Solving Problem of Ambiguity Terms using Ontology

Herlina Jayadianti*,**, Lukito Edi Nugroho*, Carlos Sousa Pinto**, Paulus Insap Santosa* and Wahyu Widayat***

**Department of Electrical Engineering and Information System, Universitas Gadjah Mada, Yogyakarta, Indonesia*
E-mail: herlinajayadianti@gmail.com
*** Departemento de Sistemas da Informacao, Universidade do Minho, Campus de Azurem, Guimaraes, Portugal*
**** Economic Development, Gadjah Mada University, Yogyakarta, Indonesia*

Abstract

Humans as God's creatures were endowed with brain used to think and to process the received knowledge. Human brain links knowledge into a giant network of ideas, memories, predictions and beliefs. Everything is interconnected in the brain. Human can understand the meaning of the same variable even if that variable has an ambiguous meaning. In another scenario, humans can understand the same thing based on different terms. People can argue and debate with one another about any problem to get a better solution. The different understanding of terms is not a big problem for human brain but imagine if it happens between computers or between humans and computers! Based on this problem we want to provide a model of managing knowledge in situations where terms can generate ambiguity. Knowledge representation process is a good approach to organize and share knowledge. We use ontologies as a technique to represent the particular knowledge stored in each computer and to find correspondences between the concepts used in those ontologies.

Key Words: Ontology, perceptions, understanding, semantic.

1. Introduction

Knowledge is a consequence of learning from experience. People can easily manage experience gained, into new knowledge. Knowledge is made of a set of vocabulary or terms that can be combined in order to construct a meaningful sentence. What about computers? Computers are better than humans in terms of remember things and perform complex numerical calculations, but not in terms of integrating and sharing knowledge from many different sources with different variables, different vocabulary and different semantics. Computers are not as intelligent as humans even tough they have an artificial “brain”. Regarding to this problem we will focus our work on computers intelligence. They can share, integrate, and understand different semantics or different terms like humans.

Semantic integration [1],[2] intends to resolve *different semantic* among various computers.

The main problem is the lack of specification of semantic heterogeneity and ambiguity [3], e.g. the process of finding "apples" in search engine today. When we conduct a search with Google, we cannot easily get the pages about fruit "*blackberry*" — the most highly ranked pages are about the computer company "*blackberry Inc.*", not the fruit. Google is a search engine and not a knowledge engine. Nowadays, there are several good knowledge engines such as Swoogle[1], Sindice, and Watson. They can push machine to understand what user want, e.g : User want to search a "*blackberry fruit*" not the enterprise "*blackberry Inc*". Another scenario is about the high number of online book stores, with each of them having their own knowledge base containing the information about the books it sells such as "Cost" and "Price" or *"Item"* and *"Number of item"*.

The ability of the human brain can not be replaced by computers, but researchers try to continue their research to attempt to obtain the most intelligent computer. It is a challenge that must be faced to make computers think like humans; of course it is certainly not an easy thing to implement. The first step is to represent knowledge. Knowledge representation (KR) is designed as a connection between data in one computer or several computers, and they will use that knowledge together. Knowledge representation is usually used for making expert systems and to allow computers think and solve problems like humans. There are several KR techniques such as frames, rules, tags, and semantic networks which are originated in cognitive science. Recent developments in KR, developed with XML-based knowledge representation languages, including Resource Description Framework (RDF), RDF Schema, and Web Ontology concepts, are very important in knowledge sharing and knowledge understanding processes. In this paper we propose a preliminary work of how computer can share and understand knowledge together using a common set of terms derived from several different ontologies. We will show the result trough a small implementation project. This paper is organized as follows: (1) Introduction; (2) Knowledge representation with RDF, OWL and ontology; (3) Implementation of the solution; (4) Discussion and (5) Conclusions of work.

2 Knowledge Representations "Blackberry" vs "Blackberry"

Wordnet[2] online library is a recommended application to find relatedness among terms. WordNet implements measures of similarity and relatedness among terms. Measures of

[1] http://swoogle.umbc.edu/ Swoogle was the first search engine dedicated to online semantic data. Its development was partially supported by DARPA and NFS (National Science Foundation).

