

WEB MINING

Version: 2021.12.14

1. THÔNG TIN CHUNG

GENERAL INFORMATION

Tên học phần	<i>Web Mining</i>
Course name:	
Mã học phần	IT4868E
Code:	
Khối lượng	3(3-1-0-6)
Credit:	<ul style="list-style-type: none"> - Lý thuyết - Lecture: 45 hours - Bài tập - Exercise: 15 hours - Thí nghiệm - Experiments: 0 hours
Học phần tiên quyết	No
Prerequisite:	
Học phần học trước	<ul style="list-style-type: none"> - IT3190E: Machine Learning - IT3320E: Introduction to Deep Learning
Prior course:	
Học phần song hành	No
Paralell course:	

2. MÔ TẢ HỌC PHẦN - COURSE DESCRIPTION

Học phần cung cấp các kiến thức cơ bản và chuyên sâu về các bài toán trong khai phá cấu trúc, nội dung và hành vi người dùng trên web: Tìm kiếm thông tin, phân tích liên kết, trích rút thông tin, khai phá quan điểm, và hệ gợi ý. Học phần cũng giới thiệu các ứng dụng của khai phá web trong quảng cáo trực tuyến.

Ngoài ra môn học cũng cung cấp cho học viên các kỹ năng làm việc nhóm và thuyết trình.

This course provides fundamental and advance knowledge in mining structures, contents, and user behaviour in the web: Information retrieval, link analysis, information extraction, and recommendation systems. The course also introduces the applications of web mining in online advertisement.

The course also provides students with the ability of team working and presentation.

MỤC TIÊU VÀ CHUẨN ĐẦU RA CỦA HỌC PHẦN

OBJECTIVE AND OUTPUT REQUIREMENTS

Sinh viên hoàn thành học phần này có khả năng

After this course the student will obtain the followings:

Mục tiêu/CĐR Objective	Mô tả mục tiêu/Chuẩn đầu ra của học phần Objective description / Output requirements	CĐR được phân bổ cho HP/ Mức độ (I/T/U) Output division/ Level (I/T/U)
---------------------------	---	--

[1]	[2]	[3]
M1	Understanding fundamental and advanced techniques in web mining	[1.1,1.2]
M1.1	Having the ability to model the problems and represent data in web mining	[1.1,1.2](T)
M1.2	Having the ability to apply techniques in machine learning and data mining to web mining	[1.1,1.2](U)
M2	Understanding the applications of web mining	[1.3,2.1,2.5]
M2.1	Understanding the applications of web mining in e-commerce, business intelligence and online advertisement	[1.3,2.1,2.5](T)
M2.2	Having the ability to apply web mining techniques to solve a problem in a relevant domain	[1.3,2.1,2.5](I)
M3	Having the ability build a simple web mining systems via team working	[2.1,2.2,2.3,3.1,3.2,4.2 4.5]
M3.1	Having the ability to analyze problem, select solution for a web mining problems	[2.1,2.2,2.3](T)
M3.2	Having the ability to deploy a web mining system	[4.2 4.5](U)
M3.3	Having the ability of team working, writting report, and presentation	[3.1,3.2](U)

3. TÀI LIỆU HỌC TẬP

REFERENCES

Textbook

Reference book

- [1] Jure Leskovec, Anand Rajaraman and Jeffrey D. Ullman (2020). *Mining massive datasets*. 3rd edition. Cambridge University Press.
- [2] Bing Liu (2011). *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data*. Springer.
- [3] Ronen Feldman and James Sanger (2007). *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press.
- [4] Dietmar Jannach, Markus Zanker, Alexander Felfernig, Gerhard Friedrich (2010). *Recommender Systems: An Introduction*. Cambridge University Press.
- [5] Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze (2008). *Introduction to Information Retrieval*. Cambridge University Press.

