Explaining Answers from the Semantic Web

Deborah L. McGuinness Paulo Pinheiro da Silva

Knowledge Systems Laboratory, Stanford University
Stanford, CA 94305, USA.
e-mail: {dlm,pp}@ksl.stanford.edu

Abstract

The Semantic Web lacks support for explaining answers from web applications. When applications return answers, many users do not know what information sources were used, when they were updated, how reliable the source was, or what information was looked up versus derived. Many users also do not know how implicit answers were derived. The Inference Web (IW) aims to take opaque query answers and make the answers more transparent by providing infrastructure for presenting and managing explanations. The explanations include information concerning where answers came from (knowledge provenance) and how they were derived (or retrieved). In this article we describe an infrastructure for IW explanations. The infrastructure includes: IWBase – an extensible web-based registry containing details on information sources, reasoners, languages, and rewrite rules; PML – the Proof Markup Language specification used for encoding portable proofs; and a proof and explanation browser. Source information in the IWBase is used to convey knowledge provenance. Representation and reasoning language axioms and rewrite rules in the IWBase are used to support proofs, proof combination, and Semantic Web agent interoperability. The IW browser is used to support navigation and presentations of proofs and their explanations. The Inference Web is in use by four Semantic Web agents, three of them using embedded reasoning engines fully registered in the IW. Additional reasoning engine registration is underway in order to help provide input for evaluation of the adequacy, breadth, and scalability of our approach. Inference Web also provides explanation infrastructure for a number of DARPA and ARDA projects.

1 Introduction

Inference Web (IW) aims to enable applications to generate portable and distributed justifications for any answer they produce. IW addresses needs that arise with systems performing reasoning and retrieval tasks in heterogeneous environments such as the web. Users (humans and computer agents) need to

decide when to trust answers before they can use those answers with confidence. We believe that the key to trust is understanding. Explanations of knowledge provenance and derivation history can be used to provide that understanding McGuinness and Pinheiro da Silva (2004). In one simple case, users retrieve information from individual or multiple sources and they may need knowledge provenance (e.g., source identification, source recency, authoritativeness, etc.) before they decide to trust an answer. Users may also obtain information from systems that manipulate data and derive information that was implicit rather than explicit. Users may need to inspect information contained in the deductive proof trace that was used to derive implicit information before they trust the system answer. Many times proof traces are long and complex so users may need the proof transformed (or abstracted) into something more understandable that we call an explanation.

Some users will decide to trust the deductions if they know what reasoner was used to deduce answers and what data sources were used in the proof. Other users may need additional information including how an answer was deduced before they will decide to trust the answer. Users may also obtain information from hybrid and distributed systems and they may need help integrating answers and solutions. As web usage grows, a broader and more distributed array of information services becomes available for use and the needs for explanations that are portable, sharable, and reusable grows. Inference Web addresses the issues of knowledge provenance with its registry infrastructure called IWBase. It also addresses the issues concerned with inspecting proofs and explanations with its browser. It addresses the issues of explanations (proofs transformed by rewrite rules for understandability) with its language axioms and rewrite rules. IW addresses the needs for combination and sharing with its Proof Markup Language specification.

In this article, we include a list of explanation requirements gathered from past work, literature searches, and from surveying users. We present the Inference Web architecture and provide a description of the major IW components including the PML specification Pinheiro da Silva et al. (2004), the IWBase registry McGuinness and Pinheiro da Silva (2003b); Pinheiro da Silva et al. (2003) (containing information about inference engines, proof methods, ontologies, and languages and their axioms), the explanation mechanism, and the justification browser. We also provide some simple usage examples. We conclude with a discussion of our work in the context of explanation work and state our contributions with respect to trust and reuse. This article is an expanded and updated version of an earlier conference paper McGuinness and Pinheiro da Silva (2003a). The primary updates include the integration with the Proof Markup Language, a description of the IWBase architecture, and a broadening of the work to explain query plans and satisfiability results.

2 Background and Related Work

Recognition of the importance of explanation components for reasoning systems has existed in a number of fields for many years. For example, from the early days in expert systems (e.g., MYCIN Shortliffe (1976)), expert systems researchers identified the need for systems that understood their reasoning processes and could generate explanations in a language understandable to its users. Inference Web attempts to stand on the shoulders of past work in expert systems, such as MYCIN and the Explainable Expert System Swartout et al. (1991) on generating explanations.

IW also builds on the learnings of explanation in description logics (e.g., Borgida et al. (1999, 2000); McGuinness (1996); McGuinness and Borgida (1995)) which attempt to provide a logical infrastructure for separating pieces of logical proofs and automatically generating follow-up questions based on the logical format. IW goes beyond this work in providing an infrastructure for explaining answers in a distributed, web-based environment possibly integrating many question answering agents using multiple reasoners. IW also attempts to integrate learnings from the theorem proving community on proof presentation(e.g., Boyer et al. (1995); Felty and Miller (1987)) and explanation (e.g., Huang (1994)), moving from proof tracing presentation to abstractions and understandable explanations. IW attempts to learn from this and push the explanation component started in Huang's work and also add the emphasis on provenance and distributed environments.

The work in this article also builds on experience designing query components for frame-like systems Borgida and McGuinness (1996); McGuinness (1996); Fikes et al. (2002) to generate requirements. The foundational work in those areas typically focus on answers and only secondarily on information supporting the understanding of the answers. In our requirements gathering effort, we obtained requirements input from contractors in DARPA-sponsored programs concerning knowledge-based applications (the High Performance Knowledge Base program[1], Rapid Knowledge Formation Program[2], and the DARPA Agent Markup Language Program[3]) and more recently, the ARDA AQUAINT[4] and NIMD[5] programs and DARPA's IPTO Office programs. We also gathered requirements from work on the usability of knowledge representation systems (e.g., McGuinness and Patel-Schneider (2003)) and ontology environments (e.g., Das et al. (2002)). We have also gathered needs from the World Wide Web Consortium efforts on CWM[6] and the related reasoner effort on Euler[7]. Finally, we gathered knowledge provenance requirements from the programs above and from previous work on data provenance from the database community(e.g., Buneman et al. (2001)) and more recently from work integrating

[1] http://reliant.teknowledge.com/HPKB/
[2] http://reliant.teknowledge.com/RKF/
[3] http://www.daml.org
[4] http://www.ic-arda.org/InfoExploit/aquaint/
[5] http://www.ic-arda.org/Novel_Intelligence/
[6] http://www.w3.org/2000/10/swap/doc/cwm.html
[7] http://www.agfa.com/w3c/euler/

information from extractors such as the work in Tap[8] Guha et al. (2003) leading to our enhanced knowledge provenance infrastructure Pinheiro da Silva et al. (2003) and information integrators (e.g., ISI's Prometheus mediator[9] which uses information obtained from Fetch's[10] wrappers in appropriate domains). Additionally requirements have been more recently obtained from initial efforts to explain text analytics work (e.g., IBM's UIMA Ferrucci and Lally (2004) as well as initial efforts to explain semantic matches using satisfiability engines (e.g., Giunchiglia and Shvaiko (2003)).

