

UNIVERSITY OF INFORMATION AND TECHNOLOGY

Báo cáo khóa luận giai đoạn 8-10/2014

hoàn thành bởi

với sự hướng dẫn của

TS. Nguyễn Anh Tuấn

Đề tài khóa luận

tốt nghiệp - sinh viên khóa MMTT2010

thuộc

Chuyên ngành phát triển ứng dụng Web và di động

Khoa mạng máy tính và truyền thông

Ngày 25 tháng 10 năm 2014

Mục lục

Abbreviations	ii
Symbols	iii
1 Các nguyên nhân dẫn đến tính thiếu nhất quán trong ontology	1
1.1 Các định nghĩa cần lưu ý ^[3]	1
Unsatisfiable Class/Concept	1
Incoherent Ontology	2
Inconsistent Ontology	2
1.2 Các nguyên nhân phổ biến dẫn đến tính thiếu nhất quán (Inconsistency) ^[4]	2
1.2.1 Khởi tạo cá thể cho một Unsatisfiable Class) - (TBox + ABox)	2
1.2.2 Khởi tạo cá thể thuộc 2 class được disjoint với nhau (TBox + ABox)	3
1.2.3 Các phát biểu ABox xung đột với nhau	3
1.2.4 Phát biểu xung đột với nghĩa "oneOf" (All TBox)	3
1.2.5 Không có khả năng khởi tạo bất kì cá thể nào (all TBox)	4
Kết luận	4
2 Giải pháp để sửa chữa inconsistent ontology	6
2.1 Mục tiêu của việc debugging ontology	7
2.2 Khái niệm và các kỹ thuật cần biết	7
Axiom Pinpointing service	7
2.3 Các bước sửa chữa các phát biểu bị lỗi	8
2.3.1 Minimal Unsatisfiability Preserving Sub-TBoxes (MUPS)	8
2.3.2 Tìm MUPS của một lớp không đáp ứng	8
Axiom Pinpointing service	8
2.3.3 Chiến thuật xếp hạng các phát biểu (<i>Axioms</i>)	9
2.3.4 Đưa ra các nguyên nhân gây lỗi chính xác bằng giải thuật HST của Reiter	9
Tìm MUPS bằng giải thuật HST của Reiter	9
3 Tài liệu tham khảo	11

Abbreviations

KB	K nowledge B ase
KR	K nowledge R epresentation
DL	D escription L ogic
SHOIN	

Symbols

\models	models of	có nghĩa trong (KB)
\subseteq	is a subset of	là tập con của
\cap	intersect	giao với
$\neg A$	complement of A of	không phải A
$\exists R.E$	e.g <i>has</i> some E	
$\forall R.E$	e.g <i>has</i> only E	

For/Dedicated to/To my...

Chương 1

Các nguyên nhân dẫn đến tính thiếu nhất quán trong ontology

1.1 Các định nghĩa cần lưu ý^[3]

Unsatisfiable Class/Concept dùng để chỉ một lớp hay một khái niệm trong một ontology mà ngữ nghĩa xung đột với ngữ nghĩa khác được nêu ra trong ontology hay có thể nói là các phát biểu về lớp hay khái niệm này mâu thuẫn với nhau hoặc mâu thuẫn với những phát biểu khác trong ontology.

Ví dụ Cow

```
SubClassOf: Vegetarian
Vegetarian
SubClassOf: Animal and eats only Plant
DisjointClasses:
Plant, Animal
```

Giải thích Trong ví dụ trên thì MadCow chính là một lớp không hợp lý do trong các phát biểu logic của nó mâu thuẫn với nhau Cow là lớp con của Vegetarian mà Vegetarian chỉ ăn Plant (nghĩa là ngoài Plant, Vegetarian không ăn thứ gì khác) trong khi đó khai báo của lớp MadCow là lớp con của Cow và ăn một số Sheep (Sheep là một lớp con của Animal).

Từ đó việc lý luận có thể đưa ra giả định sai là Sheep cũng có khả năng là một phần của Plant. Điểm quan trọng là Plant và Animal là 2 DisjointClasses, nói cách khác không tồn tại một cá thể nào vừa thuộc lớp Plant và vừa thuộc lớp Animal. Như vậy trong tất cả các phát biểu logic ở ví dụ trên đã có 2 phát biểu

gây mâu thuẫn chính là `eats only Plant` và `eats some Sheep`, và chúng làm cho lớp `MadCow` trở nên bất hợp lý (*unsatisfiable*).

