

ĐỀ CƯƠNG CHI TIẾT HỌC PHẦN CÔNG NGHỆ XML VÀ WEB NGỮ NGHĨA

1. Thông tin về giáo viên

TT	Họ tên giáo viên	Học hàm	Học vị	Đơn vị công tác (Bộ môn)
1	Trần Nguyên Ngọc	Giảng viên chính	Tiến sỹ	Bộ môn Khoa học máy tính, Khoa Công nghệ thông tin
2	Trần Cao Trường	Giảng viên	Thạc sỹ	Bộ môn Khoa học máy tính, Khoa Công nghệ thông tin

Thời gian, địa điểm làm việc: 7h đến 17h30 hàng ngày, Bộ môn Khoa học máy tính, Khoa Công nghệ Thông tin Tầng 2, nhà A1

Địa chỉ liên hệ: Bộ môn Khoa học máy tính, khoa Công nghệ thông tin

Điện thoại, email: 0948435163, tonono79@yahoo.com

Các hướng nghiên cứu chính: Kỹ thuật đồ họa và xử lý ảnh, Học máy, Điều khiển tối ưu

2. Thông tin chung về học phần

- Tên học phần: Công nghệ XML và WEB ngữ nghĩa
- Mã học phần: 1226415
- Số tín chỉ: 3
- Học phần (bắt buộc hay lựa chọn): lựa chọn
- Các học phần tiên quyết: Kỹ thuật lập trình, Quản trị hệ CSDL
- Các yêu cầu đối với học phần (nếu có):
- Giờ tín chỉ đối với các hoạt động:
 - Nghe giảng lý thuyết: 45
 - Làm bài tập trên lớp: 12
 - Thảo luận: 3
 - Thực hành, thực tập (ở PTN, nhà máy, thực tập...):
 - Hoạt động theo nhóm: 15
 - Tự học: 90

- Khoa/Bộ môn phụ trách học phần, địa chỉ: Công nghệ thông tin/ Khoa học máy tính

3. Mục tiêu của học phần

- Giới thiệu về Web ngữ nghĩa và các hình thái phát triển
- Tiếp cận một số phương pháp khai phá dữ liệu trực tiếp từ internet.
- Giới thiệu phương thức tổ chức dữ liệu cho Web ngữ nghĩa SPARQL & Ngôn ngữ truy vấn RDF.

4. Tóm tắt nội dung học phần (khoảng 150 từ)

Môn học này thuộc nhóm ngành Trí tuệ nhân tạo và Quản trị tri thức. Học phần sẽ trang bị cho học viên những kiến thức về Semantic Web (hay Web có ngữ nghĩa) là thế hệ mở rộng của Web hiện tại được đưa ra bởi Tim Berners-Lee vào khoảng năm 1998. Semantic Web sẽ hỗ trợ học viên có kiến thức tổng quan về một phương pháp tìm kiếm thông tin trên mạng một cách nhanh chóng, chuẩn xác và thông minh hơn so với các công cụ tìm kiếm truyền thống. Thông qua việc tiếp cận một công nghệ tổ chức dữ liệu mới SPARQL và phương pháp ghép nối, đối sánh thông tin dựa trên Ontology và logic mô tả người học sẽ từng bước tiếp cận với WEB 2.0, 3.0 và những công nghệ mới về WEB hiện vẫn đang được các nhà khoa học nghiên cứu và phát triển.

5. Nội dung chi tiết học phần (tên các chương, mục, tiểu mục)

7.1. PHẦN GIẢNG DẠY TRÊN LỚP: 45 tiết

Chương	Nội dung	Số tiết	TLT K
1	Chương 1. Lịch sử hình thành và nhu cầu phát triển web ngữ nghĩa. 1. Phương thức trao đổi thông tin trên internet thông qua website. 2. Ngôn ngữ HTML và công nghệ XML. 3. Khai thác thông tin từ website và các yêu cầu mới. 4. Sự ra đời và một số mốc phát triển của Web	9	2,3

