

THUYẾT MINH ĐỀ TÀI
NGHIÊN CỨU KHOA HỌC CẤP CƠ SỞ

1. TÊN ĐỀ TÀI HỌC KHÁI NIỆM ĐỐI VỚI CÁC CƠ SỞ TRI THỨC TRONG LOGIC MÔ TẢ DỰA VÀO MÔ PHỎNG HAI CHIỀU			2. MÃ SỐ
3. LĨNH VỰC NGHIÊN CỨU Tự nhiên: <input checked="" type="checkbox"/> Kinh tế, XHNV: <input type="checkbox"/> Giáo dục: <input type="checkbox"/> Kỹ thuật: <input type="checkbox"/> Nông lâm: <input type="checkbox"/> Y Dược: <input type="checkbox"/> Môi trường: <input type="checkbox"/> ATLĐ: <input type="checkbox"/> Sở hữu trí tuệ: <input type="checkbox"/>		4. LOẠI HÌNH NGHIÊN CỨU Cơ bản <input checked="" type="checkbox"/> Ứng dụng <input type="checkbox"/> Triển khai <input type="checkbox"/>	
5. THỜI GIAN THỰC HIỆN: 24 tháng Từ tháng 01 năm 2013 đến tháng 12 năm 2014			
6. CƠ QUAN CHỦ TRÌ Tên cơ quan: TRƯỜNG ĐẠI HỌC KHOA HỌC Địa chỉ: 77, Nguyễn Huệ, Thành phố Huế, Tỉnh Thừa Thiên Huế. Điện thoại: 054 3823290 Fax: 054 3824901 Email: doingoidhkh@gmail.com			
7. CHỦ NHIỆM ĐỀ TÀI Họ và tên: TRẦN THANH LƯƠNG Học vị, chức danh khoa học: Thạc sĩ Chức vụ: Giảng viên Địa chỉ CQ: 77, Nguyễn Huệ, Tp Huế Địa chỉ NR: M15, Đông Nam Thủy An, Hương Thủy Điện thoại CQ: 054 3826767 Điện thoại NR: Điện thoại di động: 091 4145414 Fax: 054 3824901 E-mail: tluong@hueuni.edu.vn; tluong@gmail.com			
8. NHỮNG NGƯỜI THAM GIA THỰC HIỆN ĐỀ TÀI			
Họ và tên	Đơn vị công tác và lĩnh vực chuyên môn	Nội dung cụ thể được giao	Chữ ký
Hoàng Thị Lan Giao	Khoa Công nghệ Thông tin, Trường Đại học Khoa học	Thư ký đề tài Nghiên cứu về lý thuyết học máy. Đánh giá độ phức tạp thuật toán.	

