社団法人　情報処理学会　研究報告
IPSJ SIG Technical Reports

2005－ＩＣＳ－139 (25)
2005／3／16

既存オントロジを用いたメタデータ設計支援

曺 喜蓮† 石田 亨†

†京都大学情報学研究科 〒606-8501 京都市左京区吉田本町
E-mail: †{cho, ishida}@kuis.kyoto-u.ac.jp

あらまし セマンティック Web の実現を妨げる問題に，メタデータ設計が困難な問題がある．この問題に対し，今までの研究は，一からの設計方法論を提案しているものが多く，具体的な設計困難を明確にしたものは少ない．他方で，セマンティック Web への移行が進むにつれ，メタデータの需要が高まりつつあるものの，メタデータ設計を容易にする設計支援法も少ない．そこで本研究は，具体的な設計困難として，1)クラスの認定問題，2)包摂関係の定義問題，3)主要プロパティの定義問題の三つの困難を明らかにする．そして，既存オントロジに定義されているメタデータを取り出し，1)クラスとプロパティの定義のどちらが多いか，2)クラス階層の定義は何か，3)複数のオントロジに表れる類似プロパティは何か，を提示することで，人間設計者のメタデータ設計を支援するツールを提案する．

キーワード セマンティック Web，メタデータ，オントロジ，設計，支援，ツール

Designing Metadata with Existing Ontologies

Heeryon CHO† and Toru ISHIDA†

† Graduate School of Informatics, Kyoto University Yoshida-Honmachi, Sakyo-ku, Kyoto, 606-8501 Japan
E-mail: †{cho, ishida}@kuis.kyoto-u.ac.jp

Abstract One obstacle that hinders the realization of the Semantic Web is high hurdle placed on metadata designing. Although metadata design difficulty is acknowledged by many, few researches have analyzed design difficulties in detail. As demand for metadata increases, support for easy designing of metadata is needed. To support metadata designing, we first clarify three metadata design difficulties: 1) difficulties relating to class and property uncertainty, 2) subsumption relation uncertainty, and 3) uncertainty in comprehensive metadata coverage. A design support tool, which retrieves, organizes, and displays relevant metadata defined in existing ontologies is proposed. The tool generates design materials that address the metadata design difficulties stated above, by providing the human designer with information about 1) which class or property definition is the majority definition, 2) what subsumption relation is defined, and 3) what similar properties are defined in multiple ontologies.

Keyword Semantic Web, metadata, ontology, design, support, tool

1. Introduction

Semantic Web is expected to become a foundation for automating complex processes by enabling machines to execute tasks like setting up schedules or searching for products that meet multiple criteria [1]. Execution of such tasks will be realized once great amount of metadata are annotated to data on the WWW. But prior to the annotation process, a set of metadata describing the target domain must be designed. Designing appropriate metadata is difficult, since interoperability must be considered. Moreover, it often requires domain knowledge and design expertise on the part of the human designer.

Although metadata design difficulty is acknowledged by many, few researches have clarified metadata design difficulties in detail. A number of existing researches dealing with design methodology have proposed various ontology construction methodologies [2], but these do not give detailed analysis of metadata design difficulties.

Furthermore, there are few researches dealing with metadata design support. Although the transition of the current web to the Semantic Web is in progress, and demand for metadata is on the rise, a limited number of metadata design support is available. Support tools that assist human designer's metadata design process are in great need.

The following two issues are addressed in this paper.

- The difficulties that arise during human designer's metadata design process are unclear. To date, there have been few researches that clarified specific difficulties arising in metadata design process.
- Metadata design support tool is limited. Few researches study the problem of supporting

metadata design process, and as a result, few related tools exist.

Here on, we investigate a method of utilizing existing ontologies as design material to support human designer's metadata design process. To do this, detailed design difficulties are first clarified, and a design support tool that retrieves, organizes and displays relevant metadata contained in existing ontologies is proposed.

In the next section, a metadata design experiment is described, and section 3 summarizes the metadata design difficulties observed in the experiment. Based on the experimental findings, section 4 proposes detailed design materials which can be generated from existing ontologies. Related researches are presented in section 5, and conclusion is discussed in section 6.