Incoherent Ontology dùng để chỉ một *ontology/model* có ý nghĩa không mạch lạc rõ ràng do nó có chứa ít nhất một *Unsatisfiable Class/Concept* và với điều kiện là trong những *Unsatisfiable Class* này không được chứa bất kì một cá thể (*Individual*) nào.

Giả sử ta có ontology A chứa các phát biểu trong ví dụ trên ngoại trừ phát biểu cuối cùng `Individual: Dora type: MadCow` thì ta có thể nói ontology A không mạch lạc rõ ràng do nó chứa *unsatisfiable class* là `MadCow`. Chúng ta vẫn có sử dụng được ontology A vì nó vẫn còn tính nhất quán (*Consistency*) miễn là không có phần tử nào thuộc lớp `MadCow`.

Inconsistent Ontology dùng để chỉ một ontology chứa ít nhất một *Unsatisfiable Class* và có ít nhất một cá thể (*Individual*) thuộc một trong những lớp *unsatisfiable* này. Như đã thể hiện trong ví dụ đầu tiên thì cá thể `Dora` thuộc lớp `MadCow` (Một lớp *unsatisfiable* thì không nên phép có bất kì cá thể nào nếu như chúng ta muốn đảm bảo tính *consistency* cho ontology), như vậy bất kì ontology nào có những phát biểu trên đều được coi là không nhất quán (*inconsistency*), điều này đồng nghĩa là ontology đó không thể sử dụng được nữa.

1.2 Các nguyên nhân phổ biến dẫn đến tính thiếu nhất quán (Inconsistency)^[4]

Các nguyên nhân dẫn đến tính thiếu nhất quán trong ontology gây bởi các lỗi được phân loại thành lỗi gây ra bởi phát biểu ở mức độ lớp (Class level - TBox), các lỗi gây ra bởi phát biểu ở mức độ cá thể (Instance/Individual level - ABox) và lỗi gây ra bởi sự kết hợp của cả 2 nguyên nhân vừa nêu trên.

1.2.1 Khởi tạo cá thể cho một Unsatisfiable Class) - (TBox + ABox)

- Khởi tạo cá thể cho một *Unsatisfiable Class* được xem là nguyên nhân phổ biến nhất gây ra tính thiếu nhất quán trong ontology.
- Ví dụ:

`Individual: Dora type: MadCow`

- Chúng ta không quan tâm đâu là nguyên nhân làm cho MadCow trở nên mâu thuẫn, chỉ cần biết là một Unsatisfiable Class thì không nên có bất kì cá thể nào trong đó. Rõ ràng là không có bất kì ontology nào mà cá thể Dora có thể đáp ứng các điều kiện như trong ví dụ đầu tiên, nói cách khác không tồn tại model nào có thể thỏa được điều kiện trên. Chúng ta phát biểu đó là một ontology không nhất quán.

1.2.2 Khởi tạo cá thể thuộc 2 class được disjoint với nhau (TBox + ABox)

- Đây là một trường hợp dễ bắt gặp vì nó sai ngay trong phát biểu về logic.
- Ví dụ

Individual: Dora

Types: Vegetarian, Carnivore

DisjointClasses: Vegetarian, Carnivore

- Lớp A disjoint với B khi và chỉ khi lớp A không có chung bất kì một phần tử/cá thể nào với lớp B. Phát biểu Disjoint Classes(A B C) có nghĩa là mỗi lớp trong đó disjoint với từng lớp còn lại (mutually disjoint). Phát biểu ABox dạng DisjointClasses(Vegetarian Carnivore) là sai vì Dora vừa thuộc Vegetarian vừa thuộc Carnivore dựa vào phát biểu Individual: Dora Types: Vegetarian, Carnivore.

1.2.3 Các phát biểu ABox xung đột với nhau

- Trường hợp này thì tương tự như nguyên nhân ở trên nhưng khác ở chỗ là lần này sự mâu thuẫn nằm trong các biểu ở cấp độ cá thể (ABox).
- Ví dụ:

Individual: Dora

Types: Vegetarian, not Vegetarian

- Dễ thấy được sự mâu thuẫn trong trong phát biểu trên vừa yêu cầu Dora là Vegetarian vừa yêu cầu nó không phải Vegetarian.