	ngữ nghĩa		
2	<p>Chương 2: Ontology nền tảng xây dựng Web ngữ nghĩa</p> <ol style="list-style-type: none"> 1. Khái niệm Ontology. 2. Khung mô tả dữ liệu RDF. 3. Ontology và logic mô tả 4. Sơ đồ Ontology. 	9	2,3
3	<p>Chương 3: Ngôn ngữ SPARQL</p> <ol style="list-style-type: none"> 1. Giới thiệu về SPARQL và các ràng buộc dữ liệu. 2. Cú pháp cấu trúc của truy vấn SPARQL. 3. Các phép toán và điều kiện trên câu truy vấn. 	9	2,3,4
4	<p>Chương 4: Xây dựng hệ thống web ngữ nghĩa.</p> <ol style="list-style-type: none"> 1. Nhu cầu khai thác thông tin từ hệ thống website cũ. 2. Các phương thức khai thác thông tin trang web. 3. Hình thành các Ontology tự động. 4. Khớp các Ontology. 	9	2,3,5
5	<p>Chương 5: Webcrawler và khai thác thông tin từ Web ngữ nghĩa.</p> <ol style="list-style-type: none"> 1. Webcrawler và các ứng dụng. 2. Phân loại và tổ chức thông tin. 3. Khai thác thông tin thông qua web ngữ nghĩa 	9	2,3,5
	Tổng	45	

7.2. PHẦN LÀM BÀI TẬP TRÊN LỚP: 9 tiết

TT	Bài TH, TN	Số tiết	Địa điểm	TLTK
1	Bài tập chương 1: Tìm hiểu công nghệ XML và ứng dụng xây dựng CSDL dạng bảng XML	3	Phòng máy bộ môn	1,2,3,4
2	Bài tập chương 3: Thực hành	3	Phòng máy bộ môn	1,2,3,4

	các câu lệnh SPARQL			
3	Bài tập chương 5: Xây dựng công cụ Webcrawler để thu thập tin tự động	3	Phòng máy bộ môn	4,5,6
	Tổng	9		

7.3. PHẦN THẢO LUẬN: 15 tiết

TT	Nội dung	Số tiết	Địa điểm	TLT K
1	Mạng xã hội và các xu hướng phát triển	3	Hội trường học	1,2
2	Search engine	3	Hội trường học	3,4
	Tổng	6		

6. Giáo trình, tài liệu tham khảo

T	T	Tên tài liệu	Tình trạng tài liệu			
			Có trên thư viện	Giáo viên hoặc Khoa có, cho mượn để TV photo hoặc có File Điện tử	Đề nghị mua mới	Đề nghị biên soạn mới
1		Daniele Nardi & Ronald J.Brachman, An Introduction to Description Logics / Cambridge University Press New York, NY, USA / 2003		x		
2		Michael C.Daconta, Leo J. Obrst & Kevin T. Smith, The Semantic Web: A Guide to the Future of XML, Web Services, and		x		

	Knowledge Management./Wiley/ 2003.				
3	SPARQL Query Language for RDF - http://www.w3.org/TR/2008/REC-rdf-sparql-query-20080115/		x		
4	http://www.w3.org/TR/rdf-sparql-query/ .		x		
5	Christopher D. Manning/ Introduction to Information Retrieval/ Cambridge university press/ 2008		x		
6	http://www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.html		x		

7. Hình thức tổ chức dạy học

7.1. *Lịch trình chung:* (Ghi tổng số giờ cho mỗi cột)

Nội dung	Hình thức tổ chức dạy học môn học					Tổng
	Lên lớp			Thực hành, thí nghiệm, thực tập...	Tự học, tự nghiên cứu	
	Lý thuyết	Bài tập	Thảo luận			
Lịch sử hình thành và nhu cầu phát triển web ngữ nghĩa	6	0	3	0	0	9
Ontology nền tảng xây dựng Web ngữ nghĩa	6	3	0	0	0	9
Ngôn ngữ SPARQL	9	3	0	0	6	21
Xây dựng hệ thống web ngữ nghĩa	6	3	3	0	6	15
Webcrawler và khai thác thông tin từ	6	3	0	0	6	15

