Research Paper for the CIS203 Introduction to Artificial Intelligence

Professor Pei Wang

Semantic Web Agent

Sung Fat Freddie Kwok
Temple University

2004 FALL

Table Of Contents

1. Introduction:

The emergence of the Internet has revolutionized almost every conceivable aspect of our lives. We can simply conduct various activities such as shopping, chatting and studying online on the Net. However, the potential of the Internet is not fully realized. It is mainly because the Net only harbors data that can be understood by human beings. The incompetency of the current Internet structure undermines the growth of the Net as an mechanism for humans to share a universal space where everyone is allowed to express and retrieve ideas freely. Semantic Web, which is the next Internet, would not only presents human readable data, but also provides a way for a machine to manipulate data meaningfully. Ideally, Semantic web would act as the extension of the current Internet if we can find a way to establish a mechanism that is compatible with the current Net architecture. Sets of inference rules and the collection of information is required to induce automated reasoning. If a machine is able to understand part of the content of the data that it presents, then "Semantic Web agent" would become a feasible goal. The implication behind the creation of Semantic web agent is extraordinary. In the future, Semantic web agent would be able to set up clinical appointments, enhance the reliability of the current search engine, and exchange data with different agents on the Net. Semantic web agents, like the steam-engine, could freed us from daily drudgery. Therefore, human beings would have more time to attack the

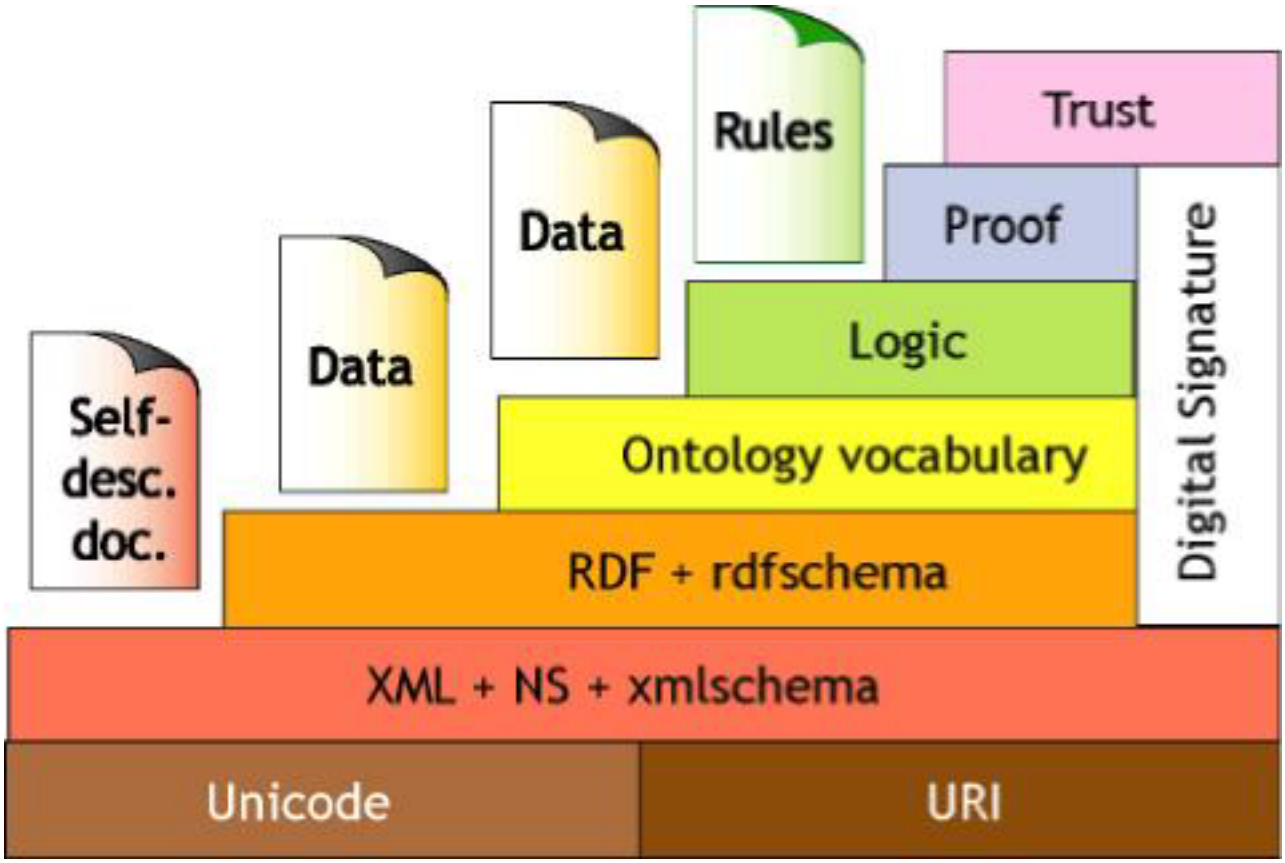
unanswered problems. This project is prepared for the undergraduate class; CIS203 Introduction to Artificial Intelligence. The paper will not present or discuss highly technical questions or lengthy codes. This paper is meant to provide a theoretical approach to issues of the AI application of Semantic Web. The main concern of the project is to explore the relation between semantic web and Artificial Intelligence ,Some samples of codes and graphics will be given to better illustrate the concepts in this paper. Also I will propose to create a A.I. middleware agent, which would help mapping ontology for the Semantic web. I sincerely hope that the paper will provide some insights into A.I. and Semantic Web.

1.1 World Wide Web and Semantic Web

The current World Wide Web provides a medium for anyone to communicate in a special way. In a sense, the Net itself is a abstract universal space everyone with a terminal and client can access to. It is not longer a dream for a computer user to conduct various activities on line, such as distant learning, E-commerce, and grid computing. But HTML per se, is not able to build up a foundation for the next Internet - the Semantic Web. The Semantic web, unlike the current web, will allow a webmaster to define definition tags. A machine-under stable semantic web will open up a thousand possibilities for computer science. And therefore it may allow Semantic web agent to understand the content

of the web. Semantic web environment would allow search engines to search more precisely and even assist human beings on dueling with daily tasks such as making appointments or paying by check. Information retrieval is not only the basic function of the Net, but it is also a crucial function of the web. Without a effective way of retrieving information, the potential of the Semantic web will be undermined.

(see the illustration of the structure of Semantic Web structure).

