

**BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI**

**CHƯƠNG TRÌNH ĐÀO TẠO
TIẾN SĨ**

**NGÀNH/CHUYÊN NGÀNH
HỆ THỐNG THÔNG TIN
MÃ SỐ: 62.48.05.01**

**Đã được Hội đồng Khoa học Viện Công nghệ thông tin và Truyền thông
thông qua ngày tháng năm 2015**

HÀ NỘI 2015

MỤC LỤC

Trang

PHẦN I TỔNG QUAN VỀ CHƯƠNG TRÌNH ĐÀO TẠO

- 1 Mục tiêu đào tạo
- 1.1 Mục tiêu chung
- 1.2 Mục tiêu cụ thể
- 2 Thời gian đào tạo
- 3 Khối lượng kiến thức
- 4 Đối tượng tuyển sinh
- 4.1 Định nghĩa
- 4.2 Phân loại đối tượng ngành phù hợp
- 4.3 Phân loại đối tượng ngành gần phù hợp
- 5 Quy trình đào tạo, điều kiện công nhận đạt
- 6 Thang điểm
- 7 Nội dung chương trình
- 7.1 Cấu trúc
- 7.2 Học phần bổ sung, chuyển đổi
- 7.2.1 Danh mục học phần bổ sung, chuyển đổi
- 7.2.2 Mô tả tóm tắt học phần bổ sung, chuyển đổi
- 7.2.3 Thời hạn hoàn thành các học phần bổ sung, chuyển đổi
- 7.3 Học phần trình độ Tiến sĩ
- 7.3.1 Danh mục học phần trình độ Tiến sĩ
- 7.3.2 Mô tả tóm tắt học phần trình độ Tiến sĩ
- 7.3.3 Kế hoạch học tập các học phần trình độ Tiến sĩ
- 7.4 Tiểu luận tổng quan
- 7.5 Chuyên đề Tiến sĩ
- 7.6 Nghiên cứu khoa học và luận án tiến sĩ
- 8 Danh sách Tạp chí / Hội nghị Khoa học

PHẦN II ĐỀ CƯƠNG CHI TIẾT CÁC HỌC PHẦN

- 9 Danh mục học phần chi tiết của chương trình đào tạo
- 9.1 Danh mục học phần bổ sung
- 9.2 Danh mục học phần trình độ Tiến sĩ
- 10 Đề cương chi tiết các học phần trình độ Tiến sĩ

PHẦN I

TỔNG QUAN VỀ CHƯƠNG TRÌNH ĐÀO TẠO

CHƯƠNG TRÌNH ĐÀO TẠO TIẾN SĨ
NGÀNH /CHUYÊN NGÀNH "HỆ THỐNG THÔNG TIN"

Tên chương trình: Chương trình đào tạo Tiến sĩ chuyên ngành "Hệ thống thông tin"
Trình độ đào tạo: Tiến sĩ
Chuyên ngành đào tạo: Hệ thống thông tin – Information Systems
Mã chuyên ngành: 62.48.05.01

(Ban hành theo Quyết định số / QĐ-ĐH BK-SĐH ngày tháng năm 2015
của Hiệu trưởng trường ĐH Bách Khoa Hà Nội)

1 Mục tiêu đào tạo

1.1 Mục tiêu chung

Đào tạo Tiến sĩ chuyên ngành "*Hệ thống thông tin*" có trình độ chuyên môn sâu cao, có khả năng nghiên cứu và lãnh đạo nhóm nghiên cứu các lĩnh vực của chuyên ngành, có tư duy khoa học, có khả năng tiếp cận và giải quyết các vấn đề khoa học chuyên ngành, có khả năng trình bày - giới thiệu các nội dung khoa học, đồng thời có khả năng đào tạo các bậc Đại học và Cao học.

1.2 Mục tiêu cụ thể

Sau khi đã kết thúc thành công chương trình đào tạo, Tiến sĩ chuyên ngành Hệ thống thông tin:

Có khả năng phát hiện và trực tiếp giải quyết các vấn đề khoa học thuộc các lĩnh vực Hệ thống thông tin nói riêng và Khoa học máy tính nói chung.

Có khả năng dẫn dắt, lãnh đạo nhóm nghiên cứu thuộc các lĩnh vực Hệ thống thông tin, Khoa học máy tính.

Có khả năng nghiên cứu, đề xuất và áp dụng các giải pháp công nghệ thuộc các lĩnh vực nói trên trong thực tiễn.

Có khả năng cao để trình bày, giới thiệu (bằng các hình thức bài viết, báo cáo hội nghị, giảng dạy đại học và sau đại học) các vấn đề khoa học thuộc hai lĩnh vực nói trên.

2 Thời gian đào tạo

Hệ tập trung liên tục: 3 năm liên tục đối với NCS có bằng ThS, 4 năm đối với NCS có bằng ĐH.

Hệ không tập trung liên tục: NCS có văn bằng ThS đăng ký thực hiện trong vòng 4 năm đảm bảo tổng thời gian học tập, nghiên cứu tại Trường là 3 năm và 12 tháng đầu tiên tập trung liên tục tại Trường.

3 Khối lượng kiến thức

Khối lượng kiến thức bao gồm khối lượng của *các học phần Tiến sĩ* và khối lượng của *các học phần bổ sung* được xác định cụ thể cho từng loại đối tượng tại mục 4.

NCS đã có bằng ThS: tối thiểu 8 tín chỉ học phần tiến sĩ + khối lượng bổ sung (nếu có).

NCS mới có bằng ĐH: tối thiểu 8 tín chỉ học phần tiến sĩ + số tín chỉ (không kể luận văn) của Chương trình Thạc sĩ Khoa học chuyên ngành "Hệ thống thông tin" (tương đương với 41 tín chỉ).

4 Đối tượng tuyển sinh

Đối tượng tuyển sinh là các thí sinh đã có bằng Thạc sĩ với chuyên ngành tốt nghiệp phù hợp hoặc gần phù hợp với chuyên ngành Hệ thống thông tin. Chỉ tuyển sinh mới có bằng ĐH với ngành tốt nghiệp phù hợp. Mức độ *phù hợp hoặc gần phù hợp* với chuyên ngành Hệ thống thông tin, được định nghĩa cụ thể ở mục 4.1 sau đây.

4.1 Định nghĩa

Ngành/chuyên ngành phù hợp (đúng): Là những hướng đào tạo chuyên sâu thuộc ngành "Công nghệ Thông tin", ngành "Hệ thống thông tin", "Khoa học máy tính", "Kỹ thuật phần mềm", "Kỹ thuật máy tính", "Truyền thông và Mạng máy tính" của chương trình đào tạo trình độ thạc sĩ trường ĐHBK HN, các chuyên ngành thuộc ngành Công nghệ thông tin của các trường đại học khác (như Khoa học máy tính, Hệ thống thông tin, Kỹ thuật phần mềm, Truyền thông và Mạng máy tính, Kỹ thuật máy tính).

Ngành/chuyên ngành gần phù hợp: Ngành "Điện tử truyền thông", "Điều khiển tự động", "Tin học công nghiệp", "Sư phạm kỹ thuật tin" của chương trình đào tạo đại học trường ĐHBK HN và các trường đại học khác

4.2 Phân loại đối tượng ngành

- Đối tượng A1: Thí sinh có bằng ThS Khoa học của ĐH Bách khoa Hà Nội, thạc sĩ khoa học các trường đại học ở nước ngoài có uy tín cấp, với ngành tốt nghiệp cao học đúng với ngành/chuyên ngành Tiến sĩ.

Đây là đối tượng không phải tham gia học bổ sung.

- Đối tượng A2: Thí sinh có bằng tốt nghiệp Đại học hệ chính quy đúng, phù hợp với ngành/chuyên ngành xếp loại "Xuất sắc" hoặc loại "Giỏi". Đối với bằng tốt nghiệp xếp loại "Giỏi" yêu cầu người dự tuyển là tác giả của ít nhất 01 bài báo đã đăng trong tạp chí/kỷ yếu hội nghị chuyên ngành có phản biện độc lập, được Hội đồng chức danh Giáo sư Nhà nước tính điểm, có trong danh mục Viện chuyên ngành quy định hoặc người dự tuyển đạt thành tích sinh viên nghiên cứu khoa học từ giải ba cấp Trường trở lên.

Đây là đối tượng phải tham gia học bổ sung toàn bộ chương trình thạc sĩ khoa học.

- Đối tượng A3: Thí sinh có bằng ThS kỹ thuật (thạc sĩ theo định hướng ứng dụng) đúng ngành hoặc có bằng ThS tốt nghiệp ngành gần phù hợp.

Đây là đối tượng phải tham gia học bổ sung.

5 Quy trình đào tạo, điều kiện công nhận đạt

Quy trình đào tạo được thực hiện theo học chế tín chỉ, tuân thủ Quyết định số 3341/QĐ-ĐHBK-SĐH ngày 21/8/2014 về tổ chức và quản lý đào tạo Sau đại học của Hiệu trưởng Trường ĐH Bách khoa Hà Nội.

Các học phần bổ sung, học phần tiến sĩ và chuyên đề tiến sĩ phải đạt mức điểm C trở lên (xem mục 6).

6 Thang điểm

Khoản 6a Điều 62 của Quy định 3341/2014 quy định:

Việc chấm điểm kiểm tra - đánh giá học phần (bao gồm các điểm kiểm tra và điểm thi kết thúc học phần) được thực hiện theo thang điểm từ 0 đến 10, làm tròn đến một chữ số thập phân sau dấu phẩy. Điểm học phần là điểm trung bình có trọng số của các điểm kiểm tra và điểm thi kết thúc (tổng của tất cả các điểm kiểm tra, điểm thi kết thúc đã nhân với trọng số tương ứng của từng điểm được quy định trong đề cương chi tiết học phần).

Điểm học phần được làm tròn đến một chữ số thập phân sau dấu phẩy, sau đó được chuyển thành điểm chữ với mức như sau:

Điểm số từ	8,5 – 10	chuyển thành điểm A (Giỏi)
Điểm số từ	7,0 – 8,4	chuyển thành điểm B (Khá)
Điểm số từ	5,5 – 6,9	chuyển thành điểm C (Trung bình)
Điểm số từ	4,0 – 5,4	chuyển thành điểm D (Trung bình yếu)
Điểm số dưới	4,0	chuyển thành điểm F (Kém)

7 Nội dung chương trình

7.1 Cấu trúc

Cấu trúc chương trình đào tạo trình độ Tiến sĩ gồm có 3 phần như bảng sau đây.

Phần	Nội dung đào tạo	A1	A2	A3
1	HP bổ sung	0	CT ThS KH	$16TC \geq \text{Bổ sung} \geq 4TC$
	HP TS	8TC		
2	TLTQ	2TC (Thực hiện và báo cáo trong năm học đầu tiên)		
	CĐTS	Tổng cộng 3 CĐTS, mỗi CĐTS 2TC		
3	NC khoa học và Luận án TS	90 TC (thực hiện trong 3 năm đối với hệ tập trung liên tục và 04 năm đối với hệ không tập trung liên tục)		

Lưu ý:

Số TC qui định cho các đối tượng trong là số TC tối thiểu NCS phải hoàn thành.

Đối tượng A2 phải thực hiện toàn bộ các học phần qui định trong chương trình ThS Khoa học của ngành tương ứng, không cần thực hiện luận văn ThS.

Các HP bổ sung được lựa chọn từ chương trình đào tạo Thạc sĩ của ngành đúng chuyên ngành Tiến sĩ.

Việc qui định số TC của HP bổ sung cho đối tượng A3 do Hội đồng khoa học Viện chuyên ngành và người hướng dẫn (NHD) quyết định dựa trên cơ sở đối chiếu các học phần trong bảng kết quả học tập ThS của thí sinh với chương trình ThS hiện tại của ngành đúng chuyên ngành Tiến sĩ nhưng phải đảm bảo số TC tối thiểu và tối đa trong bảng.

Các HP TS được NHD đề xuất từ chương trình đào tạo Thạc sĩ và Tiến sĩ của trường nhằm trang bị kiến cần thiết phục vụ cho đề tài nghiên cứu cụ thể của LATS.

7.2 Học phần bổ sung

7.2.1. Đối với NCS chưa có bằng thạc sĩ (Đối tượng A2)

NCS phải hoàn thành các học phần bổ sung trong thời hạn 2 năm kể từ ngày ký quyết định công nhận là NCS gồm các học phần ở trình độ thạc sĩ ngành "*Hệ thống thông tin*" theo chương trình cụ thể như sau:

NỘI DUNG	MÃ SỐ	TÊN HỌC PHẦN	TÍN CHỈ	KHỐI LƯỢNG
Kiến thức cơ sở bắt buộc (15 TC)	IT5310	Tính toán phân tán	3	(3-1-0-6)
	IT5320	Tích hợp dữ liệu và XML	3	(3-1-0-6)
	IT5330	Hệ cơ sở dữ liệu đa phương tiện	3	(3-1-0-6)
	IT5340	Tìm kiếm và trình diễn thông tin	3	(3-1-0-6)
	IT5350	Xử lý thông tin mờ	3	(3-1-0-6)
Kiến thức cơ sở tự chọn (6 TC)	IT5360	Xử lý ngôn ngữ tự nhiên	3	(3-1-0-6)
	IT5370	Quản trị dự án	2	(2-1-0-4)
	IT5530	Hệ thống hướng tác tử	2	(2-1-0-4)
	IT5390	Học máy	2	(2-1-0-4)
Chuyên ngành bắt buộc (8 TC)	IT6300	Các hệ cơ sở dữ liệu tiên tiến Advanced database systems	3	3(2,5-1-0-6)
	IT6320	Khai phá dữ liệu Data mining	3	3(2,5-1-0-6)
	IT6310	Tìm kiếm dữ liệu đa phương tiện	2	2(1.5-1-0-4)
Chuyên ngành tự chọn (7TC)	IT6360	Kỹ nghệ tri thức Knowledge Engineering	2	2(1.5-1-0-4)
	IT6365	Công nghệ Web ngữ nghĩa Semantic Web Technologies	3	3(2,5-1-0-6)
	IT6370	Xử lý ngôn ngữ tự nhiên nâng cao Advanced Natural Language Processing	2	2(1.5-1-0-4)
	IT6380	Logic mờ và ứng dụng Fuzzy Logic and Applications	2	2(1.5-1-0-4)
	IT6385	Lưu trữ và phân tích dữ liệu lớn Big data storage and analytics	3	3(2,5-1-0-6)

7.2.2. Đối với NCS có bằng thạc sĩ ngành gần (Đối tượng A3)

Đối với NCS có bằng thạc sĩ ngành gần với ngành/chuyên ngành "*Hệ thống thông tin*" đề nghị học các học phần bổ sung như sau:

NỘI DUNG	MÃ SỐ	TÊN HỌC PHẦN	TÍN CHỈ	KHỐI LƯỢNG
Chuyên ngành bắt buộc	IT6320	Khai phá dữ liệu Data mining	3	3(2,5-1-0-6)
	IT6310	Tìm kiếm dữ liệu đa phương tiện	2	2(1.5-1-0-4)
Chuyên ngành tự chọn	IT6360	Kỹ nghệ tri thức Knowledge Engineering	2	2(1.5-1-0-4)
	IT6365	Công nghệ Web ngữ nghĩa Semantic Web Technologies	3	3(2,5-1-0-6)

	IT6370	Xử lý ngôn ngữ tự nhiên nâng cao Advanced Natural Language Processing	2	2(1.5-1-0-4)
	IT6380	Logic mờ và ứng dụng Fuzzy Logic and Applications	2	2(1.5-1-0-4)
	IT6385	Lưu trữ và phân tích dữ liệu lớn Big data storage and analytics	3	3(2,5-1-0-6)

7.3 Học phần Tiến sĩ

Các HP TS nhằm giúp NCS cập nhật các kiến thức mới nhất của lĩnh vực chuyên môn, nâng cao trình độ lý thuyết, phương pháp luận NC và khả năng ứng dụng các phương pháp NC khoa học quan trọng, thiết yếu của lĩnh vực NC. Mỗi HP TS được thiết kế với khối lượng từ 2 đến 3 TC. Mỗi NCS phải hoàn thành tối thiểu 8 TC tương ứng với 3 HP trở lên.

7.3.1 Danh mục học phần trình độ Tiến sĩ

TT	MÃ SỐ	TÊN HỌC PHẦN	GIẢNG VIÊN	TÍN CHỈ	KHỐI LƯỢNG
1	IT7310	Kỹ nghệ dữ liệu và tri thức Knowledge and Data Engineering	1. PGS. Nguyễn Thị Kim Anh 2. TS. Thân Quang Khoát 3. TS. Nguyễn Nhật Quang	3	3(2-2-0-6)
2	IT7331	Tích hợp dữ liệu Data Integration	1. PGS. Nguyễn Thị Kim Anh 2. TS. Vũ Tuyết Trinh 3. TS. Nguyễn Bá Ngọc	3	3(2-2-0-6)
3	IT7341	Tìm kiếm thông tin Information Retrieval	1. TS. Vũ Tuyết Trinh 2. TS. Nguyễn Bá Ngọc 3. TS. Nguyễn Thị Oanh	3	3(2-2-0-6)
4	IT7351	Khai phá dữ liệu và phát hiện tri thức Data Mining and Knowledge Discovery	1. TS. Nguyễn Nhật Quang 2. PGS. Nguyễn Thị Kim Anh 3. TS. Thân Quang Khoát	3	3(2-2-0-6)

5	IT7361	Khai phá dữ liệu văn bản và dữ liệu Web Text and Web Mining	1. PGS. Lê Thanh Hương 2. PGS. Nguyễn Thị Kim Anh 3. TS. Nguyễn Bá Ngọc	3	3(2-2-0-6)
6	IT7371	Các tiếp cận logic trong biểu diễn và xử lý thông tin Logical approaches in information representation and processing	1. PGS. Trần Đình Khang 2. PGS. Lê Thanh Hương	3	3(2-2-0-6)
7	IT7381	Các chủ đề nâng cao trong xử lý dữ liệu lớn Advanced Topics in Big Data Processing	1. TS. Nguyễn Bình Minh 2. TS. Nguyễn Hữu Đức	3	3(2-2-0-6)
8	IT7391	Các chủ đề nâng cao đối với phân tích dữ liệu lớn Advanced Topics for Big Data Analytics	1. TS. Thân Quang Khoát 2. PGS. Nguyễn Thị Kim Anh	3	3(2-2-0-6)

* Nghiên cứu sinh có thể chọn một học phần tự chọn liên quan đến lĩnh vực "Hệ thống thông tin" trong các học phần do Viện Công nghệ thông tin và truyền thông phụ trách, phù hợp với yêu cầu của đề tài nghi

7.3.2 Mô tả tóm tắt học phần trình độ Tiến sĩ

IT7310 Kỹ nghệ dữ liệu và tri thức

1. Cung cấp các kỹ nghệ đối với vấn đề thu thập và quản trị các loại dữ liệu nói chung: dữ liệu có cấu trúc, bán cấu trúc và phi cấu trúc.
2. Cung cấp các kỹ nghệ đối với vấn đề thu thập và quản trị các loại tri thức nói chung: tri thức ẩn, tri thức hiện, tri thức rõ, tri thức mờ,...