9. ĐƠN VỊ PHỐI HỢP CHÍNH		
Tên đơn vị trong và ngoài nước	Nội dung phối hợp nghiên cứu	Họ tên người đại diện
Viện Tin học, Đại học Warsaw, Ba Lan.	Kiểm chứng và đánh giá thuật toán được đề xuất.	PGS. TSKH. Nguyễn Anh Linh
10. TÌNH HÌNH NGHIÊN CỨU THUỘC LĨNH VỰC CỦA ĐỀ TÀI TRONG VÀ NGOÀI NƯỚC <p>10.1. Ngoài nước (phân tích, đánh giá tình hình nghiên cứu thuộc lĩnh vực của đề tài trên thế giới, liệt kê danh mục các công trình nghiên cứu, tài liệu có liên quan đến đề tài được đánh trích dẫn khi đánh giá tổng quan).</p> <p>Học máy là một trong những lĩnh vực được nhiều nhà khoa học trên thế giới nghiên cứu trong nhiều ngữ cảnh khác nhau với nhiều phương pháp khác nhau. Đối với các hệ thống thông tin trong giả thiết thế giới đóng (Closed World Assumption), các nghiên cứu đã đạt những kết quả khá quan trọng. Tuy nhiên, đối với các hệ thống thông tin trong giả thiết thế giới mở (Open World Assumption) [1, 6], các nghiên cứu tập trung chủ yếu vào các hệ thống thông tin trong ngữ cảnh logic mô tả [2]. Logic mô tả tuy là một lĩnh vực mới nhưng được nhiều nhà khoa học nghiên cứu và có nhiều đóng góp trong thời gian gần đây. Dựa vào khả năng biểu diễn tri thức và suy luận của logic mô tả, bài toán học khái niệm trong ngữ cảnh logic mô tả đã được đặt ra như là một hướng nghiên cứu đầy hứa hẹn đối với bài toán học máy. Các nghiên cứu gần đây của các tác giả đối với bài toán này chủ yếu tiếp cận theo hướng xây dựng các toán tử làm mịn [3-5], [7-9]. Quá trình xây dựng các toán tử làm mịn tương đối phức tạp và độ chính xác trong quá trình phân lớp chưa cao.</p> <p>Logic mô tả là một biến thể của logic khả năng. Dựa vào những tính chất mô phỏng hai chiều của logic khả năng, A. L. Nguyen và Andrzej Szalas xây dựng được quan hệ mô phỏng hai chiều trong logic mô tả. Quan hệ mô phỏng hai chiều đã đem lại một khả năng mới cho việc học máy đối với hệ thống thông tin dựa trên logic mô tả [9]. Điều này mở ra một hướng mới đầy triển vọng cho việc nghiên cứu các thuật toán học khái niệm.</p> <p>Danh mục các công trình liên quan:</p> <p>[1]. Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., & Ives, Z. <i>DBpedia: A Nucleus for a Web of Open Data</i>. Proceedings of the 6th international semantic web conference (ISWC), In Lecture Notes in Computer Science, volume 4825, pages 722–735, Springer, 2008</p> <p>[2]. Baader, F., Calvanese, D., McGuinness, D. L., Nardi, D., & Patel-Schneider, P. F. (Eds.), <i>The Description Logic Handbook: Theory, Implementation, and Applications</i>, Cambridge University Press, Cambridge, 2007.</p> <p>[3]. Badea, L, <i>Perfect Refinement Operators Can be Flexible</i>. In W. Horn (Ed.), Proceedings of the 14th European Conference on Artificial Intelligence, pages 266-270, Amsterdam: IOS Press, 2000.</p> <p>[4]. Badea, L., Shan-Hwei Nienhuys-Cheng, <i>A Refinement Operator for Description Logics</i>, Proceeding ILP '00 Proceedings of the 10th International Conference on Inductive Logic Programming, pages 40-59, Springer, 2000.</p> <p>[5]. Fanizzi, N., d'Amato, C, and Esposito, F., <i>DL-FOIL Concept Learning in Description Logics</i>, Proceeding ILP '08 Proceedings of the 18th International Conference on Inductive Logic Programming,</p>		

pages 107-121, Springer, 2008.

[6]. Hellmann, S., Lehmann, J., Auer, S., *Learning of OWL Class Descriptions on Very Large Knowledge Bases*, International Journal On Semantic Web and Information Systems, 2009.

[7]. Lehmann, J., Hitzler, P., *Concept Learning in Description Logics Using Refinement Operators*, Machine Learning, volume 78, pages 203-250, Springer, 2010.

[8]. Lehmann, L., Hitzler, P., *A Refinement Operator Based Learning Algorithm for the ALC Description Logic*, In: Hendrick Blockeel, Jude W. Shavlik, Prasad Tadepalli (editors), Proceedings of the 17th International Conference on Inductive Logic Programming (ILP), Lecture Notes in Computer Science, volume 4894, pages 147-160, Springer, 2008.

[9]. L. Nguyen and A. Szałas, *Logic-based roughification*, in Rough Sets and Intelligent Systems (To the Memory of Professor Zdzisław Pawlak), A. Skowron and Z. Suraj, Eds. volume 1, pages 529–556, Springer, 2012.

10.2. Trong nước (phân tích, đánh giá tình hình nghiên cứu thuộc lĩnh vực của đề tài ở Việt Nam, liệt kê danh mục các công trình nghiên cứu, tài liệu có liên quan đến đề tài được đánh trích dẫn khi đánh giá tổng quan).

