CERIAS Tech Report 2006-76

Ontology-Based Inference Methods

by Olga Krachina, Victor Raskin
Center for Education and Research
Information Assurance and Security
Purdue University, West Lafayette, IN 47907-2086

Ontology-Based Inference Methods

Olga Krachina
NLP lab, CERIAS
Purdue University
West Lafayette, IN 47907
okrachin@purdue.edu

Victor Raskin
NLP lab, CERIAS
Purdue University
West Lafayette, IN 47907
vraskin@purdue.edu

Abstract

This research focuses on the development of an inference mechanism based on a particular variety of non-axiomatic systems known as Ontological Semantics. Systems with a heavy semantics emphasis and dynamic learning capabilities indicate a greater potential in inference-related applications, largely due to the structure of the resources which allows implementation of mixed methods: traditional deduction as well as framework-specific methods. An example illustrating the flow of the inference procedure is provided for clarity.

1 Introduction

The ultimate goal of field of automated reasoning is to model and emulate human thinking processes. The basis for any inference process is a formal system. The choice of a formal system is a crucial step, given that inference methods are, to a large degree, framework-dependent. Attempts in the past have focused mainly on the deductive systems. However, non-axiomatic frameworks prove to be a more viable and adequate choice in inference process.

Axiomatic systems are systems consisting of axioms, i.e. unquestionable atomic statements, and theorems that are derived from the axioms. Many logic systems fall into axiomatic category: first-order, temporal, predicate, modal logics, etc. Axiomatic systems are inherently closed systems, in other words, the axioms are fixed. Using an axiomatic system as a basis for the inference engine is appealing for several reasons: conceptual clarity, ability to well-define models, and generality, in the sense that theorems hold in all models. However, the advantages introduce a share of limitations in the application to the inference process: axiomatic systems, being tautological, are unable to discover new knowledge (Lakatos, 1976), closed systems do not provide the means to build complex models from simpler ones, and the inference process requires completeness of knowledge resources.

Limitations of the axiomatic systems served as the motivating factor for the non-axiomatic approach. The fundamental premise of a non-axiomatic system is the necessarily insufficient knowledge base. Other properties include:

- finiteness: the system works within its constant information processing capacity
- openness: no constraint on the knowledge and questions that the system encounters
- real-time: both the knowledge and queries may arrive at any time, and queries may have time requirements attached
- adaptive: the system self-improves its behavior on an assumption that future experience[1] will be similar to past experience (Wang, 1994).

Systems resulting from research in the non-axiomatic direction include Non-Axiomatic Reasoning System (NARS) (Wang, 1997), Ontological Semantic (OntoSem) (Raskin, 2004). The difference between the two is architectural, i.e., in the types of knowledge sources used and specifics of their interaction. This paper will focus on OntoSem and its inference capabilities.

[1] "Experience" is indicated by the history of communication with its environment

2 Paradigm for Automated Reasoning

The following components comprise a reasoning system: the formal language, semantic theory, set of inference rules, memory structure for knowledge resources, and control mechanisms (Wang, 1994). Evaluation procedure is a concluding step that is user-controlled under current model.

Many analytical tasks that involve gathering, correlating, and analyzing information can be naturally formulated as question answering problems (QA), hence ontology-based inference methods are discussed in the context of QA.

3 OntoSem: Inference Capabilities

OntoSem uses the following knowledge resources:

- ontology, a language-independent tangled hierarchy (lattice) of concepts, each with a set of properties, representing the theory of the world;
- lexicons for specific natural languages, with most lexical entries anchored in an ontological concept, often with the constraints on the property fillers;
- onomasticons for specific natural languages, which are lexicons for proper names;
- language-independent text-meaning representation (TMR) language for representing the meaning of a text in ontological terms;
- fact repository (FR), the database of recorded TMRs.

The inference process in OntoSem consists of expanding and subsequent matching of TMR modules corresponding to input-text TMR (TMR_I) and query TMR (TMR_Q).

