'SuperTypes' and Logical Segmentation of Instances

by Mike Bergman - Wednesday, September 02, 2009

http://www.mkbergman.com/759/supertypes-and-logical-segmentation-of-instances/

The Significant Advantages to a Logically Segmented TBox

The Message Understanding Conferences (MUC) were initiated in 1987 and financed by DARPA to encourage the development of new and better methods of information extraction (IE). It was a seminal series that resulted in basic measures of retrieval and semantic efficacy, recall (R) and precision (P) and the combined F-measure, and other core terminology and constructs used by IE today.

By the sixth version in the series (MUC-6), in 1995, the task of recognition of named entities and coreference was added. That initial slate of named entities included the basic building blocks of *person* (PER), *location* (LOC), and *organization* (ORG); to these were added the numeric building blocks of *time*, *percentage* or *quantity*. The very terminology of *named entity* was coined for this seminal meeting, as was the idea of inline markup [1].

What is a 'Nameable Thing'?

The intuition surrounding "named entity" and nameable "things" was that they were discrete and disjoint. A *rock* is not a *person* and is not a *chemical* or an *event*. As initially used, all "named entities" were distinct individuals. But, there also emerged the understanding that some classes of things could also be treated as more-or-less distinct nameable "things": *beetles* are not the same as *frogs* and are not the same as *rocks*. While some of these "things" might be a true individual with a discrete name, such as Kermit the Frog, or The Rock at Northwestern University, most instances of such things are unnamed.

The "nameability" (or logical categorization) of things is perhaps best kept separate from other epistemological issues of distinguishing *sets*, *collections*, or *classes* from *individuals*, *members* or *instances*.

In a closed-world system it is easier to enforce clean distinctions. The Cyc knowledge base, for example, the basis for UMBEL (*Upper Mapping and Binding Exchange Layer*), makes clear the distinction

between *individuals* and *collections*. In the semantic Web and RDF, this can become smeared a bit with the favored terminology shifting to *instances* and *classes*, and in pragmatic, real-world terms we (as humans) readily distinguish John Smith as distinct from Jane Doe but don't generally (unless we're entomologists!) make such distinctions for individual beetles, let alone entire genera or species of beetles.

Under precise conditions, these distinctions are important. The fact that Cyc, for example, is assiduous in its application of these distinctions is a major reason for the overall coherence of its knowledge base. But, for most circumstances, we think it is OK to accept a distinction between "nameable" things such as frogs and beetles, but also to accept that there may be nameable individuals at times in those groupings such as Kermit that are truly an individual in that more refined sense.

This digression sets the background for a natural progression from that first MUC-6 conference. If we could cluster *persons* or *organizations*, why not other categories of distinct and disjoint things such as *frogs* or *beetles* or *rocks*?

From the first six entity categories of MUC-6 we begin to see an expansion to broader coverage. Readers of this blog will recall that I have been a fan for quite some time of the expanded coverage of 64 classes of entities proposed by BBN or the 200 proposed by Sekine [2] (as discussed, for example in the April 2008 Subject Concepts and Named Entities article). Again, the intuition was that real things in the real world could be logically categorized into discrete and disjoint categories.

Thus, "named entities" inexorably moved to become a categorization system, where the degree of familiarity and distinction dictated whether it was the individual (with a unique name, such as *Abraham Lincoln* or *Mt. Rushmore*) or groupings such as animal or plant species and their common names (such as *beetle* or *oak*) that was the standard "handle" for assigning a name to the "nameable thing".

While many can argue these individual <--> grouping distinctions and whether we are talking about true, unique, named individuals or names of convenience, I think that (at least for this blog post and discussion), that misses the real, fundamental point.

The real, fundamental point is that some "things" (whether *individuals*, *instances* or *classes*) are distinct from other "things". Such disjoint distinctions are a powerful concept that should not be lost sight of by "angels dancing on the head of a pin" epistemological arguments. A *frog* is not a *rock*, despite neither are "individuals", and how can we take advantage of that realilty?

What Works for Entities, Works for Concepts

Nearly from the outset of our work with UMBEL as a 'TBox' [3] -- that is, as a set of 20,000 or so common "subject concepts" -- the natural question was what the relation or correspondence was of these concepts to the underlying "things" (entities) that they organized. As we probed the disjoint categories within the Sekine 200 entity types, for example, we began to see significant parallels and overlap. Also gnawing at our sense of order was the rather artificial and arbitrary class of concepts in UMBEL that we termed "Abstract Concepts".