3 Requirements

If humans and agents need to make informed decisions about when and how to use answers from applications, there are many things to consider. Decisions will be based on the quality of the source information, the suitability and quality of the reasoning/retrieval engine, and the context of the situation. Particularly for use on the web, information needs to be available in a distributed environment and be interoperable across applications.

3.1 Support for Knowledge Provenance Information

Even when search engines or databases simply retrieve asserted or "told" information, users (and agents) may need to understand where the source information came from with varying degrees of detail. Similarly, even if users are willing to trust the background reasoner in a question answering environment, they may need to understand where the background reasoner obtained its ground facts. Information about the origins of asserted facts, sometimes called provenance, may be viewed as meta information about told information. Knowledge provenance requirements may include:

- Source name (e.g., CIA World Fact Book). If facts are encountered in multiple sources, any integrated solution needs to have a way of identifying from which source information was taken.
- Date and author(s) of original information and any updates
- Authoritativeness of the source (is this knowledge store considered or certified as reliable by a third party?)
- Degree of belief (is the author certain about the information?)
- Degree of completeness (Within a particular scope, is the source considered complete. For example, does this source have information about all of the employees of a particular organization up until a some date? If so, not finding information about a particular employee would mean that this person is not employed, counting employees would be an accurate response to number of employees, etc.)

[8] http://tap.stanford.edu
[9] http://www.isi.edu/info-agents/Prometheus/
[10] http://www.fetch.com

The information above could be handled with meta information about content sources and about individual assertions. Additional types of information may be required if users need to understand the meaning of terms or implications of query answers.

- Term or phrase meaning (in natural language or a formal language)
- Term inter-relationships (ontological relations including subclass, superclass, part-of, etc.)

As a system addresses meta information about information, many additional issues come into play such as security, access, efficiency, and usage. We have separated these into a separate list since they may appear to be a secondary in that they arise as a result of meeting the needs of the initial knowledge provenance demands. There is overlap on many of these requirements with those placed on sophisticated database applications.

- Unique identifiers for provenance information
- Effective methods for indexing, storing, and querying provenance information
- Persistence of provenance information
- Support for privacy levels in storage and access
- Support for views based on a number of criteria such as privacy level, topic, thread, etc.
- Support for reuse of provenance information – tool support may be required for relating meta-information using unique identifiers and querying methods

3.2 Support for Reasoning Information

Once systems do more than simple retrieval, additional requirements result. If information is manipulated as a result of integration, synthesis, abstraction, deduction, etc., then users may need access to a trace of the manipulations performed along with information about the manipulations as well as information about the provenance. We refer to this as reasoning traces or proof traces. Requirements as a result of reasoning may include the following:

- The reasoner used
- Reasoning method (e.g., tableaux, model elimination, etc.)
- Inference rules supported by the reasoner
- Reasoner soundness and completeness properties
- Reasoner assumptions (e.g., closed world vs. open world, unique names assumption, etc.)

- Reasoner authors, version, etc.

The previous points all address meta information concerning the reasoner. The next set of requirements arise from using a reasoner and working with it in the context of an answer. These include:

- Detailed trace of inference rules applied (with appropriate variable bindings) to provide conclusion
- Term coherence (is a particular definition incoherent?)
- Were assumptions used in a derivation? If so, have the assumptions changed?
- Source consistency (is there support in a system for both A and $\neg$A)
- Support for alternative reasoning paths to a single conclusion
- Support for accessing alternative reasoning paths to the same conclusion
- Support for accessing the implicit information that can be made explicit from any particular reasoning path

3.3 Support for Explanation Generation

While knowledge provenance and proof traces may be enough for expert logicians when they attempt to understand why an answer was returned, usually they are inadequate for a typical user. For our purposes, we view an explanation as a transformation of a proof trace into an understandable justification for an answer. With this view in mind, we consider techniques for taking proofs and proof fragments and rewriting them into abstractions that produce the foundation for what is presented to users. In order to handle rewriting, details of the representation and reasoning language must be captured along with their intended semantics. Additionally, users may need to know both *what* manipulations were done (i.e., what rules of inference were used) as well as *how* manipulations were done (i.e., what was the plan used to obtain information, were resource limitations in place, etc.) Support for both kinds of proof traces and their abstractions into explanations are needed in many applications. Requirements for explanations may include:

- Representation language identification
- Representation language descriptions (e.g., DAML+OIL, OWL, RDF, ...)
- Axioms capturing the semantics of the representation languages
- Description of rewriting rules based on language axioms

Much of the past work on explanation, whether from expert systems, theorem proving, or description logics, has focused on single systems or integrated

systems that either use a single reasoner or use one integrated reasoning system. Systems being deployed on the web are moving to distributed environments where source information is quite varied and sometimes question answering systems include hybrid reasoning techniques. Additionally multi-agent systems may provide inference by many applications. Thus many additional requirements for proofs and their explanations may arise from a distributed architecture. Some requirements we are addressing are listed below:

- Reasoner result combinations (if a statement is proved by one system and another system uses that statement as a part of another proof, then the second system needs to have access to the proof trace from the first system).
- Portable proof interlingua (if two or more systems need to share proof fragments, they need an language for sharing proofs).
- Support for registering translators and comparators that can be used to translate statements from one language to another and can be used to identify similarities and differences between statements.

3.4 Support for Proof Presentation

If humans are expected to view proofs and their explanations, presentation support needs to be provided. Human users will need some help in asking questions, obtaining manageable size answers, asking follow-up question, etc. Additionally, even agents need some control over proof requests. If agents request very large proofs, they may need assistance in breaking them into appropriate size portions and also in asking appropriate follow-up questions. Requirements for proof presentation may include:

- A method for asking for explanations (or proofs)
- A way of breaking up proofs into manageable pieces
- A method for pruning proofs and explanations to help the user find relevant information
- A method for allowing proof and explanation navigation (including the ability to ask followup questions)
- A presentation solution compatible with web browsers
- A way of seeing alternative justifications for answers
- A way of translating into different presentation formats (e.g., natural language, graphs, etc.)

