

Ontology-based web annotation framework for hyperlink structures

Naing, Myo Myo; Lim, Ee Peng; Goh, Dion Hoe-Lian

2002

Naing, M. M., Lim, E. P., & Goh, D. (2002). Ontology-based web annotation framework for hyperlink structures. Proceedings of the International Workshop on Data Semantics in Web Information Systems, (December 11-14, Singapore), 184-193.

<https://hdl.handle.net/10356/79615>

<https://doi.org/10.1109/WISEW.2002.1177862>

Proceedings of the International Workshop on Data Semantics in Web Information Systems @ 2002 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE. This material is presented to ensure timely dissemination of scholarly and technical work. Copyright and all rights therein are retained by authors or by other copyright holders. All persons copying this information are expected to adhere to the terms and constraints invoked by each author's copyright. In most cases, these works may not be reposted without the explicit permission of the copyright holder.

Naing, M.M., Lim, E. P., and Goh, D.H. (2002). Ontology-based web annotation framework for hyperlink structures. *Proceedings of the International Workshop on Data Semantics in Web Information Systems*, (December 11-14, Singapore), 184-193.

Ontology-based Web Annotation Framework for HyperLink Structures *

Myo-Myo Naing Ee-Peng Lim[†] Dion Goh Hoe-Lian
Centre for Advanced Information Systems
School of Computer Engineering
Nanyang Technological University, Singapore 639798
{p118008, aseplim, ashlgoh}@ntu.edu.sg

Abstract

Heterogeneous and autonomous web sites contain information that cannot be easily processed by automated tools. To address the heterogeneity issues, one attempt is to add metadata to Web pages to provide semantic description of the web pages and other kinds of web objects. This is also known as web annotation. Ontology-based Web annotation (OWA) is a kind of web annotation that associates ontologies with annotations. Since each ontology is usually adopted by some user community, OWA therefore facilitates sharing of knowledge among users sharing the ontology. In our survey, we noted that most existing OWA approaches only annotate web pages, but not the hyperlinks among them. This restricts the association of useful semantics to the hyperlinks between web pages. In this paper, we therefore propose to annotate hyperlink information as part of OWA. Users can create annotations for hyperlink structures and to utilize such annotations in browsing links between web pages and querying web sites. We first present a general OWA framework consisting different components. An XML-based annotation scheme for annotating hyperlink structures is also proposed.

1. Introduction

The World Wide Web (WWW) consists of huge number of web servers providing abundant information for sharing and exploration. As these web sites are operated by autonomous organizations providing a wide range of content, locating useful information among the web sites can be a taunting task. This is because web sites and their pages are often structured in different styles and formats making them more eye-pleasing but also harder for their content to

be automatically extracted and processed. Without a good semantic knowledge about the web sites and pages, it is difficult to build sophisticated applications using the web information.

Web annotation refers to adding metadata to Web pages to provide higher level interpretation of the web pages and other web objects. It can be any comments, notes, explanations, references, examples, advices, correction or any other type of external remarks that can be attached to the whole or part of web pages without modifications to the original pages. Other than web pages and their text content, web annotations can also be assigned to web sites, tables and figures within web pages. In most existing web annotation systems, e.g. ComMentor¹, Annotator², Third Voice³, CritLink⁴, CoNote⁵ and Futplex⁶, the web annotations do not conform to some common semantic structures pertinent to the underlying domains of the web content. Since these web annotations are created without a controlled and well-defined model, they are simply another layer of data which share the same heterogeneity and autonomy characteristics with the underlying web [4]. To overcome this limitation, some attempts have been made to create Web annotations based on some well-defined semantic structures or models, known as **ontologies** [13, 8].

Ontology has been defined in many ways in the research literature [1, 17, 9, 13, 12]. In this paper, ontology is a specification of semantically interrelated entities within a given domain. Using ontologies as predefined semantic structures, users interpret web information as concepts and relationships within the ontologies, and create annotations to assign the pre-defined semantics to the web information [19]. To distinguish such annotation approach from the rest, we call this the **Ontology-based Web Anno-**

¹<http://hci.stanford.edu/commentor/>

²<http://www.foresight.org/WebEnhance/Annotate.html>

³<http://www.thirdvoice.com/>

⁴<http://crit.org/http://crit.org/critlink.html>

⁵<http://www.cs.cornell.edu/home/dph/annotation/annotations.html>

⁶<http://gewis.win.tue.nl/applications/futplex/>

*This work is partially supported by the SingAREN21 research grant M48020004.

[†]Dr. Ee-Peng Lim is currently on leave at Dept. of SEEM, Chinese University of Hong Kong, China.

tation (OWA). For example, a university ontology may include the concepts of ‘Professor’ and ‘PhDStudent’. Based on this ontology, the homepages of professors and graduate students at a university web site could be annotated as the instances of ‘Professor’ and ‘PhDStudent’ concepts respectively.

In recent years, several OWA schemes and languages have been developed. However, most of the existing OWA approaches [10, 13, 20, 15, 21] annotate web pages only, by associating web pages with concepts from some given ontologies. If a web page is related to another web page according to some ontological relationship, the existing approaches usually treat the URI of the latter web page as an attribute value of the annotation instance created for the former web page. The hyperlink structure between these two web pages, which actually corresponds to the relationship, is usually not captured by the annotation. This clearly prevents us from fully understanding the semantics behind the hyperlink structure and using them for more intelligent processing of web information.

In this paper, we first propose a generic annotation framework for studying the overall OWA problem. We focus on the ontological annotation of hyperlink structures, and propose some extensions to the existing annotation language to represent such annotations. Several examples using an University Ontology are also given to illustrate our proposed extensions.