We introduced Abstract Concepts in the first release of UMBEL. When introduced, we defined "*Abstract concepts* [as] representing abstract or ephemeral notions such as truth, beauty, evil or justice, or [as]

thought constructs useful to organizing or categorizing things but are not readily seen in the experiential world." In pragmatic terms, Abstract Concepts in UMBEL were often pivotal nodes in the UMBEL subject graph necessary to maintain a high degree of concept interconnectivity.

In any world view that attempts to be more-or-less comprehensive, there is a gradation of concepts from the concrete and observable to the abstract and ephemeral. The recognition that some of these concepts may be more abstract, then, was not the issue. The issue was that there was no definable basis for segregating a concrete Subject Concept from the more Abstract Concept. Where was the bright line? What was the actionable distinction?

Off and on we have probed this question for more than a year, and have looked at what might constitute a more natural and logical ordering and segmentation within UMBEL. After many tests and detailed analysis, we are now releasing the first results of our investigations.

For, like nameable entities or things, we can see a logical segmentation of (mostly) disjoint concepts within the UMBEL TBox. Here are the summary percentages of these high-level splits:

Disjoint Concepts	90%
Attributes	1%
Classifications	9%
TOTAL	100%

(Because the analysis is still being refined, exact counts and percentages for the 20,000 concepts in UMBEL are not provided.)

Why a Logical Segmentation?

As we dove deeper into these ideas, not only could we see the basis for a logical segmentation within UMBEL's concepts, but manifest benefits from doing so as well. Remember that UMBEL's concept structure performs two main roles. It: 1) provides a coherent framework for relating and "mapping" other external ontologies; and 2) provides conceptual binding points for organizing entities and instances [4]. Via logical segmentation, we get benefits for both roles.

Here are some of the broad areas of benefit from a logical UMBEL segmentation that we have identified:

- Template-driven -- as we discuss elsewhere, Structured Dynamics also uses its ontologies to "drive applications" and the user interfaces (UI) that support them. By proper segmentation of UMBEL concepts, we are able to determine to what "cluster" of things (which we call either *dimensions* or *superTypes*; see below) a given thing belongs. This identification means we can also determine how best to display information about that "thing". This determination can include either the attributes or the display templates appropriate for that thing. For example, location-based things or time-based things might invoke map or calendar or timeline type displays.

Moreover, because of the logical segmentation of concepts, we can also use the power of the concept graph to infer more generic display templates when specific matches are absent
- Computational Efficiency -- as the percentages above indicate, once we identify what *superType* concept to which a given instance belongs, we can eliminate nearly all remaining UMBEL concepts from consideration. This logical winnowing leads to computational efficiencies at all levels in the system. The fastest computational work is not to do it, and when large chunks of data are removed from consideration, many performance advantages accrue
- Disambiguation -- via this approach we now can assess concept matches in addition to entity matches. This means we can triangulate between the two assessments to aid disambiguation. Because of these logical segmentations, we also have multiple "clusters" (that is, either the *concept*, *type*, *superType* or *dimension*) upon which to do our disambiguation evaluations, either between concepts and entities or within the various concept clusters. We can do so via either multiple semantic vectors (for statistical-based methods) or multiple features (for machine learning methods). In other words, because of logical segmentation, we have increased the informational power of our concept graph
- Structure and Integrity Testing -- the very mindset of looking for logical segmentation has led to much learning about the UMBEL structure and OpenCyc upon which it is based. In the process, missing nodes (concepts), erroneous assignments, and superfluous nodes are all being discovered. Further, many of these tests can be automated using basic logical and inference approaches. The net result is a constant improvement to the scope and completeness of the structure. Lastly, these same approaches can be applied when mapping external ontologies to UMBEL, providing similar consistency benefits.

With these benefits in mind, we have undertaken concerted analysis of UMBEL to discern what this "logical segmentation" might be. This investigation has occurred over three concentrated periods over the past year. (Intervening priorities or other work prevented concentrating solely on this task.)

We are now complete with our first full iteraton of investigation. In this post, and then the subsequent release of UMBEL version 0.80 in the coming weeks, the fruits of this effort should be evident. However, it should also be noted that we are still learning much from this new mindset and approach. UMBEL structure refinement may be likely for some time to come.

UMBEL Analysis

Most things and concepts about them are based on real, observable, physical things in the real world. Because most of these things can not occupy both the same moment in time and the same location in physical space, a useful criterion for looking at these things and concepts is disjointedness.

In a broad sense, then, we can split our concepts of the world between those ideas that are disjoint because they pertain to separable objects or ideas and those that are cross-cutting or organizational or classificatory. Attributes, such as color (pink, for example), are often cross-cutting in that they can be used to describe quite disparate things. Inherent classification schemes such as academic fields of study or library catalog systems -- while useful ways to organize the world -- are not themselves in-and-of the world or discrete from other ideas. Thus, classificatory or organizational concepts are inherently not disjoint.

