

MỘT GIẢI PHÁP CHUYỂN ĐỔI TỪ CƠ SỞ DỮ LIỆU QUAN HỆ SANG MÔ HÌNH DỮ LIỆU CHO WEB NGỮ NGHĨA

LƯƠNG THÁI NGỌC*, NGUYỄN HỮU DUYỆT**

TÓM TẮT

Web ngữ nghĩa là một hướng phát triển tương lai của Web hiện tại, trong đó RDF là chuẩn cho phép đặc tả dữ liệu cho Web ngữ nghĩa. Trong bài báo này, tác giả sẽ trình bày một hướng tiếp cận cho phép chuyển đổi dữ liệu từ cơ sở dữ liệu (CSDL) quan hệ sang mô hình dữ liệu cho Web ngữ nghĩa được biểu diễn bằng đồ thị RDF, kết quả là tạo ra được một mô hình dữ liệu tương ứng được biểu diễn dưới dạng các bộ ba (Triple). Quá trình thực nghiệm chuyển đổi với CSDL quan hệ Quản lý bán hàng cũng được trình bày nhằm đánh giá kết quả nghiên cứu.

Từ khóa: CSDL, RDB, RDF, dữ liệu, quan hệ, web ngữ nghĩa.

ABSTRACT

A solution to convert Relation Database to data model for Semantic Web

Semantic Web (Web of Data) is a development trend of current Web, which is based on several standards. RDF is one of the most important standard for data description in Semantic Web. Transforming data from relational databases (RDB) to RDF is considered an important process for building the Semantic Web. In this article, the author will introduce a new approach allowing the conversion of relation database model to new data model using in Semantic Web, which is represented by RDF graph with Triples. The authors also experiments transferring a relation database of sale management system to a new Semantic Web's data model to examine article result.

Keywords: RDB, RDF, data, database, relation, semantic web.

1. Giới thiệu

John Naisbitt đã nói “Chúng ta đang chìm ngập trong thông tin nhưng lại khao khát tri thức”. Thật vậy, Web hiện tại đang chứa một lượng thông tin khổng lồ, được tạo ra từ các cá nhân, tổ chức và cộng đồng. Người sử dụng Web có thể dễ dàng truy cập thông tin thông qua địa chỉ URL hoặc các liên kết. Tính đơn giản của Web hiện tại dẫn đến một số hạn chế, chẳng hạn như chúng ta dễ dàng bị “lạc” vì đối mặt với một lượng thông tin khổng lồ có rất ít (hoặc không) liên quan từ kết quả tìm kiếm.

Web ngữ nghĩa còn gọi là Web của dữ liệu, được hình thành từ ý tưởng của Tim Berners-Lee người phát minh WWW. Web ngữ nghĩa là hướng phát triển mở rộng của Web hiện tại, mô hình dữ liệu cho Web ngữ nghĩa có thể khắc phục các hạn chế đã nêu.

* ThS, Trường Đại học Đồng Tháp; Email: ltnhoc@dthu.edu.vn

**ThS, Trường Đại học Đồng Tháp

Vì vậy, việc nghiên cứu nhằm đưa ra các giải pháp cho phép chuyển đổi dữ liệu từ các cơ sở dữ liệu quan hệ của Web hiện tại sang mô hình dữ liệu cho Web ngữ nghĩa đang được nhiều tác giả quan tâm.

Bizer C và cộng sự đề xuất D2RQ [3, 4] cho phép tạo môi trường tích hợp với nhiều tùy chọn để truy cập dữ liệu trong RDB dựa trên Jena và Sesame API được phát triển từ SQL và SPARQL trên D2RQ Server. Các ánh xạ dữ liệu được xác định bởi người dùng nên cho phép kết hợp ngữ nghĩa theo miền, và được khai báo bằng ngôn ngữ ánh xạ.

Cullot N và cộng sự đề xuất R₂O [2] là một ngôn ngữ cho phép thể hiện ánh xạ giữa RDB và ontology trên nền tảng XML. Ánh xạ R₂O được sử dụng có thể phát hiện mâu thuẫn và tính không rõ ràng trong lược đồ định nghĩa dữ liệu.

