

**ĐẠI HỌC QUỐC GIA TP. HCM
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**



TRƯƠNG HẢI BẰNG

**TÍCH HỢP ONTOLOGY MỜ
TRÊN CƠ SỞ LÝ THUYẾT ĐỒNG THUẬN**

Chuyên ngành: **KHOA HỌC MÁY TÍNH**
Mã số: **62 48 01 01**

TÓM TẮT LUẬN ÁN TIẾN SĨ KHOA HỌC MÁY TÍNH

TP. HỒ CHÍ MINH NĂM 2016

Công trình được hoàn thành tại:

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN,
ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH**

Người hướng dẫn khoa học

1. GS. TSKH. Nguyễn Ngọc Thành

2. PGS. TS. Nguyễn Phi Khứ

Phản biện 1: PGS. TS. Lê Hoài Bắc

Phản biện 2: PGS. TS. Đỗ Văn Nhơn

Phản biện 3: PGS. TS. Hồ Bảo Quốc

Phản biện độc lập 1: PGS. TS. Trần Đình Khang

Phản biện độc lập 2: PGS. TS. Lê Anh Cường

Luận án được bảo vệ trước Hội đồng chấm luận án họp tại:
Trường Đại học Công nghệ thông tin, Đại học Quốc gia Tp.
Hồ Chí Minh

Vào lúc: 8 giờ 30 ngày 17 tháng 3 năm 2016

Có thể tìm hiểu luận án tại thư viện:

- Thư viện Quốc gia Việt Nam
- Thư viện Đại học Quốc gia Tp. Hồ Chí Minh
- Thư viện Trường Đại học công nghệ thông tin,
Đại học Quốc gia Tp. Hồ Chí Minh.

DANH MỤC CÁC CÔNG TRÌNH ĐÃ CÔNG BỐ

Các bài báo Tạp chí quốc tế (SCI-E):

-
- [1] Van du Nguyen, Ngoc Thanh Nguyen, Hai Bang Truong: A Preliminary Analysis of the Influence of the Inconsistency Degree on the Quality of Collective Knowledge. *Cybernetics and Systems*47(1-2): 69-87 (2016)

Impact Factor: 0.84

-
- [2] Hai Bang Truong, Trong Hai Duong, Ngoc Thanh Nguyen: “A Hybrid Method for Fuzzy Ontology Integration”. *Jour. Cybernetics and Systems*, 44(2-3), 133-154. (2013).

DOI: 10.1080/01969722.2013.762237

Impact Factor: 0.973

-
- [3] Trong Hai Duong, Ngoc Thanh Nguyen, Hai Bang Truong, Van Huan Nguyen: “A collaborative algorithm for semantic video annotation using a consensus-based social network analysis”. *Jour. Expert Systems with Applications*, 42(1), 246-258. (2015). DOI: 10.1016/j.eswa.2014.07.046

Impact Factor: 1.965

Các bài báo Hội nghị quốc tế:

-
- [1] Ngoc Thanh Nguyen, Hai Bang Truong: “A consensus-based method for fuzzy ontology integration”, in *Computational Collective Intelligence. Technologies and Applications* (pp. 480-489). Springer. DOI: 10.1007/978-3-642-16732-4_51. Series ISSN: 0302-9743. Print ISBN: 978-3-642-16731-7. Online ISBN: 978-3-642-16732-4. (2010).

-
- [2] Hai Bang Truong, Ngoc Thanh Nguyen: “A framework of an effective fuzzy ontology alignment technique”, in *Systems, Man, and Cybernetics (SMC), International Conference on* (pp. 931-935). IEEE.

DOI: 10.1109/ICSMC.2011.6083788. Series ISSN: 1062-922X. Print ISBN: 978-1-4577-0652-3. (2011).

-
- [3] Hai Bang Truong, Ngoc Thanh Nguyen, Phi Khu Nguyen: “Fuzzy ontology building and integration for fuzzy inference systems in
-

-
- weather forecast domain”, in *Intelligent Information and Database Systems* (pp. 517-527): Springer.
DOI: 10.1007/978-3-642-20039-7_52. Series ISSN: 0302-9743.
Print ISBN: 978-3-642-20038-0. Online ISBN: 978-3-642-20039-7.
(2011).
-
- [4] Trong Hai Duong, Hai Bang Truong, Ngoc Thanh Nguyen: “Local neighbor enrichment for ontology integration”. In *Intelligent Information and Database Systems* (pp. 156-166). Springer. DOI: 10.1007/978-3-642-28487-8_16. Series ISSN: 0302-9743. Print ISBN: 978-3-642-28486-1. Online ISBN: 978-3-642-28487-8. (2012).
-
- [5] Hai Bang Truong, Ngoc Thanh Nguyen: “A multi-attribute and multi-valued model for fuzzy ontology integration on instance level”, in *Intelligent Information and Database Systems* (pp. 187-197). Springer. DOI: 10.1007/978-3-642-28487-8_19. Series ISSN: 0302-9743. Print ISBN: 978-3-642-28486-1. Online ISBN: 978-3-642-28487-8. (2012).
-
- [6] Hai Bang Truong, Quoc Uy Nguyen, Ngoc Thanh Nguyen, Trong Hai Duong: “A new graph-based flooding matching method for ontology integration”, in *Cybernetics (CYBCONF), International Conference on* (pp. 86-91). IEEE.
DOI: 10.1109/CYBConf.2013.6617467
INSPEC Accession Number: 13826106. (2013).
-
- [7] Hai Bang Truong, Hung Quach: “An Overview of Fuzzy Ontology Integration Methods Based on Consensus Theory”, in *Advanced Computational Methods for Knowledge Engineering* (pp. 217-227). Springer. DOI: 10.1007/978-3-319-06569-4_16. Series ISSN: 2194-5357. Print ISBN: 978-3-319-06568-7. Online ISBN: 978-3-319-06569-4. (2014).
-
- Các bài báo hội nghị trong nước:*
-
- [1] Trương Hải Bằng, Nguyễn Phi Khứ: “Các Phương Pháp Lập Luận và Tích hợp Ontology Mờ”, Hội nghị khoa học và công nghệ quốc gia: Nghiên cứu cơ bản và ứng dụng Công nghệ thông tin (FAIR). Thừa Thiên- Huế. (Tr. 71-79). Print ISBN: 978-604-913-165-3. (2013).
-

CHƯƠNG 1: MỞ ĐẦU

1.1 Động cơ nghiên cứu

Ontology có vai trò quan trọng trong việc tổ chức và quản lý thông tin tri thức ở các lĩnh vực nghiên cứu và ứng dụng khác nhau: trong tích hợp cơ sở dữ liệu, thương mại điện tử, các dịch vụ web ngữ nghĩa, các mạng xã hội... Để phát triển các hệ thống ứng dụng này, bài toán tích hợp ontology đã được nhiều công trình tập trung nghiên cứu. Cùng với các kết quả nghiên cứu về lý thuyết các công cụ tích hợp ontology đã được xây dựng và phát triển: Anchor-PROMPT (Noy & Musen, 2001), RiMOM (J. Li, Tang, Li, & Luo, 2009), (Y. Jean-Mary & Kabuka, 2007), FCA-Merge (Stumme & Maedche, 2001), Chimaera (McGuinness, Fikes, Rice, & Wilder, 2000), (M. Seddiqui, Aono, M., 2008), Falcon-AO (Hu et al., 2007)...

Trong các hệ thống tri thức dựa trên ontology, các khái niệm, mối quan hệ và các thực thể của các đối tượng luôn luôn được thể hiện một cách chính xác. Điều này không hoàn toàn phù hợp trong thế giới thực thường chứa các thông tin mơ hồ, không chắc chắn và không đầy đủ. Mặt khác, do sự phân tán thông tin đến từ nhiều nguồn khác nhau và ngày càng gia tăng về số lượng các ontology dẫn đến sự không nhất quán thông tin và dữ liệu. Các phương pháp và công cụ tích hợp ontology rõ không còn phù hợp, dẫn đến sự ra đời của bài toán tích hợp ontology mờ.

Các công trình nghiên cứu về tích hợp ontology mờ hiện nay được thực hiện theo phương pháp so khớp/liên kết hoặc ánh xạ trên ontology mờ được định nghĩa theo logic mô tả mờ hoặc phi logic bằng cách mở rộng các thành phần của ontology truyền thống. Trong số không nhiều các nghiên cứu đề cập đến bài toán xử lý mâu thuẫn ontology mờ, Abulaish & Dey (2006) đã đề xuất độ đo khái niệm nhất quán giữa hai ontology để xác định một khái niệm nhất quán

giữa các ontology bị mâu thuẫn. Một tiếp cận khác, Ferrara et al., (2008) đề xuất phương pháp giải quyết mâu thuẫn ở mức quan hệ trong ánh xạ ontology mờ. Các mô hình ontology mờ được xây dựng chủ yếu phục vụ cho một ứng dụng cụ thể, chưa có các giải pháp xử lý mâu thuẫn trên cấu trúc phức tạp của ontology mờ trong bài toán tích hợp. Để giải quyết bài toán này, cần phải định nghĩa một mô hình ontology mờ và các thuật toán tích hợp trên mô hình tri thức đã xây dựng. Tuy nhiên tích hợp ontology mờ là bài toán phức tạp vì tính không chắc chắn và không đầy đủ của thông tin, sự mâu thuẫn tri thức và cấu trúc đa dạng của nó. Những khó khăn thách thức này thúc đẩy động cơ nghiên cứu của luận án. Các kết quả nghiên cứu và thử nghiệm được công bố: [1-10]. Các thử nghiệm được tiến hành theo tiêu chí đánh giá của *OAEI* và đánh giá chất lượng tích hợp dựa trên lý thuyết đồng thuận. Dữ liệu thử nghiệm là các ontology của *OAEI*¹ và ontology mờ thời tiết² được mờ hóa theo chuẩn *OWL2* (Bobillo & Straccia, 2011).

