The Why's, How's, and What-of's of Natural Language Ontology

Victor Raskin[1, 2], Christian F. Hempelmann[1, 2], Julia M. Taylor[1, 2],
Maxim S. Petrenko[3], Katrina E. Triezenberg[1], Brian Buck[1]

[1] RiverGlass, Inc.
2001 South First Street, Champaign, IL 61820, USA
{khempelmann, jtaylor, ktriez, bbuck}@riverglassinc.com
http://www.riverglassinc.com
[2] Purdue University
West Lafayette, IN 47907, USA
vraskin@purdue.edu
[3] Princess Ekaterina R. Dashkova Moscow Humanities Institute
6 Leskov Street, Moscow 127349, Russia
mpetrenk@gmail.com

ABSTRACT. After almost two decades of the design, development, justification, and testing/implementation of a natural language ontology in the Direct Meaning Access (DMA) ontological semantic school, the paper revisits the crucial issues of the relationship of its ontology-cum-lexicon to formal, engineering, and philosophical ontologies. The pertinent questions include but are not limited to: the legitimacy of its neutrality to and compatibility with any reasonably well-developed formalism; the constraints that the connection to natural language puts on its engineering; its relationship to the weak (yes) and strong (no) AI theses. An interesting characteristic of DMA is the extreme richness of its property branch, subsuming the standard logico-philosophical properties and including all the conceptual attribute material covered by the scalar adjectives of languages as well as by the non-scalars, and all the binary and non-binary relations. The attributes introduce a problematic but convenient distinction between their symbolic, absolute- and relative-measured values; the relations raise the non-trivial issue of inheritance among their domains and ranges. Most significantly, though, the paper discusses what it means for an ontology to be semantically interpreted for the computer. The appendix briefly describes an example of a major set of applications by a high-tech company which is majorly committed to the DMA technology.

1. DMA School of Ontological Semantics (Purdue Flavor)

Ontological Semantics is a system of meaning representation, whose static resources consist of a parsimonious language-independent ontology (PLIO) and a comprehensive lexicon for each natural language. Together, it is claimed, they provide a better coverage of the domains than the single controlled-vocabulary-type ontologies (CVTO) that dominate the industry because the former structure and relate the terms much more richly and reliably and enable a greater set of applications. Their machine-tractable formalism is easily translatable into virtually any other format, thus enabling almost seamless merges on the basis of a well-developed and implemented methodology. They provide an easy hybrid of human common-sense general ontology, specific domain ontologies, and scientific ontologies, all extensible and mergeable seamlessly to blend with any CVT ontology, which is treated as a lexicon, with each entry interpreted ontologically. Together with an ontological parser/analyzer/text meaning representation generator, they also allow for a comprehensive meaning processing (meaning-based natural language processing) functionality, on which many NLP applications, standard and new, can be based.

PLIO is a constructed model of reality, a theory of the world. It is a highly structured system of concepts covering the processes, objects, and properties of a domain in all of their pertinent complex relations, to the grain size determined by an application or considerations of computational complexity. Thus, an ontology may divide the root concept as shown in figure 1; EVENTs as in figure 2; OBJECTs as in figure 3; PROPERTYs as in figure 4.

Figure 1.

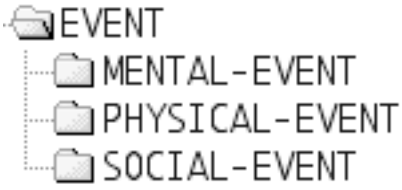


Figure 2.

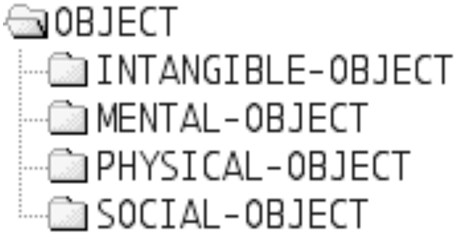


Figure 3.

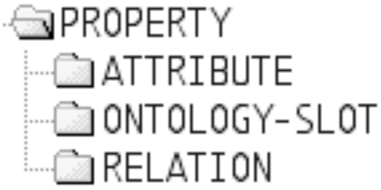


Figure 4.