Tại Việt Nam, logic mô tả nói chung chỉ mới được quan tâm nghiên cứu trong những năm gần đây bởi các nhà khoa học ở Đại học Bách khoa Thành phố Hồ Chí Minh; Đại học Công nghệ thuộc Đại học Quốc Gia Hà Nội; Đại học Huế và tập trung chủ yếu vào việc ứng dụng khả năng biểu diễn tri thức, sử dụng bài toán suy luận trong web ngữ nghĩa. Các nghiên cứu chuyên sâu về logic mô tả như xây dựng các thuật toán suy luận hiệu quả, tối ưu hóa các thuật toán suy luận,... gần như chưa có những kết quả cụ thể nào. Đặc biệt bài toán học máy cho các hệ thống thông tin trong ngữ cảnh logic mô tả (học khái niệm) chưa có tác giả nào quan tâm nghiên cứu.

10.3. Danh mục các công trình đã công bố thuộc lĩnh vực của đề tài của chủ nhiệm và những thành viên tham gia nghiên cứu (Họ và tên tác giả, bài báo, ấn phẩm, các yếu tố về xuất bản).

[1]. Quang-Thuy Ha, Thi-Lan-Giao Hoang, Linh Anh Nguyen, Hung Son Nguyen, Andrzej Szałas and Thanh-Luong Tran, *A Bisimulation-based Method of Concept Learning for Knowledge Bases in Description Logics*. SoICT 2012 - 3rd International Symposium on Information and Communication Technology, Pages 241-249, Ha Long, August 23-24, 2012.

[2]. Thanh-Luong Tran, Quang-Thuy Ha, Thi-Lan-Giao Hoang, Linh Anh Nguyen, Hung Son Nguyen and Andrzej Szałas, *Concept Learning for Description Logic-based Information Systems*, KSE 2012-International Conference on Knowledge and Systems Engineering, Pages 65-73, Da Nang August 17-19, 2012.

[3]. Hoàng Thị Lan Giao, Trần Thanh Lương, *Phân lớp dữ liệu đa nhãn dựa trên các thuật toán học máy*, Kỷ yếu Trường Đại học Khoa học, 2011.

[4]. Hồ Thuần, Hoàng Thị Lan Giao, *Mở rộng một số toán tử quan hệ lên cơ sở dữ liệu thiếu thông tin*, Tạp chí Tin học và Điều khiển học, Số 4, 2003.

[5]. Hồ Thuần, Hoàng Thị Lan Giao, *Một thuật toán tìm tập rút gọn sử dụng ma trận phân biệt được*, Chuyên san Các công trình nghiên cứu triển khai Viễn thông và CNTT, Số 15, 2005.

[6]. Hồ Thuần, Hoàng Thị Lan Giao, *Khám phá phụ thuộc và phụ thuộc xấp xỉ dựa vào ma trận phụ thuộc*, Tạp chí Tin học và Điều khiển học, Số 1, 2006.

11. TÍNH CẤP THIẾT CỦA ĐỀ TÀI

Học máy là một lĩnh vực của trí tuệ nhân tạo liên quan đến việc thiết kế và phát triển các thuật toán cho phép máy tính phân tích được các thông tin dựa trên các dữ liệu thực nghiệm. Học máy được áp dụng trong nhiều lĩnh vực như chẩn đoán y tế, phân tích thị trường chứng khoán, phân lớp dữ liệu văn bản, ... Một trong những nhánh của học máy là học có giám sát, tức là phải học một hàm dự đoán từ các mẫu cho trước. Cụ thể là cho một tập các mẫu $(x, f(x))$ của một hàm f , xây dựng một hàm h sao cho h xấp xỉ với f . Trong học máy, mỗi đối tượng thường được đặc tả bởi một tập các thuộc tính và x được xem như một vector các giá trị thuộc tính. Nếu f là một hàm nhị phân thì bài toán học máy được gọi là bài toán phân lớp nhị phân.