IT7310 Knowledge and Data Engineering

1. Provide techniques for data collection and data management: structured data, semi-structured data and unstructured data.
2. Provide techniques for knowledge collection and knowledge management: hidden knowledge, present knowledge, clear knowledge, fuzzy knowledge, ...

IT7331 Tích hợp dữ liệu

1. Cung cấp các kỹ thuật đối với vấn đề thu thập dữ liệu từ các nguồn dữ liệu thuần nhất và không thuần nhất
2. Cung cấp các kỹ thuật đối với vấn đề làm sạch dữ liệu để đảm bảo tính nhất quán của dữ liệu
3. Cung cấp các kỹ thuật đối với vấn đề sinh các câu trả lời nhất quán

IT7331 Data Integration

1. Provide techniques for data collection from homogeneous and heterogeneous data sources
2. Provide techniques for data cleaning to ensure data consistency
3. Provide techniques for generating consistent answers

IT7341 Tìm kiếm thông tin

1. Cung cấp các kỹ thuật đối với vấn đề thu thập, trích chọn, biểu diễn, lưu trữ các tài liệu đối với một hệ tìm kiếm thông tin
2. Cung cấp các kỹ thuật đối với vấn đề xử lý các yêu cầu tìm kiếm và trình diễn các kết quả tìm kiếm

IT7341 Information Retrieval

1. Provide techniques for collecting, extracting, representing, storing data for an search engine.
2. Provide techniques for processing search queries and presenting search result.

IT7351 Khai phá dữ liệu và phát hiện tri thức (Data Mining and Knowledge Discovery)

1. Cung cấp các kỹ thuật đối với vấn đề chuẩn bị dữ liệu cho khai phá dữ liệu
2. Cung cấp các kỹ thuật khai phá dữ liệu nói chung: phát hiện luật kết hợp, phân cụm dữ liệu, phân loại dữ liệu và tóm tắt dữ liệu

IT7351 Data Mining and Knowledge Discovery

1. Provide techniques for data preparation in data mining
2. Provide techniques for data mining: discovering association rules, data clustering, data categorization and data summarization

IT7361 Khai phá dữ liệu văn bản và dữ liệu Web (Text and Web Mining)

Học phần này cung cấp các kỹ thuật và các công cụ sử dụng trong khai phá dữ liệu văn bản và dữ liệu web. Trên cơ sở đó, NCS có thể phát triển các ứng dụng như máy tìm kiếm, thương mại điện tử, các thư viện số, các hệ thống quản lý tri thức, v.v.

1. Cung cấp các kỹ thuật đối với vấn đề tiền xử lý và biểu diễn văn bản cho khai phá văn bản
2. Cung cấp các kỹ thuật khai phá văn bản nói chung: phân cụm văn bản, phân loại văn bản và tóm tắt văn bản
3. Cung cấp các kỹ thuật khai phá Web

IT7361 Text and Web Mining

This course provides techniques and tools used in text mining and web mining. From that point of view, PhD student can develop applications such as search engines, e-commerce, digital libraries, systems for knowledge management, etc.

1. Provide techniques for preprocessing and representing documents for text mining.
2. Provide text mining techniques: text clustering, text categorization and summary documents
3. Provide Web mining techniques

IT7371 Các tiếp cận logic trong biểu diễn và xử lý thông tin

Trình bày tổng quan về logic tính toán, với các logic kinh điển và logic bậc cao; phân loại các ngôn ngữ khái niệm theo cú pháp và cấu trúc cùng với các thủ tục tính toán. Đưa ra các biểu diễn ngữ nghĩa cho các nhãn ngôn ngữ và các phép toán xử lý.

IT7371 Logical approaches in information representation and processing

Presenting an overviewing of computational logic, including the classic logic and the higher-order logic; classifying conceptual languages by their syntax and their structure along with computational procedures. Providing semantic representations for language labels and processing operators.

IT7381 Các chủ đề nâng cao trong xử lý dữ liệu lớn

Học phần này sẽ trình bày các kỹ thuật và hệ thống cho việc xử lý dữ liệu lớn theo hướng nâng cao. Các chủ đề liên quan tới nền tảng tính toán cho dữ liệu lớn sẽ bao gồm lưu trữ dữ liệu lớn, các hệ thống cho phép xử lý dữ liệu lớn. Các chủ đề liên quan tới xử lý dữ liệu lớn bao gồm truy hồi thông tin (information retrieval) với MapReduce, xử lý đồ thị (graph processing) với MapReduce, quản lý dữ liệu với MapReduce, khai phá dữ liệu với MapReduce và một số mô hình toán học thiết kế thuật toán trên dữ liệu lớn. Mục tiêu của học phần này là giúp học viên có được các kiến thức nâng cao cùng với những phân tích sâu về các chủ đề nói trên.

IT7381 Advanced Topics in Big Data Processing

This course surveys techniques and systems for efficiently processing massive datasets with the focus on advanced knowledge. Topics related to computing platforms for big data include: big data storage, big data processing systems. Topics related to large dataset processing include: information retrieval with MapReduce, graph processing with MapReduce, data management with MapReduce, data mining with MapReduce and several mathematical models for designing algorithms for big data. The goal of the course is to gain advanced knowledge along with in-depth discussions of the topics covered.

IT7391 Các chủ đề nâng cao đối với phân tích dữ liệu lớn

Học phần này sẽ trình bày các phương pháp và kỹ thuật đối với phân tích dữ liệu lớn theo hướng nâng cao. Các chủ đề liên quan tới phân tích dữ liệu lớn sẽ bao gồm: phân loại và hồi qui, các mô hình đồ thị xác suất, giảm chiều, mô hình thưa.... đối với dữ liệu lớn. Mục tiêu của học phần này là giúp học viên có được các kiến thức nâng cao cùng với những phân tích sâu về các chủ đề nói trên.

IT7391 Advanced Topics for Big Data Analytics

This course surveys methods and techniques for massive datasets analytics with the focus on advanced knowledge. Topics related to big data analytics include: Large-scale classification and regression, Probabilistic graphical models, Nonlinear dimensionality reduction, Sparse modeling..... The goal of the course is to gain advanced knowledge along with in-depth discussions of the topics covered.

7.3.3 Kế hoạch học tập các học phần trình độ Tiến sĩ

Nghiên cứu sinh phải hoàn thành các học phần Tiến sĩ trong vòng 24 tháng kể từ ngày Ký quyết định công nhận NCS và theo kế hoạch năm học. HP TS được coi là đạt nếu điểm kết thúc đạt từ C trở lên

Các HP TS được thực hiện theo các bước sau:

Bước 1: Khi NCS nhập học, NCS phải đăng ký học các HP TS và nộp cho Viện ĐT Sau đại học.

Bước 2: Viện Công nghệ thông tin và truyền thông lên kế hoạch tổ chức lớp và thông báo cho giáo viên phụ trách học phần và giao cho giáo viên phụ trách HP trong tuần thứ 5 của học kỳ.

Bước 3: NCS thực hiện các HP TS theo đúng qui định và yêu cầu của môn học.

Bước 4: Giáo viên giảng dạy có trách nhiệm nộp cho Viện Công nghệ thông tin và truyền thông kết quả học phần chậm nhất 2 tuần sau khi kết thúc học kỳ để Viện chuyên ngành nộp kết quả cho Viện Đào tạo Sau đại học.

7.4. Tiểu luận tổng quan

Bài TLTQ về tình hình NC và các vấn đề liên quan đến đề tài luận án: thể hiện kết quả NC phân tích, đánh giá các công trình NC đã có của các tác giả trong và ngoài nước liên quan mật thiết đến đề tài luận án, nêu những vấn đề còn tồn tại, chỉ ra những vấn đề mà luận án cần tập trung NC giải quyết. NCS thực hiện bài TLTQ dưới sự hướng dẫn của NHD luận án.

Tiểu luận tổng quan được đánh giá kết thúc thông qua hình thức báo cáo trước đơn vị chuyên môn (báo cáo trình bày trong khoảng 15 phút), tranh luận và trả lời câu hỏi, sau đó đơn vị chuyên môn sẽ đánh giá bài TLTQ đạt yêu cầu hay chưa đạt yêu cầu, có ghi biên bản buổi báo cáo.

NCS phải hoàn thành bài TLTQ với kết quả đạt yêu cầu trong vòng 12 tháng kể từ ngày được triệu tập trúng tuyển. Tiểu luận tổng quan tương đương với 2 tín chỉ.

7.5 Chuyên đề Tiến sĩ

Các CĐTS đòi hỏi NCS tự cập nhật kiến thức mới liên quan trực tiếp đến đề tài của NCS, nâng cao năng lực NC khoa học, giúp NCS giải quyết trực tiếp một số nội dung của đề tài luận án. Mỗi nghiên cứu sinh phải hoàn thành 3 chuyên đề Tiến sĩ, có thể tùy chọn từ danh sách hướng chuyên sâu. Mỗi hướng chuyên sâu đều có người hướng dẫn do Hội đồng Xây dựng chương trình đào tạo chuyên ngành của Viện quyết định.

Người hướng dẫn khoa học luận án của nghiên cứu sinh sẽ đề xuất đề tài cụ thể. Ưu tiên đề xuất đề tài gắn liền, thiết thực với đề tài của luận án Tiến sĩ.

Sau khi đã có đề tài cụ thể, NCS thực hiện đề tài đó dưới sự hướng dẫn khoa học của người hướng dẫn chuyên đề.

Danh mục hướng chuyên sâu cho Chuyên đề Tiến sĩ

TT	MÃ SỐ	HƯỚNG CHUYÊN SÂU	NGƯỜI HƯỚNG DẪN	TÍN CHỈ
1	IT7410	Các hệ thống thông minh Intelligent Systems	PGS. Trần Đình Khang PGS. Nguyễn Thị Kim Anh	2
2	IT7421	Các hệ CSDL thông minh Intelligent Database Systems	PGS. Nguyễn Thị Kim Anh PGS. Lê Thanh Hương	2
3	IT7431	Tìm kiếm dựa trên ngữ nghĩa Intelligent Information Retrieval	TS. Lê Thanh Hương TS. Nguyễn Bá Ngọc	2
4	IT7441	Hệ dựa trên logic Logic-based Systems	PGS. Trần Đình Khang	2
5	IT7451	Hệ gợi ý Recommender Systems	TS. Nguyễn Nhật Quang PGS. Nguyễn Thị Kim Anh	2
6	IT7461	Hệ thích nghi dựa trên ngữ cảnh Context-based adaptive Systems	TS. Vũ Tuyết Trinh TS. Nguyễn Nhật Quang	2
7	IT7471	Ontology và Web ngữ nghĩa Ontology and semantic Web	TS. Nguyễn Bá Ngọc TS. Lê Thanh Hương	2
8	IT7481	Tính toán hiệu năng cao High performance Computing	TS. Nguyễn Hữu Đức TS. Nguyễn Bình Minh	2

9	IT7491	Các phương pháp hiện đại đối với phân tích dữ liệu Modern Methods for Data Analysis	TS. Thân Quang Khoát TS. Nguyễn Nhật Quang PGS. Nguyễn Thị Kim Anh	2
10	IT7495	Biểu diễn và tìm kiếm thông tin đa phương tiện Multimedia Information Representation and Retrieval	TS. Nguyễn Thị Oanh TS. Vũ Tuyết Trinh	2

CĐTS được coi là đạt nếu kết quả trung bình của các thành viên hội đồng đạt từ C trở lên.

7.6. Nghiên cứu khoa học và luận án tiến sĩ

NC khoa học là giai đoạn chính, mang tính bắt buộc trong quá trình NCS thực hiện LATS. Đây là giai đoạn mà NCS có thể đạt tới tri thức mới hoặc giải pháp mới, hình thành các cơ sở quan trọng nhất để viết nên LATS. Trên cơ sở tính chất của lĩnh vực NC thuộc khoa học tự nhiên hay khoa học kỹ thuật – công nghệ, các Viện chuyên ngành, các BM và NHD có các yêu cầu cụ thể đối với việc NC khoa học của NCS:

Đánh giá hiện trạng tri thức, hiện trạng giải pháp công nghệ liên quan đến đề tài luận án.

Yêu cầu điều tra, thực nghiệm để bổ sung các dữ liệu cần thiết.

Yêu cầu suy luận khoa học hoặc thiết kế giải pháp, gắn liền với thí nghiệm.

Phân tích, đánh giá các kết quả thu được từ quá trình suy luận khoa học hay thí nghiệm.

NCS phải chủ động thực hiện nhiệm vụ NCKH và kết quả nghiên cứu phải được công bố chính thức thành các bài báo khoa học theo đúng quy định của Quy chế đào tạo tiến sĩ. Các đề tài NCKH và bài viết công bố phải phù hợp với mục tiêu của luận án, đảm bảo tính trung thực, tính khoa học và tính mới. Nội dung các bài báo không được trùng lặp và phản ánh các nội dung chính của luận án. Các bài báo, phát minh, sáng chế là kết quả nghiên cứu, nghiên cứu sinh phải đứng tên của Trường Đại học Bách khoa Hà Nội.

Luận án tiến sĩ phải là một công trình NC khoa học sáng tạo của chính NCS, có đóng góp về mặt lý luận và thực tiễn trong lĩnh vực nghiên cứu hoặc giải pháp mới có giá trị trong việc phát triển, gia tăng tri thức khoa học của lĩnh vực nghiên cứu, giải quyết sáng tạo các vấn đề của ngành khoa học hay thực tiễn kinh tế - xã hội. Luận án tiến sĩ thực hiện đúng quy cách và đảm bảo các yêu cầu cơ bản theo quy định của Quy chế đào tạo tiến sĩ.

NCS chịu trách nhiệm về tính trung thực, chính xác, tính mới của kết quả nghiên cứu của luận án, chấp hành các quy định về sở hữu trí tuệ của Việt Nam và quốc tế.

8 Danh sách Tạp chí / Hội nghị khoa học

Các diễn đàn khoa học trong nước trong bảng dưới đây là nơi NCS có thể chọn công bố các kết quả nghiên cứu khoa học phục vụ hoàn thành luận án Tiến sĩ.

Số TT	Tên diễn đàn	Địa chỉ liên hệ	Định kỳ xuất bản / họp
1	Tạp chí Khoa học và Công nghệ	ĐH Bách Khoa Hà Nội; Số 1, phố Đại Cồ Việt, Hai Bà Trưng, Hà Nội	Hai tháng 1 lần
2	Tạp chí Khoa học và Công nghệ trường Đại học quốc gia Hà nội	Trường Đại học Quốc Gia Hà Nội	Hai tháng 1 lần
3	Tạp chí Khoa học và Công nghệ trường Đại học quốc gia thành phố Hồ Chí Minh	Trường Đại học Quốc Gia Thành phố Hồ Chí Minh	Hai tháng 1 lần

4	Tạp chí Báo chính viễn thông, Chuyên san Các nghiên cứu và triển khai ứng dụng trong viễn thông và Công nghệ thông tin	Viện Khoa học và Công nghệ Việt Nam; đường Hoàng Quốc Việt, Hà Nội	Ba tháng 1 lần
5	Tạp chí Công nghệ thông tin 3 tháng 1 lần	Viện Khoa học và Công nghệ Việt Nam; số 18 đường Hoàng Quốc Việt, Hà Nội	Ba tháng 1 lần
6	Hội thảo khoa học quốc gia về nghiên cứu, phát triển và ứng dụng CNTT và Truyền thông ICT RDA	Ban chương trình quốc gia	Hàng năm
7	Hội nghị khoa học “Một số vấn đề chọn lọc trong CNTT”	Ban chương trình quốc gia	Hàng năm
8	Hội thảo Khoa học Quốc gia lần thứ nhất – "Nghiên cứu cơ bản và ứng dụng Công nghệ thông tin" FAIR	Ban chương trình quốc gia	Hàng năm
9	Các Hội nghị quốc tế tổ chức tại Việt Nam	Ban chương trình quốc gia	Hàng năm

PHẦN II

ĐỀ CƯƠNG CHI TIẾT CÁC HỌC PHẦN

9 Danh mục học phần chi tiết của chương trình đào tạo

9.1 Danh mục học phần bổ sung

Danh mục học phần bổ sung tương ứng với một đối tượng NCS cụ thể có thể xem chi tiết trong mục 7.2.

Đề cương chi tiết các học phần bổ sung

IT5310 Tính toán phân tán

1. Tên học phần: TÍNH TOÁN PHÂN TÁN

2. Mã số: IT5310

3. Khối lượng: 3(3-1-0-6)

- Lý thuyết: 45 tiết
- Bài tập: 15 tiết

4. Yêu cầu học phần: bắt buộc, học vào kỳ 1

5. Đối tượng tham dự: Học viên cao học chuyên ngành Hệ thống thông tin

6. Điều kiện học phần:

7. Mục tiêu học phần: Cung cấp cho học viên phương pháp luận thực hiện phân tích, thiết kế các hệ thống tính toán phân tán.

Sau khi học xong học phần này, yêu cầu sinh viên có khả năng:

- Các mô hình tính toán phân tán
- Mô hình lập trình tính toán phân tán
- Lựa chọn giải pháp công nghệ để phát triển các hệ thống tính toán phân tán.

8. Nội dung tóm tắt học phần: Nghiên cứu mô hình mô hình tính toán phân tán, phát triển chương trình phân tán theo mô hình truyền thông điệp, phát triển chương trình phân tán theo mô hình bộ nhớ chia, hệ thống tính toán lưới, hệ thống tính toán tình nguyện, tính toán phân tán dựa trên mô hình MapReduce.

9. Nhiệm vụ của sinh viên:

- Dự lớp: Theo quy định của Bộ GD&ĐT và trường ĐHBKHN.
- Bài tập: Theo quy định của Bộ GD&ĐT và trường ĐHBKHN.
- Thí nghiệm:.

10. Đánh giá kết quả: KT/BT(0.30)-T(BTL:0.70)

- Điểm quá trình: trọng số 0.30
 - Bài tập làm đầy đủ
 - Hoàn thành bài tập lớn
 - Kiểm tra giữa kỳ
- Thi cuối kỳ (bảo vệ bài tập lớn): trọng số 0.70

11. Tài liệu học tập:

- Sách giáo khoa chính:
- Sách tham khảo: Xem phân tài liệu tham khảo.