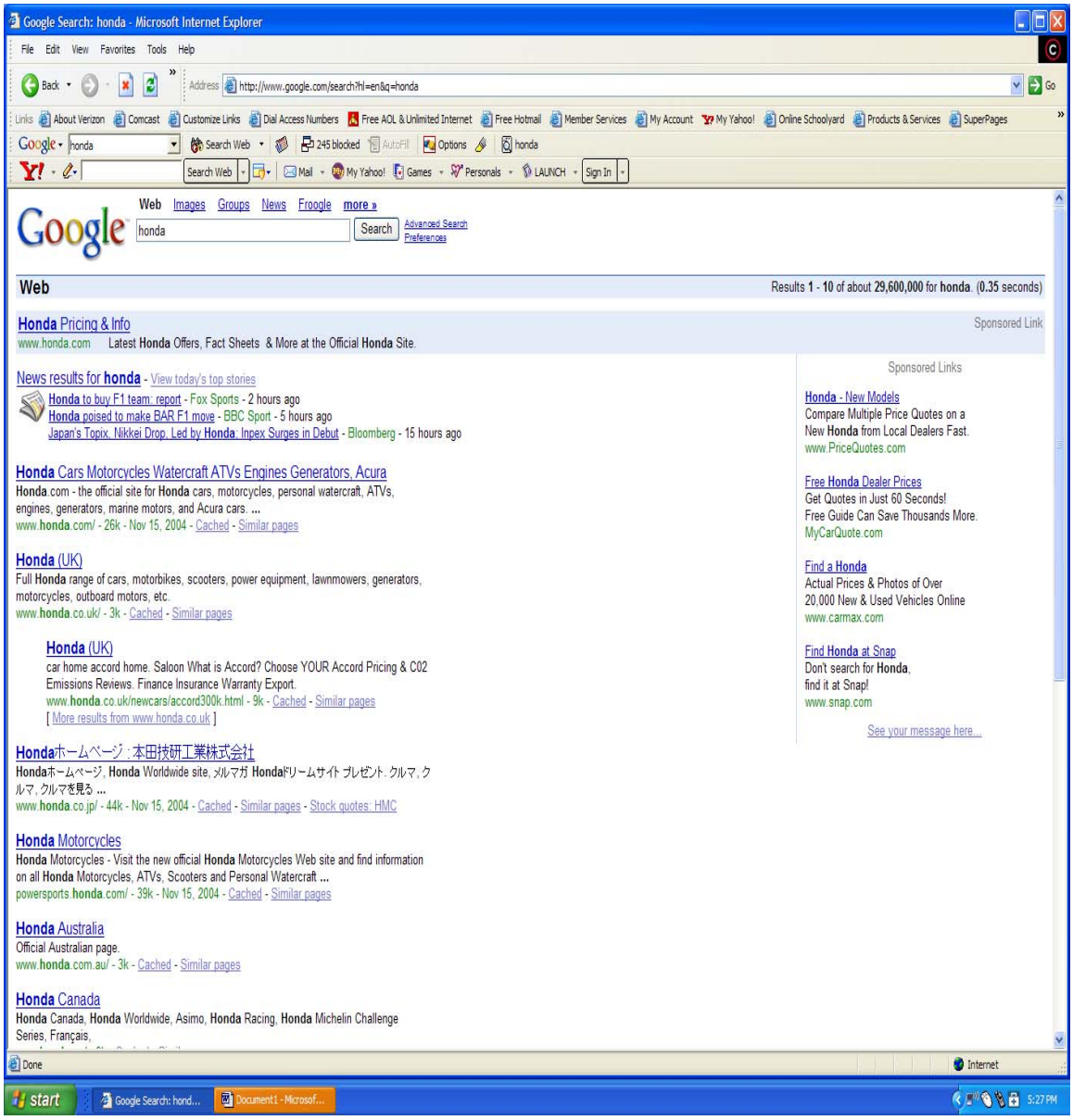


Graph taken from Tim Berners-Lee web.

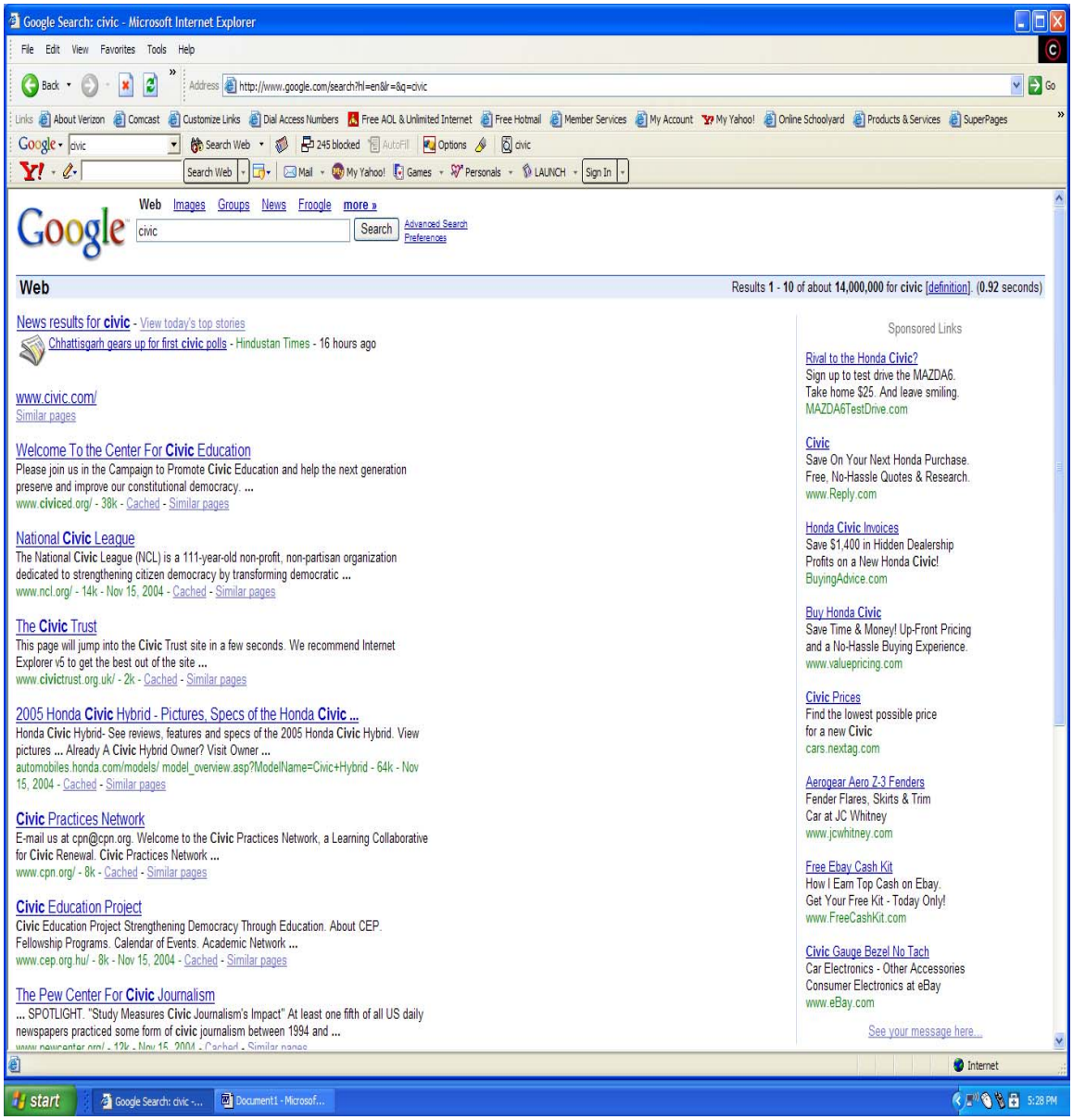
1.3 Information retrieval on the current Internet

