# 1. Bối cảnh

## 1.1 Giới thiệu web ngữ nghĩa và Ontology

Semantic Web, đôi khi được gọi là Web 3.0, là một phần mở rộng của World Wide Web thông qua các tiêu chuẩn do World Wide Web Consortium (W3C) đặt ra. Mục tiêu của Semantic Web là làm cho dữ liệu Internet có thể đọc được bằng máy.

Web ngữ nghĩa được phát triển dựa trên Ontology. Bản thể học [[1]](#footnote-1)(hoặc từ vựng) xác định các khái niệm và mối quan hệ (còn được gọi là "thuật ngữ") được sử dụng để mô tả và đại diện cho một lĩnh vực quan tâm. Từ vựng được sử dụng để phân loại các thuật ngữ có thể được sử dụng trong một ứng dụng cụ thể, mô tả các mối quan hệ có thể có và xác định các ràng buộc có thể có đối với việc sử dụng các thuật ngữ đó. Trong thực tế, các từ vựng có thể rất phức tạp (với vài nghìn thuật ngữ) hoặc rất đơn giản (chỉ mô tả một hoặc hai khái niệm)

Ontology cũng có thể hiểu là **mô hình dữ liệu.**

Vai trò của bản thể học là giúp tích hợp dữ liệu khi, ví dụ, có thể tồn tại sự mơ hồ đối với các thuật ngữ được sử dụng trong các tập dữ liệu khác nhau hoặc khi một chút kiến thức bổ sung có thể dẫn đến việc khám phá các mối quan hệ mới

Tham khảo: <https://arxiv.org/pdf/1006.4567.pdf>

## 1,2 Ba

Bộ ba [[2]](#footnote-2)là một bộ ba thực thể mã hóa một câu lệnh về dữ liệu ngữ nghĩa dưới dạng biểu thức **chủ ngữ – vị ngữ – đối tượng** (ví dụ: "Bob 35 tuổi" hoặc "Bob biết John"). Định dạng này cho phép kiến thức được trình bày theo cách mà máy có thể đọc được.

Bộ ba có thể được biểu diễn ở nhiều định dạng [[3]](#footnote-3)như RDF / XML, RDFa, Notation3 [[4]](#footnote-4)và Turtle [[5]](#footnote-5).

**rùa** điển hình :

Vị ngữ chủ ngữ1 tân ngữ;

vị ngữ2 đối tượng2;

vị ngữ3 tân ngữ3, tân ngữ4.

Ví dụ, trong bộ ba sau

< <http://www.example.org/index.html>> <hasCreator> "John Smith".

* Chủ đề là trang web **http://www.example.org/index.html**
* Vị từ là thuộc tính **hasCreator**
* Đối tượng là Chuỗi **"John Smith"** .

## 1.3 Các thành phần Ontology

Ontology bao gồm ba thành phần chính: khái niệm (hoặc lớp), thể hiện và thuộc tính

Khái niệm (còn được gọi là một lớp hoặc một thuật ngữ) là một nhóm, tập hợp hoặc tập hợp các đối tượng trừu tượng. Nó là phần tử cơ bản của miền và thường đại diện cho một nhóm hoặc lớp có các thành viên chia sẻ các thuộc tính chung. Thành phần này được biểu diễn dưới dạng đồ thị phân cấp, sao cho nó trông tương tự như **các hệ thống hướng đối tượng** . Khái niệm được đại diện bởi một “siêu lớp”, đại diện cho lớp cao hơn hoặc cái gọi là “lớp cha” và “lớp con” đại diện cho cấp dưới hoặc cái gọi là “lớp con”.

Các cá thể (cá thể hoặc đối tượng) là các thành phần cơ bản, "mức cơ bản" của một bản thể luận. Các cá thể trong bản thể luận có thể bao gồm các đối tượng cụ thể như người, động vật, bàn, ô tô, phân tử và hành tinh. Nói một cách chính xác, bản thể luận không cần bao gồm bất kỳ cá nhân nào, nhưng một trong những mục đích chung của bản thể luận là cung cấp một phương tiện phân loại các cá thể, ngay cả khi những cá nhân đó không rõ ràng là một phần của bản thể luận.

