A Case for Embedded Natural Logic for Ontological Knowledge Bases

Troels Andreasen[1] and Jørgen Fischer Nilsson[2]
[1]*Computer Science, Roskilde University, Roskilde, Denmark*
[2]*Department of Mathematics and Computer Science, Technical University of Denmark, Lyngby, Denmark*

Keywords: Ontological Engineering, Natural Language Processing, Natural Logic, Domain Modeling.

Abstract: We argue in favour of adopting a form of natural logic for ontology-structured knowledge bases as an alternative to description logic and rule based languages. Natural logic is a form of logic resembling natural language assertions, unlike description logic. This is essential e.g. in life sciences, where the large and evolving knowledge specifications should be directly accessible to domain experts. Moreover, natural logic comes with intuitive inference rules. The considered version of natural logic leans toward the closed world assumption (CWA) unlike the open world assumption with classical negation in description logic. We embed the natural logic in DATALOG clauses which is to take care of the computational inference in connection with querying.

1 INTRODUCTION

This position paper discusses a novel specification language framework for logical knowledge bases. We have in mind in particular but not exclusively the life science application domains.

We take as departure the following desiderata for ontological knowledge base languages:

- Readability of the knowledge base for domain experts
- Appropriate level and range of expressivity balanced against computational complexity concerns
- Semantic rigor: a precise logical semantics, affording intuitive and computationally manageable reasoning rules
- Ability to cope with classes and relationships intensionally in querying, that is not just as extensional sets.

We argue that these desiderata may be met with the following set-up: A form of natural logic embedded in clausal logic, the latter as known from logic programming. The natural logic endorses arbitrarily complex formulations by recursive forms reflecting natural language phrase forms.

In this set-up the innermost level of natural logic serves to represent domain assertions, whereas the outermost logic provide the means of formulating logical inference rules as well as *ad hoc* domain rules. In other words we advocate a metalogic framework, where the interaction between the two language layers is facilitated by the variable free form of natural logic. Thus there is no confusion between quantified variables at the metalevel and the natural logic level. It should be mentioned that the proposal is not an attempt to merge the two mentioned logical languages into one language, cf. (Grosof et al., 2003).

As a key point in our approach the compound natural logic assertions are broken down into atomic assertions in the form of triples admitting a labeled graph representation conducive to pathway computations in the entire semantic graph.

The ideas in our approach appeared in seminal form in (Andreasen and Fischer Nilsson, 2004; Nilsson, 2011; Andreasen et al., 2013). The brief presentation here draws on and reflects on our (Andreasen et al., 2014a), which focusses on life science application domains and (Andreasen et al., 2014b), which describes the natural logic forms, and the semantic net internal representation for pathway computations in the knowledge base.

We consider knowledge bases as conventionally conceived as classes of entities and relationships expressed as logical assertions. As usual the backbone ontology is formed by the class inclusion relationship conventionally known as isa. The inclusion relation comes with inheritance of attached properties.

In addition there may be introduced domain specific relationships such as locative, causative, prop-

Andreasen T. and Fischer Nilsson J..
A Case for Embedded Natural Logic for Ontological Knowledge Bases.
DOI: 10.5220/0005156504230427
In *Proceedings of the International Conference on Knowledge Engineering and Ontology Development* (KEOD-2014), pages 423-427
ISBN: 978-989-758-049-9

erty ascribing, and partonomic as well as more domain specific relationships.

2 A NATURAL LOGIC

Natural logic is a common term for logics which resemble natural language and which are further supposed to apply more intuitive reasoning rules than mathematical logics (van Benthem, 1986; van Benthem, 2011; Muskens, 2011; Sanchez Valencia, 2004, MacCartney and Manning, 2009). The price paid is limited expressivity compared, say, with first order logic.