1.2 Thách thức của bài toán tích hợp ontology mờ

Các khó khăn thách thức cho bài toán tích hợp ontology mờ là: (1) Cần phải xác định mô hình tri thức ontology mờ cho bài toán tích hợp, (2) Tìm và xác định những điểm tương đồng và khác biệt giữa các ontology (so khớp ontology). (3) Xử lý mâu thuẫn giữa các ontology mờ trong quá trình tích hợp. Xử lý mâu thuẫn ontology mờ cần thỏa các điều kiện sau: Bảo toàn thông tin; bảo toàn cấu trúc; giải quyết được mâu thuẫn và chất lượng của các phương pháp xử lý mâu thuẫn.

¹ <http://oaei.ontologymatching.org/>

² <http://www.sensormeasurement.appspot.com/?p=ontologies#weather;http://paul.staroch.name/thesis/SmartHomeWeather.owl#>

1.3 Mục tiêu và phạm vi nghiên cứu

1.3.1 Mục tiêu nghiên cứu

Mục tiêu nghiên cứu của luận án là xây dựng mô hình ontology mờ cho bài toán tích hợp và các phương pháp tích hợp ontology mờ dựa trên cơ sở của lý thuyết đồng thuận.

1.3.2 Phạm vi nghiên cứu

- Xây dựng định nghĩa ontology mờ cho bài toán tích hợp.
- Xây dựng thuật toán so khớp khái niệm giữa các ontology mờ dựa trên phần chung tiềm năng.
- Xây dựng các thuật toán tích hợp ontology mờ dựa trên lý thuyết đồng thuận theo các mức khái niệm, quan hệ và thực thể.

1.4 Nội dung nghiên cứu

- Xây dựng định nghĩa ontology mờ (Mục 3.1, Chương 3).
- Bài toán PCP: Xây dựng các thuật toán so khớp ontology mờ dựa trên phần chung tiềm năng (Mục 3.2, Chương 3).
- Bài toán FOI-1: Định nghĩa bài toán mâu thuẫn mức khái niệm; Xây dựng thuật toán tích hợp ontology mờ dựa trên lý thuyết đồng thuận mức khái niệm (Mục 3.4, Chương 4).
- Bài toán FOI-2.1, FOI-2.2: Định nghĩa bài toán mâu thuẫn mức quan hệ; Xây dựng thuật toán tích hợp ontology mờ dựa trên lý thuyết đồng thuận mức quan hệ (Mục 3.5, Chương 3).
- Bài toán FOI-3: Định nghĩa bài toán mâu thuẫn mức thực thể; Xây dựng thuật toán tích hợp ontology mờ dựa trên lý thuyết đồng thuận mức thực thể (Mục 3.6, Chương 3).
- Thử nghiệm (Chương 4).

1.5 Các đóng góp chính của luận án

- Xây dựng mô hình tri thức ontology mờ.
- Xây dựng thuật toán so khớp ontology mờ.
- Xây dựng các thuật toán xử lý mâu thuẫn trong tích hợp ontology

mở ở ba mức khái niệm, quan hệ và thực thể dựa trên lý thuyết đồng thuận.

– Thực hiện cài đặt và đánh giá các thuật toán tích hợp ontology mở trên ontology mở thời tiết và dữ liệu ontology của *OAEI*.

1.6 Phương pháp nghiên cứu

Xây dựng mô hình tri thức ontology mở trên cơ sở lý thuyết mở của Zadeh (1965) và sử dụng lý thuyết đồng thuận (Nguyen, 2008a) trong các thuật toán tích hợp ontology mở.

1.7 Bố cục của luận án

Chương 1. Mở đầu; Chương 2. Tích hợp ontology mở và các nghiên cứu liên quan; Chương 3. Tích hợp ontology mở trên cơ sở lý thuyết đồng thuận; Chương 4. Thử nghiệm và đánh giá; Chương 5. Kết luận và hướng phát triển; Tài liệu tham khảo; Phụ lục A, B và C.

CHƯƠNG 2: TÍCH HỢP ONTOLOGY MỞ VÀ CÁC NGHIÊN CỨU LIÊN QUAN

2.1. Ontology và tích hợp ontology

2.1.1. Ontology

Định nghĩa 2.1 (Ontology):

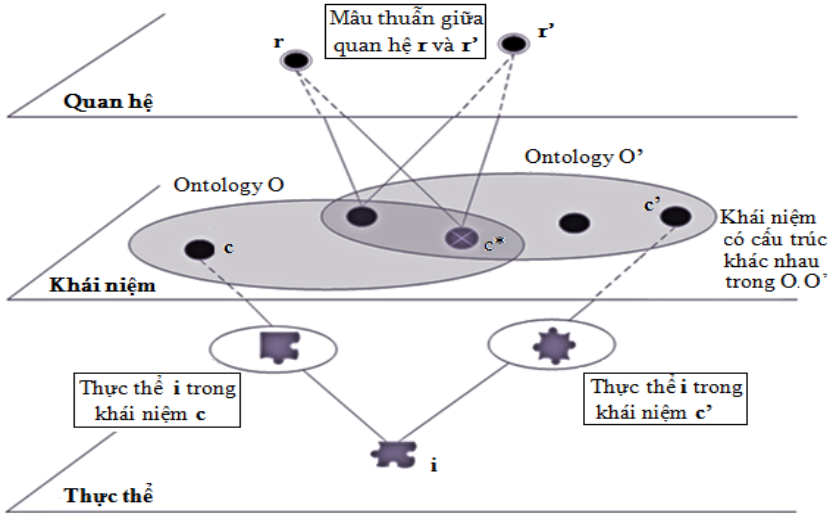
Theo (Gruber, 1993, 1995) (Nguyen, 2008), ontology là một bộ bốn được có các thành phần như sau: $O = (C, I, R, Z)$, trong đó:

- C là tập hợp các khái niệm;
- I là tập hợp các thực thể;
- R là tập hợp các quan hệ được định nghĩa trên C ;
- Z là tập hợp các tiên đề biểu diễn các ràng buộc toàn vẹn hoặc các mối quan hệ giữa các thực thể và các khái niệm.

2.1.2. Tích hợp ontology

Tích hợp tri thức là một quá trình mà trong đó sự không nhất quán của tri thức từ các nguồn khác nhau được hợp nhất nhằm mang lại một sự thống nhất của tri thức (Nguyen, 2008a).

Định nghĩa 2.2 (Nguyen, 2008a): Tích hợp ontology là quá trình xác định một ontology O^* tốt nhất đại diện cho các ontology $O_i, i=1, \dots, n$. Tích hợp ontology được thực hiện theo các mức khái niệm, quan hệ và thực thể (Hình 2.1).



Hình 2.1. Các mức mâu thuẫn ontology (Nguyen, 2008a)

Định nghĩa về tích hợp ontology và cấu trúc phân cấp về xử lý mâu thuẫn theo ba mức (Nguyen, 2008a) là cách tiếp cận của phương pháp tích hợp ontology mờ của luận án.

2.2. Ontology mờ

Hiện nay khái niệm ontology rõ không đủ khả năng để mô tả và biểu diễn các thông tin mơ hồ, không đầy đủ và không chắc chắn. Liên quan đến các nghiên cứu về vấn đề mờ hóa ontology có thể chia thành theo hai nhóm. *Nhóm thứ nhất*: các tác giả đề xuất các hướng tiếp cận dựa trên logic mô tả mờ. *Nhóm thứ hai*: các tác giả đề xuất một kiến trúc chung cho ontology mờ trong đó các khái niệm và các mối quan hệ được làm mờ bằng cách sử dụng các hàm thành viên mờ Zadeh (1965). Khảo sát các công trình liên quan đến mô hình

ontology mờ cho bài toán tích hợp, có thể thấy rằng không có định nghĩa ontology mờ duy nhất. Trong trường hợp đơn giản, theo Dey & Abulaish (2008) một ontology mờ là một cặp (C, R) , trong đó C là tập khái niệm mờ và R là một tập quan hệ mờ. Hầu hết các định nghĩa ontology mờ phụ thuộc vào ứng dụng cụ thể. Để xây dựng mô hình ontology mờ cho bài toán tích hợp, cần phải xem xét các vấn đề sau:

- Một số định nghĩa không mờ hóa các khái niệm, trong khi có định nghĩa khác không mờ hóa các quan hệ. Hầu hết các định nghĩa không trình bày ý nghĩa của tập Z (*thành phần mô tả các quy tắc, ràng buộc của ontology*).
- Các định nghĩa ontology mờ được xây dựng chủ yếu phục vụ cho bài toán tích hợp ở các mức so khớp, liên kết hoặc ánh xạ. Mô hình ontology mờ cho bài toán xử lý mâu thuẫn chưa có nhiều nghiên cứu quan tâm.
- Tùy theo ứng dụng và mục tiêu nghiên cứu bộ (C, R) có thể được mở rộng theo nhiều cách khác nhau, *cần phải có một mô hình ontology mờ chung để có sự chia sẻ, tương tác và tái sử dụng giữa các hệ thống tri thức ontology mờ*.
- Trong tương lai các ngôn ngữ ontology mờ sẽ cung cấp khả năng mở rộng mà các định nghĩa này không cho phép mờ hóa ontology theo yêu cầu. Điều này dẫn đến hạn chế khả năng mở rộng và tái sử dụng của các định nghĩa ontology mờ.

2.3. Tích hợp ontology mờ

Các công trình liên quan đến tích hợp ontology mờ hiện nay chủ yếu thực hiện ở mức so khớp, liên kết, ánh xạ ontology mờ, có rất ít công trình đề cập đến bài toán xử lý mâu thuẫn. Các công trình này chỉ đề cập đến xử lý mâu thuẫn mức khái niệm hoặc quan hệ, không đề cập

đến các tiêu chí tích hợp. Các bài toán trong tích hợp ontology mờ cần phải giải quyết là:

1. *So khớp/liên kết hoặc ánh xạ ontology mờ;*
2. *Xử lý mâu thuẫn ontology mờ trong bài toán tích hợp.*

Bài toán thứ nhất: So khớp/liên kết hoặc ánh xạ ontology mờ: (Patrice Buche, 2008); (Xu et al., 2005); (Bahri, Bouaziz, & Gargouri, 2007); (Ferrara et al., 2008); (Todorov, Hudelot, Popescu, & Geibel, 2014); (Abulaish & Dey, 2006). Đặc điểm nổi bật của các công trình này là:

- Xây dựng độ tương đồng giữa các khái niệm mờ của ontology trên mô hình logic mô tả mờ.
- Các thuật toán ánh xạ ontology mờ được thực hiện theo phương pháp vét cạn.