1.2.4 Phát biểu xung đột với nghĩa "oneOf" (All TBox)

- Phát biểu bao gồm hoặc một trong (oneOf trong cú pháp của OWL) cho phép sử dụng các cá thể trong khai báo phát biểu ABox, sự kết hợp này có thể dẫn đến sự thiếu nhất quán.

- Lấy ví dụ sau:

```
Class: MyFavouriteCow
    EquivalentTo: {Dora}
Class: AllMyCows
    EquivalentTo: {Dora, Daisy, Patty}
DisjointClasses: MyFavouriteCow, AllMyCows
```

- Phần đầu tiên của các phát biểu trên tất cả các thể thuộc lớp **MyFavouriteCow** phải tương đương với cá thể tên Dora, nói cách khác là **SameIndividual** với Dora. Phần thứ hai cũng tương tự tất cả các cá thể thuộc lớp **AllMyCows** buộc phải tương đương với một trong 3 cá thể tên Dora, Daisy hoặc Patty. Do 2 phát biểu trên chúng ta đã nói Dora thuộc cả 2 lớp **MyFavoriteCow** và **AllMyCows** nên mâu thuẫn với phát biểu cuối cùng khi nói 2 lớp này không có chung một cá thể nào. Vì vậy dẫn tới ontology bị thiếu nhất quán (inconsistent).

1.2.5 Không có khả năng khởi tạo bất kì cá thể nào (all TBox)

- Ví dụ:

```
Vegetarian or not Vegetarian
SubClassOf: Cow and not Cow
```

- Đây chỉ là một ví dụ đơn giản để minh họa cho trường hợp này. Thực tế sẽ ít người dùng nào tạo ra một phát biểu ngớ ngẩn như vậy nhưng nó vẫn có khả năng xảy ra khi phát biểu trên là kết quả từ suy luận (reasoning) của những phát biểu lớn và phức tạp hơn.
- Có thể giải thích ví dụ trên như sau. Đầu tiên để đáp ứng ý nghĩa dòng đầu tiên yêu cầu cá thể vừa là **Vegetarian** hoặc không phải **Vegetarian** - bất kỳ phát biểu nào dạng này, "cá thể thuộc hoặc không thuộc một lớp" chính là tất cả cá thể xuất hiện trong ontology. Dòng thứ hai yêu cầu cá thể vừa là Cow vừa không phải là Cow, phát biểu này rơi vào một trong các nguyên nhân vừa nêu ở trên. Tổng hợp lại chúng yêu cầu tất cả cá thể vừa là Cow vừa không phải Cow, điều này gây ra mâu thuẫn trên toàn ontology do phát biểu đầu tiên chỉ tới tất cả các cá thể.

Kết luận Trên đây chúng em đã liệt kê những nguyên nhân phổ biến dẫn đến thiếu nhất quán qua những ví dụ đã được đơn giản hoá để dễ dàng nắm bắt được đâu là căn nguyên gây ra sự mâu thuẫn về logic. Trên thực tế với những ontology có số lượng phát biểu lớn và phức tạp rất khó để người dùng có thể nhận diện được đâu là nguyên nhân

chính xác gây ra mâu thuẫn, do vậy sự ra đời của một công cụ giúp chúng ta phát hiện chính xác nguyên nhân gây lỗi là rất cần thiết. Vì vậy trong nội dung chương 2, chúng em sẽ đề cập tới Ontology Debugging một khía cạnh rất được chú trọng khi số lượng phát biểu của ontology ngày càng tăng.