12. Nội dung chi tiết học phần:

TÍNH TOÁN PHÂN TÁN

Người biên soạn: TS. Nguyễn Hữu Đức

MỞ ĐẦU

1. Mục đích môn học
2. Nội dung môn học
3. Sách giáo khoa và tài liệu tham khảo

Chương 1: Tổng quan

- 1.1 Định nghĩa và các thành phần
- 1.2 Tại sao cần đến hệ thống phân tán
- 1.3 Quan hệ giữa hệ thống phân tán và hệ thống song song
- 1.4 Hệ thống truyền thông điệp và hệ thống chia sẻ bộ nhớ
- 1.5 Cơ bản về truyền thông phân tán
- 1.6 Các vấn đề về thiết kế hệ thống phân tán

Chương 2: Mô hình tính toán phân tán

- 2.1 Khái niệm chương trình phân tán
- 2.2 Mô hình thực thi phân tán
- 2.3 Các mô hình mạng truyền thông
- 2.4 Mô hình truyền thông tiến trình

Chương 3: Tính toán theo mô hình truyền thông điệp

- 3.1 Tổng quan mô hình truyền thông điệp
- 3.2 Truyền thông điểm điểm
- 3.3 Truyền thông công cộng
- 3.4 Một số ví dụ

Chương 4: Tính toán theo mô hình bộ nhớ chia sẻ

- 4.1 Thread
- 4.2 Mutex
- 4.3 Locks
- 4.4 OpenMP
- 4.5 Một số ví dụ

Chương 5 : Tính toán lưới

- 5.1 Tổng quan tính toán lưới
- 5.2 Kiến trúc dịch vụ lưới mở
- 5.3 Tổng quan về gobus toolkit
- 5.4 Lưới dữ liệu
- 5.5 Quản lý thực thi trên lưới
- 5.6 Bảo mật trong lưới và tổ chức ảo

Chương 6: Tính toán tình nguyện

- 6.1 Giới thiệu về tính toán tình nguyện
- 6.2 Giới thiệu về BOINC

Chương 7: MapReduce

7.1 Định nghĩa

7.2 Mô hình tính toán với MapReduce

13. Tài liệu tham khảo

1. Ajay D.Kshemkayani; Mukesh Singha. “Distributed Computing, Principles, Algorithms, and Systems”. Cambridge University Press, 2008.
2. Ian Foster, Carl Kesselman. “The Grid: Blueprint for a New Computing Infrastructure”, 1st edition, Morgan Kaufmann Publishers, San Francisco, USA (1 November 1998), ISBN: 1558604758
3. Cameron Hughes, Tracey Hughes. “Parallel and Distributed Programming Using C++”. Addison Wesley, August 25, 2003. ISBN 0-13-101376-9.

IT5320 Tích hợp dữ liệu và XML

1. Tên học phần: TÍCH HỢP DỮ LIỆU VÀ XML

2. Mã số: IT5320

3. Khối lượng: 3(3-1-0-6)

- Lý thuyết: 45 tiết
- Bài tập: 15 tiết

4. Yêu cầu học phần: bắt buộc, học vào kỳ 2

5. Đối tượng tham dự: Học viên cao học chuyên ngành Hệ thống thông tin

6. Điều kiện học phần:

7. Mục tiêu học phần: Học phần này nhằm mục đích giới thiệu các vấn đề chung về tích hợp dữ liệu và vai trò của XML trong các hệ tích hợp dữ liệu.

8. Nội dung tóm tắt học phần: Giới thiệu XML, DTD, XSD, XPath, XQuery, XSL, XSLT, XLink, XPointer, các cách tiếp cận để xây dựng một hệ tích hợp dữ liệu: các kỹ thuật đối sánh lược đồ, đối sánh dữ liệu, xây dựng lược đồ tích hợp, xử lý yêu cầu truy vấn, XML và vai trò của nó trong các hệ tích hợp dữ liệu, một số hệ thống tích hợp dữ liệu.

9. Nhiệm vụ của sinh viên:

- Dự lớp: Theo quy định của Bộ GD&ĐT và trường ĐHBKHN.
- Bài tập: Theo quy định của Bộ GD&ĐT và trường ĐHBKHN.
- Thí nghiệm:.

10. Đánh giá kết quả: KT/BT(0.30)-T(BTL:0.70)

- Điểm quá trình: trọng số 0.30
 - Bài tập làm đầy đủ
 - Hoàn thành bài tập lớn
 - Kiểm tra giữa kỳ
- Thi cuối kỳ (bảo vệ bài tập lớn): trọng số 0.70

11. Tài liệu học tập:

- Sách giáo khoa chính:
- Sách tham khảo: Xem phần tài liệu tham khảo.

12. Nội dung chi tiết học phần:

TÍCH HỢP DỮ LIỆU VÀ XML

Người biên soạn: TS. Vũ Tuyết Trinh

MỞ ĐẦU

1. Mục đích môn học
2. Nội dung môn học
3. Sách giáo khoa và tài liệu tham khảo

Chương 1. Giới thiệu chung

1. Vài nét về lịch sử
2. Các vấn đề liên quan đến XML
3. Các vấn đề liên quan đến tích hợp dữ liệu

Chương 2. Căn bản về XML

1. Các khái niệm cơ bản
2. Mô hình hóa dữ liệu XML
3. Mô hình dữ liệu của W3C
4. Định nghĩa kiểu tài liệu DTD
5. XSL, XSLT, XLink, XPointer, XSD

Chương 3. Truy vấn dữ liệu với XML

1. Ngôn ngữ truy vấn dữ liệu bán cấu trúc
2. XPath
3. XQuery
4. Lưu trữ và đánh chỉ số dữ liệu XML

Chương 4. Lược đồ và các ngôn ngữ biểu diễn lược đồ

1. Khái niệm
2. SQL
3. DTD
4. XSD
5. DSD
6. XDuce
7. RELAX Core
8. TREX
9. OWL

Chương 5. Đối sánh lược đồ

1. Đặt vấn đề
2. Các ứng dụng của đối sánh lược đồ
3. Phân loại các cách tiếp cận
4. Cách tiếp cận bộ đối sánh riêng rẽ
5. Cách tiếp cận kết hợp các bộ đối sánh
6. Chiến lược đối sánh
7. Xây dựng bộ đối sánh
8. Các tiêu chuẩn đánh giá

Chương 6. Tích hợp lược đồ và dữ liệu

1. Vai trò của tích hợp lược đồ và dữ liệu
2. Tích hợp lược đồ
3. Tích hợp dữ liệu
4. Xử lý truy vấn trong hệ tích hợp dữ liệu
5. Sự không chắc chắn trong tích hợp dữ liệu

Chương 7. Hiện trạng và hướng phát triển

1. Một số hệ thống thử nghiệm
2. Một vài sản phẩm thương mại

3. Các vấn đề mở và hướng phát triển

13. Tài liệu tham khảo

[1] Do Hong Hai, "Schema Matching and Mapping-based Data Integration", PhD Thesis, University of Leipzig, 2005.

[2] Zohra Bellahsene, Angela Bonifati, Erhard Rahm, "Schema Matching and Mapping", Springer-Verlag Berlin Heidelberg, 2011. ISBN: 978-3-642-16517-7.

[3] Avigdor Gal, "Uncertain Schema Matching", Morgan & Claypool Publishers, 2011. ISBN: 978-1-608-45433-4.

IT5330 Hệ cơ sở dữ liệu đa phương tiện

1. Tên học phần: **HỆ CƠ SỞ DỮ LIỆU ĐA PHƯƠNG TIỆN**

2. Mã số: **IT5330**

3. Khối lượng: **3(3-1-0-6)**

- Lý thuyết: 45 tiết
- Bài tập: 15 tiết

4. Yêu cầu học phần: bắt buộc, học vào kỳ 1

5. Đối tượng tham dự: Học viên cao học chuyên ngành Hệ thống thông tin

6. Điều kiện học phần:

7. Mục tiêu học phần: Giới thiệu các kỹ thuật cơ bản và các thuật toán hỗ trợ để xây dựng các hệ cơ sở dữ liệu đa phương tiện.

Sau khi hoàn thành học phần này, yêu cầu sinh viên:

- Hiểu các kiến thức cơ bản khi xây dựng hệ CSDL ĐPT
- Có khả năng phân tích và đề xuất, phát triển giải pháp kỹ thuật cho các vấn đề liên quan
- Rèn luyện kỹ năng thảo luận và làm việc theo nhóm

8. Nội dung tóm tắt học phần: Học phần này giới thiệu với sinh viên các kỹ thuật về tổ chức lưu trữ, đánh chỉ mục và tìm kiếm dữ liệu đa phương tiện (text, audio, video, ...), các cách biểu diễn, mô hình hoá và xử lý truy vấn dữ liệu đa phương tiện, các cách phân phối và trình diễn dữ liệu đa phương tiện ...

9. Nhiệm vụ của sinh viên:

- Dự lớp: Theo quy định của Bộ GD&ĐT và trường ĐHBKHN.
- Bài tập: Theo quy định của Bộ GD&ĐT và trường ĐHBKHN.
- Thí nghiệm:.

10. Đánh giá kết quả: **KT/BT(0.30)-T(BTL:0.70)**

- Điểm quá trình: trọng số 0.30
 - Bài tập làm đầy đủ
 - Hoàn thành bài tập lớn
 - Kiểm tra giữa kỳ
- Thi cuối kỳ (bảo vệ bài tập lớn): trọng số 0.70

11. Tài liệu học tập:

- Sách giáo khoa chính:
- Sách tham khảo: Xem phần tài liệu tham khảo.

12. Nội dung chi tiết học phần:

HỆ CƠ SỞ DỮ LIỆU ĐA PHƯƠNG TIỆN

Người biên soạn: TS. Nguyễn Thị Oanh

MỞ ĐẦU

1. Mục đích môn học
2. Nội dung môn học
3. Sách giáo khoa và tài liệu tham khảo

Chương 1: Các khái niệm cơ bản

1. Các khái niệm
2. Các dạng dữ liệu đa phương tiện
3. Một số vấn đề cơ bản
4. Tổng quan về CSDL hướng đối tượng (Object oriented databases)

Chương 2: Các cấu trúc dữ liệu đa chiều

1. k-d cây
2. Cây tứ phân dạng điểm
3. Cây tứ phân MX
4. R-cây
5. Một số biến thể
6. So sánh các cấu trúc dữ liệu khác nhau

Chương 3: Biểu diễn và tìm kiếm dữ liệu đa phương tiện

1. CSDL văn bản
2. CSDL ảnh
3. CSDL video
4. CSDL đa phương tiện

Chương 4: Truy nhập dữ liệu đa phương tiện

1. Truy nhập dữ liệu đa phương tiện từ đĩa
2. Truy nhập dữ liệu đa phương tiện từ CD-ROM

Chương 5: Trình diễn dữ liệu đa phương tiện

1. Các đối tượng trong trình diễn đa phương tiện
2. Dữ liệu đa phương tiện với ràng buộc thời gian
3. Dữ liệu đa phương tiện với ràng buộc không gian
4. Giải pháp hiệu quả đối với trình diễn phụ thuộc thời gian

13. Tài liệu tham khảo

[1] VS. Subrahmanian, Principles of Multimedia Database Systems, Morgan Kaufmann, San Francisco, 1998

[2] Multimedia Database Management Systems : Research issues and future directions, 1997

[3] Multimedia Database Management : Perspective and Challenges, Advances in Databases, 1995

IT5340 Tìm kiếm và trình diễn thông tin

1. Tên học phần: TÌM KIẾM VÀ TRÌNH DIỄN THÔNG TIN

2. Mã số: IT5340

3. Khối lượng: 3(3-1-0-6)

- Lý thuyết: 45 tiết
- Bài tập: 15 tiết

4. **Yêu cầu học phần:** bắt buộc, học vào kỳ 2

5. **Đối tượng tham dự:** Học viên cao học chuyên ngành Hệ thống thông tin

6. Điều kiện học phần:

7. **Mục tiêu học phần:** Sinh viên nắm được các khái niệm cơ bản trong tìm kiếm thông tin và trình diễn thông tin. Hiểu các mô hình và ứng dụng trong các bài toán liên quan.

8. **Nội dung tóm tắt học phần:** Đề cập đến hai vấn đề lớn thường thấy khi phải xử lý thông tin với một khối lượng lớn: tìm kiếm thông tin (information retrieval) và trình diễn thông tin (information visualization), các kỹ thuật căn bản được sử dụng trong hai lĩnh vực này, ví dụ như mô hình logic, mô hình vector trong tìm kiếm thông tin hoặc các phương pháp vẽ cây và đồ thị trong trình diễn thông tin.

9. Nhiệm vụ của sinh viên:

- Dự lớp: Theo quy định của Bộ GD&ĐT và trường ĐHBKHN.
- Bài tập: Theo quy định của Bộ GD&ĐT và trường ĐHBKHN.
- Thí nghiệm:.

10. Đánh giá kết quả: KT/BT(0.30)-T(BTL:0.70)

- Điểm quá trình: trọng số 0.30
 - Bài tập làm đầy đủ
 - Hoàn thành bài tập lớn
 - Kiểm tra giữa kỳ
- Thi cuối kỳ (bảo vệ bài tập lớn): trọng số 0.70

11. Tài liệu học tập:

- Sách giáo khoa chính:
- Sách tham khảo: Xem phần tài liệu tham khảo.

12. Nội dung chi tiết học phần:

TÌM KIẾM VÀ TRÌNH DIỄN THÔNG TIN

Người biên soạn: TS. Nguyễn Bá Ngọc

MỞ ĐẦU

1. Mục đích môn học
2. Nội dung môn học
3. Sách giáo khoa và tài liệu tham khảo

Chương 1: Giới thiệu chung

1. Tổng quan về tìm kiếm thông tin
2. Các khái niệm cơ bản
3. Quy trình tìm kiếm thông tin

Chương 2: Các mô hình tìm kiếm thông tin

1. Mô hình Boolean
2. Mô hình không gian vec-tơ
3. Mô hình xác suất

4. Một số mô hình khác

Chương 3: Đánh giá hiệu quả tìm kiếm

1. Các thông số cơ bản đánh giá kết quả tìm kiếm: Precision, Re-call, Fall-out, ...
2. Thử nghiệm và đánh giá hệ tìm kiếm thông tin với các bộ dữ liệu

Chương 4: Ngôn ngữ tìm kiếm

1. Giới thiệu chung
2. Tìm kiếm dựa trên từ khóa
3. Các phép toán tìm kiếm

Chương 5: Tổ chức lưu trữ

1. Tổ chức chỉ mục ngược
2. Các kiểu chỉ mục khác

Chương 6: Tìm kiếm

1. Tìm kiếm có cấu trúc
2. Đối sánh mẫu
3. Tìm kiếm với các kiểu dữ liệu phức tạp

Chương 7: Trình diễn thông tin

1. Các khái niệm cơ bản về tương tác người máy
2. Mô hình hóa tương tác
3. Phản hồi của người dùng về kết quả tìm kiếm
4. Vai trò giao diện trong quá trình tìm kiếm thông tin

Chương 8: Các xu hướng phát triển

1. Mô-tơ tìm kiếm trên web
2. Tìm kiếm đa ngôn ngữ
3. Tìm kiếm thông tin đa phương tiện
4. Tìm kiếm thông tin hướng người dùng
5. Các ứng dụng tiêu biểu
6. Một số vấn đề mở

13. Tài liệu tham khảo

- [1] Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, *Introduction to Information Retrieval*, Cambridge University Press. 2008.
- [2] Ricardo Baeza-Yates, Berthier Ribeiro-Neto, *Modern Information Retrieval*. Addison Wesley. 2004
- [3] Baecker, R., Grudin, J., Buxton, W., and Greenberg, S. (1995) *Readings in Human Computer Interaction: Towards the Year 2000 (2nd Edition)*, 950 pages, Morgan Kaufmann Publishers Inc, California. ISBN 1-55860-246-1, Dewey Catalog QA76.9.H85R43.

IT5350 Xử lý thông tin mờ

1. Tên học phần: XỬ LÝ THÔNG TIN MỜ

2. Mã số: IT5350

3. Khối lượng: 3(3-1-0-6)

- Lý thuyết: 45 tiết
- Bài tập: 15 tiết

4. Yêu cầu học phần: bắt buộc, học vào kỳ 1

5. Đối tượng tham dự: Học viên cao học chuyên ngành Hệ thống thông tin

6. Điều kiện học phần:

7. Mục tiêu học phần: Sinh viên có được các kiến thức cơ bản về lý thuyết tập mờ và ứng dụng xử lý các thông tin không chắc chắn, không đầy đủ, không chính xác.

8. Nội dung tóm tắt học phần: Các khái niệm cơ bản về tập mờ, các phép toán với tập mờ, quan hệ mờ, logic mờ và suy luận xấp xỉ, suy diễn mờ; Các ứng dụng của hệ thống mờ và xây dựng hệ mờ.

9. Nhiệm vụ của sinh viên:

- Dự lớp: Theo quy định của Bộ GD&ĐT và trường ĐHBKHN.
- Bài tập: Theo quy định của Bộ GD&ĐT và trường ĐHBKHN.
- Thí nghiệm:.

10. Đánh giá kết quả: KT/BT(0.30)-T(BTL:0.70)

- Điểm quá trình: trọng số 0.30
 - Bài tập làm đầy đủ
 - Hoàn thành bài tập lớn
 - Kiểm tra giữa kỳ
- Thi cuối kỳ (bảo vệ bài tập lớn): trọng số 0.70

11. Tài liệu học tập:

- Sách giáo khoa chính:

[1] Hồ Thuần, Đặng Thanh Hà, Logic mờ và ứng dụng, NXB ĐHQG Hà nội, 2007

- Sách tham khảo: Xem phần tài liệu tham khảo.

12. Nội dung chi tiết học phần:

XỬ LÝ THÔNG TIN MỜ

Người biên soạn: PGS. Trần Đình Khang

MỞ ĐẦU

1. Mục đích môn học
2. Nội dung môn học
3. Sách giáo khoa và tài liệu tham khảo

Chương 1 – Nhập môn

1. Thông tin mờ
2. Biến ngôn ngữ

Chương 2 – Tập mờ

1. Định nghĩa tập mờ
2. Các phép toán với tập mờ trên cùng không gian tham chiếu
3. Nguyên lý mở rộng
4. Số mờ
5. Họ các toán tử Hamacher, Yager, Dubois & Prade, độ đo mờ, tính mờ, tập mờ kiểu LR

Chương 3 – Quan hệ mờ

1. Quan hệ mờ
2. Phép hợp thành

Chương 4 – Logic mờ

1. Các mở rộng của logic kinh điển
2. Logic mờ, suy luận xấp xỉ
3. Các phép kéo theo mờ
4. Miền giá trị chân lý

Chương 5 – Suy diễn mờ

1. Suy diễn mờ đơn điều kiện
2. Tiêu chuẩn suy diễn “tốt”
3. Suy diễn mờ mở rộng
4. Nội suy mờ

Chương 6 – Hệ mờ

1. Kiến trúc chung của hệ mờ
2. Xây dựng hệ mờ
3. Các thủ tục xây dựng hệ mờ từ bộ dữ liệu vào ra
4. Học và điều chỉnh tham số hệ mờ

Chương 7 – Xu hướng phát triển

13. Tài liệu tham khảo

- [1] H.J. Zimmermann, Fuzzy set theory and its applications, Kluwer Academic Publishers, 1991.
- [2] James J. Buckley, Esfandiar Eslami, An Introduction to fuzzy logic and fuzzy sets, Physica-Verlag Heidelberg New York, 2002

IT5360 Xử lý ngôn ngữ tự nhiên

1. Tên học phần: XỬ LÝ NGÔN NGỮ TỰ NHIÊN

2. Mã số: IT5360

3. Khối lượng: 3(3-1-0-6)

- Lý thuyết: 45 tiết
- Bài tập: 15 tiết

4. Yêu cầu học phần: tự chọn, học vào kỳ 1 hoặc kỳ 2

5. Đối tượng tham dự: Học viên cao học chuyên ngành Hệ thống thông tin

6. Điều kiện học phần:

7. Mục tiêu học phần: Sau khi hoàn thành học phần này, sinh viên phải nắm được các kiến thức sau:

- Cách thức biểu diễn các cấu trúc (từ, cú pháp) của văn bản.
- Một số thuật toán để xây dựng cấu trúc của văn bản.
- Cách xây dựng các mô hình xác suất để lựa chọn kết quả tốt nhất.
- Phương pháp xây dựng các ứng dụng thực tế của xử lý ngôn ngữ trên cơ sở các kiến thức nền tảng đã học.