2. Metadata Design Experiment

Metadata is data about data. In particular, metadata is machine understandable information for the web [3]. A set of metadata that forms a semantic structure can be viewed as ontology. Ontology is an explicit specification of a conceptualization [4].

RDF (Resource Description Framework) [5] and OWL (Web Ontology Language) [6] are standard web-based frameworks for metadata and ontology respectively. RDF provides a framework for describing data about data. It describes metadata in triples, which consist of a subject, a predicate and an object. OWL is an ontology description language. Three sublanguages of OWL are available: OWL Lite, OWL DL, and OWL Full. RDF and OWL are W3C (World Wide Web Consortium) recommendations.

Metadata must assure some form of interoperability in order to be effectively used. Designing metadata with interoperability in mind is difficult, and it requires expertise such as domain knowledge and design skill on the part of the human designer. Such requirement places a hurdle on easy metadata creation.

To overcome this hurdle, design difficulties that arise during metadata design process must first be clarified. In this experiment, design difficulties that arise within a limited information environment are clarified through a trial design experiment.

2.1. Objective

The goal of this experiment is to clarify what information needs arise during metadata design process. Difficulties in the design process partly arise from the designer's lack of information concerning metadata, so by supplementing this lack of information with appropriate outside information, difficulties faced during the metadata design process can be mitigated. In this experiment we seek to identify what information is available at hand regardless of designer's knowledge and skill, and what information is not available, and hence is needed. If this "not available" information can be supplemented during the metadata design process, it could help the designer to make better design decisions.

A trial metadata design experiment was conducted by the author to identify the difficulties in metadata design process. A set of metadata for a target document was designed. Detailed information needs that arise during the design process were observed and documented. For the description language, OWL DL [7] was used to describe metadata.

2.2. Data and Method

1) Target Document

A user input form of an online social networking service called Orkut [8] was selected as the target document to be designed. Orkut allows registered users to input various information about themselves. The user information is received through six different user input pages: general page (receives name, gender, etc.), contact page (email, phone, etc.), photo page (photo), professional page (education, occupation, etc.), and personal page (personal traits like eye and hair color). In this experiment, general page was chosen as the target document. The author designed a set of metadata that describes each input/selection items contained in the target document. The entire general page items are listed in Table 2.1. Items on the left such as first name and last name are located in the first column of the table; detailed items on the right and HTML form control types like text boxes and radio buttons are located in the second column.

2) Metadata Design Step

With the target document at hand, the metadata design experiment was carried out in the following steps. Since the author did not have prior experience in metadata design, the following documents were referred: OWL Language Specification [6], a user's guide to creating ontology [9], and a wine ontology in OWL specification [10]. Other than these documents, no other resources, including any existing ontologies on the WWW, were referred. This restriction was placed deliberately on the author to highlight the information needs that arise during the design process. As for the actual design steps, six of the seven steps in [9] were performed.