Formally, then, an ontology is a tangled hierarchy (lattice) of conceptual nodes, each of which can be represented as:

concept-label
    (property-slot property-value)+

In other words, a concept has one or (usually) more properties. Every concept but the root ALL has the property IS-A, and the value of the property is the parent of this concept, the higher node—so the concept MENTAL-PROCESS, a child of PROCESS, is, on partial view, as follows:

mental-process
    is-a process
        (property-slot property-value)+

Here is a typical, albeit simplified example of an ontological concept with its properties:

inform-v1

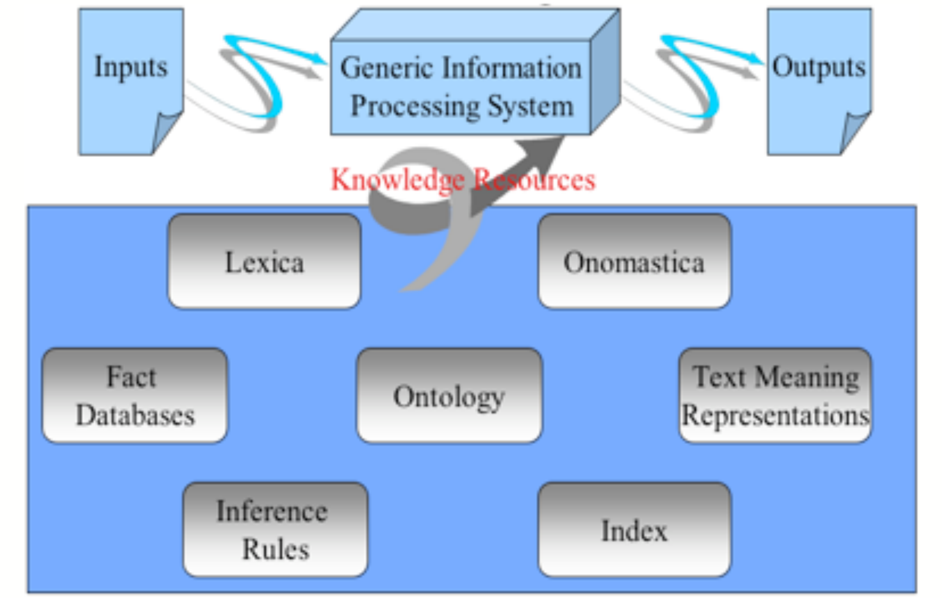
is-a	communicative-act
agent	human
theme	event
instr	communication device

The OntoSem system consists of static resources and processing software (Figure 5). The static resources are an ontology of language-independent concepts, which are linked to the words of a specific language in the lexicon, and, of course, that lexicon. Proper names may be contained in a separate onomasticon. The processing components of the system handle punctuation, morphology, and syntax in the incoming text, and assign the appropriate fillers to the property slots of the (usually) EVENT template, resulting in a Text Meaning Representation (TMR) for each sentence. Let us look at this example:

(1) The outlaws ran cocaine into the U.S.

This is a simple sentence with only one clause and one event, that of "run" in the sense of running a smuggling operation. The subject of this sentence is "outlaw", which maps to the concept CRIMINAL, which is a descendent of HUMAN, and the theme is "cocaine" which maps to the concept COCAINE, a descendent of CONTROLLED-DRUG. The parser will transform "ran" into its present singular form of "run," then examine the concept named in the sem-struc for each of the senses of "run" (in this case, there are eight possibilities). It will see that there is a sense run-v6 which has an agent that is a HUMAN or ORGANIZATION, and a theme of GUN, CONTROLLED-DRUG, or IMMIGRANT. Given this nice fit, it will choose run-v6 as the correct sense. Below is the TMR as it currently stands, and the lexicon entries for run-v6 (a verb, with its argument structure defined) and its corresponding concept SMUGGLE.

Figure 5



```
(2) SMUGGLE
        agent           value   CRIMINAL
        theme           value   COCAINE
```

```
(3) (run-v6
        (cat(v))
        (anno(def "to smuggle"))
        (syn-struc(
                (subject((root($var1))(cat(n))))
                (root($var0))(cat(v))
                (directobject((root($var2))(cat(n))))))
        (sem-struc(smuggle
                (agent(value(^$var1)))
                (theme(value(^$var2)))))
        )
```

```
(4) (smuggle
        (definition
                (value("to bring into or take out of a country illegally")))
        (is-a
                (value(crime-against-property change-location)))
        (theme
                (default(weapon controlled-drug immigrant)))
```

```
          (agent
                    (sem (human organization)))*
          (source
                    (sem (place)))*
          (destination

                    (sem (place)))*
          )
```

* Indicates that the slot was inherited from an ancestor.

