
11	8
10	9
9	11
8	13
7	11
6	33
5	54
4	74
3	308
2	1201
1	111960

Table 17: Component distribution for `DIRECT-HAS-PART`.

- The second largest component with 967 members as 0.8% of the structure.

The number of roots and leaves of the components with nine or more elements is shown in Table 18, with the roots shown for any with one or two roots. This list requires more analysis, as in many cases the height and the number of roots and leaves is not as was to have ben expected.

Because of the complex component structure, the rank structure is not very meaningful, and we show the parents/children distribution in Table 19 unioned over all components. We see a moderate amount of multiple inheritance, comprising 6.0% of the structure, and including some nodes with up to 19 parents.

8 Conclusions and Next Steps

The purpose of this initial work was to understand the nature of the Omega ontology, and to verify its appropriateness for subsequent development of dispersion measures of query results.

We have verified that Omega is broadly hierarchical, with 93.2% of link instances participating in hierarhical link types. Additionally, Omega is is-a complete, dominated by a single class hierarchy with multiple inheritance.

Next steps in our work include:

New Version of Omega: A new version of Omega is due to be released in January, 2010. We look forward to repeating this analysis on that version.

Inherit Relations: In Omega properties are inherited down the subclass hierarchy. We can calculate the link distributional statistics modified by such inheritance.

Compid	Size	Height	# Roots	# Leaves	Root(s)
228	2564	7	38	1980	
131	967	9	250	530	
417	786	9	57	523	
440	96	7	6	63	
4893	89	3	16	60	
8454	88	3	5	70	
5456	36	3	5	24	
3819	31	3	1	27	—Bible—
15374	28	2	1	25	—body_armor—
10301	27	4	7	14	
15084	25	2	22	2	
20049	24	10	3	3	
3290	23	2	1	20	—electromagnetic_spectrum—
6087	23	9	1	11	—day¿today—
7817	20	2	1	18	—welkin—
8028	20	1	18	2	
38458	20	8	4	9	
10639	19	3	2	14	—meiosis—,—mitosis—
21380	19	1	2	17	—tag<<game—,—baseball—
401	18	2	1	12	—Jewish_calendar—
3876	17	1	3	14	
6806	16	2	2	12	—paper<press—,—mag—
11138	16	4	1	7	—angiosperm—
18632	15	2	1	13	—wind_scale—
35688	15	13	1	2	—t—
384	14	2	1	12	—church—
87	13	4	6	4	
3820	13	2	1	11	—publication<work—
5160	13	2	1	8	—church_calendar—
6655	13	1	1	12	—Hindu_calendar—
11364	13	12	1	1	—mym—
32158	13	1	1	12	—Revolutionary_calendar—
56711	13	1	1	12	—Muhammadan_calendar—
1441	12	2	3	7	
2846	12	2	3	7	
3163	12	2	3	8	
5037	12	2	1	10	—adulthood—
1269	11	2	3	7	
2063	11	3	3	5	
4784	11	2	1	7	—meal—
5934	11	4	1	7	—cows—
14205	11	3	3	4	
14799	11	3	1	7	—temple<building—
18442	11	1	1	10	—ATHLETICS-DECATHLON—
21786	11	2	1	9	—harness<tack—
1484	10	2	4	5	
1911	10	2	2	7	—road¿line—,—driveway—
2188	10	2	7	2	
5148	10	1	4	6	
10138	10	2	2	7	—amphibian¿toad—,—fish¿spawner—
19468	10	1	9	1	
20048	10	9	1	1	—megaton—
37889	10	2	1	7	—Cenozoic—
40632	10	2	2	7	—chimney—,—cookstove—
307	9	1	1	8	—cards—
1118	9	4	1	5	—space—
2311	9	1	3	6	
9206	9	2	1	7	—lower_respiratory_tract—
11425	9	6	1	2	—circumference<length—
14212	9	2	1	7	—links—
18615	9	3	1	5	—atmosphere<gas—
20351	9	1	6	3	
31949	9	2	1	7	—prehistory—
34651	9	2	1	7	—Paleozoic—
37987	9	8	1	1	—cubic_kilometer—

Table 18: Component details for PART-OF components with nine or more elements; roots shown for those with one or two roots.

# parents/children	Count children	Count parents
0	117950	114358
1	2409	6467
2	591	639
3	226	119
4	132	40
5	69	8
6	58	10
7	41	3
8	40	3
9	26	1
10	16	4
11	13	1
12	20	2
13	8	2
14	8	1
15	6	2
16	7	0
17	4	0
18	3	1
19	3	1
20	2	0
21	3	0
22	1	0
23	2	0
25	3	0
29	4	0
31	3	0
33	1	0
36	2	0
37	2	0
41	1	0
42	1	0
44	2	0
49	1	0
50	1	0
61	1	0
62	1	0
64	1	0
76	1	0

Table 19: Part-Of parent/children node distribution unioned across all components.

Long `HAS-PART` Components: Some components of the `HAS-PART` link type are very odd. For example, component ID 11364, headed by the single root `mym`, has size 13 and height 12. Essentially, it is a single chain. Other components are anomolously high. This needs to be examined.

Combine Link Types: While each of the three hierarchical link types `SUBCLASSES`, `DIRECT-HAS-MEMBER` and `DIRECT-PART-OF` is individually hierarchical, together they may or may not be. We will examine each of the four unions available (the three pairs and the single three-way union) to understand if they introduce any cycles. If not, it will enrich the amount of multiple inheritance in the class hierarchy.

Examine Queries: We will work with the sponsor to receive and understand appropriate test queries and/or result sets.

Centroid and Dispersion Measures: Finally we will proceed on our central task, to develop measures of centroid and dispersion appropriate for hierarchically-structured ontologies.

References

[1] Davey, BA and Priestly, HA: (1990) *Introduction to Lattices and Order*, Cambridge UP, Cambridge UK, 2nd Edition

[2] Joslyn, Cliff: (2004) "Poset Ontologies and Concept Lattices as Semantic Hierarchies", in: *Conceptual*

Structures at Work, Lecture Notes in Artificial Intelligence, v. **3127**, ed. Wolff, Pfeiffer and Delugach, pp. 287-302, Springer-Verlag, Berlin

[3] Joslyn, Cliff; Mniszewski, SM; Smith, SA; and Weber, PM: (2006) "Spindle-Viz: A Three Dimensional, Order Theoretical Visualization Environment for the Gene Ontology", in: *Joint BioLINK and 9th Bio-Ontologies Meeting (JBB 06)*, `http://www.bio-ontologies.org.uk/2006/download/Joslyn2EtAlSpindleviz.pdf`

[4] Philpot, Andrew; Hovy, Eduard; and Pantel, Patrick: (2005) "The Omega Ontology", Proc. OntoLex 2005 - Ontologies and Lexical Resources