Table 2.1: User input items in Orkut general page

Left Side Items	Right Side Items and/or HTML Form Control Types ◉ Radio button (choose one) [▾1] Menu (choose one) ☑ checkbox (multiple choice) [Textbox]
first name	[textbox]
last name	[textbox]
gender	◉ female, male
relationship status	[▾1] single, married, committed, open_marriage, open_relationship
birth day	[▾1] jan, feb, mar, apr, may, jun, jul, aug, sep, oct, nov, dec
	[▾1] 1 to 31
birth year	[▾1] 1920 to 1986
zip/ postal code	[textbox]
country	[▾1] United States, Canada, Afghanistan, ..., Zimbabwe (Total of 227 countries)
primary language	[▾1] English, Afrikaans, Ainu, Albanian, ..., Zulu (Total of 115 languages)
I'm interested in	☑ friends, activity partners, business networking, dating ([▾1] men&women, men, women)
children	[▾1] no answer, no, yes-at home full time, yes-at home part time, yes-not at home
ethnicity	[▾1] no answer, african american (black), asian, caucasian (white), east indian, hispanic/latino, middle eastern, native american, pacific islander, multi-ethnic, other
religion	[▾1] no answer, Agnostic, Atheist, Buddhist/Taoist, Christian/Catholic, Christian/LDS, Christian/Protestant, Christian/Other, Hindu, Jewish, Islam, Spiritual but not religious, Religious humanism, other
political view	[▾1] no answer, right-conservative, very right-conservative, centrist, left-liberal, very left-liberal, libertarian, very libertarian, authoritarian, very authoritarian, depends, not political
sense of humor	☑ campy/cheesy, goofy/slapstick, dry/sarcastic, obscure, clever/quick witted, raunchy, friendly
sexual orientation	[▾1] no answer, straight, gay, bisexual, bi-curious
fashion	☑ alternative, casual, classic, contemporary, designer, minimal, natural, outdoorsy, smart, trendy, urban
smoking	[▾1] no answer, no, socially, occasionally, regularly, heavily, trying to quit, quit
drinking	[▾1] no answer, no, socially, occasionally, regularly, heavily
pets	[▾1] no answer, i love my pet(s), i like them at the zoos, I like pet(s), i don't like pets
living	☑ alone, with kid(s), with roommate(s), with parents, with partner, friends visit often, with pet(s), party every night
home town	[textbox] (in) [▾1] no answer, Alabama, Alaska, ..., Wyoming (50 States in the U.S.)
web page	[textbox]
describe yourself	[textbox]

Note that the second step, which recommends ontology reuse, was skipped because reusing other ontologies will violate the restriction placed on the experiment. Below are steps 1 through 7 given in [9]. Actual design steps that the author performed are explained in the steps 1e through 7e.

```
Step 1: Determine the domain and scope of the ontology.
  -> Step 1e: Set the domain of the ontology to social
  networking service.
Step 2: Consider reusing existing ontologies.
  -> Step 2e: Skip.
Step 3: Enumerate important terms in the ontology.
  -> Step 3e: Let important terms be equal to the
  user input items given by the Orkut general page.
  Enumerate the input items as important terms.
Step 4: Define the classes and the class hierarchy.
  -> Step 4e: If items are defined as classes, define
  the class hierarchy.
Step 5: Define the properties of classes-slots.
  -> Step 5e: Define properties of the classes.
Step 6: Define the facets of the slots.
  -> Step 6e: Define cardinality, datatype, domain,
  range of the properties.
Step 7: Create instances.
  -> Step 7e: Create instances.
```

3. Metadata Design Difficulties

After performing design steps 1e through 7e, a sample ontology containing metadata of the target document was created. The syntax of the resulting ontology was checked using an OWL Validator [11]. The following difficulties were observed during the design process.

- Determining whether an item should be designed as a class or a property is difficult.
- Determining what subsumption relation holds between two classes is difficult.
- Being confident that most major properties of the target are designed is difficult.

Each of the difficulties can be regarded as information needs problem. In the next section, each of the problems is explained in detail.

3.1. Class/Property Uncertainty

Determining whether an item should be designed as a class or simply as a property was difficult. Classes provide an abstraction mechanism for grouping resources with similar characteristics. A class description describes an OWL class, either by a class name or by specifying the class extension of an unnamed anonymous class. During the trial design process, it was difficult to determine whether items like first name, last name, gender, and web page should be modeled as classes.

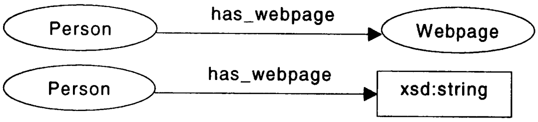

Figure 3.1: Webpage item designed as class (top) and as property (bottom)

Figure 3.1 shows two different ways of designing the webpage item. The top graph shows webpage item designed as Webpage Class and linked to Person Class through has_webpage property, and the bottom graph shows webpage defined as property without any classes. Similar design dilemma exists in the case of the gender item and name item. These variations to design can raise uncertainty in the designer since it is difficult to know which might be the better design. If the author had a chance to refer to existing ontologies for actual examples of name, gender and webpage metadata definitions, some design hints might have been obtained.