Soren A và cộng sự đề xuất Triplify [1, 9] cho phép chuyển sang RDF từ cơ sở dữ liệu quan hệ. Triplify sử dụng nguyên tắc ánh xạ các yêu cầu URI dựa trên RDB. Triplify cho phép chuyển đổi các mối quan hệ vào RDF và xuất bản dữ liệu RDF trên Web, cấu trúc gọn nhẹ nên có thể sử dụng để xuất bản bộ dữ liệu lớn.

R2RML [6] cho phép chuyển đổi trực tiếp từ RDB sang RDF, cấu trúc của đồ thị RDF phản ánh trực tiếp cấu trúc của cơ sở dữ liệu, từ vựng RDF phản ánh tên của phần tử trong lược đồ cơ sở dữ liệu. Ánh xạ trong R2RML phù hợp với lược đồ cơ sở dữ liệu và chỉ mục từ vựng, kết quả thể hiện dưới dạng đồ thị RDF.

Trong phần tiếp theo, bài báo sẽ trình bày cơ sở lý thuyết RDF và đồ thị RDF, phần 3 trình bày giải pháp cho phép thực hiện chuyển đổi từ RDB sang mô hình dữ liệu cho Web ngữ nghĩa, phần 4 là thực nghiệm và cuối cùng là kết luận.

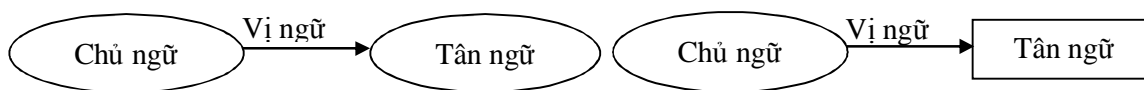
2. Đồ thị RDF

a) Giới thiệu về RDF

Khung mô tả tài nguyên (RDF) là tập hợp các nguyên tắc, cung cấp mô hình dữ liệu với cú pháp đơn giản dùng để mô tả tài nguyên. Đồng thời, RDF được thiết kế cho phép máy tính có thể hiểu và đọc được thông tin, chứ không chỉ đơn giản là để trình bày dữ liệu cho người dùng. Mọi tài nguyên đều có những thuộc tính và các thuộc tính đó đều có giá trị. [5]

Các tài nguyên có thể được mô tả thông qua các phát biểu cho phép xác định thuộc tính và giá trị. RDF sử dụng bộ ba (Triple) để nói về những thành phần khác nhau của phát biểu. Một bộ ba được mô tả dưới dạng nút-cung-nút, cung có hướng bắt đầu từ chủ ngữ đến tân ngữ (hình 1).

- **Chủ ngữ (Subject):** Là thành phần xác định đối tượng mà phát biểu đề cập;
- **Vị ngữ (Predicate):** Là thành phần xác định thuộc tính của chủ ngữ trong phát biểu;
- **Tân ngữ (Object):** Là thành phần xác định giá trị của thuộc tính, nó có thể là một tài nguyên (URI), hoặc một giá trị (Literal).

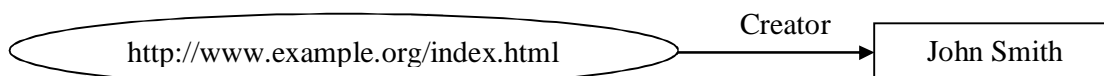


a) Tân ngữ (O) là một tài nguyên

b) Tân ngữ (O) là một giá trị

Hình 1. Mô hình biểu diễn bộ ba [5]

Ví dụ 1. Xét phát biểu *http://www.example.org/index.html has a creator whose value is John Smith*. Ta có chủ ngữ là *http://www.example.org/index.html*, vị ngữ là *creator* và tân ngữ là *John Smith*.



Hình 2. Bộ ba mô tả tài nguyên tại ví dụ 1 [7]

b) Đồ thị RDF

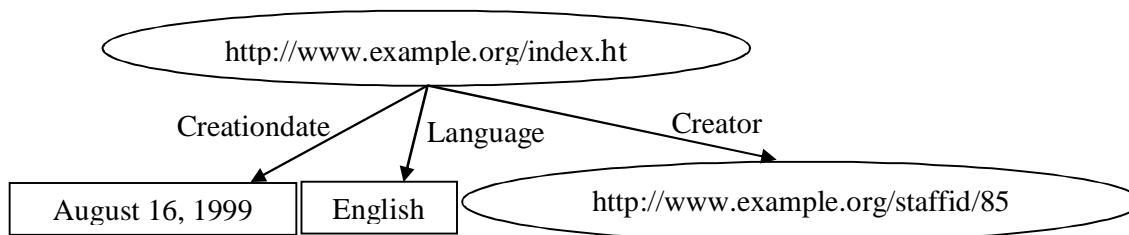
Tập các bộ ba hợp lại tạo thành đồ thị RDF, các nút trong đồ thị có thể là các chủ ngữ hoặc tân ngữ, các cung trong đồ thị là vị ngữ.