What is left over in this sentence is the prepositional phrase “into the United States.” The parser will recognize that the prepositional phrase “into” denotes a DESTINATION slot, and will recognize “United States” in the onomasticon as the name of a COUNTRY. Looking into the concept for the central event SMUGGLE, the parser finds that smuggling is a change-of-location event, and as such has a SOURCE and a DESTINATION that are both loosely constrained as being places. This is in agreement with the interpretation of the sentence, so a DESTINATION slot is added to the event SMUGGLE. Below are included the TMR as it now stands, as well as the ontological entry for COUNTRY, which may be of interest because of its large number of unique slots (inherited slots have been omitted).

```
(5) SMUGGLE
          agent          value    CRIMINAL
          theme          value    COCAINE
          destination    value    COUNTRY
                                            has-name        value    “United States”

(6) (country
          (theme-of
                    (inv(transfer-possession lend economic-support trade-sanction
                    commerce-event trade-liberalization ally grant produce promise)))
          (definition
                    (value("a sovereign geopolitical entity also known as a state")))
          (has-object-as-part
                    (sem(city)))
          (is-a
                    (value(large-geopolitical-entity)))
          (standard-location
                    (sem(geographical-entity)))
          (standard-location-of
                    (inv(county)))
          (origin-of
                    (inv(asylee refugee)))
          (theme-of
                    (inv(trade-liberalization trade-sanction)))
          (has-member
```

```
                (inv(citizen)))
        (producer-of
                (inv(document)))
        )
```

Because there is only one event in this sentence there is no temporal ordering information to be included in this TMR. The sentence vaguely but not definitely indicates that the smuggling operation was successful, however the lack of a definitive indicator ("the outlaws did smuggle cocaine into the United States") indicates that it is probably wiser to not insert an epiteuctic modality into the TMR. Therefore, this sentence's very short and simple TMR will be:

```
(7) SMUGGLE
        agent           value   CRIMINAL
        theme           value   COCAINE
        destination     value   COUNTRY
                                        has-name        value   "United States"
```

2. DMA and Formal Ontology: The (In)Significance of the Formalism Choice

Many, if not most ontological approaches spend a considerable amount of energy and enthusiasm on developing and improving their own formalisms, as well as arguing the supremacy of those often very elegant and subtle formalisms over others. DMA has never participated in these discussions. Its own formalism is an outmoded and pretty transparent quasi-LISP notation, fully compatible with just about any reasonably developed formalism. If not exactly isomorphic, those are at least homomorphic with respect to that of DMA. If the Tarski principle of effability holds among formalism as it does among natural languages and—equally unintended by Tarski—programming languages, then the distinctions among formalisms are matters of convenience, taste, preference, and application-conditioned grain size.

Because the intended application of DMA is the comprehensive representation of the meaning of the text, asymptotically closing on human understanding, its primitive formalism has proven its sufficiency for expressing the grain size of any fineness, capable of extracting and explicitly representing the NL meaning of any subtlety and complexity.

The DMA formalism does deviate from the binary property slot/property filler format by adding a small set of facets in between, but the alternative formalisms can easily match it also by making the facet value and its corresponding filler value into a binary object and bifurcating each property into set of pairs, with the property slot being the first member of each such pair. Another apparent complication, equally easily manageable by alternative formalisms, is the use of bound variables between syn-struc and sem-struc in the lexicon and especially a special variable, refsemX, used inside the syn-struc, mostly purely mechanically, to avoid excessive nesting but not so in (8):

```
(8) (apotheosis-n1
          …
          (sem-struc(change-event
                    (precondition
                              (sem(event
                                        (agent(value(refsem1))(sem(human))))))
                                        (effect(sem(event
                                                  (agent(value(refsem1))(sem(deity))))))))
```

Fun and games aside, which is probably a tactless thing to say to those dealing with formalism, the crucial issues is how a formalism is not the content and how, therefore, the formalism serves the content. The lack of full understanding of this complex issue has already led to the removal of the Semantic Web from its main claim and ambition by its founder, who declared it at a recent NY “meetup” to be syntactic and usable only for structured databases. We will essay our own, far from perfect interpretation in a later section.

3. DMA and Philosophical Ontology: The Weak and Strong AI Theses

In its TMRs, DMA aspires to emulate human understanding. If successful, it will then become, at least, a black-box model of the conceptual system and lexicon, internalized in the minds of the native speakers. This means that DMA is an attempt to implement the weak-thesis AI hypothesis and thus to contribute to the proof of its feasibility. Naturally, if successful to the planned 95+% of accuracy, as evaluated by the application performances and inference/reasoning validity, the issue of the strong-thesis AI will arise. DMA is obviously not indifferent to this hypothesis: who would not give anything to learn how the mind works, and any light shed on this “last frontier” is extremely valuable, but there is no specific effort in DMA, other than optimizing the resources and TMR-generating software to achieve the highest possible degree of accuracy, to make the strong AI claim about itself or to strive to support or justify it.