[2] http://wordnet.princeton.edu/

similarity use information found in an *is–a* hierarchy of concepts, and quantify how much concept *"A"* is equivalent to concept *"B" or* concept *"A"* is not equivalent to concept *"B"*, e.g. *"Item"* is equivalent to *"Number of item"*. Another example is *"blackberry"* as a fruit is not equivalent to *"blackberry Inc"* (See Fig.1).

These situations shows that there are two conditions that can happen in a system: (1) Same terms with different meaning (ambiguity [3] terms) or (2) Different terms with same meaning [4], [5]. Trough this paper, we focused on ambiguity terms.

Fig.1 Different concepts of term Blackberry

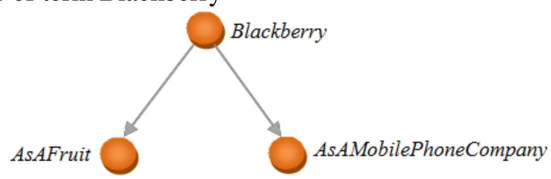
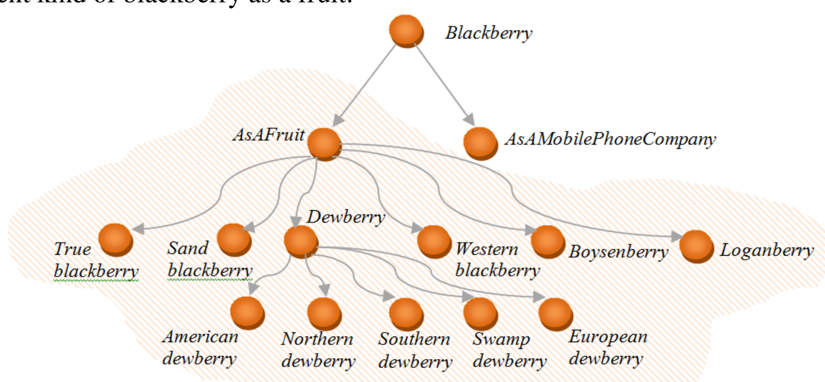
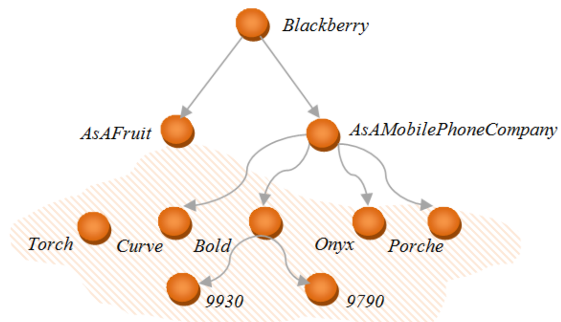


Fig2. Different kind of blackberry as a fruit.



As a kind of a fruit, there are a lot of different perceptions of it, such as; *"True blackberry (Rubus fruticosus)"*, the sweet edible black or dark purple from Europe; *"Sand blackberry (Rubus cuneifolius)"* from United States; *"Dewberry (dewberry bush)"* from North America; *"American dewberry" (Rubus Canadensis)* from North America; *"Northern dewberry (Rubus flagellaris)"* from eastern North America; *"Southern dewberry (Rubus trivialis)"* from southern North America; *"Swamp dewberry (Rubus hispidus)"* from eastern North America; *"European dewberry (Rubus caesius)"*; *"Western blackberry (Rubus ursinus)"*; *"Boysenberry (boysenberry bush)"*; *"Loganberry (Rubus loganobaccus or Rubus ursinus loganobaccus)"* from California.

Fig 3. Different kind of blackberry as a mobile phone.



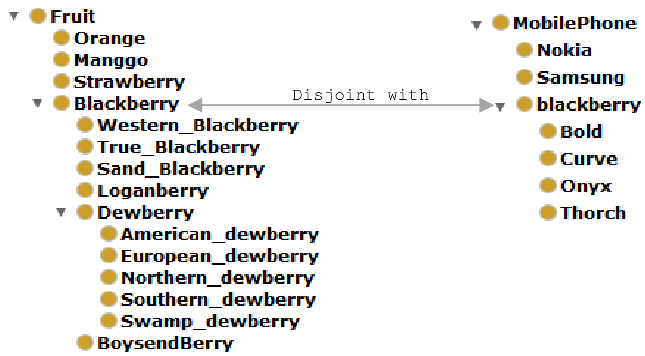
Example: The "*Bold blackberry*" as a type of mobile phone has the following DataProperties: `Color: Black and Grey, Price: 500USD, Type: 9930.`