Ví dụ 2. Ta xét thêm các phát biểu *http://www.example.org/index.html has a creation-date whose value is August 16, 1999* và *http://www.example.org/index.html has a language whose value is English*.

Bảng 1. Bảng phân tích các thành phần của phát biểu

Subject	Predicate	Object
http://www.example.org/index.html	http://purl.org/dc/elements/1.1/creator	http://www.example.org/staffid/85
http://www.example.org/index.html	http://www.example.org/terms/creationdate	August 16, 1999
http://www.example.org/index.html	http://purl.org/dc/elements/1.1/language	English

Bảng 1 mô tả các thành phần trong phát biểu tại ví dụ 2. Đồ thị RDF tương ứng được trình bày trong hình 3.



Hình 3. Đồ thị RDF tương ứng ví dụ 2

3. Chuyển đổi từ CSDL quan hệ sang đồ thị RDF

a) Thuật toán RDB-to-RDF

Trong RDB, mỗi bộ tương ứng với một dòng của bảng gồm các giá trị tương ứng với mỗi cột. Khi thực hiện chuyển đổi sang đồ thị RDF ta có các lớp (Class) tương ứng với từng bảng, mỗi dòng trong bảng được biểu diễn bằng một phát biểu dưới dạng bộ ba (S, P, O) tạo thành đồ thị RDF, trong đó chủ đề (S) là giá trị khóa, các vị ngữ (P) tương ứng các nhãn của mỗi cột, tân ngữ (O) là giá trị tại cột tương ứng vị ngữ. Thuật toán 1 trình bày chi tiết chuyển đổi từ CSDL quan hệ sang đồ thị RDF.

Thuật toán 1: Chuyển đổi từ CSDL quan hệ sang đồ thị RDF.

Input: CSDL quan hệ RDB. **Output:** Đồ thị RDF.

BEGIN

```

1.   RDFGraph G = new RDFGraph();
      //Tạo một lớp RDF cho từng bảng
      For (int i=0; i < RDB.Tables.Count(); i++){
2.       Table T = RDB.Tables[i];
3.       C[i] = CreateClassRDF(T);
      //Mỗi dòng của bảng tương ứng với một Triple trong RDF
4.       For(int j=0; j < T.Rows.Count(); j++){
5.           Row R = T.Rows[j];
6.           Triple Tp = New Triple();
      //Chủ thể (S) của Tp là giá trị khóa của mỗi bộ
7.           Tp.Subject = R.id;
8.           For (int h=0; h < R.Cells.Count(); h++){
9.               If ((R.Cells[i].ColumnHead() != ID) &&
                  (R.Cells[i].ColumnHead() Not In R.GetForeignKeys())){
10.                  Predicate P = new Predicate(R.Cells[i].ColumnHead());
11.                  Tp.Predicate[h] = P;
12.                  Object O = new Object(R.Cells[i].Value());
13.                  Tp.Object[h] = O;
                }
      //Tạo Triple cho phụ thuộc hàm giữa 2 quan hệ
14.      If (R.Cells[i].ColumnHead() In R.GetForeignKeys()){

```

```

//Duyệt qua các bảng có quan hệ với T
15.   Foreach (Table tb in T.GetRelatedTables()){
16.       Predicate P =T.Name + "-" + tb.Name;
17.       Tp.Predicate[h] = P;
18.       String vl = R.Cells[i].ColumnHead();
19.       Triple Tem = FindTripleWithSubject(vl, tb);
20.       URI uri = (URI) Tem.Subject;
21.       Tp.Object[h] = uri;
           }
       }
           }
//Thêm bộ ba vào đồ thị
22.   G.AddTriple(Tp);
           }
}
23.   Return (G);
END

```

b) Thảo luận

Vấn đề chuyển đổi từ RDB sang mô hình dữ liệu cho Web ngữ nghĩa được nhiều tác giả quan tâm nghiên cứu, đã có một số hướng tiếp cận được đề xuất trong thời gian qua, chi tiết của các giải pháp được tổng hợp trong bảng 2.