Logic mô tả là một họ các ngôn ngữ hình thức thích hợp cho việc biểu diễn và suy luận tri thức. Trong logic mô tả, một miền được mô tả thông qua tập các cá thể, khái niệm và vai trò. Một khái niệm đại diện cho một tập các đối tượng, một vai trò đại diện cho một quan hệ hai ngôi giữa các đối tượng.

Bài toán phân lớp nhị phân trong ngữ cảnh logic mô tả được gọi là học khái niệm, hàm phân loại ở đây được đặc trưng bởi một khái niệm. Học khái niệm trong logic được nghiên cứu bởi nhiều nhà nghiên cứu. Badea và Nienhuys-Cheng¹ nghiên cứu học khái niệm trong logic mô tả thông qua các toán tử làm mịn. Cohen và Hirsh² đã đề xuất một thuật toán học khái niệm LCSLearn dựa trên “bao hàm phổ biến nhỏ nhất”. Lambrix và Larocchia³ đưa ra thuật toán học khái niệm đơn giản dựa trên sự chuẩn hóa khái niệm.

Tuy nhiên, những nghiên cứu trong các công trình đã đề cập trên áp dụng cho các hệ thống thông tin trong ngữ cảnh logic mô tả với bộ ký tự và các cấu tử khái niệm còn hạn chế. Do vậy việc tổng quát hóa và mở rộng phương pháp học khái niệm cho các cơ sở tri thức trong ngữ cảnh logic mô tả dựa trên mô phỏng hai chiều là một vấn đề cấp thiết hiện nay.

12. MỤC TIÊU ĐỀ TÀI

Mở rộng cơ sở lý thuyết và phương pháp học khái niệm cho hệ thống thông tin trong ngữ cảnh logic mô tả dựa trên mô phỏng hai chiều. Đề xuất mô hình học khái niệm dựa trên quan hệ mô phỏng hai chiều để giải quyết bài toán phân lớp dữ liệu trong logic mô tả.

13. ĐỐI TƯỢNG, PHẠM VI NGHIÊN CỨU

13.1. Đối tượng nghiên cứu

- Các logic mô tả và ngữ nghĩa của các logic mô tả,
- Các hệ thống thông tin dựa trên logic mô tả,
- Học khái niệm và các thuật toán học khái niệm dựa trên mô phỏng hai chiều.

13.2. Phạm vi nghiên cứu

- Nghiên cứu logic mô tả, hệ thống thông tin trong ngữ cảnh logic mô tả,
- Nghiên cứu các hệ thống cơ sở tri thức, các tập dữ liệu và các thuật toán học máy trong ngữ cảnh

¹ L. Badea and S.-H. Nienhuys-Cheng, “A refinement operator for description logics,” in Proceedings of ILP’2000, ser. LNCS, vol. 1866. Springer, 2000, pp. 40–59.

² W. Cohen and H. Hirsh, “Learning the Classic description logic: Theoretical and experimental results,” in Proceedings of KR’1994, 1994, pp. 121–133.

³ P. Lambrix and P. Larocchia, “Learning composite concepts,” in Proceedings of DL’1998, 1998.

14. CÁCH TIẾP CẬN, PHƯƠNG PHÁP NGHIÊN CỨU

14.1. Cách tiếp cận

- Bắt đầu với các hệ thống thông tin trong ngữ cảnh logic mô tả và khảo sát các thuật toán học khái niệm, sau đó phân tích, tổng hợp, hệ thống hóa các tài liệu liên quan đến vấn đề học máy, phân lớp dữ liệu, logic mô tả và học khái niệm nhằm xây dựng cơ sở lý thuyết về vấn đề nghiên cứu.
- Xây dựng quan hệ mô phỏng hai chiều và đề xuất phương pháp học khái niệm dựa vào mô phỏng hai chiều.
- Thu thập, nghiên cứu, đánh giá các cơ sở tri thức cụ thể gồm những cơ sở tri thức truyền thống, tức là những cơ sở tri thức không có tập vai trò và những tập cơ sở tri thức như gen ontology, protein ontology, ... trên các phương diện kích thước dữ liệu, độ phức tạp của dữ liệu để nắm bắt các đặc trưng, tính chất của các cơ sở tri thức đó nói riêng và các tập cơ sở tri thức tổng quát nói chung.

