

# Semantic Structure Content for Dynamic Web Pages

Mamdouh Farouk, Mitsuru Ishizuka

Creative Informatics Department

Graduate School of Information Science & Technology, The University of Tokyo

Tokyo, Japan

mamdouh@mi.ci.i.u-tokyo.ac.jp, ishizuka@i.u-tokyo.ac.jp

**Abstract**—Representing web data into a machine understandable format is a curtail task for the next generation of the web. Most of solutions are relying on ontologies. However, there are many problems of using ontologies. This paper proposes an approach to represent dynamic web page contents retrieved from underlying database, into Concept Description Language (CDL) semantic format. This format does not depend on ontologies. However, CDL describes semantic structure of web content based on a set of predefined concepts and semantic relations. A prototype of the proposed approach is implemented to show visibility of the proposed approach.

*Semantic representation, semantic database, Concept Description Language*

## I. INTRODUCTION

Representing web data and resources into a standard semantic format is an important step for semantic web [1]. This semantic representation enables intelligent agents to interact web resources semantically [2][3][4]. Work that addresses the issue of representing dynamic web content into a semantic format uses semantic web languages (RDF, DAML, OWL ...), which depend on ontology [5]. However, this paper proposes a technique to represent dynamic web content into a new semantic structure format which does not depend on ontologies.

Many researchers use ontologies to represent web data into a machine understandable format [6][7]. However, there is no agreement on an ontology. In other words, there are a lot of ontologies available on the Internet that may use same terminology to refer different concepts and vice versa [8]. Moreover, the problem of ontology interoperability is still an open problem. So using current web ontologies in representing web data is not an optimal solution.

This paper focuses on representing dynamic web content that are generated from underlying database into CDL semantic format which does not depend on ontologies. The generated CDL semantic representation can be used by intelligent search engine to get accurate search results.

This paper is organized as follows: section 2 defines the semantic language which used in the proposed technique. Section 3 explains the architecture of the proposed system. Section 4 shows a prototype for the proposed approach.

Comparison with related work is discussed in Section 5. Finally, section 6 contains the conclusion of this research.

## II. CONCEPT DESCRIPTION LANGUAGE (CDL)

Our approach uses a new semantic language which called Concept Description Language (CDL). This semantic format, which proposed by Institute of Semantic Computing, describes semantic/conceptual structure of contents (resources) and can deal with natural languages, mathematical expressions, movie, music, etc [9]. The aims of CDL are to realize machine understandability of web text contents, and to overcome language barrier on the web [10].

CDL is one of the three forms that can be used to express CWL (Common Web Language). Moreover, CWL is a common language for exchanging information through the web and also for enabling computers to process information semantically. CWL is a part of the Incubator Activity of W3C [11].

This new representation bases on Concept Description Language for natural language (CDL.nl) which describes the concept structure of the text based on a set of predefined semantic relations [9]. The main advantage of CDL is that it does not depend on ontologies. However it depends on the Universal Networking Language Knowledge Based (UNLKB) and a set of universal relations so it can be used without facing similar problems of using ontologies. CDL depends on a complete dictionary which contains around 120000 words. Moreover, this dictionary contains the definitions of these words represented into CDL form. This means that the computer can understand the word definition as well as the semantic relation between words that is also contained in that dictionary. Moreover, reasoning agents can use these definitions to get better understanding and more accurate results.

For example, representation of a statement such as “John bought a computer yesterday” in CDL is: {#A Event tmp='past';  
{#a1 buy;}  
{#a2 computer;}  
{#a3 yesterday;} {John John;}  
[#a1 agt John] [#a1 obj #a2] [#a1 tim #a3]}

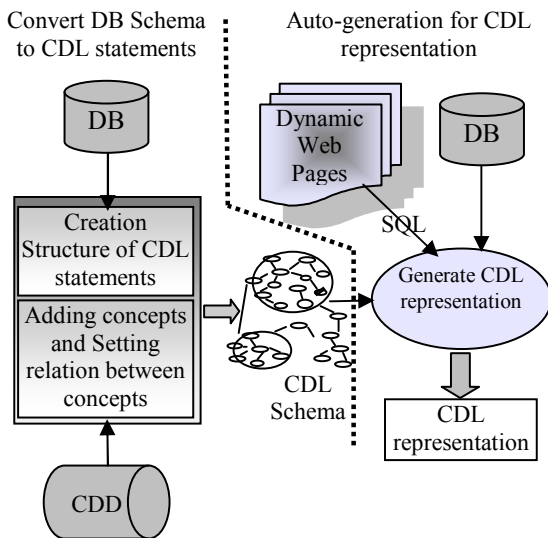


Figure 1. System architecture

### III. PROPOSED SYSTEM ARCHITECTURE

The architecture of our system consists of two main components, DB converter and CDL generator as shown in fig.1. The first step to represent dynamic web content, which generated from DB, into CDL is converting the underlying DB schema to a semantic network. This semantic network is represented as a set of CDL statements which contain references to DB objects. This process of converting DB schema to a semantic network occurs only once at installation time. The second step is auto-generation for the semantic of the webpage dynamic content using those CDL statements.