The general linguistic form of assertions in the natural logic considered here, called NATURALOG, is

$$Q_1\ CNterm_1\ Verb\ Q_2\ CNterm_2$$

where Q_i is either of the determiners every and some, $CNterm_i$ are common noun terms, and *Verb* is a transitive verb. In the simplest case $CNterm_i$ is a class name.

As an example, in the context of a knowledge base for biological cells we have

every eukariot has-part some nucleus

where the verb form "has part" is stylized into has-part, cf. the handling of partonomies in (Smith and Rosse, 2004).

In compound noun terms a class name forming the head noun is attached modifiers corresponding to adjectives, compound nouns, adnominal prepositional phrases, and restrictive relative clauses. The latter two modifiers have recursive syntactic structure. Individual class entities are handled by being re-conceived as singleton classes. This complies with scientific terms such as substance names being considered as class names.

In the logical conception the above NATURALOG assertion becomes

$$Q_1\ Cterm_1\ Rterm\ Q_2\ Cterm_2$$

where $Cterm_i$ are class or concept terms and *Rterm* is a binary relation name coming from the transitive verb. In compound concept terms a class name is attached modifiers. Here we consider the form [that] *Rterm* some *Cterm*, where that is an optional keyword, serving to improve readability, only. Modifiers are supposed always to restrict a class to a subclass. Presence of multiple modifiers in a class term forms a conjunction.

The above propositional form yields four quantifier cases:

every c r some d
every c r every d
some c r some d
some c r every d

The first one covers most cases in knowledge base practice. By appropriate default rules for quantifiers the sample

alphacell secrete glucagon

is interpreted logically as the proposition

every alphacell secrete some glucagon

that is, in the predicate logic explication

$$\forall x(alphacell(x) \rightarrow \exists y(secrete(x,y) \land glucagon(y)))$$

As a slightly more complex example, let us consider the natural language sentence

cells that produce glucagon reside in pancreas

In predicate logic it would be

$$\forall x(cell(x) \land \exists y(glucagon(y) \land produce(x,y)) \rightarrow \exists z(pancreas(z) \land residein(x,z))$$

In description logic:

$$cell \sqcap \exists produce.glucagon \sqsubseteq \exists residein.pancreas$$

In natural logic

(cell that produce glucagon) reside-in pancreas

or simply

cell that produce glucagon reside-in pancreas

Our $\forall\forall$ natural logic sentence every c r every d should not be confused with the description logic sentence $c \sqsubseteq \forall r.d$.

2.1 Class Inclusion

The key relationship of class inclusion, conventionally denoted isa, actually comes about as a special case of the natural logic forms every c r some d, namely with the relation r being equality. However, we use c isa d for every c equals some d.

As an example of class inclusion we can state

alphacell isa cell

By contrast we need not state that pancreatic cell isa a cell explicitly because pancreatic acts as a restrictive modifier.

By default two classes (simple or compound) are conceived to be disjoint unless either

- one is a subclass of the other, or
- they have a common subclass.

Unlike description logic classes are formally considered nonempty: $\exists x c(x)$ for all classes c. A common subclass cd is readily obtained by stating the two assertions cd isa c and cd isa d.

Unlike (Smith and Rosse, 2004) we accept taxonomic cross categories (common subclasses). For instance in our ontology the blood (cf. bloodstream) is conceived of as a bodily organ as well as an substance coming in quantities (figure 1).

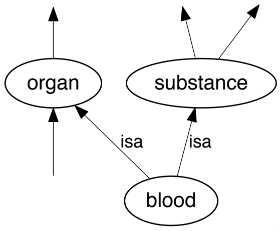

Figure 1: Blood as cross categories in the ontology.

Another example, there is presumably peptide hormone, where there are peptides which are not hormones and hormones which are not peptides. The common noun peptide hormone is an example of a compound noun formed by two class names where peptide acts as a restrictive modifier.

2.2 Reduction to Predicate Logic

Let us consider the above main quantifier case $\forall\exists$, that is

every c r some d

where c is the class name c_0 possibly with modifiers r_i c_i and analogously for d.