Bài toán thứ hai: Xử lý mâu thuẫn ontology mờ:

- Phương pháp xử lý mâu thuẫn của Abulaish & Dey (2006): Xây dựng *độ đo khái niệm nhất quán* giữa hai ontology để xác định một *khái niệm nhất quán* của các ontology có mâu thuẫn, không đề cập đến mâu thuẫn mức quan hệ và thực thể.
- Phương pháp ánh xạ ontology mờ theo các quan hệ tương đương và bao hàm (Ferrara et al., 2008) không đề cập xử lý mâu thuẫn mức khái niệm và thực thể và các tiêu chí tích hợp.

Khó khăn và thách thức: Theo Klein (2001), Mitra và Wiederhold (2002) một thuật toán tích hợp được đánh giá dựa trên hai tiêu chí chính: chất lượng của phương pháp tích hợp và độ phức tạp của chương trình, trong đó chất lượng của phương pháp tích hợp được đánh giá dựa trên mức độ giải quyết mâu thuẫn ontology.

2.4. Lý thuyết đồng thuận

Phần này trình bày tổng quan về lý thuyết đồng thuận trong xử lý mâu thuẫn tri thức (Nguyen, 2008a).

Định nghĩa 2.4.1 (Hàm khoảng cách): Gọi U là tập biểu diễn các đối tượng mâu thuẫn, hàm khoảng cách d được định nghĩa như sau:

$d: U \times U \rightarrow [0, 1]$, thỏa các điều kiện:

- Không âm: $\forall x, y \in U: d(x, y) \geq 0$
- Phản xạ: $\forall x, y \in U: d(x, y) = 0$ nếu $x = y$
- Đối xứng: $\forall x, y \in U: d(x, y) = d(y, x)$.

Định nghĩa 2.4.2 (Lược tả): Một lược tả X được định nghĩa như sau: $X = \{ r_i \in TUPLE(T_i) : T_i \subseteq A, i = 1, \dots, n \}$, với A là tập các thuộc tính, $TUPLE(T_i)$ là tập hợp các bộ kiểu T_i .

Định nghĩa 2.4.3 (Lược tả mâu thuẫn): Cho U là tập hữu hạn biểu diễn các đối tượng cho sự mâu thuẫn. Ký hiệu $\Pi_k(U)$ là tất cả các tập con k -phần tử của U với $k \in N$, N là tập số tự nhiên. Ký hiệu: $\prod(U) = \cup_{k \in N} \Pi_k(U)$ là tập tất cả các tập con khác rỗng của U , khi đó mỗi phần tử của $\prod(U)$ được gọi là một lược tả mâu thuẫn.

2.4.1. Định nghĩa sự đồng thuận

Định nghĩa 2.4.4 (Sự đồng thuận): Cho vũ trụ U và d đã được định nghĩa (Định nghĩa 2.4.1) sự đồng thuận được định nghĩa như sau: Cho ánh xạ $C: \prod(U) \rightarrow 2^U$, với mỗi lược tả mâu thuẫn $X \in \prod(U)$, một phần tử của $C(X)$ được gọi là một đồng thuận của lược tả X .

2.4.2. Các tiêu chí đồng thuận

Ký hiệu $Con(U)$ là tập hợp tất cả các hàm lựa chọn sự đồng thuận trong (U, d) . Với mọi $X, Y \in \prod(U)$ và $x \in U$, một số công thức và khái niệm được định nghĩa như sau:

- $d(x, X) = \sum_{y \in X} d(x, y)$
- $d^n(x, y) = \sum_{y \in X} (d(x, y))^n, n > 1, n \in N$
- X được gọi là đồng đều nếu tất cả các phần tử của X là đồng nhất, nghĩa là $X = \{n * x\}, x \in U, n \in N$.

Các tiêu chí của lý thuyết đồng thuận:

Với $C \in Con(U)$

1. Tiêu chí (Re): $C(X) \neq \emptyset$

2. Tiêu chí (Un): $C(\{n*x\}) = \{x\} \forall n \in N, x \in U$

3. Tiêu chí (Si):

(Lược tả X là bội của Lược tả Y) $\Rightarrow C(X) = C(Y)$

4. Tiêu chí (Qu):

$(x \notin C(X)) \Rightarrow (\exists n \in \mathcal{N}: x \in C(X \cup \{n*x\}), \forall x \in U$

5. Tiêu chí (Co):

$(x \in C(X)) \Rightarrow (x \in C(X \cup \{x\})) \forall x \in U$

6. Tiêu chí (Cc):

$(C(X_1) \cap C(X_2) \neq \emptyset) \Rightarrow (C(X_1 \cup X_2) = C(X_1) \cap C(X_2)) \forall X_1, X_2 \in \prod(U)$

7. Tiêu chí (Gc):

$C(X_1) \cap C(X_2) \subseteq C(X_1 \cup X_2) \subseteq C(X_1) \cup C(X_2), \forall X_1, X_2 \in \prod(U)$

8. Tiêu chí (Pr):

$(X_1 \subseteq X_2 \wedge x \in C(X_1) \wedge y \in C(X_2)) \Rightarrow (d(x, X_1) \leq d(y, X_2))$

9. Tiêu chí C_1 :

$(x \in C(X)) \Rightarrow (d(x, X) = \min_{y \in U} d(y, X)), \forall X \in \prod(U)$

10. Tiêu chí C_2 :

$(x \in C(X)) \Rightarrow (d^2(x, X) = \min_{y \in U} d^2(y, X)), \forall X \in \prod(U)$

2.4.3. Lý thuyết đồng thuận và mâu thuẫn lược tả mờ

2.4.3.1. Các khái niệm cơ bản

Định nghĩa 2.4.5 (Mâu thuẫn lược tả mờ):

Cho U là tập vũ trụ hữu hạn, một tập vũ trụ mờ ký hiệu U_F được định nghĩa như sau: $U_F = U \times [0, 1]$. Mỗi phần tử của $\prod(U_F)$ được gọi là một mâu thuẫn lược tả mờ.

Định nghĩa 2.4.6 (Hàm khoảng cách):

Khoảng cách giữa các phần tử mờ của U_F được định nghĩa như sau:

- $d_F(x, y) \geq 0$, với mọi $x, y \in U_F$ (tính không âm)
- $d_F(x, y) = 0$ khi và chỉ khi $x = y$ (tính phản xạ)

- $d_F(x, y) = d_F(y, x)$, với mọi $x, y \in U_F$ (tính đối xứng)
- $d_F((x, 1), (y, 1)) = d(x, y)$
- $d_F((x, v_1), (x, v_2)) = |v_1 - v_2|$, $\forall x, y \in U, v_1, v_2 \in [0, 1]$

2.4.3.2. Các tiêu chí đồng thuận mờ

Định nghĩa 2.4.7.

Cho ánh xạ $C: \prod U_F \rightarrow 2^{U_F}$, với mỗi lược tả mờ $X \in \prod U_F$

Định nghĩa các hàm $d_F(z, X)$ và $d_F^2(x, X)$:

- $d_F(z, X) = \sum_{y \in X} d_F(z, y)$, $z \in U_F$
- $d_F^2(x, X) = \sum_{y \in X} (d_F(z, y))^2$, $z \in U_F$

Định nghĩa các tập $S_X, X(x)$ và \tilde{X} :

- $S_X = \{x \in U: x \text{ có trong } X\}$
- $X(x) = \{(x, v): (x, v) \in X \text{ và } x \in S_X\}$.
- $\tilde{X} = \{x: (x, v) \in X \text{ và } v \in [0, 1]\}$, \tilde{X} là lược tả rõ là một trường hợp của lược tả mờ $X \in \prod(U_F)$, trong đó: X là lược tả đồng đều nếu \tilde{X} là đồng đều.

Các tiêu chí đồng thuận cho các mâu thuẫn lược tả mờ:

1. Tiêu chí (Pr):

$$(X_1 \subseteq X_2 \wedge x \in C(X_1) \wedge y \in C(X_2)) \Rightarrow (d_F(x, X_1) \leq d_F(y, X_2))$$

2. Tiêu chí C_1 :

$$(x \in C(X)) \Rightarrow (d_F(x, X) = \min_{y \in U_F} d_F(y, X)), \forall X \in \prod(U_F)$$

3. Tiêu chí C_2 :

$$(x \in C(X)) \Rightarrow (d_F^2(x, X) = \min_{y \in U_F} d_F^2(y, X)), \forall X \in \prod(U_F)$$

4. Tiêu chí (Cl): $S_{C(X)} \subseteq S_X$

5. Tiêu chí (TLC):

$$X \text{ là lược tả mờ chính quy} \Rightarrow C(X) = C(\bigcup_{x \in S_X} C(X(x)))$$

6. Tiêu chí (Fa): X là lược tả mờ đồng đều \Rightarrow

$$C(X) = \{(x, v^*): v^* = \frac{\sum_{(x,v) \in X} v}{\text{card}(X)}\}$$

7. Tiêu chí (Fs) : $(\forall (x, v), (x', v') \in X: v = v' = v^*)$

$$\Rightarrow C(X) = \{(x, v^*): x \in C'(\tilde{X})\} \text{ với } C' \in \text{Con}(U).$$

2.4.4. Nhận xét

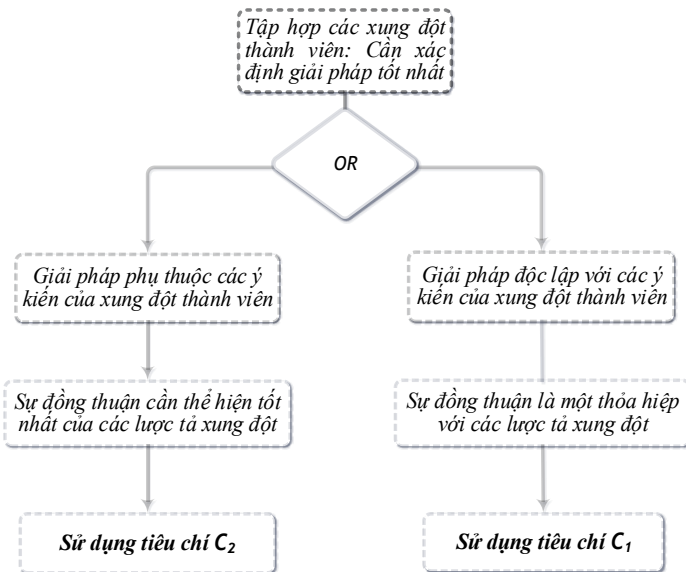
– Các tiêu chí (Re , Co , Qu , Cl , TLC) độc lập với cấu trúc của U và U_F , các tiêu chí Pr , C_1 , C_2 phụ thuộc các hàm khoảng cách d và d_F .