#### A. Converting DB schema to CDL semantic network

In this stage, the schema of the underlying DB is converted to semantic related concepts. The output of this process is a semantic network in which nodes represent concepts and arcs represent semantic relations between concepts. As shown in fig. 1, there are two steps for this conversion.

First, a semantic network structure that reflects database schema is created automatically by DB2CDL tool, fig. 2. In this step, for each data field in the DB a CDL statement is created.

Second, the user manually completes a CDL statement for each field by adding concepts and relations to the statement to make the overall meaning of the DB schema. For example, to represent the field *address* in table *members*, the user may insert a new event (concept) such as "live" and use CDL semantic relations to get the semantic representation for this field as the following notation: "M.name←agt live plc →M.address". This notation means that a member is an agent for the event *live* and the place of this event is the member's address. Moreover, when the user tries to add a concept to a CDL semantic network, the DB2CDL tool shows different usage forms of this concept according to CDD (Concept

Definition Dictionary) of CDL language. The user should select one choice depending on the meaning. For example, if the user wants to add the verb *live*, he/she should select according to his meaning from the following alternatives.

- live(agt>person,obj>thing)
- live(agt>person,obj>food)
- live(agt>person)
- live(agt>thing,obj>state)
- live(aoj>behavior)
- ...

In addition, it is important to include DB relations into the outputted semantic network. So the final generated semantic CDL is accurate and reflects the overall meaning of the data. Using the proposed tool (DB2CDL) the user can convert the DB field which represents a foreign key such as *deptID* field in *member* table, by creating a CDL statement which relates information from both tables (*member* and *department*). For instance, the user may add an event "work", make a relation between this event and the *name* field in *researcher* table, and add another relation between the new event and the *name* field in *department* table. Finally, the statement will be as the following:

```
<S field='deptID' ref='department'>
  <cdl>{ [a0:work(agt>person)]
    [a1:"?member.name?"]
    [a2:department(pof>organization)]
    [a4:"?department.name?"]
    [a1 agt a0][a2 plc a0][a4 mod a3] };:uw</cdl>
</S>
```

The references to database objects in outputted CDL statements are replaced with the appropriate data in the CDL generation phase. Converting DB schema to CDL statements is not a time consuming or tedious work because the user does this only once. The proposed tool (DB2CDL), fig. 2, helps