4 Use Cases

Every query-answering environment is a potential new context for the Inference Web. We provide two motivating scenarios and use the second scenario for our

examples throughout the article. Consider the situation where someone has analyzed a situation previously and wants to retrieve this analysis. In order to present the findings, the analyst may need to defend the conclusions by exposing the reasoning path used along with the source of the information. In order for the analyst to reuse the previous work, s/he will also need to decide if the source information used previously is still valid (and possibly if the reasoning path is still valid).

Another simple motivating example arises when a user asks for information from a web application and then needs to decide whether to act on the information. For example, a user might use a search engine interface or a query language such as OWL-QL[11] for retrieving information such as "zinfandels from Napa Valley" or "wine recommended for serving with a spicy red meat meal" (as exemplified in the wine agent example in the OWL guide document Smith et al. (2003)). A user might ask for an explanation of why the particular wines were recommended as well as why any particular property of the wine was recommended (like flavor, body, color, etc.). The user may also want information concerning whose recommendations these were (a wine store trying to move its inventory, a wine writer, etc.). In order for this scenario to be operationalized, we need to have the following:

- A way for applications (reasoners, retrieval engines, etc.) to dump justifications for their answers in a format that others can understand. This supports the distributed proofs requirements above. To solve this problem we introduce a portable and sharable proof specification called the Proof Markup Language.

- A place for receiving, storing, manipulating, annotating, comparing, and returning meta information used to enrich proofs and proof fragments. To address this requirement, we introduce the IWBase for storing the meta information and the Inference Web registrar web application for handling IWBase data. This addresses the issues related to knowledge provenance.

- A way to present justifications to the user. Our solution to this has multiple components. First the IW browser is capable of navigating through proof dumps provided in PML format. It can display multiple formats including KIF[12] and English. Additionally, it is capable of using rewrite rules (or tactics) to abstract proofs in order to provide more understandable explanations. Using this combination, we address issues related to reasoning, explanations, and presentation.

5 Inference Web

The Inference Web framework contains the following:

[11] http://ksl.stanford.edu/projects/owl-ql/
[12] http://logic.stanford.edu/kif/kif.html

- data used for representing proofs, explanations, and meta information about proofs and explanations;
- software tools and services used for building, maintaining, presenting, and manipulating proofs.

In terms of data, the Inference Web provides the Proof Markup Language (PML) – an OWL-based specification for documents representing both proofs and proof meta information. PML classes are OWL classes (thus they are subclasses of `owl:Class`) and they can are either *proof elements* (proof level concepts) or *provenance elements* (provenance level concepts).

Inference Web proofs and explanations are PML documents built using proof elements and referring to provenance elements, as described in Section 5.1. Proofs using the PML format become a portion of the Inference Web data used for combining and presenting proofs and for generating explanations. Figure 1 presents an abstract and partial view of the Inference Web framework[13] showing proofs and explanations published anywhere in the web. Inference Web data also has a distributed repository of PML documents describing provenance information about proof elements such as sources, inference engines and inference rules, as described in Section 5.2. IWBase is an infrastructure within the Inference Web framework for proof meta information, as described in Section 5.3.

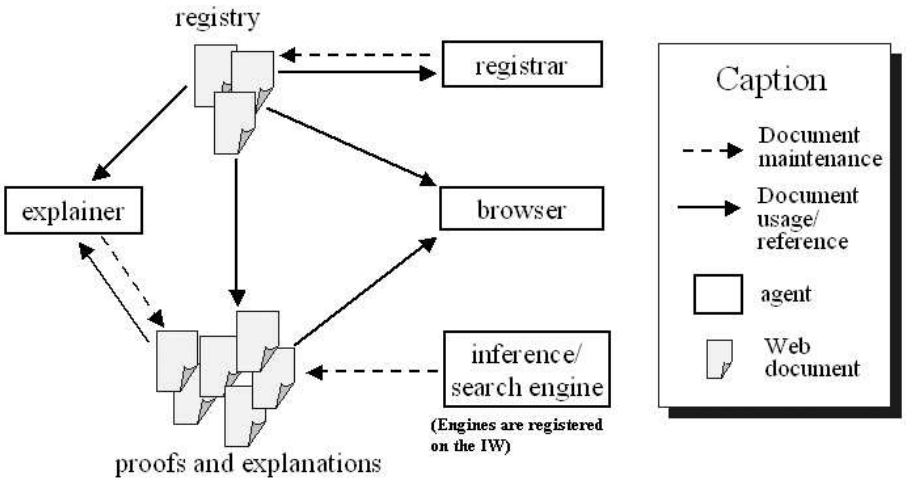

Figure 1: Inference Web framework overview.

In terms of software, Inference Web tools include: the registrar for handling IWBase entries (also described in Section 5.3); the explainer for abstracting proofs into explanations, as described in Section 5.4; the browser for displaying proofs, as described in Section 5.5; and planned future tools such as proof

[13] A more detailed view is available at http://ksl.stanford.edu/software/IW/details.shtml

web-search engines, proof verifiers, proof combinators, and truth maintenance systems. Figure 1 presents how IW data is used by some of the IW tools mentioned above. For instance, it shows that the explainer has PML proofs as inputs and outputs. Explainer outputs are proof abstractions, or explanations. In this article, we limit our discussion to the PML specification (and an associated API), IWBase architecture (and the associated registrar tools and proof generation services), explanations, and the browser.

5.1 PML Proof Elements

Our PML specification includes two major components for building proof trees: `NodeSets` and `InferenceSteps`. Figure 2 presents a typical dump of an IW node set. A `NodeSet` represents a step in a proof whose conclusion is justified by any of a set of inference steps associated with the `NodeSet`. PML adopts the term "node set" since each instance of `NodeSet` can be viewed as a set of nodes gathered from one or more proof trees having the same conclusion. The `iw:hasConclusion` property of a node set represents the expression concluded by the proof step. Every node set has one conclusion, and a conclusion of a node set is represented in the language specified by the `iw:hasLanguage` property of the node set. In the example, the node set has a conclusion stating that the color of `W1` is `?x` or the value of the color property of `Wine1` is the item of interest. The node set represents a statement and the last step in a deductive path that led a system to derive the statement.

In general, each node set can be associated with multiple or single inference steps as presented by the `iw:isConsequentOf` property of the node set in Figure 2. A proof can then be defined as a tree of inference steps explaining the process of deducing the consequent sentence (a more formal definition of proofs within PML documents is described in Pinheiro da Silva et al. (2004)). In terms of number of files, an IW proof can physically vary from a single PML file containing all its node sets to many PML files, each one containing a single node set. Also, PML files containing node sets can be distributed in the web. Considering the IW requirement that proofs need to be combinable, it is important to emphasize that an IW proof is a forest of trees since the nodes of a proof tree are sets of inference steps. In contrast with typical proof trees that are composed of nodes rather than node sets, every theorem in an IW proof can have multiple justifications.