Google's success is due to its unique combination of link evaluation, text-matching, and page ranking. Its algorithm evaluates web pages by checking how many links are linked to it. The more links that are pointed to a specific web page implies that the content of the web page is what people seek. Then, the web robot ranks that page based on its popularity. Unfortunately, due to the infrastructure of the web page, Web content is not understood by any program. For instance, if I want to search keyword "CIVIC" on the Internet, I would get clusters of a search result. With the help of the Boolean and clear search target expression, ("CIVIC & HONDA"), I would have a better chance to get to the desired web site. A user can also search within the web page by inputting www.honda.com"civic". This experiment illustrates a flaw within the current world wide web. Ontology-Based knowledge discovery on the World Wide Web. Search engine approach is to use lexical and syntactic content as the textmatching. Yahoo uses another approach, The company hires on-line expect to categorize each web on the Internet. It is obviously not the best answer for the exponential growth of the web pages. Most of the conventional search engines use Ad-hoc robots. Gathering information by checking the existing HTML tags. This method allow Webmaster to embedded a specific keyboard in their homepages. Web-wandering robot with ad-hoc machinery for specialized search tasks. One of the foreseen application of Semantic web agent is to brew a more effective and user-tailed search engine.
(see graph for search result by using Google)

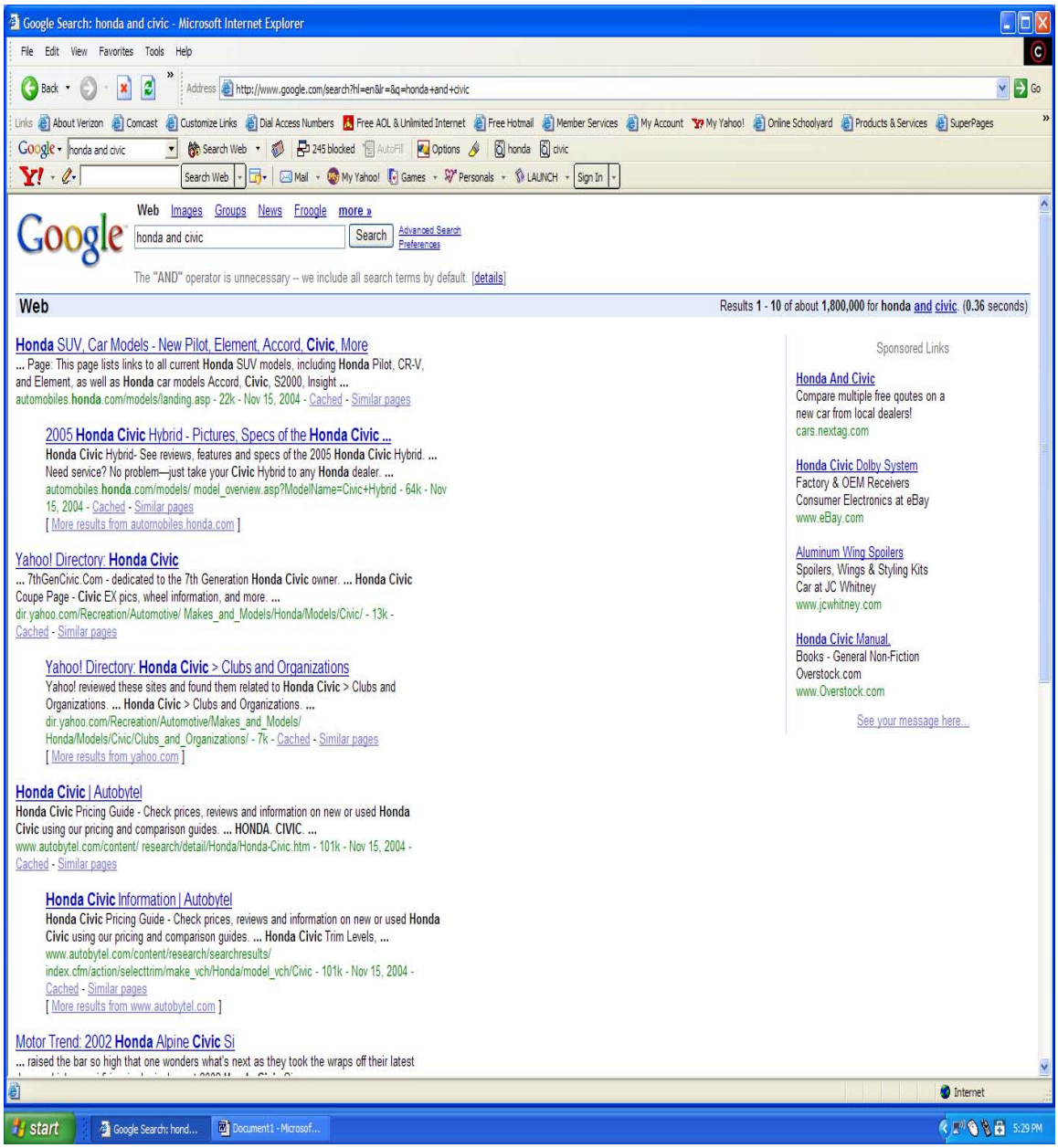
Graph 1



Search result of the keyword “Honda” by using Google Search Engine.