This case has the following backtranslation to predicate logic:

$$\forall x(c_{comp}(x) \rightarrow \exists y(r_0(x,y) \wedge d_{comp}(y)))$$

with

$$\forall x(c_{comp}(x) \leftrightarrow c_0 \wedge \bigwedge_{i=1}^{m} \exists y_i(r_i(x,y_i) \wedge c_i(y_i)))$$

$$\forall x(d_{comp}(x) \rightarrow d_0 \wedge \bigwedge_{i=1}^{n} \exists y_i(s_i(x,y_i) \wedge d_i(y_i)))$$

where c_i and d_i simple or complex with recursively the same structure

The special class inclusion case:

$$\forall x(c(x) \rightarrow \exists y(x = y) \wedge d(y)))$$

reduces to

$$\forall x(c(x) \rightarrow d(x))$$

It should be stressed that this reduction to predicate logic is for explicatory reasons, only. The applied variable-free forms are subject to reasoning at the metalogic level as to be explained next.

3 META LEVEL

We now turn to the clausal logic level into which the natural logic assertions are embedded. At the meta level NATURALOG knowledge base assertions appear encoded as data. The applied meta level logic consist of the well known DATALOG clauses

$$p_0(t_{01},...,t_{0n_0}) \leftarrow \bigwedge_{i=1}^{k} p_i(t_{i1},...,t_{in_i})$$

where the logical terms t are either constants or universally quantified variables. The variables are distinguished by upper case initial letter. The case of k being 0 yields an atomic fact $p(t_1,...,t_n)$. These clauses are occasionally enriched with use of stratified negation by non-provability for which is used the symbol $\nvdash$ referred to as DATALOG$^{\nvdash}$.

A NATURALOG assertion with simple classes, every c r some d, may then be represented straightforwardly at the meta level, say, in principle with the ground atomic

$assert\forall\exists(c,r,d)$

We remind that this framework is not an attempt to extend the natural logic with the rule language cf. (Grosof et al., 2012), since the two languages are kept at two different levels.

3.1 Decomposition of Natural Logic Assertions

Compound class terms (class names with modifiers) call for encoding with the functional logic terms being available in the general form of definite clauses. However, since we stick to DATALOG we have to decompose compound class terms.

Consider again the assertion

cell that produce glucagon reside-in pancreas

This assertion is decomposed into the DATALOG fact

$assert\forall\exists$(cell-that-produce-glucagon,
reside-in, pancreas)

where cell-that-produce-glucagon is conceived of as a new, auxiliary class name, which is in turn defined by the ground atomic facts

isa(cell-that-produce-glucagon, cell)

def(cell-that-produce-glucagon,
produces, glucagon)

This decomposition principle admits representation of unlimited complex assertions as labeled graphs as illustrated in figure 2, where the adjoined arc tells that the two edges form a definition. The graph for one NATURALOG assertion forms part of the graph conception for the entire knowledge base, cf. semantic nets (Sowa, 1991; Sowa, 2000).

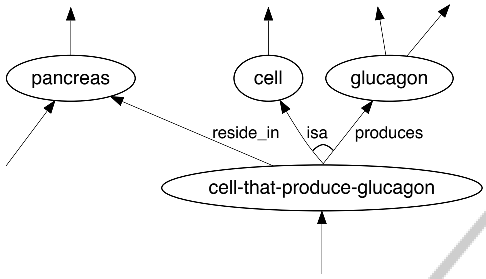

Figure 2: Complex assertions represented as labeled graphs.