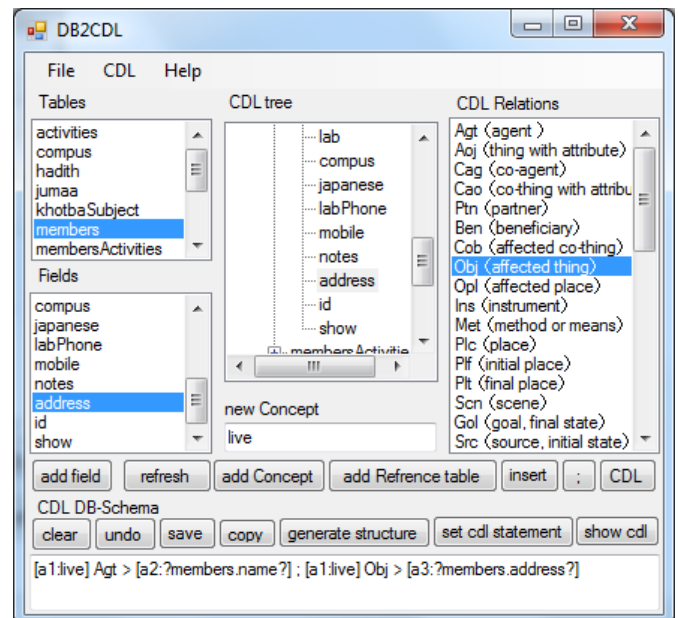


Figure2. DB2CDL tool

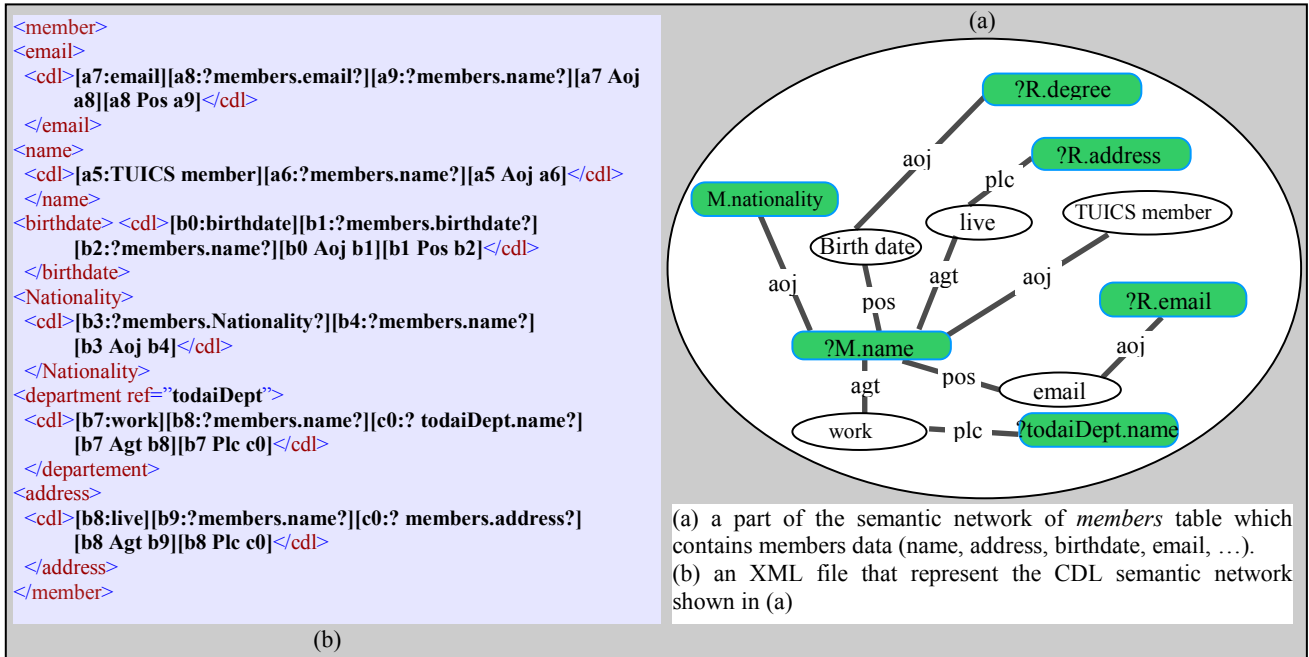


Figure.3 CDL semantic schema example

users to manage this conversion correctly and easily. Fig. 3 shows an example for the process of converting a DB table such as "member" table, which contains data of student society members, to CDL semantic statements.

The generated CDL statements will not be changed even though the stored data is changed. However, if the database schema changed, these CDL statements should be adapted to reflect schema changes.

B. CDL data generation

This stage automatically converts a dynamic web page which retrieves its contents from DB to CDL format. This conversion is based on the semantic DB schema represented into CDL statements, which is the output of the first stage. This step is maintenance free. This means that there is no change in this step even though the database schema is changed.

A dynamic web page which retrieves its contents from underlying database contains SQL queries that are executed on server side to generate the page content. In order to represent this dynamic content into a semantic format, SQL query result should be represented into CDL format.

	<b>Mohamed Farghaly</b>
<b>Nationality</b>	Egyptian
<b>status</b>	D1
<b>Department</b>	Creative Informatics
<b>Contacts</b>	
<b>Telephone</b>	090-6493-433
<b>Address</b>	Minato-ku, Tokyo
<b>E-mail</b>	<a href="mailto:Ali@hotmail.com">Ali@hotmail.com</a>

The left side shows an example of a web page while the right side shows the corresponding CDL semantic statements for that web page.

```

[a5:TUICS member][a6:Mohamed Farghaly][a5 Aoj a6]

[b3:Egyptian][b4:Mohamed Farghaly][b3 Aoj b4]

[b5:D1][b6:Mohamed Farghaly][b5 Aoj b6]

[b7:work][b8:Mohamed Farghaly][b9:work][c0:Creative Informatics][b7 Agt b8][b9 Plc c0]

[c4:mobile][c5:090-6493-433][c6:Mohamed Farghaly][c4 Aoj c5][c5 Pos c6]

[c1:live][c2:Mohamed Farghaly][c1:live][c3:Minato-ku, Tokyo][c1 Agt c2][c1 Plc c3]

[a7:email][a8: Ali@hotmail.com][a9:Mohamed Farghaly][a7 Aoj a8][a8 Pos a9]
```

Figure 4. final CDL representation example

For example, consider a dynamic web page, which shows member information, contains an SQL query, “*Select member.name, phone, address, email, researchPoints, degree, department.name from member, department where member.id = Pr and member.deptID = department.ID;*”. If the web server receives a request to this page with the parameter Pr=42, it will show the information of a member with id = 42, fig. 4.