2. Annotation of Hyperlink Structures between Web Pages

Consider the web pages in Figure 1. The figure shows a professor homepage belonging to Dr Lim Ee Peng. The web page contains a link to another web page containing a list of students supervised by Dr Lim. This web page also provides the hyperlinks to the student homepages. Among them is the homepage of Myo Myo, a PhD student. Let us assume that we are also given a University ontology depicted in Figure 2. Informally, the boxes represent concepts and the arrowed lines between them represent relationships. The attributes of concepts and relationships are listed within the boxes and on the lines respectively.

To perform ontological web annotation, a typical existing annotation approach will create an annotation instance for Dr Lim’s homepage associating it with the Professor concept, and another annotation instance for Myo Myo’s homepage associating it to the PhDStudent concept. The Supervision relationship between these two web pages will be annotated by having the URI of Myo Myo’s homepage inserted as an attribute value in the annotation instance of Dr Lim’s homepage. The annotation approach does not annotate the actual hyperlink structure connecting Dr Lim’s homepage with Myo Myo’s homepage corresponding to the

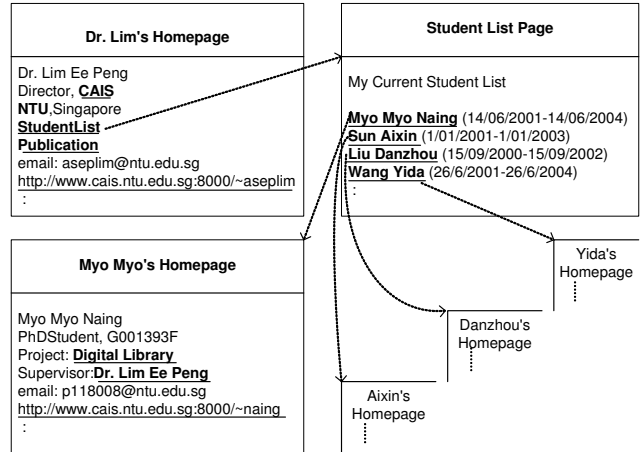


Figure 1. Example Web Pages with Hyperlink Structures

Supervision relationship.

In the above example, the hyperlink structure connecting from Dr Lim’s homepage to Myo Myo’s homepage represents an essential piece of information about the Supervise relationship. It allows us to know which hyperlink or chain of hyperlinks constitutes the Supervise relationship. With such a knowledge, navigation among the web pages can be more intelligently done.

Consider Figure 3 which depicts another chain of hyperlinks from Dr Lim’s homepage to Myo Myo’s homepage via the homepage of the research center (i.e., CAIS) Dr Lim belongs to, the project listing web page of the center, the Digital Library project web page, and the Digital Library Project People web page. Note that while the same Professor and PhDStudent web pages are involved, this chain of hyperlinks does not carry any semantics about the Supervise relationship. This particular chain of hyperlinks merely states that Dr Lim and Myo Myo are members of the research center. Hence, it is important to distinguish between the two different chains of hyperlinks between the Professor’s and PhDStudent’s web pages. In this paper, we introduce the annotation of hyperlink structure to capture such a knowledge.

To the best of our knowledge, although most of the existing OWA approaches attempt to annotate relationships between web pages, none of them provides the abilities to annotate hyperlink structures as relationships between web pages. Such hyperlink structures can be very useful in a number of applications. We cite two obvious ones below:

- **Browsing:**
When a user browses a web page, annotated as an instance of some concept in an ontology, he or she can

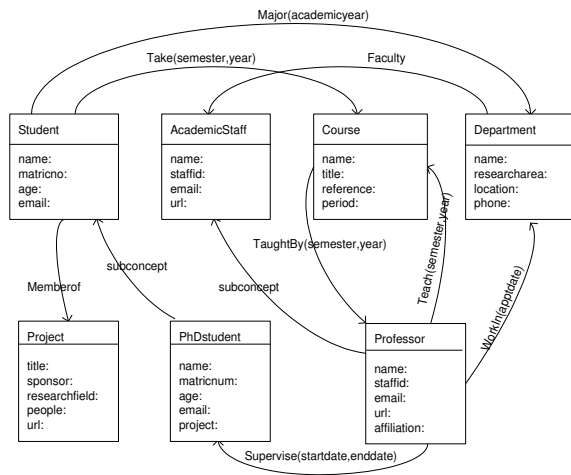


Figure 2. An example part of University Ontology

examine the relationship semantics of the links of the web page before deciding the link to be traversed next. In other words, the user can be presented with the relationship(s) associated with each annotated link in the web page. With the additional relationship information, the user can make more informed and better decision on navigating the web site.

• Query :

With hyperlinks annotated, users can now query information associated with hyperlinks and their assigned semantics. Examples of such queries include:

- Find the links that connect a 'School' homepage to it's 'AcademicStaff' homepages.
- Find the homepages of PhDStudents which are not connected with any project web page by hyperlinks annotated with 'Memberof' relationship, suggesting that the students have not inserted links to their projects.

3. OWA Framework

Before we present our proposed annotation approach for hyperlink structures, we first introduce a framework for ontological web annotation as shown in Figure 4. The framework, represented as an extended ER diagram, aims to describe the different entities involved in OWA and their inter-relationships. It allows us to study an annotation approach systematically and to put our study on annotation of hyperlink structures in right perspective. Note that we have suppressed the attribute information in the diagram to ease reading.

