Common Sense Reasoning with the Semantic Web

Christopher C. Johnson and Push Singh
MIT Summer Research Program
Massachusetts Institute of Technology, Cambridge, MA 02139
CJohnson26@mail.utexas.edu, push@media.MIT.edu

Abstract
Current HTML content on the World Wide Web has no real meaning to the computers that display the content. Rather, the content is just fodder for the human eye. This is unfortunate as in fact Web documents describe real objects and concepts, and give particular relationships between them. The goal of the World Wide Web Consortium's (W3C) Semantic Web initiative is to formalize web content into Resource Description Framework (RDF) ontologies so that computers may reason and make decisions about content across the Web. Current W3C work has so far been concerned with creating languages in which to express formal Web ontologies and tools, but has overlooked the value and importance of implementing common sense reasoning within the Semantic Web. As Web blogging and news postings become more prominent across the Web, there will be a vast source of natural language text not represented as RDF metadata. Common sense reasoning will be needed to take full advantage of this content. In this paper we will first describe our work in converting the common sense knowledge base, ConceptNet, to RDF format and running N3 rules through the forward chaining reasoner, CWM, to further produce new concepts in ConceptNet. We will then describe an example in using ConceptNet to recommend gift ideas by analyzing the contents of a weblog. Finally, we will speculate about our ongoing common sense research and how we will further extend this project in the future.
1 Introduction

"The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation." [1]

The main motivation of the Semantic Web is to provide automation. Current web content exists in the form of HTML and XML layouts. Though this content provides an efficient visual picture for humans, computers cannot reason about the content. The Semantic Web attempts to formalize and link web content such that computers can make decisions about web data without the assistance of humans. The W3C is attempting to achieve this goal by relating web concepts according to a specific information representation called Resource Description Framework (RDF). RDF relates concepts according to triples consisting of a subject, verb, and object. When many triples consisting of similar subjects and objects are combined they form an ontology. Initially, the syntax for RDF was an application of XML called RDF/XML. However, this syntax was difficult to read and discuss in simple terms, and thus an accompanying human-readable syntax called Notation3 (N3)[5] was created. From now on, when we refer to RDF triples we will use N3 notation.

2 A Motivation for Common Sense Knowledge

As RDF becomes more widely accepted by the web development community, developers will be motivated to formally define their information and services using RDF metadata. However, given that users rarely fill out the metadata in their Word documents or edit their mp3’s ID3 tags, it is a very optimistic approach to believe that users will correctly build the logical triples needed for the Semantic Web. [2] Also, weblogs, news postings, and other pre-existing natural language websites will create a vast resource of natural language text. To fully take advantage of this data we will need tools to access this natural language text and make sense of it. A second motivation of bringing common sense reasoning to the semantic web is that to intelligently complete tasks for a user, a software agent must have a vast knowledge about the world we live in. A plumbing service website may express the concept of ”plumber” using the correct URI in an RDF triple, but the computer will not be able to associate a clogged toilet with ”plumber” without access to a common sense knowledge base. The idea here is not to replace existing RDF ontologies with a massive common
sense knowledge base, but rather to use common sense data along with existing RDF ontologies to make more intelligent decisions.

3 Using ConceptNet to bring Common Sense Reasoning to the Semantic Web

The Open Mind common sense web site [6], developed by Push Singh, contains a vast array of natural language statements entered by the general public over the past 3 years. Using this array of natural language statements, Push was able to compose the common sense knowledge base, ConceptNet [7], which currently contains approximately 1.6 million assertions. Statements in ConceptNet are represented as triples, though its semantics are represented in a much more informal natural language format than RDF. Fig. 1 gives an example of how data is represented in ConceptNet.

Figure 1: An excerpt from ConceptNet’s semantic network of commonsense knowledge.

The first step in using this data was to convert the database to RDF format. We used only a small subset of ConceptNet (approximately 18 thousand assertions rather than 1.6 million) in order to speed up run time. We also chose to use the N3 syntax for its readability. Due to ConceptNet’s data already being represented in a triples format, converting ConceptNet to N3 was a rather simple task requiring a
short Python script.