Bảng 2. Một số nghiên cứu đã công bố trong thời gian qua [8]

Giải pháp	T₁	T₂	T₃	T₄	T₅
D2RQ	•	•	•	•	•
R ₂ O	•		•	•	
Triplify	•	•	•	•	•
R2RML	•	•	•	•	•
RDB-to-RDF	•		•		•

Trong đó:

- Bảng thành lớp (T₁): Đặc điểm này cho phép chuyển đổi các bảng trong RDB thành các lớp trong từ vựng RDF;

- Mỗi quan hệ giữa các bảng (T_2): Mỗi quan hệ giữa các bảng trong RDB được biểu diễn dưới dạng 1-1, 1-n hoặc n-n. Đặc điểm này cho phép chuyển đổi mỗi quan hệ thành các thuộc tính trong từ vựng RDF;
- Truy vấn SPASQL (T_3): Hỗ trợ ngôn ngữ truy vấn có tính chất đồ thị, đây là một chuẩn để truy vấn dữ liệu RDF;
- Kiểu dữ liệu (T_4): Đặc điểm này cho phép thiết lập kiểu dữ liệu tương ứng của tên ngữ trong từ vựng RDF.
- Linked Data (T_5): Đặc điểm này cho phép tạo ra các liên kết định kiểu giữa các dữ liệu từ nhiều nguồn tài nguyên khác nhau.

4. Thực nghiệm

Bài báo tiến hành thực nghiệm trên CSDL quan hệ Quản lí bán hàng (QLBH). Khóa chính của các quan hệ là các thuộc tính in đậm và gạch chân (hình 4), dữ liệu trong các quan hệ được lưu trữ như hình 5.

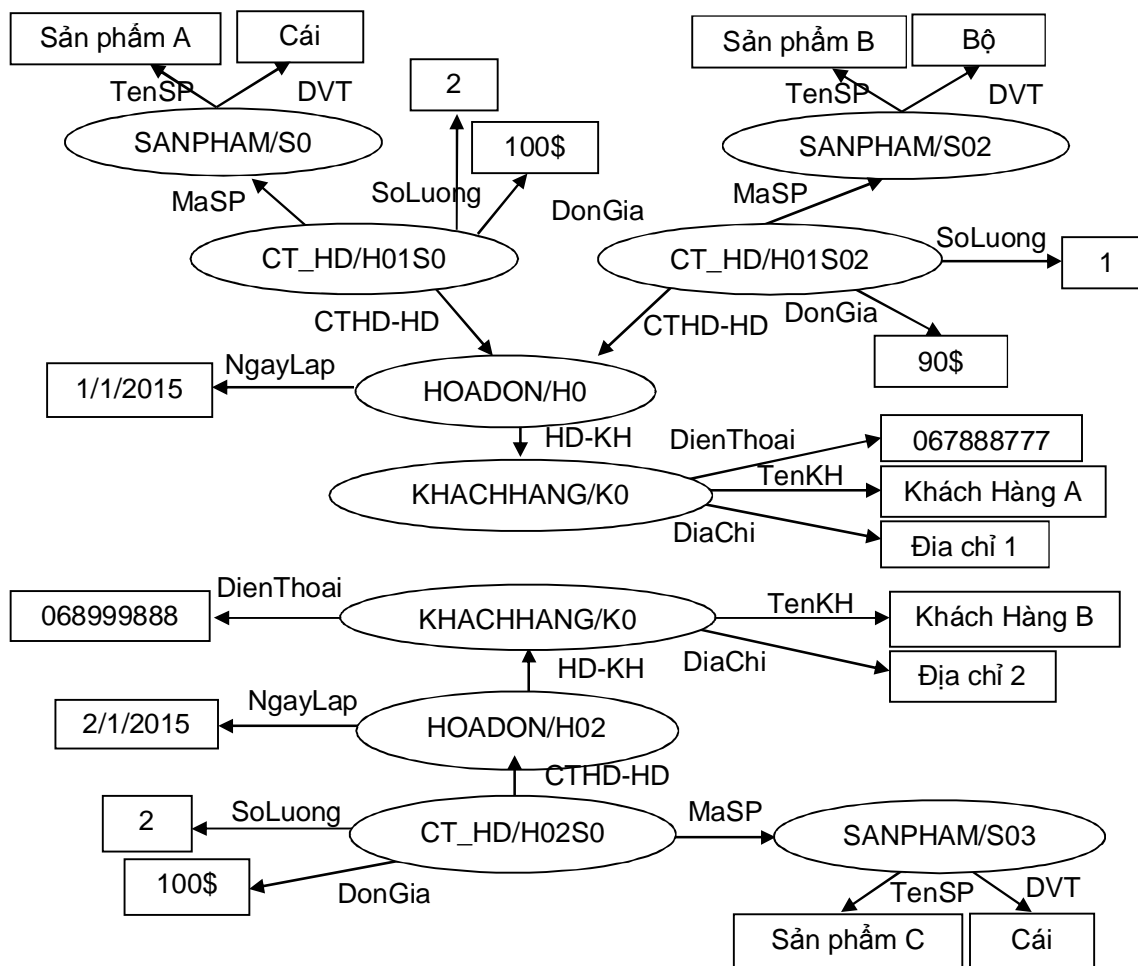
KHACHHANG (<u>MaKH</u> , TenKH, DiaChi, DienThoai)
SANPHAM (<u>MaSP</u> , TenSP, DVT)
HOADON (<u>SoHD</u> , MaKH, NgayLap)
CHITIET_HD (<u>SoHD</u> , <u>MaSP</u> , SoLuong, Dongia)