Web ngữ nghĩa						
---------------	--	--	--	--	--	--

7.2. Lịch trình tổ chức dạy học cụ thể

Bài giảng 1: Giới thiệu tổng quan về Web ngữ nghĩa và các ứng dụng

Chương, mục: 1.1, 1.2

Tiết thứ: 1-3 Tuần thứ: 1

Mục đích, yêu cầu: Giới thiệu những khái niệm cơ bản của môn học, tính cấp thiết của các nội dung nghiên cứu cũng như phạm vi ứng dụng thực tế

Hình thức tổ chức dạy học: Giảng dạy lý thuyết trên lớp, thảo luận mở dạng hỏi đáp trực tiếp

Thời gian: 3 tiết

Địa điểm: H5415

Nội dung chính: Giới thiệu tổng quan về Lịch sử hình thành và nhu cầu phát triển web ngữ nghĩa.

Yêu cầu sinh viên chuẩn bị: Mượn tài liệu tham khảo; Tìm hiểu ngôn ngữ HTML, ôn tập lập trình WEB

Ghi chú (tài liệu tham khảo): TL1, TL5

Bài giảng 2: Giới thiệu về công nghệ XML

Chương, mục: 1.1, 1.2

Tiết thứ: 4-6 Tuần thứ: 2

Mục đích, yêu cầu: Trang bị cho học viên những kiến thức cơ bản về ngôn ngữ HTML và các dạng phát triển kế tiếp. Khảo sát các trang web thực tế để đánh giá mức độ phát triển của công nghệ

Hình thức tổ chức dạy học: Giảng dạy lý thuyết trên lớp

Thời gian: 3 tiết

Địa điểm: H5415

Nội dung chính: Ngôn ngữ HTML và công nghệ XML; Tìm hiểu công nghệ XML và ứng dụng xây dựng CSDL dạng bảng XML

Yêu cầu sinh viên chuẩn bị: Tham khảo các tài liệu từ cộng đồng WEB 3.0 trên mạng Internet

Ghi chú (tài liệu tham khảo): TL2, TL4

Bài giảng 3: **Quá trình WEB ngữ nghĩa hình thành và phát triển**

Chương, mục: 1.3, 1.4

Tiết thứ: 7-9 Tuần thứ: 3

Mục đích, yêu cầu: Giới thiệu và phân tích về mạng xã hội từ đó phân tích xu hướng hình thành và phát triển WEB ngữ nghĩa

Hình thức tổ chức dạy học: Giảng dạy lý thuyết trên lớp

Thời gian: 3 tiết

Địa điểm: H5415

Nội dung chính: Sự ra đời và một số mốc phát triển của Web ngữ nghĩa; Mạng xã hội và các xu hướng phát triển.

Yêu cầu sinh viên chuẩn bị: Tìm hiểu về một số công nghệ tìm kiếm của Google, Yandex, trang mạng xã hội Facebook

Ghi chú (tài liệu tham khảo): TL2, TL4

Bài giảng 4: **Ontology và vấn đề mô tả tri thức**

Chương, mục: 2.1

Tiết thứ: 10-12 Tuần thứ: 4

Mục đích, yêu cầu: Trang bị kiến thức cơ sở về ontology và giới thiệu một số công cụ thực hành xây dựng ontology

Hình thức tổ chức dạy học: Giảng dạy lý thuyết trên lớp

Thời gian: 3 tiết

Địa điểm: H5415

Nội dung chính: Ontology nền tảng xây dựng Web ngữ nghĩa; Tìm hiểu về công cụ cho phép mô tả ontology trực tuyến.

Yêu cầu sinh viên chuẩn bị: Tìm và cài đặt một số công cụ xây dựng ontology

Ghi chú (tài liệu tham khảo): TL1, TL3

Bài giảng 5: **Khung mô tả dữ liệu RDF**

Chương, mục: 2.2,2.3,2.4

Tiết thứ: 13-15 Tuần thứ: 5

Mục đích, yêu cầu: Trang bị kiến thức về khung mô tả dữ liệu, phục vụ cho bài toán xây dựng web ngữ nghĩa và các ứng dụng phân tích quan điểm, khai phá tri thức

Hình thức tổ chức dạy học: Giảng dạy lý thuyết trên lớp

Thời gian: 3 tiết

Địa điểm: H5415

Nội dung chính: Khung mô tả dữ liệu RDF; Tìm hiểu thêm về opinion mining và sentiment analysis.