An `InferenceStep` represents a justification for the conclusion of a node set. Inference steps are anonymous OWL classes defined within node sets. For this reason, it is assumed that applications handling PML proofs are able to identify the node set of an inference step. Also for this reason, inference steps have no URIs. For an IW proof, an `InferenceStep` is a single application of an inference rule, whether the rule is primitive or derived as discussed in Section 5.3. Inference rules (such as modus ponens) can be used to deduce a conclusion from any number of antecedents (that are the conclusions of other node sets). Inference steps contain URI references to node sets concluding its antecedents, the inference rule used, the supporting sources for the justification,

```
<?xml version='1.0'?> <rdf:RDF (...)>
<iw:NodeSet rdf:about='../sample/IW1.owl#IW1'>
    <iw:hasConclusion rdf:datatype='string'>
         (wines:COLOR W1 ?x)
    </iw:hasConclusion>
    <iw:hasLanguage>
      <iw:Language rdf:about='.../registry/LG/KIF.owl'/>
    </iw:hasLanguage>
    <iw:isConsequentOf>
         (a NodeSet can be associated to a set of Inference steps)
      <rdf:List>
        <rdf:first>
        <iw:InferenceStep>
          <iw:hasInferenceRule>
            <iw:DeclarativeRule rdf:about='../registry/IR/GMP.owl'/>
          </iw:hasInferenceRule>
          <iw:hasInferenceEngine>
            <iw:InferenceEngine rdf:about='../registry/IE/JTP.owl'/>
          </iw:hasInferenceEngine>
            (...)
          <iw:hasAntecedent>
    (inference step antecedents are IW files with their own URIs)
            <rdf:List>
              <rdf:first>
                <iw:NodeSet rdf:about='../sample/IW3.owl#IW3'/>
              </rdf:first>
              <rdf:rest>
                <rdf:List>
                  <rdf:first>
                    <iw:NodeSet rdf:about='../sample/IW4.owl#IW3'/>
                  </rdf:first>
                  <rdf:rest rdf:resource="rdf:nil"/>
                </rdf:List>
              </rdf:rest>
            </rdf:List>
          </iw:hasAntecedent>
          (...)
        </iw:InferenceStep>
        </rdf:first>
        <rdf:rest rdf:resource="rdf:nil"/>
       </rdf:List>
    </iw:isConsequentOf>
  </iw:NodeSet>
</rdf:RDF>
```

Figure 2: A PML node set.

and any variable bindings used in the step. There is no source associated with the node set in Figure 2 since it is derived (although it could be derived and associated with a source). If it had been asserted, it would require an association to a source, which is typically an ontology that contains it. The antecedent sentence in an inference step may come from conclusions in other node sets, existing ontologies, extraction from documents, or they may be assumptions.

With respect to a query, a logical starting point for an IW proof is a proof fragment containing the last inference step used to derive a node set that contains an answer sentence for the query. Any node set can be presented as a stand alone, meaningful proof fragment as it contains at least one inference step, and each one of its inference steps has the inference rule used along with links to the inference step antecedents, sources and variable bindings.

The IW infrastructure can automatically generate follow-up questions for any proof fragment by asking how each antecedent sentence was derived. The individual proof fragments may be combined together to generate a complete proof, i.e., a set of inference steps culminating in inference steps containing only asserted (rather than derived) antecedents. When an antecedent sentence is asserted, there are no additional follow-up questions required and that ends the complete proof generation. The specification of IW concepts used in Figure 2 is available at http://www.ksl.stanford.edu/software/IW/pmlspec/2004/02/iw.owl.

5.2 PML Provenance Elements

Every IWBase entry is an instance of an PML provenance element. *InferenceEngine*, *Language* and *Source* are the core provenance elements. Other PML provenance elements are related to one of these core elements.

The *InferenceEngine* is a core concept since every inference step should have a link to at least one entry of *InferenceEngine* that was responsible for instantiating the inference step itself. For instance, Figure 2 shows that the `iw:hasInferenceEngine` property of `iw:InferenceStep` has a pointer to `JTP.owl`, which is the IWBase meta information about Stanford's JTP[14] model-elimination theorem prover. Inference engines currently may have the following properties associated with them: name, URL, author(s), date, version number, organization, etc.

InferenceRule is one of the more important concepts associated with *InferenceEngine*. Inference rules basically tell a user or agent what kind of manipulations a particular inference engine may perform. With respect to an inference engine, registered rules can be either primitive or derived from other registered rules. Figure 3 contains a screen shot from an IW browser interface presenting the entry for the modus ponens (MP) rule. Thus, MP may be a primitive rule for some inference engines[15]. Each of the inference rules may

[14] http://www.ksl.stanford.edu/software/jtp/

[15] MP or any rule may be primitive for one reasoner while it may be derived for another reasoner.

include a name, description, optional example, and optional formal specification. Figure 3 shows that the MP inference rule can be formally specified by the string "`%p, (implies %p %q) |- %q;; (Sent %p %q)`" that is written in Meta-SCL[16]. Meta-SCL is built on top of the evolving SCL[17]. Thus, the meaning of MP rule comes from the Meta-SCL semantics and the MP specification in Meta-SCL. Given a rule: it has "`%p`" and "`(implies %p %q)`" as premises; "`%q`" as conclusion; and "`(Sent %p %q)`" as side-condition. Moreover, premises and conclusion are sentence patterns since they have meta-variables on them (e.g., `%p` and `%q`). Indeed, the `(Sent %p %q)` side condition says that the arguments `%p` and `%q` used in the premises and conclusion can be bound to a sentence.

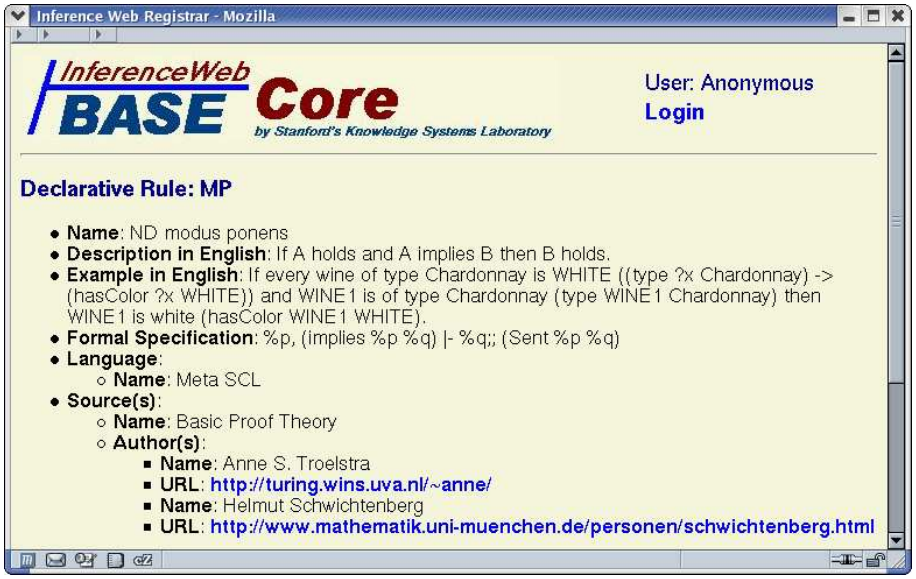

Figure 3: Sample IWBase entry for an inference rule.