3.2. Subsumption Relation Uncertainty

Determining what subsumption relation holds between two classes was difficult. Subsumption relation, which is the basis of a taxonomy, is an extremely useful tool for imparting structure on an ontology [12]. Subsumtion relation can be defined by using "rdfs:subClassOf" element. Determining whether one class subsumes another, however, is difficult to determine, and [12] also points out that subsumption relation is often misused. For example in the experiment, difficulty was observed in designing the class hierarchy of Female and Male Class.

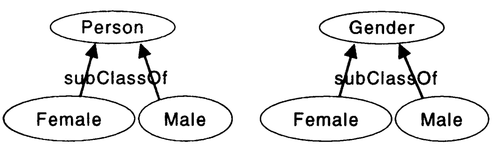

Figure 3.2: Person subsuming Female and Male (left) and Gender subsuming Female and Male (right)

Figure 3.2 shows two design possibilities to define class hierarchy for Male and Female Class. The left ovals show Person Class subsuming Female and Male Class. The right ovals show Gender Class subsuming Female and Male Class. Both, one or none may be correct. If the author were able to refer to other ontologies for class hierarchy definitions, it might have helped the author to make better design decision.

3.3. Uncertainty of Comprehensive Metadata Coverage

The target document used for the experiment was general page of the Orkut user input form. Apart from the general page, there are five other pages that deal with person related information. Although these pages can be easily obtained in the Orkut website, it may not always be the case that relevant documents or resources are easily obtained in other metadata design situations. In such cases, the designer may not have enough relevant information to design necessary metadata. As a result, the designer may end up designing only limited kinds of metadata without covering all basic metadata. This is the comprehensive metadata coverage problem.

Supposedly, if other documents in Orkut website were not acquired, basic properties for describing Person Class will be defined as that shown in Figure 3.3.

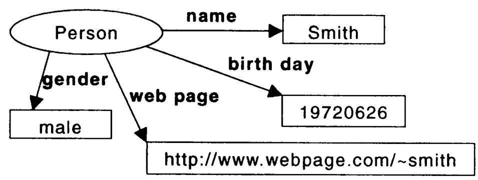

Figure 3.3: Basic properties for describing "Person"

4. Ontology As Design Material

The difficulties observed during the trial metadata design experiment can be alleviated by referring to existing ontologies on the WWW. If somewhere on the WWW, there exists metadata that closely relates to the metadata the human designer has difficulty is designing, by referring to them, the designer can know how others have designed similar metadata. These actual metadata examples may aid the human designer to make better design decisions. In this section, detailed design materials will be proposed for the three design problems presented in section 3.

4.1. Majority Rule as Hints

The uncertainty of whether to design a target item as class or property can be reduced by comparing existing class and property descriptions of that item in other ontologies. By looking at the multiple class and property examples at one glance, the human designer can know which definition is the majority case. Table 4.1 shows design material generated for the webpage item. In the first column, the metadata name is provided. Here, "homepage" is provided. In the second column, the type of metadata, whether it is a class or a property, is indicated. The third column shows ontology URL where the metadata is defined. The support tool that we have implemented can generate this table automatically from existing ontologies with keyword input from the designer.