Based on semantic CDL schema that generated in the first phase we can generate CDL semantic of the query result. For example, it is stated in the CDL schema that the member name is the agent of an event 'live' and the place of this event is the member's address. Consequently, by replacing DB references we obtain this CDL { live:0D >agt Ali Saber:01; :0D >plc "Minato-ku, Tokyo":1D }::uw. Moreover, in order to auto-generate CDL data for query result the following steps should be executed.

1. get tables list stated in the *from* clause
2. get list of fields retrieved from each table
3. get CDL statement for each field from CDL semantic schema
4. for each field, generate the corresponding CDL statement by replacing DB references in the schema statement with values of query result.
5. find the DB relations stated in both *where* clause and CDL schema
6. for each field represent a relation (foreign key) find its CDL statement. If there is missing information for any statement, make a new query to get this information.

By applying these steps on the previous query the representation of the web page contents will be as shown in fig. 4. These CDL statements contain semantic of the web page contents. Consequently, this dynamic web page can be accessed semantically.

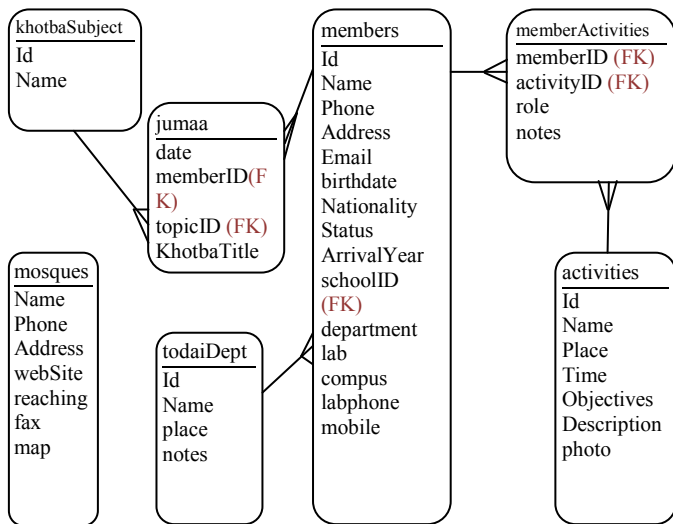


Figure 5. TUICS DB schema

This phase of automatic CDL generation for a dynamic web page contents was implemented as a set of APIs. These APIs can be used in the script of the web page. This means that if a request is received by the web server to a specific page, the

```

<?php
$sql = "select id, name,
Nationality from members
where `show`= 1";
$res = mysql_query($sql);
while($row =
mysql_fetch_array($res))
{
echo
"<tr><td>".$row['name']."
</td><td>".$row['National
ity']."</td><td><a
href='memberDetails.php?
memberId=".$row['id']."'>
Details</a></td></tr>";
}
?>

<?php
$sql = "select id, name, Nationality
from members where `show`= 1";
$res = mysql_query($sql);
while($row =
mysql_fetch_array($res))
{
include("CDLGenerator.php");
$generator=new GenCDL();
$generator->
initialize("tuicsCDLSchema.xml
", $con);
$CDL = $generator->
getCDLRepresentation($sql);
echo "<!--"; $generator->
putInPage($CDL); echo "-->";
....
  
```

Figure 6. CDL generation APIs

CDL generation process will be run as a step of the requested pages script execution on the server side. The generated CDL form is sent to the client as a part of page content. As a result, web agents can easily understand this kind of web pages.

#### IV. IMPLEMENTATION AND EXAMPLES

This section shows a prototype for the proposed approach the applied on a real web site. TUICS web site (www.tuics.com) was selected to implement the proposed approach. TUICS is a website for a students society. It shows information of members of this society and their activities. It contains a lot of dynamic web pages developed using php. This prototype uses TUICS web site, which contains a simple DB, to show the visibility of the proposed approach.