Chương 2

Giải pháp để sửa chữa inconsistent ontology

- Như đã được đề cập trong chương 1, trong các nguyên nhân dẫn đến tính thiếu nhất quán (*Inconsistency*) trong ontology thì **Unsatisfiable Class** (lớp không thỏa về tính logic) là nguyên nhân nếu có thể được phát hiện sớm để loại bỏ hoặc sửa lại các phát biểu gây mâu thuẫn thì giúp cho ontology tránh bị inconsistent.
- Đã có rất nhiều nghiên cứu thành công trong việc tìm và phát hiện lỗi (các phát biểu mâu thuẫn) trong ontology. Trong đó có một nghiên cứu nổi bật^[1], không chỉ có khả năng phát hiện gần như chính xác các nguyên nhân gây lỗi mà còn được đưa ra các giải pháp tối ưu* để sửa lỗi. Nghiên cứu này đã được ứng dụng để đưa ra các giải thích về các lớp không thỏa về nghĩa (unsatisfiable classes) trong bộ thực viện lập trình ontology thông dụng hiện nay là OWL-API^[2]. Sau đây chúng em xin được trình bày lại những điểm quan trọng trong nghiên cứu vừa được đề cập**

* Tối ưu có nghĩa là hạn chế tối đa các thay đổi về ý nghĩa mà việc xóa hoặc thay đổi phát biểu mâu thuẫn có thể gây ra cho các phát biểu khác (other axioms) trong ontology.

** Mọi quan điểm và ý tưởng trình bày ở phần sau của Chương 2 đều thuộc sở hữu của các tác giả bài báo^[1] và ^[9]. Chúng em chỉ trình bày lại sau khi đã đọc và nắm được ý tưởng chính yếu của bài báo.

2.1 Mục tiêu của việc debugging ontology

Mục tiêu chính của việc debugging ontology gồm hai phần quan trọng. Thứ nhất, với một ontology có số lượng lớn các lớp unsatisfiable, cần tìm và nhận dạng được nguyên nhân gây ra mâu thuẫn và các lớp bị ảnh hưởng bởi sự mâu thuẫn đó trong ontology. Thứ hai, cho biết trước một Unsatisfiable Class cụ thể, trích xuất và trình bày cho người sử dụng ontology (*modeler*) một tập hợp tối thiểu các phát biểu (*minimal set of axioms*) từ ontology hay nguyên nhân chính xác chịu trách nhiệm trong việc gây ra sự mâu thuẫn về logic.

2.2 Khái niệm và các kỹ thuật cần biết

Các hệ thống Description Logic thường cung cấp một tập hợp các tác vụ suy luận đã được chuẩn hóa như phân loại các khái niệm (*concept classification*), kiểm tra tính đáp ứng về logic (*concept satisfiability*) và kiểm tra tính nhất quán của knowledge base (KB). Hầu hết các reasoner thông dụng hiện nay đều buộc phải cung cấp đủ 3 tác vụ nêu trên, nhưng tất cả chúng đều không thân thiện với người dùng. Do tất cả những gì chúng ta biết được đều là kết quả (hay output) từ sự suy luận (reasoning) của reasoner.

Để giúp cho các tác vụ suy luận (reasoning) trở nên thân thiện với người dùng hơn, một hệ thống DL-based Knowledge Representation (KR) phải mở rộng thêm các lựa chọn về các tác vụ không nằm trong tiêu chuẩn của DL. Một ví dụ cụ thể là việc tạo ra các giải thích tại sao một lớp lại bị reasoner đánh giá là unsatisfiable. Thêm một tình huống mà người dùng cần được giải thích là tại sao reasoner đánh giá một lớp là lớp con của một lớp khác - đâu là lý do. Việc ra đời tác vụ giải thích nguyên nhân và kết quả là thật sự cần thiết trong bối cảnh sự phát triển nhanh của Semantic Web và cộng đồng người dùng/nhà phát triển Ontology ngày càng tăng nhanh.

Axiom Pinpointing service chính là dịch vụ có khả năng thực hiện tác vụ giải thích vừa được đề cập, với một KB và bất kì kết quả suy luận nào từ KB, dịch vụ này sẽ trả về tập các chứng minh/giải thích cho suy luận đó bằng những phát biểu đã được khai báo trong KB.

Có thể giải thích ngắn gọn như sau, cho một phát biểu kết quả họ SHOIN α được suy ra từ một knowledge base K , tập các giải thích/chứng minh cho α trong K là một phần tối thiểu $K' \subseteq K$ chịu trách nhiệm cho α xảy ra. Chứng minh K' là tối thiểu với điều kiện α là một kết quả logic được suy ra từ K' , hay nói cách khác K' tối thiểu khi và chỉ khi bất kì tập con nào của K' đều không suy ra được α . Nói chung có thể tồn tại

nhiều giải thích/chứng minh cho α trong K .