With the criterion of disjointedness in hand, then, we began an evaluation process of the UMBEL subject concepts. We looked to organizational schema such as the entity types of Sekine or BBN for some starting guidance. We also kept in mind that we also wanted our categories to inform logical clusterings of possible data presentation, such as media types or locations or time.

For terminology, we adopted the term *superType* to denote the largest cluster designation upon which this disjointedness may occur. As a way to test the basic coherence of these *superTypes*, we also collected them into larger groups which we termed *dimensions*.

Our analysis process began with branch-by-branch testing of the UMBEL concept graph using automated scripts, attempting to find pivotal nodes where child instance members were disjoint from other *superTypes*. This we term the "top-down" method.

This automated analysis was then supplemented with a complete manual inspection of all unassigned and assigned concepts, with a "bottom up" assignment of concepts or corrections to the automated approach. This inspection then led to new insights and identification of missing concepts that needed to be added into UMBEL.

We are still converging between these two methods. Optimally, we should be able to tease out all UMBEL *superTypes* with a relatively few number of **union**, **intersection**, or **complement** set operations. In its current form, we are close, but there are still some rough spots.

Nonetheless, this analysis method has led us to identify some 33 *superTypes* [5], clustered into 9 dimensions. Of these, 29 *superTypes* and 8 dimensions are mostly disjoint. The one dimension of Classificatory includes the four cross-cutting *superTypes* of attributes and organizational schema that can apply to any of the 29 disjoint *superTypes*.

UMBEL superTypes

Here is the schema, with the descriptions of each:

Dimension	superType	Description/Sub-types
Natural World	**Natural Phenomena**	This *superType* includes natural phenomena and natural proces erosion, fires, lightning, earthquakes, tectonics, etc. Clouds and specifically included. Also includes climate cycles, general natu that are not specifically named, and biochemical processes and
	Natural Substances	Notable inclusions are minerals, compounds, chemicals, or phy outcome of purposeful human effort, but are found naturally oc (such as rock, fossil, etc.) are also found under this *superType*.
	Earthscape	The Earthscape *superType* consists mostly of the collection of on the surface of the Earth. Positive examples include Mountai features such as canals are excluded. Most instances of these fe space.Underground and underwater are also explicitly containe

		disjoint with Extraterrestrial (see below).
	Extraterrestrial	This *superType* includes all natural things not specifically terres (planets, asteroids, stars, galaxies, etc., that can be located with
Living Things	**Prokaryotes**	The Prokaryotes include all prokaryotic organisms, including th Bacteria, and Blue-green algas. Also included in this *superType*
	Protists or Fungus	This is the remaining cluster of eukaryotic organisms, specifica protista (protozoans and slime molds).
	Plants	This *superType* includes all plant types and flora, including flow flowering plants, gymnosperms, cycads, and plant parts and boo are also included.
	Animals	This large *superType* includes all animal types, including speci invertebrates, insects, crustaceans, fish, reptiles, amphibia, bird parts. Animal parts are specifically included. Also, groupings o Humans, as an animal, are included (versus as an individual Per excluded.
	Diseases	Diseases are atypical or unusual or unhealthy conditions for (m generally known as conditions, disorders, infections, diseases o living things and sometimes are caused by living things. This *su* impairments, disease vectors, wounds and injuries, and poisoni
	Person Types	The appropriate *superType* for all named, individual human bei includes the assignment of formal, honorific or cultural titles gi individuals. It further includes names given to humans who con (the latter case is known as an avocation). Examples include ste plumber, artisan. Ethnic groups are specifically included.
Human Activities	**Organizations**	Organization is a broad *superType* and includes formal collecti legal means, charter, agreement or some mode of formal unders geopolitical entities such as nations, municipalities or countries governments, universities, militaries, political parties, game gro trade associations, etc. All institutions, for example, are organiz informal collections of humans. Informal or less defined groupi ethnicity or tribes or nationality or from shared interests (such a lists) or expertise ("communities of practice"). This dimension identifiable human groups with set members at any given point groups, cast members of a play, directors on a corporate Board, mobs, juries, generations, minorities, etc.Finally, Organizations Industries and Programs and Communities.
	Finance & Economy	This *superType* pertains to all things financial and with respect