In our framework, web objects are the items to be annotated. A web object can be a web site, a web page, or any item within a web page such as a multimedia object, table, etc.. An annotation instance can be created for a single web object or a set of web objects. The latter is required when the web objects are closely related and may be better to be annotated together. An example is the annotation of both personal and official homepages of a person since they describe the same person with different focuses.

Figure 4 also shows that multiple annotation instances (or simply annotation) can be created for each web object, but each annotation instance is represented in some annotation language. Ideally, one would like to see all annotation instances created using the same annotation language. Nevertheless, the reality is quite different since each annotation language provides different set of annotation features. The existing annotation languages includes Ontobroker [11], SHOE [13], RDF [16, 18], and others.

To adopt OWA, ontology must be used in the creation of annotation instances. There are many ways ontology can be

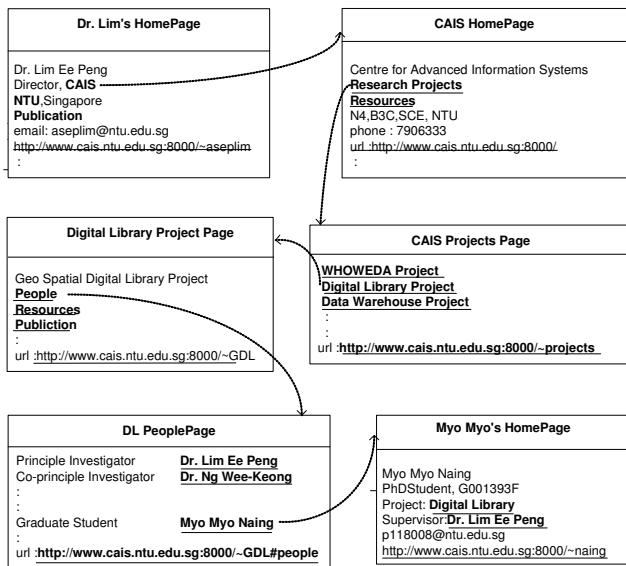


Figure 3. Example Web Pages with Additional Hyperlink Structures

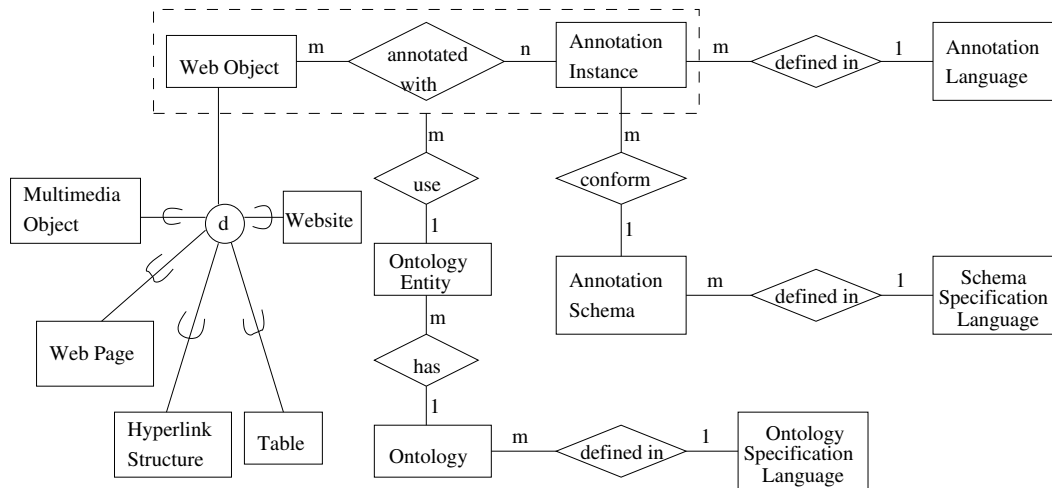


Figure 4. OWA Framework

defined. In our framework, an ontology consists of ontology entities. In Section 4.1, we give our own formal definition of ontology which consists of concepts, relationships, attributes and axioms. An ontology specification language is used to define ontologies. As we create an annotation instance, the annotated web object is associated with some ontology entity. This association is recorded within the annotation instance.

Very often, using the same annotation language and ontology, one can still create many different annotation instances for the same web object. For example, the eXtensible Markup Language (XML) itself is an annotation language but it can be used in different ways to represent an annotation instance as shown in Figures 5 and 6.

```
<concept name="Professor">
  <attribute>
    <name>Ee-Peng Lim</name>
    <email>aseplim@ntu.edu.sg</email>
    <url>http://www.cais.ntu.edu.sg:8000/~aseplim</url>
    <affiliation>CAIS</affiliation>
  </attribute>
</concept>
<relationship name="Supervise">
  <source>http://www.cais.ntu.edu.sg:8000/~aseplim</source>
  <target>http://www.cais.ntu.edu.sg:8000/~naing</target>
</relationship>
:
:
```

Figure 5. Example XML Annotation instance(1)

To unify the representation, the annotation schema is introduced. Annotation schema describes how the annotation instances are formatted. In other words, annotation schema

```
<Professor>
  <name>Ee-Peng Lim</name>
  <email>aseplim@ntu.edu.sg</email>
  <url>http://www.cais.ntu.edu.sg:8000/~aseplim</url>
  <affiliation>CAIS</affiliation>
</Professor>
<Supervise>
  <source>http://www.cais.ntu.edu.sg:8000/~aseplim</source>
  <target>http://www.cais.ntu.edu.sg:8000/~naing</target>
</Supervise>
:
:
```

Figure 6. Example XML Annotation instance(2)

is a common template for instantiating annotation instances. With such a template, it is then possible for any applications to parse and extract information from annotation instances. Note that the annotation schema itself can be represented in some annotation schema language.