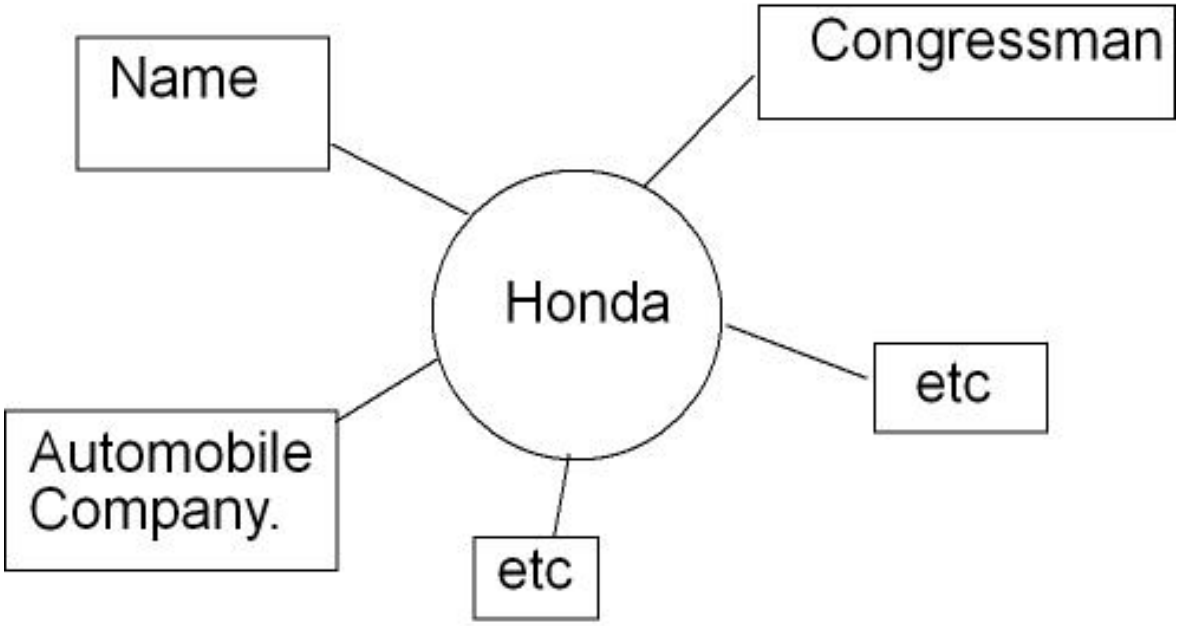


Search result of the keyword “civic” by using Google Search Engine.

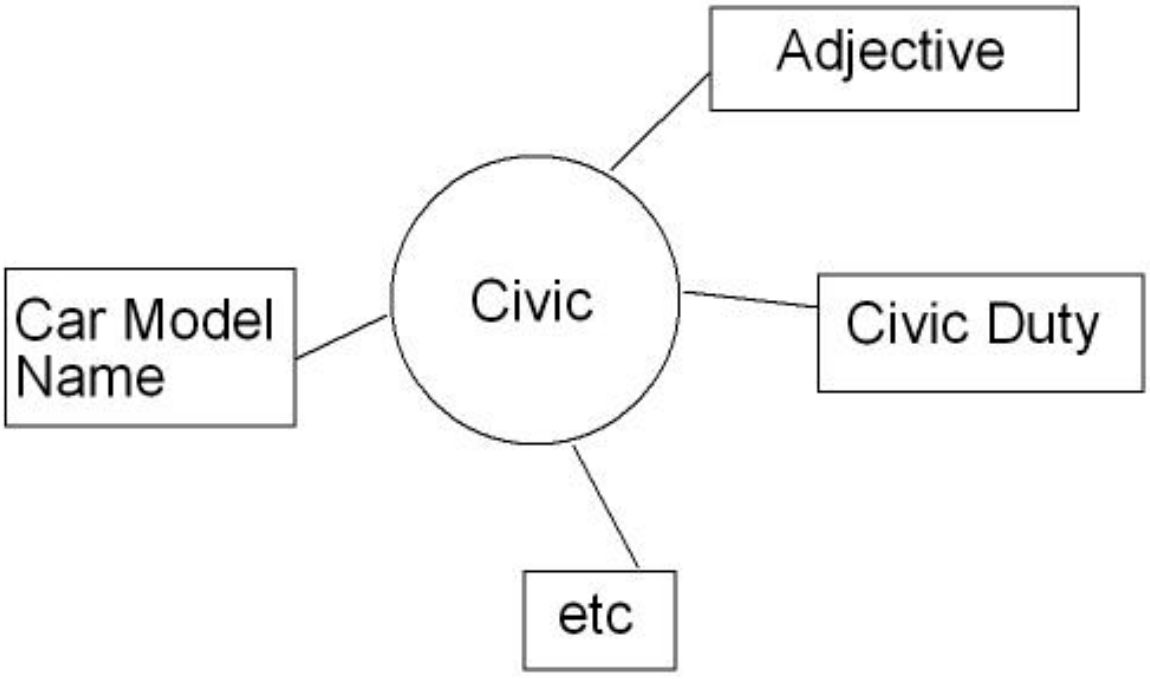


Search result of the keyword “Honda” and “civic” by using Google Search Engine.

G r a p h y 2



Everything related to “Honda”



Everything related to “Civic”

1.3 Data representation of the web

HTML(Hypertext Markup language)

HTML - simple markup language for specifying the structure
and content of information in Web documents. Tim Bernets-Lee created the first version of the World Wide Web program at CERN in 1990. A protocol was created to allow different machines to communicate and describe the document. Hypertext is a collection of documents with cross-reference (links) between them. In a sense, it resembles a human brain. Our senses could act as a link to retrieve our memory in our brain. For instance, the color of red may remind a Chinese descendent New Year. And the smell of kielbasa may remind a Polish descendent of his/her parents.factors act as links that would HTML is still the predominant language to create web page and is defined by DTD (Document Type Definition), within the context of SGML. HTML allows us to format the basic structure of the web by declaring contents within different tags (<tag>). A valid HTML declaration includes an open tag (e.g. <HTML>) and end tag (</HTML). A user can define font size ,color ,background and tables by utilizing HTML. But there's no understanding between computer and tags.