14.2. Phương pháp nghiên cứu

Qua các tài liệu đã có và qua quá trình nghiên cứu các tài liệu đó, xây dựng các định lý và chứng minh các định lý nêu ra nhằm làm nền tảng vững chắc cho cơ sở lý luận của đề tài.

Xây dựng các phương pháp, thuật toán, kỹ thuật có chứng minh tính đúng đắn dựa trên cơ sở lý thuyết mà đề tài sẽ đưa ra.

15. NỘI DUNG NGHIÊN CỨU VÀ TIẾN ĐỘ THỰC HIỆN

15.1. Nội dung nghiên cứu (trình bày dưới dạng đề cương nghiên cứu chi tiết)

- Hệ thống thông tin trong ngữ cảnh logic mô tả
 - Tổng quan về logic mô tả
 - Logic mô tả cơ bản ALC và các logic mở rộng
 - Suy luận trong logic mô tả
 - Hệ thống thông tin trong ngữ cảnh logic mô tả
 - Biểu diễn tri thức trong logic mô tả
- Quan hệ mô phỏng hai chiều
 - Giới thiệu về mô phỏng hai chiều
 - Xây dựng quan hệ mô phỏng hai chiều
 - Tính phân biệt được của các đối tượng
 - Phân hoạch thông qua quan hệ mô phỏng hai chiều
- Thuật toán học máy dựa trên mô phỏng hai chiều
 - Học máy trong ngữ cảnh logic mô tả
 - Học khái niệm cho các hệ thống thông tin dựa trên mô phỏng hai chiều
 - Thuật toán học khái niệm trong cơ sở tri thức
- Kết luận và hướng phát triển

15.2. Tiến độ thực hiện				
STT	Các nội dung và công việc thực hiện chủ yếu	Sản phẩm	Thời gian bắt đầu-kết thúc	Người thực hiện
1	Hệ thống hóa các lý thuyết về logic mô tả và hệ thống thông tin.	Lý thuyết về logic mô tả	Từ 01/2013 đến 06/2013	Trần Thanh Lương
2	Xây dựng lý thuyết về thuật toán dựa trên mô phỏng hai chiều trong ngữ cảnh logic mô tả.	Các định lý về mô phỏng hai chiều (bài báo 1)	Từ 07/2013 đến 12/2013	Hoàng Thị Lan Giao, Trần Thanh Lương
3	Xây dựng thuật toán dựa trên lý thuyết đã đề ra.	Đưa ra thuật toán (bài báo 2)	Từ 01/2014 đến 05/2014	Hoàng Thị Lan Giao, Trần Thanh Lương
4	Thu thập dữ liệu thực nghiệm, cài đặt thuật toán và đánh giá.	Chương trình thực nghiệm	Từ 06/2014 đến 10/2014	Hoàng Thị Lan Giao, Trần Thanh Lương
5	Tập hợp kết quả, viết báo cáo tổng kết đề tài.	Báo cáo tổng kết	Từ 11/2014 đến 12/2014	Trần Thanh Lương