		chartable company performance, stock index entities, money, lo accounts and accounting, mortgages and property.
	Culture, Issues, Beliefs	This category includes concepts related to political systems, law governing societal or community behavior, or doctrinal, faith or as gods, angels, totems) governing spiritual human matters. Cul activisms (most -isms) are included
	Activities	These are ongoing activities that result (mostly) from human ef organizations to assist other organizations or individuals (in wh services, such as medicine, law, printing, consulting or teaching for leisure, fun, sports, games or personal interests (activities)
Human Works	**Products**	This is the largest *superType* and includes any instance offered commercial service. Often physical object made by humans tha facility, such as vehicles, cars, trains, aircraft, spaceships, ships weapons. Products also include the concept of 'state' (e/g/., on/o
	Food or Drink	This *superType* is any edible substance grown, made or harvest specifically includes the concept of cuisines
	Drugs	This *superType* is an drug, medication or addictive substance
	Facilities	Facilities are physical places or buildings constructed by human institutions, markets, museums, amusement parks, worship plac carstops, lines, railroads, roads, waterways, tunnels, bridges, pa All can be geospatially located.Facilities also include animal pe human "activity" areas (golf course, archeology sites, etc.). Imp infrastructure systems such as roadways and physical networks component parts that go into making them (such as foundations
Information	**Chemistry (n.o.c)**	This *superType* is a residual category (n.o.c., not otherwise cate chemical composition groupings, and the like. It is formed by w living thing (organic) substance.
	Audio Info	This *superType* is for any audio-only human work. Examples in record albums, or radio shows or individual radio broadcasts
	Visual Info	This *superType* includes any still image or picture or streaming without audio. Examples include graphics, pictures, movies, TV TV show, etc.
	Written Info	This *superType* includes any general material written by human manuscripts, but any written information conveyed via text.