4. DMA and Engineering Ontology: Natural Language Constraints

In engineering an ontology for a specific endeavor or an application [see also Nirenburg and Raskin 2004 for further discussion and references], one is motivated largely by common sense, by the experience, especially negative, of one's predecessors, and by some notion of what will contribute to optimizing the success of the endeavor or application, which is where some testing and establishing evaluation criteria may become available.

The economy of effort is always a consideration, and one is probably best off with the most parsimonious resources that are capable of achieving the same results. With DMA, there is the necessity, imposed categorically by the nature of the enterprise, to include every sense of every word and phrasal of a language in the lexicon and to represent it adequately enough ontologically for each resulting TMR, in turn, to represent adequately the meaning of every sentence (and, eventually, text). This is, of course, an extremely tall order, and the practical implementations of DMA's predecessor, Ontological Semantics (OntoSem), namely the academic proof-of-concept implementations at the Center for Machine Translation (CMT), Carnegie Mellon University, in 1987-94, Computing Research Laboratory (CRL), New Mexico State University, in 1994-2002, at the Natural Processing Laboratory, Purdue University (PNLPL), since, roughly, 1984, at the Center for Education and Research in Information Assurance and Security (CERIAS), Purdue University, since its inception in 1999, and at the Institute of Language and Information Technology (ILIT), University of Maryland Baltimore County, since 2002. "The Purdue flavor" and "the UMBC flavor," headed by the first author and Sergei Nirenburg, respectively, and actively developed by the second through fifth author for the former flavor, and Marjorie MacShane, for the latter, while both originated in Raskin's work since the 1960s and developed together in the 1980-90s, separated amicably soon after the completion of [Nirenburg and Raskin 2004] in 2001 on the role of syntax in the technology (more with UMBC, including the use of grammatical cases, than with Purdue), on the significance of theory-based non-adhoc incremental improvements in applications (Purdue, yes; UMBC, no), and on the interest in the development of commercial applications (Purdue, active; UMBC, less so). The product-level implementation of the Purdue flavor has been undertaken by AnswerChase, Inc., in 2000-2001, hakia.com in 2004-2008, RiverGlass, Inc., in 2007-2008, and Knowledge Based Systems, Inc. (KBSI) in 2007-2008. RiverGlass, Inc., took leadership in developing the new DMA version of the Purdue flavor in the Fall of 2008 (see the Appendix).

The meaning-driven and goal-oriented nature of human semantic competence places a significant constraint on the engineering of any ontology-based computational system.

Humans process natural language abductively. To generalize over a wide range of sources, abduction can be defined as an inference-based mechanism of deriving a most plausible yet potentially defeasible solution [see Aliseda 2006, Gabbay and Woods 2005 for an extensive review of the literature on the subject, and Petrenko and Raskin 2008 for the description of the nature and structure of abduction in the context of NLP].

When processing natural language elliptic input, human agents reconstruct omitted segments by projecting their background knowledge on the immediate data and proceeding inferentially to the reconstruction of a most plausibly missing segment. An NLP system should thus be rich and versatile enough to emulate the three major components of human competence: the background knowledge (i.e. the world model), the immediate knowledge (i.e. situation-dependent goals) and the set of inference rules. For the ontology engineer, several criteria should be taken into account:

- Ontology should be richly specified, i.e. contain enough concepts and properties to represent major entities of the world;
- Ontology should be versatile enough to allow the TMR to represent various semantic relations;
- Ontology should be naturally structured: the hierarchy of concepts should parallel fundamental semantic relations in human.

Within DMA, the ontological semantics-based resources must emulate human-like processing of natural language input. Most striking similarities between human competence and DMA can be observed in the functioning of the TMR-generating software. [Petrenko and Raskin 2008] illustrate how the parser would process elliptic input inferentially, i.e. through identifying case-role dependencies among certain fillers across clauses. We will focus on the features of PLIO-cum-lexicon that make it a rich knowledge resource for emulating abduction within DMA.

Abductive reasoning implies resolving ambiguity by selecting a most appropriate meaning out of a number of other available candidates (for discussion of abduction as an “inference to the best explanation” see [Aliseda 2006], [Gabbay and Woods 2005], [Walton 2004]. Richly and naturally (i.e., in a hypero-hyponymic manner but also well beyond) structured knowledge resources thus constitute an indispensable prerequisite for an abduction-based NLP application.