(See code and graph)

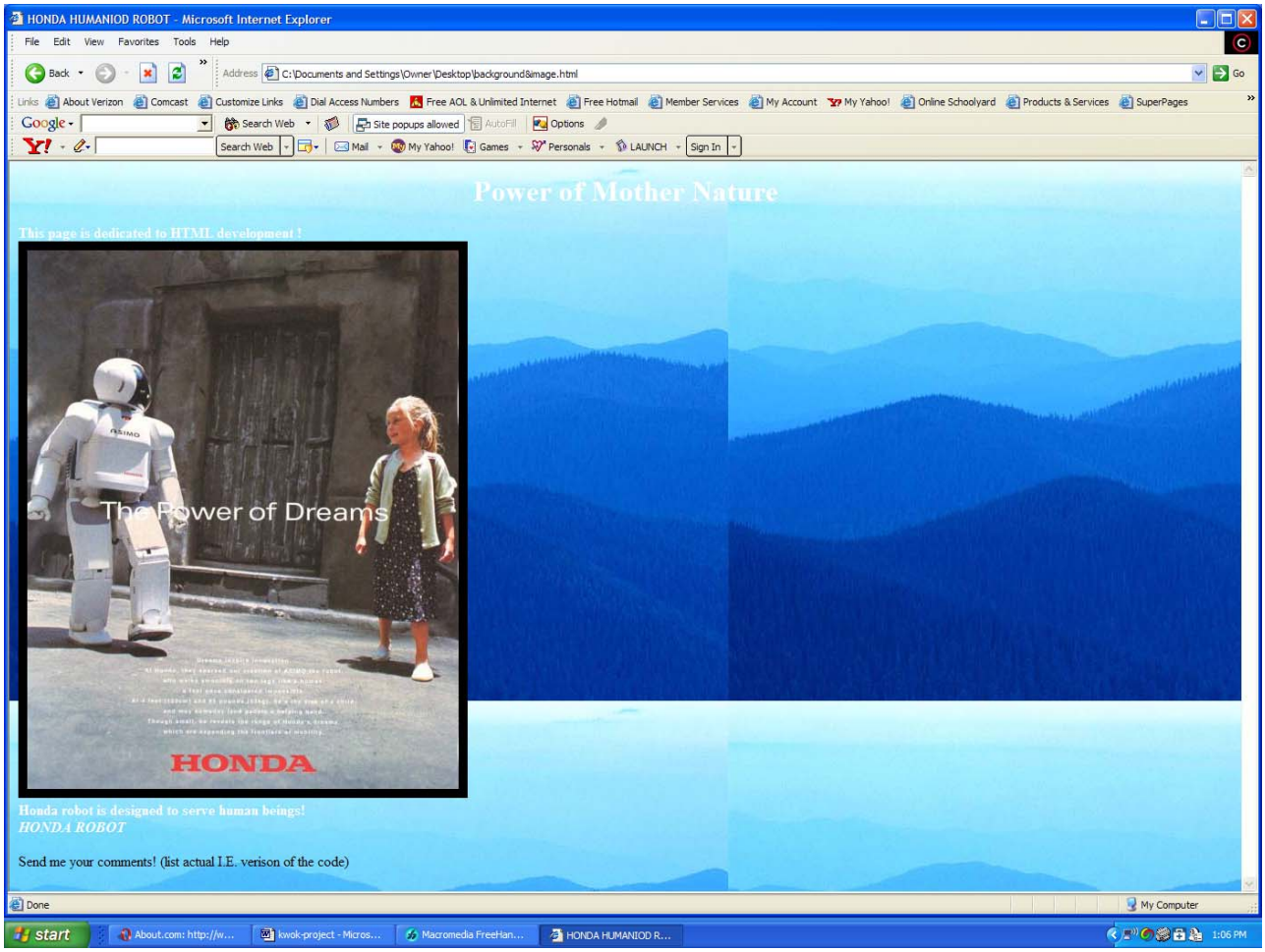
HTML code sample:

```
<HTML>
<HEAD>
<TITLE> HONDA Humanoid ROBOT </TITLE>
</HEAD>
<BODY BACKGROUND = "Blue Hills.jpg">

<FONT COLOR = "#FFFFFF">
<H1><CENTER> Power of Mother Nature </CENTER></H1>
<P ALIGH = "center"><B> This page is dedicated to HTML development !
<BR><IMG SRC = "honda-robot.jpg" ALIGH = "center" BORDER = "10">
<BR> Honda robot is designed to serve human beings!
<BR><I>HONDA ROBOT</I></B></P>

</FONT>
<A HERF="fredkwok@temple.edu ?subject=comments">Send me your
comments!</A>
</BODY>
</HTML>
```

Displaying the code with web browser.



Cascading Style Sheet:

CSS declares how document is displayed in the web browser. It can be applied to HTML

and XML-based language. It is created by W3C. CSS rules determine the color,

font, text formatting and colors and numerous background settings. CSS only

works with what is presented in the markup file and is backward compatible. The

look of the web page can be displayed independently from the information on the

page.

CSS rules: each rule is a single statement that identifies how and what to display.
The basic structure of CSS is:
-structure: It consists of two part,
-selector(s) and declaration(s). eg : selector { property: value;}

Here's a basic CSS style sheet with color, size, and font declaration.

```
/* basicCss.css
/* written by S.F.Freddie Kwok <fredkwok@temple.edu>

   body{ font-family: Arial;
         color: Orange;
         background-color: Black;}
/* orange headlines */
h1,h2,h3,h4,h5,h6{ font-family: orange, sans-serif;}

/* This puts the second level heading in read */
h2{color: black;}

address{font-family:Verdana, sans-serif;
        font-size: smaller;}
```

we can link the style sheet to XML by using <link> in both

HTML and XML

```
<?xml-stylesheet type = "text/css" href="filename"?>
```

in the XML document.

XML (Extensible Markup Language):

Sir Barnets-Lee also created the XML which is like an extension of the current HTML. It is also human-readable and self-describing. XML is to create a more solid and flexible foundation for the current Net. The Web master can descript a syntax that user create his own language. XML documents are able to represent hierarchies of concept. It descripts the structure of data and is built upon Unicode characters and URL which identifies concepts .XML only provides syntactic interoperability. One of the advantage of using XML is it allows Web authors to focus on structure rather than the format of documents to style sheets. The disadvantage of XML is it doesn't tell the web browser how to display information. Therefore it needs style sheet. Both HTML and XML are standards from World Wide Web Consortium. Information is useless without a lot of work to decipher it. XML was Computer scientist's effort to answer this question of Repackaging data. Flexibility, longevity and accessibility of data.