Some users can use their own knowledge to search data about a *blackberry* as a brand of mobile phone. They can use, for example, the term *"Curve blackberry black" in Google search engine. As* another example we can consider another user that already have knowledge about *"sweet black blackberry"* and he want to search data about sweet black blackberry from Europe with term *"True blackberry"*. We should highlight that this is an opportunity for user to search a sweet blackberry only with term *"blackberry"*, maybe he don't have enough knowledge and experience about any kind of blackberry as a fruit (such as `Latin:` *`Rubus fruticosus`*; `Color: Black`; `Taste: Sweet`; and `Origin:Europe`), on the other hand computer as a machine don't have enough knowledge in their "artificial brain" to understand the semantic of "blackberry". If the user uses the term *"blackberry"* for a target fruit blackberry and push machine to understand the meaning of what he want, the machine can give to him information about both blackberry concepts - as a *sweet fruit blackberry* and as a *blackberry mobile phone.* How to include knowledge in machines allowing them to perform "thinking" somewhat like humans is the main purpose of this paper.

3 Implementation of the solution

Let's consider the referred case study of two different ontologies both representing a different domain. One ontology focus on domain `fruit` and another one focus on domain `MobilePhone.` The problem is that they use same terms with different semantics e.g. `Blackberry.` This section shows the simple implementation of the problem with knowledge representations using RDF and OWL. First process is class design for domain fruit (See Fig 4 and Fig 5).

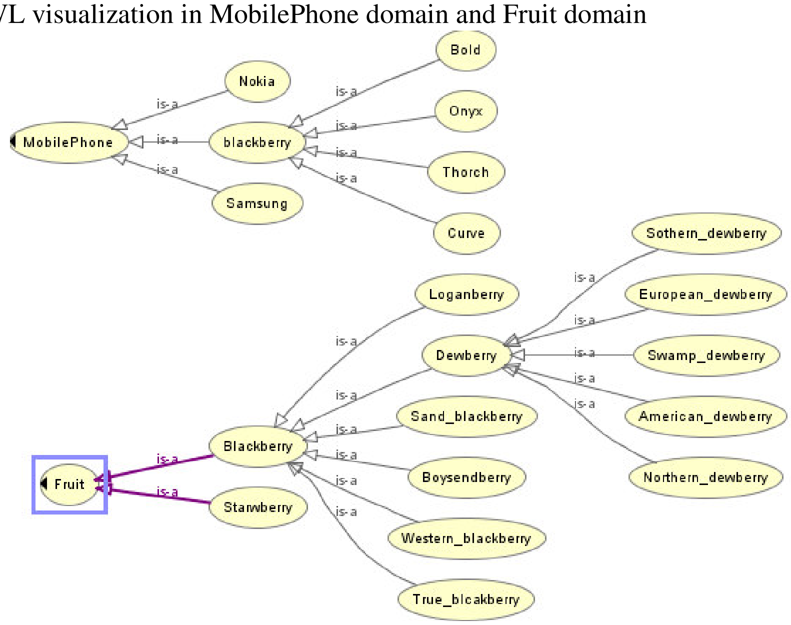
Fig 4. Class design



We can see in Fig 4 and Fig 5 that there are two (2) ontologies about different domains - ontology `Fruit` and ontology `MobilePhone`. Each of them uses the same class named “`blackberry`”, but both are disjoint.

```
 <owl:Class rdf:about="&blackberry;Fruit">
<rdfs:label xml:lang="en">Blackberry</rdfs:label>
 <rdfs:subClassOf rdf:resource="&Fruit"/>
 <owl:disjointWith rdf:resource="&blackberry"/>
<owl:disjointWith rdf:resource="&blackberry;MobilePhone"/>
</owl:Class>
```

Fig 5. OWL visualization in MobilePhone domain and Fruit domain



In the domain of fruit-subclass blackberry, we will focus on class “`True_blackberry`”, a kind of blackberry fruit. “`True_blackberry`” has some properties (datatype):

```
hasColor(True_blackberry,black);
hasLatin(True_blackberry,Rubus fruticosus);
```

```
hasOriginFrom(True_blackberry,Europe).
hasTaste(True_blackberry,sweet)
```

In the domain `mobilephone`, considering the subclass blackberry, we will focus on the subclass `"bold"` (kind of blackberry phone). `"bold"` has as properties (datatype): `hasColor(bold,black);` The next step after designing classes, individuals and datatype properties is testing. We will use sparql[3] as a query language for Resource Description Framework (RDF)[4] databases.

```
Prefix : <http://www.semanticweb.org/user/blackberry#>
SELECT ?Blackberry ?Category
WHERE {?Blackberry :HasTaste ?Category.
WHERE {?Blackberry :HasColor ?Category.
?Category :TypeCategory ?value.
…
}
```

Fig 6 .Query results in Sparql testing query.

