A Semantic Faceted Search with Rule-based Inference

Yung-Hsin Wang and Pin-Siou Jhuo

Abstract—**Semantic Search has become an active research of Semantic Web in recent years. The classification methodology plays a pretty critical role in the beginning of search process to disambiguate irrelevant information. However, the applications related to Folksonomy suffer from many obstacles. This study attempts to eliminate the problems resulted from Folksonomy using existing semantic technology. We also focus on how to effectively integrate heterogeneous ontologies over the Internet to acquire the integrity of domain knowledge. A faceted logic layer is abstracted in order to strengthen category framework and organize existing available ontologies according to a series of steps based on the methodology of faceted classification and ontology construction. The result showed that our approach can facilitate the integration of inconsistent or even heterogeneous ontologies. This paper also generalizes the principles of picking appropriate facets with which our facet browser completely complies so that better semantic search result can be obtained.**

Index Terms—**Faceted Classification, Folksonomy, Ontology, Semantic Search, Semantic Web.**

I. INTRODUCTION

The development of Semantic Search was intent on solving the existing problems of traditional information retrieval by using Semantic Web technology [1]. Guha *et al.* [2] assumed that Semantic Search is a Research Search which discovers objects related to keywords rather than particular documents. In general, this terminology is referred as long as the semantics involve in three phases of search process: query construction, search algorithm and result presentation [3].

In the aspect of query construction, keywords, natural language and facets [4]–[10] are usually used by human to express what information they need. In the search algorithm place, semantic matching and ranking algorithms [6], [11] are responsible to pick objects and to sort these candidate objects respectively by knowledge meaning. Regarding the result presentation, semantic annotations [12], [13] are usually utilized to visualize the knowledge objects concerning user requirements in the mankind understandable form.

In recent years, many researches [6]–[10], [13]–[17] have worked on how to apply the methodology of classification to semantic search for acquiring the optimization of search results based on knowledge ontology. However, one popular social classification method—Folksonomy has suffered from many obstacles, such as the lack of organization and precision [16], even though its contribution towards Web 2.0 is undoubted. Therefore, some compromises [15], [17] between Folksonomy and Taxonomy have emerged for solving the problems. Even so, the issues regarding precision still exist due to the open environment. On the other hand, the policies [13], [14] combined Folksonomy and Ontology appeared for the same purpose. Nevertheless, the potential conflicts between Ontology and Folksonomy as well as inconsistent ontologies rose dramatically because of no synonym control.

There are numerous heterogeneous ontologies existing over the Internet, yet how to effectively combine them rather than redesigning again and again for the integrity of domain knowledge is also our another concern. Since Faceted Classification is a methodology appropriate for managing organizational knowledge [18], we came up an idea if it is possible to organize existing available ontologies by abstracting a faceted logic layer in fully flexible notation.

The objective of this study is to construct an inference-based semantic faceted search browser from existing available ontologies while solving the problem of precision and organization resulted from Folksonomy. Firstly, the category framework is strengthened by employing Description Logic to thoroughly control hierarchical or parts-whole relationships between categories. Next, a faceted logic layer is abstracted for establishing faceted classification according to a series of steps we proposed based on the combination of Vickery's method [19] and Ontology Development 101 [20]. Rule-enhanced categories, which produced at the last stage, are reused to integrate existing RDF documents over the Internet. The separation of semantic rules from distributed programming logic would facilitate the integration of inconsistent ontologies annotated by various users over the Internet or even heterogeneous ontologies represented distinct objects or projects.

In the rest of this paper, Section II will explore several classification methodologies including their potential drawbacks. Section III describes our approach to solving the precision problem and integrating inconsistent heterogeneous ontologies. In order to prove the feasibility of our method, Section IV then experiments on an inferable semantic faceted browser based on an example case. Finally, Section V concludes this paper with future work suggestions.

II. CLASSIFICATION METHODOLOGIES

Four strategies of classification applied in Semantic Search: Taxonomy, Folksonomy, compromises between Taxonomy and Folksonomy, and Facets (see Fig. 1) along with their shortages are described in what follows.

Manuscript received September 1, 2008. This work was supported in part by Tatung University, Taipei, Taiwan, under Grant B96-N01-044.

Y.-H. Wang is with the Tatung University, Taipei, Taiwan (phone: 886-2-25925252; fax: 886-2-25853966; e-mail: ywang@ ttu.edu.tw).

P.-S. Jhuo was graduated from Tatung University. She is now studying in the University of Southampton, UK (e-mail: psjhuo@gmail.com).

Taxonomy, a science of classification, organizes information in a ranked hierarchical structure consisting of controlled vocabularies defined by experts, as shown in Fig. 1(a). For instance, the most familiar application known as Linnaean Taxonomy classified organisms into kingdom, phylum, class, order, family, genus, and species.

Folksonomy is a social tagging system which classifies context by user contributed tags. Due to Semantic Web maturation and Social Web growth it has become the popular classification methodology on the Web since 2004. The earlier Web 2.0 applications, for example, Flicker [21], YouTube [22], 43Things [23] and Wikipedia [17], allow photos, videos or articles to be annotated and browsed in user defined keywords instead of controlled vocabularies. However, Kroski [16] argues that this mechanism lacks of organization and precision, even though it supports query by observing user behavior. In addition, each Folksonomy's tag is unconnected with each other, as shown in Fig. 1(b).