Hình 4. Lược đồ CSDL quan hệ QLBH

MaKH	TenKH	DiaChi	DienThoai	MaSP	TenSP	DVT
K01	Khách Hàng A	Địa chỉ 1	067888777	S01	Sản phẩm A	Cái
K02	Khách Hàng B	Địa chỉ 2	068999888	S02	Sản phẩm B	Bộ
SoHD	MaSP	SoLuong	DonGia	S03	Sản phẩm C	Cái
H01	S01	2	100\$	SoHD	NgayLap	MaKH
H01	S02	1	90\$	H01	1/1/2015	K01
H02	S03	2	110\$	H02	2/1/2015	K02

Hình 5. Chi tiết dữ liệu lưu trữ trong các quan hệ

Từ CSDL quan hệ Quản lí bán hàng tại hình 5, tác giả tiến hành chuyển đổi sang đồ thị RDF dựa trên thuật toán 1 đã trình bày, kết quả thu được như hình 6.



Hình 6. Đồ thị RDF tương ứng với CSDL quan hệ QL BH

Thực hiện chuyển đổi từ đồ thị RDF (hình 6) sang mô hình dữ liệu cho Web ngữ nghĩa được biểu diễn dưới dạng các bộ ba. Kết quả được lưu trữ như trong bảng 3 với QName ex = <http://www.dthu.edu.vn/qlbh/>.

Bảng 3. Bảng lưu trữ dữ liệu RDF hiển thị dưới dạng bộ ba

Subject	Predicate	Object
ex/sanpham/S01	ex/sanpham#masp	S01
ex/sanpham/S01	ex/sanpham#tensp	Sản Phẩm A
ex/sanpham/S01	ex/sanpham#dvt	Cái
ex/sanpham/S02	ex/sanpham#masp	S02
ex/sanpham/S02	ex/sanpham#tensp	Sản Phẩm B