3.2 Logical Inference Rules

The natural logic is supported by the logic inference rules stated in DATALOG and comprising

- partial order rules for class inclusion isa
- monotonicity rules (van Benthem, 1986, 2011) e.g.

$$assert\forall\exists(C_{sub},R,D) \leftarrow isa(C_{sub},C) \wedge assert\forall\exists(C,R,D)$$
$$assert\forall\exists(C,R,D_{sup}) \leftarrow isa(D,D_{sup}) \wedge assert\forall\exists(C,R,D)$$

- a subsumption rule which adds isa relationships following logically from the given assertions. For instance alphacell is subsumed by cell-that-produce-glucagon, cf. figure 3, because alphacell isa cell and alphacell produces glucagon
- *ad hoc* domain rules e.g. quasi-transitivity for causation and partonomy
- integrity constraints e.g. for location (it being a functional relation)

The disjointness of two classes is verified with

$$disjoint(C,D) \leftarrow \nvdash overlap(C,D)$$
$$overlap(C,D) \leftarrow isa(CD,C) \wedge isa(CD,D)$$

where the variable CD ranging over applied class names may be conceived to be existentially quantified to the right of the inverse implication.

From the point of view of ontology development the non-monotonic negation by non-provability implies that addition of new overlapping classes to the knowledge base incurs retraction of previous disjointness.

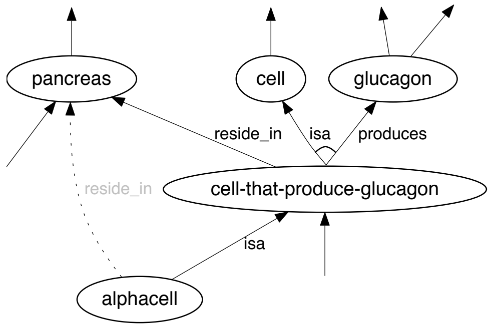

Figure 3: Inferred relationship alphacell reside-in pancreas.

3.3 Intensional Querying and Pathfinding

The NATURALOG knowledge base may now be queried deductively via the clause language appealing to the above-mentioned inference rules. The given class names are introduced by

$$class(c)$$

The concepts (simple or complex) may be queried, say, with

$$\leftarrow isa(X,c)$$

giving for variable X all concept terms below c, – or more restrictively with

$$\leftarrow class(X) \wedge isa(X,c)$$

giving all applied subordinate class names.

Figure 3 illustrates derivation of the assertion alphacell reside-in pancreas using a monotonicity inference rule.

Furthermore, the framework affords conceptual pathfinding between a pair of stated terms to be answered by computing shortest paths in the graph between the two terms, see further (Andreasen et al., 2014). In the graph conception derived assertions may act as shortcuts in the pathways. For instance, giving the pair of terms pancreas and alphacell, the inferred sentence forms a shortcut.

4 POSITION SUMMARY

We summarise our position as follows:

- Natural logic is like a stylized form of natural language and thus easy to read for domain experts.
- Predicate logic is "unreadable" and complex for practical reasoning tasks.
- Natural logic possesses intuitive reasoning rules.

- Description logics is "unnatural" from a knowledge engineering point of view by enforcing copula form (c' ISA c'') As an alternative to our natural logic approach, (de Azevedo, 2014) generates internal description logic representations.
- We wish to distinguish definitional (analytic) and empirical (synthetic) facts.
- We prefer CWA in favour of classical negation e.g. for class disjointness.
- Our natural logic comes with a semantic graph form facilitating computational pathfinding and intensional querying.

5 CONCLUDING REMARKS

We have advocated use of natural logic embedded in clausal logic for ontology structured knowledge bases as an alternative to the prevailing use of description logic dialects and derivatives. Our approach differs from description logic approaches primarily in that it recognizes and supports the role of the main verb in knowledge base assertions. Moreover, the two level set-up affords an intensional view in which class terms appear in query answers.

We are in the process of building a prototype system as a test bed for domain specifications within selected bio- and medico-domains for ascertaining whether this is a viable approach meeting the desiderata in the introduction. The DATALOG level in a scaling up of the prototype may readily be realized by appealing to state-of-the-art relational database technology offering efficient access to massive data.