3.1 TMR: Construction

To quote a not totally accurate but convenient characterization, a TMR is "a network of frames, each of which has a head and a list of binary relations that link the head to a frame, a pointer to another frame, a simple value, or a more complex combination for defaults, semantic types, relaxable types, etc. Each TMR is a set of six kinds of frames: one or more *propositions*, zero or more *discourse relations*, zero or more *modalities*, one *style*, zero or more *references*, and one *TMR time*" (Sowa, 2005).

Due to the limited scope of this paper, the process of TMR construction is not described in full detail-- for an extensive discussion, see Nirenburg and Raskin, 2004. The terminology utilized in the framework is briefly illustrated below:

(HEAD
    (SLOT   FILLER)$^+$

The flow of the inference process is best demonstrated on a particular example of an input text and a query. Two major categories of queries that will be examined are the wh- and yes-no types; the input as well as queries are provided below.

Input: *He asked UN to authorize the war (from a text about Colin Powell).*

Query 1: *What did Colin Powell ask UN to do?*

Query 2: *Has Colin Powell spoken with Emyr Jones Parry*[2]*?*

The underlying algorithm for process of inference is given in listing 1.

Listing 1. Algorithm Underlying Inference Process in Ontological Framework.

```
TARGET = THEME-OF REQUEST-INFORMATION
if ((TARGET := OBJECT) || (TARGET := EVENT)
          && (HAS-NAME (non-empty slot)))
    INFO = SEARCH_FDB(TARGET)
    case_1: SINGLE_ENTRY
          return INFO
    case_2: MULTIPLE_ENTRIES
          do REFINE_VARIABLE(TARGET)
    case_3: NO_ENTRY
          do DIRECT_TMR_MATCHING
          APPEND_FDB(TARGET)
          RESULT = TARGET
else
          do TMR_EXPANSION (PREMISE_SET)
          RESULT = EVALUATE(PREMISE_SET)
          do ERROR_ESTIMATION
          RESULT = RESULT + ERROR
return (RESULT)
```

The corresponding initial TMR_I and TMR_Q are given in listing 2(a) and 2(b), respectively.

Listing 2. TMR_I and TMR_Q corresponding to the input and the queries.

(a) TMR_I

REQUEST-ACTION

[2] Head of Security Council (2003-present)

```
        AGENT              HUMAN
        THEME              ACCEPT
        BENEFICIARY        ORGANIZATION
        SOURCE-ROOT-WORD   ask
        TIME               (< (FIND-ANCHOR-TIME))

ACCEPT
        THEME              WAR-ACTIVITY
        THEME-OF           REQUEST-ACTION
        SOURCE-ROOT-WORD   authorize
ORGANIZATION
        HAS-NAME           UNITED-NATIONS
        BENEFICIARY-OF     REQUEST-ACTION
        SOURCE-ROOT-WORD   UN
HUMAN
        HAS-NAME           COLIN POWELL
        AGENT-OF           REQUEST-ACTION
        SOURCE-ROOT-WORD   he
WAR-ACTIVITY
        THEME-OF           ACCEPT
        SOURCE-ROOT-WORD   war
```

(cf. Beale *et al.*, 2004)

(b) TMR_Q

```
REQUEST-ACTION
        AGENT              HUMAN
        THEME              EVENT
        BENEFICIARY        ORGANIZATION
        SOURCE-ROOT-WORD   ask
        TIME               (<(FIND-ANCHOR-TIME))
ORGANIZATION
        HAS-NAME           UNITIED-NATIONS
        BENEFICIARY-OF     REQUEST-ACTION
        SOURCE-ROOT-WORD   UN
HUMAN
        HAS-NAME           COLIN POWELL
        AGENT-OF           REQUEST-ACTION
        SOURCE-ROOT-WORD   he
REQUEST-INFORMATION
        THEME              request-action.theme
```

(c) TMR_Q

```
CONVERSATION
        AGENT              HUMAN-1
        SOURCE-ROOT-WORD   speak
        ACCOMPANIER        HUMAN-2
HUMAN-1
        HAS-NAME           COLIN POWELL
        AGENT-OF           CONVERSATION
        SOURCE-ROOT-WORD   he
HUMAN-2
        HAS-NAME           EMIR JONES PARRY
        ACCOMPANIER-OF     HUMAN-1
REQUEST-INFOMRATION
        THEME              conversation.accompanier[3]
```

It is worth noting that REQUEST-INFORMATION head indicates the fact that the TMR represents a query, while the type of the query is reflected in the slots and fillers of REQUEST-INFORMATION, where the dot notation is used to formulate query in terms of the properties of the object or event in question.