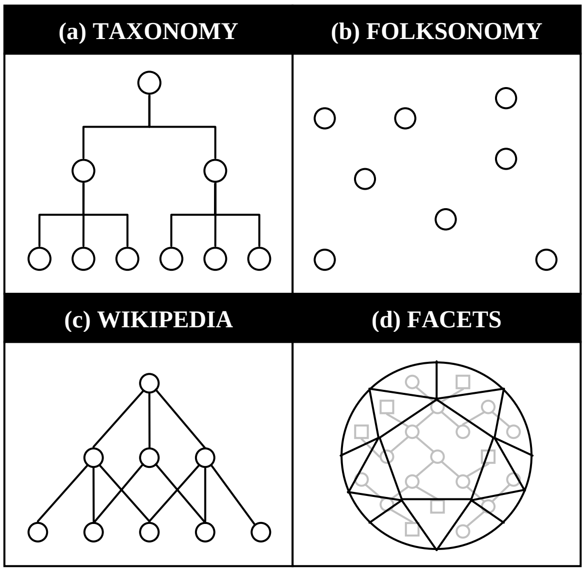
In order to overcome the problems of Folksonomy, the compromises between Taxonomy and Folksonomy surfaced, e.g., Wikipedia organized a hierarchy of categories defined by the article authors, yet it is not strict hierarchy or tree structure in that it allows one article to belong to multiple categories (see Fig. 1(c)). Likewise, Taxonomy-Directed-Folksonomy [15] relies on the user interfaces to suggest tags from a formal Taxonomy, but allows users to use their own tags. Even so, the precision problem still exist due to the open environment wherein users who cannot understand the architecture of huge categories will likely pick frequent one, random one, or even create new one. Wikipedia had no choice but to restrict the growth of categories for preventing explosion.

Recently, there are policies combined Ontology and Folksonomy. Gruber [14] considered Ontology as a methodology for representing materials while Folksonomy as an aggregation for supporting discovery, each one does its duty. Fig. 2 illustrates the operations of Semantic Wikipedia [13]: (1) significant information is annotated by users in semantic extension based on Wiki markup [24] and then forms ontologies after aggregating; (2) the builders identify the hierarchical relationships regarding articles as categories; and (3) the users classify articles by tags with keywords known as Folksonomy. However, the potential conflicts between them emerged, as Ontology is strict structure in the closed world, whereas Folksonomy is loose aggregation in the open world; the new category is isolated from these class hierarchies, providing that users annotate in tags which are not defined ahead by builders. Besides, the lack of synonym control leads to the inconsistency of information.

This study proposed a method to strengthen the Category frameworks of Semantic Wikipedia by using an existing semantic Description Logic to express categories with (a) hierarchical, and (b) parts-whole relationships (as depicted in Fig. 2).

Faceted Classification does not assign fixed slots to certain subjects in specific sequence, yet uses clearly defined, mutually exclusive, and collectively exhaustive aspects, properties, or characteristics of a class or subject known as facets, as shown in Fig. 1(d). In fact, this classification method is appropriate for managing organizational knowledge [18], which can reach available objects by adding a new facet at any time even if the object's name is unknown.

Vickery [19] stated a four-step scheme for making a faceted classification: (1) analyze homogeneous and mutually exclusive facets; (2) assign an order to these facets for constructing compound subject headings; (3) find the fit schedules with a fully flexible notation; and (4) plan to use the faceted scheme by a deep survey. Additionally, Denton [25] expanded Vickery's four-step scheme to form a seven-step scheme by adding start and finish phases.



(a) TAXONOMY

(b) FOLKSONOMY

(c) WIKIPEDIA

(d) FACETS

Fig. 1. The Strategies of Classification.

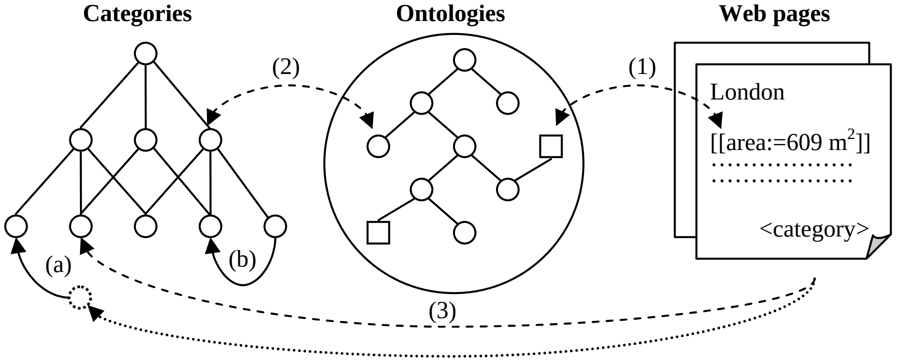


Fig. 2. The Operations of Semantic Wikipedia.