Sau đây là một ví dụ cho ý tưởng vừa nêu. Cho KB K với các phát biểu như sau:

1. $A \subseteq B \cap C$
2. $B \subseteq \neg E$
3. $A \subseteq D \cap \exists R.E$
4. $D \subseteq C \cap \forall R.B$

Trong KB trên, A, B, C, D, E là atomic concepts và R là atomic role. Chúng ta sẽ dùng số thứ tự của từng câu phát biểu trên thay vì lặp lại nguyên văn. Từ các phát biểu trên ta có $K \models (A \subseteq C)$. Tuy nhiên, điều kiện cần và đủ để suy ra được một kết quả tương tự từ 2 phần nhỏ hơn của KB K là $K_1 = 1$ và $K_2 = 3, 4$. Chúng ta nói K_1 và K_2 là các giải thích/chứng minh cho kết luận nói C là tập con của A - $A \subseteq C$.

2.3 Các bước sửa chữa các phát biểu bị lỗi

2.3.1 Minimal Unsatisfiability Preserving Sub-TBoxes (MUPS)

Khái niệm MUPS lần đầu được giới thiệu trong^[6]. Một MUPS của một lớp nguyên tử (*atomic class or concept*) A là một phần nhỏ nhất trong Knowledge Base mà trong đó A không đáp ứng (*unsatisfiable*). Mỗi lớp không đáp ứng sẽ có nhiều MUPS khác nhau trong một ontology. Tìm tất cả MUPS của một lớp không đáp ứng là một nhiệm vụ cực kì khó đứng ở khía cạnh debugging, từ đó để sửa lại một lớp không đáp ứng (*unsatisfiable class*) chúng ta cần loại bỏ tối thiểu ít nhất một phát biểu từ từng tập các phát biểu trong MUPS diện cho unsatisfiable class đó.

2.3.2 Tìm MUPS của một lớp không đáp ứng

Axiom Pinpointing service Một thành phần quan trọng trong giải pháp debugging dùng để tìm và trích xuất MUPS từ một lớp không đáp ứng. Họ cũng đã nâng cấp chức năng này nhằm xác định được chính xác những phần phát biểu gây lỗi, họ định nghĩa MUPS tìm được là *precise* MUPS.

Ý tưởng của kĩ thuật này có thể được mô tả khái quát như sau (chi tiết hơn về cách áp dụng xem^[8]). Mục tiêu là nhận diện những phần có liên quan của các phát biểu, nên họ đã định nghĩa một hàm nhằm chia nhỏ các phát biểu trong KB \mathbf{K} thành những phát biểu nhỏ hơn để thu được KB K_s có nghĩa tương đương với KB \mathbf{K} , với số lượng các phát

biểu được chia nhỏ nhiều nhất có thể.

Ví dụ: $A \subseteq C \cap D$ tương đương với $A \subseteq C$, $A \subseteq D$

Trong một vài trường hợp, kĩ thuật này bắt buộc phải giới thiệu ra một tên lớp mới, chỉ với mục đích là chia các phát biểu thành những mảnh nhỏ hơn.

Ví dụ: $A \subseteq \exists R.(C \cap D)$ không tương đương với $A \subseteq \exists R.C$, $A \subseteq \exists R.D$

Để chia nhỏ phát biểu trên chúng ta sẽ giới thiệu một tên lớp mới, gọi là E .

Như vậy ta có: $A \subseteq \exists R.(C \cap D)$ sẽ tương đương với $A \subseteq \exists R.E$, $E \subseteq C$, $E \subseteq D$, $C \cap D \subseteq E$. Cuối cùng nhiệm vụ tìm kiếm *precise* MUPS của lớp không đáp ứng trong KB **K** được đơn giản hóa thành vấn đề tìm MUPS trong những phiên bản đã được tách nhỏ trong KB K_s . Chi tiết hơn về kĩ thuật này xin đọc^[8].

2.3.3 Chiến thuật xếp hạng các phát biểu (*Axioms*)

Đây là một phần chính yếu trong tiến trình chỉnh sửa lại các phát biểu gây lỗi, quyết định xem nên loại bỏ phát biểu nào từ MUPS để sửa được lớp không đáp ứng.