XML and the role of interoperability mechanism.

Data	Metadata
Freddie	Name
Narberth	City
Pennsylvania	State

The basic structure of XML includes tags, element, head, footing and body.

XML schema(replacement of DTD) , like RDF, is a separate document whose purpose is to define the legal elements, attributes and structure of XML instance. document.(XML schema, like HTML and XML, is recommended by W3C.)

XML sample code and graph

```
<?xml version="1.0"?>

<INVENTORY>
<BOOK>
        <TITLE>OLENKA IN POLSKA</TITLE>
        <AUTHOR>Sung Fat Freddie Kwok</AUTHOR>
        <BINDING>mass market paperback</BINDING>
        <PAGES>298</PAGES>
        <PRICE>$5.49</PRICE>

</BOOK>
```

Cascading Style Sheets(CSS) specify the format and layer for an XML document. The new version of Internet browser needs to understand the content of XML.

```
<name nickname = "Fat">
        <first> Sung Fat Freddie </first>
        <middle> none </middle>
        <last> Kwok </last>
</name>
```

Tags, elements and attributes in XML.XML as simple form of SGML. Web master use different Attributes to separate different type of information, though it adds complexity to the data structure, meta data is the description of the data. RDF (Resource Descripting Framework):RDF: Resource Description Framework is a W3C recommended format that attempts to address XML's semantic limitation. It represents node connected by labeled arcs,

nodes = resource, arc = properties.

RDF allows the interchange of XML more effectively.

<rdf:Description about='http://www.temple.edu/RDF/Why-RDF.html'>
<Author>Sung Fat Kwok</Author>
<Home-Page rdf:resource='http://www.temple.ed' />
</rdf:Description>

Containers for objects

: bag, sequence and alternative.
reification: descripting data (metadata)

RDF schema allows user to create schemas of standard classes

and properties using RDF.

URI identify resource such as classes and properties.

- no polysemy problem, it gives resource an ID.
- definition of class is a collection of statement about its properties.

Axiom could be used to infer data that has not been presented

implicitly. It can map definition of different concepts and

layers of Semantic web.

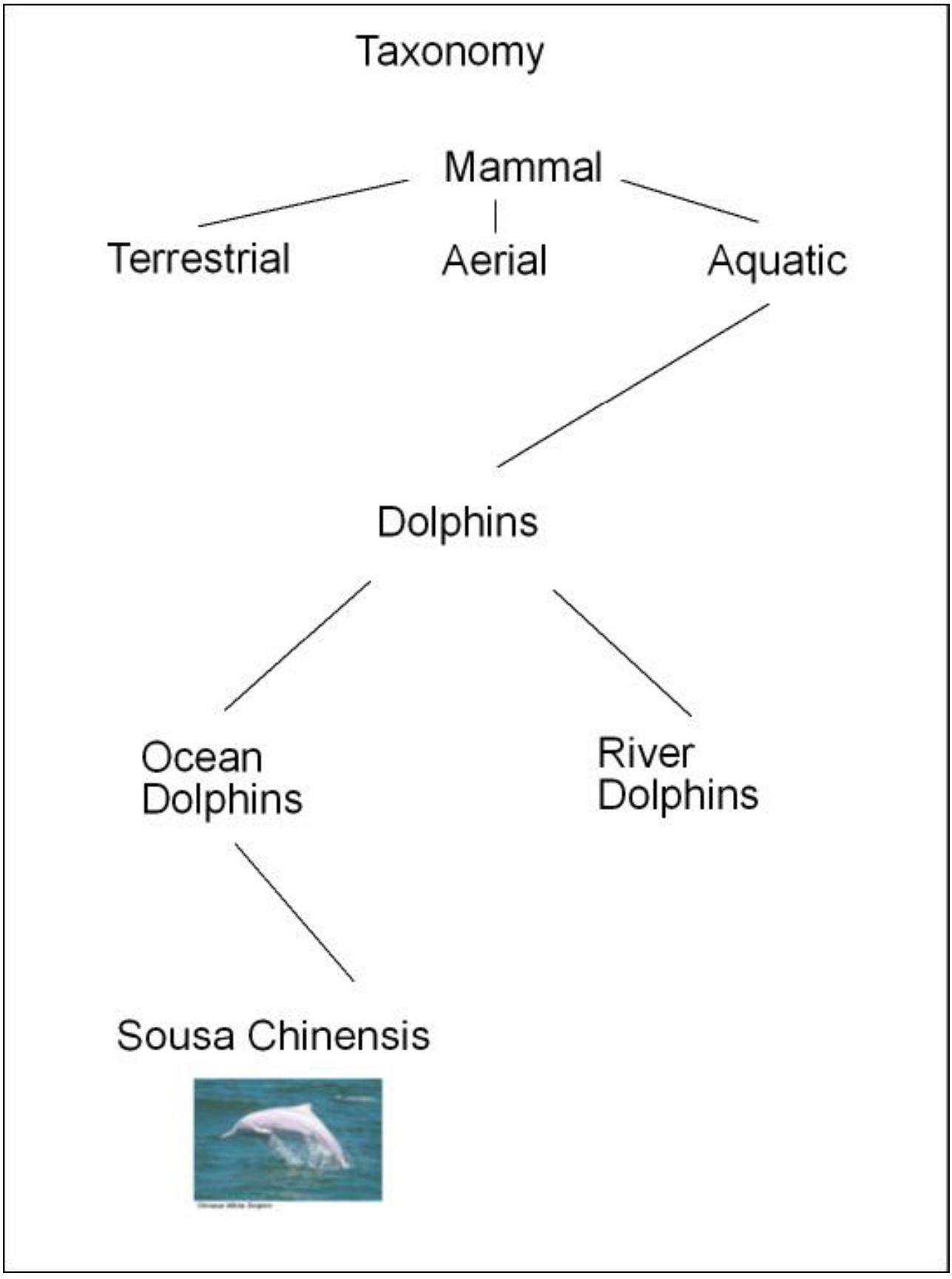
Paper Overview:

This paper will investigate feasible solutions for achieving a semantic web agent Which would further assist human beings by freeing up their time for leisure or intellectual activities. To achieve this goal, the paper would analyze the web markup language (HTML,XML) and the relation between XML and ontology. This is how middleware might solve this problem. The paper will discuss the potential usage of the A.I. Semantic Web agent so that the reader will have a better vision of

the power of the next generation of Internet.