III. THE APPROACH TO ESTABLISHING RULES

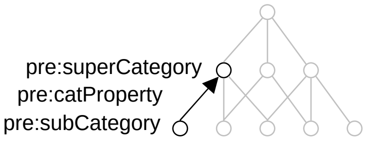
A. *Category rules*

This section illustrates how to use Description Logic to thoroughly control (a) hierarchical or (b) parts-whole relationships between categories (as illustrated in Fig. 2) in order to strengthen category framework to solve the precision and organization problem resulted from Folksonomy. We analyze two possible situations as follows, yet the realization of an example case will be presented in the next section. Note that *pre:catProperty* represents the relationships between categories in this section, whereas *pre:property* describes the connection among objects in the next section.

Case 1. Categories with hierarchical relationship. The principal deficiency of Folksonomy is that tags with hierarchical relationship are unconnected with each other. This case illustrates how to integrate these tags with hierarchical relationship into category framework without changing the structure of ontology. As shown in Fig. 3 and 4, subclasses are regarded as premises and superclasses as conclusions. It implies that these instances are the member of *pre:superCategory*, as inference engine matches triples with the *pre:subCategory*'s instances. Therefore, this type of rules could make sure promoting the organization of Folksonomy.

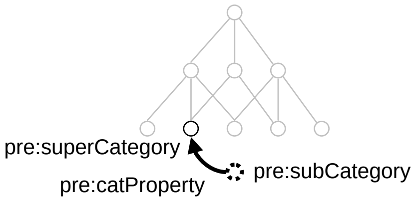
Case 2. Categories with parts-whole relationship. Another deficiency of Folksonomy is that there are numerous similar meaning tags which point at same group instances. This case illustrates how to integrate these instances with similar meaning tags into category framework without changing the structure of Ontology. As shown in Fig. 5, classes with fewer instances are considered as premises and classes with more instances as conclusions. More precisely, the instances of *pre:moreInstanceCategory* consist of *pre:lessInstanceCategory*'s instances. Consequently, this kind of rules offers a synonym control mechanism to improve the precision of Folksonomy.

Conceptually, Case 1 is similar to Taxonomy that organizes domain knowledge from several ranked and hierarchical concepts, whereas Case 2 summarizes a concept from similar concepts. More example rules will be demonstrated later.



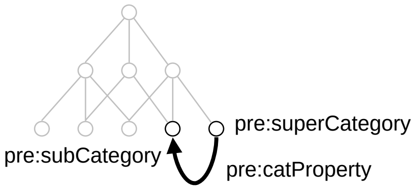
[H: (?I pre:catProperty pre:subCategory) ->
(?I pre:catProperty pre:superCategory)]

Fig. 3. The Categories with Hierarchical Relationship.



[H: (?I pre:catProperty pre:newSubCategory) ->
(?I pre:catProperty pre:superCategory)]

Fig. 4. The New Categories with Hierarchical Relationship.



[U: (?I pre:catProperty pre:lessInstanceCategory) ->
(?I pre:catProperty pre:moreInstanceCategory)]

Fig. 5. Categories with Parts-whole Relationship.

B. *Facet rules*

A logic layer for establishing Faceted Classification is abstracted by reusing rule-enhanced categories described above and collecting available ontologies over the Internet. The following construction steps are based on the combined method of Vickery's four-step process [19] and Ontology Development 101 [20]. The idea is to take advantage of the Ontology concept of the latter to supplement the former.

Step 1. Determine the domain. In general, a domain is a sphere of knowledge identified by a name; thus we consider using a noun to describe it. What domain is in our case study should be figured out before we can discover available and suitable ontologies over the Internet. In this study, we assume a term *Domain* as the example case's domain.

Step 2. Rephrase existing facets. Facets are common aspects, properties, or characteristics regarding an observing domain. Picking suitable ones from defined faceted classification and further rephrasing them as what we want is undoubtedly a better way to avoid mistakes. It is significant to note that these facets should be mutually exclusive and the quantity of them should be appropriate to identify domain. In this step, we suppose that *Domain* consist of three facets *Facet1*, *Facet2* and *Facet3*. By the way, the facets produced in this stage are the content of elements <*facet*> in XFML [26] (eXchangeable Faceted Metadata Language).

Step 3. List instances. Instances indicate person, events or things existing in real world. We extract directly instances from every ontology hierarchical class in order to hold the possibility of various hierarchical instances. In this step, we assume variable *?I* as the instances of domain which can be observed by facets. By the way, the instances exposed in this stage are the content of elements <*topic*> in XFML.

Step 4. Discover features. There are two sources of facets, categories and relation. We propose to construct a layer of features after facets, because the classes represented by distinct predicates should be generalized for synonym control. This is to cater for our promise to integrating existing heterogeneous ontologies, i.e., those dissimilar objects described in the same projects or located in different projects.

For categories, we should pick suitable candidates from higher level classes, which contain more instances. Thus, we assume *pre:superCategory*, the parent of *pre:subCategory*, as the candidate. On the other hand, the fitting properties should be gathered from available ontologies. Because the same type properties originated from different classes are significant viewpoints to observe instances, we identify two possible conditions: (1) the distinct classes represented by the same predicate, such as *pre:class1* and *pre:class2* described by *pre:property1*, and (2) the distinct classes represented by the similar predicate, *pre:class1* and *pre:class3* described by *pre:property1* and *pre:property2* respectively for example. Afterwards, we must generalize a feature *Feature1* from the related candidate category (*pre:superCategory*) and properties (*pre:class1*, *pre:class2* and *pre:class3*).