Với mục tiêu này, một nhân tố đáng quan tâm là các phát biểu trong MUPS có thể được *xếp hạng* dựa theo mức độ quan trọng của chúng hay không. Việc sửa chữa các nguyên nhân gây lỗi được trở thành một vấn đề cần được tối ưu với mục tiêu quan trọng nhất là loại bỏ tất cả các lỗi thiếu nhất quán trong ontology, trong khi vẫn chắc chắn rằng những phát biểu có thứ hạng cao, nói cách khác là có giá trị quan trọng nghĩa sẽ được giữ lại và các phát biểu có thứ hạng thấp nhất sẽ được xóa đi.

Tiêu chí đơn giản nhất để xếp hạng các phát biểu là đếm số lần chúng xuất hiện trong MUPS từ những lớp không đáp ứng xuất hiện trong một ontology,

2.3.4 Đưa ra các nguyên nhân gây lỗi chính xác bằng giải thuật HST của Reiter

Qua các bước trên, họ đã đưa ra một quy trình để tìm được ...

Tìm MUPS bằng giải thuật HST của Reiter Một kỹ thuật tối ưu hơn kỹ thuật **Axiom Pinpointing** đã được đề cập ở mục trên nhằm tìm kiếm MUPS của một unsatisfiable class, họ đã sử dụng giải thuật Hitting Set của Reiter^[7], nhằm để xác định căn nguyên(*root cause*) của một vấn đề từ một bộ(*collection*) gồm nhiều tập hợp đựng độ chứa các nguyên nhân dẫn tới vấn đề, giải thuật này sẽ tạo ra những tập tối thiểu (*minimal hitting set*) chứa các nguyên nhân gây ra vấn đề. Một tập hợp đựng độ (*hitting set*) trong một bộ **C** các tập hợp là tập hợp giao (có chung phần tử) với từng tập hợp trong **C**. Một tập hợp đựng độ là tối thiểu nếu không có bất kì tập con nào của nó lại là một tập đựng độ cho **C**. Để áp dụng vào trường hợp này khi chúng ta đã có một bộ chứa

tập hợp chứa các nguyên nhân gây lỗi(*MUPS*), họ áp dụng giải thuật này để tạo ra các tập đùng độ từ MUPS - ý tưởng ở đây là loại bỏ tất cả các phát biểu trong tập đùng độ tối thiểu sẽ giúp loại bỏ từng phát biểu gây lỗi trên từng tập phát biểu của MUPS và cuối cùng giúp cho sửa chữa được unsatisfiable class.

Chương 3

Tài liệu tham khảo

1. Aditya Kalyanpur, Bijan Parsia, Evren Sirin , Bernardo Cuenca-Grau.Repairing Unsatisfiable Concepts in OWL Ontologies, 2007. Available online at <http://www.cs.ox.ac.uk/people/bernardo.cuencagrau/publications/repair.pdf>
2. The OWL API. The OWL API is a Java API for creating, parsing, manipulating and serialising OWL Ontologies. Available online at <https://github.com/owlcs/owlapi>
3. Pellet - OWL 2 Reasoner for Java.Pellet is an OWL 2 reasoner. Pellet provides standard and cutting-edge reasoning services for OWL ontologies. Available online at <http://clarkparsia.com/pellet/>
4. Robert Stevens. Ontogenesis, (I can't get no) satisfiability Available online at <http://ontogenesis.knowledgeblog.org/1329>
5. Samantha Bail.Ontogenesis, Common reasons for ontology inconsistency.Available online at <http://ontogenesis.knowledgeblog.org/1343>
6. S. Schlobach and R. Cornet. Non-standard reasoning services for the debugging of description logic terminologies. In Proceedings of IJCAI, 2003.
7. R. Reiter. A theory of diagnosis from first principles. 1987. Artificial Intelligence 32:57-95.
8. A. Kalyanpur, B. Parsia, B. Cuenca-Grau, and E. Sirin. Axiom pinpointing: Finding (precise) justifications for arbitrary entailments in SHOIN (owl-dl). Technical report, UMIACS, 2005-66, 2006. Technical Report