Inference Web does not have a specific standard language for formalizing inference rule specifications. Instead, through PML, Inference Web provides a mechanism for registering rule specifications and the languages used to state the rule specifications. For instance, in the MP example above the rule specification is written in Meta-SCL, which is an appropriate language for describing rules used on proofs where conclusions are written in KIF. Indeed, any valid instantiation of the premises and conclusion of the MP rule specification above are valid KIF sentences. Meta-SCL is a convenient choice for rule specification since it was designed for this purpose, is built on the next generation of KIF, and is understood by Inference Web, thereby enabling Inference Web to provide

[16]The Meta-SCL language we are using is joint work in progress between Stanford/KSL and Pat Hayes.

[17]http://cl.tamu.edu/docs/scl/scl-latest.html

proof abstraction services. It is not the only choice however that can be registered – rules may be specified in other languages as well. Some IW tools such as proof checkers, however, may be unable to provide as many services for proofs applying rules formally specified on languages that the tools cannot understand.

Our experience specifying primitive rules in the Inference Web has demonstrated that a significant proportion of them can be formalized completely by a declarative specification language for rules such as Meta-SCL. PML calls these rules that can be formalized completely `DeclarativeRules`. Rules that can not be fully specified formally are called `MethodRules` since rules of this category may need to rely on an additional method for deciding whether an inference step based on a method rule is a valid application of the rule. Method rules are specified in PML in order to accommodate meta-information about rules often called *procedural attachments*. For method rules, in addition to the formal specification, PML allows the registration of a method and a language used to write the method. In fact, the formal specification string of a method rule can still be used for matching premises with conclusions while the rule method can be applied later at proof-checking time to verify the correct use of the rule in a proof.

Many reasoners also use a set of derived rules that may be useful for optimization or other efficiency concerns. One individual reasoner may not be able to provide a proof of any particular derived rule but it may point to another reasoner's proof of a rule. Thus, reasoner-specific rules can be explained in the IWBase before the reasoner is actually used to generate IW proofs. Inference Web thus provides a way to use one reasoner to explain another reasoner's inference rules. (This was the strategy used in Borgida et al. (1999) for example where the performance tableaux reasoner was explained by a set of natural-deduction style inference rules in the explanation system.) This strategy may be useful for explaining heavily optimized inference engines. IWBase, when fully populated, will contain inference rule sets for many common reasoning systems. In this case, users may view inference rule sets to help them decide whether to use a particular inference engine. Today IWBase contains rule sets for JTP, JTP's special purpose reasoners for DAML/OWL and temporal reasoning, SNARK, JSAT, and ISI's Mediator. It also has partial rule sets for some other reasoners.

Inference engines may use specialized language axioms to support a language such as OWL or RDF. *Language* is a core IWBase concept. Axiom sets such as the one specified in Fikes and McGuinness (2001) may be associated with a *Language*. The axiom set may be used as a source and specialized rewrites of those axioms may be used by a particular theorem prover to reason efficiently. Thus proofs may depend upon these language-specific axioms sets called *LanguageAxiomSets* in the IW. It is worth noting that an entry of *Language* may be associated with a number of entries of *LanguageAxiomSet* as different reasoners may find different sets of axioms to be more useful. For example, JTP uses a horn-style set of DAML axioms for its DAML reasoner while another reasoner may use a slightly different set for efficiency, stylistic, interoperability, or presentation reasons. Also, an entry of an *Axiom* can be included in multiple

entries of *LanguageAxiomSet*. The content attribute of *Axiom* entries contains the axiom stated in the language specified by the language attribute of *Axiom*.

Source is the other core IWBase concept and it is a provenance element since it is used to identify the origin of a piece of information. *Source* is specialized into five basic classes: *Person*, *Team*, *Publication*, *Ontology*[18], and *Organization*. At the moment, we are expanding the specification of (authoritative) sources as required. We have begun with a minimal description of these sources in the initial specification used in the IW and are expanding as needed based on empirical usage studies. Entries of *Ontology*, for example, describe stores of assertions that may be used in proofs. It can be important to be able to present information such as ontology source, date, version, URL (for browsing), etc. IW uses ontology in a broad sense McGuinness (2003) and includes both conceptual models as well as individual information and thus both knowledge bases and domain models are registered as ontologies in IWBase. Figure 4 contains a sample ontology registry entry for the ontology used in our wine examples.

Figure 4: Sample IWBase entry for an ontology.

5.3 IWBase

IWBase is an interconnected network of distributed repositories of proof and explanation meta information. Each repository of the network is an *IWBase node* residing in a web server. An IWBase node entry is a URI on the residing web server containing an OWL document of a provenance element. The content of each IWBase node URI is also mirrored in a database system. Therefore, PML proofs and explanations can have direct access to their meta information by resolving their URI references to IWBase entries.

[18] *LanguageAxiomSet* is a subclass of *Ontology*.

IWBase node services, however, may need more sophisticated ways of querying the entries since they may not know exactly which entry to retrieve, for example, when an IWBase user needs to browse the entries in a node. In these cases, the services can take advantage of the underlying database system for querying node entries. For instance, in order to interact with IWBase, each node provides a collection of services collectively called a node *registrar* that supports users in updating or browsing the registry. The registrar may grant update or access privileges on a provenance element basis and the node administrator may define and implement policies for accessing the IWBase node. The generation of proof fragments is a straightforward task once inference engine data structures storing proof elements are identified as IW components. To facilitate the generation of proofs, IWBase provides a set of SOAP-based web services that dump proofs from IW components and uploads IW components from proofs. This service is a language-independent facility used to dump proofs. Also, it is a valuable mechanism for recording the usage of IWBase entries.