Step 5. Create rules. The rule is an efficient and flexible pattern matching algorithm to be employed to replace XFML for the inference-based semantic search. According to Jena's rule syntax [27], three types of rules domain-to-facets, facets-to-features, and features-to-instances are composed from four objects: domain, facets, features, and instances. In this step, facts that include candidate categories and properties are regarded as premises, whereas triples that contain features are regarded as conclusions. It is significant to note that all properties placed in conclusions are implicit relationships between instances and the other three objects. In other words, these predicates were undefined in advance, such as *wiki:domain*, *wiki:facet* and *wiki:feature*.

IV. IMPLEMENTATION AND DEMONSTRATION

A. *Case description*

In order to present and verify our method proposed in this study, we implemented a case project wherein the investors who would like to preliminarily understand what countries are suitable for them to build factories can use the faceted browser established. Before founding faceted classification, we have to survey the materials regarding requirement. From the perspective of investors, the geographic location, joined political entities, and labor force affect their determination. The details of each composing factor are shown in Table I.

B. *Implementation flow*

To begin with, the available OWL files of Semantic Wikipedia, which was employed as ontology resources, were exported and further analyzed. Next, system components were designed by employing Jena. Eventually, the method and steps we developed were utilized to establish Semantic Faceted Browser from existing ontologies (see Fig. 6).

C. *Analyzing available ontologies*

Völkel *et al.* [13] considered that Wikipedia's content is barely machine-interpretable and built Semantic Wikipedia accordingly. We employ the available RDF documents from Semantic Wikipedia for our analysis. Because Semantic Wikipedia changes exported facts by different subjects or categories, we must choose ontologies of several articles and categories. In order to completely understand entire domain knowledge, we adopted the reverse engineering of Ontology Development 101 to analyze Semantic Wikipedia ontologies.

Step 1: Determine the domain → What is the domain? The observing subject *Country* was regarded as our domain.

Step 2: Reuse existing ontologies → List reused Ontologies. In this step, existing ontologies, such as *SWiVT* (Semantic Wiki Vocabulary and Terminology) [28], were discovered from prefixes as shown in Table II, because XML Namespaces use URN (Universal Resource Names) to represent controlled vocabularies in specific domain. However, *wiki* and *property* belong to the prefixes of the Semantic Wikipedia ontology.

Step 3: Define classes and hierarchies → Organize category frameworks. The category framework is organized in Fig. 7 by listing classes related to the domain and tracking their hierarchies. As can be seen, a significant category *wiki:C3ACountry_in_Europe* related to *Country* was exhibited and divided into two classes (subcategories).

Table I. The Preliminary Survey.

Observing Subject	Country		
User Perspectives	Region	Organization	Population (x)
Composing Factors	Africa	EU	x < 1m
	Antarctica	Mercosur	1m < x < 100m
	Asia	NATO	x > 100m
	Australia	NU	
	Europe	Schengen	
	North_America		
	South_America		

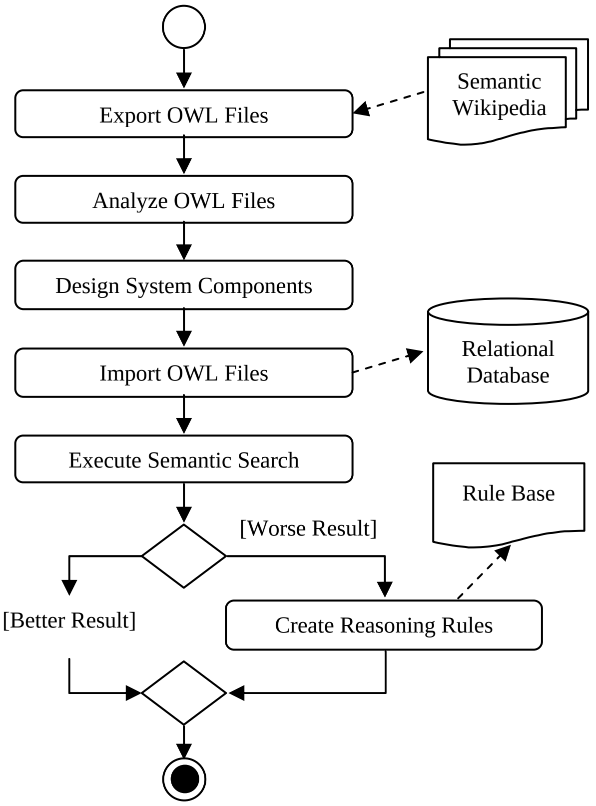


Fig. 6. The Implementation Flow.

Table II. The Prefixes of Semantic Wikipedia.

PREFIX	URI
swivt	http://semantic-mediawiki.org/swivt/1.0#
wiki	http://www.semanticweb.org/id/
property	http://www.semanticweb.org/id/Property-3A

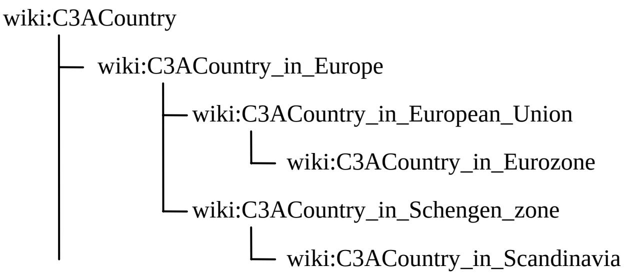


Fig. 7. The Category Framework of Semantic Wikipedia.