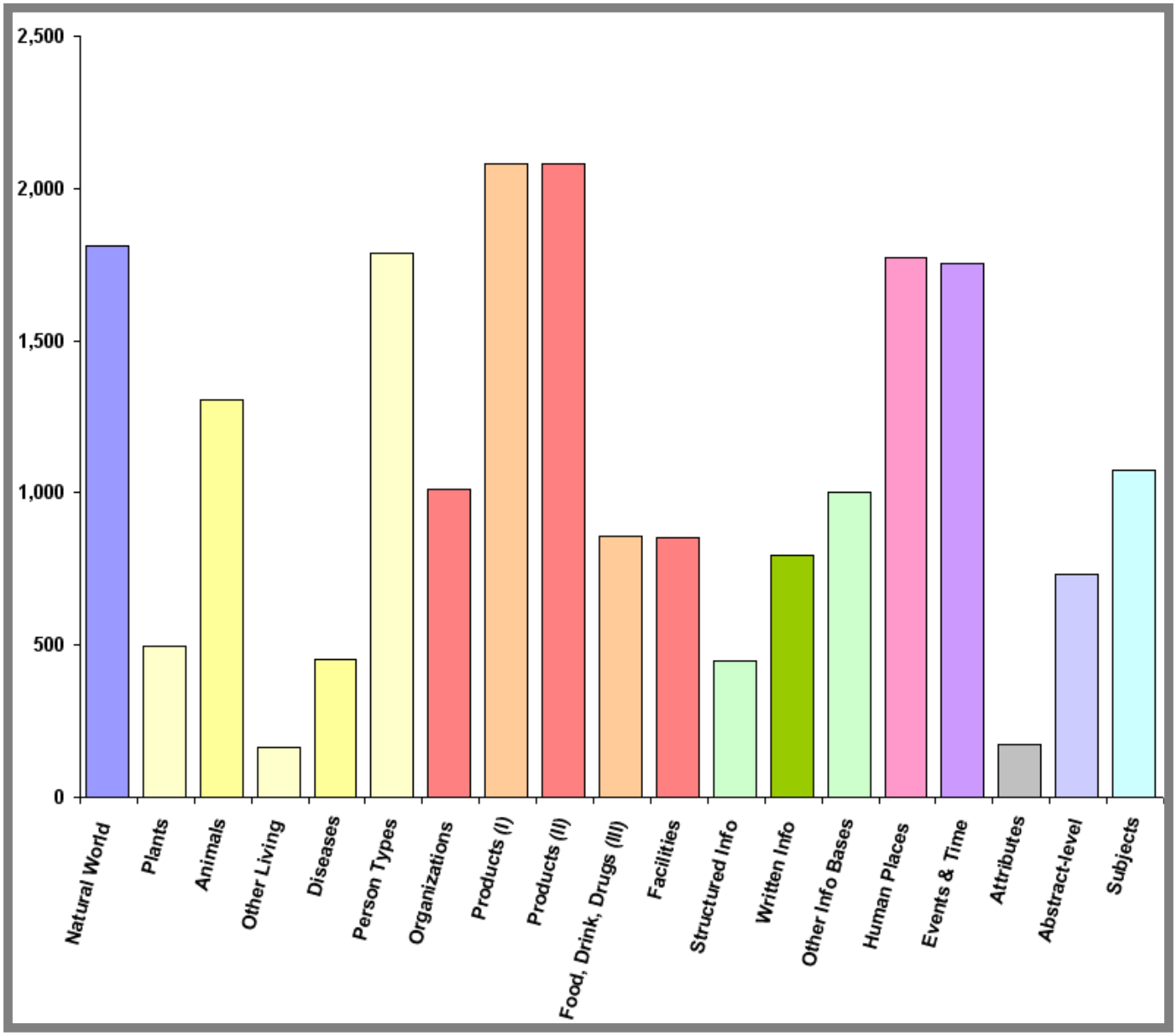
	Structured Info	This information *superType* is for all kinds of structured inform computer programs, databases, files, Web pages and structured tabular form
	Notations & References	Akin to conceptual works, these are codified means of human e human languages themselves, to more domain-specific cases su code (A-G-C-T), protocols, and computer languages, mathemat etc.Identifiers (numeric or alphanumeric identifiers for objects, such as phone numbers, URLs, zip and postal codes, SKUs, pro the various ways in which measurement, space, volume, weight calories, siesmic intensity or other quantitative descriptions of p reference types are also included in this *superType*
	Numbers	This unique *superType* is for any abstract representation of num
Human Places	**Geopolitical**	Named places that have some informal or formal political (auth Important subcollections include Country, IndependentCountry and Province.
	Workplaces, etc.	These are various workplaces and areas of human activities, ran workstations to large aggregations of people (but which are not
Time-related	**Events**	These are nameable occasions, games, sports events, conferenc natural disasters, wars, incidents, anniversaries, holidays, or not in time
	Time	This *superType* is for specific time or date or period (such as er type intervals) references in various formats
Descriptive	**Attributes**	This general *superType* category is for descriptive attributes of specific attributes in Wikipedia "infoboxes" to understand the p this *superType*. It includes colors, shapes, sizes, or other descrip an object
Classificatory	**Abstract-level**	This general *superType* category is largely composed of former some of the more abstract upper-level nodes for connecting the *superType* also includes theories or processes or methods for hu technology
	Topics/Categories	This largely subject-oriented *superType* is a means for using co classification schemes for characterizing what content "is about category are Types, Classifications, Concepts, Topics, and cont
	Markets & Industries	This *superType* is a specialized classificatory system for marke combined with the *superType* above, but is kept separate in ord oriented system.