Các cá nhân có tài sản, để lưu trữ thông tin của họ. Có hai loại thuộc tính:

**Thuộc tính dữ liệu**

Thuộc tính dữ liệu kết nối các cá nhân với các chữ. Trong một số hệ thống biểu diễn tri thức, các thuộc tính dữ liệu chức năng được gọi là thuộc tính. Ví dụ về các thuộc tính dữ liệu là: hasName, hasPhoneNumber, v.v.

**Thuộc tính đối tượng**

Thuộc tính đối tượng kết nối hai cá nhân. Ví dụ về các thuộc tính đối tượng là isTaughtBy, giám sát, v.v.

Mỗi thuộc tính có miền và phạm vi

* **Miền thuộc tính** s: chỉ định các lớp mà thuộc tính đó thuộc về
* **Phạm vi thuộc tính** : chỉ định giá trị có thể có của thuộc tính

Ví dụ:

: age a owl: DatatypeProperty;

rdfs: miền foaf: Person;

rdfs: phạm vi xsd: nonNegativeInteger.

: isTaughtBy a owl: ObjectProperty;

rdfs: domain: Course;

rdfs: range: AcademicStaffMember;

rdfs: subPropertyOf: liên quan đến.

## 1.4 Bản thể học Du lịch Việt Nam

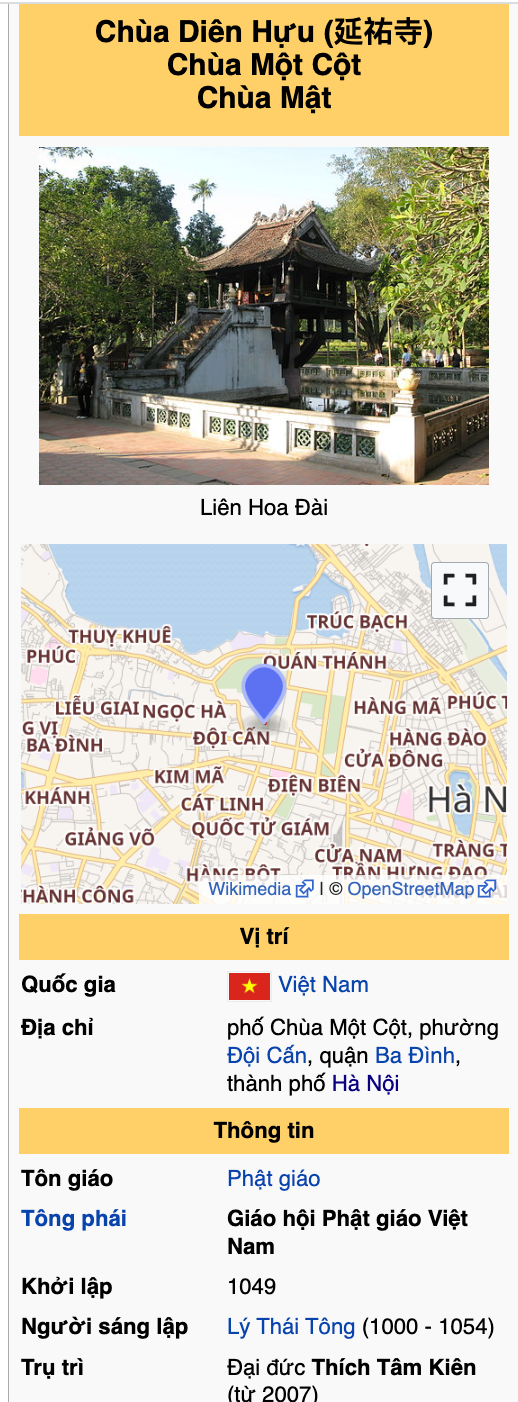
Bản thể học du lịch Việt Nam là một đại diện thống nhất và chung của dữ liệu trong lĩnh vực du lịch. Nó liên quan đến nhiều chủ đề lịch sử và văn hóa như một đặc điểm tự nhiên của du lịch.