Step 4: Define properties → Draw knowledge maps. The knowledge map was drawn in Fig. 8 by listing properties concerning user perspectives and *rdf:type* which bind the relation between instances and categories for better understanding the domain knowledge.

Step 5: Create instances→List instances. Instances only belonged to certain classes were listed in the form of triple for later experimental design use, as shown in Table III.

D. System architecture

Fig. 9 illustrates the system architecture of this study which includes three layers: Semantics, JavaBeans and WebPages. The first two layers are realized in Jena. Web pages are only used to present the functions of Java Beans.

We store RDF documents in the TripleStore relational database. One reason is that it provides off-the-shelf solution, scalability, formulated query, efficiency, optimization, and organization [29], [30]. Note that Semantic Wikipedia also adopted MySQL relational database. In the aspect of implementation, Jena produces a table to store statements (triples) consisting of fields of *Subject*, *Property*, and *Object*, as well as *Graph_Id*, which identifies distinct domains.

Since ontologies existing alone are merely a data structure that describes the objects on the real world, it must coordinate with software such as the RDF Readers. Therefore, we designed several Java Beans in the software layer on top of the semantic layer.

Three of JavaBeans are designed for understanding triples and rules. ***ARTExhibiter*** parses semantic annotations based on Wiki markup [24] and shows the facts related to a subject. ***CATExhibiter*** exhibits all categories and further infers quantity, subcategories and articles of each category. ***DOMExhibiter*** demonstrates Faceted Browser by parsing the domain, facts and features according to rules we developed (described shortly) and reasoning instances from TripleStore. The other three components are designed for contacting Jena's query engine and rule-based inference engine. ***RDBOperator*** employs Jena's RDF/OWL APIs for importing RDF Documents to database. ***RDBQuerist*** utilizes Jena's ARQ in order to search articles by the subjects and categories. Finally, ***RDBReasoner*** makes use of Jena's inference engine towards performing our developed rules.

E. Reasoning rules

1) Example case's category rules

We now illustrate with several examples how to use Jena's rule syntax to strengthen category organization and eliminate precision problem mentioned earlier.

Case 1. Categories with hierarchical relationship. Consider the category framework of Fig. 7 as example. As can be seen in Rules H1.1 and H1.2, a significant super-category *wiki:C3ACountry_in_Europe* related to example case's domain *Country* was regarded as conclusion, and subcategories *wiki:C3ACountry_in_European_Union* and *wiki:C3ACountry_in_Schengen_zone* were as premises.

```
[H1:   (?I rdf:type wiki:C3ACountry_in_Europe) ->
       (?I rdf:type wiki:C3ACountry)]
[H1.1: (?I rdf:type wiki:C3ACountry_in_European_Union) ->
       (?I rdf:type wiki:C3ACountry_in_Europe)]
[H1.2: (?I rdf:type wiki:C3ACountry_in_Schengen_zone) ->
       (?I rdf:type wiki:C3ACountry_in_Europe)]
```

However, the rules should be adjusted immediately to compromise the folk's behavior, e.g., a new created category *C3ACountry_in_Asia* was annotated as the subclass of *C3ACountry*. The former was considered as a premise and the latter as a conclusion, as can be seen in Rule H2.

```
[H2:   (?I rdf:type wiki:C3ACountry_in_Asia) ->
       (?I rdf:type wiki:C3ACountry)]
```

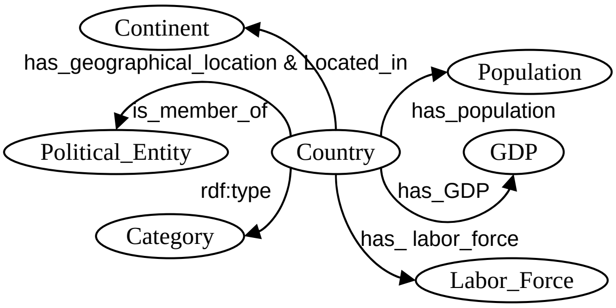


Fig. 8. The Knowledge Map of Semantic Wikipedia.

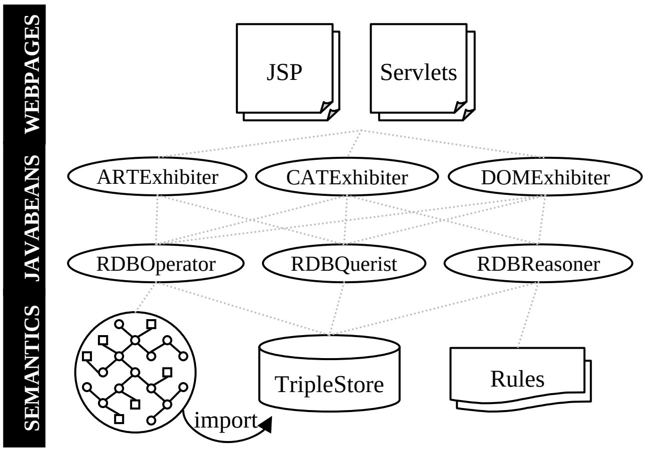


Fig. 9. System Architecture.