With its 8000+ concept-large hierarchy and hundreds of properties for each concept, PLIO constitutes a computer-emulated world model. Concepts are interrelated in a natural fashion by filling each other’s properties-slots. Each concept, in turn, has multiple fillers for one property-slot. Similarly to a wide pool of potential meanings, from which abducing humans pinpoint effortlessly the correct candidate, PLIO interprets each lexical sense of the lexicon and supplies the software with a range of possible slot fillers, at which point the inference module takes over and locates the most appropriate filler.

The versatile nature of PLIO also comes from the “slot-facet-filler” format of ontological properties, mentioned briefly in the formal-ontology section. Four basic facets—value, default, sem, and relaxable-to—capture the degree of probability of a slot-filler: from numeric/literal values and names through default and regular fillers to less probable fillers motivated metonymically or metaphorically [see also Nirenburg and Raskin 2004: 165 on facets]. When processing input, the TMR-generating software is thus able to prioritize the case role-filling procedure, first testing default fillers and then, if needed, advancing to secondary ones.

With an impoverished, domain-limited and/or artificially structured ontology, the processing potential of the inference-based software plummets significantly, its disambiguation capacity reduces adversely, and, as the result, the whole enterprise turns into another toy-system with “controlled vocabularies” and multimillion entries-large pseudo-ontologies.

5. Is It Reasonable to Dismiss Controlled Vocabulary Type Ontology as Non-?

This section responds directly to the indignation caused by the previous sentence. Most of the ontologies developed in the government, academe, and industry are, in fact, controlled vocabularies, as GeneOntology clearly states. They are developed for human manual use, primarily to unify and standardize terminology and to avoid duplication and contradiction. Other than the hierarchical IS-A property, few of them rely on any additional relations. One can find a little synonymy indication here and there, but the main bragging right is comprehensive coverage, hundreds of thousands and even over a million entries. There are the usual problems of the rules of acquisition, uniformity, consistency, and compatibility. Many CVTOs have multi-institutional acquisition committees, whose membership typically has no expertise, either theoretical or descriptive, in working with natural language and, in many cases, do not even realize that linguistics is a highly sophisticated and technical discipline, with a huge body of knowledge, very much like physics or mathematics. This leads also to the feeling of being overwhelmed, quick disappointment and loss of direction and functionality.

In DMA, the comprehensive coverage claim and the numbers game are borne out by the lexicon, which is supposed to have tens or even hundreds of thousands of entries, even though, unlike GeneOntology, it does not store all multi-word expressions as entries but rather only those whose meanings are not compositional, i.e., not directly computable from the meanings of the components. PLIO contains the minimum number of conceptual nodes that is sufficient to represent the meanings of all lexical entries in these terms. Consequently and highly desirably, PLIO has a couple of orders of magnitude fewer entries—usually, in the single digits of thousands. PLIO is, then, in an important sense, a hybrid actual and virtual (potential) ontology, with a built-in descriptive/explanatory mechanism that the CVTO, with few notable exceptions, has not yet implemented nor even discovered the need for.

The bullets and related brief comments below capture the major differences between the PLIO-cum-lexicon and a typical CVTO:

- full meaning representation, in ontological terms, of each lexical entry;
- nodes with interrelated properties: as shown above, each concept is a set of ontological properties;
- multiple properties as essential part of ontology;
- non-monotonic inheritance;
- focus on content, not formalism: results in easy compatibility with any reasonable formalism;
- acquisition toolbox—hybrid semi-automatic: severely limits subjectivity; ensures uniformity;
- tested and implemented expedient domain extension capability;

- fully automatic applications: going far beyond the noble task of terminology unification into a large set of meaning processing applications (see the last section) and on to non-natural-language ontological support for specific domains.

An interesting issue, occasionally debated within the small but quickly growing DMA community, is whether the bifurcation of the ontological resource into the ontology per se, the PLIO, and the lexicon is necessary, i.e., whether the whole resource is a complexly organized ontology. [Nirenburg and Raskin 2004: 261-265] and other DMA publications [e.g., Triezenberg 2006] discuss the distribution of the semantic material between the ontology and the lexicon and present it as an occasional judgment call: thus, there are pros and cons to setting up the ontological concepts for each kind of cat as the children of the concept FELINE, on the one hand, and to having all the appropriate literal, zoological senses of all the words for those cats to be anchored in that one concept and all the additional properties distinguishing them to be stated in the lexicon rather than in the ontology. As long as all the properties are captured correctly in either place the TMRs will be equivalent.