In addition to the generic properties of IWBase nodes described above, the IWBase architecture specifies that each node is either a *core* node or a *domain* node. In fact, some provenance elements such as *inference engine*, *inference rule* and *language* are so generic that it may be appropriate to gather them in a single node, the core node, that is also publicly available for the other IWBase nodes. The current demonstration registrar for the core node is available at: http://inferenceweb.stanford.edu/iwregistrar/ and is one example core node. The core node architecture is convenient when there is one set of entries that describe the main meta information about the common engines, their rules, and representation and reasoning languages. Empirically, we have found our current uses of Inference Web benefit from such a core node. However, domain ontologies and their related meta information can vary widely from project to project thus these are appropriate to maintain in project-specific domain nodes. Just as the notion of upper ontologies is both popular and contentious, we anticipate some upper level ontologies to emerge that may be beneficial for inclusion in core nodes. We expect these decisions to evolve with usage.

The IWBase architecture also specifies some services supporting the collaboration between the nodes. Basic services for making local copies of node entries are provided by a concurrent versions system (CVS) repository where the OWL URIs are stored. Thus, using the CVS services, users can check out personal copies of other node entries (whether they are entries from the core or a domain node) and store them locally. For instance, we expect that domain node administrators will keep local copies of the core node for efficiency and/or privacy issues. More sophisticated services are provided for interacting domain nodes. For instance, the administrator of a domain node (e.g, a node specialized on meta information about laptops) can specify that the node has visibility of another domain node. So, if a domain node for laptop computers may benefit from reusing some meta information already stored in a domain node about computers in general, it may use that node. These services between domain nodes also provide a solution for the problem of deciding where to store

meta information about the so-called *upper level ontologies*. In fact, it is up to the users of another domain node to decide whether they want to reuse meta information about other ontologies and thus they may decide what they would like to include.

The current IWBase provides support for provenance information a the level of knowledge bases and ontologies. However, we are on the process of extending the IWBase infrastructure in order to provide support for provenance information whenever it is possible to identify some document or document element to which we can associate provenance information as described in Pinheiro da Silva et al. (2003).

5.4 Explanations

Although essential for automated reasoning, inference rules such as those used by theorem provers and registered in the IWBase as *InferenceRule* entries are often inappropriate for "explaining" reasoning tasks. Moreover, syntactic manipulations of proofs based on atomic inference rules may also be insufficient for abstracting machine-generated proofs into some more understandable proofs Huang (1994). Proofs, however, can be abstracted when they are rewritten using rules derived from axioms and other rules. Axioms in rewriting rules are the elements responsible for recognizing patterns and providing rewritten abstracted versions of the rules. Entries of *DerivedRule* are the natural candidates for storing specialized sets of rewriting rules. In IW, tactics are rewrite rules associated with axioms, and are used independent of whether a rule is atomic or derived.

Many intermediate results are "dropped" along with their supporting axioms, thereby abstracting the structure of proofs. The user may always ask follow-up questions and still obtain the detail, however the default explanation provides abstracted explanations. The general result is to hide the core reasoner rules and expose abstractions of the higher-level derived rules. An example of an IW explanation is described in the Inference Web web page at: http://www.ksl.stanford.edu/software/iw/Ex1/. The implementation of the IW explainer is work in progress. The explainer algorithm generates explanations in a systematic way using the derived rules related to a selected language axiom set. We have used the current rewrite rule set to abstract presentations of answers obtained from JTP in analysis applications.

5.5 Browser

Inference Web includes a browser that can display both proofs and their explanations in a number of proof styles and sentence formats. Initially, we include the "English", "Proof" and "Dag" styles and the restricted "English", "KIF" and "Raw" formats[19]. We also expect that some applications may implement their own displays using the IW API and one of our projects uses this model.

[19]Current investigations are underway for an N3 format as well.

The IW browser implements a lens metaphor responsible for rendering a fixed number of levels of inference steps depending on the lens magnitude setting. The prototype browser allows a user to see up to five levels of inference steps simultaneously along with their derived sentences and antecedent sentences.

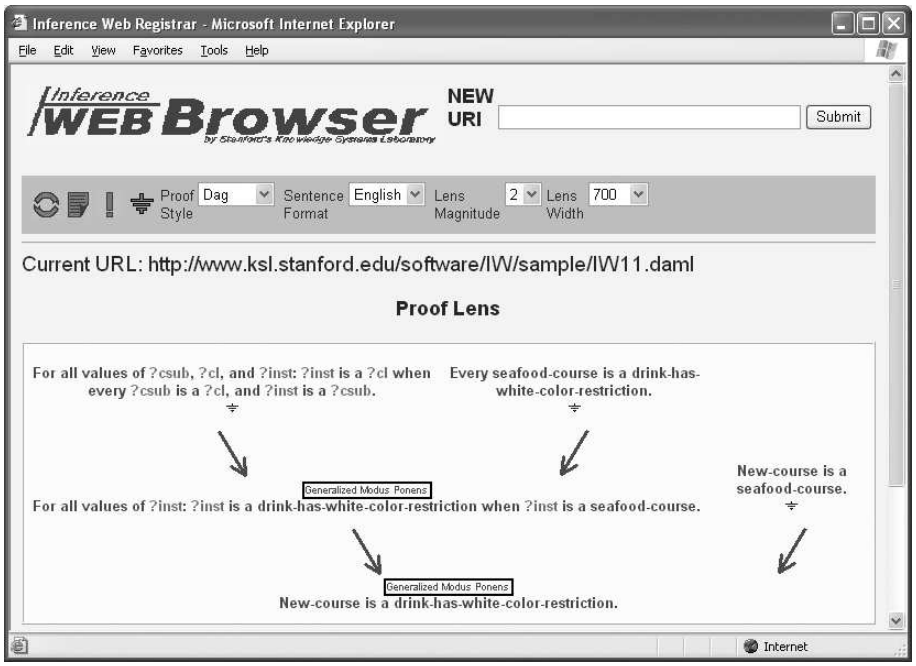

Figure 5: An Inference Web Browser screen.

Figure 5 presents two levels of inference step for one wine use case in Section 4. Prior to this view, the program has asked what wine to serve with a seafood course. In Figure 5, one can see that `New-course`, which is the selected meal course, requires a drink that has a white color since it is a seafood course. The sentences are formatted in English and the lens magnitude is two, thus the browser displays the inference steps used to derive it including its antecedents and the antecedent's derivations. Concerning sentence formats, the browser supports some restricted translations between sentences that can be requested by the user. For example, the "raw" format indicates that the user wants to see node set conclusions as originally stated. However, if the user selects "KIF", then if node set conclusions are not already in KIF, and the browser has a translator from the original language into KIF, then the browser translates and presents the sentences in KIF. Otherwise, it presents the sentence in its original language. The same method is used for translating other formats into English, for example. We currently have a KIF to limited English translator for such needs.