Table III. The Instances of Country in Semantic Wikipedia.

Instance	Property	Class
Monaco, Ukraine, Vatican_City	type	Country_in_Europe
Czech_Republic, Hungary, Poland		Country_in_European_Union
Iceland		Country_in_Schengen_zone
Egypt, Ethiopia, Nigeria	Has_geographical_location	Africa
Congo, South_Africa, Sudan	Located_in	

Case 2. Categories with parts-whole relationship. Consider a special phenomenon that all Schengen countries participated in European Union and the member of European Union consists of the member of Schengen. As in Rule U, *C3ACountry_in_Schengen_zone* was regarded as a premise and *C3ACountry_in_European_Union* as a conclusion.

```
[U:     (?I rdf:type wiki:C3ACountry_in_Schengen_zone) ->
        (?I rdf:type wiki:C3ACountry_in_European_Union)]
```

2) *Example Case's Facet Rules*

We now proceed to establish Faceted Classification by reusing those rule-enhanced categories (shown in the first line of the facet rules below) and the analysis results of Semantic Wikipedia's ontologies.

Step 1. Determine the domain. The observing subject *Country* was considered as domain.

Step 2. Rephrase existing facets. In this step, *Space* and *Personality* was chosen from Colon Classification [31], and then was rephrased as *Region* and *Organization* in accordance with the user perspectives in Table I.

Step 3. List instances. Instance of *wiki:C3ACountry*, America, Instance of *wiki:C3ACountry_in_Europe*, Ukraine, and so on were selected in this step.

Step 4. Discover features. *wiki:C3ACountry_in_Europe* was picked as a candidate category, since it is the super class of *wiki:C3ACountry_in_European_Union* and *wiki:C3ACountry_in_Schengen_zone*. *wiki:Central_Asia*, *wiki:Asia* (both representd by *wiki:P3ALocated_in*) and *wiki:East_Asia* (by *wiki:P3AHas_geographical_location*) were chosen as candidate properties. Afterwards, *Europe* was summarized from *wiki:C3ACountry_in_Europe* and *Asia* was generalized from *wiki:C3ACountry_in_Asia*, *wiki:Asia*, *wiki:Central_Asia*, and *wiki:East_Asia*.

Step 5. Create rules. The facts that include *wiki:Asia*, *wiki:C3ACountry_in_Asia*, *wiki:Central_Asia*, and *wiki:East_Asia* were regarded as premises, whereas implied triples such as *Europe* and *Asia* are regarded as conclusions. The facet rules are as follows:

```
@include <http://localhost:9080/SemanticSearch/rdf/category.rules>
[F1:     (?I wiki:facet Region) -> (?I wiki:domain Country)]
[F2:     (?I wiki:facet Organization) -> (?I wiki:domain Country)]
[F3:     (?I wiki:facet Population) -> (?I wiki:domain Country)]
[F1.3:   (?I wiki:feature Asia) -> (?I wiki:facet Region)]
[F1.5:   (?I wiki:feature Europe) -> (?I wiki:facet Region)]
[C1.3.1: (?I rdf:type wiki:Category-3ACountry_in_Asia) ->
         (?I wiki:feature Asia)]
[P1.3.3: (?I wiki:P3AHas_geographical_location wiki:East_Asia) ->
         (?I wiki:feature Asia)]
[P1.3.7: (?I wiki:P3ALocated_in wiki:Asia) ->
         (?I wiki:feature Asia)]
[P1.3.8: (?I wiki:P3ALocated_in wiki:Central_Asia) ->
         (?I wiki:feature Asia)]
[C1.5.1: (?I rdf:type wiki:Category-3ACountry_in_Europe) ->
         (?I wiki:feature Europe)]
[P3.3.1: (?I wiki:P3APopulation ?N), greaterThan(?N, 100000000) ->
         (?I wiki:feature > 100m)]
```

F. Rule evolution

The following illustrates the evolution processes with user scenarios. Scenario 1 (Table IV) shows that an investor wants to seek countries that are located in Europe, belong European Union and have medium-sized labor force for plant installation. The system will (1) obtain firstly the European countries from categories *Country_in_Schengen_zone*, *Country_in_European_Union* and *Country_in_Europe* according to the hierarchical rules H1, H1.1 and H1.2, (2) derive the participators of European Union from category *Country_in_European_Union* including the members of *Country_in_Schengen_zone* based on the parts-whole rule U, (3) acquire countries with population between one million and a hundred million, and (4) intersect these countries to gain the inferred results: *Czech Republic* and *Poland*.

For Scenario 2 (Table V), a businessman prefers cheaper and more labor force, hence considering region *Africa*. The system will merge all Africa's instances with the properties *Has_geographical_location* and *Located_in* according to facets rules P1.1.1 and P1.1.2, and then find population more than 100 million (P3.3.1). Eventually, the inference results tell that *Nigeria* seems the best choice. It goes without saying that the query order could change but get the same outcome.

Table IV. The Evolution Process of Scenario 1.