2.Background

2.1 Representing information on the Semantic Web. The self-defining tags of XML also make it extremely difficult to unify standard, and cope with the problems synonymy(different words are used for the same meaning) and polysemy (same word used for different meaning).But the freedom of XML is a threat to the reliability of information on the Net. Semantic web is a net with infinite capacity. Kurt Godel, the famous logician, once said that human thoughts are sophisticated and sometimes human languages fail to capture the essence of the thought. What the paper is trying to do achieve is to create an agent that changes with the different contents of the Semantic web. The Net is a billion-web-page database would expand nonstop as long as computers exist. The current best search engine only crawls one fourth of the indexed web. An agent in a Semantic Web environment would revolutionize everything from effective searching to new levels of E-commerce and Net application. Semantic Web harbors plenty of concepts, and each concept can be represented by a node in a graph. The nodes are connected by arcs which stand for relation. For instance, instance-of , has-a use arc to represent concepts. is-a represent subset relation, instance-of represent element of relation. (in set theory).
(see graph)


Taxonomy
Mammal
Terrestrial
Aerial
Aquatic
Dolphins
Ocean Dolphins
River Dolphins
Sousa Chinensis

Predicate logic/calculus:

Many things can be represented as objects and relationship in the world. For instance, Temple University has a Computer Science department and the department itself has Professor Wang.

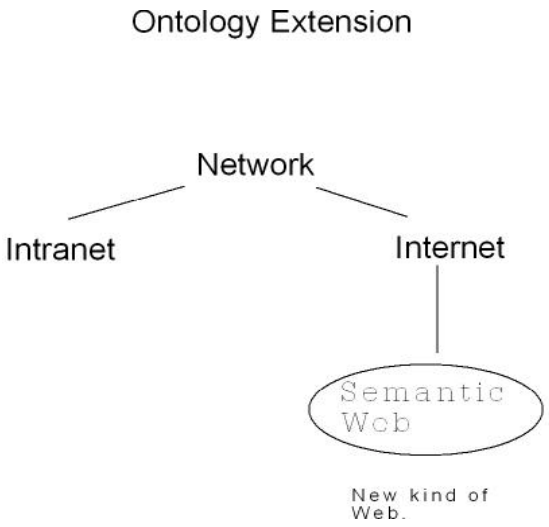
Logical symbols: quantification, implication, conjunction, disjunction.

Non-logical symbols: constants, predicates, function and variable.

Ontology:

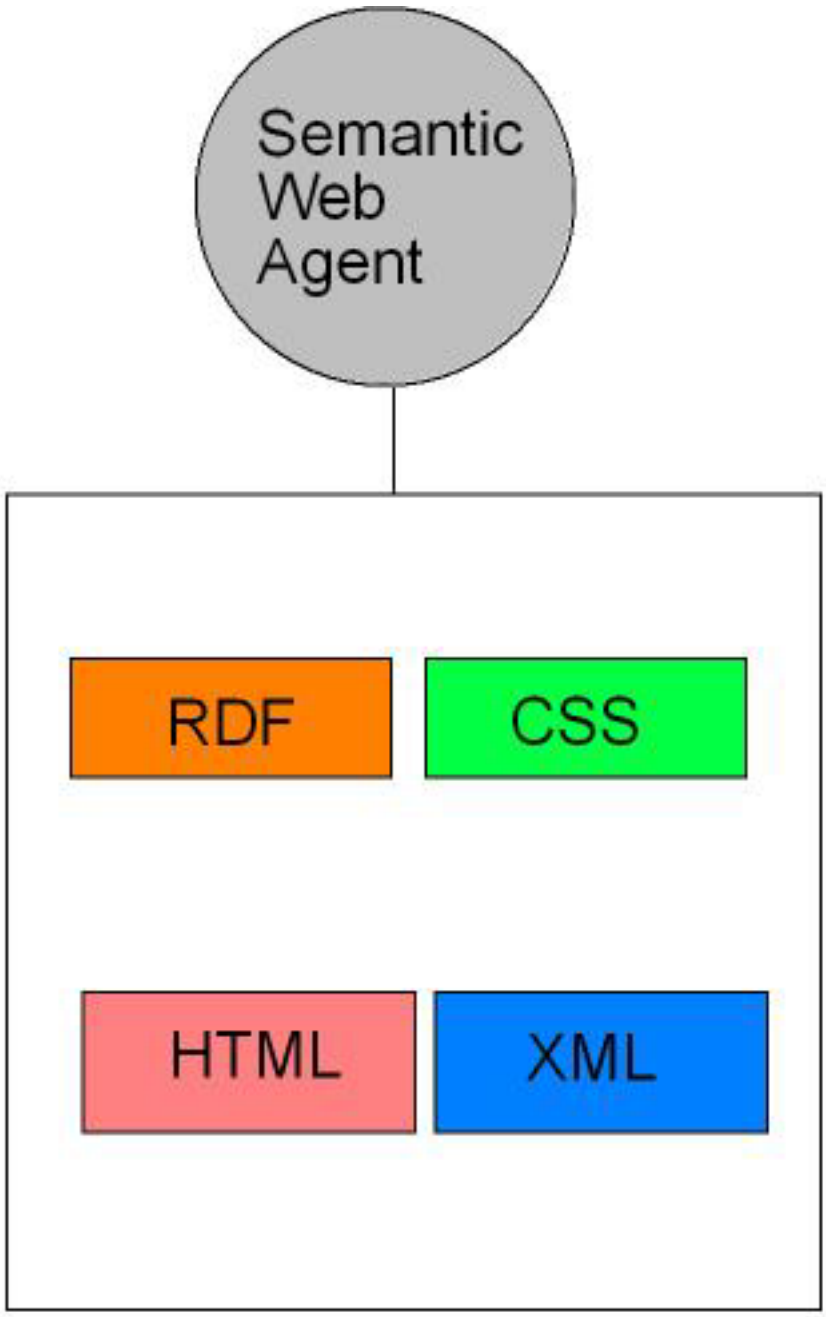
If two web documents share an understanding of the relevant domain and structure,then it should be possible for them to reuse and share their knowledge. When two parties agree on ontology, they agree on the shared insights. The term ontology is vague. Large ontology can be constructed by assembling and refining existing tag components. product(S,T,G) S is a student identifier, T is the school he/she graduated from. G is the graduate GPA of the student. Let us assume US education is used. Then 4.0 would be the highest possible GPA. But what if the website wants to include students from different nations? The web agent has to convert different GPA's to US standard. product(S,T,G,C) But then an evaluating agent has to be created to deal with this situation. A Semantic Web agent would be sent to evaluate ontology on the web and combine any of them if needed, The Domain Difference increases the semantic heterogeneity. It requires sophisticated language To define programs and functions.