– Theo định nghĩa, các tiêu chí C_1 , C_2 trong không gian (U, d) và (U_F, d_F) có sự tương đương như sau:

- Tiêu chí C_1 : tổng các khoảng cách từ một đồng thuận đến các phần tử của lược tả $\prod(U)$ là nhỏ nhất.
- Tiêu chí C_2 : tổng bình phương các khoảng cách từ một đồng thuận đến các phần tử của lược tả $\prod(U)$ là nhỏ nhất.

2.4.5. Lược đồ lựa chọn tiêu chí đồng thuận

Lược đồ lựa chọn tiêu chí đồng thuận C_1 và C_2 như sau:



Hình 2.2. Lược đồ lựa chọn tiêu chí đồng thuận (Nguyen, 2008a)

2.5. Tổng kết chương

Chương 2 trình bày các nghiên cứu liên quan đến nội dung và cơ sở lý thuyết cho nghiên cứu của luận án, bao gồm các khái niệm về

ontology và tích hợp ontology; tổng quan nghiên cứu về ontology mờ và tích hợp ontology mờ và cơ sở lý thuyết đồng thuận cho bài toán tích hợp ontology mờ.

CHƯƠNG 3: TÍCH HỢP ONTOLOGY MỜ TRÊN CƠ SỞ LÝ THUYẾT ĐỒNG THUẬN

Trong quá trình thực hiện tích hợp ontology mờ cần phải giải quyết hai bài toán quan trọng: *Bài toán thứ nhất* là tìm và xác định những điểm tương đồng và khác biệt giữa các ontology (so khớp/ liên kết ontology). *Bài toán thứ hai* là phương pháp xử lý mâu thuẫn giữa các ontology mờ trong quá trình tích hợp. Chương 3 trình bày các nội dung đóng góp mới của luận án: (1) xây dựng mô hình tri thức ontology mờ cho bài toán tích hợp, (2) phương pháp so khớp ontology mờ dựa trên phân chung tiềm năng và (3) các thuật toán tích hợp ontology mờ dựa trên các tiêu chí của lý thuyết đồng thuận. Các kết quả nghiên cứu và thực nghiệm được công bố tại [1, 3, 4, 5, 9, 10].

3.1. Mô hình ontology mờ cho bài toán tích hợp

Định nghĩa 3.1.1 (Ontology mờ):

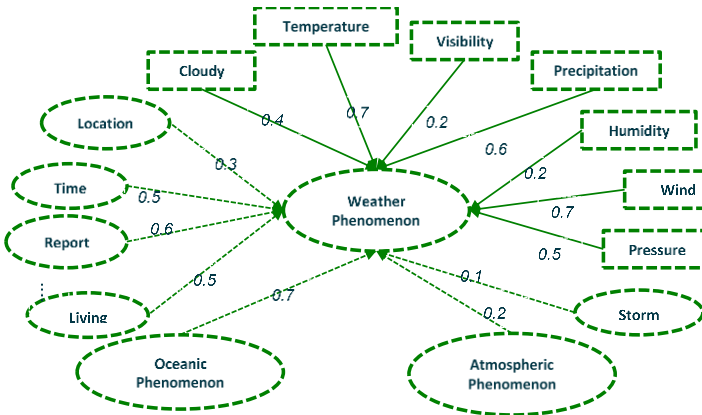
Cho (A, V) là một thể giới thực, trong đó A là tập hữu hạn các thuộc tính, V miền giá trị của A . Ontology mờ được định nghĩa là bộ bốn (C, R, I, Z) , trong đó:

- C là tập hữu hạn các khái niệm. Một khái niệm c của ontology mờ được định nghĩa là một bộ tứ: (c, A^c, V^c, f^c) , với c là tên duy nhất của khái niệm, $A^c \subseteq A$ là tập các thuộc tính mô tả khái niệm, $V^c \subseteq V$ là miền giá trị của thuộc tính: $V^c = \bigcup_{a \in A} V_a$ (V_a là miền giá trị của thuộc tính a) và f^c là hàm thành viên mờ: $f^c: A^c \rightarrow [0, 1]$ biểu diễn mức độ của thuộc tính được mô tả trong khái niệm c . Bộ (A^c, V^c, f^c) được gọi là cấu trúc mờ của c . Minh họa (Hình 3.1).

- R là tập các quan hệ mờ giữa các khái niệm, $R = \{R_1, R_2, \dots, R_m\}$,

$R_i \subseteq C \times C \times [0,1], i = 1, \dots, m$. Một quan hệ là một tập bao gồm một cặp khái niệm và giá trị mờ biểu diễn mức độ quan hệ giữa chúng. Mỗi quan hệ R_i giữa hai khái niệm trong ontology c được biểu diễn bằng một giá trị mờ duy nhất, nghĩa là nếu $(c, c', v) \in R_i$ và $(c, c', v') \in R_i$ thì $v = v'$. Ví dụ (Hình 3.2).

- I là tập các thực thể mờ của khái niệm c được mô tả bởi các thuộc tính của tập A^c . là một cặp (i, v) , với i là định danh của thực thể, v là tập các giá trị mờ của thực thể. (Bảng 3.1).
- Z là tập hợp các quy tắc, các ràng buộc toàn vẹn để xác định các khái niệm C các quan hệ R .



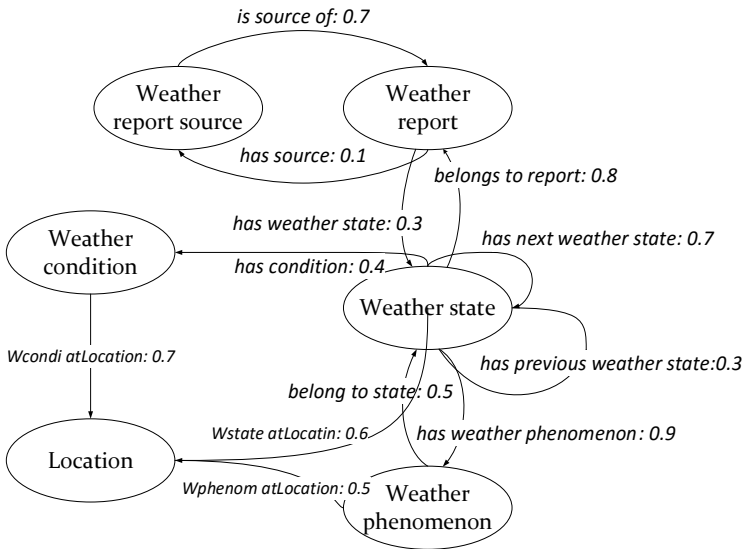
Hình 3.1. Ví dụ về khái niệm mờ thời tiết

Ví dụ 3.1.1. (Khái niệm mờ): Theo định nghĩa một khái niệm (c, A^c, V^c, f^c) của ontology thời tiết được mô tả như sau (Hình 3.2):

Khái niệm c mô tả hiện tượng thời tiết *WeatherPhenomenon* với các thuộc tính: *Nhiệt độ, áp suất khí quyển, lượng mưa, pressure, độ ẩm, gió, tầm nhìn, tuyết*, được ký hiệu như sau: *Weather Phenomenon*

(*temperature, pressure, precipitation, humidity, wind, visibility, snow*), trong đó:

- Tập $A^c = \{temperature, precipitation, pressure, humidity, wind, visibility, snow\}$.
- V^c là tập các miền giá trị của các thuộc tính tương ứng với các thuộc tính thuộc tập $A^c = \{temperature, precipitation, pressure, humidity, wind, visibility, snow\}$.
- Các giá trị mờ của thuộc tính tương ứng của tập A^c là $f^c: A^c \rightarrow [0, 1]$ biểu diễn mức độ của các thuộc tính được mô tả trong khái niệm *Weather Phenomenon*.



Hình 3.2. Ví dụ về quan hệ mờ thời tiết

Ví dụ 3.1.2. (Quan hệ mờ). Quan hệ mờ giữa khái niệm *Weather Phenomenon* và các khái niệm $C = \{Report, Location, Time, WeatherState, WeatherReport, \dots\}$ (Hình 3.2). Các quan hệ mờ $R =$

$\{R_1, R_2, \dots, R_{12}\}, R_i \subseteq C \times C \times [0, 1]: R_1 = \text{is source of}: 0.7; R_2 = \text{has weather state}: 0.3; \dots; R_{12} = \text{Wphenom atLocation}: 0.5$

Ví dụ 3.1.3. (Thực thể mờ). Khái niệm *Location*(*temperature, precipitation, humidity, wind*) có các thực thể mờ: *Hanoi, QuangBinh, HCMC, Paris* như sau (Bảng 3.1):

Bảng 3.1. Ví dụ về thực thể mờ thời tiết

Thực thể → Thuộc tính ↓	Hanoi	QuangBinh	HCMC	Paris
<i>temperature</i>	0.4	0.3	0.3	0.3
<i>precipitation</i>	0.4	0.6	0.7	0.7
<i>pressure</i>	0.2	0.5	0.2	0.4
<i>humidity</i>	0.7	0.5	0.5	0.4
<i>Wind</i>	0.3	0.8	0.3	0.2
<i>visibility</i>	0.6	0.4	0.3	0.5
<i>snow</i>	1.0	1.0	1.0	0.6

3.2. So khớp ontology mờ dựa trên phần chung tiềm năng

Thuật toán so khớp ontology mờ dựa trên phần chung tiềm năng được phát triển từ *độ quan trọng khái niệm* trên ontology rõ (Duong et al., 2012) và khái niệm *phần chung tiềm năng* (Xu et al., 2005). Các kết quả nghiên cứu và thử nghiệm được công bố tại [1, 4, 9, 10].