Country_in_Europe (3)	Monaco, Ukraine, Vatican_City
Country_in_European_Union (3)	Czech_Republic, Hungary, Poland
Country_in_Schengen_zone (1)	Iceland
Europe (7)	Czech_Republic, Hungary, Iceland, Monaco, Poland, Ukraine, Vatican_City
Country_in_European_Union (3)	Czech_Republic, Hungary, Poland
Country_in_Schengen_zone (1)	Iceland
EU (4)	Czech_Republic, Hungary, Iceland, Poland
1m < Population < 100m(2)	Czech_Republic, Poland
Inference Results (2)	**Czech_Republic, Poland**

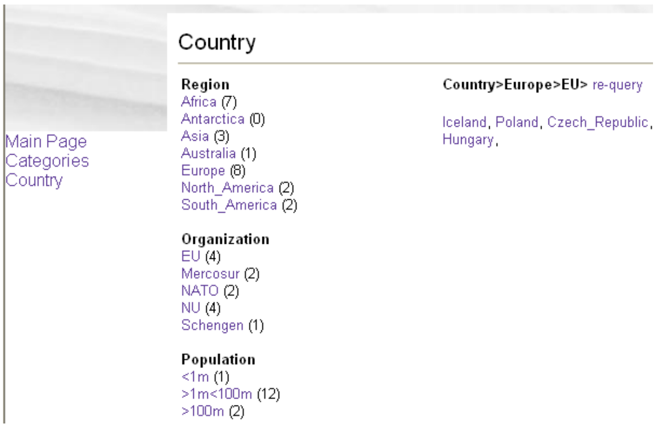
Table V. The Evolution Process of Scenario 2.

Has_geographical_location (3)	Egypt, Ethiopia, Nigeria
Located_in (3)	Congo, South Africa, Sudan
Africa (6)	Congo, Egypt, Ethiopia, Nigeria, South Africa, Sudan
Population > 100m (1)	Nigeria
Inference Results (1)	**Nigeria**

G. System demonstration and evaluation

The system main page (omitted due to space limitation) includes the interfaces for importing RDF files and keyword search. The category overview (also omitted) exhibits the root categories of tree structures and the instances belonged to these categories by parsing and inferring category rules. Additionally, the specific category presents not only the articles regarding current category but also its subcategories.

There are two reasons why our Faceted Browser is better than other general faceted browsers. One reason is that our Faceted Browser is not restricted to fixed sequences. Fig. 10 shows European Countries joined in EU (the result is the same if browsing from EU to Europe). Another reason is that the display of facets, features and instances belonged to the features is based on the result of rule inference which has improved the original categories framework of Semantic Wikipedia. This leads to the flexibility of maintenance because the rule-based Faceted Browser facilitates extension only by creating new facet rules to add new factures or build a new domain. This also leads to the progress of precision because organization problem is relieved and synonym control is provided by category rules.



Main Page
Categories
Country

Country

Region
Africa (7)
Antarctica (0)
Asia (3)
Australia (1)
Europe (8)
North_America (2)
South_America (2)

Organization
EU (4)
Mercosur (2)
NATO (2)
NU (4)
Schengen (1)

Population
<1m (1)
>1m<100m (12)
>100m (2)

Country>Europe>EU> re-query

Iceland, Poland, Czech_Republic, Hungary,

Fig. 10. The Faceted Browsing Result from Europe to EU.

In sum, picking appropriate facets should consider the relativity with domain, mutual exclusivity, a fitting quantity, broader sources (both categories and properties), and a centralized and flexible policy. More precious, domain knowledge should be understood thoroughly, user perspectives should be surveyed deeply, available ontologies should be collected as soon as possible, and flexible notation should be thought over. In other words, the establishment of faceted classification neither extracts directly all classes or properties in existing ontologies nor remakes repeatedly new ontologies. Our approach complies with the above principles to form a better Faceted Classification for Semantic Search.

V. CONCLUSION AND FUTURE WORKS

This study has established an inference-based faceted browser for semantic search where an example case was implemented to demonstrate how the category rules and facet rules are created and evolved. Our approach is an independent solution which separates semantic rules from distributed programming logic to facilitate the integration of inconsistent or heterogeneous ontologies that can promote organization and precision of Folksonomy along with resolving the potential conflicts between Folksonomy and Ontologies.

Two future research suggestions are: (1) Our current method of manually picking facet is expected to be replaced by a powerful algorithm. Constructing a powerful algorithm based on the steps we proposed would be beneficial to reduce human mistake; (2) Web services can be deployed from the Java Beans we realized. This will enable system builders to establish different domain knowledge of semantic faceted search by just offering the RDF files and rules.

REFERENCES

[1] E. Mäkelä, “Survey of Semantic Search Research,” in *Proceedings of the Seminar on Knowledge Management on the Semantic Web*, Department of Computer Science, University of Helsinki, 2005.

[2] R. Guha, R. McCool, and E. Miller, “Semantic Search,” in *Proceedings of the 12th International World Wide Web Conference*, Budapest, Hungary, 2003.

[3] M. Hildebrand, J. van Ossenbruggen, and L. Hardman, “An Analysis of Search-based User Interaction on the Semantic Web,” Technical Report INS-E0706, CWI, Amsterdam, The Netherlands, May 2007.

[4] A. Bernstein, E. Kaufmann, C. Kaiser, and C. Kiefer, “Ginseng: A Guided Input Natural Language Search Engine for Querying Ontologies,” in *Proceedings of the 2006 Jena User Conference*, Bristol, UK, 2006.