4. Proposed Annotation Schema

In this section, we present our proposed annotation schema that can be used to annotate web objects with concepts, relationships and attributes from a given ontology. We first give the definition of ontology. We then describe the different types of hyperlink structures to be covered by our proposed annotation schema. We finally define our proposed annotation schema. To allow wide adoption of our proposed annotation approach, we have chosen XML [3] as the annotation language and XML Schema [22] as the annotation schema language respectively. There are several advantages for using XML:

- XML syntax are simple and their parsers are widely available;
- Since annotations are intended for querying web information, having them represented in XML allows us to apply some XML query languages on them without inventing a new set of query languages; and
- Database systems supporting XML can be readily used to implement the XML-based annotation approaches.

4.1. Definition of Ontology

In our OWA framework, ontologies serve as the domain knowledge for organizing semantically interrelated data on the web. Compared to the traditional relational database schema, an ontology can describe more complex objects due to its richer semantic constructs.

In the following, we give the formal definition of ontology.

Definition 1 (Ontology) An ontology consists of 6 elements $\{C, A^C, R, A^R, H, X\}$, where C represents a set of concepts; A^C represents a collection of attribute sets, one for each concept; R represents a set of relationships; A^R represents a collection of attribute sets, one for each relationship; H represents a concept hierarchy; and X represents a set of axioms.

Each concept c_i in C represents a set of objects of the same kind, and can be described by the same set of attributes denoted by $A^C(c_i)$. Each relationship $r_i(c_p, c_q)$ in R represents a binary association between concepts c_p and c_q , and the instances of such a relationship are pairs of (c_p, c_q) concept objects. The attributes of r_i can be denoted by $A^R(r_i)$. H is a concept hierarchy derived from C and it is a set of parent-child (or superclass-subclass) relations between concepts in C . $(c_p, c_q) \in H$ if c_q is a subclass, or sub-concept, of c_p . Each axiom in X is a constraint on the concept's and relationship's attribute values or a constraint on the relationships between concept objects. Each constraint can be expressed like a Prolog-like rule [2].

In the following, we describe the simple University ontology shown in Figure 2. This ontology example will also be used throughout this paper.

Example 1 $Univ_Ontology = \{C_{univ}, A_{univ}^C, R_{univ}, A_{univ}^R, H_{univ}, X_{univ}\}$ where

- $C_{univ} = \{Student, PhDStudent, AcademicStaff, Professor, Department, Course, Project, \dots\}$
- $A_{univ}^C = \{A_{univ}^C(Student), A_{univ}^C(PhDStudent), A_{univ}^C(AcademicStaff), A_{univ}^C(Professor),$

$$A_{univ}^C(Department), A_{univ}^C(Course), \dots\}$$

$$A_{univ}^C(Student) = \{name, matricnum, email, \dots\}$$

$$A_{univ}^C(PhDStudent) = \{name, matricnum, email, project, \dots\}$$

$$A_{univ}^C(AcademicStaff) = \{name, staffid, \dots\}$$

$$A_{univ}^C(Professor) = \{name, staffid, email, \dots\}$$

$$A_{univ}^C(Department) = \{name, researcharea, \dots\}$$

$$A_{univ}^C(Course) = \{name, title, period, \dots\}$$

- $R_{univ} = \{Supervise(Professor, PhDStudent), Teach(Professor, Course), TaughtBy(Course, Professor), Faculty(AcademicStaff, Department), TakeCourse(Student, Course), \dots\}$
- $A_{univ}^R = \{A_{univ}^R(Supervise), A_{univ}^R(WorkIn), A_{univ}^R(Major), A_{univ}^R(Teach), A_{univ}^R(Take), \dots\}$

$$A_{univ}^R(Supervise) = \{startdate, enddate, \dots\}$$

$$A_{univ}^R(WorkIn) = \{apptdate, \dots\}$$

$$A_{univ}^R(Major) = \{academicyear, \dots\}$$

$$A_{univ}^R(Teach) = \{semester, year, \dots\}$$

$$A_{univ}^R(Take) = \{semester, year, \dots\}$$
- $H_{univ} = \{(Student, PhDStudent), (AcademicStaff, Professor), \dots\}$
- $X_{univ} = \{TaughtBy(X, Y) \leftarrow Teach(Y, X), \dots\}$

This ontology can be represented in any ontology representation language such as RDF(S) [5], DAML+OIL [6] and others. In this paper, we only focus on the concept, relationship and attribute entities of an ontology since they are the usual entities involved in annotation. Axioms, on the other hand, are useful when reasoning based on ontology is required.

4.2 Types of Hyperlink Structures

We now review the types of hyperlink structures that can be annotated. Hyperlink structure between two given web pages can be represented as follows:

$$Source \quad Link \quad (Ip \quad Link) * \quad Target$$

The source page (*Source*) is linked to the target page (*Target*) with at least one link (*Link*). When more than one link is involved, there will be intermediate

page(s)(I_P) between every adjacent links pair. Sometimes the relationship between two web pages is described by text string in the source page without any hyperlink. In our approach, we assume that two web pages are linked by at least one hyperlink so that we could be able to annotate the hyperlink information.

We classify the linkage between a source web page and a target web page into two categories: *Single-path Link* and *Multi-path Link*.