8. Nội dung tóm tắt học phần: Học phần này cung cấp các kiến thức cơ sở về các phương pháp xử lý ngôn ngữ tự nhiên thông qua máy tính như phân tích hình thái từ, phân tích từ loại, phân tích cú pháp. Một số hướng nghiên cứu mới trong xử lý ngôn ngữ tự nhiên cũng được giới thiệu trong học phần này.

9. Nhiệm vụ của sinh viên:

- Dự lớp: Theo quy định của Bộ GD&ĐT và trường ĐHBKHN.
- Bài tập: Theo quy định của Bộ GD&ĐT và trường ĐHBKHN.
- Thí nghiệm:.

10. Đánh giá kết quả: KT/BT(0.30)-T(BTL:0.70)

- Điểm quá trình: trọng số 0.30
 - Bài tập làm đầy đủ
 - Hoàn thành bài tập lớn
 - Kiểm tra giữa kỳ
- Thi cuối kỳ (bảo vệ bài tập lớn): trọng số 0.70

11. Tài liệu học tập:

- Sách giáo khoa chính:
 - S. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach (3rd Edition)*. Prentice Hall, 2009.
- Sách tham khảo: Xem phần tài liệu tham khảo.

12. Nội dung chi tiết học phần:

XỬ LÝ NGÔN NGỮ TỰ NHIÊN

Người biên soạn: PGS. Lê Thanh Hương

MỞ ĐẦU

1. Mục đích môn học
2. Nội dung môn học
3. Sách giáo khoa và tài liệu tham khảo

PHẦN I. CÁC KHÁI NIỆM CƠ BẢN

1. Mở đầu

- 1.1 Giới thiệu chung
- 1.2 Các hướng nghiên cứu trong xử lý ngôn ngữ tự nhiên
- 1.3 Các mức phân tích trong xử lý ngôn ngữ tự nhiên
- 1.4 Các ứng dụng của xử lý ngôn ngữ tự nhiên

2 Hình thái từ

- 2.1 Định nghĩa
- 2.2 Các dạng hình thái từ
- 2.3 Phân tích hình thái từ dựa trên ôtomat hữu hạn

3 Từ loại

- 3.1 Định nghĩa
- 3.2 Các cách tiếp cận trong phân tích từ loại
 - 3.2.1 Phương pháp Hidden Markov Model
 - 3.2.2 Phương pháp dựa trên ràng buộc ngữ pháp
 - 3.2.3 Phương pháp chuyển đổi

4 Cú pháp

- 4.1 Một số khái niệm
- 4.2 Phân tích cú pháp dựa trên văn phạm phi ngữ cảnh

- 4.2.1 Giới thiệu
- 4.2.2 Phân tích kiểu trên xuống
- 4.2.3 Phân tích kiểu dưới lên
- 4.2.4 Một số vấn đề trong phân tích cú pháp
- 4.3 Phân tích cú pháp theo cách tiếp cận xác suất
 - 4.3.1 Một số khái niệm về xác suất
 - 4.3.2 Phân tích cú pháp sử dụng xác suất
- 4.4 Một số cách tiếp cận hiện đại trong phân tích cú pháp

PHẦN 2. CÁC HƯỚNG NGHIÊN CỨU MỚI TRONG XLNNTN

5 Dịch máy

- 5.1 Các mô hình dịch máy cổ điển
- 5.2 Cách tiếp cận dịch máy thống kê

6 Phân loại văn bản

- 6.1 Bài toán phân loại văn bản
- 6.2 Một số cách tiếp cận trong phân loại văn bản

7 Tìm kiếm thông tin

- 7.1 Định nghĩa và các ứng dụng của tìm kiếm thông tin
- 7.2 Một số cách tiếp cận trong tìm kiếm thông tin

13. Tài liệu tham khảo

1. Nguyễn Thanh Thủy, *Trí Tuệ Nhân Tạo*. Nhà xuất bản giáo dục, Nhà xuất bản KHKT.
2. Phan Huy Khánh. *Lập trình logic trong Prolog*. NXB Đại học quốc gia Hà Nội. 2004.
3. N. J. Nilsson. *Artificial Intelligence: A New Synthesis*. Morgan Kaufmann, 1998.
4. T. M. Mitchell. *Machine Learning*. McGraw-Hill, 1997.

IT5370 Quản trị dự án

1. Tên học phần: QUẢN TRỊ DỰ ÁN

2. Mã số: IT5370

3. Khối lượng: 2(2-1-0-4)

- Lý thuyết: 30 tiết
- Bài tập: 15 tiết

4. Yêu cầu học phần: tự chọn, học vào kỳ 1 hoặc kỳ 2

5. Đối tượng tham dự: Học viên cao học chuyên ngành Hệ thống thông tin

6. Điều kiện học phần:

7. Mục tiêu học phần: QLDA là một trong những công cụ quản lý hữu hiệu, nhằm trợ giúp cho các loại hình DA đạt được những hiệu quả khoa học, kỹ thuật cũng như kinh tế góp phần bảo vệ lợi ích của nhà nước cũng như các cơ quan xí nghiệp có liên quan.

8. Nội dung tóm tắt học phần: Ngữ cảnh QLDA; Các chu trình QLDA: Khởi đầu, Lập kế hoạch, Kiểm soát, Thực hiện, Kết thúc; Các lĩnh vực tri thức QLDA: Tích hợp, Phạm vi, Thời gian, Chi phí, Chất lượng, Nhân lực, Truyền thông, Rủi ro, Thu mua; QLDA CNTT.

9. Nhiệm vụ của sinh viên:

- Dự lớp: Theo quy định của Bộ GD&ĐT và trường ĐHBKHN.
- Bài tập: Theo quy định của Bộ GD&ĐT và trường ĐHBKHN.
- Thí nghiệm:.

10. Đánh giá kết quả: KT/BT(0.30)-T(BTL:0.70)

- Điểm quá trình: trọng số 0.30
 - Bài tập làm đầy đủ
 - Hoàn thành bài tập lớn
 - Kiểm tra giữa kỳ
- Thi cuối kỳ (bảo vệ bài tập lớn): trọng số 0.70

11. Tài liệu học tập:

- Sách giáo khoa chính:
- Sách tham khảo: Xem phần tài liệu tham khảo.

12. Nội dung chi tiết học phần:

QUẢN TRỊ DỰ ÁN

Người biên soạn: TS. Phạm Văn Hải

MỞ ĐẦU

1. Mục đích môn học
2. Nội dung môn học
3. Sách giáo khoa và tài liệu tham khảo

NỘI DUNG

1. Ngữ cảnh QLDA
2. Các chu trình QLDA: Khởi đầu, Lập kế hoạch, Kiểm soát, Thực hiện, Kết thúc
3. Các lĩnh vực tri thức QLDA: Tích hợp, Phạm vi, Thời gian, Chi phí, Chất lượng, Nhân lực, Truyền thông, Rủi ro, Thu mua
4. Quản lý tích hợp dự án
5. Quản lý phạm vi dự án
6. Quản lý thời gian dự án
7. Quản lý chi phí dự án
8. Quản lý chất lượng dự án
9. Quản lý nhân lực dự án
10. Quản lý truyền thông dự án
11. Quản lý rủi ro dự án
12. Quản lý thu mua dự án
13. QLDA: Case study
14. Quản lý dự án Công nghệ Thông tin

13. Tài liệu tham khảo

1. Pankaj Jalote, Software Project Management in Practice, Pearson Education, Delhi 110092. India, 2002
2. Pankaj Jalote, CMM in practice, processes for excuting Software Project at Infosys, Addison-Wesley Edition, 2000
3. William R. Duncan, A guide to the project management body of knowledge, Project Management Institute, Four Campus Boulevard Newtown Square, PA 19073-3299 USA

IT5530 Hệ thống hướng tác tử

1. Tên học phần: **HỆ THỐNG HƯỚNG TÁC TỬ**
2. Mã số: **IT5530**
3. Khối lượng: **2(2-1-0-4)**

- Lý thuyết: 30 tiết

- Bài tập: 15 tiết

4. Yêu cầu học phần: tự chọn, học vào kỳ 1 hoặc kỳ 2

5. Đối tượng tham dự: Học viên cao học chuyên ngành Hệ thống thông tin

6. Điều kiện học phần:

7. Mục tiêu học phần: Cung cấp khái niệm về tác tử cũng như một phương pháp luận mới trong việc xây dựng các ứng dụng phần mềm thông minh và tự chủ. Giúp sinh viên nhận thức các hệ thống hướng tác tử như những sản phẩm phần mềm có tri thức để thực hiện các tác vụ theo cách thức cộng tác, phối hợp và cạnh tranh.

8. Nội dung tóm tắt học phần: Khái niệm tác tử và hệ đa tác tử. Các đặc điểm về hệ thống hướng tác tử, ưu nhược điểm và khả năng ứng dụng. Kiến trúc tác tử. Lý thuyết tương tác giữa các tác tử, cộng tác phối hợp giữa các tác tử. Các công cụ phát triển tác tử. Giao thức và phương pháp liên lạc giữa các tác tử. Các kỹ thuật và thư viện tiêu biểu để phát triển phần mềm dựa tác tử. Kỹ thuật phân tích và thiết kế các hệ hướng tác tử trong công nghệ phần mềm.

9. Nhiệm vụ của sinh viên:

- Dự lớp: Theo quy định của Bộ GD&ĐT và trường ĐHBKHN.
- Bài tập: Theo quy định của Bộ GD&ĐT và trường ĐHBKHN.
- Thí nghiệm:.

10. Đánh giá kết quả: KT/BT(0.30)-T(BTL:0.70)

- Điểm quá trình: trọng số 0.30
 - Bài tập làm đầy đủ
 - Hoàn thành bài tập lớn
 - Kiểm tra giữa kỳ
- Thi cuối kỳ (bảo vệ bài tập lớn): trọng số 0.70

11. Tài liệu học tập:

- Sách giáo khoa chính:

1. Tác tử công nghệ phần mềm hướng tác tử, **Lê Tấn Hùng, Từ Minh Phương, Huỳnh Quyết Thắng**, NXB KHKT 2006.
2. Agent-Based Software Engineering, **Michael Wooldridge**, Mitsubishi Electric Digital Library Group, 1997
3. *An Introduction to Multiagent Systems* by **Michael Wooldridge**. Published in February 2002 by **John Wiley & Sons** (Chichester, England). ISBN 0 47149691X.

- Sách tham khảo: Xem phần tài liệu tham khảo.

12. Nội dung chi tiết học phần:

HỆ THỐNG HƯỚNG TÁC TỬ

Người biên soạn: TS. Cao Tuấn Dũng

MỞ ĐẦU

1. Mục đích môn học
2. Nội dung môn học
3. Sách giáo khoa và tài liệu tham khảo

Chương 1. Các khái niệm cơ sở về tác tử và hệ đa tác tử

- 1.1 Khái niệm tác tử
- 1.2 So sánh với một số khái niệm khác
- 1.3 Các đặc điểm của agent
- 1.4 Tác tử và môi trường
- 1.5 Hệ hướng tác tử và hệ đa tác tử
- 1.6 Tác tử và các lĩnh vực nghiên cứu liên quan
- 1.7 Phân loại tác tử
- 1.8 Một số ứng dụng

Chương 2. Kiến trúc tác tử và hệ thống dựa tác tử

- 2.1 Kiến trúc chung của tác tử
- 2.2 Cảm nhận và tác động
- 2.3 Cơ chế ra quyết định
- 2.4 Kiến trúc phản xạ
- 2.5 Kiến trúc suy diễn
- 2.6 Kiến trúc lai.

Chương 3. Tương tác cạnh tranh

- 3.1 Lý thuyết Trò chơi và tác tử
- 3.2 Hành động ưu thế và cân bằng Nash.
- 3.3 Tác tử tư lợi
- 3.4 Tiêu chí thiết kế giao thức tương tác
- 3.5 Tương tác qua Đấu giá
- 3.6 Thương lượng

Chương 4. Tương tác Phối hợp

- 4.1 Khái niệm phối hợp
- 4.2 Chia sẻ công việc
- 4.3 Chia sẻ kết quả

Chương 5. Giao thức và ngôn ngữ liên lạc giữa các tác tử

- 5.1 Giao thức liên lạc
- 5.2 Lý thuyết hành động lời nói (Speech-Act)
- 5.2 Ngôn ngữ ACL
- 5.3 Ontology và ngôn ngữ ACL

Chương 6. Công nghệ phần mềm hướng tác tử

- 6.1 Khái quát về phần mềm dựa Agent
- 6.2 Khả năng và đặc điểm phân tích thiết kế dựa trên agent
- 6.3 Phương pháp luận hướng tác tử
- 6.4 Mở rộng UML cho đặc tả agent

Chương 7. Các nền tảng và công cụ phát triển hệ thống hướng tác tử.

- 7.1 Nền tảng JADE
- 7.2 NETLOGO
- 7.3 JASON

13. Tài liệu tham khảo

1. Multiagent Systems, A Modern Approach to Distributed Artificial Intelligence, **Weiss, Gerhard** (Ed.), MIT Press, 1999.
2. Software Agents, AAI **Bradshaw, J.** (Ed.), Press/MIT Press, Cambridge, MA, 1997.

IT5390 Học máy

- 1. Tên học phần: HỌC MÁY**
- 2. Mã số: IT5390**
- 3. Khối lượng: 2(2-1-0-4)**

- Lý thuyết: 30 tiết
- Bài tập: 15 tiết

4. Yêu cầu học phần: tự chọn, học vào kỳ 1 hoặc kỳ 2

5. Đối tượng tham dự: Học viên cao học chuyên ngành Hệ thống thông tin

6. Điều kiện học phần:

7. Mục tiêu học phần: Học phần này giúp sinh viên có kiến thức và hiểu biết về các phương pháp và giải thuật học máy cơ bản, các điểm mạnh (ưu điểm) và các điểm yếu (nhược điểm) của các giải thuật học máy, các phương pháp tiến hành thí nghiệm và đánh giá hiệu năng của các hệ thống học máy, và việc sử dụng học máy để giải quyết các bài toán thực tế. Học phần này cũng giúp sinh viên làm quen và sử dụng được công cụ WEKA.

8. Nội dung tóm tắt học phần: Giới thiệu chung về học máy và công cụ WEKA. Đánh giá hiệu năng hệ thống học máy. Các phương pháp học dựa trên xác suất: Định lý Bayes (Bayes theorem), Xác suất hậu nghiệm cực đại (maximum a posteriori), Đánh giá khả năng có thể nhất (maximum likelihood estimation), Phân loại Naïve Bayes (naïve Bayes classifier). Các phương pháp học có giám sát: Học cây quyết định (decision tree learning), Học dựa trên láng giềng gần nhất (nearest neighbor learner), Mạng nơ-ron nhân tạo (perceptron & neural network), Máy vectơ hỗ trợ (support vector machine). Các phương pháp học không giám sát: Phân cụm dựa trên phân tách (partition-based clustering), Phân cụm dựa trên tích tụ phân cấp (hierarchical agglomerative clustering).

9. Nhiệm vụ của sinh viên:

- Dự lớp: Theo quy định của Bộ GD&ĐT và trường ĐHBKHN.
- Bài tập: Theo quy định của Bộ GD&ĐT và trường ĐHBKHN.
- Thí nghiệm:.

10. Đánh giá kết quả: KT/BT(0.30)-T(BTL:0.70)

- Điểm quá trình: trọng số 0.30
 - Bài tập làm đầy đủ
 - Hoàn thành bài tập lớn
 - Kiểm tra giữa kỳ
- Thi cuối kỳ (bảo vệ bài tập lớn): trọng số 0.70

11. Tài liệu học tập:

- Sách giáo khoa chính:
 1. T. M. Mitchell. Machine Learning. McGraw-Hill, 1997.
 2. Công cụ WEKA, <http://www.cs.waikato.ac.nz/ml/weka>.
- Sách tham khảo: Xem phân tài liệu tham khảo.

12. Nội dung chi tiết học phần:

HỌC MÁY

Người biên soạn: TS. Nguyễn Nhật Quang

MỞ ĐẦU

1. Mục đích môn học
2. Nội dung môn học
3. Sách giáo khoa và tài liệu tham khảo

NỘI DUNG

1. Giới thiệu về Học máy
2. Giới thiệu về công cụ WEKA
3. Đánh giá hiệu năng của hệ thống học máy
4. Các phương pháp học dựa trên xác suất
5. Các phương pháp học dựa trên xác suất
6. Học cây quyết định
7. Học cây quyết định
8. Học dựa trên láng giềng gần nhất
9. Mạng nơ-ron nhân tạo
10. Mạng nơ-ron nhân tạo
11. Mạng nơ-ron nhân tạo
12. Máy vectơ hỗ trợ
13. Máy vectơ hỗ trợ
14. Phân cụm dựa trên phân tách
15. Phân cụm dựa trên tích tụ phân cấp

13. Tài liệu tham khảo

1. B. Liu. Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data. Springer, 2006.

IT6300 Các hệ cơ sở dữ liệu tiên tiến

1. Tên học phần: CÁC HỆ CƠ SỞ DỮ LIỆU TIÊN TIẾN

2. Mã số: IT6300

3. Khối lượng: 3(2.5-1-0-6)

- Lý thuyết: 37.5 tiết
- Bài tập: 15 tiết

4. Yêu cầu học phần: bắt buộc, học vào kỳ 3 (có thể là học phần tự chọn của chuyên ngành khác).

5. Đối tượng tham dự: Học viên cao học chuyên ngành Hệ thống thông tin

6. Điều kiện học phần:

7. Mục tiêu học phần: Kết thúc học phần này, học viên có kiến thức chuyên sâu về các nguyên lý và các kỹ thuật thiết kế các hệ cơ sở dữ liệu tiên tiến; các xu hướng phát triển trong lĩnh vực cơ sở dữ liệu và giúp cho học viên có thể phát triển các ứng dụng nâng cao, đáp ứng nhu cầu khai thác dữ liệu phong phú và đa dạng của các tổ chức, xí nghiệp.

8. Nội dung tóm tắt học phần: Nội dung học phần bao gồm: giới thiệu nhanh về các hệ cơ sở dữ liệu quan hệ, các chủ đề nâng cao về các hệ cơ sở dữ liệu hiện đại bao hàm các cơ sở dữ liệu phân tán, suy diễn, hướng đối tượng và xử lý phân tích trực tuyến. Học viên được thảo luận về mô tả dữ liệu và ngôn ngữ truy vấn, thiết kế cơ sở dữ liệu, xử lý và tối ưu hóa truy vấn, các kho dữ liệu và khai phá dữ liệu.