Graph 3



2.2 Concept of Semantic Web agent

If most of the existing ontology can be captured and mapped by a Semantic Web agent, then an Internet user would be able to extract information from the web. More precise data-mining ability means that different academic institutions can interchange data dynamically. The Semantic Web would definitely empower supersized grid computing .which allows each personal computer to tackle a tiny piece of a complex problem such as studying drug resistance of various diseases. (see graph)



Agent to map ontologies on Semantic Web.

2.3 Architecture of Semantic Web

LOGICAL FOUNDATION OF THE SEMANTIC WEB:

Giving out denotation semantics for web pages such as applying predicate calculus to describe the website. Domain, objects of interest. Internet source as IS, the domain of discourse. potential Internet resource : PIR, Infinite set of well-formed formulas could be constructed within IS.

3. Hypothetical Case study

Case application in influenza:

Different web pages provide certain degree of information accuracy. It is difficult to allocate relevant information. Map between terminology and contextual information. Rule to identify the symmetric, inverse, and transitive relationship .Source material for TSE risks: source material, processing, and end-product use.

```
...
<body>
<ontology id="flu ontology" version ="1,0">
<use-ontology id="base ontology" versoin = "1.0" prefix="base">

<def-category name="disease_agent" ISA="base.shoeentity:>
<def-category name="disease_agent" ISA="disease_agent">

....
<relation name="hasInput">
        <ARG POS = 1 TYPE="Process">
        <ARG POS = 2 TYPE="Material">
</relation>
<RELATION NAME = "hasOutput">
        <ARG POS=1 TYPE="Process">
        <ARG POS=2 TYPE="Material">
</RELATION>

...
```

</ONTOLOGY>
</BODY>

3.1.1 The ontology

A well-defined Ontology Design reflects category hierarchies, relations, and inference rules. The source code of the case study defines names that are attached with categories and relations. Declaring and extending ontology is critical to the success of the Semantic Web. The code from the case study determines:

-Defining has-a and is-a relations between concepts and objects.
- What sorts of answer to be presented to users.
- Relationship between objects and ontology.

The code from the case study also shows rules that try to map tags with similar meanings together. Defining classes and relationships is a good start toward establishing a workable ontology. Includes similar meaning to the oncology. It is time-consuming to construct a well-defined ontology mechanism. Use inclusion-exclusion theory. The challenge comes when ontology software need rule to identify the symmetric, inverse, and transitive relationship .
(e.g.: x=y, y=z, x=z , a transitive property)

3.1.2 Annotation

ANNOTATION is the process of adding semantic markup to a web page. An instance would be identified by a key,(form by URL).Knowledge annotator display instances ,ontology and claims.

-Different objects can exist in different categories.

-Every single meaningful words.

-Relationship between ontology would also be descried as well.

Category and Relation and Instance.

```
<HTML>
<BODY>
..
<INSTANCE KEY="http:://www.health.temple.edu">
<USE-ONTOLOGY ID = "flu-Ontology" VERSION="1.0" PREFIX="flu"
<CATEGORY NAME="flu.Process">
<RELATION NAME="flu.name">
        <ARG POS="TO" VALUE="Rendering">
</RELATION>

....
</RELATION>
</INSTANCE>
</BODY>
</HTML>
```

3.1.3 How to process the annotation.

The most challenging part of the creation of Semantic Web is to process the annotation. Annotation describes URL or the relation between objects that web master creates. An manual web annotation maybe a workable solution. But processing them requires a much more effective approach.

3.1.4 Information analysis

Software such as SHOE (Simple HTML Ontology Extensions) and OWL can work out a certain amount of ontology. But they're still far from perfect. If Computer scientists can find a way to establish a program that analysis contents of web sites,

then a Semantic Web agent should be able to map ambiguous ontology and even combine them.

3.1.5 Summary of the case study

4. Proposal of the Semantic Web middle agent.

I propose to create a web middle agent that could distinguish the relationship and hierarchy of different tags on the Semantic web. Mapping, reusing and dynamic communicating with different web. By embedding the software like WordNet, a human lexical memory developed by Princeton University, in the functions of the Semantic Web agent, a more accurate mapping between web ontology may become possible. In addition, a better software has to be developed to meet the almost infinite ontology and information in the world. Hopefully, an agent can act as a "middle-man" between all these communication layers.

5. Summary

Due to the complexity of both human and computer language and its flexibility ,It is extremely difficult to accomplish Semantic web agent at this time. The progress of transferring from the current Net to the Semantic Web would speed up if everyone starts annotating his/her web pages. The application of Semantic web by using Semantic web middle agent may be feasible. Similar to the emergence of popularity of computer, it may take several years for the general public to

recognized its importance. To realize the potential of the Internet, an Semantic Web agent is needed to solve the problems of communication levels. A workable Semantic Web agent is indeed a complicated problem that requires lots of effort from the Computer Science community. We not only have to work out the infrastructure of the net, but also a standard medium to make them work. I believe that successful implementation of Semantic web agent would be another breakthrough for the Computer science community. It would transform any field that uses computer system.

Reference

Lee, Berners-Lee(2002)
Commemorative Lecture, The World Wide Web - Past Present and Future,
For Japan Prize Commemorative Lecture.

Berners-Lee, Tim & Hendler, James & Lassila, Ora (2001-5-17)
The Semantic Web, A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities.
Scientific American.

Berners-Lee, Tim
Why RDF is more than XML.
http://www.w3.org/DesignIssues/RDF-XML.html

B.Palmer, Sean(2001-09)
The Semantic Web: An Introduction .

E science: The Grid and the Semantic Web
IEEE: distributed systems online.

De Roure, David: University of Southampton
Hendler, James A. University of Maryland

Hendler, James (2001-3)
IEEE Intelligent System Journal
Department of Computer science, University of Maryland, COllege Park. Agents and the Semantic Web

Godel Incompleteness Theroem,
THe Time 100 ; Jurt Godel by Time magazine.

The Semantic Web: A Primer , O'Reilly XML.com.

Math Awareness Month (2004-4)
The Mathematics of Networks - The small-World Phenomenon and Decentralized Search

Hypersearching the Web by Members of the Clever Project, IBM Almaden Research Center,
San Jose.