- *Single-path Link*: In *Single-path Link*, there is only one path available starting from the source page to a target page. The source page is connected to a target page by a single path represented by a sequence of links and intermediate pages. For example, a single-path link that describes the Supervise relationship between a Professor homepage and a PhDStudent homepage is shown as follows:

$$S_A \xrightarrow{L_1} I_{P1} \xrightarrow{L_2} I_{P2} \xrightarrow{L_3} I_{P3} \xrightarrow{L_4} T_B$$

Where $S_A = \text{Professor homepage}$

$T_B = \text{PhDStudent homepage}$

$I_{P1} = \text{project list page}$

$I_{P2} = \text{Project homepage}$

$I_{P3} = \text{list of people page}$

The intermediate pages are pages describing information related to the Professor and PhDStudent. They may or may not be annotated as concept instances. For example, the project homepage may be annotated as an instance of ‘Project’ concept. If an intermediate page in the chain has been annotated as a concept instance, this information is also kept as the *sc-name* attribute in the linkchain information. The *sc-name* value is *none* by default if the intermediate page is not a concept instance.

- *Multi-path Link*: When a source page is connected to a target page by traversing through more than one path of the links between them, and they all capture the same semantic relationship between the source and target pages, we say that there is a *Multi-path Link* between them. Some of the links and intermediate pages in the different paths may be the same while others are different. For example, multi-path link consisting two paths below may represent the ‘Supervise’ relationship:

$$S_A \xrightarrow{L_1} I_{P1} \xrightarrow{L_2} I_{P2} \xrightarrow{L_3} I_{P3} \xrightarrow{L_4} T_B$$

$$S_A \xrightarrow{L_{1n}} I_{P1n} \xrightarrow{L_{2n}} I_{P2n} \xrightarrow{L_{3n}} I_{P3} \xrightarrow{L_4} T_B$$

4.3 Sharing of Annotated Subpaths

Very often, different linkchains may share some common intermediate pages and links. If they are annotated in every linkchain containing them, the amount of annotation information will be very large.

To address this problem, we introduce the Shared-path Link. A shared-path is a sub-path between a source and a target page that has been annotated and the subpath is also used in the path between another pair of source page and target page. For example, a shared-path link from S_A to I_{P3} is shared between the (S_A, T_B) pair and (S_A, T_C) pair.

$$S_A \xrightarrow{L_1} I_{P1} \xrightarrow{L_2} I_{P2} \xrightarrow{L_3} I_{P3} \xrightarrow{L_4} T_B$$

$$S_A \xrightarrow{L_1} I_{P1} \xrightarrow{L_2} I_{P2} \xrightarrow{L_3} I_{P3} \xrightarrow{L_5} T_C$$

S_A may denote a Professor homepage, T_B and T_C are two student pages. The sub-path from the Professor homepage to the list of people page(I_{P3}) is common.

4.4 Proposed Annotation Approach

The core of the our annotation framework is the annotation schema which defines the template for creating annotation instances. We earlier mentioned that there are existing annotation approaches using a variety of annotation languages. Most of them adopt some implicit annotation schema commonly agreed by a community of users. They however focus only on annotating web pages with concepts but not the hyperlink structures.

Instead of inventing an entirely new annotation approach, we have chosen to extend the data annotation approach presented in[4] with the capability of annotating hyperlink structures. In addition to the use of the concept tag for associating web pages with concepts, we introduce the relationship tag to allow hyperlink structures of relationship instances and their relevant attributes to be annotated. Figure 7 depicts the BNF of the annotation schema. In the figure, the optional elements are enclosed by [’s and]’s, and repetitive elements by {’s and }’s. To distinguish non-terminal elements from terminal ones, the former are enclosed in angle brackets, i.e., <’s and >’s. The * symbol outside the repeating element indicates zero or more repeating occurrences, and the + symbol for one or more repeating occurrences.

According to the schema, each annotation instance is assigned a unique identifier(a-id), instance URL (inst-url) and is created for a web object. The annotation instance consists of a concept part associated with the annotated web page and a relationship part associated with the links with other web pages. The concept part is self-explanatory. The relationship part consists of one or more relationships. Each


```

annotation ::= a-id inst-url <concept> {<relationship>}*
concept    ::= c-id c-name {<c-attrib>}*
c-attrib  ::= name value
relationship ::= r-id r-name [source-inst] [target-inst] {<r-attrib>}* {<linkchain>}+
r-attrib  ::= name value
linkchain  ::= l-id {<link> | <shared-path> | <use-shared-path>}+
link       ::= desturl label [sc-name]
shared-path ::= s-id {<link>}+
use-shared-path ::= path value

```

Figure 7. BNF for Annotation Schema

relationship consists of the id of the corresponding relationship entity in the ontology, the relationship name, the source and target annotation instance ids, their attribute values, and the hyperlink structures linking the source web page to the target web page if any.

We annotate the hyperlink structures by treating them as link chains. Each link chain is identified with link id (l-id) and includes a sequence of hyperlinks. At present, we identify each hyperlink in the chain by the destination web page URL, the hyperlink label and the source concept name if it is the another concept instance. There may be more than one link chains in each relationship if there exists multi-path link between source page and target page. In addition, shared-path link can also be identified within the link chain if it has been annotated by other link chains.

To facilitate sharing of annotation schema among different applications, the annotation schema is represented using the XML Schema language. The XML Schema for our annotation example is shown in Figure 8.

Based on the annotation schema, each annotated web page is associated with an annotation instance that includes a unique identifier, URI of the annotated web object, URI and other attributes of the assigned concept and annotation information of the relationships originating from that web page. Based on the assigned concept, the values of concept attributes can be indicated in the annotation. For each relationship originating from the web page, the relationship attributes, and the chain of hyperlink information leading to the target web page are included. The chainType complex type in the schema annotates the chain of hyperlinks as a <linkchain> element. Each <linkchain> can have <shared-path> if some of the links are common to other <linkchain>. The annotated shared-path can be used by using the <use-shared-path> tag. This way of identifying hyperlinks is simple and easy to implement.