9. Nhiệm vụ của sinh viên:

- Dự lớp: Theo quy định của Bộ GD&ĐT và trường ĐHBKHN.

- Bài tập: Theo quy định giáo viên giảng dạy và trường ĐHBKHN.

10. Đánh giá kết quả: KT/BT(0.30)-T(TL:0.70)

Điểm quá trình: trọng số 0.30

- Bài tập làm đầy đủ
- Hoàn thành bài tập lớn, có báo cáo và bảo vệ

Thi cuối kỳ (tự luận): trọng số 0.70

11. Tài liệu học tập:

- Sách giáo khoa chính:
- Sách tham khảo: Xem phần tài liệu tham khảo.

12. Nội dung chi tiết học phần:

CÁC HỆ CƠ SỞ DỮ LIỆU TIÊN TIẾN

Người biên soạn: PGS. Nguyễn Kim Anh

MỞ ĐẦU

1. Mục đích môn học
2. Nội dung môn học
3. Sách giáo khoa và tài liệu tham khảo

CHƯƠNG 1. CÁC XU THẾ PHÁT TRIỂN TRONG CƠ SỞ DỮ LIỆU

- 1.1. Ưu điểm và nhược điểm của các hệ cơ sở dữ liệu cổ điển, tập trung
- 1.2. Các ứng dụng nâng cao
- 1.3. Xu hướng phát triển trong lĩnh vực cơ sở dữ liệu

CHƯƠNG 2. CÁC HỆ CƠ SỞ DỮ LIỆU PHÂN TÁN

- 2.1. Đại cương về các hệ cơ sở dữ liệu phân tán
- 2.2. Phân loại các hệ cơ sở dữ liệu phân tán
- 2.3. Thiết kế cơ sở dữ liệu phân tán thuần nhất
- 2.4. Thiết kế cơ sở dữ liệu phân tán không thuần nhất
- 2.5. Tối ưu hoá câu hỏi phân tán
- 2.6. Quản trị giao dịch phân tán và điều khiển tương tranh

CHƯƠNG 3. CÁC HỆ CƠ SỞ DỮ LIỆU SUY DIỄN

- 3.1. Logic và cơ sở dữ liệu
- 3.2. Hệ quản trị cơ sở dữ liệu suy diễn
- 3.3. Ngôn ngữ luật DATALOG
- 3.4. Đánh giá câu hỏi suy diễn

CHƯƠNG 4. CÁC HỆ CƠ SỞ DỮ LIỆU HƯỚNG ĐỐI TƯỢNG

- 4.1. Mô hình dữ liệu hướng đối tượng
- 4.2. Ngôn ngữ định nghĩa và truy vấn hướng đối tượng
- 4.3. Xử lý và tối ưu hoá truy vấn hướng đối tượng
- 4.4. Các chỉ dẫn trong hệ cơ sở dữ liệu hướng đối tượng

CHƯƠNG 5. XỬ LÝ PHÂN TÍCH TRỰC TUYẾN

- 5.1 Sự cần thiết đối với OLAP
- 5.2 Mô hình dữ liệu đa chiều
- 5.3 Xây dựng mô hình dữ liệu đa chiều từ mô hình dữ liệu khái niệm của các hệ cơ sở dữ liệu tác nghiệp
- 5.4 Các thao tác OLAP trong mô hình dữ liệu đa chiều
- 5.5 Cập nhật các chiều OLAP

- 5.6 Tổ chức lưu trữ dữ liệu OLAP
- 5.7 Xử lý hiệu quả các yêu cầu OLAP
- 5.8 Tích hợp OLAP và khai phá dữ liệu (DM)

13. Nội dung các bài thí nghiệm (thực hành, tiểu luận, bài tập lớn)

Các bài tập lớn sẽ hướng sinh viên vào việc giải quyết một số vấn đề liên quan đến các hệ cơ sở dữ liệu phân tán, hướng đối tượng, suy diễn và OLAP.

14. Tài liệu tham khảo

1. R. Elmasri and S. Navathe, *Fundamentals of Database Systems*, 2004 (4th edition), Addison-Wesley.
2. M.T.Ozsu, P. Valduriez, *Principles of Distributed Database Systems*, 1999 (2nd edition), Prentice Hall
3. J.D.Ullman, *Principle of Knowledge-base and Database Systems*, 1988
4. C.T.Yu, W.Meng, *Principles of Database Query Processing for Advanced Applications*, 1998

IT6310 Tìm kiếm dữ liệu đa phương tiện

1. Tên học phần: TÌM KIẾM DỮ LIỆU ĐA PHƯƠNG TIỆN

2. Mã số: IT6310

3. Khối lượng: 2 (1.5-1-0-4)

- Lý thuyết: 22.5 giờ
- Bài tập: 15 giờ

4. Yêu cầu học phần: bắt buộc, học vào kỳ 3 (có thể là học phần tự chọn của chuyên ngành khác).

5. Đối tượng tham dự: Học viên cao học chuyên ngành Hệ thống thông tin

6. Điều kiện học phần:

7. Mục tiêu học phần: Giới thiệu các kỹ thuật cơ bản và các phương thức hỗ trợ xây dựng hệ thống tìm kiếm dữ liệu đa phương tiện.

8. Nội dung vắn tắt học phần: Học phần này trình bày hệ thống tìm kiếm dữ liệu đa phương tiện, chủ yếu tập trung vào tìm kiếm ảnh/vidéo dựa trên nội dung. Các thành phần chính trong hệ thống cũng như các kỹ thuật cơ bản để xây dựng các thành phần này sẽ được giới thiệu : biểu diễn dữ liệu, tổ chức lưu trữ, tìm kiếm, ... Trong học phần này chúng tôi cũng đề cập đến một số khái niệm cơ bản về xử lý, phân tích dữ liệu đa phương tiện.

9. Nhiệm vụ của sinh viên:

Dự lớp: đầy đủ theo quy chế

Bài tập: hoàn thành các bài tập của học phần

10. Đánh giá kết quả:

- Quá trình : 0.3
- Cuối kỳ : tự luận hoặc bảo vệ bài tập lớn : 0.7

11. Tài liệu học tập

Sách, giáo trình chính: Bài giảng trên lớp và các tài liệu giáo viên yêu cầu

Sách tham khảo: xem đề cương chi tiết

12. Nội dung chi tiết học phần:

TÌM KIẾM DỮ LIỆU ĐA PHƯƠNG TIỆN

Người biên soạn : TS. Nguyễn Thị Oanh

Mở đầu

Chương 1: Tổng quan

- 1.1. Dữ liệu đa phương tiện và các khái niệm cơ bản
- 1.2. Tổng quan về tìm kiếm dữ liệu đa phương tiện

Chương 2 : Tìm kiếm dữ liệu văn bản

- 2.1. Biểu diễn văn bản
- 2.2. Đánh chỉ mục văn bản
- 2.3. Các mô hình tìm kiếm văn bản
- 2.4. Phản hồi thông tin
- 2.5. Đánh giá hiệu năng

Chương 3 : Tìm kiếm dữ liệu ảnh

- 3.1. Các khái niệm và phép xử lý, phân tích ảnh cơ bản
- 3.2. Trích chọn đặc trưng và biểu diễn dữ liệu ảnh
- 3.3. Tìm kiếm dựa trên từ khóa
- 3.4. Tìm kiếm dựa trên nội dung

Chương 4 : Tìm kiếm dữ liệu video

- 5.1. Các khái niệm cơ bản
- 5.2. Phân tích video
- 5.3. Biểu diễn nội dung: mục tiêu và đặc trưng
- 5.4. Biểu diễn truy vấn và thực hiện tìm kiếm : ví dụ cụ thể

Chương 5: Tổ chức dữ liệu đa chiều và các phương pháp tìm kiếm

- 6.1. Các tổ chức dữ liệu
- 6.2. Giảm chiều dữ liệu
- 6.3. Các phương pháp tìm kiếm

Chương 6: Đánh giá hiệu năng

- 7.1. Dữ liệu thử nghiệm
- 7.2. Các phép đo đánh giá hiệu năng

13. Nội dung các bài thí nghiệm (thực hành, tiểu luận, bài tập lớn)

Các bài tập lớn sẽ hướng sinh viên vào việc giải quyết một số vấn đề liên quan trong hệ thống truy vấn theo nội dung.

14. Tài liệu tham khảo

- [1] Michael S. Lew, *Principles of Visual Information Retrieval*, Springer 2001
[2] Blanken, H.M.; de Vries, A.P.; Blok, H.E.; Feng, L., *Multimedia Retrieval*, Springer 2007
[3] Patrick Gros, *L'indexation Multimédia*, LAVOIRSIER 2007
[4] Rafael C. Gonzalez and Richard E. Woods, *Digital Image Processing*, 2002
[5] Ricardo Baeza-Yates and Berthier Ribeiro-Neto, *Modern Information Retrieval*, 1999

IT6320 Khai phá dữ liệu

1. Tên học phần: KHAI PHÁ DỮ LIỆU

2. Mã số: IT6320

3. Khối lượng: 3(2.5-1-0-6)

- Lý thuyết: 37.5 tiết
- Bài tập: 15 tiết

4. Yêu cầu học phần: bắt buộc, học vào kỳ 3 (có thể là học phần tự chọn của chuyên ngành khác).

5. Đối tượng tham dự: Học viên cao học chuyên ngành Hệ thống thông tin

6. Điều kiện học phần:

7. Mục tiêu học phần: Kết thúc học phần này, học viên có kiến thức về quá trình khai phá dữ liệu và các kỹ thuật quan trọng và phổ biến nhất trong khai phá dữ liệu.

8. Nội dung tóm tắt học phần: Giới thiệu về quá trình khai phá dữ liệu. Tiền xử lý dữ liệu. Kỹ thuật phát hiện các luật kết hợp. Các kỹ thuật phân lớp và dự đoán. Các kỹ thuật phân nhóm. Khai thác công cụ WEKA để đánh giá các kỹ thuật khai phá dữ liệu.

9. Nhiệm vụ của sinh viên:

- Dự lớp: Đầy đủ theo quy chế.
- Bài tập: Hoàn thành các bài tập của học phần.

10. Đánh giá kết quả: KT/BT(0.3)-T(TL:0.7)

Điểm quá trình: trọng số 0.3

- Làm đầy đủ các bài tập.

Thi cuối kỳ (tự luận): trọng số 0.7

11. Tài liệu học tập:

- Sách giáo khoa chính:
Pang-Ning Tan, Michael Steinbach, and Vipin Kumar, *Introduction to Data Mining*, Addison-Wesley, 2005.
- Sách tham khảo:
Xem phần tài liệu tham khảo.

12. Nội dung chi tiết học phần:

KHAI PHÁ DỮ LIỆU

Người biên soạn: TS. Nguyễn Nhật Quang

MỞ ĐẦU

1. Mục đích môn học
2. Nội dung môn học
3. Sách giáo khoa và tài liệu tham khảo

CHƯƠNG 1. GIỚI THIỆU

- 1.1 Khai phá dữ liệu là gì?

- 1.2 Các nhiệm vụ chính trong khai phá dữ liệu
- 1.3 Các bước chính của quá trình khai phá dữ liệu
- 1.4 Các ứng dụng tiêu biểu của khai phá dữ liệu
- 1.5 Các vấn đề thách thức trong khai phá dữ liệu
- 1.6 Giới thiệu về công cụ khai phá dữ liệu WEKA

CHƯƠNG 2. TIỀN XỬ LÝ DỮ LIỆU

- 2.1 Mô tả dữ liệu
- 2.2 Tại sao phải tiền xử lý dữ liệu?
- 2.3 Làm sạch dữ liệu
- 2.4 Lấy mẫu dữ liệu
- 2.5 Tích hợp và chuyển đổi dữ liệu
- 2.6 Thu nhỏ và giảm bớt dữ liệu

CHƯƠNG 3. PHÁT HIỆN CÁC LUẬT KẾT HỢP

- 3.1 Định nghĩa bài toán phát hiện các luật kết hợp
- 3.2 Giải thuật phát hiện các luật kết hợp Apriori
 - 3.2.1 Sản sinh ra các tập mục thường xuyên
 - 3.2.2 Sản sinh ra các luật kết hợp
- 3.3 Độ phức tạp tính toán của giải thuật Apriori
- 3.4 Các cải tiến đối với giải thuật Apriori

CHƯƠNG 4. PHÂN LỚP VÀ DỰ ĐOÁN

- 4.1 Định nghĩa bài toán phân lớp và dự đoán
- 4.2 Lý thuyết Bayes và phân lớp dựa trên phương pháp học Bayes
- 4.3 Phân lớp bằng cây quyết định
- 4.4 Phân lớp và dự đoán dựa trên các lân cận gần nhất
- 4.5 Phân lớp và dự đoán bằng mạng nơ ron nhân tạo
- 4.6 Đánh giá các mô hình phân lớp và dự đoán thu được

CHƯƠNG 5. PHÂN NHÓM

- 5.1 Định nghĩa bài toán phân nhóm
- 5.2 Phân nhóm dựa trên chia cắt (k-Means)
- 5.3 Phân nhóm dựa trên tích tụ phân cấp (HAC)
- 5.4 Đánh giá các mô hình phân nhóm thu được

13. Nội dung các bài tập

Bài 1: Các bài tập về phát hiện luật kết hợp (có sử dụng công cụ WEKA).

Bài 2: Các bài tập về phân lớp và dự đoán (có sử dụng công cụ WEKA).

Bài 3: Các bài tập về phân nhóm (có sử dụng công cụ WEKA).

14. Tài liệu tham khảo

1. Jiawei Han and Micheline Kamber, *Data Mining: Concepts and Techniques*, 2nd Edition, Morgan Kaufmann, 2006.
2. Ian H. Witten and Eibe Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd Edition, Morgan Kaufmann, 2005.
3. WEKA (<http://www.cs.waikato.ac.nz/ml/weka/>).

IT6360 Kỹ nghệ tri thức

1. Tên học phần: KỸ NGHỆ TRI THỨC

2. Mã số: IT6360

3. Khối lượng: 2 (1.5-1-0-4)

- Lý thuyết: 22.5 giờ
- Bài tập: 15 giờ

4. Yêu cầu học phần: tự chọn đối với chương trình thạc sĩ khoa học chuyên ngành Hệ thống Thông tin.

5. Đối tượng tham dự: Học viên cao học chuyên ngành Hệ thống Thông tin

6. Điều kiện học phần: không có

7. Mục tiêu học phần: Kỹ nghệ tri thức nghiên cứu việc xây dựng và tích hợp tri thức vào các hệ thống máy tính nhằm mục đích giải quyết các bài toán thực tế phức tạp, thường đòi hỏi một mức độ am hiểu chuyên môn sâu. Bài giảng sẽ cung cấp một cái nhìn đầy đủ và toàn diện các vấn đề cốt lõi trong kỹ nghệ tri thức. Các chủ đề cốt lõi bao gồm các hệ thống dựa trên tri thức; học, khai phá, và biểu diễn tri thức; suy diễn tự động; các vấn đề về việc lưu trữ, tương tác, tích hợp và bảo mật. Ngoài ra bài giảng cũng minh họa một vài hệ thống dựa trên tri thức. Sau khóa học, sinh viên sẽ có một cái nhìn tổng thể về lĩnh vực.

8. Nội dung tóm tắt học phần: Giới thiệu tri thức và kỹ nghệ tri thức; Tổng quan về kỹ nghệ tri thức; giới thiệu một vài hệ thống dựa trên tri thức; học, khai phá, và biểu diễn tri thức; suy diễn tự động; các vấn đề về việc lưu trữ, tương tác, tích hợp và bảo mật; Thảo luận các chuyên đề nghiên cứu và ứng dụng kỹ nghệ tri thức: trí tuệ nhân tạo, hệ chuyên gia và hệ trợ giúp quyết định; Bài tập dự án về kỹ nghệ tri thức.

9. Nhiệm vụ của sinh viên:

- Dự lớp: Theo quy định của Bộ GD&ĐT và trường ĐHBKHN.
- Bài tập: Theo quy định của Bộ GD&ĐT và trường ĐHBKHN.

10. Đánh giá kết quả: KT/BT(0.30)-T(TL:0.70)

Điểm quá trình: trọng số 0.30

- Bài tập (ở lớp, ở nhà)

Thi cuối kỳ (bài tập lớn): trọng số 0.70

- Đánh giá bài tập lớn
- Bảo vệ bài tập lớn

11. Tài liệu học tập:

- Sách giáo khoa chính:
- Sách tham khảo: Xem phần tài liệu tham khảo.

12. Nội dung chi tiết học phần:

KỸ NGHỆ TRI THỨC

Người biên soạn:

TS. Thân Quang Khoát
TS. Nguyễn Nhật Quang

MỞ ĐẦU

CHƯƠNG 1. Tổng quan về kỹ nghệ tri thức

- 1.1. Khái niệm về dữ liệu, thông tin, tri thức
- 1.2. Phân loại tri thức
- 1.3. Khái niệm kỹ nghệ tri thức
- 1.4. Ứng dụng của kỹ nghệ tri thức
- 1.5. Các vấn đề cơ bản trong kỹ nghệ tri thức

CHƯƠNG 2. Một vài hệ thống dựa trên tri thức

- 2.1. Hệ chuyên gia
- 2.2. Mạng nơron
- 2.3. Tác tử thông minh
- 2.4. Hệ thống thông minh lai

CHƯƠNG 3. Học và biểu diễn tri thức

- 3.1. Các phương pháp biểu diễn tri thức cổ điển
- 3.2. Các phương pháp biểu diễn tri thức hiện đại, học biểu diễn
- 3.3. Học tri thức và phân loại
- 3.4. Học tri thức từ chuyên gia
- 3.5. Học tri thức tự động từ dữ liệu bằng các phương pháp học máy
- 3.6. Học tri thức qua các phương pháp khai phá dữ liệu
- 3.7. Học tri thức tự động từ Dữ liệu lớn
- 3.8. Kiểm tra sự phù hợp và chính xác của tri thức

CHƯƠNG 4. Suy diễn tự động dựa trên tri thức

- 4.1. Các phương pháp suy diễn
- 4.2. Suy diễn nhờ vào luật định sẵn
- 4.3. Suy diễn trong môi trường không chắc chắn
- 4.4. Suy diễn xác suất theo luật Bayes

CHƯƠNG 5. Kiến trúc lưu và tương tác với kho tri thức và dữ liệu

- 5.1. Các kiến trúc lưu tri thức và dữ liệu
- 5.2. Các kỹ thuật tương tác với kho tri thức và dữ liệu
- 5.3. Tích hợp tri thức mới
- 5.4. An toàn và bảo mật kho tri thức và dữ liệu

13. Nội dung bài tập lớn : Tùy chọn 1 trong 2 hướng sau :

- Tiểu luận nghiên cứu : tìm hiểu, đề xuất các phương án giải quyết các vấn đề liên quan đến môn học
- Triển khai ứng dụng tiếp cận các công cụ điển hình xây dựng hệ thống tri thức

14. Tài liệu tham khảo

1. S. Kendal and M. Creen. *An Introduction to Knowledge Engineering*. Springer-Verlag, 2007.
2. G. Antoniou and F. van Harmelen. *A Semantic Web Primer*, 2nd Edition. MIT Press, 2008.
3. S. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach*. 2nd edition. Prentice-Hall, 2003.