These may undergo some further refinement prior to release of UMBEL v 0.80, and some of the definitions will be tightened up.

(Note: It should also be mentioned that some of these *superTypes* further lend themselves to further splits and analysis. The Product *superType*, for example, is ripe for such treatment.)

Distribution of superTypes

The following diagram shows the distribution of these 20,000 UMBEL concepts across major area. By far the largest *superType* is Products, even with further splits into Food and Drinks and Pharmaceuticals. The next largest categories are Person and Places and Events *superTypes*, with Organizations and Animals not far behind:

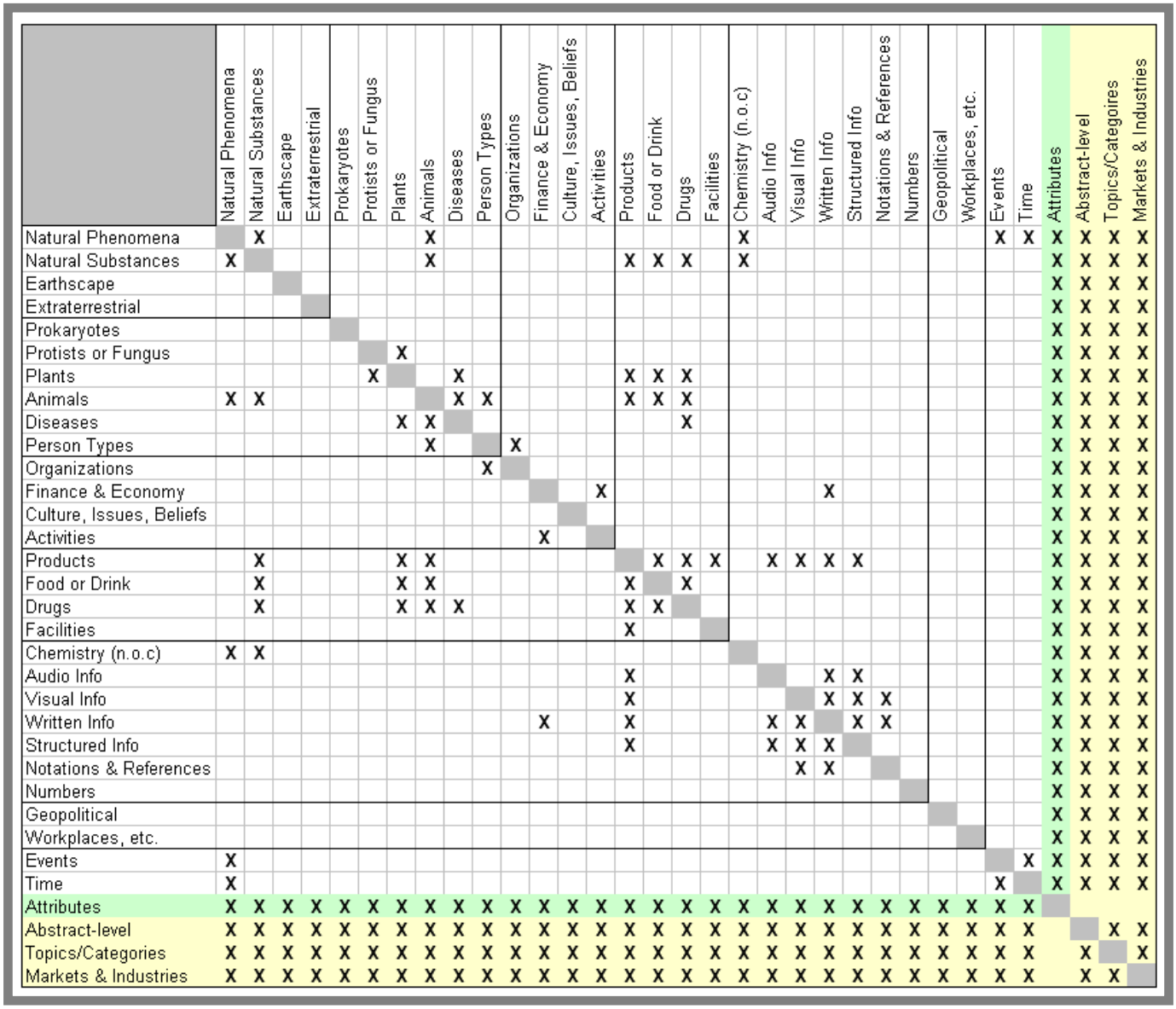


Even in its generic state, UMBEL provides a very rich vocabulary for describing things or for tying in more detailed external ontologies. There are nearly 5,000 concepts across products of all types, for example.

Possible Overlaps (non-disjoint) between superTypes

You may recall that our analysis showed 29 of the *superTypes* to be "mostly disjoint." This is because there are some concepts -- say, `MusicPerformingAgent` -- that can apply to either a person or a group (band or orchestra, for example). Thus, for this concept alone, we have a bit of overlap between the normally disjoint Person and Organization *superTypes*.

The following shows the resulting interaction matrix where there may be some overlap between *superTypes*:



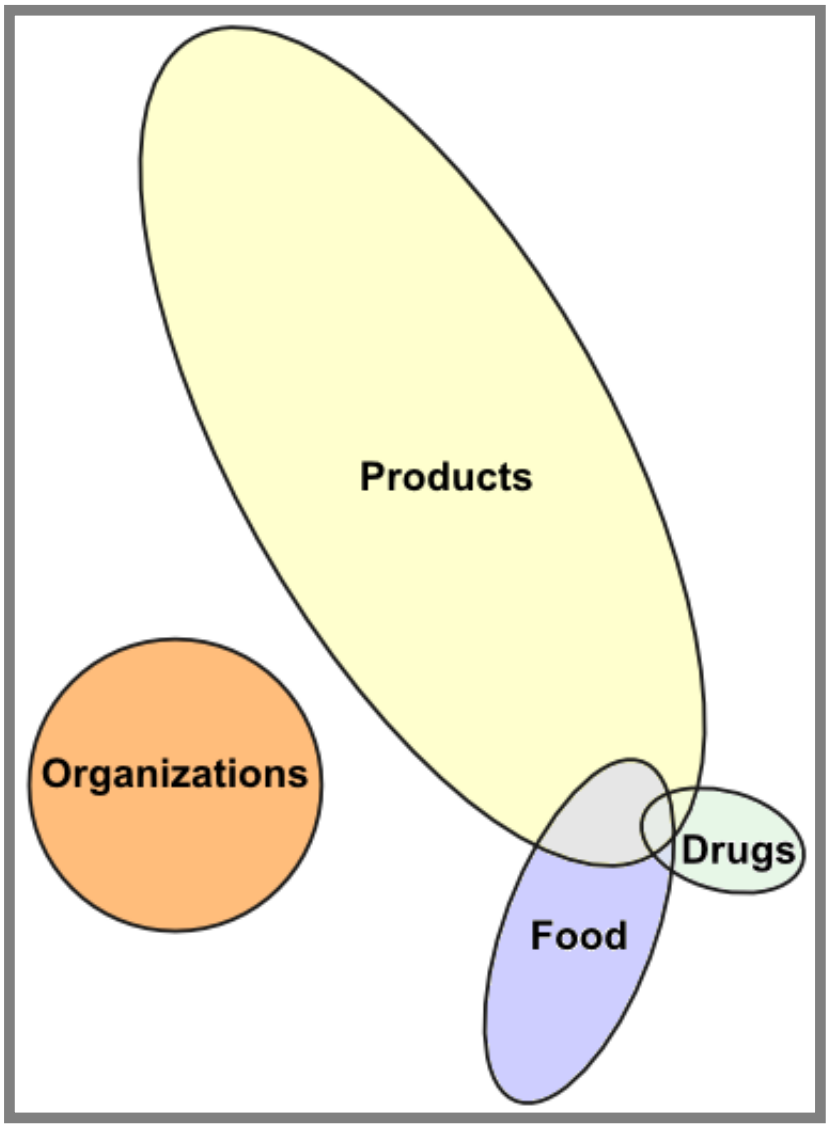
	Natural Phenomena	Natural Substances	Earthscape	Extraterrestrial	Prokaryotes	Protists or Fungus	Plants	Animals	Diseases	Person Types	Organizations	Finance & Economy	Culture, Issues, Beliefs	Activities	Products	Food or Drink	Drugs	Facilities	Chemistry (n.o.c)	Audio Info	Visual Info	Written Info	Structured Info	Notations & References	Numbers	Geopolitical	Workplaces, etc.	Events	Time	Attributes	Abstract-level	Topics/Categoires	Markets & Industries
Natural Phenomena		X						X											X									X	X	X	X	X	X
Natural Substances	X							X							X	X	X		X									X	X	X	X	X	X
Earthscape																														X	X	X	X
Extraterrestrial																														X	X	X	X
Prokaryotes																														X	X	X	X
Protists or Fungus							X																							X	X	X	X
Plants						X			X						X	X	X													X	X	X	X
Animals	X	X							X	X					X	X	X													X	X	X	X
Diseases							X	X									X													X	X	X	X
Person Types								X			X																			X	X	X	X
Organizations										X																				X	X	X	X
Finance & Economy														X								X								X	X	X	X
Culture, Issues, Beliefs																														X	X	X	X
Activities												X																		X	X	X	X
Products		X					X	X								X	X	X		X	X	X	X							X	X	X	X
Food or Drink		X					X	X							X		X													X	X	X	X
Drugs		X					X	X	X						X	X														X	X	X	X
Facilities															X															X	X	X	X
Chemistry (n.o.c)	X	X																												X	X	X	X
Audio Info															X							X	X							X	X	X	X
Visual Info															X							X	X	X						X	X	X	X
Written Info												X			X					X	X		X	X						X	X	X	X
Structured Info															X					X	X	X								X	X	X	X
Notations & References																					X	X								X	X	X	X
Numbers																														X	X	X	X
Geopolitical																														X	X	X	X
Workplaces, etc.																														X	X	X	X
Events	X																													X	X	X	X
Time	X																												X	X	X	X	X
Attributes	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X				
Abstract-level	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X			X	X
Topics/Categories	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		X		X
Markets & Industries	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		X	X	

This kind of interaction diagram is also useful for further analyzing the concept graph structure, as well.

Even Where Overlaps Occur, They are Minor

Of the 29 "mostly" disjoint *superTypes*, only a relatively few show potential interactions, and then only in

minor ways. We can illustrate this (drawn to scale) for the interaction between the Product, Food & Drink and Drug (Pharmaceuticals) *superTypes*, with the fully disjoint Organization *superType* thrown in for comparison:



Across all 20,000 concepts, then, fully 85% are disjoint from one another (5% is lost due to overlaps between "mostly" disjoint *superTypes*). This is a surprising high percentage, with even better likelihood to deliver the benefits previously noted.

Interim Conclusions and Observations

These are exciting findings that bode well for UMBEL's ongoing role and usefulness. Also, the very detailed analysis that has led to these interim findings very much reaffirms the wisdom of basing UMBEL on Cyc. Cyc showed itself to be admirably coherent and remarkably complete. (It also appears that the first versions of UMBEL were also extracted well in terms of good coverage.)

This approach now gives us an understandable and defensible basis for logical segementation of UMBEL. It also provides a much-desired alternative to the earlier Abstract Concepts, which will now be dropped

entirely as a schema concept.