5. Annotation Examples of Hyperlink Structures

The annotation instance for Dr Lim's homepage using our proposed annotation schema is shown in Figure 9 where the homepage of Dr Lim is assigned to Professor concept with three relationships, namely two *Supervise* relation-

ships to his PhD student Myo Myo and Danzhou, and one *Teach* relationship to the course he taught. The first *Supervise* relationship includes two hyperlink structures from Dr Lim's homepage to Myo Myo's homepage have been represented by two link chains. That means multi-path link exists between Dr Lim's homepage to Myo Myo's homepage and each link chain includes shared-path s01 and s02. Shared-path s02 is shared to Danzhou's homepage under the same *Supervise* relationship. The third hyperlink structure is a single-path link for a *Teach* relationship. It simply includes the hyperlink structure from Dr Lim's homepage to a Teaching label in the homepage, and later to the course (SC304) web page.

```

<annotation a-id = "&a111">
  <inst-url=http://www.cais.ntu.edu.sg:8000/~aseplim</inst-url>
  <concept c-id = "&c001">
    <c-name>Professor</c-name>
    <c-attrib name = "name" value = "Dr. Lim Ee Peng"/>
    <c-attrib name = "staffid" value = "12345"/>
    <c-attrib name = "email" value = "aseplim@ntu.edu.sg"/>
  </concept>
  <relationship r-id = "&r002">
    <r-name>Supervise</r-name>
    <source-inst=http://www.cais.ntu.edu.sg:8000/~aseplim</source-inst>
    <target-inst=http://www.cais.ntu.edu.sg:8000/~naing</target-inst>
    <r-attrib name = "start-date" value = "14-06-2001"/>
    <r-attrib name = "end-date" value = "14-06-2004"/>
    <linkchain l-id = "l01">
      <shared-path s-id = "s01">
        <link desturl = "http://www.cais.ntu.edu.sg:8000/~aseplim/#student" label = "StudentList"/>
      </shared-path>
      <link desturl = "http://www.cais.ntu.edu.sg:8000/~naing" label = "Myo Myo Naing"/>
    </linkchain>
    <linkchain l-id = "l02">
      <shared-path s-id = "s02">
        <link desturl = "http://pipe.cais.ntu.edu.sg:8000/~aseplim/#research" label = "Research"/>
        <link desturl = "http://www.cais.ntu.edu.sg:8000/~aseplim/geo.html"
          label = "Digital Libraries" sc-name = "Project"/>
      </shared-path>
      <link desturl = "http://www.cais.ntu.edu.sg:8000/~naing" label = "Myo Myo Naing"/>
    </linkchain>
  </relationship>
  <relationship r-id = "&r002">
    <r-name>Supervise</r-name>
    <r-attrib name = "start-date" value = "30-09-2000"/>
    <r-attrib name = "end-date" value = "30-09-2002"/>
    <linkchain l-id = "l03">
      <use-shared-path path = "s-id" value = "s02"/>
      <link desturl = "http://www.cais.ntu.edu.sg:8000/~danzhou" label = "Liu Danzhou"/>
    </linkchain>
  </relationship>
  <relationship r-id = "&r003">
    <r-name>Teach</r-name>
    <r-attrib name = "semester" value = "1"/>
    <r-attrib name = "year" value = "2001"/>
    <linkchain l-id = "l04">
      <link desturl = "http://www.cais.ntu.edu.sg:8000/~aseplim/#teaching" label = "Teaching"/>
      <link desturl = "http://www.cais.ntu.edu.sg:8000/~aseplim/sc304.htm" label = "SC304"/>
    </linkchain>
  </relationship>
</annotation>

