



# NGHIÊN CỨU TÍCH HỢP MÔ HÌNH DỮ LIỆU TRONG TRUNG TÂM DỮ LIỆU NGÀNH DẦU KHÍ VIỆT NAM

Nghiên cứu sinh: Vũ Ngọc Trinh <sup>(1)</sup>; Cán bộ hướng dẫn: PGS. TS. Hà Quang Thụy <sup>(2)</sup> và GS. TSKH. Nguyễn Hùng Sơn <sup>(3)</sup>

<sup>(1)</sup> Viện Dầu khí Việt Nam, Tập đoàn Dầu khí Quốc gia Việt Nam; <sup>(2)</sup> Trường Đại học Công nghệ, Đại học Quốc gia Hà Nội; <sup>(3)</sup> Đại học Warsaw University, Ba Lan



## Giới thiệu

Trong bài trình bày này chúng tôi trình bày các nội dung chính như sau:

- ▶ Thực trạng và các giải pháp:
  - Hiện trạng quản lý dữ liệu ngành dầu khí Việt Nam
  - Tình hình nghiên cứu về tích hợp mô hình dữ liệu trên thế giới và trong nước
- ▶ Phương pháp nghiên cứu của luận án
- ▶ Các kết quả nghiên cứu đã đạt được
  - Kết quả về mặt lý thuyết
  - Kết quả cài đặt và triển khai
  - Các công trình đã công bố

## Mục tiêu

- ▶ Đề xuất mô hình và giải pháp về tích hợp dữ liệu:
  - Đảm bảo tính khoa học, tính hiện đại và tính thực tiễn về tích hợp dữ liệu (tập trung vào tích hợp mô hình dữ liệu)
  - Phát hiện tri thức từ dữ liệu
- ▶ Đề xuất và thi hành các mô hình, các giải pháp tích hợp dữ liệu và phát hiện tri thức từ dữ liệu vào Trung tâm dữ liệu ngành dầu khí Việt Nam

## Phương pháp

- ▶ Phân tích và hệ thống hóa nội dung chủ đề tích hợp mô hình dữ liệu từ tài liệu.
- ▶ Đánh giá phương pháp tích hợp mô hình dữ liệu đã có, đề xuất phương pháp mới. Thực nghiệm đánh giá các thuật toán và giải pháp đã đề xuất.
- ▶ Khảo sát hiện trạng, đưa ra đánh giá chi tiết về khối lượng, chất lượng và công tác quản lý dữ liệu của ngành Dầu khí tại Việt Nam.
- ▶ Thực hiện công tác quốc tế theo tập thể các nhà khoa học hướng dẫn luận án.
- ▶ Công bố khoa học quốc tế và quốc gia xác nhận các kết quả nghiên cứu

## Kết quả

- a) Nghiên cứu và nắm được nội dung các kỹ thuật/phương pháp tích hợp dữ liệu**
- Các kỹ thuật/phương pháp tích hợp dữ liệu/tích hợp ontology.
  - Độ đo Google trong tích hợp dữ liệu.
  - Nghiên cứu, thử nghiệm một số kỹ thuật tích hợp ontology sử dụng học máy, áp dụng trong lĩnh vực y sinh; trong chuyên ngành dầu khí.
- b) Nghiên cứu và nắm được hiện trạng quản lý dữ liệu thăm dò khai thác dầu khí**
- Khảo sát được tình hình quản lý dữ liệu của ngành Dầu khí nói chung và dữ liệu thăm dò khai thác nói riêng.
  - Tìm hiểu và thu thập các sơ đồ dữ liệu về thăm dò khai thác trong một số phần mềm chuyên ngành dầu khí.
  - Nghiên cứu các chuẩn dữ liệu thăm dò khai thác dầu khí, các ontology về thăm dò khai thác, và mô hình quản lý dữ liệu dầu khí (PPDM), Từ điển dầu khí.
  - Xác định các mô hình dữ liệu từ các phần mềm quản lý dữ liệu thăm dò khai thác phổ biến hiện nay.
  - Xây dựng một Ontology thăm dò khai thác dầu khí và một số công cụ xây dựng ontology trên thế giới.
  - Tập trung nghiên cứu và thực hành xây dựng ontology về thăm dò khai thác dựa trên Từ điển dầu khí.
- c) Các công trình khoa học đã công bố:**
1. Van-Hien Tran, Thi-Huyen-Trang Doan, Ngoc-Trinh Vu, Hoang-Quynh Le, and Mai-Vu Tran (2014). *A Method for Building an Annotated Named Entity Recognition Corpus in Phenotype Domain*. MLAVN Workshop, ACML 2014, Nhatrang, November 2014.
  2. Ngoc-Trinh Vu, Van-Hien Tran, Thi-Huyen-Trang Doan, Hoang-Quynh Le, and Mai-Vu Tran (2015). *A Method for Building a Labeled Named Entity Recognition Corpus Using Ontologies*, Proceedings of ICCSAMA 2015, pp 141-149, ISBN 978-3-319-17995-7, 2015.
  3. Trinh Vu Ngoc, Ha Quang Thuy, Tran Trong Hieu. *Độ đo GOOGLE trong tích hợp dữ liệu*. Kỷ yếu hội nghị FAIR 2015 trang 224-231.
  4. Trong Hieu Tran, Thi Hong Khanh Nguyen, Quang Thuy Ha, and Ngoc Trinh Vu. *Argumentation framework for merging stratified belief bases*. Proceeding of ACIIDS 2016. pp. 43-53, 2016

## Kết luận

- ▶ Một số vấn đề nghiên cứu được trình bày.
- ▶ Phương pháp nghiên cứu được đề xuất
- ▶ Các kết quả nghiên cứu được trình bày
- ▶ Các công việc tương lai:
  - Xây dựng và hoàn chỉnh các ontology dầu khí
  - Viết phần mềm hỗ trợ chuyển đổi dữ liệu giữa các phần mềm dầu khí hiện đang sử dụng
  - Tích hợp dữ liệu trong các HTTT dầu khí

## Tài liệu tham khảo

- AnHai Doan, Alon Y. Halevy, Zachary G. Ives (2012). Principles of Data Integration. *Morgan Kaufmann*, ISBN 978-0-12-416044-6, pp. I-XVIII, 1-497.
- Bing Liu (2011). *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data*. Second Edition. Data-Centric Systems and Applications, Springer, ISBN 978-3-642-19459-7, pp. 1-603
- Pavel Shvaiko, Jérôme Euzenat (2013). *Ontology Matching: State of the Art and Future Challenges*. *IEEE Trans. Knowl. Data Eng.* 25(1): 158-176.
- Zarko Aleksovski (2008). *Using background knowledge in ontology matching*, *PhD Thesis*, Vrije Universiteit Amsterdam.
- Andrew R. Cohen, Paul M. B. Vitányi (2013). *Normalized Google Distance of Multisets with Applications*. *CoRR* abs/1308.3177.
- Philip A. Bernstein, Jayant Madhavan, Erhard Rahm (2011). *Generic Schema Matching, Ten Years Later*, *PVLDB* 4(11): 695-701.
- Tamir Bechor, Seev Neumann, Moshe Zviran, Chanan Glezer (2010). *A contingency model for estimating success of strategic information systems planning*. *Information & Management* 47 (2010):17–29.
- AnHai Doan (2002). *Learning to Map between Structured Representations of Data*, *PhD Thesis*, The University of Washington.
- Jayant Madhavan, Philip A. Bernstein, Pedro Domingos, Alon Y. Halevy (2002). *Representing and Reasoning about Mappings between Domain Models*, *AAAI/IAAI 2002*: 80-86.