Yêu cầu sinh viên chuẩn bị: Tìm kiếm và down load một số công cụ tách từ (VNtokenizer)

Ghi chú (tài liệu tham khảo): TL1, TL3, TL6

Bài giảng 6: **Truy vấn dữ liệu ngữ nghĩa**

Chương, mục: 3.1

Tiết thứ: 16-18 Tuần thứ: 6

Mục đích, yêu cầu: Trang bị kiến thức về ngôn ngữ truy vấn cho bài toán khai phá tri thức từ thông tin có tổ chức

Hình thức tổ chức dạy học: Giảng dạy lý thuyết trên lớp

Thời gian: 3 tiết

Địa điểm: H5415

Nội dung chính: Dữ liệu liên kết và cách thức truy vấn; Giới thiệu về SPARQL và các ràng buộc dữ liệu;

Yêu cầu sinh viên chuẩn bị: Chuẩn bị môi trường semweb và một số hàm truy vấn SPARQL

Ghi chú (tài liệu tham khảo): TL4

Bài giảng 7: Truy vấn dữ liệu ngữ nghĩa với SPARQL

Chương, mục: 3.1

Tiết thứ: 19-21 Tuần thứ: 7

Mục đích, yêu cầu: Thực hành với các thao tác truy vấn dữ liệu sử dụng SPARQL

Hình thức tổ chức dạy học: Hướng dẫn thực hành trên lớp

Thời gian: 3 tiết

Địa điểm: H5415

Nội dung chính: Cú pháp cấu trúc của truy vấn SPARQL; Thực hành các câu lệnh SPARQL trong môi trường SEMWEB cho .NET;

Yêu cầu sinh viên chuẩn bị: Chuẩn bị môi trường semweb và một số hàm truy vấn SPARQL trên .NET

Ghi chú (tài liệu tham khảo): TL4

Bài giảng 8: Ôn tập và đánh giá giữa học phần

Chương, mục: 1,2,3

Tiết thứ: 22-24 Tuần thứ: 8

Mục đích, yêu cầu: Ôn tập và làm bài kiểm tra giữa môn

Hình thức tổ chức dạy học: Hướng dẫn ôn tập và kiểm tra giữa học phần

Thời gian: 2 tiết ôn tập, 1 tiết kiểm tra

Địa điểm: H5415

Nội dung chính: Các phép toán và điều kiện trên câu truy vấn; Ôn tập về lập trình WEB, XML trong môi trường JAVA, .NET;

Yêu cầu sinh viên chuẩn bị: Chuẩn bị tốt các nội dung lý thuyết cho bài kiểm tra về tổ chức dữ liệu RDF và truy vấn SPARQL

Ghi chú (tài liệu tham khảo): TL4

Bài giảng 9: Xây dựng hệ thống WEB ngữ nghĩa

Chương, mục: 4.1,4.2

Tiết thứ: 25-27 Tuần thứ: 9

Mục đích, yêu cầu: Giới thiệu các bước chung để có thể xây dựng hệ thống WEB ngữ nghĩa

Hình thức tổ chức dạy học: Giảng dạy lý thuyết và thảo luận mở

Thời gian: 2 tiết giảng lý thuyết, 1 tiết thảo luận về tổ chức nhóm

Địa điểm: H5415

Nội dung chính: Xây dựng hệ thống web ngữ nghĩa; Một số công cụ và cộng đồng hỗ trợ phát triển WEB ngữ nghĩa; Tổ chức và phân nhóm phát triển các modul

Yêu cầu sinh viên chuẩn bị: Ôn tập về kiến trúc tổng thể WEB ngữ nghĩa. Tìm hiểu về JENA framework cho môi trường JAVA