Link tải: <https://drive.google.com/drive/folders/1QGoKV4hWqQzi2LGo8T8Fu30LzZPc3nwL?usp=sharing>

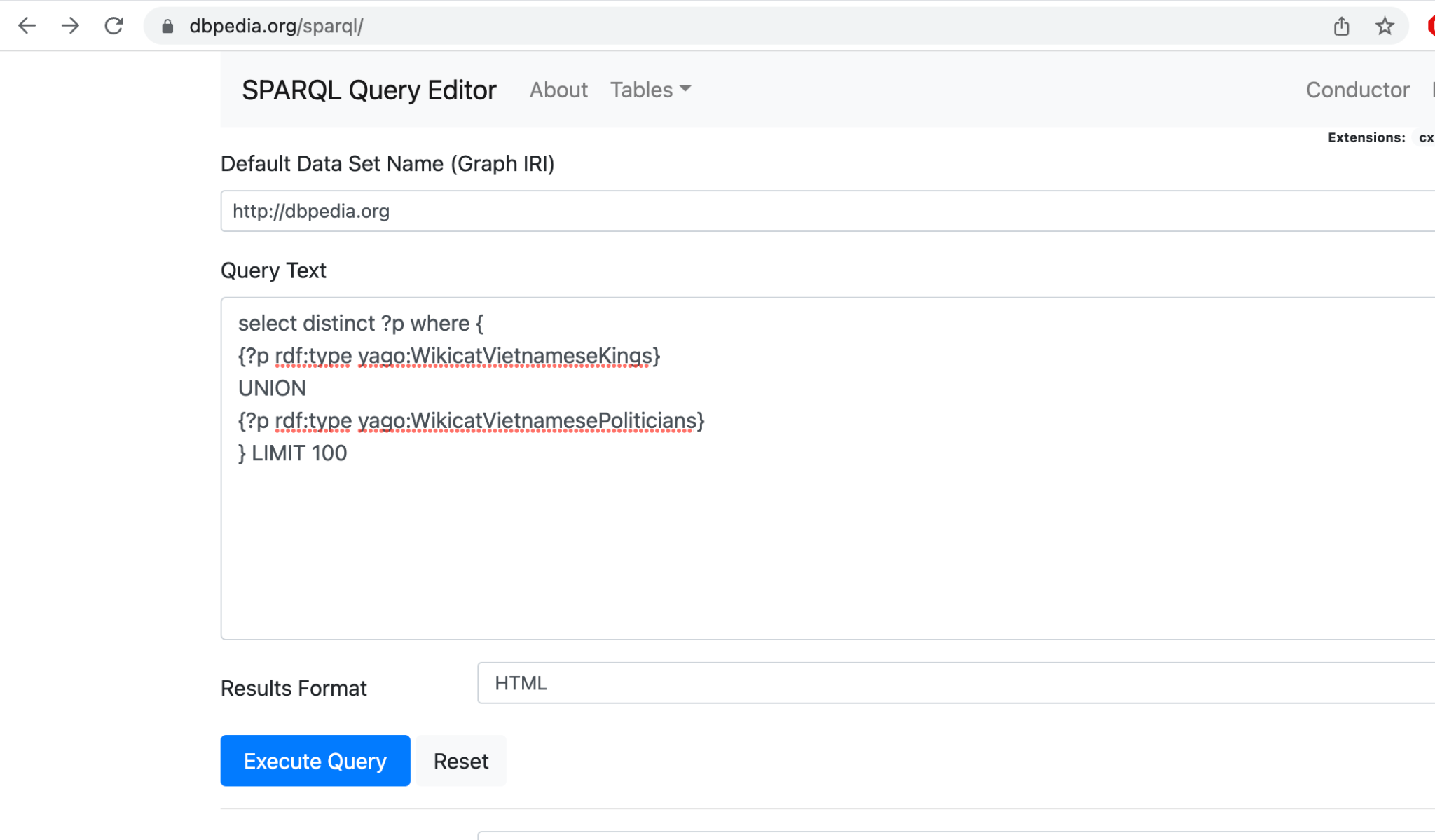
Bản thể học có thể được mở bằng phần mềm **protege** <https://protege.stanford.edu/>

## 1.5 Các trang web để thu thập dữ liệu Du lịch

Có rất nhiều trang web nơi chúng tôi có thể thu thập dữ liệu du lịch. Ví dụ như Wikipedia, Wikidata.org, DBpedia.org, và các trang web của sở văn hóa và du lịch ở nhiều tỉnh thành ở Việt Nam.