In order to represent TUICS dynamic content into CDL, as a first step, TUICS DB schema (fig. 5) should be converted to CDL semantic network using DB2CDL tool. The resulted semantic network of TUICS DB stored in the web server as an XML file named *tuicsCDLSchema.xml*. The second step is to modify php pages to use the implemented APIs that generate CDL representation. Fig. 6 shows php code of *member.php* web page before and after using the implemented APIs. It is a small piece of code that should be added to generate CDL representation for the dynamic content which retrieved from the DB. The main function, *getCDLRepresentation*, takes an SQL query as an input and returns the CDL representation for the query result. The generated CDL is added as a comment in the web page so a web browser can show the original page and web agents can get the semantic representation of the content. The final CDL representation which generated by the proposed approach is available on the Internet at [www.tuics.com](http://www.tuics.com).

#### V. RELATED WORK

Comparing to RDB2RDF approach, which converts relational database to RDF [12], CDL is richer than RDF. This is because RDF represents data in triples format (object, property, value). However, CDL represents data in a semantic structure that enables users to express more internal relations.

For example, RDF representation for the result of a query such as “select name, address from members where id =3;” is represented in two triple as follows: (member, name, *Khaled*) (member, address, *Tokyo*). However, CDL representation for the same query result looks like (*Khaled*←agt-live—plc→*Tokyo*). There is additional relation between name and address in CDL representation. This kind of relations is useful in answering queries semantically. Finally, the generated CDL does not depend on a domain ontology.

## VI. CONCLUSION

This paper proposes an approach to represent dynamic web content that generated from DB into a semantic format (CDL) which does not depend on ontologies. The proposed technique semi-automatically transforms the database schema to CDL semantic network using the implemented tool (DB2CDL). Based on this CDL semantic network the proposed approach auto-generates semantic description of a dynamic web page according to the retrieved data the page contains. Semantic agent can use the generated CDL to answer question and search the original content semantically. The implemented prototype on a real web site, [www.tuics.com](http://www.tuics.com), shows the visibility of the proposed approach.

## REFERENCES

- [1] T. Berners-Lee, J. Hendler, O. Lassila, “The Semantic Web”, *Scientific American*, Vol. 284, No. 5, 2001, pp. 34-43.
- [2] Siegfried Handschuh , Raphael Volz , Steffen Staab, Annotation for the Deep Web, *IEEE Intelligent Systems*, v.18 n.5, September 2003, pp.42-48.
- [3] Mamdouh Farouk, Samhaa R. El-Beltagy, Mahmoud Rafea, "On-the Fly Annotation of Dynamic Web ", *Proceedings of the First International Conference on Web Information Systems and Technologies (WEBIST 2005)*, Miami (USA), may 2005, pp 327-332.
- [4] Yulan Yan, Yutaka Matsuo, Mitsuru Ishizuka, Toshio Yokoi. "Annotating Extension layer of semantic structure for natural language text", *The IEEE International conference on semantic computing*. 2008, pp.174-181.
- [5] Zhuoming Xu, Shichao Zhang, and Yisheng Dong, Mapping between Relational Database Schema and Owl Ontology for Deep Annotation, *WT'06: Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence*, IEEE Computer Society, 2006, pp. 548-552.
- [6] Ismael Navas Delgado, María del Mar Roldán García, José Francisco Aldana Montes: “Deep Crawling in the Semantic Web: In Search of Deep Knowledge”. *WISE 2004*, pp. 541-546
- [7] Ismael Navas Delgado, Nathalie Moreno Vergara, Antonio C. Gomez Lora, María del Mar Roldán García, Iván Ruiz Mostazo, José Francisco Aldana Montes: “Embedding Semantic Annotations into Dynamic Web Contents”. *Proceeding of 15th international workshop on database and Expert Systems Applications*, 2004, pp. 231-235
- [8] Li Wenjie. Study of Semantic Web-Oriented Ontology Integration Technologies. In : *Proceedings of the WRI World Congress on Software Engineering (WCSE'09)* . Xiamen, China, May 2009, 2 : 142-145
- [9] T. Yokoi, H. Uchida, K. Hasida, et al. CDL (Concept Description Language): A Common Language for Semantic Computing, *www2005 workshop on the semantic computing initiative (SeC2005)*
- [10] Mitsuru Ishizuka, "A Common Concept Description of Natural Language Texts as the Foundation of Semantic Computing on the Web", *IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing*, Taiwan, June 2008, p.385
- [11] H. Uchida, T. Yokoi, M. Zhu, N. Saito, V. Avetisyan, “Common Web Language”, *W3C Incubator Group Report*, <http://www.w3.org/2005/Incubator/cwl/XGR-cwl/>, March 2008
- [12] Svihla, M., Jelinek, I.: The Database to RDF Mapping Model for an Easy Semantic Extending of Dynamic Web Sites. *Proceedings of IADIS International Conference WWW/Internet*, Lisbon, Portugal, 2005, pp.27-34