16. SẢN PHẨM

16.1. Sản phẩm khoa học

Sách chuyên khảo:☐

Bài báo đăng Tạp chí nước ngoài:☐

Sách tham khảo:☐

Bài báo đăng Tạp chí trong nước:☒

Giáo trình:☐

Bài báo đăng HN/HT quốc tế:☒

16.2. Sản phẩm đào tạo

Nghiên cứu sinh:☐

Thạc sĩ:☐

Cử nhân:☒

16.3. Sản phẩm ứng dụng

Mẫu☐ Tài liệu dự báo☐ Qui phạm☐ Luận chứng kinh tế☐

Vật liệu☐ Bản kiến nghị☐ Giống cây trồng☐ Qui trình công nghệ☐

Thiết bị máy móc☐ Đề án☐ Báo cáo phân tích ...☐ Dây chuyền công nghệ ..☐

Tiêu chuẩn☐ Phương pháp☐ Bản quy hoạch☐ Chương trình máy tính ..☐

Sơ đồ, bản thiết kế ..☐ Giống vật nuôi☐

16.4. Các sản phẩm khác (ghi rõ sản phẩm gì)

16.5. Tên sản phẩm, số lượng và yêu cầu khoa học đối với sản phẩm

Số TT	Tên sản phẩm	Số lượng	Yêu cầu khoa học
1	Bài báo khoa học	02	Mới và có tính thời sự, đăng ở các hội nghị, tạp chí trong và ngoài nước.
2	Cử nhân	02	

17. HIỆU QUẢ (giáo dục và đào tạo, kinh tế-xã hội) Làm tài liệu tham khảo cho sinh viên đại học và học viên cao học và những người nghiên cứu chuyên ngành Khoa học Máy tính nói chung và logic mô tả cũng như học máy nói riêng.		
18. PHƯƠNG THỨC CHUYỂN GIAO KẾT QUẢ NGHIÊN CỨU VÀ ĐỊA CHỈ ỨNG DỤNG Báo cáo tổng kết đề tài làm tài liệu tham khảo cho sinh viên đại học, học viên cao học chuyên ngành khoa học máy tính. Công bố các kết quả trên một số tạp chí chuyên ngành, báo cáo tại các hội nghị, hội thảo khoa học chuyên ngành trong nước hoặc quốc tế. Địa chỉ ứng dụng: Khoa Công nghệ Thông tin của các trường đại học trong cả nước.		
19. KINH PHÍ THỰC HIỆN ĐỀ TÀI VÀ NGUỒN KINH PHÍ Tổng kinh phí (triệu đồng): 40.000.000 đồng (<i>Bốn mươi triệu đồng chẵn</i>) Trong đó: Ngân sách Nhà nước: 40.000.000 đồng (<i>Bốn mươi triệu đồng chẵn</i>) Các nguồn kinh phí khác: 0 đồng (<i>Không đồng</i>) Nhu cầu kinh phí từng năm: - Năm 2013: 20.000.000 đồng (<i>Hai mươi triệu đồng chẵn</i>) - Năm 2014: 20.000.000 đồng (<i>Hai mươi triệu đồng chẵn</i>)		
Dự trù kinh phí theo các mục chi phù hợp với nội dung nghiên cứu:		
Số TT	Mục chi	Số tiền (triệu đồng)
1	Thuê khoán chuyên môn	26.0
	• Tổng hợp về cơ sở lý thuyết	4.0
	• Xây dựng thuật toán	6.0
	• Thu thập và xây dựng các bộ dữ liệu thử nghiệm	4.0
	• Cài đặt thuật toán thực nghiệm	4.0
	• Đánh giá kết quả thực nghiệm	4.0
	• Thù lao điều hành	4.0
3	Dự hội nghị, hội thảo, seminar khoa học	7.0
	• Dự hội nghị khoa học KSE 2013 tại Hà Nội	5.0
	• Seminar tại bộ môn 1 lần/năm × 2 năm × 20 người × 50.000đ	2.0
4	Chi phí nghiệm thu	3.0
	• Nghiệm thu cấp cơ sở	1.5
	• Nghiệm thu cấp Đại học Huế	1.5

5	Quản lý phí (5%)	2.0
6	Mua tài liệu, dịch thuật, in ấn, văn phòng phẩm	2.0
Tổng cộng:		40.000.000 đồng
<i>(Bốn mươi triệu đồng chẵn)</i>		

Ngày tháng năm 2012

Cơ quan chủ trì

(Ký tên, đóng dấu)

Ngày 22 tháng 10 năm 2012

Chủ nhiệm đề tài

(Họ và tên, ký)

TRẦN THANH LƯƠNG

Cơ quan chủ quản duyệt

GIÁM ĐỐC ĐẠI HỌC HUẾ