ex/sanpham/S02	ex/sanpham#dvt	Bộ
ex/sanpham/S03	ex/sanpham#masp	S03
ex/sanpham/S03	ex/sanpham#tensp	Sản Phẩm C
ex/sanpham/S03	ex/sanpham#dvt	Cái
ex/khachhang/K01	ex/khachhang#makh	K01
ex/khachhang/K01	ex/khachhang#tenkh	Khách Hàng A
ex/khachhang/K01	ex/khachhang#diachi	Địa chỉ 1
ex/khachhang/K01	ex/khachhang#dienthoai	067888777
ex/khachhang/K02	ex/khachhang#makh	K02
ex/khachhang/K02	ex/khachhang#tenkh	Khách Hàng B
ex/khachhang/K02	ex/khachhang#diachi	Địa chỉ 2
ex/khachhang/K02	ex/khachhang#dienthoai	068999888
ex/hoadon/H01	ex/hoadon#sohd	H01
ex/hoadon/H01	ex/hoadon#ngaylap	1/1/2015
ex/hoadon/H01	ex/hoadon#makh	ex/khachhang/K01
ex/hoadon/H02	ex/hoadon#sohd	H02
ex/hoadon/H02	ex/hoadon#ngaylap	2/1/2015
ex/hoadon/H02	ex/hoadon#makh	ex/khachhang/K02
ex/chitiethd/H01S01	ex/chitiethd#sohd	ex/hanghoa/H01
ex/chitiethd/H01S01	ex/chitiethd#masp	ex/sanpham/S01
ex/chitiethd/H01S01	ex/chitiethd#soluong	2
ex/chitiethd/H01S01	ex/chitiethd#dongia	100\$
ex/chitiethd/H01S02	ex/chitiethd#sohd	ex/hanghoa/H01
ex/chitiethd/H01S02	ex/chitiethd#masp	ex/sanpham/S02
ex/chitiethd/H01S02	ex/chitiethd#soluong	1
ex/chitiethd/H01S02	ex/chitiethd#dongia	90\$
ex/chitiethd/H02S03	ex/chitiethd#sohd	ex/hanghoa/H02
ex/chitiethd/H02S03	ex/chitiethd#masp	ex/sanpham/S03
ex/chitiethd/H02S03	ex/chitiethd#soluong	2
ex/chitiethd/H02S03	ex/chitiethd#dongia	110\$

5. Kết luận

Như vậy, bài báo đã trình bày một phương pháp cho phép chuyển đổi từ CSDL quan hệ sang mô hình dữ liệu cho Web ngữ nghĩa. Qua thực nghiệm, ta thấy hướng tiếp cận này cho phép chuyển đổi dữ liệu trong RDB sang đồ thị RDF, đồng thời biểu diễn dữ liệu từ đồ thị RDF sang mô hình dữ liệu cho Web ngữ nghĩa dưới hình thức các bộ ba. Vấn đề chuyển đổi từ RDB sang RDFs nhằm thỏa mãn các ràng buộc toàn vẹn dữ liệu như trong RDB là hướng nghiên cứu tiếp theo của tác giả trong tương lai.

Ghi chú: Nghiên cứu này được hỗ trợ bởi đề tài có mã số CS2015.01.27 do tác giả thực hiện tại Trường Đại học Đồng Tháp.

TÀI LIỆU THAM KHẢO

1. Auer S, Dietzold S, Lehmann J, Hellmann S, Aumueller D (2009), "Triplify – LightWeight Linked Data Publication from Relational Databases", *In Proceedings of the 18th International World Wide Web Conference*, ACM 978-1-60558-487-4/09/04, 621-630.
2. Barrasa J, Corcho O, Gomez-Perez A (2004), "R₂O an Extensible and Semantically Based Database-to-Ontology Mapping Language", *In Proceedings of the 2nd Workshop on Semantic Web and Databases, Toronto, Canada*.
3. Bizer C, Cyganiak R (2007), "D2RQ - Lessons Learned", *Position paper for the W3C Workshop on RDF Access to Relational Databases, Cambridge (USA)*, 25-36.
4. Bizer C, Cyganiak R, Garbers J, Maresch O, Becker C (2009), "The D2RQ Platform v0.7 - Treating Non-RDF Relational Databases as Virtual RDF Graphs", *On the Move to Meaningful Internet Systems: OTM 2010 Workshops*.
5. Blakeley C (2007), *RDF Views of SQL Data (Declarative SQL Schema to RDF Mapping)*, OpenLink Software.
6. Das S, Sundara S, Cyganiak R(2010), "R2RML: RDB to RDF Mapping Language", *W3C Working*. <http://www.w3.org/TR/2010/WD-r2rml-20101028/>.
7. Java, JRuby, Scala, and Clojure (2011), *Practical Semantic Web and Linked Data Applications*, Mark Watson.
8. Matthias H, Gerald R, Harald C G (2011), "A Comparison of RDB-to-RDF Mapping Languages", *Conf. on Semantic Systems Austria, ACM 978-1-4503-0621-8*.
9. Soren A, Sebastian D, Jens L (2009), "Triplify – Light Weight Linked Data Publication from Relational Databases", *IW3C2 Madrid (Spain)*, ACM 978-1-60558-487-4/09/04.

(Ngày Tòa soạn nhận được bài: 14-8-2015; ngày phản biện đánh giá: 09-9-2015;
ngày chấp nhận đăng: 24-9-2015)