3.2 TMR: Matching and Expansion

In general, the processing of wh-type queries is less complex as compared to yes-no-type queries, i.e., the query can be satisfied through a direct search of FR. In this case, the target variable is initialized with the filler of slot in REQUEST-INFORMATION field of TMR_Q.

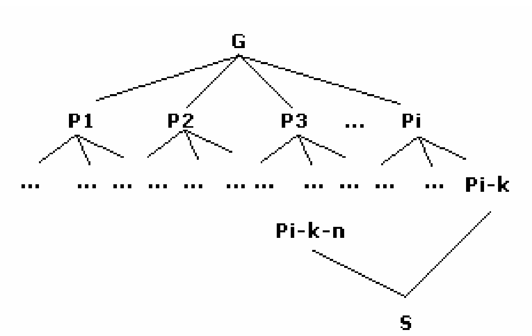
Query 1 is an example of the wh-query. In this case, the third condition of the if-statement does not hold, hence the flow of the process is transferred to the *direct TMR matching*. Direct TMR matching is a process of finding the answer to the query in the TMR_I. The procedure is ‘direct’ because one of the heads of the TMR_I matches the target variable. It is worth noting, that the answer is a chain of dependencies starting from the target variable itself and ending with the last relevant head of the frame. The dependency chain for the example discussed in this paper is

REQUEST-ACTION.THEME = ACCEPT (THEME WAR-ACTIVITY).

Query 2 is a yes-no query. Tracing the algorithm outlined in listing 1, control will be transferred to the “else” part of the “if-else” statement. Furthermore, the direct TMR matching procedure will fail to produce a result, since the CONVERSATION head in the TMR_I is not available. At this point, the PREMISE-SET procedure would be invoked. PREMISE-SET is a procedure that updates two relations in TMR_Q: the CASE-ROLES proper (including INVERSES) and the subsumption relations. Thus, comparing TMR_I and TMR_Q *a priori*, the task before the system is to establish a connection between BENEFICIARY—ACCOMPANER and REQUEST-ACTION – CONVERSATION. Given the target variable as CONVERSATION.ACCOMPANIER and the failure of the direct TMR procedure, the system attempts to find the *principal heads*, i.e., the heads of TMR_I and TMR_Q that are related through the Most Common Intermediate Node (MCIN). Given a parent node G and P*i* children nodes and P*i-k-n* being children nodes of P*i-k* nodes, MCIN is a node such that the distance between P*i* and P*i-n* is minimal, as demonstrated in figure 2.

[3] ACCOMPANIER and ACCOMPANIER-OF are assumed to be isomorphic, i.e., inverses of each other that map identically.

Figure 2. MCIN demonstrated.



First, the subclasses[4] of concept REQUEST-ACTION are examined and matched against TMR_I. If not found, the properties (slots and fillers) of the parent node of the TMR_Q concept are examined and compared to those of the heads of the TMR_I. In the given example, COMMUNICATIVE-EVENT is the parent of CONVERSATION. Two candidates from TMR_I, namely ACCEPT and REQUEST-ACTION, that inherit properties of COMMUNICATIVE-EVENT are identified. However, the number of principal heads needs to be reduced to one per TMR. Since ACCEPT is THEME of REQUEST-ACTION, it will be evaluated after REQUEST-ACTION. REQUEST-ACTION inherits properties of COMMUNICATIVE-EVENT. The distinctive property of COMMUNICATIVE-EVENT with respect to its sister nodes is INSTRUMENT (COMMUNICATION-DEVICE, NATURAL-LANGUAGE). Error estimation is based on the depth difference of the two ontological concept nodes in question. Currently, a basic probability assignment is assumed where the children of the same parent node have the same probability. The error associated with the connection between REQUEST-ACTION and CONVERSATION amounts to be $1/2\times1/5\times1/6\times1/2 = 8.33\times10^{-3}$. Thus, the connection between REQUEST-ACTION and CONVERSATION is complete.