We believe that one of the keys to presentation of justifications is breaking proofs into separable pieces. Since we present fragments, automatic follow-up question support is a critical function of the IW browser. Every element in the viewing lens can trigger a browser action. The selection of an antecedent re-focuses the lens on an antecedent's inference step. For other lens elements, associated actions present IWBase meta information. The selection of the consequent presents details about the inference engine used to derive the actual theorem. The selection of an inference rule presents a description of the rule. The selection of the source icon beside sentences associated with source documents presents details about sources where the axiom is defined. In Figure 5, selecting the consequent would present information about JTP – the inference engine used to derive it. Selecting GMP – the inference rule, would present information about JTP's Generalized Modus Ponens rule.

6 Contributions and Future Work

The Wine Agent[20] and the DAML Query Language Front-End[21] are two example Semantic Web agents supported by the Inference Web. These agents are based on the Stanford's JTP theorem prover that produces PML proofs. The IWBase is populated with JTP information: one *InferenceEngine* entry for the reasoner itself, nine entries for its primitive inference rules, one entry for its set of DAML axioms, and 56 entries for the axioms. Using this registration of JTP and the fact that JTP dumps PML proofs, Inference Web can be used to present proofs and explanations of any of JTP's answers.

Beyond just explaining a single system, Inference Web attempts to incorporate best in class explanations and provide a way of combining and presenting proofs that are available. It does not take one stance on the form of the explanation since it allows deductive engines to dump single or multiple explanations of any deduction in the deductive language of their choice. It provides the user with flexibility in viewing fragments of single or multiple explanations in multiple formats. IW simply requires inference rule registration and PML format. It does not limit itself to only explaining deductive engines. It provides a proof theoretic foundation on which to build and present its explanations, but any question answering system may be registered in the Inference Web and thus explained. More recently, we have begun integrating with query planners and extractors and focus more on explaining the process by which an answer was determined rather than the exact inference rules used to obtain a particular answer. This lets the Inference Web provide explanations for tasks that need to know what was done and also provide explanations for how something was done.

Revisiting the Inference Web requirements in Section 3, we can identify the following contributions:

[20] http://onto.stanford.edu:8080/wino/
[21] http://onto.stanford.edu:8080/dql/servlet/DQLFrontEnd

- *Support for knowledge provenance* is provided by: the PML specification that allows node sets to be associated with sources; and the IWBase that supports meta information for annotating sources and provides database storage and access to the meta information.

- *Support for reasoning information* is provided by: the proof specification that supports a comprehensive representation of proof trees; and the IWBase that supports meta information for annotating inference engines along with their primitive inference rules. Also, the proof specification provides support for alternative justifications by allowing multiple inference steps per node set and the proof browser supports navigation of the information.

- *Support for explanation generation* is provided by the IWBase that supports both formal and informal information about languages, axioms, axiom sets, derived and rewrite rules. Rewrite rules provide the key to abstracting complicated proofs into more understandable explanations. The proof support for alternative justifications allows derivations to be performed by performance reasoners with explanations being generated by alternative reasoners aimed at human consumption.

- *Support for distributed proofs* is provided by the IW architecture. Proofs are specified in PML (using the emerging web standard OWL so as to leverage XML-, RDF-, and OWL-based information services) and are interoperable. Proof fragments as well as entire proofs may be combined and interchanged.

- *Support for proof presentation* is provided by a lightweight proof browsing using the lens-based IW browser. The browser can present either pruned justifications or guided viewing of a complete reasoning path.

We have registered a few theorem provers and updated them so that they produce PML proofs. Inference Web can then be used to browse proofs and explanations of any answer produced by those reasoners. Inference web was originally aimed at explaining answers from theorem provers that encode a set of declaratively specified inference rules. More recently, we have looked at other kinds of reasoning engines such as JSAT and can now browse proofs generated from this satisfiability reasoner Shvaiko et al. (2004). In joint work with Ambite, Knoblock and Muslea, we have also recently registered the ISI Mediator so that query plans can be presented by the Inference Web. These two efforts along with initial discussions with IBM (Ferrucci, Murdock, and Welty) about registering text analytics engines, we have broadened our notion of what kinds of inference rules and reasoners can be registered and thus broadened the kinds of question answering systems that can be explained.

Future work includes the registration of more question answering systems – whether they are theorem provers, planners, extractors, or of other types. We have encoded some rewrite rules that enable explanations to be generated from

the more detailed proofs. Some simple examples of explanations can be seen on the Inference Web web site (e.g., http://www. ksl.stanford.edu/software/IW/Ex1). Currently, we are developing tools for generating tactics that are required for explaining other proofs. We also intend to provide specialized support for why-not questions expanding upon Chalupsky and Russ (2002) and McGuinness (1996). We have also begun an effort to provide specialized support for explaining contradictory information. We are also looking at additional support for proof browsing and pruning. We have also initiated conversations with the verification community in order to provide a PML format that meets their needs as well as meeting the needs of the applications that require explanation. Initial discussions at least for utilizing IWBase inference rule information with "correct-by-construction" software environments such as Specware[22] appear promising.

7 Conclusion

Inference Web enables applications to generate portable explanations of their conclusions. We identified the support for knowledge provenance, reasoning information, explanation generation, distributed proofs, and proof presentation as requirements for explanations in the web. We described the major components of IW – the PML specification based on the emerging web language – OWL supporting proofs and their explanations, the IWBase, and the IW proof browser. We described how Inference Web features provide infrastructure for the identified requirements for web explanations. We facilitated use in a distributed environment by providing IW tools for registering and manipulating proofs, proof fragments, inference engines, ontologies, and source information. We also facilitated interoperability by specifying the PML format and providing tools for manipulating proofs and fragments. We have implemented the IW approach for four Semantic Web agents (three of them based on JTP and one based on JSAT) and are in discussions with additional reasoner authors to include more reasoning engines. We have presented the work at government sponsored program meetings (RKF, DAML, PAL, AQUAINT, and NIMD) to gather input from other reasoner authors/users and have obtained feedback and interest. Current registration work includes IBM's UIMA, ISI's Mediator, SRI's SNARK, and W3C's CWM.

Acknowledgments Many people have provided valuable input to our work. Thanks in particular go to current and past colleagues at KSL including Richard Fikes, Jessica Jenkins, Cynthia Chang, Gleb Frank, Eric Hsu, Bill MacCartney, Rob McCool, Sheila McIlraith, and Yulin Li for input on JTP, our specification or applications. Thanks also go to Pat Hayes for his reviews of our work and joint work on Meta-SCL. Thanks to many people from our government sponsored research programs for providing requirements and comments. Thanks

[22] http://www.kestrel.edu/HTML/prototypes/specware.html

to current collaborators on recent integrations, in particular Jose-Luis Ambite, Fausto Giunchiglia, Craig Knoblock, Pavel Shvaiko, and Mark Stickel. All errors, of course are our responsibility.