Ghi chú (tài liệu tham khảo): TL1, TL4

Bài giảng 10: **Thực hành Xây dựng Ontology**

Chương, mục: 4.2, 4.3

Tiết thứ: 28-30 Tuần thứ: 10

Mục đích, yêu cầu: Tìm hiểu và thao tác trên các công cụ tạo dựng ontology và lưu trữ dưới các định dạng khác nhau

Hình thức tổ chức dạy học: Giảng dạy lý thuyết và hướng dẫn thực hành

Thời gian: 2 tiết giảng lý thuyết, 1 tiết thực hành mẫu

Địa điểm: H5415

Nội dung chính: Hình thành các Ontology tự động; Giới thiệu quy trình cơ bản trong Protégé để tạo Ontology về học tập.

Yêu cầu sinh viên chuẩn bị: Chuẩn bị cài đặt Protégé bản từ 3.5 trở lên; Tìm hiểu một số nội dung liên quan đến xử lý tiếng Việt

Ghi chú (tài liệu tham khảo): TL1, TL4, TL6

Bài giảng 11: **Khớp Ontology**

Chương, mục: 4.4

Tiết thứ: 31-33 Tuần thứ: 11

Mục đích, yêu cầu: Trang bị kiến thức lập trình cần thiết và các thuật toán cơ bản cho việc khai phá dữ liệu từ các ontology

Hình thức tổ chức dạy học: Giảng dạy lý thuyết và hướng dẫn thực hành

Thời gian: 2 tiết giảng lý thuyết, 1 tiết thực hành mẫu

Địa điểm: H5415

Nội dung chính: Thuật toán Khớp các Ontology; Lập trình khớp ontology.

Yêu cầu sinh viên chuẩn bị: Ôn tập các thuật toán đối sánh xâu ký tự, tìm xâu con chung dài nhất của môn học Cấu trúc dữ liệu và giải thuật; Ôn tập các kỹ năng lập trình C# hoặc Java. Chú ý kỹ năng phân tích xâu ký tự.

Ghi chú (tài liệu tham khảo): TL1, TL4, TL6

Bài giảng 12: **WEBCRAWLER và bài toán thu thập dữ liệu**

Chương, mục: 5.1

Tiết thứ: 34-36 Tuần thứ: 12

Mục đích, yêu cầu: Trang bị kỹ năng thu thập thông tin tự động, phân tích cấu trúc các trang web, từ đó xây dựng công cụ trong môi trường lập trình .NET

Hình thức tổ chức dạy học: Giảng dạy lý thuyết và hướng dẫn thực hành

Thời gian: 2 tiết giảng lý thuyết, 1 tiết thực hành mẫu

Địa điểm: H5415

Nội dung chính: Webcrawler và các ứng dụng; Xây dựng công cụ crawler các trang web.

Yêu cầu sinh viên chuẩn bị: Tìm hiểu các công cụ thu thập tin tự động; Nghiên cứu công nghệ RSS.

Ghi chú (tài liệu tham khảo): TL1, TL4, TL6

Bài giảng 13: **WEB ngữ nghĩa và các hướng ứng dụng hiện nay**

Chương, mục: 5.2, 5.3

Tiết thứ: 37-39 Tuần thứ: 13

Mục đích, yêu cầu: Giúp hiểu rõ hơn mục đích của web ngữ nghĩa trong các ứng dụng thực tế

Hình thức tổ chức dạy học: Giảng dạy lý thuyết và DEMO minh họa

Thời gian: 2 tiết giảng lý thuyết, 1 tiết DEMO minh họa

Địa điểm: H5415

Nội dung chính: Khai thác thông tin thông qua web ngữ nghĩa; Các xu hướng phát triển của các hệ thống Gợi ý tự động.

Yêu cầu sinh viên chuẩn bị: Tìm hiểu thêm về cơ chế gợi ý mua hàng, xem video trực tuyến của Youtube, đặt phòng online agoda.