4. Trevor Hastie, Robert Tibshirani, Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2009
5. T. M. Mitchell. *Machine Learning*. McGraw-Hill, 1997.
6. Frank van Harmelen, Vladimir Lifschitz, and Bruce Porter. *Handbook of Knowledge Representation*. Elsevier, 2008.

IT6365 Công nghệ Web ngữ nghĩa

1. Tên học phần: CÔNG NGHỆ WEB NGỮ NGHĨA

2. Mã số: IT6365

3. Khối lượng: 3(2,5-1-0-6)

- Lý thuyết: 37,5 tiết
- Bài tập lớn: 15 tiết

4. Yêu cầu học phần: tự chọn

5. Đối tượng tham dự: Học viên cao học chuyên ngành Hệ thống thông tin

6. Điều kiện học phần:

7. Mục tiêu học phần: Học phần nhằm trang bị cho học viên kiến thức chuyên sâu về những công nghệ Web ngữ nghĩa, các công cụ và môi trường phát triển phần mềm để sử dụng hiệu quả những tài nguyên Web ngữ nghĩa.

8. Nội dung tóm tắt học phần: Định hướng phát triển Web ngữ nghĩa; kiến trúc tổng quan của web ngữ nghĩa; những công nghệ nền tảng của web ngữ nghĩa: RDF(S), OWL, SPARQL; khai thác dữ liệu mở.

9. Nhiệm vụ của học viên:

- Dự lớp: Theo quy định của Bộ GD&ĐT và trường ĐHBKHN.
- Bài tập: Theo quy định của giáo viên giảng dạy và trường ĐHBKHN.

10. Đánh giá kết quả: KT/BT(0.3)–T(TL, TN: 0,7)

- Điểm quá trình: Trọng số 0.3
 - Bài tập làm đầy đủ
 - Hoàn thành bài tập lớn, có báo cáo và bảo vệ
- Thi cuối kỳ (tự luận, trắc nghiệm): Trọng số 0.7

11. Tài liệu học tập:

- Sách giáo khoa chính:
- Sách tham khảo: Xem phần tài liệu tham khảo

12. Nội dung chi tiết học phần:

MỞ ĐẦU

1. Mục đích môn học
2. Nội dung môn học
3. Sách giáo khoa và tài liệu tham khảo

CHƯƠNG 1. GIỚI THIỆU WEB NGỮ NGHĨA VÀ XML

- 1.1. Định hướng Web ngữ nghĩa
- 1.2. Kiến trúc phân tầng của Web ngữ nghĩa
- 1.3. Những kiến thức cơ bản về XML

CHƯƠNG 2. MÔ TẢ TÀI NGUYÊN VỚI RDF

- 2.1. Tính cấp thiết của RDF
- 2.2. Các định dạng RDF
- 2.3. Mô tả cấu trúc với RDFS
 - 2.3.1. Tính cấp thiết của RDFS
 - 2.3.2. Cú pháp RDFS
 - 2.3.3. Ngữ nghĩa hình thức của RDF(S)
 - 2.3.4. Suy diễn RDFS dựa trên luật

CHƯƠNG 3. MÔ TẢ ONTOLOGY VỚI OWL VÀ SUY DIỄN

- 3.1. Khái niệm ontology
- 3.2. Khái quát ngôn ngữ OWL
- 3.3. Cú pháp, các phiên bản và bài toán suy diễn OWL
- 3.4. OWL và logic mô tả
- 3.5. Suy diễn OWL dựa trên các giải thuật Tableau
- 3.6. OWL và Luật
- 3.7. Giới thiệu OWL 2

CHƯƠNG 4. NGÔN NGỮ TRUY VẤN SPARQL

- 4.1. Cú pháp SPARQL
- 4.2. Ngữ nghĩa SPARQL
- 4.3. Đại số SPARQL
- 4.4. Giới thiệu SPARQL 1.1
- 4.5. Hệ suy diễn cho SPARQL

CHƯƠNG 5. DỮ LIỆU MỞ

- 5.1. Tài nguyên dữ liệu mở trên Web
- 5.2. Công bố dữ liệu mở

13. Nội dung bài tập lớn (15 tiết):

Tiểu luận nghiên cứu: Tìm hiểu, đề xuất các phương pháp giải quyết những vấn đề liên quan đến nội dung học phần

14. Tài liệu tham khảo:

1. Fensen and J. Hendler, *Spinning the Semantic Web*, 2003
2. F. Baader, D. Calvanese, D.L. McGuinness, D. Nardi, P.F. Patel-Schneider, *Description Logic Handbook*, 2003.
3. Grigoris Antoniou, Frank van Harmelen, *A Semantic Web Primer, (third edition)*, 2012.

IT6370 Xử lý ngôn ngữ tự nhiên nâng cao

1. Tên học phần: Xử lý ngôn ngữ tự nhiên nâng cao

2. Mã số: IT6370

3. Khối lượng: 2 (1.5-1-0-4)

- Lý thuyết: 22.5 giờ
- Bài tập: 15 giờ

4. Yêu cầu học phần: tự chọn

5. Đối tượng tham dự: Học viên cao học chuyên ngành Hệ thống thông tin.

6. Điều kiện học phần:

7. Mục tiêu học phần: Kết thúc học phần này, học viên có kiến thức chuyên sâu về xử lý ngôn ngữ tự nhiên trên máy tính và phương pháp xây dựng các ứng dụng dựa trên các kỹ thuật xử lý ngôn ngữ tự nhiên.

8. Nội dung tóm tắt học phần:

Các kiến thức cơ bản về ngôn ngữ. Các cách tiếp cận dựa trên luật và dựa trên thông kê trong phân tích từ và phân tích cú pháp. Một số ứng dụng tiêu biểu của xử lý ngôn ngữ: tóm tắt văn bản và dịch máy.

9. Nhiệm vụ của sinh viên:

- Dự lớp: Theo quy định của Bộ GD&ĐT và trường ĐHBKHN.
- Bài tập lớn: Theo quy định của Bộ GD&ĐT và trường ĐHBKHN.

10. Đánh giá kết quả: KT/BT(0.40)-T(TL:0.60)

Điểm quá trình: trọng số 0.40

- Hoàn thành bài tập lớn

Thi cuối kỳ (tự luận hoặc bảo vệ bài tập lớn): trọng số 0.60

11. Tài liệu học tập:

- Sách giáo khoa chính: chưa có
- Sách tham khảo: Xem phần tài liệu tham khảo.

12. Nội dung chi tiết học phần:

XỬ LÝ NGÔN NGỮ TỰ NHIÊN NÂNG CAO

Người biên soạn: PGS. Lê Thanh Hương

MỞ ĐẦU

1. Mục đích môn học
2. Nội dung môn học

3. Sách giáo khoa và tài liệu tham khảo

PHẦN I. CÁC KHÁI NIỆM CƠ BẢN

CHƯƠNG 1. TỔNG QUAN

- 1.1 Giới thiệu chung
- 1.2 Các hướng nghiên cứu trong xử lý ngôn ngữ tự nhiên
- 1.3 Các mức phân tích trong xử lý ngôn ngữ tự nhiên
- 1.4 Các ứng dụng của xử lý ngôn ngữ tự nhiên

CHƯƠNG 2. CÁC KIẾN THỨC CƠ BẢN VỀ NGÔN NGỮ

- 2.1 Từ và từ loại
- 2.2 Cụm từ
 - 2.2.1 Các loại cụm từ
 - 2.2.2 Cấu trúc cụm từ tiếng Anh
 - 2.2.3 Cấu trúc cụm từ tiếng Việt

CHƯƠNG 3. GÁN NHÃN TỪ LOẠI

- 3.1 Định nghĩa
- 3.2 Các cách tiếp cận trong phân tích từ loại
 - 3.2.1 Phương pháp Hidden Markov Model
 - 3.2.2 Phương pháp dựa trên ràng buộc ngữ pháp

CHƯƠNG 4. PHÂN TÍCH CÚ PHÁP

- 4.1 Văn phạm và cấu trúc câu
- 4.2 Phân tích cú pháp dựa trên văn phạm phi ngữ cảnh
 - 4.2.1 Giới thiệu
 - 4.2.2 Phân tích kiểu trên xuống
 - 4.2.3 Phân tích kiểu dưới lên
- 4.3 Tập ngữ liệu có chú giải ngữ pháp
- 4.4 Phân tích cú pháp dựa trên văn phạm phi ngữ cảnh xác suất
 - 4.4.1 Văn phạm phi ngữ cảnh xác suất
 - 4.4.2 Phân tích cú pháp dựa trên văn phạm phi ngữ cảnh xác suất
- 4.5 Nét và văn phạm gia tổ
 - 4.5.1 Hệ thống nét và văn phạm gia tổ
 - 4.5.2 Một số hệ thống nét cho tiếng Anh
 - 4.5.3 Phân tích cú pháp sử dụng nét

PHẦN 2. MỘT SỐ ỨNG DỤNG CỦA XỬ LÝ NGÔN NGỮ TỰ NHIÊN

CHƯƠNG 5. TÓM TẮT VĂN BẢN

- 5.1 Tóm tắt văn bản là gì?
- 5.2 Các tiêu chí đánh giá
- 5.3 Phân loại bài toán tóm tắt văn bản
- 5.4 Mô hình hệ thống tóm tắt văn bản
 - 5.4.1 Mô hình chung
 - 5.4.2 Các phương pháp áp dụng trong các pha
 - 5.4.3 Đánh giá kết quả tóm tắt
- 5.5 Giới thiệu một số hệ thống tóm tắt văn bản

CHƯƠNG 6. DỊCH MÁY

- 6.1 Các khó khăn trong dịch máy
- 6.2 Một số cách tiếp cận trong dịch máy
- 6.3 Phương pháp dựa trên luật chuyển đổi
- 6.4 Phương pháp đóng hàng văn bản
 - 6.4.1 Đóng hàng dựa trên từ
 - 6.4.2 Đóng hàng dựa trên câu

6.5 Giới thiệu một số hệ thống dịch máy

13. Nội dung tiểu luận, bài tập lớn

Phần thực hành tạo cho học viên có khả năng chủ động sáng tạo trong nghiên cứu và hiểu sâu hơn về môn học. Học viên sẽ có một bài tập lớn để thực hành các kiến thức đã học theo hai hướng:

- Viết tiểu luận so sánh các cách tiếp cận mới trong một lĩnh vực nghiên cứu cụ thể của xử lý ngôn ngữ tự nhiên.
- Viết chương trình sử dụng các kỹ thuật xử lý ngôn ngữ (có thể phát triển từ phần mềm mã nguồn mở) và đánh giá hiệu năng của chương trình đó.

14. Tài liệu tham khảo

1. Christopher Manning and Hinrich Schütze, *Foundations of Statistical Natural Language Processing*, The MIT Press, 1999.
2. Dan Jurafsky and James Martin, *Speech and Language Processing*, PrenticeHall, 2000.
3. James Allen, *Natural Language Understanding*, The Benajmins/Cummings Publishing Company Inc, 1994.
4. Barton, E., Berwick, R., & Ristad, E, *Computational Complexity and Natural Language*, The MIT Press, ISBN 0-26-02266-4. 1987.

IT6380 Logic mờ và ứng dụng

1. Tên học phần: LOGIC MỜ VÀ ỨNG DỤNG

2. Mã số: IT6380

3. Khối lượng: 2 (1.5-1-0-4)

- Lý thuyết: 22.5 giờ
- Bài tập: 15 giờ

4. Yêu cầu học phần: tự chọn

5. Đối tượng tham dự: Học viên cao học chuyên ngành Hệ thống thông tin.

6. Điều kiện học phần:

7. **Mục tiêu học phần:** Kết thúc học phần này, học viên có kiến thức về các khái niệm cơ bản về tập mờ, logic mờ, suy diễn mờ, hệ logic mờ và xây dựng hệ mờ.

8. **Nội dung tóm tắt học phần:** Cung cấp kiến thức về tập mờ, logic mờ và ứng dụng cho các học viên cao học chuyên ngành Công nghệ Thông tin, bao gồm các khái niệm, phương pháp suy diễn và khả năng ứng dụng hệ logic mờ trong thực tế.

9. Nhiệm vụ của sinh viên:

- Dự lớp: Theo quy định của Bộ GD&ĐT và trường ĐHBKHN.
- Bài tập: Theo quy định của Bộ GD&ĐT và trường ĐHBKHN.
- Thí nghiệm:.

10. Đánh giá kết quả: KT/BT(0.40)-T(BTL:0.60)

Điểm quá trình: trọng số 0.40

- Bài tập làm đầy đủ
- Hoàn thành bài tập lớn
- Kiểm tra giữa kỳ

Thi cuối kỳ (bảo vệ bài tập lớn): trọng số 0.60

11. Tài liệu học tập:

- Sách giáo khoa chính: Hồ Thuần, Đặng Thanh Hà, *Logic mờ và ứng dụng*, NXB ĐHQGHN 2007.
- Sách tham khảo: Xem phần tài liệu tham khảo.

12. Nội dung chi tiết học phần:

LOGIC MỜ VÀ ỨNG DỤNG

Người biên soạn: PGS. Trần Đình Khang

MỞ ĐẦU

1. Mục đích môn học
2. Nội dung môn học
3. Sách giáo khoa và tài liệu tham khảo

CHƯƠNG 1. TẬP MỜ VÀ BIẾN NGÔN NGỮ

- 1.1 Các khái niệm cơ sở
 - 1.1.1 Tập mờ
 - 1.1.2 Các phép toán trên tập mờ
 - 1.1.3 Chuẩn và đối chuẩn tam giác
 - 1.1.4 Nguyên lý mở rộng
- 1.2 Quan hệ mờ
 - 1.2.1 Quan hệ mờ
 - 1.2.2 Phép hợp thành mờ
- 1.3 Biến ngôn ngữ và mệnh đề mờ
 - 1.3.1 Biến ngôn ngữ
 - 1.3.2 Mệnh đề mờ
 - 1.3.3 Đại số gia tử

Bài tập: Biểu diễn tri thức sử dụng tập mờ

CHƯƠNG 2. LOGIC MỜ VÀ LẬP LUẬN XẤP XỈ

- 2.1 Logic kinh điển
 - 2.1.1 Cú pháp và ngữ nghĩa
 - 2.1.2 Chuẩn hoá
 - 2.1.3 Dẫn xuất
 - 2.1.4 Các mở rộng của logic kinh điển
- 2.2 Logic mờ
 - 2.2.1 Các phép kéo theo mờ
 - 2.2.2 Modus ponens suy rộng
- 2.3 Suy diễn mờ
 - 2.3.1 Suy diễn mờ đơn điều kiện
 - 2.3.2 Suy diễn mờ mở rộng
 - 2.3.3 Nội suy mờ

Bài tập: Suy luận xấp xỉ

CHƯƠNG 3. HỆ LOGIC MỜ VÀ ỨNG DỤNG

- 3.1 Hệ mờ
 - 3.1.1 Cơ sở luật mờ
 - 3.1.2 Tích hợp
 - 3.1.3 Mờ hoá và khử mờ
- 3.2 Xây dựng hệ mờ
- 3.3 Điều khiển mờ
- 3.4 Một số ứng dụng

Bài tập: Hệ mờ ứng dụng

13. Tài liệu tham khảo

1. H.J. Zimmermann, *Fuzzy set theory and its applications*, Kluwer Academic Publishers, 1991.
2. James J. Buckley, Esfandiar Eslami, *An Introduction to fuzzy logic and fuzzy sets*, Physica-Verlag Heidelberg New York, 2002
3. Hồ Thuần, Đặng Thanh Hà, *Logic mờ và ứng dụng*, NXB ĐH QGHN, 2007.

IT6385 Lưu trữ và phân tích dữ liệu lớn

1. Tên học phần: LƯU TRỮ VÀ PHÂN TÍCH DỮ LIỆU LỚN

2. Mã số: IT6385

3. Khối lượng: 3(2.5-1-0-6)

- Lý thuyết: 37.5 tiết
- Bài tập: 15 tiết

4. Yêu cầu học phần: tự chọn

5. Đối tượng tham dự: Học viên thạc sĩ khoa học chuyên ngành Hệ Thống Thông tin.

6. Điều kiện học phần:

7. Mục tiêu học phần: Kết thúc học phần này, học viên có kiến thức về dữ liệu lớn, các công nghệ lưu trữ dữ liệu lớn – phân tán và không ràng buộc; cơ sở dữ liệu NoSQL; các phương pháp trong phân tích và xử lý các dữ liệu lớn đã được thu thập và lưu trữ. Mỗi liên hệ giữa dữ liệu lớn và điện toán đám mây. Kiến thức thu nhận được sau học phần này cho phép học viên có thể tham gia vào các dự án công nghiệp và nghiên cứu liên quan đến sử dụng và khai thác dữ liệu lớn.

8. Nội dung tóm tắt học phần: Các chủ đề liên quan đến lưu trữ dữ liệu lớn bao gồm: khái niệm, công nghệ và mô hình lưu trữ dữ liệu lớn (HDFS), cơ sở dữ liệu NoSQL, các dạng lưu trữ NoSQL, giới thiệu về Map Reduce/ Hadoop. Các chủ đề liên quan đến phân tích dữ liệu lớn: khối dữ liệu, phân tích đa chiều, khai phá mẫu và kết hợp, khai phá hàng đợi và chuỗi thời gian, phân tích phân cụm, các thuật toán đồ thị. Mỗi liên hệ giữa dữ liệu lớn và điện toán đám mây cũng được trình bày một cách cụ thể.

9. Nhiệm vụ của sinh viên:

- Dự lớp: Theo quy định của Bộ GD&ĐT và trường ĐHBKHN
- Bài tập: Theo quy định của Bộ GD&ĐT và trường ĐHBKHN

10. Đánh giá kết quả: KT/BT(0.40)-T(TL:0.60)

- Điểm quá trình: trọng số 0.4
- Bài tập làm đầy đủ
- Thi cuối kỳ (tự luận): trọng số 0.6

11. Tài liệu học tập:

- Sách giáo khoa chính: Tài liệu bài giảng phát trên lớp và các tài liệu, bài báo nghiên cứu giáo viên yêu cầu.
- Sách tham khảo: Xem phần tài liệu tham khảo.