This work is supported by the following grants DARPA F30602-00-2-0579, N66001-00-C-8027, NBCHD030010, and ARDA H278000*000 and H768000* 000/4400049114.

References

Borgida, A., Franconi, E., Horrocks, I., 2000. Explaining *ACL* Subsumption. In: Proc. of the 14th European Conf. on Artificial Intelligence (ECAI2000). IOS Press, pp. 209–213.

Borgida, A., Franconi, E., Horrocks, I., McGuinness, D. L., Patel-Schneider, P. F., July 1999. Explaining *ALC* Subsumption. In: Proc. of the International Workshop on Description Logics (DL'99). Linköping, Sweden, pp. 33–36.

Borgida, A., McGuinness, D. L., November 1996. Asking Queries about Frames. In: Proceedings of Fifth International Conference on the Principles of Knowledge Representation and Reasoning. Morgan Kaufmann, Cambridge, Massachusetts.

Boyer, R., Kaufmann, M., Moore, J., 1995. The Boyer-Moore Theorem Prover and Its Interactive Enhancements. Computers and Mathematics with Applications 29 (2), 27–62.

Buneman, P., Khanna, S., Tan, W.-C., January 2001. Why and Where: A Characterization of Data Provenance. In: Proceedings of 8th International Conference on Database Theory. pp. 316–330.

Chalupsky, H., Russ, T., 2002. WhyNot: Debugging Failed Queries in Large Knowledge Bases. In: Proc. of the 14th Innovative Applications of Artificial Intelligence Conference (IAAI-02). pp. 870–877.

Das, A., Wu, W., McGuinness, D. L., 2002. Industrial Strength Ontology Management. In: Cruz, I., Decker, S., Euzenat, J., McGuinness, D. L. (Eds.), The Emerging Semantic Web. IOS Press.
URL `http://www.iospress.nl/site/html/boek-1381825766.html`

Felty, A., Miller, D., 1987. Proof Explanation and Revision. Tech. Rep. MSCIS8817, University of Pennsylvania.
URL `http://cm.bell-labs.com/who/felty/abstracts/proof87.html`

Ferrucci, D., Lally, A., 2004. UIMA: An Architectural Approach Unstructured Information Processing in the Corporate Research Environment. . Journal of Natural Language EngineeringTo appear.

Fikes, R., Hayes, P., Horrocks, I., 2002. DAML Query Language (DQL) Abstract Specification. Tech. rep., W3C.
URL `http://www.daml.org/2002/08/dql`

Fikes, R., McGuinness, D. L., December 2001. An Axiomatic Semantics for RDF, RDF-S, and DAML+OIL (March 2001). Tech. Rep. Note 18, W3C. URL http://www.w3.org/TR/daml+oil-axioms

Giunchiglia, F., Shvaiko, P., 2003. Semantic matching. Tech. Rep. Vol. 71, CEUR - WS. URL http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-71/Giunchiglia.pdf

Guha, R. V., McCool, R., Miller, E., May 2003. Semantic search. In: Proceedings of the Twelfth International World Wide Web Conference (WWW2003). ACM Press, Budapest, Hungary, pp. 700–709.

Huang, X., 1994. Reconstructing Proofs at the Assertion Level. In: Proceedings of CADE-94. LNAI-814. Springer, pp. 738–752.

McGuinness, D. L., 1996. Explaining Reasoning in Description Logics. Ph.D. thesis, Rutgers University.

McGuinness, D. L., 2003. Ontologies Come of Age. In: Fensel, D., Hendler, J. A., Lieberman, H., Wahlster, W. (Eds.), Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential. MIT Press, pp. 171–194.

McGuinness, D. L., Borgida, A., August 1995. Explaining Subsumption in Description Logics. In: Proc. of the 14th International Joint Conference on Artificial Intelligence. Morgan Kaufmann, Montreal, Canada, pp. 816–821.

McGuinness, D. L., Patel-Schneider, P., 2003. From Description Logic Provers to Knowledge Representation Systems. In: Baader, F., Calvanese, D., McGuinness, D., Nardi, D., Patel-Schneider, P. (Eds.), The Description Logic Handbook: Theory, Implementation, and Applications. Cambridge University Press, pp. 265–281.

McGuinness, D. L., Pinheiro da Silva, P., October 2003a. Infrastructure for Web Explanations. In: Fensel, D., Sycara, K., Mylopoulos, J. (Eds.), Proceedings of 2nd International Semantic Web Conference (ISWC2003). LNCS-2870. Springer, Sanibel, FL, USA, pp. 113–129.

McGuinness, D. L., Pinheiro da Silva, P., August 2003b. Registry-Based Support for Information Integration. In: Proceedings of IJCAI-2003 Workshop on Information Integration on the Web (IIWeb-03). Acapulco, Mexico, pp. 117–122.

McGuinness, D. L., Pinheiro da Silva, P., 2004. Trusting Answers on the Web. In: Maybury, M. T. (Ed.), New Directions in Question Answering. AAAI/MIT Press, to appear.

Pinheiro da Silva, P., McGuinness, D. L., Fikes, R., January 2004. A Proof Markup Language for Semantic Web Services. Tech. Rep. KSL-04-01, Knowledge Systems Laboratory, Stanford University, Stanford, CA, USA.

Pinheiro da Silva, P., McGuinness, D. L., McCool, R., December 2003. Knowledge Provenance Infrastructure. IEEE Data Engineering Bulletin 25 (2), 179–227.

Shortliffe, E. H., 1976. Computer-Based Medical Consultations: MYCIN. Elsevier/North Holland, New York, NY, USA.

Shvaiko, P., Giunchiglia, F., Pinheiro da Silva, P., McGuinness, D. L., January 2004. Web Explanations for Semantic Heterogeneity Discovery. Tech. Rep. KSL-04-02, Knowledge Systems Laboratory, Stanford University, Stanford, CA.

Smith, M., McGuinness, D. L., Volz, R., Welty, C., 2003. Web Ontology Language (OWL) Guide Version 1.0. Tech. Rep. Working Draft, World Wide Web Committee (W3C).
URL http://www.w3.org/TR/owl-guide

Swartout, W., Paris, C., Moore, J., June 1991. Explanations in Knowledge Systems: Design for Explainable Expert Systems. IEEE Intelligent Systems.
URL http://www.computer.org/intelligent/ex199/x3058abs.htm