Table 4.1: Majority rules that homepage is a property

Metadata	Class Property	Ontology URL
Personal Homepage	class	http://www.atl.external.lmco.com/projects/ontology/ontologies/comsci/csA.rdf
Personal Homepage	class	http://www.cs.umd.edu/projects/plus/DAML/onts/docmnt1.0.daml
Homepage	class	http://www.cs.man.ac.uk/~lopatena/cerif/cerif.daml
Homepage	class	http://www.cs.man.ac.uk/~lopatena/cerif.daml
homepage	Datatype property	http://daml.umbc.edu/ontologies/webofbelief/Foaf.owl
homepage	Datatype property	http://daml.umbc.edu/ontologies/webofbelief/0.81/foaf.owl
homepage	Datatype property	http://daml.umbc.edu/ontologies/webofbelief/0.8/foaf-lite.owl
School Homepage	Property	http://www.cs.umd.edu/~hendler/2003/MindPeople4-30.rdf
workInfo Homepage	property	http://bblfish.net/work/atom-owl/2004-08-12/Atom.old2.owl
School Homepage	property	http://simile.mit.edu/repository/ontologies/official/foaf.owl
School Homepage	property	http://gemini.doosh.net/foaf/index.rdf
homepage	property	http://ilrt.org/discovery/2001/05/ical/index.rdf
course_ Homepage	property	http://www.arches.uga.edu/~vstaub/GlobalInfoSys/project/ontology/Could_have_been.rdfs
homepage	property	http://www.aifb.uni-karlsruhe.de/WBS/meh/mapping/data/swrc1a.rdf
homepage	property	http://lithwww.epfl.ch/teaching/docmul/seance17/people.rdfs
homepage	property	http://www.swed.org.uk/swed/data/swed/organisation_v1.2.owl
homepage	property	http://lithwww.epfl.ch/teaching/docmul/seance17/labpeople.rdfs
School Homepage	Object property	http://www.cs.vu.nl/~pmika/foaf/foaf.owl
School Homepage	Object property	http://svn.mindswap.org/pychinko/allogtests/mindswapRealized.rdf

4.2. Seeing How Others Define Subsumption Relations

The sample class subsumption problem observed in section 4 was that of designing superclass of Male and Female Class. Table 4.2 shows the output of the design material generated upon keyword input "male". The first column shows the subclass metadata. In this case, Male and Female Class are the subclass metadata. The second column shows the superclass metadata. And the third column shows the ontology URLs where the metadata are defined. Based on the result presented in the table, it may be interpreted that many ontologies define Animal Class as the superclass of Male and Female Class.

Notice that a new design direction, which was unthought-of, was presented by the tool through reusing existing ontologies. The initial design possibilities were Person Class and Gender Class. The support tool can also generate this table from keyword input.

Table 4.2: Male Class is mostly defined as subclass of Animal Class

Subclass	Super class	Ontology URL
Male	Individual	http://www.tt.cs.titech.ac.jp/~fukatani/kadai/yasuda/genealogy.owl
Male	Animal	http://www.daml.org/validator/examples/ont3.daml
female	animal	http://www.cs.man.ac.uk/~horrocks/ESSLLI2003/Ontologies/sane_cows.daml
Male	Animal	http://www.atl.external.lmco.com/projects/ontology/ontologies/animals/animalsB.owl
Male	Animal	http://www.atl.external.lmco.com/projects/ontology/ontologies/animals/animalsA.owl
Female	Animal	http://www.daml.org/2002/06/webont/owl-ex.owl
Male	Animal	http://www.w3.org/2000/10/swap/test/dpo/daml+oil-ex.daml
Female	Animal	http://cvs.sourceforge.net/viewcvs.py/jena/jena2/testing/ontology/daml/daml_oil_2001_03/daml+oil-ex.daml?rev=1.2
Male	Animal	http://www.srdc.metu.edu.tr/~yildiray/example.daml
Female	Animal	http://www.daml.org/2000/12/daml+oil-ex.daml
Male	Animal	http://www.daml.org/validator/examples/ont1.daml
Female	Animal	http://www.cs.man.ac.uk/~horrocks/DAML+OIL/Datatypes/daml+oil+dt-ex.daml
Female	Animal	http://www.cs.vu.nl/~mcaklein/onto/example11.daml
Male	Animal	http://www.daml.org/2000/10/daml-ex.daml
Female	Sex	http://www.csc.fi/kielipankki/puhe/schemas/official/recording.rdfs
Male Person	Person	http://www.cs.umd.edu/~evren/cmsc828y/hw1/ontology.daml

4.3. Comparing Properties Through Table

The last design difficulty discussed in section 3 is the problem of comprehensive property coverage. During the design process, the metadata designer may be unsure if important or major properties of the target subject have all been defined.

One way to supplement this lack of knowledge is to provide sorted properties of the target subject defined in multiple ontologies. The target subject in this case is defined as class. By retrieving properties of that class from multiple ontologies and sorting them according to similar properties, the metadata designer can know what properties are frequently defined. Based on this, the designer may interpret that these frequently defined properties constitute the basic property of that class.