12. Nội dung chi tiết học phần:

LƯU TRỮ VÀ PHÂN TÍCH DỮ LIỆU LỚN

PHẦN MỞ ĐẦU

1. Mục đích môn học
2. Nội dung môn học
3. Sách giáo khoa và tài liệu tham khảo
4. Khái niệm về dữ liệu lớn
5. Các tính chất của dữ liệu lớn: khối lượng (volume), đa dạng (variety), tốc độ (velocity)
6. Các nguồn tạo ra và các ứng dụng trên dữ liệu lớn
7. Các công nghệ lưu trữ và phương pháp phân tích dữ liệu lớn

PHẦN 1: LƯU TRỮ DỮ LIỆU LỚN

CHƯƠNG 1: Lưu trữ dữ liệu lớn trên hệ thống tệp phân tán

1. Tổng quan về dữ liệu lớn trên hệ thống tệp phân tán
2. Các tính chất của lưu trữ dữ liệu lớn trên hệ thống tệp phân tán
3. Công nghệ lưu trữ Hadoop Distributed File System (HDFS)

CHƯƠNG 2: Cơ sở dữ liệu NoSQL

1. Giới thiệu tổng quan về NoSQL
2. Phân loại và so sánh các cơ sở dữ liệu NoSQL
3. Các hạn chế của SQL

CHƯƠNG 3: Khái niệm cơ bản, công nghệ và mô hình mẫu của NoSQL

1. Tính thống nhất (consistency)
2. Phân vùng (partitioning)
3. Cấu trúc lưu trữ (storage layout)
4. Các mô hình truy vấn (query models)

CHƯƠNG 4: Các dạng cơ sở dữ liệu NoSQL và ứng dụng

1. Khóa-/giá trị lưu trữ (Key-/Value-Store), giới thiệu Amazon Dynamo
2. Cơ sở dữ liệu văn bản (document databases), giới thiệu Apache CouchDB
3. Cơ sở dữ liệu hướng đối tượng cột (Column-Oriented databases), giới thiệu Cassandra.

CHƯƠNG 5: Dữ liệu lớn và điện toán đám mây

1. Cơ sở về điện toán đám mây (cloud fundamentals)
2. Lưu trữ đám mây (cloud storage)
3. Cơ sở dữ liệu trong đám mây (database in cloud)
4. Một số mô hình của dữ liệu lớn và điện toán đám mây (case studies)

PHẦN 2: PHÂN TÍCH DỮ LIỆU LỚN

CHƯƠNG 1: Các công nghệ trong phân tích dữ liệu lớn

1. Kỹ thuật xử lý dữ liệu phân tán thông qua Map/Reduce
2. Hadoop - môi trường song song thực thi map-reduce dựa trên Java
3. Giới thiệu Pig and Hive

CHƯƠNG 2: Một số bài toán phân tích dữ liệu lớn

1. Khối dữ liệu, phân tích đa chiều (data cube, multidimensional analysis)
2. Khai phá mẫu và kết hợp (frequent pattern and association mining)
3. Khai phá hàng đợi và chuỗi thời gian (sequence and time series mining)
4. Phân tích phân cụm (clustering)
5. Các thuật toán đồ thị (graph algorithms)

13. Nội dung các bài thí nghiệm (thực hành, tiểu luận, bài tập lớn)

- Phần lưu trữ dữ liệu lớn, học viên xây dựng các ứng dụng với Hadoop/Map Reduce.
- Phần phân tích dữ liệu lớn, học viên làm bài tập liên quan tới các phương pháp phân tích dữ liệu.

14. Tài liệu tham khảo

1. Nathan Marz (Twitter) and James Warren. 2012. Big Data: Principles and best practices of scalable realtime data systems. Manning Publications Co., Greenwich, CT, USA.
2. J. Lin and C. Dyer: Data-Intensive Text Processing with MapReduce, Morgan & Claypool Publishers, 2010
3. Chuck Lam. 2010. Hadoop in Action (1st ed.). Manning Publications Co., Greenwich, CT, USA.
4. Gates, Alan. Programming Pig. O'Reilly, 2011.
5. Capriolo, Edward, Dean Wampler, and Jason Rutherglen. Programming Hive. O'Reilly, 2012.
6. Ron Bekkerman. 2011. Scaling up Machine Learning Parallel and Distributed Approaches
7. J. Han, M. Kamber, and J. Pei: Data Mining: Concepts and Techniques (3rd Ed), Morgan Kaufmann, 2012
8. Jothy Rosenberg and Arthur Mateos. 2011. The Cloud at Your Service. Manning Publications Co., Greenwich, CT, USA.

9.2 Danh mục học phần trình độ Tiến sĩ

Số TT	MÃ SỐ	TÊN HỌC PHẦN	TÊN TIẾNG ANH	KHỐI LƯỢNG	Khoa/Viện Bộ môn	Đánh giá
1	IT7310	Kỹ nghệ dữ liệu và tri thức	Knowledge and Data Engineering	3(2-2-0-6)	Bm HTTT	KT0,4-T0,6
2	IT7331	Tích hợp dữ liệu	Data Integration	3(2-2-0-6)	Bm HTTT	KT0,4-T0,6
3	IT7341	Tìm kiếm thông tin	Information Retrieval	3(2-2-0-6)	Bm HTTT	KT0,4-T0,6
4	IT7351	Khai phá dữ liệu và phát hiện tri thức	Data Mining and Knowledge Discovery	3(2-2-0-6)	Bm HTTT	KT0,4-T0,6
5	IT7361	Khai phá dữ liệu văn bản và dữ liệu Web	Text and Web Mining	3(2-2-0-6)	Bm HTTT	KT0,4-T0,6
6	IT7371	Các tiếp cận logic trong biểu diễn và xử lý thông tin	Logical approaches in information representation and procesing	3(2-2-0-6)	Bm HTTT	KT0,4-T0,6
7	IT7381	Các chủ đề nâng cao trong xử lý dữ liệu lớn	Advanced topics in big data procesing	3(2-2-0-6)	Bm HTTT	KT0,4-T0,6

8	IT7391	Các chủ đề nâng cao đối với phân tích dữ liệu lớn	Advanced topics for big data analytics	3(2-2-0-6)	Bm HTTT	KT0,4-T0,6
---	--------	---	--	------------	---------	------------

10 Đề cương chi tiết các học phần trình độ Tiến sĩ

IT7310 Kỹ nghệ dữ liệu và tri thức

Knowledge and Data Engineering

1. Tên học phần: Kỹ nghệ dữ liệu và tri thức

2. Mã học phần: IT7310

3. Tên tiếng Anh: Knowledge and Data Engineering

4. Khối lượng: 3 (2-2-0-6)

- Lý thuyết: 30 tiết

- Bài tập: 30 tiết

- Thí nghiệm:

5. Đối tượng tham dự: NCS thuộc chuyên ngành Hệ thống thông tin

6. Mục tiêu của học phần: Học phần này nhằm mang lại cho NCS:

- Các kiến thức nâng cao về lý luận chuyên ngành Hệ thống thông tin

- Rèn luyện khả năng tư duy

- Rèn luyện kỹ năng thử nghiệm của chuyên ngành Hệ thống thông tin

- Có khả năng lựa chọn các kỹ nghệ thích hợp cho phép thu thập và quản trị dữ liệu và tri thức đối với một Hệ thống thông tin cụ thể

7. Nội dung tóm tắt:

Học phần này cung cấp các kỹ nghệ dữ liệu bao gồm: các kỹ thuật thu thập, biểu diễn, lưu trữ và truy vấn dữ liệu và các kỹ nghệ tri thức bao gồm: biểu diễn tri thức, lập luận không chắc chắn và phát hiện tri thức sử dụng học máy và khai phá dữ liệu.

8. Nhiệm vụ của NCS:

- Dự lớp: đầy đủ theo quy chế

- Bài tập: hoàn thành các bài tập của học phần

- Thí nghiệm:

9. Đánh giá kết quả:

- Mức độ dự giờ giảng:

- Kiểm tra định kỳ: trọng số 0.4

- Thi kết thúc học phần: tự luận, trọng số 0.6

10. Nội dung chi tiết học phần:

PHẦN MỞ ĐẦU

Giới thiệu môn học

Giới thiệu đề cương môn học

Giới thiệu tài liệu tham khảo

Phần 1. Kỹ nghệ dữ liệu (DE)

Chương 1: Giới thiệu

1. Các khái niệm cơ bản về DE

2. Các hoạt động chính của DE

3. Các hệ CSDL

4. Các ứng dụng nâng cao

Chương 2: Biểu diễn dữ liệu

1. Các kiểu dữ liệu

2. Mô hình dữ liệu quan hệ

3. Mô hình dữ liệu hướng đối tượng

4. Mô hình dữ liệu đa chiều

5. Mô hình dữ liệu XML

6. Thiết kế và chuẩn hóa dữ liệu

Chương 3: Lưu trữ dữ liệu

1. Tổ chức lưu trữ dữ liệu quan hệ
2. Tổ chức lưu trữ dữ liệu hướng đối tượng
3. Tổ chức lưu trữ dữ liệu OLAP
4. Lưu trữ và đánh chỉ số dữ liệu XML

Chương 4: Truy vấn dữ liệu

1. Ngôn ngữ truy vấn đối với CSDL quan hệ
2. Ngôn ngữ truy vấn đối với CSDL hướng đối tượng
3. Các thao tác OLAP trong mô hình dữ liệu đa chiều
4. Ngôn ngữ truy vấn dữ liệu bán cấu trúc
5. Các kỹ thuật xử lý và tối ưu hóa truy vấn

Phần 2. Kỹ nghệ tri thức (KE)

Chương 1: Giới thiệu

1. Các khái niệm cơ bản về KE
2. Các hoạt động chính của KE
3. Các hệ thống dựa trên tri thức (KBSs)

Chương 2: Biểu diễn tri thức

1. Biểu diễn dựa trên luật
2. Biểu diễn dựa trên frame
3. Biểu diễn dựa trên mạng ngữ nghĩa

Chương 3: Lập luận không chắc chắn

1. Lý thuyết xác suất
2. Lập luận xác suất
3. Các mạng Bayes
4. Logic mờ

Chương 4: Phát hiện tri thức sử dụng Học máy và Khai phá dữ liệu

1. Học cây quyết định
2. Quy nạp luật
3. Học dựa trên mẫu
4. Phát hiện luật kết hợp

11. Tài liệu học tập:

12. Tài liệu tham khảo:

1. S. Kendal and M. Creen. *An Introduction to Knowledge Engineering*. Springer-Verlag, 2007.
2. S. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach. 2nd edition*. Prentice-Hall, 2003.
3. T. M. Mitchell. *Machine Learning*. McGraw-Hill, 1997.
4. S. Chakrabarti. *Mining the Web: Discovering Knowledge from Hypertext Data*. Morgan Kaufmann Publishers, 2003.
5. P.N. Tan, M. Steinbach, and V. Kumar. *Introduction to Data Mining*. Addison-Wesley, 2005.
6. J. Han and M. Kamber. *Data Mining: Concepts and Techniques, 2nd Edition*. Morgan Kaufmann, 2006.
7. H. Witten and E. Frank. *Data Mining: Practical Machine Learning Tools and Techniques, 2nd Edition*. Morgan Kaufmann, 2005.

IT7331 Tích hợp dữ liệu

Data Integration

1. Tên học phần: Tích hợp dữ liệu

2. Mã học phần: IT7331

3. Tên tiếng Anh: Data Integration

4. Khối lượng: 3 (2-2-0-6)

- Lý thuyết: 30 tiết

- Bài tập: 30 tiết

- Thí nghiệm:

5. Đối tượng tham dự: NCS thuộc chuyên ngành Hệ thống thông tin

6. Mục tiêu của học phần: Học phần này nhằm mang lại cho NCS:

- Các kiến thức nâng cao về các phương pháp và các thách thức trong tích hợp dữ liệu

- Rèn luyện khả năng tư duy

- Rèn luyện kỹ năng thử nghiệm xây dựng Hệ tích hợp dữ liệu.

7. Nội dung tóm tắt:

Học phần này cung cấp các cách tiếp cận để xây dựng một hệ tích hợp dữ liệu: các kỹ thuật đối sánh lược đồ, đối sánh dữ liệu, xây dựng lược đồ tích hợp, xử lý yêu cầu truy vấn trong hệ tích hợp dữ liệu, các kỹ thuật nâng cao về làm sạch dữ liệu và trả lời truy vấn nhất quán.

8. Nhiệm vụ của NCS:

- Dự lớp: đầy đủ theo quy chế

- Bài tập: hoàn thành các bài tập của học phần

- Thí nghiệm:

9. Đánh giá kết quả:

- Mức độ dự giờ giảng:

- Kiểm tra định kỳ: trọng số 0.4

- Thi kết thúc học phần: tự luận, trọng số 0.6

10. Nội dung chi tiết học phần:

PHẦN MỞ ĐẦU

Giới thiệu môn học

Giới thiệu đề cương môn học

Giới thiệu tài liệu tham khảo

Chương 1: Giới thiệu chung

1. Vài nét về lịch sử

2. Các cách tiếp cận tích hợp dữ liệu

3. Các vấn đề liên quan đến tích hợp dữ liệu

3.1 Tích hợp lược đồ

3.2 Tích hợp dữ liệu

3.3 Viết lại truy vấn

Chương 2: Tích hợp lược đồ & dữ liệu, xử lý truy vấn trong hệ tích hợp dữ liệu

1. Tích hợp lược đồ

1.1. Các cách tiếp cận đối sánh lược đồ

1.2. Đánh giá đối sánh lược đồ

- 1.3. Các ứng dụng của đối sánh lược đồ
- 1.4. Các vấn đề còn mở
2. Tích hợp dữ liệu
 - 2.1. Tiếp cận GENMAPPER
 - 2.2. Tiếp cận tích hợp lai
3. Xử lý truy vấn trong hệ tích hợp dữ liệu
 - 3.1 Mô hình chung
 - 3.2 Viết lại truy vấn
 - 3.3 Tối ưu hóa xử lý truy vấn trong một hệ tích hợp dữ liệu

Chương 3: Làm sạch dữ liệu và trả lời truy vấn nhất quán

1. Các dị thường dữ liệu và chất lượng dữ liệu
2. Các phương pháp được sử dụng đối với làm sạch dữ liệu
3. Các cách tiếp cận đang tồn tại đối với làm sạch dữ liệu
4. Trả lời truy vấn nhất quán đối với các hệ tích hợp dữ liệu
5. Các thách thức và các vấn đề còn mở

11. Tài liệu học tập:

12. Tài liệu tham khảo:

1. Do Hong Hai, Schema Matching and Mapping-based Data Integration, VDM Verlag Dr. Muller, 2006.
2. Maurizio Lenzerini (2002). "Data Integration: A Theoretical Perspective". *PODS 2002*: 233-246.
3. Patrick Ziegler and Klaus R. Dittrich (2004). "Three Decades of Data Integration - All Problems Solved?". *WCC 2004*: 3-12.
4. C. Koch (2001). "*Data Integration against Multiple Evolving Autonomous Schemata*".
5. Jeffrey D. Ullman (1997). "Information Integration Using Logical Views". *ICDT 1997*: 19-40.
6. A.Y. Halevy (2001). "Answering queries using views: A survey". *The VLDB Journal*: 270-294.
7. Alon Y. Halevy, Naveen Ashish, Dina Bitton, Michael J. Carey, Denise Draper, Jeff Pollock, Arnon Rosenthal, Vishal Sikka (2005). "Enterprise information integration: successes, challenges and controversies". *SIGMOD 2005*: 778-787.

IT7341 Tìm kiếm thông tin
Information Retrieval

1. **Tên học phần:** Tìm kiếm thông tin
2. **Mã học phần:** IT7341
3. **Tên tiếng Anh:** Information Retrieval
4. **Khối lượng:** 3(2-2-0-6)
 - Lý thuyết: 30 tiết
 - Bài tập: 30 tiết
 - Thí nghiệm:
5. **Đối tượng tham dự:** Tất cả NCS thuộc chuyên ngành Hệ thống thông tin
6. **Mục tiêu của học phần:** Học phần này nhằm mang lại cho NCS:
 - các kiến thức cơ bản và nâng cao trong tìm kiếm thông tin
 - các chủ đề nghiên cứu liên quan đến tìm kiếm thông tin
7. **Nội dung tóm tắt:**
Biểu diễn và lưu trữ; Mô hình tìm kiếm; Chỉ mục; Phân lớp, phân cụm; Các kỹ thuật tìm kiếm; Đánh giá hệ thống tìm kiếm thông tin, máy tìm kiếm trên web; tìm kiếm đa phương tiện; các xu hướng phát triển
8. **Nhiệm vụ của NCS:**
 - Dự lớp: theo quy định
 - Bài tập: hoàn thành đầy đủ bài tập
 - Thí nghiệm:
9. **Đánh giá kết quả:**
 - Mức độ dự giờ giảng: 0.05
 - Kiểm tra định kỳ: 0.25
 - Thi kết thúc học phần: 0.7
10. **Nội dung chi tiết học phần:**

PHẦN MỞ ĐẦU

- Giới thiệu môn học
- Giới thiệu đề cương môn học
- Giới thiệu tài liệu tham khảo

Chương 1: Giới thiệu chung

1. Tổng quan về tìm kiếm thông tin
2. Các khái niệm cơ bản
3. Quy trình tìm kiếm thông tin

Chương 2: Các mô hình tìm kiếm thông tin và chỉ mục

1. Mô hình Boolean
2. Mô hình không gian vec-tơ
3. Mô hình xác suất
4. Từ điển và posting list

Chương 3: Phân loại và phân cụm

1. Naive Bayes, Bayesian Networks
2. Nearest Neighbour and SVM
3. Decision tree learning
4. Ensemble learning
5. kMeans & EM

6. DBSCAN

7. Hierarchical clustering

Chương 4: Đánh giá hiệu quả tìm kiếm

1. Các thông số cơ bản đánh giá kết quả tìm kiếm

Precision, Re-call, Fall-out, ...

2. Thử nghiệm và đánh giá hệ tìm kiếm thông tin với các bộ dữ liệu

2.1 TREC

2.2 CACM & CISI

2.3 Cystic Fibrosis

Chương 5: Kỹ thuật tìm kiếm và lọc

1. Phản hồi của người dùng về kết quả tìm kiếm

2. Hồ sơ người dùng

3. Lọc cộng tác

Chương 6: Mô-tơ tìm kiếm Web (search engine)

1. Giới thiệu chung về tìm kiếm web

2. Thu thập và đánh chỉ mục

3. Phân tích liên kết

Web spam, SEO ; Reference models ; Google's Pagerank ; Hub and authorities

Chương 7: Các xu hướng phát triển

1. Tìm kiếm thông tin đa phương tiện

2. Tìm kiếm đa ngôn ngữ

3. Tìm kiếm thông tin hướng người dùng

4. Các ứng dụng tiêu biểu

5. Một số vấn đề mở

11. Tài liệu học tập:

12. Tài liệu tham khảo:

[1] Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, *Introduction to Information Retrieval*, Cambridge University Press. 2008.

[2] Ricardo Baeza-Yates, Berthier Ribeiro-Neto, *Modern Information Retrieval*. Addison Wesley. 2004

và các bài báo

IT7351 Khai phá dữ liệu và phát hiện tri thức Data Mining and Knowledge Discovery

- 1. Tên học phần:** Khai phá dữ liệu và phát hiện tri thức
- 2. Mã học phần:** IT7351
- 3. Tên tiếng Anh:** Data Mining and Knowledge Discovery
- 4. Khối lượng:** 3 (2-2-0-6)
 - Lý thuyết: 30 tiết
 - Bài tập: 30 tiết
 - Thí nghiệm:

5. Đối tượng tham dự: NCS thuộc chuyên ngành Hệ thống thông tin

6. Mục tiêu của học phần: Học phần này giới thiệu tổng quan về khai phá dữ liệu và phát hiện tri thức, trình bày tóm tắt các kỹ thuật và các phương pháp cơ bản trong khai phá dữ liệu và phát hiện tri thức. Ngoài ra, học phần sẽ trình bày một số vấn đề và phương pháp nâng cao trong khai phá dữ liệu và phát hiện tri thức. Trong học phần này, nghiên cứu sinh cũng sẽ được yêu cầu nghiên cứu và nắm được kiến thức thông qua việc tự đọc các bài báo chọn lọc và trình bày hiểu biết thu được.