Ví dụ: sử dụng Wikipedia, bạn có thể phân tích cú pháp dữ liệu được trình bày trong hộp thông tin.<https://vi.wikipedia.org/wiki/Ch%C3%B9a_M%E1%BB%99t_C%E1%BB%99t>

Sử dụng truy vấn sau với điểm cuối DBPedia SPARQL, chúng ta có thể tìm được những người nổi tiếng ở Việt Nam. Truy cập <https://www.w3.org/TR/rdf-sparql-query/>để tìm hiểu SPARQL.



chọn khác biệt? p ở đâu {

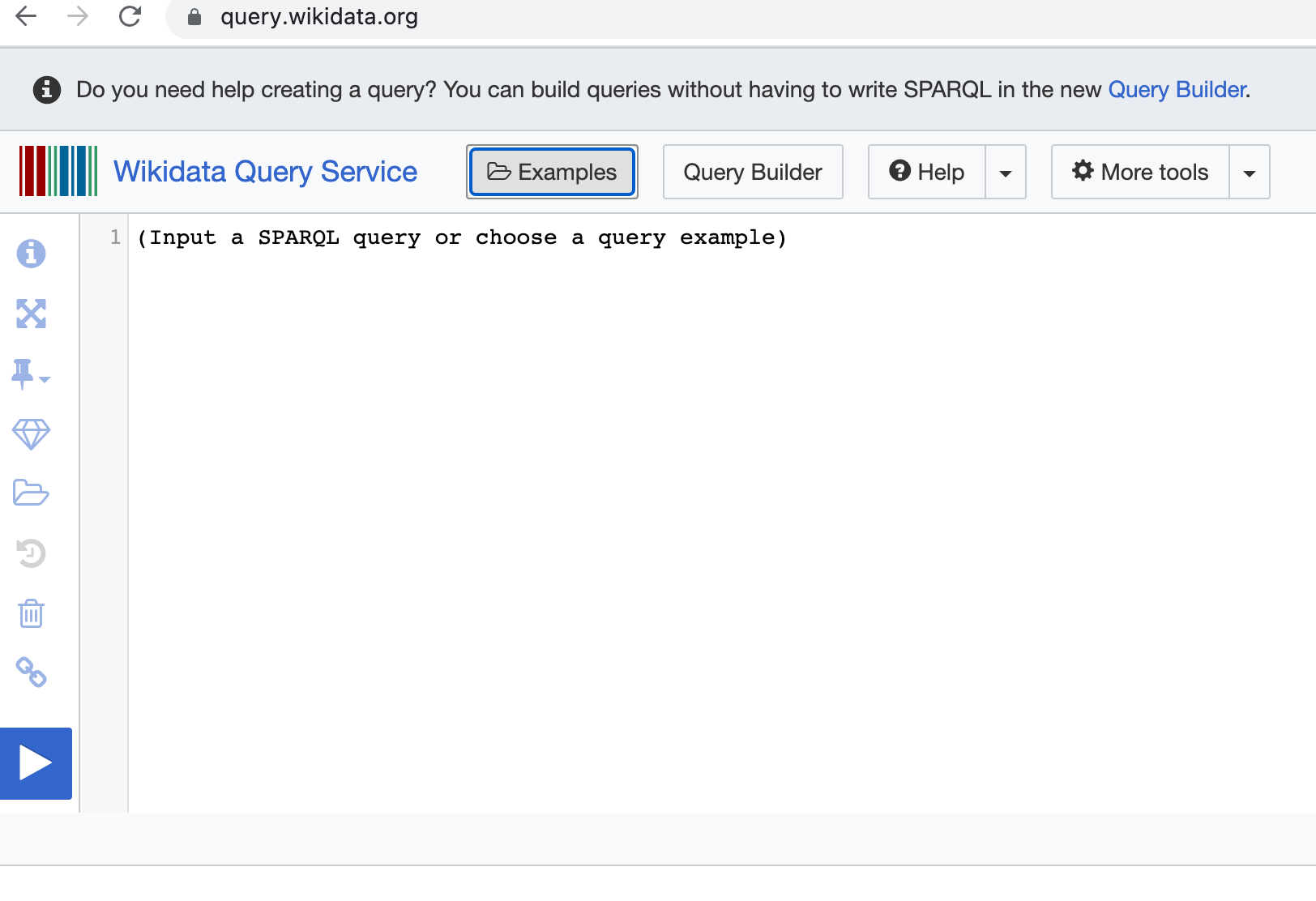
{? p rdf: gõ yago: WikicatVietnameseKings}