5. Related Research

Existing research that closely relates to this research is a research on ontology development toolkit called SWOOPed [13]. SWOOPed also focuses on the problem of reusing existing ontologies, and as an answer to this, provides an easy ontology construction environment to enable human designer to directly reuse metadata in existing ontologies. This research differs with SWOOPed in that this research aims at providing design material to the human designer rather than providing an editing environment. Moreover, organizing relevant metadata contained in multiple ontologies is the challenge that this research faces.

Another research that takes up the ontology reengineering problem is [14]. Here, a rough ontology generated from database schema is grounded to a foundational ontology to guarantee its consistency. Our research, however, does not go as far as to deal with the matter of ontology consistency. It will remain focused on specific metadata handling.

Researches that propose various ontology construction methodologies [2] are also related to this research in that they deal with the designing of ontologies. However, their focus is mainly on designing ontology from scratch, rather than referring existing ontologies, which this research focuses on.

6. Conclusion

We clarified metadata design difficulties and proposed a support tool that utilizes metadata retrieved from existing ontologies. The contributions of this research are as follows.

1. Three metadata design difficulties were identified.
2. A design support tool, which organizes and displays relevant metadata defined in existing ontologies, was proposed.

The three metadata difficulties that were identified are as follows.

- Determining whether an item should be designed as class or property is difficult.
- Determining what subsumption relation holds between two classes is difficult.
- Being confident that most major properties of the target class are defined is difficult.

The support tool displays three kinds of metadata information that corresponds to the three metadata design difficulties: The first kind shows which class or property definition is the majority definition. The second kind shows what subsumption relation is defined in existing ontologies. And the third kind shows what similar properties are defined in multiple ontologies. All of the three metadata information is displayed in table format. Since various metadata defined in existing ontologies reflect other designers' design decisions, design material generated from existing ontologies provide integrated result of multiple designers' design decisions. Such metadata information may provide hints or ideas about possible design directions.

The quality of the metadata, designed with the help of the support tool, should be evaluated in the future.

References

[1] T. Berners-Lee, J. Hendler, and O. Lassila, "The Semantic Web," Scientific American, vol.284, no.5, pp. 34-43, May 2001.

[2] M. Fernadez-Lopez and A. Gomez-Perez, "Overview and Analysis of Methodologies for Building Ontologies, The Knowledge Engineering Review, vol.17, no.2, pp.129-156, June 2002.

[3] World Wide Web Consortium (W3C), Metadata and Resource Description, http://www.w3.org/Metadata/, 2001.

[4] T. Gruber, "A Translation Approach to Portable Ontology Specifications," Knowledge Acquisition, vol.5, no.2, pp. 199-220, June 1993.

[5] World Wide Web Consortium (W3C), Resource Description Framework (RDF), http://www.w3.org/RDF/.

[6] World Wide Web Consortuim (W3C), Web Ontology Language (OWL), http://www.w3.org/2004/OWL/.

[7] OWL DL, http://www.w3.org/TR/2004/REC-owl-guide-20040210/#term_OWLDL

[8] Orkut, http://www.orkut.com/.

[9] F. Noy and D. L. McGuinness, "Ontology Development 101: A Guide to Creating Your First Ontology," Stanford Knowledge Systems Laboratory Technical Report KSL-01-05, March 2001.

[10] Wine Ontology, http://www.w3.org/TR/owl-guide/wine.rdf.

[11] WonderWeb OWL Ontology Validator, http://phoebus.cs.man.ac.uk:9999/OWL/Validator.

[12] N. Guarino and C. Welty, "Evaluating Ontological Decisions with ONTOCLEAN," Comm. of the ACM, vol.45, no.2, pp.61-65, Feb. 2002.

[13] A. Kalyanpur, N. Hashmi, J. Golbeck, and B. Parsia, "Lifecycle of a Casual Web Ontology Development Process," Workshop on Application Design, Development and Implementation Issues in the Semantic Web, New York, USA, May 2004.

[14] S. Bechhofer et al, "Tackling the Ontology Acquisition Bottleneck: An Experiement in Ontology Re-Engineering," Oct. 2003.