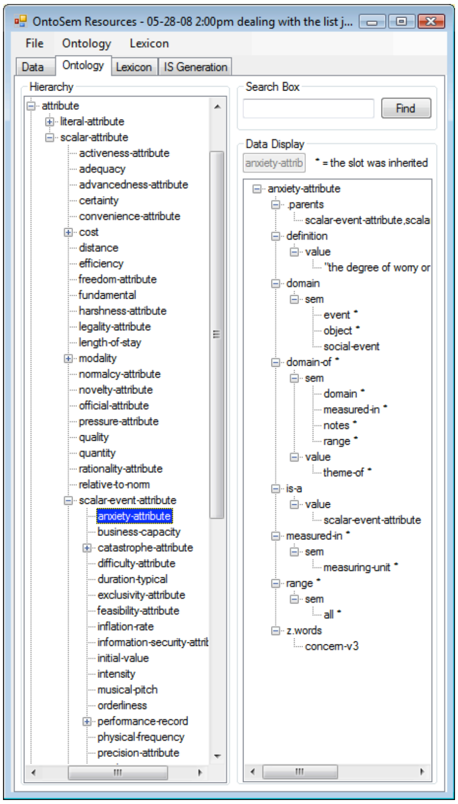
6. Hundreds of Properties?

The DMA ontology uses hundreds of properties, in a well-developed branch (see Figures 6 and 7 below).

The large number of ontological properties, by far exceeding the typical time/space/size/etc. set of logico-philosophical properties, in DMA ontology can be justified at least in two ways.

Evolutionally speaking, the DMA ontology marks the most recent phase in the development of linguistic metalanguage of semantic description. A most cursory view at the history of natural language semantics reveals the intention to design a lexicon that could capture infinitely many meanings through a limited set of semantic terms [see Raskin 1983]. Componential Analysis, especially after having been adopted by transformational theory, put this objective on a solid academic ground [see Katz and Fodor 1964; Weinreich 1970] but was hobbled by the still unmanageable profusion of semantic features during the cross-domain study [see Raskin and Weiser 1987 for details]. The theory of scripts constituted another attempt to describe linguistic meaning through a parsimonious set of non-linearly interconnected properties [see Schank and Abelson 1977, and Raskin 1985 for theoretical discussion, and Raskin et al. 2003a for description of scripts within the framework of DMA].

In its current incarnation, PLIO has enough capacity to represent any natural linguistic meaning in virtually any natural language in any domain. Most of its descriptive power comes from a highly structured network of interrelated ontological properties. The versatility of the descriptive apparatus of DMA ontology is only restricted by the purposes of an application (e.g. to cite a classical example, for an ontological domain unrelated to SPORT the information about ball weights and dimensions may be redundant.



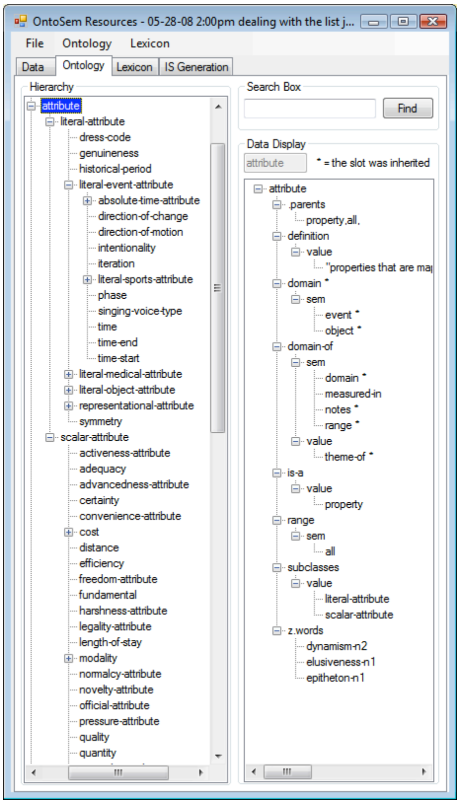


Figure 6. Figure 7.

An interesting development has occurred recently in the DMA ontology. Some ontological properties reveal a different degree of transitivity and are thus capable of forming chain-like relations of linear and/or circular structures of various lengths. To illustrate, the properties of IS-A, the mereologically well-known CONTAINS and HAS-OBJECT-AS-PART, HAS-EVENT-AS-PART, MADE-OF, etc., as well as their inverse counterparts, are transitionally projected through the whole classes and subclasses thus allowing deductive calculations of the format:

(CONCEPT-1
 (transitive-relation-1(value(CONCEPT-2
 (transitive-relation-1(value(CONCEPT-3)))))))
→ (CONCEPT-1(transitive-relation-1(value(CONCEPT-3))))

The issue of uncovered intrinsic systemic regularities in the DMA ontology undoubtedly requires attention as it would allow developing fast-working “sliding” automated mechanisms, which would enhance (albeit, at the price of increased complexity) the computational capacity of the TMR-generating software and thus further approximate the DMA NLP output to human competence.