7. Nội dung tóm tắt:

Giới thiệu tổng quan về khai phá dữ liệu và phát hiện tri thức. Các kỹ thuật cơ bản và một số vấn đề nâng cao trong tiền xử lý dữ liệu. Các phương pháp cơ bản và một số vấn đề nâng cao trong phát hiện luật kết hợp. Các phương pháp cơ bản và một số vấn đề nâng cao trong phân lớp và dự đoán. Các phương pháp cơ bản và một số vấn đề nâng cao trong phân nhóm.

8. Nhiệm vụ của NCS:

- Dự lớp: đầy đủ theo quy chế
- Bài tập: hoàn thành các bài tập của học phần
- Thí nghiệm:

9. Đánh giá kết quả:

- Mức độ dự giờ giảng:
- Kiểm tra định kỳ: trọng số 0.3
- Thi kết thúc học phần: tự luận, trọng số 0.7

10. Nội dung chi tiết học phần:

PHẦN MỞ ĐẦU

Giới thiệu học phần

Giới thiệu đề cương học phần

Giới thiệu tài liệu tham khảo

CHƯƠNG 1: Tổng quan về khai phá dữ liệu và phát hiện tri thức

- 1.1. Giới thiệu về khai phá dữ liệu và phát hiện tri thức
- 1.2. Các bước chính của quá trình khai phá dữ liệu và phát hiện tri thức
- 1.3. Các ứng dụng tiêu biểu
- 1.4. Các vấn đề thách thức
- 1.5. Các công cụ tiện ích

CHƯƠNG 2: Tiền xử lý dữ liệu

- 2.1. Giới thiệu về tiền xử lý dữ liệu
- 2.2. Các kỹ thuật cơ bản trong tiền xử lý dữ liệu
- 2.3. Chọn lọc một số vấn đề và kỹ thuật nâng cao trong tiền xử lý dữ liệu

CHƯƠNG 3: Phát hiện các luật kết hợp

- 3.1. Giới thiệu về phát hiện luật kết hợp
- 3.2. Các phương pháp cơ bản trong phát hiện luật kết hợp
- 3.3. Chọn lọc một số vấn đề và phương pháp nâng cao trong phát hiện luật kết hợp

CHƯƠNG 4: Phân lớp và dự đoán

- 4.1. Giới thiệu về phân lớp và dự đoán
- 4.2. Các phương pháp cơ bản trong phân lớp và dự đoán
- 4.3. Chọn lọc một số vấn đề và phương pháp nâng cao trong phân lớp và dự đoán

CHƯƠNG 5: Phân nhóm

- 5.1. Giới thiệu về phân nhóm
- 5.2. Các phương pháp cơ bản trong phân nhóm
- 5.3. Chọn lọc một số vấn đề và phương pháp nâng cao trong phân nhóm

11. Tài liệu học tập:

12. Tài liệu tham khảo:

- [1] Pang-Ning Tan, Michael Steinbach, and Vipin Kumar. *Introduction to Data Mining*. Addison-Wesley, 2005.
- [2] Jiawei Han and Micheline Kamber. *Data Mining: Concepts and Techniques*, 2nd Edition. Morgan Kaufmann, 2006.
- [3] *Các bài báo chọn lọc liên quan đến các vấn đề nâng cao.*

IT7361 Khai phá dữ liệu văn bản và dữ liệu web Text and Web Mining

- 1. Tên học phần:** Khai phá dữ liệu văn bản và dữ liệu web
- 2. Mã học phần:** IT7361
- 3. Tên tiếng Anh:** Text and Web Mining
- 4. Khối lượng:** 3 (2-2-0-6)
 - Lý thuyết: 30 tiết
 - Bài tập: 30 tiết
 - Thí nghiệm:
- 5. Đối tượng tham dự:** NCS thuộc chuyên ngành Hệ thống thông tin
- 6. Mục tiêu của học phần:** Học phần này nhằm mang lại cho NCS:
 - Các kiến thức nâng cao về lý luận chuyên ngành Hệ thống thông tin
 - Rèn luyện khả năng tư duy
 - Rèn luyện kỹ năng thí nghiệm của chuyên ngành Hệ thống thông tin
 - Có khả năng lựa chọn phương pháp thích hợp để cài đặt ứng dụng cho phép xử lý dữ liệu web hoặc văn bản cho một ứng dụng cụ thể
- 7. Nội dung tóm tắt:**

Học phần này cung cấp các kỹ thuật và các công cụ sử dụng trong khai phá dữ liệu văn bản và dữ liệu web. Trên cơ sở đó, NCS có phát triển các ứng dụng như máy tìm kiếm, thương mại điện tử, các thư viện số, các hệ thống quản lý tri thức, v.v.
- 8. Nhiệm vụ của NCS:**
 - Dự lớp: đầy đủ theo quy chế
 - Bài tập: hoàn thành các bài tập của học phần
 - Thí nghiệm:
- 9. Đánh giá kết quả:**
 - Mức độ dự giờ giảng:
 - Kiểm tra định kỳ: trọng số 0.4
 - Thi kết thúc học phần: tự luận, trọng số 0.6
- 10. Nội dung chi tiết học phần:**

PHẦN MỞ ĐẦU

Giới thiệu môn học

Giới thiệu đề cương môn học

Giới thiệu tài liệu tham khảo

CHƯƠNG 1: Khai phá văn bản

1.1. Giới thiệu

1.2. Các thao tác khai phá văn bản

1.2.1. Các thao tác cơ bản

1.2.2. Sử dụng các tri thức nền tảng trong khai phá văn bản

1.2.3. Ngôn ngữ truy vấn trong khai phá văn bản

1.3. Các kỹ thuật tiền xử lý khai phá văn bản

1.4. Các công cụ trong khai phá văn bản

1.5. Các ứng dụng của khai phá văn bản

CHƯƠNG 2: Khai phá dữ liệu web

2.1. Khai phá nội dung trang web

2.1.1. Đánh chỉ số trang và tìm kiếm trong môi trường web

2.1.2. Xếp hạng kết quả tìm kiếm

2.1.3. Phân loại và gộp nhóm tài liệu web

2.2. Khai phá cấu trúc web

2.2.1. Phân tích liên kết web

2.2.2. Xếp hạng trang web

2.3. Khai phá việc sử dụng web

2.3.1. Ứng xử của người dùng

2.3.2. Quảng cáo trên mạng

2.4. Trích rút thông tin

2.6. Hệ tư vấn

11. Tài liệu học tập:

12. Tài liệu tham khảo:

[4] Sullivan, Dan. (2001) *Document Warehousing and Text Mining*: Wiley.

[5] Brin, Sergey, and Lawrence Page. (1998) *The Anatomy of a Large-Scale Hypertextual Web Search Engine*. Computer Networks and ISDN Systems 30 : 107-117.

[6] Croft, W. Bruce. (1995) *What Do People Want from Information Retrieval?* D-Lib Magazine November.

[7] Ronen Feldman, James Sanger. (2006) *The text mining handbook: advanced approaches in analyzing unstructured data*. Cambridge University Press.

IT7371 Các tiếp cận logic trong biểu diễn và xử lý thông tin

Logical approaches in information representation and processing

- 1. Tên học phần:** Các tiếp cận logic trong biểu diễn và xử lý thông tin
2. Mã học phần: IT7371
- 3. Tên tiếng Anh:** Logical approaches in information representation and processing
4. Khối lượng: 3(2-2-0-6)
 - Lý thuyết: 30 tiết
 - Bài tập: 30 tiết (BTL)
 - Thí nghiệm:
- 5. Đối tượng tham dự:** NCS thuộc chuyên ngành Hệ thống thông tin
- 6. Mục tiêu của học phần:** Học phần này nhằm mang lại cho NCS:
 - Các kiến thức nâng cao về lý luận chuyên ngành Hệ thống thông tin
 - Rèn luyện khả năng tư duy
 - Rèn luyện kỹ năng thí nghiệm của chuyên ngành Hệ thống thông tin
 - Có khả năng lựa chọn các tiếp cận logic thích hợp trong biểu diễn và xử lý thông tin
- 7. Nội dung tóm tắt:** Trình bày tổng quan về logic tính toán, với các logic kinh điển và logic bậc cao; phân loại các ngôn ngữ khái niệm theo cú pháp và cấu trúc cùng với các thủ tục tính toán. Đưa ra các biểu diễn ngữ nghĩa cho các nhân ngôn ngữ và các phép toán xử lý.
- 8. Nhiệm vụ của NCS:**
 - Dự lớp: đầy đủ, theo quy chế
 - Bài tập: hoàn thành các bài tập
 - Thí nghiệm:
- 9. Đánh giá kết quả:** (cách cho điểm giống như quy định đối với Cao học)
 - Mức độ dự giờ giảng:
 - Kiểm tra định kỳ: trọng số 0.4
 - Thi kết thúc học phần: trọng số 0.6
10. Nội dung chi tiết học phần:

PHẦN MỞ ĐẦU

Giới thiệu môn học

Giới thiệu đề cương môn học

Giới thiệu tài liệu tham khảo

CHƯƠNG 1: Tổng quan về logic tính toán

1.1. Logic mệnh đề

1.2. Logic vị từ bậc nhất

1.3. Logic bậc cao

CHƯƠNG 2: Các tiếp cận logic trong biểu diễn và xử lý thông tin

2.1. Các ngôn ngữ khái niệm

2.2. Suy diễn không đơn điệu

CHƯƠNG 3: Biểu diễn và xử lý thông tin dạng ngôn ngữ

3.1. Tính toán với từ

3.2. Đại số gia tử

3.3. Suy luận ngôn ngữ

CHƯƠNG 4: Các vấn đề khác

4.1. Ứng dụng của logic trong điều khiển học

4.2. Ứng dụng logic trong xử lý ngôn ngữ tự nhiên

- 11. Tài liệu học tập:** (danh mục các giáo trình, nếu không có thì bỏ trống)

12. Tài liệu tham khảo:

- [1] Franz Baader, Diago Calvanese, Deborah McGuinness, Daniele Nardi, Peter Patel-Schneider, The Description Logic Handbook, Theory, Implementation and Applications, Cambridge University Press 2003
- [2] Rene Cori, Daniel Lascar, Mathematical Logic, Oxford University Press, 2001
- [3] LOTFI A. ZADEH, FROM COMPUTING WITH NUMBERS TO COMPUTING WITH WORDS - FROM MANIPULATION OF MEASUREMENTS TO MANIPULATION OF PERCEPTIONS, Int. J. Appl. Math. Comput. Sci., 2002, Vol.12, No.3, 307–324
- [4] Jerry M. Mendel, Computing with words and its relationships with fuzzistics, Information Sciences: an International Journal, Volume 177 Issue 4, February, 2007
- [5] Cat-Ho Nguyen, Dinh Khang Tran, Van Nam Huynh, Hai Chau Nguyen, Hedge Algebras, Linguistic-Valued Logic and Their Application to Fuzzy Reasoning. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 7(4): 347-361 (1999)

IT7381 Các chủ đề nâng cao trong xử lý dữ liệu lớn

Advanced topics in big data processing

1. Tên học phần: Các chủ đề nâng cao trong xử lý dữ liệu lớn

2. Mã học phần:

3. Tên tiếng Anh: Advanced topics in big data processing

4. Khối lượng: 3(2-2-0-6)

- Lý thuyết: 30 tiết

- Bài tập: 30 tiết (BTL)

- Thí nghiệm:

5. Đối tượng tham dự: NCS thuộc chuyên ngành Hệ thống thông tin

6. Mục tiêu của học phần: Học phần này nhằm mang lại cho NCS:

- Các kiến thức nâng cao về lý luận chuyên ngành Hệ thống thông tin

- Rèn luyện khả năng tư duy

- Rèn luyện kỹ năng thí nghiệm của chuyên ngành Hệ thống thông tin

- Có khả năng lựa chọn các tiếp cận thích hợp trong lưu trữ và xử lý dữ liệu lớn.

7. Nội dung tóm tắt: Học phần này sẽ trình bày các kỹ thuật và hệ thống cho việc xử lý dữ liệu lớn theo hướng nâng cao. Các chủ đề liên quan tới nền tảng tính toán cho dữ liệu lớn sẽ bao gồm lưu trữ dữ liệu lớn, các hệ thống cho phép xử lý dữ liệu lớn. Các chủ đề liên quan tới xử lý dữ liệu lớn bao gồm truy hồi thông tin (information retrieval) với MapReduce, xử lý đồ thị (graph processing) với MapReduce, quản lý dữ liệu với MapReduce, khai phá dữ liệu với MapReduce và một số mô hình toán học thiết kế thuật toán trên dữ liệu lớn. Mục tiêu của học phần này là giúp học viên có được các kiến thức nâng cao cùng với những phân tích sâu về các chủ đề nói trên.

8. Nhiệm vụ của NCS:

- Dự lớp: đầy đủ, theo quy chế

- Bài tập: hoàn thành các bài tập

- Thí nghiệm:

9. Đánh giá kết quả: (cách cho điểm giống như quy định đối với Cao học)

- Mức độ dự giờ giảng:

- Kiểm tra định kỳ: trọng số 0.4

- Thi kết thúc học phần: trọng số 0.6

10. Nội dung chi tiết học phần:

PHẦN MỞ ĐẦU

Giới thiệu môn học

Giới thiệu đề cương môn học

Giới thiệu tài liệu tham khảo

CHƯƠNG 1: Tổng quan dữ liệu lớn và lưu trữ dữ liệu lớn

1.1. Các khái niệm cơ bản về dữ liệu lớn và xử lý dữ liệu lớn

1.2. Khái niệm, công nghệ và mô hình NoSQL

1.3. Các hệ cơ sở dữ liệu NoSQL cho dữ liệu lớn và một số ứng dụng

CHƯƠNG 2: Các hệ thống lưu trữ và xử lý dữ liệu lớn

2.1. Điện toán đám mây

2.2. Các hệ thống lưu trữ phân tán

2.3. Hadoop/MapReduce

CHƯƠNG 3: Xử lý dữ liệu lớn trên MapReduce

3.1. Truy hồi thông tin

3.2. Xử lý đồ thị

3.3. Quản lý dữ liệu

3.4. Khai phá dữ liệu

CHƯƠNG 4: Một số mô hình toán học thiết kế thuật toán cho dữ liệu lớn

4.1. Mô hình toán học thiết kế thuật toán cho MapReduce

4.2. Thuật toán và cấu trúc dữ liệu giảm I/O cho dữ liệu lớn

4.3. Một số kỹ thuật giảm chiều dữ liệu vẫn giữ cấu trúc

11. Tài liệu học tập: (danh mục các giáo trình, nếu không có thì bỏ trống)

12. Tài liệu tham khảo:

9. Nathan Marz (Twitter) and James Warren. 2012. Big Data: Principles and best practices of scalable realtime data systems. Manning Publications Co., Greenwich, CT, USA.
10. J. Lin and C. Dyer: Data-Intensive Text Processing with MapReduce, Morgan & Claypool Publishers, 2010
11. Chuck Lam. 2010. Hadoop in Action (1st ed.). Manning Publications Co., Greenwich, CT, USA.
12. Ron Bekkerman. 2011. Scaling up Machine Learning Parallel and Distributed Approaches
13. J. Han, M. Kamber, and J. Pei: Data Mining: Concepts and Techniques (3rd Ed), Morgan Kaufmann, 2012
14. Jothy Rosenberg and Arthur Mateos. 2011. The Cloud at Your Service. Manning Publications Co., Greenwich, CT, USA.

IT7391 Các chủ đề nâng cao đối với phân tích dữ liệu lớn

Advanced topics for big data analytics

1. **Tên học phần:** Các chủ đề nâng cao đối với phân tích dữ liệu lớn
2. Mã học phần:
3. **Tên tiếng Anh:** Advanced topics for big data analytics
4. Khối lượng: 3(2-2-0-6)
 - Lý thuyết: 30 tiết
 - Bài tập: 30 tiết (BTL)

- Thí nghiệm:

5. Đối tượng tham dự: NCS thuộc chuyên ngành Hệ thống thông tin

6. Mục tiêu của học phần: Học phần này nhằm mang lại cho NCS:

- Các kiến thức nâng cao về lý luận chuyên ngành Hệ thống thông tin
- Rèn luyện khả năng tư duy
- Rèn luyện kỹ năng thí nghiệm của chuyên ngành Hệ thống thông tin
- Có khả năng lựa chọn các phương pháp và kỹ thuật thích hợp đối với phân tích dữ liệu lớn.

7. Nội dung tóm tắt: Học phần này sẽ trình bày các phương pháp và kỹ thuật đối với phân tích dữ liệu lớn theo hướng nâng cao. Các chủ đề liên quan tới phân tích dữ liệu lớn sẽ bao gồm: phân loại và hồi qui, các mô hình đồ thị xác suất, giảm chiều, mô hình thưa.... đối với dữ liệu lớn. Mục tiêu của học phần này là giúp học viên có được các kiến thức nâng cao cùng với những phân tích sâu về các chủ đề nói trên.

8. Nhiệm vụ của NCS:

- Dự lớp: đầy đủ, theo quy chế
- Bài tập: hoàn thành các bài tập
- Thí nghiệm:

9. Đánh giá kết quả: (cách cho điểm giống như quy định đối với Cao học)

- Mức độ dự giờ giảng:
- Kiểm tra định kỳ: trọng số 0.4
- Thi kết thúc học phần: trọng số 0.6

10. Nội dung chi tiết học phần:

PHẦN MỞ ĐẦU

Giới thiệu môn học

Giới thiệu đề cương môn học

Giới thiệu tài liệu tham khảo

CHƯƠNG 1: Tổng quan dữ liệu lớn và phân tích dữ liệu lớn

1. Các khái niệm cơ bản về dữ liệu lớn và xử lý dữ liệu lớn
2. Các bài toán phân tích dữ liệu lớn
3. Các ứng dụng của phân tích dữ liệu lớn

CHƯƠNG 2: Các chủ đề nâng cao đối với phân tích dữ liệu lớn

1. Phân loại và hồi qui dữ liệu lớn
2. Các mô hình đồ thị xác suất
3. Giảm gradient ngẫu nhiên
4. Học online, suy diễn ngẫu nhiên
5. Giảm chiều phi tuyến
6. Mô hình hóa chủ đề
7. Mô hình thưa
8. Học đồ thị/mạng
9. Học biểu diễn
10. Học phân tán không đồng bộ

11. Tài liệu học tập: (danh mục các giáo trình, nếu không có thì bỏ trống)

12. Tài liệu tham khảo:

1. Izenman. Modern Multivariate Statistical techniques, Springer 2008
2. Trevor Hastie, Robert Tibshirani, Jerome Friedman. The Elements of Statistical Learning, Springer 2009
3. Ethem Alpayđın. Introduction to Machine Learning, MIT press 2010