```

Figure 9. Annotation instance of Dr. Lim's Homepage

Another annotation instance is given in Figure 10. Here, Myo Myo's homepage is annotated with the *PhDStudent* concept. The hyperlink structures corresponding to the *SupervisedBy* and *TakeCourse* relationships are annotated. The hyperlink structure for the *SupervisedBy* relationship includes only one direct hyperlink to Dr Lim's homepage. The hyperlink structure for the *TakeCourse* relationship consists of a chain from Myo Myo's homepage to the Subject label in the web page, and later to the course (H6405) web-

```

<?xml version="1.0"?>
<xsd:schema xmlns:xsd = "http://www.w3.org/2001/XMLSchema">
<xsd:element name="annotation" type="annotationType"/>
<xsd:complexType name="annotationType">
  <xsd:sequence>
    <xsd:element name="inst-uri" type="xsd:anyURI"/>
    <xsd:element name="concept" type="conceptType" maxOccurs="1"/>
    <xsd:element name="relationship" type="relationshipType" maxOccurs="unbounded"/>
  </xsd:sequence>
  <xsd:attribute name="a-id" type="xsd:anyURI" />
</xsd:complexType>
<xsd:complexType name="conceptType">
  <xsd:sequence>
    <xsd:element name="c-name" type="xsd:string"/>
    <xsd:element name="c-attrib" type="c-attribType" minOccurs="0" maxOccurs="unbounded"/>
  </xsd:sequence>
  <xsd:attribute name="c-id" type="xsd:anyURI"/>
</xsd:complexType>
<xsd:complexType name="c-attribType">
  <xsd:attribute name="name" type="xsd:string"/>
  <xsd:attribute name="value" type="xsd:string"/>
</xsd:complexType>
<xsd:complexType name="relationshipType">
  <xsd:sequence>
    <xsd:element name="r-name" type="xsd:string"/>
    <xsd:element name="source-inst" type="xsd:anyURI" minOccurs="0" maxOccurs="1"/>
    <xsd:element name="target-inst" type="xsd:anyURI" minOccurs="0" maxOccurs="1"/>
    <xsd:element name="r-attrib" type="r-attribType" minOccurs="0" maxOccurs="unbounded"/>
    <xsd:element name="linkchain" type="chainType" maxOccurs="unbounded"/>
  </xsd:sequence>
  <xsd:attribute name="r-id" type="xsd:anyURI"/>
</xsd:complexType>
<xsd:complexType name="r-attribType">
  <xsd:attribute name="name" type="xsd:string"/>
  <xsd:attribute name="value" type="xsd:string"/>
</xsd:complexType>
<xsd:complexType name="chainType" >
  <xsd:sequence>
    <xsd:element name="link" maxOccurs="unbounded">
      <xsd:complexType>
        <xsd:attribute name="desturl" type="xsd:anyURI"/>
        <xsd:attribute name="label" type="xsd:string"/>
        <xsd:attribute name="sc-name" type="xsd:string" default="none" />
      </xsd:complexType>
    </xsd:element>
    <xsd:element name="shared-path" type="shared-pathType" minOccurs="0" maxOccurs="unbounded"/>
    <xsd:element name="use-shared-path">
      <xsd:complexType>
        <xsd:attribute name="path" default="s-id"/>
        <xsd:attribute name="value" type="xsd:string"/>
      </xsd:complexType>
    </xsd:element>
  </xsd:sequence>
  <xsd:attribute name="l-id" type="xsd:string"/>
</xsd:complexType>
<xsd:complexType name="shared-pathType">
  <xsd:sequence>
    <xsd:element name="link" maxOccurs="unbounded">
      <xsd:complexType>
        <xsd:attribute name="desturl" type="xsd:anyURI"/>
        <xsd:attribute name="label" type="xsd:string"/>
        <xsd:attribute name="sc-name" type="xsd:string" default="none" />
      </xsd:complexType>
    </xsd:element>
  </xsd:sequence>
  <xsd:attribute name="s-id" type="xsd:string"/>
</xsd:complexType>
</xsd:schema>

```

Figure 8. Annotation Schema

page.

```
<annotation a-id = "&a125">
  <inst-url>http://www.cais.ntu.edu.sg:8000/~naing</inst-url>
  <concept c-id = "&c002">
    <c-name>PhDStudent</c-name>
    <c-attrb name = "name" value = "Myo Myo Naing"/>
    <c-attrb name = "matricnum" value = "G0001393F"/>
    <c-attrb name = "email" value = "p118008@ntu.edu.sg"/>
  </concept>
  <relationship r-id = "&r005">
    <r-name>SuperviseBy</r-name>
    <r-attrb name = "start-date" value = "14-06-2001"/>
    <r-attrb name = "end-date" value = "14-06-2004"/>
    <linkchain l-id = "107">
      <link desturl = "http://www.cais.ntu.edu.sg:8000/~aseplim" label = "Dr. Lim Ee Peng"/>
    </linkchain>
  </relationship>
  <relationship r-id = "&r006">
    <r-name>TakeCourse</r-name>
    <r-attrb name = "semester" value = "1"/>
    <r-attrb name = "year" value = "2002"/>
    <linkchain l-id = "108">
      <link desturl = "http://www.cais.ntu.edu.sg:8000/~naing/#subjects" label = "Subjects"/>
      <link desturl = "http://www.ntu.edu.sg/home/awkgng/H6405.htm" label = "Data Mining"/>
    </linkchain>
  </relationship>
</annotation>
```

Figure 10. Annotation of Myo Myo's Home-page

The above two examples illustrate that hyperlink structures can be annotated together with the web pages in an integrated manner. With these annotated hyperlink structures, it will much easier to process the relationships between web pages.

6. Related Work

In this section, we will give a brief overview on some existing ontology-based web annotation approaches and to compare our work with them.

Ontology-based web annotation has been studied in several research projects such as Ontobroker [10], SHOE [13], OntoAnnotate [20], CREAM [21] and the data annotation project [4]. None of them addresses the issues of annotating hyperlink structures.

To conduct OWA, several annotation languages have been introduced. They can be roughly divided into HTML-based annotation languages, XML-based annotation languages and RDF-based annotation languages.

Example of HTML-based annotation languages are Ontobroker [10] and SHOE [13]. Both of them do not support explicit annotation schema, they just use one single implicit annotation schema for representing the annotations. In addition, annotations need to be embedded in the original web pages. Text portions of the web pages can be annotated as concept's attribute values. The annotation of relationships between web pages only includes the URL of the related web pages but not the hyperlink structures.

XOML [7] is an example of XML-based annotation language, in which the annotation schema is defined by Document Type Definition(DTD). Original HTML document are

transformed into modified XML annotated documents including the relevant concept and relationship tags. Again, no hyperlink information is included in annotation of relationships.

A RDF-based annotation language has been adopted by OntoAnnotate [20] and Annotea [14]. Their annotation schemas are represented in RDF(Schema). Annotation instances are external to the annotated items and are stored separately in the RDF databases. OntoAnnotate allows only the web page annotation but in Annotea annotated items can be any web objects with URI or referenced by XPointer. Relationship assignment between different annotated items are allowed but do not include the hyperlink information.