Ghi chú (tài liệu tham khảo): TL1, TL4, TL6

Bài giảng 14: Thực hành tương tác với WEB ngữ nghĩa

Chương, mục: 5.3

Tiết thứ: 40-42 Tuần thứ: 14

Mục đích, yêu cầu: Tiếp tục hoàn thiện công cụ webcrawler và đối chiếu thực tế để thấy rõ ý nghĩa của WEB ngữ nghĩa trong các ứng dụng tìm kiếm thông minh

Hình thức tổ chức dạy học: Giảng dạy lý thuyết và DEMO minh họa

Thời gian: 2 tiết giảng lý thuyết, 1 tiết DEMO minh họa

Địa điểm: H5415

Nội dung chính: Xây dựng công cụ Webcrawler để thu thập tin tự động; Ứng dụng tìm kiếm thông tin trên Ontology tự xây dựng

Yêu cầu sinh viên chuẩn bị: Tổng hợp các thông tin, hoàn thiện Ontology về các môn học, liên kết với WebServer và tổ chức công cụ tìm kiếm trên nền giao diện Web.

Ghi chú (tài liệu tham khảo): TL1, TL4, TL6

Bài giảng 15: Ôn tập và hướng dẫn thi hết học phần

Chương, mục: 1,2,3,4,5

Tiết thứ: 43-45 Tuần thứ: 15

Mục đích, yêu cầu: Hệ thống lại toàn bộ kiến thức, định hướng ôn tập và giải đáp các vướng mắc

Hình thức tổ chức dạy học: Giảng dạy lý thuyết

Thời gian: 3 tiết giảng lý thuyết

Địa điểm: H5415

Nội dung chính: Tổng ôn tập toàn môn học; Hướng dẫn đánh giá qua làm việc nhóm

Yêu cầu sinh viên chuẩn bị: Ôn tập các nội dung đã học.

Ghi chú (tài liệu tham khảo): TL1-TL6

8. Chính sách đối với môn học và các yêu cầu khác của giáo viên

Tham gia học tập đầy đủ: có điểm danh, không mất trật tự, đi học đúng giờ

Làm bài tập, tham gia thảo luận: làm đủ bài tập, tham gia các buổi thảo luận

Kiểm tra, thi: Luôn bao gồm 2 phần lý thuyết và thực hành với trọng số như nhau

Học viên có thể được đặc cách miễn thi nếu tham gia tốt quá trình học tập và đạt kết quả xuất sắc trong các bài kiểm tra định kỳ, tích cực tham gia thảo luận.

9. Phương pháp, hình thức kiểm tra - đánh giá kết quả học tập môn học

Phân chia các mục tiêu cho từng hình thức kiểm tra - đánh giá

9.1. Kiểm tra – đánh giá thường xuyên

Kiểm tra cả nội dung lý thuyết và kết quả làm bài tập nhóm, thực hành lập trình.

9.2. Kiểm tra - đánh giá định kì: Bao gồm các phần sau (trọng số của từng phần do giảng viên đề xuất, chủ nhiệm bộ môn thông qua):

- Tham gia học tập trên lớp (đi học đầy đủ, chuẩn bị bài tốt và tích cực thảo luận,...) 10%;
- Phần tự học, tự nghiên cứu (hoàn thành tốt nội dung, nhiệm vụ mà giảng viên giao cho cá nhân/ tuần; bài tập nhóm / tháng; bài tập cá nhân/ học kì,...) 20%;
- Hoạt động theo nhóm 10%;
- Kiểm tra - đánh giá giữa kì 10%;
- Kiểm tra - đánh giá cuối kì 50%;
- Các kiểm tra khác.

9.3. Tiêu chí đánh giá các loại bài tập

- Chương trình đảm bảo chạy đúng yêu cầu, học viên phải trả lời được các câu hỏi lý thuyết liên quan
- Học viên có thể không dùng các ngôn ngữ lập trình mà giáo viên đã giới thiệu trên lớp để trả bài.
- Khuyến khích các bài tập có tính sáng tạo cao và ứng dụng độc đáo.

Chủ nhiệm Khoa
(Ký và ghi rõ họ tên)

Chủ nhiệm Bộ môn
(Ký và ghi rõ họ tên)

Giảng viên biên soạn
(Ký và ghi rõ họ tên)

4// Đào Thanh Tĩnh

2// Ngô Hữu Phúc

TS Trần Nguyên Ngọc