LIÊN HIỆP

{? p rdf: gõ yago: WikicatVietnamesePoliticians}

} GIỚI HẠN 100

Tương tự, bạn có thể sử dụng điểm cuối SPARQL từ Wikidata



Nhiều trang web khác cung cấp dữ liệu về du lịch, nhưng dữ liệu không có cấu trúc. Ví dụ là

* Hoidulich.net
* Khamphadisan.com
* http://lehoi.info/

2 Mô tả vấn đề

Bạn cần hiểu bản thể học du lịch, thu thập dữ liệu du lịch từ các trang web du lịch và lưu trữ dữ liệu trong tệp sử dụng [[6]](#footnote-6)định dạng rùa. Mỗi chủ đề cần được lưu trữ trong một tệp riêng biệt.

Bạn sẽ đạt điểm cao nếu thu thập được một lượng lớn dữ liệu về các chủ đề khác nhau và dữ liệu phải chính xác.

Ví dụ về đại diện rùa:

@prefix vntourism: <http://www.semanticweb.org/minhn/ontologies/2021/0/vntourism#>.

vntourism: Cổ\_Loa\_Citadel rdf: gõ cú: NamedIndividual,

vntourism: CitadelArchitecture;

vntourism: ChooseCapitalBy vntourism: An\_Dương\_Vương,

vntourism: Ngô\_Quyền,

vntourism: Ngô\_Xương\_Văn;

vntourism: hasAdministrativeDivision vntourism: Hà\_Nội;

vntourism: hasBuildBy vntourism: An\_Dương\_Vương;

vntourism: hasBuildTime "thế kỷ thứ 3" @en;

vntourism: hasCountry vntourism: AuLac;

vntourism: lễ hội vntourism: Lễ\_hội\_Cổ\_Loa;

vntourism: hasHistoricalSiteLevel vntourism: SepcialNationalLevel;

vntourism: hasPeriod vntourism: NgoDyosystem;

rdfs: label "Cổ Loa Citade" @en,

"Thành Cổ Loa" @vn.

: Ly\_Thuong\_Kiet\_Working\_TimeLine

rdf: gõ cú: NamedIndividual,: TemporalEntity;

: hasBeginning [

rdf: type: Instant;

: inTimePosition [

rdf: type: TimePosition;

: hasJob: Kinh\_Phong\_Su\_Lý\_Thường\_Kiệt;

: năm "1061" ^^ xsd: gYear

]

];

: hasDurationDescription [

rdf: type: GeneralDurationDescription;

: năm "8" ^^ xsd: gYear

];

: hasEnd [

rdf: type: Instant;

: inTimePosition [

rdf: type: TimePosition;

: năm "1069" ^^ xsd: gYear

]

].

1. <https://www.w3.org/standards/semanticweb/ontology> [↑](#footnote-ref-1)
2. <https://en.wikipedia.org/wiki/Semantic_triple> [↑](#footnote-ref-2)
3. <https://ontola.io/blog/rdf-serialization-formats/> [↑](#footnote-ref-3)
4. <https://www.w3.org/TeamSubmission/n3/> [↑](#footnote-ref-4)
5. Https://www.w3.org/TR/turt [le /](https://www.w3.org/TR/turtle/) [↑](#footnote-ref-5)
6. Https://www.w3.org/TR/turtle/ [↑](#footnote-ref-6)