7. What Difference Does Direct Meaning Access Make?

Yet another argument for the multiplicity of properties discussed in the previous section is that they make it possible to anchor each sense of every lexical entry accurately in a set of ontological concepts. This anchoring constitutes the true semanticity of the DMA environment as per axiomatic theories: a semantic theory interprets its syntactic terms and propositions in entities of another plane. The lexicon represents natural language; the DMA ontology, the real world (material or its mental images—fit your favorite philosophical view!) that natural languages expresses.

Most well-developed ontologies, however, lack this second plane—or, rather, it is provided by the mediating humans, who relate the NL ontological labels to the real world by virtue of the human knowledge of the NL. The absence of any second plane definitely underlies all the CVTO efforts, especially those that are keen on automatic acquisition.

There are two major advantages to the DMA effort that we are aware of—but the issue is still open to active discussion both within and without the cross-institutional government/academia/industry DMA development team. First, the DMA ontology-cum-lexicon is has a built-in NLP capability that the other technologies will take a separate effort to develop. Second and enabling the first, the DMA environment makes the meaning of NL entities explicit.

The success of NCOR clearly indicates, however, that these two advantages are not seen as critical by the serious ontological community. This remains a puzzle and warrants a vigorous bilateral discussion.

References

[Aliseda, A. 2006] Atocha Aliseda. Abductive reasoning: logical investigations into discovery and explanation. National Autonomous University of Mexico.

[Attallah et al 2001a] Mikhail J. Atallah, Victor Raskin, Michael Crogan, Christian F. Hempelmann, Florian Kerschbaum, Dina Mohamed, Sanket Naik. *Natural language watermarking: Design, analysis, and a proof-of-concept implementation*. In: I. S. Moskowitz (ed.), Information Hiding. Berlin-New York: Springer Verlag, 2001. Also CERIAS TR 2001-13.

[Attallah et al. 2001b] Mikhail J. Atallah, Victor Raskin, Christian F. Hempelmann and Dina Mohamed. *Hybrid Data and Text System for Downgrading Sensitive Documents*. CERIAS Technical Report, Purdue University, West Lafayette, Indiana. CERIAS TR 2001-19.

[Attallah et al. 2002] Mikhail J. Atallah, Victor Raskin, Christian F. Hempelmann, Mercan Karahan, Radu Sion, Umut Topkara, and Kartina E. Triezenberg. *Natural language watermarking and tamperproofing*. In: F. A. P. Petitcolas (ed.). Information Hiding. Berlin-New York: Springer Verlag, 2002. Also CERIAS TR 2002-23.

[Gabbay and Woods 2005] Dov M. Gabbay and John Woods. The Reach of Abduction. Insight and Trial. A Practical Logic of Cognitive Systems, 2. Elsevier Science, Amsterdam, the Netherlands.

[McDonough 1999] Craig McDonough. Sing-A-Password jingles: Mnemonic software for strong passwords, implemented in Visual Basic/Windows platform, Version Beta 2

[Mummoorthy 2008] Murugesan Mummoorthy. *Privacy and anonymity in text and search*. A preliminary examination document, Department of Computer Science and CERIAS, TR-2008-63, Purdue University, West Lafayette, Indiana.

[Katz and Fodor 1964] Jerrold Katz and Jerry Fodor. *The structure of a Semantic Theory*. The structure of language;readings in the philosophy of language. Englewood Cliffs: New Jersey.

[Nirenburg and Raskin 2004] Sergei Nirenburg and Victor Raskin. Ontological Semantics. Cambridge, MA: MIT Press. A pre-publication draft, chapter by chapter, can be found on ontological semantics.com.

[Petrenko and Raskin 2008] Max Petrenko and Victor Raskin. *Modeling Abduction within Ontological Semantics*. Proceedings of Midwestern Computational Linguistics Colloquium-5. Michigan State University. May 2008.

[Raskin 1983] Victor Raskin. A Concise History of Linguistic Semantics. An unpublished manuscript. Third Edition. Purdue University.

[Raskin 1985]. Victor Raskin. Semantic Mechanisms of Humor. Dordrecht: D. Reidel Publishing Company.