Blackberry	Category
western_blackberry	Fruit
true_blackberry	Fruit
sand_blackberry	Fruit
loganberry	Fruit
swamp_dewberry	Fruit
southern_dewberry	Fruit
northern_dewberry	Fruit
european_dewberry	Fruit
american_dewberry	Fruit
boysendberry	Fruit

Blackberry	Category
Torch	MobilePhone
Curve	MobilePhone
Bold	MobilePhone

Fig 6 shows the result of a simple query test with Sparql. The result of that test shows that ambiguity can be addressed by ontology.

4. Discussions

To build an "artificial brain" in computer as intelligent as "human brain" is not a simple thing. Mapping out all perceptions, visualizations, and arguments from human brain into a complete artificial knowledge in computer program was a difficult work. Present above, users enter value of "blackberry" based on perceptions, experience and knowledge they have, such as `color`, `flavor`, `taste`, `type`, and others. Computer will understand what users request based on `dataProperty` as a value, such as

[3] http://www.w3.org/TR/rdf-sparql-query/

[4] RDF is a standard model for data interchange on the web. http://www.w3.org/RDF/

```
"Hastaste".
WHERE {?Blackberry :HasTaste ?Category.
WHERE {?Blackberry :HasColor ?Category.
```

Based on DataProperty `"HasTaste(Sweet)"` the target is blackberry as a `"fruit"` instead of `"mobilePhone"`.

This work is different from our previous works about semantic equivalency [2],[3]. In the present article we show that representation of knowledge with ontology can resolve some semantic problems. Ontology is responsible for discovering semantic relationships between concepts from various information models.

5 Conclusions

Computers can not learn by observation and experiences like human brain do. They also can not easily share and understand knowledge from different user using different terms and semantic. It is not easy to build a system that can accommodate what all users want, especially if we have to consider ambiguity. The two situations that can be happen are: (1) Same terms with different meanings (ambiguity terms) or (2) Different terms with same meaning. Trough this paper, we focused on ambiguity terms. To represent knowledge, RDF and Ontology are recommended technologies that allow better solutions to make search processes easier to humans and to computers.

6. Acknowledgments

We would like to admit the funding from the Erasmus Mundus EuroAsia program as a research foundation to this project, and also to acknowledge Universidade do Minho and Universitas Gadjah Mada for the collaboration.

References

[1] Y. Xue, "Ontological View-driven Semantic Integration in Open Environments," Book, The University of Western Ontario, 2010.

[2] N. F. Noy, "Semantic integration: a survey of ontology-based approaches," *Acm Sigmod Rec.*, vol. 33, no. 4, pp. 65–70, 2004.

[3] C. Brewster and K. O'Hara, "Knowledge representation with ontologies: the present and future," *Intell. Syst. Ieee*, vol. 19, no. 1, pp. 72–81, 2004.

[4] H. Jayadianti, C. Pinto, L. Nugroho, and P. Insap, "Integrating large knowledge repositories in multiagent ontologies," *Proceedings of the 2nd International Workshop on Exploiting Large Knowledge Repositories (E-LKR'12), in conjunction with the 1st International Workshop on Automatic Text Summarization for the Future (ATSF-2012), CEUR-WS.org/Vol-882 - Exploiting Large Knowledge Repositories (E-LKR'12)*, 2012. [Online]. Available: http://ceur-ws.org/Vol-882/. [Accessed: 18-Jan-2013].

[5] H. Jayadianti, L. Nugroho, C. Pinto, P. Insap, and W. Widayat, “Solving Problems of Data Heterogeneity, Semantic Heterogeneity and Data Inequality using Ontologies,” *Elpub Conference 2012*, 2012. [Online]. Available: http://www3.dsi.uminho.pt/elpub2012/Elpub_2012/Programme.html. [Accessed: 18-Jan-2013].