7. Conclusion and Future Work

We have identified that the annotation of hyperlink structures is missing in the current state-of-the-art ontology-based web annotation (OWA). Such hyperlink structures can be useful not only in query processing but also in intelligent web browsing and visualization. We therefore propose an annotation schema to support such kind of annotations. Some examples are given to illustrate our proposed approach which is based on an extension of the Bremer's annotation approach. We also present a conceptual framework for ontology-based web annotation. This framework allows us to study OWA components independently and their inter-relationships.

Our research on hyperlink structure annotation approach is still in the early phase and several future work needs to be done to fully realise the goal of annotating hyperlink structures. They are:

- Integrating our annotation approach into other existing annotation languages.
- Developing automatic or semi-automatic techniques to annotate hyperlink structures together with other web objects.
- Building a system infrastructure to support the registration and sharing of annotation schemas.
- Current approach does not include how to handle the changes of hyperlink structures. We need to develop some extensive approaches to reflect the changes of hyperlink structures in the original web objects.

References

- [1] V. R. Benjamins, D. Fensel, and S. Decker. (KA)2: Building Ontologies for the Internet: A Mid Term Report, 1999.
- [2] I. Bratko. *PROLOG Programming for Artificial Intelligence*. Pearson Education Limited, third edition, August 2000.

- [3] T. Bray, J. Paoli, C. M. Sperberg, and E. Maler. Extensible Markup Language (XML) 1.0 (Second Edition), October 2000. <http://www.w3.org/TR/REC-xml>.
- [4] M. Bremer and M. Gertz. Web data indexing through external semantic-carrying annotations. In *Eleventh International Workshop on Research Issues in Data Engineering: Document Management for Data Intensive Business and Scientific Applications, Heidelberg, Germany, 1-2 April 2001*, pages 69–76, 2001.
- [5] D. Brickley and R. Guha. RDF Vocabulary Description Language 1.0: RDF Schema, April 2002. http://www.w3.org/TR/rdf-schema/#ch_appendix_rdfs.
- [6] D. Connolly, F. V. Harmelen, I. Horrocks, D. McGuinness, P. F. P-Schneider, and L. A. Stein. Annotated DAML+OIL Ontology Markup, December 2001. <http://www.w3.org/TR/daml+oil-walkthru/>.
- [7] B. Czejdo, J. Dinsmore, C. Hwang, R. Miller, and M. Rusinkiewicz. Automatic generation of ontology based annotations in xml and their use in retrieval systems. In *Proceedings of 1st International Conference on Web Information Systems Engineering*, Hong Kong, China, June 2000.
- [8] S. Decker, M. Erdmann, D. Fensel, and R. Studer. Ontobroker: Ontology based access to distributed and semi-structured information. In *Semantic Issues in Multimedia Systems Conference*, Boston, MA, USA, 1999.
- [9] D. Fensel, F. V. Harmelen, M. Klein, and H. Akkermans. Onto-knowledge: ontology-based tools for knowledge management. In *Proceedings of the eBusiness and eWork 2000 Conference*, Madrid, Spain, 2000.
- [10] D. Fensel, J. Angele, S. Decker, M. Erdmann, H.-P. Schnurr, S. Staab, R. Studer, and A. Witt. On2broker: Semantic-based access to information sources at the WWW. In *World Conference on the WWW Internet (WebNet99)*, pages 366–371, Honolulu, Hawaii, 1999.
- [11] D. Fensel, S. Decker, M. Erdmann, and R. Studer. Ontobroker: Or how to enable intelligent access to the WWW. April 1998. URL: <http://www.aifb.uni-karlsruhe.de/WBS/broker>.
- [12] N. Guarino. Formal Ontology and Information Systems. In *Proceedings of the 1st International Conference on Formal Ontologies in Information Systems*, pages 3–15, June 1998.
- [13] Jeffrey Douglas Heflin. *Towards the Semantic WEB: Knowledge Representation in a Dynamic, Distributed Environment*. PhD thesis, University of Maryland, 2001.
- [14] J. Kahan and M.-R. Koivunen. Annotea: an open rdf infrastructure for shared web annotations. In *The tenth international World Wide Web Conference on World Wide Web*, Hong Kong, May 2001.
- [15] P. Martin and P. Eklund. Knowledge retrieval and the world wide web, 2000.
- [16] E. Miller. An Introduction to the Resource Description Framework. *D-Lib Magazine*, May 1999.
- [17] S. Staab, J. Angele, S. Decker, M. Erdmann, A. Hotho, A. Maedche, H. P. Schnurr, R. Studer, and Y. Sure. *Semantic Community Web Portals*, volume 1-6, pages 473–491. WWW9/computer network, 2000.
- [18] S. Staab, M. Erdmann, A. Maedche, and S. Decker. An extensible approach for modeling Ontologies in RDF(S). In *The First ECDL'2000 Semantic Web Workshop*, Lisbon, Portugal, 2000.
- [19] S. Staab, A. Maedche, and S. Handschuh. An Annotation Framework for the Semantic Web. In *Proceedings of the First workshop on Multimedia Annotation*, Tokyo, Japan, 2001.
- [20] S. Staab, A. Mdche, and S. Handschuh. Creating Metadata for the Semantic Web—An Annotation Environment and the Human Factor.
- [21] S. H. Steffen. CREAM - Creating Relational Metadata with a Component-based, Ontology-driven annotation Framework. In *First International Conference on Knowledge Capture*, Victoria, BC, Canada, 2001.
- [22] E. van der Vlist. Using W3C XML Schema, October 2001. <http://www.xml.com/pub/a/2000/11/29/schemas/part1.html>.