The next step is to examine the REQUEST-ACTION entry. Direct TMR matching is used to identify any common properties which are directly recorded in the PREMISE-SET if they are in the form of case-roles proper. Next, the proper names, if any, are identified and searched for in FR. The result of the searches is expressed as a set of property-based premises. TMR_Q is being updated after every FR search. If during the TMR expansion process the added slots are not CASE-ROLES but rather HAS-NAME slots that in turn are filled with proper names, the FR search continues.

The target search entry is determined on the basis of the relation of concept fillers of previously searched entry since slots do not necessarily indicate a hierarchy. Thus in current example, relevant part of the Onomasticon entry for Emyr Jones Parry is given in listing 3.

Listing 3. Relevant Part of Onomasticon Entry for E.J. Parry

```
EMYR JONES PARRY
        ...
        EMPLOYED-BY     UN
        HEAD-OF         SECURITY-COUNCIL
        …
```

Listing 4 shows updated TMR_Q.

Listing 4. Updated TMR_Q after the FR search of E.J. Parry entry.

```
CONVERSATION
        AGENT                   HUMAN-1
        SOURCE-ROOT-WORD        speak
        ACCOMPANIER             HUMAN-2
HUMAN-1
        HAS-NAME                COLIN POWELL
        AGENT-OF                CONVERSATION
        SOURCE-ROOT-WORD        he
HUMAN-2
        HAS-NAME                EMIR JONES PARRY
        ACCOMPANIER-OF          HUMAN-1
        EMPLOYED-BY             ORGANIZATION
        HEAD-OF                 ORGANIZATION-DIVISION
ORGANIZATION
        HAS-NAME                UN
ORGANIZATION-DIVISION
        HAS-NAME                SECURITY-COUNCIL
REQUEST-INFOMRATION
        THEME                   conversation.accompanier
```

The question is which of the proper names is to be searched next: UN or SECURITY-COUNCIL? Hierarchical structure of Ontology as well the inheritance mechanism employed in the framework allows making the optimal decision.

A more accurate relation between the filler concepts of HUMAN-2 is established through looking at the filler concepts. The MCIN of the fillers is found and the depth with respect to the MCIN is calculated, and the HAS-NAME filler of the concept with the greater depth is the next candidate to be set as a search variable. Following the example, the final expanded TMR_Q is given in listing 5. PREMISE-SET is updated every time a CASE-ROLE is encoun-

[4] The reader is suggested to refer to the KBAE tool to follow the discussion of particular examples of ontological entries.

tered; the final premise set for this example is given in listing 6.

Listing 5. Expanded TMR_Q.

```
CONVERSATION
          AGENT              HUMAN-1
          SOURCE-ROOT-WORD   speak
          ACCOMPANIER        HUMAN-2
HUMAN-1
          HAS-NAME           COLIN POWELL
          AGENT-OF           CONVERSATION
          SOURCE-ROOT-WORD   he
HUMAN-2
          HAS-NAME           EMIR JONES PARRY
          ACCOMPANIER-OF     HUMAN-1
          EMPLOYED-BY        ORGANIZATION
          HEAD-OF            ORGANIZATION-DIVISION
ORGANIZATION
          HAS-NAME           UN[5]
ORGANIZATION-DIVISION
          HAS-NAME           SECURITY-COUNCIL
          PART-OF-OBJECT     ORGANIZATION
          AGENT-OF           ALLOW
REQUEST-INFOMRATION
          THEME              conversation.accompanier
```

Listing 6. Final Premise Set.