3.2.1. Khái niệm phần chung tiềm năng:

Khái niệm phần chung tiềm năng *PCP* (Potentially Common Parts) được đề xuất bởi Xu, Kang, Lu, Li, & Jiang (2005) là phần chung của hai ontology cùng lĩnh vực với đa số các khái niệm có khả năng tương đồng với nhau.

Giả sử $P_1 = \{c_1, c_2, \dots, c_m\}$ và $P_2 = \{c'_1, c'_2, \dots, c'_n\}$ là tập các khái niệm của các *PCP* thuộc hai ontology tương ứng O_1 và O_2 . Khoảng cách ban đầu giữa các *PCP* được định nghĩa như sau:

$$D_0(P_1, P_2) = \sum_{c_i \in P_1} \sum_{c'_j \in K} |CI(c_i) - CI(c'_j)| \quad (3.1)$$

Nếu $D_0(P_1, P_2)$ là nhỏ nhất, suy ra cặp khái niệm (c_i, c_j) có khả năng tương đương.

- Với $K \subset P_2$ là tập hợp các phần tử được xác định bởi khoảng cách giữa các khái niệm trên P_1 và P_2 phụ thuộc vào khoảng cách $D_k(P_1, P_2)$: $K = \{c_j \in P_2 \mid d(c_i, c_j) \leq \eta, \forall c_i \in P_1, \eta$ là một ngưỡng cho trước $\}$.
- $CI(c)$ là độ đo tầm quan trọng của một khái niệm được thể hiện sự đóng góp của các thuộc tính của khái niệm c so với các khái niệm khác trong ontology mờ được định nghĩa như sau:

$$CI(c) = AI(c) + RI(c) \quad (3.2)$$

$$AI(c) = \sum_{a \in A^c} f^c(a) \quad (3.3)$$

$$RI(c) = \sum_{cRc'} f^c(R) * CI(c') \quad (3.4)$$

Trong đó $f^c(a)$ là giá trị mờ của các thuộc tính, $f^c(R)$ là giá trị mờ của quan hệ giữa các khái niệm c và c' .

3.2.2. Thuật toán so khớp ontology mờ PCP

3.2.2.1 Phát biểu bài toán (So khớp hai ontology mờ):

Cho cặp khái niệm $(c_i, c_j) \in (O_1, O_2)$ và độ sâu *DepthLevel*, cần xác định tập các khái niệm tương đồng giữa các ontology O_1 và O_2 .

Ý tưởng chính của thuật toán là bắt đầu từ việc so khớp một cặp khái niệm ban đầu (cặp *medoids*) để xác định phần chung tiềm năng ban đầu giữa các ontology mờ. Quá trình so khớp được thực hiện từ các PCP và kết quả thu được là một tập hợp các cặp khái niệm tương đồng. Để tìm ra các cặp khái niệm có khả năng tương đồng mới, sử dụng công thức đệ quy sau:

$$D_{k+1}(P_1, P_2) = D_k(P_1, P_2) + E(P_1, P_2), k = 0, 1, 2.. \quad (3.5)$$

trong đó: $E(P_1, P_2)$ là giá trị sai khác bởi sự sắp xếp lại P_1 hoặc P_2 .

3.2.2.2 Thuật toán so khớp ontology mờ dựa trên phần chung tiềm năng:

Các biến sử dụng trong thuật toán:

- η^* là ngưỡng thấp nhất thỏa sự tương đồng giữa hai khái niệm;
- η là ngưỡng sao cho $D_{k+1} - D_0 > \eta$;
- RealMatch là tập hợp các cặp khái niệm có độ tương đồng:

Similarity (c_i, c_j) $\geq \eta^*$

INPUT: Cặp khái niệm Medoids (c_1, c_2) và độ sâu DepthLevel

OUTPUT: Tập các khái niệm tương đồng RealMatch

BEGIN

1. Từ cặp khái niệm (c_1, c_2) xây dựng các PCP P_1, P_2 thuộc các ontology O_1, O_2 ;
2. Khởi tạo bảng MinTable từ các PCP P_1, P_2
3. $k = 0$;
4. $D_{k+1} = D_0$;
5. while $D_{k+1} - D_0 \geq \eta$ & MinTable! = Null do
6. if tồn tại cycle-cross cell (c_i, c_j) trong MinTable
7. if *Similarity* (c_i, c_j) $\geq \eta^*$
8. Inserts (c_i, c_j) to RealMatch;
9. Loại bỏ dòng và cột tương ứng (c_i, c_j);
10. Tính lại D_0 ;
11. $k = 0$;
12. $D_{k+1} = D_0$;
13. else
14. Marks (c_i, c_j);
15. Di chuyển dòng và cột đến giá trị nhỏ nhất trong MinTable;
16. Tính lại D_{k+1} ;
17. else
18. Chọn cell (c_i, c_j) có giá trị nhỏ nhất
19. if *Similarity* (c_i, c_j) $\geq \eta^*$
20. Inserts (c_i, c_j) to RealMatch;
21. Loại bỏ dòng và cột chứa (c_i, c_j)
22. Updates MinTable;
23. Tính lại D_0 ;
24. $k = 0$;
25. $D_{k+1} = D_0$;

```

26.   else
27.       Marks ( $c_i, c_j$ );
28.       Di chuyển dòng và cột đến giá trị nhỏ
        nhất trong MinTable;
29.       Tính lại  $D_{k+1}$ ;
30. Return (RealMatch);
END.

```

Độ phức tạp của thuật toán là $O(n \cdot \log n)$. Bảng 3.1. so sánh độ phức tạp của thuật toán so khớp PCP với các phương pháp khác.

Bảng 3.1. Độ phức tạp của phương pháp PCP

Phương pháp	Độ phức tạp
PROMPT	$O(n^2)$
Anchor-PROMPT	$O(n^2 \cdot \log^2 n)$
Glue	$O(n^2)$
PCP	$O(n \cdot \log n)$

3.3. Chất lượng của tri thức đồng thuận trong tích hợp ontology mờ

Chất lượng của một thuật toán tích hợp tri thức được hiểu là sự khác biệt giữa *tri thức nhóm* (*Collective Knowledge*) và *tri thức đúng* (*Real State of Knowledge*). Cho các mâu thuẫn lược tả X_1, \dots, X_n , ký hiệu $d(X^*, r)$ là khoảng cách từ r đến X^* , trong đó r là *tri thức đúng* và X^* là *tri thức nhóm*.

Định nghĩa 3.3.1. (Chất lượng của sự đồng thuận)

Với các ký hiệu $Con(U)$, $\Pi(U)$ và $C(X)$ được trình bày tại mục 2.4. Chất lượng của sự đồng thuận x trong lược tả X được định nghĩa như sau (Nguyen, 2008a):

$$d^*(x, X) = 1 - \frac{d(x, X)}{|X|} \quad (3.6)$$

Trong đó: $X \in \Pi(U)$, $C \in Con(U)$ và $x \in C(X)$,

Các thuật toán tích hợp tri thức theo các tiêu chí đồng thuận C_1 và C_2 được xây dựng dựa trên các *Định lý* (Nguyen, 2008a) sau đây:

Định lý 3.1 (Sử dụng tiêu chí C_2):

Cho lược tả $X = \{x^{(i)} = x_1^{(i)}, x_2^{(i)}, \dots, x_m^{(i)}\}: i=1, 2, \dots, n\}$. Vector $x = (x_1, x_2, \dots, x_m)$ thỏa tiêu chí đồng thuận C_2 khi và chỉ khi: $x_j = \frac{1}{n} \sum_{i=1}^n x_j^{(i)}, j = 1, 2, \dots, m$.

Định lý 3.2 (Sử dụng tiêu chí C_1):

Cho lược tả $X = \{x^{(i)} = x_1^{(i)}, x_2^{(i)}, \dots, x_m^{(i)}\}: i=1, 2, \dots, n\}$. Vector $x = (x_1, x_2, \dots, x_m)$ thỏa tiêu chí đồng thuận C_1 nếu không tồn tại bất kỳ vector y nào sao cho: $d(y, x) > d(y, x^{(i)}), \forall i=1, 2, \dots, n$.

Nhận xét

Để xác định tiêu chí đồng thuận cho các thuật toán tích hợp ontology mờ, cần phải thực hiện các bước như sau:

- Xem xét cấu trúc của các mâu thuẫn lược tả để lựa chọn các tiêu chí đồng thuận C_1 và C_2 theo lược đồ ở mục 2.4.5.
- Sử dụng các tiêu chí C_1 và C_2 :
 - Đối với tiêu chí C_1 : Sử dụng định lý 3.2
 - Đối với tiêu chí C_2 : Sử dụng định lý 3.1.

3.4. Tích hợp ontology mờ ở mức khái niệm

3.4.1. Mâu thuẫn ontology mờ mức khái niệm

Định nghĩa 3.4.1 (Mâu thuẫn ontology mờ mức khái niệm):

Cho hai ontology mờ O_1 và O_2 , khái niệm $(c_1, A^{c_1}, V^{c_1}, f^1)$ thuộc O_1 và khái niệm $(c_2, A^{c_2}, V^{c_2}, f^2)$ thuộc O_2 . Mâu thuẫn khái niệm xảy ra nếu $A^{c_1} \neq A^{c_2}$ hoặc $V^{c_1} \neq V^{c_2}$ hoặc $f^1 \neq f^2$.

3.4.2. Thuật toán tích hợp ontology mờ mức khái niệm

Phát biểu bài toán FOI-1:

Cho tập $X = \{(A^i, V^i, f^i) | (A^i, V^i, f^i)\}$ là cấu trúc mờ của khái niệm c trong ontology $O_i, i=1, \dots, n$, cần xác định bộ ba: $c^* = (A^*, V^*, f^*)$ tốt nhất thỏa tiêu chí của lý thuyết đồng thuận.