[5] V. Lopez, M. Pasin, and E. Motta, “AquaLog: An Ontology-Portable Question Answering System for the Semantic Web,” in *Proceedings of the 2nd European Semantic Web Conference*, Heraklion, Greece, 2005.

[6] M. Hildebrand, J. van Ossenbruggen, and L. Hardman, “/facet: A Browser for Heterogeneous Semantic Web Repositories,” in *Proceedings of the 5th International Semantic Web Conference*, Athens, GA, USA, Nov. 5–9, 2006.

[7] E. Hyvönen, E. Mäkelä, M. Salminen, A. Valo, K. Viljanen, S. Saarela, M. Junnila, and S. Kettula, “MUSEUMFINLAND—Finnish Museums on the Semantic Web,” *Journal of Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 3, issues 2–3, Oct. 2005, pp. 224–241.

[8] S. R. Kruk, “Social Semantic Search and Browsing,” *DERI Research Seminar*, Galway, Ireland, 2005.

[9] E. Oren, R. Delbru, and S. Decker, “Extending faceted navigation for RDF data,” in *Proceedings of the 5th International Semantic Web Conference*, Athens, USA, 2006.

[10] O. Suominen, K. Viljanen, and E. Hyvönen, “Semantic Faceted Search in a Citizens’ Health Portal,” in *Proceedings of European Semantic Web Conference*, Innsbruck, Austria, 2007.

[11] L. Ding, T. Finin, A. Joshi, R. Pan, R. S. Cost, Y. Peng, P. Reddivari, V. Doshi, and J. Sachs, “Swoogle: A Search and Metadata Engine for the Semantic Web,” in *Proceedings of the 13th ACM International Conference on Information and Knowledge Management*, Washington, D.C., USA, 2004, pp. 652–659.

[12] S. Auer, S. Dietzold, and T. Riechert, “OntoWiki – A Tool for Social, Semantic Collaboration,” in *Proceedings of the 5th International Semantic Web Conference*, Athens, GA, USA, Nov. 5–9, 2006.

[13] M. Völkel, M. Krötzsch, D. Vrandecic, H. Haller, and R. Studer, “Semantic Wikipedia,” in *Proceedings of the 2006 International Conference on World Wide Web*, Edinburgh, Scotland, 2006.

[14] T. Gruber, “Ontology of Folksonomy: A Mash-up of Apples and Oranges,” *International Journal on Semantic Web and Information Systems*, vol. 3, issue 1, 2007, pp. 1–11.

[15] S. Hayman and N. Lothian, “Taxonomy-Directed-Folksonomy,” in *Proceedings of the World Library and Information Congress: 73rd IFLA General Conference and Council*, Durban, South Africa, Aug. 19–23, 2007.

[16] E. Kroski. (2005). “The Hive Mind: Folksonomies and User-Based Tagging,” Available: http://infotangle.blogsome.com/2005/12/07/the-hive-mind-folksonomies-and-user-based-tagging/.

[17] Wikimedia Foundation. (2001–2008). “Wikipedia,” Available: http://www.wikipedia.org/.

[18] P. Murray. (2002). “Faceted Classification of Information,” Available: http://www.kmconnection.com/DOC100100.htm.

[19] B. C. Vickery, *Faceted Classification: a Guide to the Construction and Use of Special Schemes*, Aslib, London, 1960.

[20] N. F. Noy and D. L. McGuinness. (2001, March). “Ontology Development 101: A Guide to Creating Your First Ontology,” Technical Report KSL-01-05, Stanford Knowledge Systems Laboratory and Technical Report SMI-2001-0880, Stanford Medical Informatics, Available: http://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html.

[21] Yahoo! Inc. (2008). “Flicker,” Available: http://www.flickr.com/.

[22] YouTube, LLC. (2008). “YouTube,” Available: http://youtube.com/.

[23] Robot Co-op. (2004–2008). “43Things,” Available: http://www.43things.com/.

[24] Wikimedia Foundation. (2001–2008). “Wikipedia: How to edit a page,” Available: http://en.wikipedia.org/wiki/Wikipedia:How_to_edit_a_page.

[25] W. Denton. (2003). “How to Make a Faceted Classification and Put It On the Web,” Available: http://www.miskatonic.org/library/facet-web-howto.html.

[26] P. V. Dijck. (2002). “XFML Core – eXchangeable Faceted Metadata Language,” Available: http://petervandijck.com/xfml/spec/1.0.html.

[27] HP Labs Semantic Web Research. (2000–2008). “Jena – A Semantic Web Framework for Java,” Available: http://jena.sourceforge.net/.

[28] M. Krötzsch and D. Vrandecic. (2007). “SWIVT Ontology Specification,” Available: http://semantic-mediawiki.org/swivt/.

[29] Q. M. Ilyas, Y. Z. Kai, and M. A. Talib, “A Conceptual Architecture for Semantic Search Engine,” in *Proceedings of the 8th International Multitopic Conference*, Lahore, Pakistan, 2004.

[30] S. Melnik. (2000). “Storing RDF in a relational database,” Available: http://infolab.stanford.edu/~melnik/rdf/db.html.

[31] S. R. Ranganathan, *Elements of library classification*, Asia Publishing House, New York, 1962.