An open issue in knowledge bases is the handling of denials. The use of CWA seems appealing since it departures with classes being born disjoint in accordance with scientific practice in classification. Moreover, it opens for means of dealing with exceptions in the non-monotonic fashion.

REFERENCES

Andreasen, T. and Fischer Nilsson, J. (2004) Grammatical Specification of Domain Ontologies. *Data & Knowledge Engineering*, vol: 48, issue: 2, pages: 221-230.

Andreasen, T., Bulskov, H., Fischer Nilsson, J., Jensen, P. A., Lassen, T. (2013). Conceptual Pathway Querying of Natural Logic Knowledge Bases from Text Bases. In *10th international conference on Flexible Query Answering Systems*, Springer-Verlag, Berlin, Heidelberg, pages 1-12.

Andreasen, T., Bulskov, H., Fischer Nilsson, J., Jensen, P. A. (2014a). Computing Pathways in Bio-Models Derived from Bio-Science Text Sources. In *IWBBIO International Work-Conference on Bioinformatics and Biomedical Engineering*, Proceedings, Granada April 7-9, 2014, ISBN 84-15814-84-9, pages 217-226.

Andreasen, T., Bulskov, H., Fischer Nilsson, J., Anker Jensen, P. (2014b). Computing Conceptual Pathways in Bio-Medical Text Models. In *Foundations of Intelligent Systems - 19th International Symposium, ISMIS 2014*, Roskilde, Denmark, June 28-30.

de Azevedo, R. R., Freitas, F., Rocha, R. Menezes, J. A., Pereira, L. F. A. (2014). Generating Description Logic $\mathcal{ALC}$ from Text in Natural Language In proceedings of *Foundations of Intelligent Systems - 21th International Symposium*, ISMIS 2014, Roskilde, Denmark, June 25-27.

van Benthem, J.(1986). Essays in Logical Semantics, *Studies in Linguistics and Philosophy*, Vol. 29, D. Reidel Publishing Company.

van Benthem, J. (2011). Natural Logic, Past And Future, Workshop on Natural Logic, Proof Theory, and Computational Semantics 2011, CSLI Stanford. http://www.stanford.edu/~icard/logic&language/index.html

Fischer Nilsson, J. (2011). Querying class-relationship logic in a metalogic framework. In Proceedings of the 9th international conference on Flexible Query Answering Systems (FQAS'11), Henning Christiansen, H., De Tré, G., Yazici, A., Zadrozny, S., and Andreasen, T. (Eds.). Springer-Verlag, Berlin, Heidelberg (2011) 96-107

Grosof, B. N. G., Horrocks, I., Volz, R., and Decker, S. (2003): Description Logic Programs: Combining Logic Programs with Description Logic. In Proceedings of the *12th international conference on World Wide Web (WWW '03)*. ACM, New York, NY, USA, pages 48-57.

MacCartney, B. and Manning, C. (2009): An Extended Model of Natural Logic. In Bunt, H., Petukhova, V., and Wubben, S., (eds), Proceedings of the *8th IWCS*, Tilburg, pages 140-156.

Muskens, R. (2011). Towards Logics that Model Natural Reasoning, Program Description Research program in Natural Logic, http://lyrawww.uvt.nl/~rmuskens/natural/

Sanchez Valencia, V. (2004). The Algebra of Logic, in Gabbay, D. M. & Woods, J. (eds.). *Handbook of the History of Logic*, Vol. 3 The Rise of Modern Logic: From Leibniz to Frege, Elsevier.

Smith, B. and Rosse, C. (2004). The Role of Foundational Relations in the Aligment of Biomedical Ontologies, *MEDINFO 2004*, M. Fieschi *et al.*.

Sowa, J. F. (ed.). (1991). *Principles of Semantic Networks*, Morgan Kaufmann.

Sowa, J. F. (2000). *Knowledge Representation, Logical, Philosophical, and Computational Foundations*, Brooks/Cole Thomson Learning.