Thuật toán FOI-1:

INPUT: $X = \{(A^i, V^i, f^i) | (A^i, V^i, f^i)\}$ là tập các cấu trúc mờ của khái niệm c trong ontology $O_i, i=1, \dots, n$

OUTPUT: $c^* = (A^*, V^*, f^*)$ biểu diễn tốt nhất từ X thỏa các tiêu chí đồng thuận.

BEGIN

```

1.    $A^* = \bigcup_{i=1}^n A_i$ ;
2.    $V^* = \bigcup_{i=1}^n V_i$ ;
3.   foreach pair  $a_1, a_2 \in A^*$  do
4.     if  $R(a_1, \Leftrightarrow, a_2)$  then  $A^* \setminus \{a_2\}$  and  $X_a = X_a \cup \{f^i(a_2)\}$ ; /* eg., job  $\Leftrightarrow$  occupation */
5.     if  $R(a_1, \pm, a_2)$  then  $A^* \setminus \{a_1\}$  and  $X_a = X_a \cup \{f^i(a_1)\}$ ; /* eg., age  $\pm$  birthday */
6.     if  $R(a_1, \pm, a_2)$  then  $A^* \setminus \{a_1\}$  and  $X_a = X_a \cup \{f^i(a_1)\}$ ; /* eg., sex  $\pm$  female */
7.     if  $R(a_1, \perp, a_2)$  then  $A^* \setminus \{a_1\}$  and  $X_a = X_a \cup \{f^i(a_2)\}$ ; /* eg., single  $\perp$  married */
8.   end
9.   foreach attribute  $a$  from set  $A^*$  do
10.    if the number of occurrences of  $a$  in triple  $(V^{c_i}, f^{c_i})$  is smaller than  $n/2$ 
        then set  $A^* := A^* \setminus \{a\}$ ;
11.  end
12.  foreach attribute  $a$  from set  $A^*$  do
13.    Determine multi-set  $X_a = \{f^i(a) : \text{if } f^i(a) \text{ exists and } i = 1, \dots, n\}$ ;
14.    Calculate  $f^*(a) := \frac{1}{\text{card}(X_a)} \sum_{v \in X_a} v$ ;
15.  end
16.  Return  $(A^*, V^*, f^*)$ ;
END.
```

3.4.3. Đánh giá thuật toán

Trong bài toán tích hợp mức khái niệm, mâu thuẫn giữa các *lược tả mờ thuộc tính* là phụ thuộc, vì vậy theo lược đồ lựa chọn tiêu chí đồng thuận (mục 2.4), tiêu chí C_2 được sử dụng. Từ việc phân tích các tiêu chí đồng thuận và *định lý 3.1*, chứng minh được rằng thuật toán *FOI-1* thỏa mãn các tiêu chí đồng thuận sau: Un, Si, Qu, Co, C_2 . Thuật toán có độ phức tạp là $O(n^2)$.

3.5. Tích hợp ontology mờ ở mức quan hệ

3.5.1. Mâu thuẫn ontology mờ mức quan hệ

Định nghĩa 3.5.1 (Mâu thuẫn ontology mờ mức quan hệ):

Cho hai ontology mờ O_1 và O_2 có các khái niệm c và c' . Mâu thuẫn mức quan hệ xảy ra nếu $R_{i1}(c, c') \neq R_{i2}(c, c')$, $i \in \{1, \dots, m\}$.

3.5.2. Thuật toán tích hợp ontology mờ mức quan hệ

Phát biểu bài toán FOI-2.1:

Cho $i \in \{1, \dots, m\}$ và tập các quan hệ $X = \{R_{ij}(c, c') : i=1, \dots, m; j = 1, \dots, n\}$ giữa 2 khái niệm c và c' trong n ontology, cần xác định $R_i(c, c')$ - quan hệ tốt nhất giữa c và c' trong tập các quan hệ đã cho thỏa tiêu chí đồng thuận.

Thuật toán FOI-2.1:

INPUT: Cho tập các quan hệ giữa 2 khái niệm c và c' trong n ontology:

$X = \{R_{ij}(c, c') : j = 1, \dots, n\}$

OUTPUT: Quan hệ $R_i(c, c') = (c, c', v)$ tốt nhất trong tập các quan hệ X thỏa tiêu chí đồng thuận.

BEGIN

1. Order set X in increasing order giving $X = \{x_1, x_2, \dots, x_n\}$;

2. Set interval $\langle X_{\lfloor \frac{n+1}{2} \rfloor}, X_{\lfloor \frac{n+2}{2} \rfloor} \rangle$

3. Set v as a value belonging to the above defined interval: $\langle X_{\lfloor \frac{n+1}{2} \rfloor}, X_{\lfloor \frac{n+2}{2} \rfloor} \rangle$

END.

Phát biểu bài toán FOI-2.2. (Tích hợp quan hệ bậc cầu).

Cho $i = 1, \dots, m$ và tập các quan hệ giữa hai khái niệm c và c' trong n ontology: $X = \{R_{ij} \subseteq C \times C \times (0, 1] : j = 1, \dots, n\}$. Cần xác định $R_i(c, c')$ quan hệ $R_i \subseteq C \times C \times (0, 1]$ tốt nhất trong tập các quan hệ X đã cho.

Thuật toán FOI-2.2:

INPUT: - Tập quan hệ cùng loại giữa các khái niệm trong n ontology $X = \{R_{ij} \subseteq C \times C \times (0, 1] : j =$

$1, \dots, n\}$

- Quan hệ có tính bắc cầu.

OUTPUT: Quan hệ $R_i \subseteq C \times C \times (0, 1]$ tốt nhất của X thỏa tiêu chí đồng thuận.

BEGIN

1. Set $R_i = \phi$;

2. for each pair $(c, c') \in C \times C$ do

3. Determine multi-set $X(c, c') = \{v: \langle c, c', v \rangle \in R_{ij}; i=1, \dots, m; j= 1, \dots, n\}$;

4. Order set $X(c, c')$ in increasing order giving $X = \{x_1, x_2, \dots, x_k\}$;

5. Set interval $(x_{\lfloor \frac{k+1}{2} \rfloor}, x_{\lfloor \frac{k+2}{2} \rfloor})$;

6. Set v as a value belonging to the above defined interval;

7. Set $R_i = R_i \cup \{\langle c, c', v \rangle\}$;

8. end

9. for each $(c, c', c'') \in C \times C \times C$ do

10. if $\langle c, c', v_1 \rangle \in R_i, \langle c', c'', v_2 \rangle \in R_i$ and $\langle c, c'', v_3 \rangle \in R_i$ then change $v_3 = \min(v_1, v_2)$;

11. if only $\langle c, c', v_1 \rangle \in R_i$ and $\langle c', c'', v_2 \rangle \in R_i$ then Set $R_i = R_i \cup \{\langle c, c'', v_3 \rangle\}$ where $v_3 = \min(v_1, v_2)$;

12. end

13. Return(R_i);

END.

3.5.3. Đánh giá thuật toán

Trong bài toán tích hợp mức quan hệ, mâu thuẫn giữa các lược tả mờ quan hệ là độc lập, vì vậy theo lược đồ lựa chọn tiêu chí đồng thuận (mục 2.4), tiêu chí C_1 được sử dụng. Từ việc phân tích các tiêu chí đồng thuận, *định lý 3.2* và *thuật toán 3.1*, chúng minh được rằng các thuật toán *FOI-2.1* và *FOI-2.2* thỏa các tiêu chí đồng thuận sau:

Un, Si, Qu, Co, Pr, C_1 .

• *Thuật toán FOI-2.1* có độ phức tạp là $O(n^2)$.

• *Thuật toán FOI-2.2* có độ phức tạp là $O(n^3)$.

3.6. Tích hợp ontology mờ ở mức thực thể

3.6.1. Mâu thuẫn ontology mờ mức thực thể

Định nghĩa 3.6.1 (Thực thể mờ): Một thực thể mờ của khái niệm c được mô tả bởi các thuộc tính của tập A^c có các giá trị thuộc tập 2^{V^x} ($X = A^c$) là cặp (i, v) , trong đó:

- i là định danh của thực thể, v là giá trị của thực thể,
- v là một bộ có kiểu A^c được biểu diễn là một hàm $v: A \rightarrow \bar{2}^{A^c}$ với $v(a) \in 2^{V^a}, \forall a \in A^c$.

Định nghĩa 3.6.2 (Mâu thuẫn ontology mờ mức thực thể):

Cho hai ontology O_1 và O_2 và các khái niệm $(c_1, A^{c_1}, V^{c_1}, f^1)$ thuộc O_1 , $(c_2, A^{c_2}, V^{c_2}, f^2)$ thuộc O_2 . Cho các thực thể $(i, v_1) \in (O_1, c_1)$ và $(i, v_2) \in (O_2, c_2)$. Mâu thuẫn thực thể xảy ra nếu $v(a) \neq v'(a)$, $a \in A^c \cap A^{c'}$.

3.6.2. Tiêu chí đồng thuận cho tích hợp ontology mờ mức thực thể mờ

Các tiêu chí đồng thuận cho xử lý mâu thuẫn tri thức (Nguyen, 2008) được sử dụng cho bài toán tích hợp ontology mờ ở mức thực thể như sau:

P₁. *Thực thể đóng:* $t^* < \bigcup_{i=1}^n t_i$

P₂. *Thực thể nhất quán:* $\bigcap_{i=1}^n t_i < t^*$

P₃. *Thực thể ưu việt:* Nếu tập các thuộc tính T_i ($i = 1, 2, \dots, n$) là rời nhau thì: $t^* = [\bigcup_{i=1}^n t_i]_{T^*}$, với $[\bigcup_{i=1}^n t_i]_{T^*}$ là tổng $\bigcup_{i=1}^n t_i$ thu hẹp về các thuộc tính của T^* .

P₄. *Độ tương đồng tối đa:* Gọi d_a là hàm khoảng cách giữa các thuộc tính $a \in A$, sự khác biệt giữa tích hợp t^* và các phần tử của lược tả cần được tối thiểu: $\forall a \in T^*$ thì $\sum_{r \in Z_a} d(t_a^*, r)$ là nhỏ nhất, với $Z_a = \{r_{ia}: r_i \in Z, i = 1, 2, \dots, n\}$.