One area deserving further attention is in the Attribute *superType*. We are in the process, for example, of analyzing attributes across Wikipedia and need to look through a slightly different lens at this *superType* [6]. This area is further important in its strong interaction with the Instance Record Vocabulary that is accompanying this effort on the entity side.

Another lesson for us has been to back away from the terminology of named entity, introduced at MUC-6. The expansions of that idea into other "nameable" things has caused us to embrace the "instance" nomenclature, as evidenced by our emerging IRV.

It is rewarding to prepare this next iteration release of UMBEL with its new mindset of logical segmentation and disjointedness. But -- what is also clear -- there are many treasures left to mine still hidden in the inherent structure of UMBEL and its Cyc parent.

---

[1] The original labels were ENAMEX for *entity named expression* and NUMEX for *numeric expression*. The markup format specified was also SGML. For an interesting history of this MUC-6 watershed, see Ralph Grishman and Beth Sundheim, 1996. Message Understanding Conference - 6: A Brief History, in *Proceedings of the 16th International Conference on Computational Linguistics (COLING)*, I, Kopenhagen, 1996, 466–471.

[2] In a *named entity*, the word *named* applies to entities that have a "rigid designators" as defined by Kripke for the referent. For instance, the automotive company created by Henry Ford in 1903 is referred to as Ford or Ford Motor Company. Rigid designators include proper names as well as certain natural kind of terms like biological species and substances.

Sekine's extended hierarchy proposed in 2002 is made up of 200 subtypes, with 32 larger clusters within that. Here is the top level of the Sekine type system:

Name-Other	Title	Timex	Frequency
Person	Unit	Periodx	Rank
Organization	Vocation	Numex-Other	Age
Location	Disease	Money	School Age
Facility	God	Stock Index	Latitude Longitude
Product	ID Number	Point	Measurement
Event	Color	Percent	Countx
Natural Object	Time-Other	Multiplication	Ordinal Number

Though developed separately and for different purposes, BBN categories also proposed in 2002 consists of 29 types and 64 subtypes. Here are the BBN types (Note: BBN claims 29 types because there are double entries or considerations for the first five entries):

Person	Time	Animal
NORP (adjectival GPEs)	Percent	Substance

Facility	Money	Disease
Organization	Quantity	Work of Art
GPE (geopolitical places)	Ordinal	Law
Location	Cardinal	Language
Product	Events	Contact Info
Date	Plant	Game

Of course, other entity extraction systems have similar clusterings and approaches. Though less formal in the sense of a hierarchy or purported complete entity coverage, here for example is the listing of entity types within Calais:

Anniversary	FaxNumber	NaturalFeature	RadioProgram
City	Holiday	OperatingSystem	RadioStation
Company	IndustryTerm	Organization	Region
Continent	MarketIndex	Person	SportsEvent
Country	MedicalCondition	PhoneNumber	SportsGame
Currency	Movie	Position	SportsLeague
EmailAddress	MusicAlbum	Product	Technology
EntertainmentAwardEvent	MusicGroup	ProgrammingLanguage	TVShow
Facility	NaturalDisaster	ProvinceOrState	TVStation
		PublishedMedium	URL

See further the Wikipedia entry on named entity recognition.

[3] We use the reference to "TBox" in accordance with our working definition for description logics:

"Description logics and their semantics traditionally split *concepts* and their relationships from the different treatment of *instances* and their attributes and roles, expressed as fact assertions. The concept split is known as the TBox (for *terminological* knowledge, the basis for *T* in *TBox*) and represents the schema or taxonomy of the domain at hand. The TBox is the structural and intensional component of conceptual relationships. The second split of instances is known as the ABox (for *assertions*, the basis for *A* in *ABox*) and describes the attributes of instances (and individuals), the roles between instances, and other assertions about instances regarding their class membership with the TBox concepts."

[4] UMBEL also provides a SKOS-based vocabulary extension for describing other domains and mappings between classes and instances. This purpose, however, is outside of the scope of this current article.

[5] As a reference roadmap, UMBEL was specifically designed ***not*** to include meronymous (part of) relationships (see further this reference). Thus, all "part of" type concepts were assigned to the whole *superType* category for which they are a part. Thus, "animal parts" are assigned to the *superType* Animal; "car parts" to the *superType* Product.

[6] For a general discussion of attributes and their relation to entities, see Satoshi Sekine, 2008. Extended Named Entity Ontology with Attribute Information, in *Proceedings of the 6th edition of the Language Resources and Evaluation Conference (LREC 2008)*. Marrakech, Morocco. See http://www.lrec-conf.org/proceedings/lrec2008/pdf/21_paper.pdf.