4 Using CWM to expand ConceptNet’s Knowledge Base

CWM, which stands for Closed World Machine, is a forward chaining inference engine designed by the W3C’s Semantic Web Advanced Development (SWAD) project to be used in conjunction with RDF/XML, N3, and N3 rules. Now that we had an RDF version of ConceptNet, our first experiment was to attempt to expand ConceptNet’s knowledge base by running various transitivity rules on ConceptNet through CWM. We first ran transitivity on the IsA relation and saw great success, as the results yielded a high signal-to-noise ratio [8]. The IsA transitivity rule as well as triples represented in ConceptNet before and after application of the rule are shown below

@prefix : <http://dig.csail.mit.edu/2005/08/conceptNet#> .
#Before application of rule

#After application of rule
:germanshepherd :IsA :canine,
    :commonpet,
    :dog,
    :domesticanimal,
    :fourleggedanimal,
    :fourleggedmammal,
    :mansbestfriend,
    :packanimal,
    :territorialanimal .

Other transitivity rules were not as simple to run, and some did not prove to have such a high signal-to-noise ratio [8]. For example, in order to run the transitivity of LocationOf, we needed a way to unify things like inhouse with house. We need to remember that when talking about these concepts in RDF terms that "inhouse" is actually a URI. Thus, we cannot simply look at the string "inhouse" and unify
it with "house". The hack around this was to add a new predicate to ConceptNet, PropositionOf, which would apply to any subjects that had a proposition in their title.

@prefix : <http://dig.csail.mit.edu/2005/08/conceptNet#> .

Now that we had this new verb we could easily unify propositions with the objects that they referred to. The transitivity rule for LocationOf as well as an example of data that fits the rule are shown below.

@prefix : <http://dig.csail.mit.edu/2005/08/conceptNet#> .

:anchor :LocationOf :onboat.
:boat :LocationOf :onsea.

#therefore

:anchor :LocationOf :onsea.

The result of the LocationOf transitivity yielded a satisfactory signal-to-noise ratio [8], however the size of the result was almost as large as ConceptNet itself! If we stop for a moment to think about this, it does make sense, as locations are very transitive in nature. Consider a penny inside your pocket as you sit on your bed. That penny is also on you, on your bed, in your room, in your house, in your neighborhood, in your town, in your state, etc. Other rules run that did not yield satisfactory signal-to-noise ratios are listed below.

@prefix : <http://dig.csail.mit.edu/2005/08/conceptNet#> .
Transitivity of SubeventOf for one was very noisy. It seems as though each iteration through the rule yields a noisier result. For example, consider the following triples contained in ConceptNet.

```

It is somewhat reliable to think that since flying in an airplane has a subevent of sleep, and sleep has a subevent of snore, that flying in an airplane has a subevent of snore. However, it is not very reliable to say that flying in an airplane has a subevent of annoying one’s spouse. Thus, each iteration through the transitivity of SubeventOf produces less reliable data.

5 An Example of Using ConceptNet to Make Common Sense Decisions about Web Content

Suppose your brother Matt’s birthday is coming up and he has just moved into a new apartment. Using common sense knowledge you may determine that he needs furniture since he has just moved into a new place. Thus, you go out and purchase your brother a new love seat for his new apartment. This decision required several steps of common sense reasoning. First you realized that Matt has just moved into a new apartment, then you determined that moving into a new apartment brings with it the necessity of furniture, and finally you determined that having such a necessity would yield a favorable gift.

Suppose that you were to prompt a computer for gift recommendations for your brother. How can the computer make the same common sense decisions as above? First, the computer will need specific information about Matt. This can be achieved by running a natural language parser (NLP) through his weblog. Fig. 2 displays an example of what Matt’s weblog may look like.

Using a natural language parser (NLP) we would be able to parse out relevant data regarding Matt in RDF format. Such triples would include:

:Matt :PropertyOf :MovedIntoApartment .
```
Now that the computer has relevant data about Matt it needs a set of rules that it will use to determine gift recommendations. We decided to implement 3 main rules for requirement: requirement due to PropertyOf, requirement due to IsA, and requirement due to LocationOf. The N3 rules that we used to represent these requirement rules are listed below.

@prefix : <http://dig.csail.mit.edu/2005/08/conceptNet#> .
#require due to property

#require due to IsA

{ :Matt :IsA ?type . ?type :PropertyOf ?prop . ?prop :CapableOf

#require due to Location
Christopher C. Johnson


Important to note is the necessityOf verb. This verb was added to ConceptNet’s knowledge base so that one could unify things like ”NeedFurniture” with ”Furniture”.