3.6.3. Thuật toán tích hợp ontology mờ mức thực thể

Phát biểu bài toán FOI-3:

Cho tập hợp các thực thể $X = \{(i, v_1), \dots, (i, v_n)\}$, với $v_i \in A_i \subseteq A$, $v_i: A_i \rightarrow V_i (i = 1, \dots, n)$, $V_i = \bigcup_{a \in V A_i} V_a$, cần xác định cặp (i, v) tốt nhất trong tập X thỏa các tiêu chí của lý thuyết đồng thuận.

Thuật toán FOI-3:

INPUT: tập các mô tả của các thực thể: $X = \{r_i \in TUPLE(T_i) : T_i \subseteq A, i = 1, 2, \dots, n\}$

và hàm khoảng cách d_a cho các thuộc tính $a \in A$, $d_a: 2^{V_a} \times 2^{V_a} \rightarrow [0, 1]$

d_a là hàm được xác định như sau $d_a: 2^{V_a} \times 2^{V_a} \rightarrow [0, 1]$

OUTPUT: bộ $t^* \in T^* \subseteq A$ là đại diện tốt nhất theo các tiêu chí đồng thuận.

BEGIN

1. $A = \bigcup_{i=1}^n T_i$;

2. For each $a \in A$ determine a set with repetitions $X_a = \{t_{ia} : t_i \in X \text{ for } i = 1, 2, \dots, n\}$;

3. For each $a \in A$ using distance function d_a determine a value $v_a \subseteq V_a$ such that

$$\sum_{r_{ia} \in X_a} d_a(v_a, r_{ia}) = \min_{v'_a \subseteq V_a} \sum_{r_{ia} \in X_a} d_a(v'_a, r_{ia})$$

4. Create tuple t^* consisting of values v_a for all $a \in A$;

END

3.6.4. Đánh giá thuật toán

Tại *bước 3*, cho thấy thỏa mãn các tiêu chí C_1 và P_4 . Theo kết quả tích hợp ở *bước 4*: t^* bao gồm các giá trị v_a với mọi $a \in A$, chứng tỏ rằng thuật toán thỏa P_1 , P_2 và P_3 . Từ việc phân tích các tiêu chí đồng thuận chứng minh được rằng *thuật toán FOI-3* thỏa các tiêu chí đồng thuận sau: $C_1, P_1, P_2, P_3, P_4, Un, Si, Qu, Co, Pr$.

- Thuật toán có độ phức tạp $O(n^2)$.

3.7. Tổng kết chương

Chương 3 trình bày các kết quả nghiên cứu đóng góp mới của luận án, bao gồm: phân tích các tiêu chí đồng thuận liên quan đến bài toán

tích hợp ontology mờ; xây dựng mô hình tri thức ontology mờ cho bài toán tích hợp; xây dựng phương pháp so khớp ontology mờ dựa trên phần chung tiềm năng và các thuật toán tích hợp ontology mờ theo ba mức khái niệm, quan hệ và thực thể trên cơ sở của lý thuyết đồng thuận.

CHƯƠNG 4: THỰC NGHIỆM VÀ ĐÁNH GIÁ

4.1. Thử nghiệm 1

Thuật toán so khớp ontology mờ dựa trên phần chung tiềm năng được thử nghiệm theo tiêu chí và dữ liệu của hệ thống *OAEI* được mờ hóa theo lý thuyết mờ (Zadel, 1965) và ngôn ngữ *OWL2* (Bobillo & Straccia, 2011) các kết quả thử nghiệm công bố tại [1, 9,10].

4.1.1. Dữ liệu thử nghiệm

4.1.1.1. Dữ liệu theo chuẩn *OAEI*

Tập dữ liệu thử nghiệm sử dụng để đánh giá phương pháp so khớp *PCP* theo chuẩn *OAEI* 2013 như sau:

101 - # 104 (*data set 1*) gồm 33 lớp, 24 quan hệ, 40 thuộc tính, 56 thực thể.

201- # 264 (*data set 2*) gồm 36 lớp, 26 quan hệ, 46 thuộc tính, 32 thực thể.

301 - # 304 (*data set 3*) gồm 56 lớp, 72 quan hệ, 25 thuộc tính

4.1.1.2. Dữ liệu ontology mờ thời tiết

Các dữ liệu ontology được mờ hóa theo chuẩn *OWL2* (Bobillo & Straccia, 2011). Bảng 3.1 thống kê các thành phần của các ontology mờ thời tiết đã xây dựng.

Bảng 4.1. Thống kê các thành phần của ontololy mờ thời tiết

<i>FuzzyOntology</i>	<i>#Concept</i>	<i>#DataPro</i>	<i>#ObjectPro</i>	<i>#Instance</i>
<i>Weather1.owl</i>	153	36	56	103
<i>Weather2.owl</i>	146	45	46	79
<i>Weather3.owl</i>	138	40	48	80

4.1.2. Phương pháp đánh giá

Các độ đo đánh giá trong thử nghiệm là *precision*, *recall* và *f-measure* như sau:

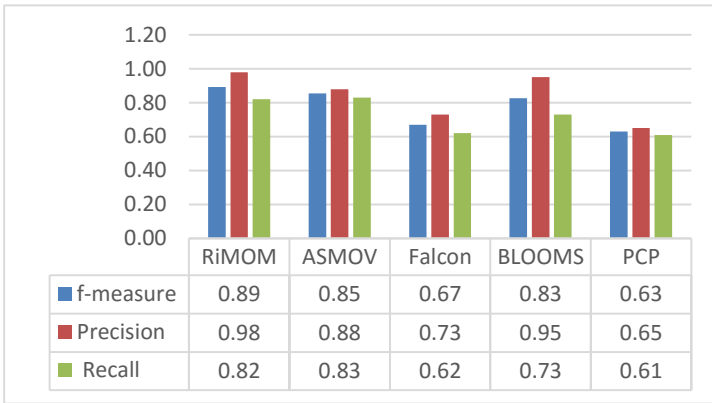
$$Precision = \frac{N_{correct}}{N_{correct} + N_{incorrect}} \quad (4.1)$$

$$Recall = \frac{N_{correct}}{N_{total}} \quad (4.2)$$

$$F = \frac{2 * Precision * Recall}{Precision + Recall} \quad (4.3)$$

4.1.3. Kết quả và nhận xét

4.1.3.1. Kết quả thử nghiệm trên dữ liệu OAEI



Hình 4.1. So sánh kết quả so khớp của *PCP* với 4 hệ thống khác trên dữ liệu thử nghiệm của *OAEI*

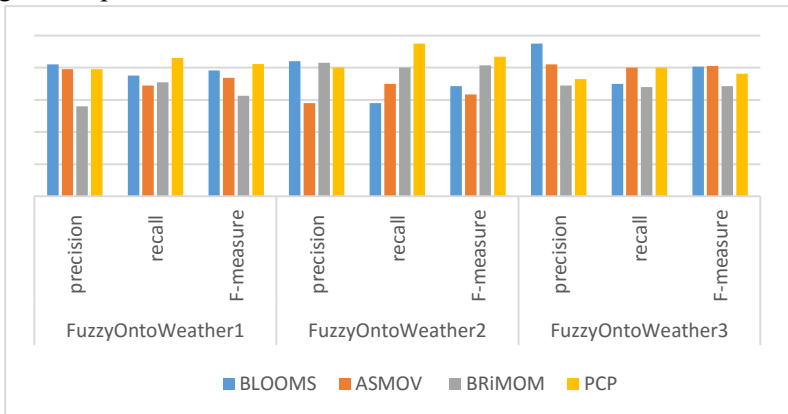
Các quan sát trong Hình 4.1 cho thấy kết quả so khớp của *PCP* với 4 hệ thống so khớp tham gia *OAEI* năm 2010 là *ASMOV*, *RiMOM*, *BLOOMS* và *Falcon-AO*. Quan sát trên đồ thị, kết quả so khớp của *RiMOM* là tốt nhất. Phương pháp *PCP* tốt hơn so với *Falcon-AO* về cả ba độ đo, trong đó độ đo *F-measure* của *PCP* cao hơn so với *RiMOM*, *ASMOV*, *BLOOMS* và *Falcon-AO*.

Kết quả đánh giá cho thấy độ đo *precision* của *PCP* tốt hơn các hệ thống trong khi đó vẫn đảm bảo được *Recall*. Lý do là thay vì phải kiểm tra trên toàn bộ các cặp khái niệm, thuật toán *PCP* đã thu giảm

không gian tìm kiếm bằng cách xác định phần chung tiềm năng với cách tính khoảng cách nhỏ nhất để tìm ra *PCP* vì vậy giảm thiểu độ phức tạp và tăng độ chính xác của thuật toán.

4.1.3.2. Kết quả thử nghiệm trên ontology mờ thời tiết

Quan sát trên hình 4.2 cho thấy kết quả đánh giá cho thấy độ đo precision của *PCP* tốt hơn ba hệ thống ngoại trừ *RiMOM* trong khi đó vẫn đảm bảo được Recall. Lý do là thay vì phải kiểm tra trên toàn bộ các cặp khái niệm, thuật toán *PCP* đã thu giảm không gian tìm kiếm bằng cách xác định phần chung tiềm năng để so khớp, vì vậy giảm độ phức tạp của thuật toán.



Hình 4.2. So sánh phương pháp *PCP* và các hệ thống khác trên ontology mờ thời tiết

4.2. Thử nghiệm 2

Phần này trình bày cài đặt thực thi quá trình tích hợp và phân tích để đánh giá chất lượng xử lý mâu thuẫn của các thuật toán tích hợp ontology mờ dựa trên lý thuyết đồng thuận.

Hệ thống tích hợp ontology mờ *FOIS* được xây dựng để đánh giá thử nghiệm các thuật toán *FOI-1*, *FOI-2* *FOI-3* có các đặc điểm sau:

- Dạng dữ liệu input: File RDF hoặc OWL
- Thao tác so khớp sẽ thực hiện trước khi thực hiện xử lý mâu thuẫn
- Tích hợp hoàn toàn tự động

- Có thể tích hợp được nhiều ontology tại một thời điểm
- Xử lý mâu thuẫn theo các mức khái niệm, quan hệ và thực thể.
- Giao diện thân thiện, dễ sử dụng

Các chức năng tích hợp của hệ thống FOIS được trình bày trong *Phụ lục C* của luận án với các kết quả thử nghiệm đã công bố [1, 9, 10].