1. REQUEST-ACTION $\subseteq$ COMMUNICATIVE-EVENT
2. CONVERSATION $\subseteq$ COMMUNICATIVE-EVENT
3. $\therefore$ CONVERSATION $\subseteq$ REQUEST-ACTION
4. COLIN POWELL AGENT-OF REQUEST-ACTION
5. PRAGANIZATION BENEFICIARY-OF REQUEST-ACTION
6. ORGANIZATION-DIVISION $\subseteq$ ORGANIZATION $\therefore$ HEAD-OF $\subseteq$ EMPLOYED-BY $\therefore$ HUMAN-2 $\subseteq$ ORGANIZATION
7. ORGANIZATION-DIVISION AGENT-OF ALLOW
8. ALLOW THEME WAR-ACTIVITY
9. REQUEST-ACTION THEME WAR-ACTIVITY
10. ALLOW $\subseteq$ ILLOCUTIONARY-ACT
11. REQUEST-ACTION $\subseteq$ ILLOCUTIONARY-ACT
12. $\therefore$ ILLOCUTIONARY-ACT THEME WAR-ACTIVITY
13. COLIN-POWELL AGENT-OF REQUEST-ACTION
14. ORGANIZATION BENEFICIARY-OF REQUEST-ACTION THEME ACCEPT THEME WAR-ACTIVITY
15. HUMAN-2 $\subseteq$ ORGANIZATION
16. HUMAN-2 BENEFICIARY-OF REQUEST-ACTION AGENT COLIN POWELL

[5] Note, that this head concept is not expanded any further since the procedure halts, i.e., returns RESULT.

The last inference, necessary to complete the premise set, is the connection between BENEFICIARY-OF and ACCOMPANIER-OF. Direct TMR matching fails in this case; moreover, TMR expansion cannot proceed since all possible inferences have been recorded in the premise set and TMR_Q was updated. The solution is to examine the slots and fillers of the entries in question. In this case, BENEFICIARY-OF and ACCOMPANIER-OF are sister nodes under CASE-ROLE-INVERSE. However, the fact that the case-roles are sisters is not a sufficient condition to conclude that the two case-roles are equivalent. In this particular example, the domains and ranges are compared, and since three of the case-roles-inverse share the same filler for DOMAIN, the error of 1/3 is assigned to the result. Thus the last inference is recorded into premise set:

17. BENEFICIARY-OF $\cong$ ACCOMPANIER-OF

Error estimation is based on the total number of tokens in the original query and number of correctly inferred tokens. Probability of error in the example discussed in this paper (five tokens) is calculated as follows:

$$1 - (3/5 + 1/3 \times 1/5 + 1/5 \times (1 - 0.00833)) = \\ = 0.1347$$

Final result:

HUMAN-2 (ACCOMPANIER-OF (HUMAN-1 (AGENT-OF CONVERSATION))) [ERROR: 0.1347]

3.3 Conclusion

The purpose of this research has been to investigate the inference capabilities of a non-axiomatic system Ontological Semantics. The advantage of using OntoSem for inference-related applications is in the structure of the system resources and consequently reduced probability of the combinatorial explosion. Even though, at the core of the inference process, lies a high-quality TMR, there still exists a certain degree of ambiguity, which is expressed as error probabilities.

Further research is directed at testing more queries as well as the development of more prescribed procedures and of a more rigorous methodology for error assignment. Additional issues include evaluation of query satisfaction and implementation-related problems.

3.4 References

Stephen Beale, Benoit Lavoie, Marjorie McShane, Sergei Nirenburg, Tanya Korelsky. N.d. Question Answering Using Ontological Semantics.

KBAE. Knowledge Base Acquisition Editor. http://kbae.cerias.purdue.edu:443/ (guest login "guest," password, "nlpgroup").

Imre Lokatos. 1976. *Proofs and Refutations*. Cambridge University Press, Cambridge, UK.

Sergei Nirenburg and Victor Raskin. 2004. *Ontological Semantics.* MIT Press, Cambridge, MA.

John Sowa. 2005. Review of Ontological Semantics by Sergei Nirenburg and Victor Raskin. *Computational Linguistics*, 31(1): 147-152.

Pei Wang. 1984. From Inheritance Relation to Non-Axiomatic Logic. *International Journal of Approximate Reasoning*, 7:1-74.

Larry Wos. 1988. *Automated Reasoning: 33 Basic research problems*. Prentice-Hall, Englewood Cliffs, NJ.