[Raskin et al. 2002] Victor Raskin, Segrei Nirenburg, Mikhail J. Atallah, Christian F. Hempelmann and Katrina E.Triezenberg. *Why NLP should move into IAS?* In Steven Krauwer (ed.), Proceedings of the Workshop on a Roadmap for Computational Linguistics at COLING 2002, Taipei University, Taipei, Taiwan.

[Raskin et al. 2003a] Victor Raskin , Sergei Nirenburg, Inna Nirenburg, Christian. F. Hempelmann and Katrina E. Triezenberg. *The genesis of a script for bankruptcy in ontological semantics*. In: G. Hirst and S. Nirenburg (eds.), Proceedings of the Text Meaning Workshop, HLT/NAACL 2003. ACL: Edmonton, Alberta, Canada, May 31.

[Raskin et al. 2003b] Victor Raskin, Christian F. Hempelmann and Katrina E. Triezenberg. *Semantic forensics: An application of ontological semantics to information assurance*. In: G. Hirst and S. Nirenburg (eds.), Proceedings of the Second Workshop on Text Meaning and Interpretation, ACL, Barcelona, Spain, July 25-26, 2004, Edmonton, Alberta, Canada: ACL Publication.

[Raskin and Weiser 1987] Victor Raskin and Irwin Weiser. Language and Writing: Applications of Linguistics to Rhetoric and Composition. Norwood, NJ: Ablex Publishing Corporation.

[Schank and Abelson 1977] Roger Schank and Robert P. Abelson. Scripts, Plans, Goals and Understanding. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

[Triezenberg 2006] Katrina E. Triezenberg. Ontology of Emotion. Ph. D. Thesis, Linguistics and CERIAS, Purdue University, W. Lafayette, IN, CERIAS TR-2006-83.

[Walton 2004] Douglas Walton. Abductive Reasoning. University of Alabama Press.

[Weinreich 1970] Uriel Weinreich. *Explorations in Semantic Theory*. Th. Sebeok (ed.). Current Trends in Linguistics. 3. The Hague, Paris: Mouton.

Appendix

This appendix describes the applications of the DMA ontological semantic resources and software in the commercial software products that RiverGlass, Inc., supplies to different communities to detect and mitigate threats. RiverGlass serves state, local, and federal intelligence fusion centers which are dealing with counterterrorism, counternarcotics, and violent crime issues by enabling them to:

- Automatically seek out and monitor websites that relate to criminal investigations or other intelligence needs. Information posted to the Internet from news or reference sites, as well as sites that post threatening or criminal content, is critical to effective intelligence analysis. The Web, however, continues to expand in size and analysts are faced with extreme information overload. The Web is the largest and fastest growing information resource that has ever existed—and companies take little or no advantage of it. The amount of data stored on network servers will increase six-fold by the year 2010—to a staggering 988 billion gigabytes of data. This will increasingly be unstructured data. IDC estimates that business workers spend half of their work week looking for, and making sense of, information. 1/3 to 1/2 of the time, searchers don't find what they are looking for. For organizations employing 1,000 knowledge workers, $11m is lost annually due to inability to find information and time lost getting information into a usable format once found—to say nothing of the cost to business' competitive advantage. RiverGlass tools allow analysts to intelligently and continuously scan the web for information relevant to their context—their mission, role and task. The contextual knowledge about mission, role, and task is modeled in a domain-specific ontology to drive the quick and precise discovery of semantically relevant content.
- Streamline the consumption of daily intel reports. State, local, and federal intelligence organizations share information with each other, but the resulting volume of text that analysts must read daily is a significant burden. DMA techniques support distilling information found in these periodic intelligence reports and identifying redundant sections across these reports so that analysts are not forced to read the same information twice.
- Find relevant entities, relationships, events contained within the data. DMA TMR extraction and other techniques are used to extract information and map it to the real-world concepts modeled in the domain ontology. Analysts view information in terms of the areas of an investigation modeled in the ontology and lexicon—groups, targets, methods, places, and more—and can also quickly find and access documents and integrate information from the ontology-mediated federation of public and internal data sources in a unified manner. Various visualizations are derived from this unified view of structured and unstructured data. Incoming data is monitored to alert analysts as soon as important information becomes available. The alert criteria can be anything from monitoring for simple changes in a suspect profile to sophisticated threat-assessments.

In a similar manner, RiverGlass uses DMA to serve corporations whose chief security officers, chief financial officers, intelligence analysts and research analysts are all faced with business challenges including IP theft; protecting brand equity; ensuring executive safety; guarding supply chain integrity; detecting threats to physical infrastructure from terrorist groups; conducting M&A, employee and partner due diligence; and supporting business continuity processes.