4.2.1. Dữ liệu

Dữ liệu thử nghiệm là các ontology mờ thời tiết được xây dựng theo phương pháp trình bày Mục 4.1 được mô tả theo các bảng sau:

Bảng 4.2. Mô tả thuộc tính của các ontology mờ thời tiết

Các Ontology → Các Thuộc tính ↓	FuzzyOnto1	FuzzyOnto2	...	FuzzyOnto5
<i>hasAltitude</i>	0.71	0.81	...	0.67
<i>hasAtoms</i>	0.82	0.62	...	0.48
<i>hasDegree</i>	0.45	0.25	...	0.11
<i>hasHumidity</i>	0.34	0.44	...	0.33
<i>hasPressure</i>	0.37	0.37	...	0.23
<i>hasProtons</i>	0.56	0.47	...	0.33
<i>hasRain</i>	0.72	0.54	...	0.36
....
<i>hasSaltPercent</i>	0.55	0.89	...	0.75

Bảng 4.3. Mô tả quan hệ của các ontology mờ thời tiết

Các quan hệ → Các ontology ↓	<i>belongs to state</i>	...	<i>has source</i>	<i>has weather phenomenon</i>	<i>weather state</i>
FuzzyOntoWeather1	0.71	...	0.88	0.80	0.91
FuzzyOntoWeather2	0.82	...	0.51	0.20	0.19
FuzzyOntoWeather3	0.68	...	0.53	0.70	0.83
FuzzyOntoWeather4	0.90	...	0.64	0.30	0.29
FuzzyOntoWeather5	0.37	...	0.15	0.60	0.40

Bảng 4.4. Mô tả thực thể của các ontology mờ thời tiết

Các Thuộc tính → Các Thực thể ↓	Wind	...	Temperature	Pressure	Rain
PARIS	0.42	...	0.63	0.50	0.58
HANOI	0.83	...	0.48	0.75	0.55
HCMC	0.78	...	0.58	0.58	0.89
KOREA	0.66	...	0.56	0.73	0.68
HUE	0.63	...	0.49	0.47	0.89

4.2.2. Phương pháp đánh giá

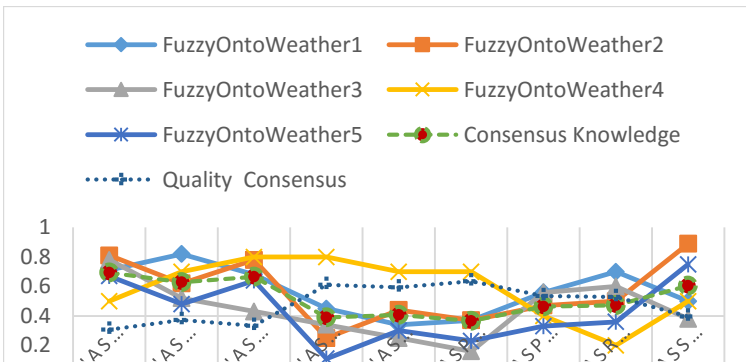
Theo Định nghĩa 3.3.1 (Mục 3.3) chất lượng của sự đồng thuận x trong lược tả X được định nghĩa theo công thức sau:

$$d^*(x, X) = 1 - \frac{d(x, X)}{|X|} \quad (4.4)$$

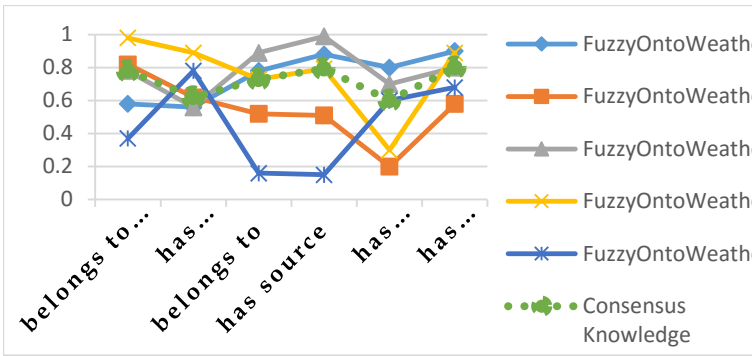
Trong đó: $X \in \Pi(U)$, $C \in \text{Con}(U)$, $x \in C(X)$ và $d(x, X) = \sum_{i=1}^n d(x, x_i)$.

Theo (Nguyen, 2008a), chất lượng đồng thuận được sử dụng để so sánh chất lượng của sự đồng thuận trong các lược tả khác nhau.

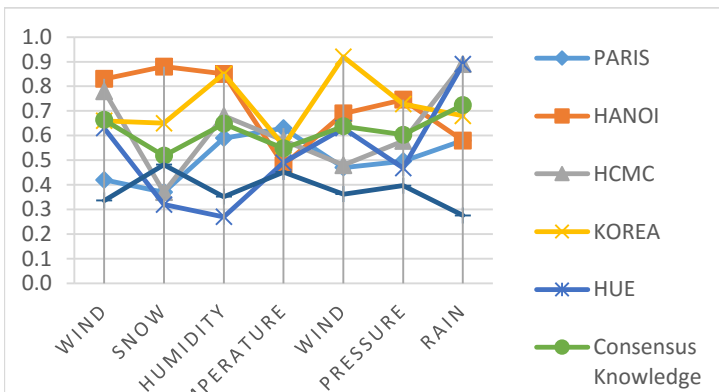
4.2.3. Kết quả và thảo luận



Hình 4.3. Chất lượng tích hợp mức Khái niệm



Hình 4.4. Kết quả tích hợp mức quan hệ



Hình 4.5. Chất lượng tích hợp mức thực thể

Từ kết quả thử nghiệm sử dụng các bộ dữ liệu ontology mở thời tiết với các mức độ mâu thuẫn khác nhau có được các nhận xét sau đây:

- Các lược tả mâu thuẫn có mức mâu thuẫn cao sẽ có chất lượng tích hợp sẽ tốt hơn các lược tả mâu thuẫn có mức mâu thuẫn thấp.
- Chất lượng tích hợp tỷ lệ thuận với số lượng thành viên mâu thuẫn tham gia tích hợp.

- Hai lược tả mâu thuẫn có mức mâu thuẫn như nhau thì chất lượng tri thức tích hợp phụ thuộc vào tổng các khoảng cách từ các mâu thuẫn lược tả đến tri thức tích hợp.
- Khoảng cách từ các lược tả mâu thuẫn đến tri thức đồng thuận không vượt quá mức trung bình của khoảng cách từ lược tả mâu thuẫn đến tri thức đồng thuận.

Các kết quả nhận xét trên đã được chứng minh và công bố tại: (Nguyen, 2008a), (Nguyen, Du & Truong, 2016) và (Nguyen & Du, 2015a, 2015b).

Bảng 4.1 cho thấy hiệu quả của phương pháp đồng thuận trong tích hợp ontology mờ.

Bảng 4.1. Chất lượng xử lý mâu thuẫn của các hệ thống

Công cụ	Recall	Precision
PROMPT	0.5	0.7
Chimaera	0.6	0.33
FOIS	0.8	1

4.3. Tổng kết chương

Chương 4 trình bày hai thử nghiệm về các thuật toán so khớp và tích hợp ontology mờ, trong đó hệ thống tích hợp ontology mờ FOIS được xây dựng để thực hiện các thử nghiệm các thuật toán tích hợp ontology mờ. Các thử nghiệm sử dụng dữ liệu ontology của OAEI được mờ hóa và các ontology mờ thời tiết. *Thử nghiệm 1* cho thuật toán so khớp ontology mờ dựa trên phần chung tiềm năng PCP sử dụng độ đo *precision*, *recall* và *F-measure* cho kết quả tương đối tốt so với các phương pháp so khớp ontology khác. *Thử nghiệm 2* Phương pháp của *thử nghiệm 2* dựa trên định nghĩa về chất lượng của tri thức đồng thuận (Nguyen, 2008a) để xây dựng phương pháp thử nghiệm. Việc đánh giá chất lượng tích hợp tri thức dựa trên lý thuyết đồng thuận đã được công bố tại (Nguyen, Du & Truong 2015) và (Nguyen & Du 2015a, 2015b). Sử dụng các kết quả thử nghiệm

của ([Lambrix & Edberg, 2003](#); [Lambrix & Tan, 2007](#)) ([Noy & Musen, 2000](#)): PROMPT, Chimaera cho thấy hệ thống tích hợp ontology mờ *FOIS* có kết quả tốt hơn.

KẾT LUẬN VÀ KIẾN NGHỊ

4.4. Kết luận

Các kết quả đạt được của luận án được tóm tắt như sau:

- Đề xuất định nghĩa ontology mờ cho bài toán tích hợp.
- Xây dựng thuật toán so khớp khái niệm giữa các ontology mờ dựa trên phần chung tiềm năng.
- Xây dựng các thuật toán tích hợp ontology mờ dựa trên lý thuyết đồng thuận theo ba mức.
- Cài đặt và thử nghiệm các thuật toán tích hợp trên ontology mờ thời tiết và *OAEI*.

4.5. Hướng phát triển

Các vấn đề cần được tiếp tục nghiên cứu là:

- Phát triển ngôn ngữ mô tả mờ theo mô hình ontology mờ đã đề xuất để cài đặt các ứng dụng tích hợp tri thức ontology mờ phân tán.
- Xây dựng các cơ sở tri thức ontology mờ trong các lĩnh vực khác nhau để phục vụ cho các thử nghiệm và ứng dụng các thuật toán tích hợp ontology mờ.
- Xây dựng các công cụ mờ hóa ontology rõ sử dụng hàm thành viên phù hợp cho các ứng dụng liên quan.
- Phát triển các ứng dụng tích hợp các hệ thống tri thức ontology mờ trên các kết quả nghiên cứu của luận án.
- Phát triển ứng dụng của các thuật toán tích hợp tri thức ontology mờ trên các hệ thống tri thức phân tán: hệ thống đa tác tử, các hệ thống mạng xã hội,...trên cơ sở lý thuyết đồng thuận.