Also of importance to note is that we wrote our rules in terms of Matt only to speed up run time and to filter results. Generally these rules should be expected to apply to any person.

Running these rules through CWM along with ConceptNet and our newly obtained Matt data should have produced desired outputs, as indeed they did. ConceptNet’s knowledge base previously contained the following relevant triples:

@prefix : <http://dig.csail.mit.edu/2005/08/conceptNet#> .
:MovedIntoApartment :CapableOf :requireFurniture .

:golfer :CapableOf :needgolfclubs .
:needgolfclubs :necessityOf :golfclubs .


From this, it should be simple to see that CWM reported the following gift recommendations:

@prefix : <http://dig.csail.mit.edu/2005/08/conceptNet#> .
:Matt :GiftIdea :furniture .
:Matt :GiftIdea :golfclubs .
:Matt :GiftIdea :sweater .

Of course, data will not always be this clean. For experimental purposes we tried running the same rules with the following added triples:
Running the rules once again with these added triples yielded the following additional results:

@prefix : <http://dig.csail.mit.edu/2005/08/conceptNet#> .
:Matt :GiftIdea :golfclubs .
:Matt :GiftIdea :food .
:Matt :GiftIdea :furniture .
:Matt :GiftIdea :glass .
:Matt :GiftIdea :hug .
:Matt :GiftIdea :sweater .
:Matt :GiftIdea :tosleep .

Obviously, a hug or sleep are not good gift ideas. However, an easy solution to this problem in the future would be to run these objects through Amazon.com’s RDF library at which point we would determine that things like golfclubs and furniture are purchasable while things like a hug or sleep or not. From this we could then modify our rules to limit gift ideas to purchasable objects, at which point hug and tosleep would be dropped from our final results list.

6 Looking Towards the Future

As more and more web sites begin to post RDF metadata, the number of applications of using ConceptNet along with existing RDF ontologies will surely increase. At this point we will be faced with 2 major problems. The first deals with linking ontologies, and the second is dealing with much of the noise that is common in working with a common sense knowledge base. The problem of linking ontologies has already been addressed by many semantic web researchers. Sergey Melnik and Stefan Decker for one have developed an RDF version of WordNet [10] to be used to link various ontologies. WordNet is a lexical database for English, where ”nouns, verbs,
and adverbs are organized into synonym sets, each representing one underlying lexical concept. Different relations link the synonym sets.” [9] Suppose ConceptNet has a triple that refers to ”pop”, but the local grocery store web site has a triple that refers to ”soda”. Using WordNet, one could link these 2 triples. Dan Brickley of the W3C has also done similar work in representing WordNet in RDF [11]. Our second problem will be how to deal with the noise produced by using a common sense knowledge base. Our solution to this is to implement a weighted version of ConceptNet in RDF. ConceptNet triples come pre-equipped with both a frequency and an iterative weight. The frequency weight refers to the number of people who submitted the concept to Open Mind while the iterative weight refers to the number of iterations through rules that were required to acquire the concept. Thus, a high frequency weight and a low iterative weight make for a more reliable triple (see [3] for a better explanation of ConceptNet weights). We believe that we can implement a weighted RDF system in which rules will have their own user programmed weight algorithms that will determine the final weight of a resulting triple. For example, in the gift recommendation example, a requirement due to PropertyOf may be more reliable than a requirement due to LocationOf and thus the algorithm to determine the PropertyOf result may return a higher frequency weight than the LocationOf result. One could then order the resulting gift ideas by weight, with the higher weighted results being more reliable gift recommendations. The implementation of such a weighted system is part of our ongoing research.

7 Final Thoughts

In order to take advantage of the vast sources of natural language text that will continue to pervade the web into the future, computers will require vast amounts of general knowledge about the world we live in. Our proposal is to take advantage of the ConceptNet knowledge base which is continually growing from user input every day. Once we have access to such a common sense knowledge base we will be able to exploit existing RDF ontologies to the highest possible level.

Acknowledgments

Christopher would like to thank Push Singh (MIT Media Lab) for conversations and guidance throughout this project, Christopher Jones (Assistant Dean for graduate
students, and MSRP chair) for continuous support, and Ralph Swick (MIT/CSAIL, DIG), Tim Berners-Lee (MIT/CSAIL, DIG), and the other members of the W3C who helped make this project possible. This work is supported in part by funds provided by the MIT Summer Research Program (MSRP).

References