4. CÁCH ĐÁNH GIÁ HỌC PHẦN - EVALUATION

Điểm thành phần Module	Phương pháp đánh giá cụ thể	Mô tả Descriptio	CDR được	Tỷ trọng Percentag
---------------------------	--------------------------------	---------------------	-------------	-----------------------

	Evaluation method	n	đánh giá Output	e
[1]	[2]	[3]	[4]	[5]
A1. Điểm quá trình Mid-term (*)	Đánh giá quá trình Progress			40%
	A1.1. Thảo luận trên lớp Discussion	Thuyết trình Presentation	M1.3; M3.1; M3.2	10%
	A1.2. Bài tập về nhà Homework	Tự luận Written	M1.3; M3.1; M3.2	10%
	A1.3. Bài tập nhóm Capstone Project	Báo cáo Presentation	M1.3; M3.1; M3.2	20%
A2. Điểm cuối kỳ Final exam	A2.1. Thi cuối kỳ Final exam	Thi viết Written exam	M1÷M2	60%

* Điểm quá trình sẽ được điều chỉnh bằng cách cộng thêm điểm chuyên cần. Điểm chuyên cần có giá trị từ -2 đến +1, theo Quy chế Đào tạo đại học hệ chính quy của Trường ĐH Bách khoa Hà Nội.

The evaluation can be adjusted with some bonus. The bonus should belong to [-2, +1], according to the policy of Hanoi University of Science and Technology.

5. KẾ HOẠCH GIẢNG DẠY - SCHEDULE

Tuần Week	Nội dung Content	CDR học phần Output	Hoạt động dạy và học Teaching activities	Bài đánh giá Evaluation
[1]	[2]	[3]	[4]	[5]
1	Intro to Web mining - Capstone project description	M4.1	Reading materials, Lecture	
2	Chapter 1 Fundamentals of Machine Learning 1. Supervised machine learning 2. Deep learning 3. Unsupervised learning		Reading materials, Lecture	
3	Chapter 1 Fundamentals of Machine Learning (cont) Chapter 2 Data visualization 1. Statics graphs 2. Visualization by pixel 3. Visualization in vector space		Reading materials, Lecture	

	4. Hyperbolic tree 5. Self-organizing maps			
4	Chapter 3 Information retrieval 1. Basic concepts in IR 2. IR models 3. Relevance feedback 4. Evaluation 5. Preprocessing 6. Inverted index 7. Latent semantic indexing 8. Web search 9. Meta search 10. Web spam	M1,M2	Reading materials, Lecture	A2.1
5	Chapter 4 Link analysis 1. Graph ranking	M1,M2	Reading materials, Lecture	A2.1
6	Chapter 4 Link analysis (cont) 2. Community detection 3. Học biểu diễn đồ thị	M1,M2	Reading materials, Lecture	A2.1
7	Chapter 5 Opinion mining 1. Intro to Opinion mining 2. Unsupervised sentiment analysis		Reading materials, Lecture	
8	Chapter 5 Opinion mining (cont) 3. Supervised sentiment analysis 4. Aspect-based sentiment analysis		Reading materials, Lecture	
9	Capstone project checkpoint		Presentation	
10	Chapter 6 Information extraction 1. IE architecture 2. Named entity recognition 3. Relation extraction 4. Phân giải đồng tham chiếu	M1,M2	Reading materials, Lecture	A2.1
11	Chapter 7 Recommendation systems 1. Intro to Recsys 2. Evaluations 3. K-nearest neighbors collaborative filtering	M1,M2	Reading materials, Lecture	A2.1
12	Chapter 7 Recommendation systems (cont) 4. Matrix factorization for collaborative filtering 5. Neural collaborative filtering 6. Session-based recommendation	M1,M2	Reading materials, Lecture	A2.1

13	Advanced topics in web mining		Reading materials, Lecture	
14	Revision			
15	<i>Capstone project</i>	M1,M3	Presentation	A1.1

6. QUY ĐỊNH CỦA HỌC PHẦN - COURSE REQUIREMENT

(The specific requirements if any)

7. NGÀY PHÊ DUYỆT - DATE:

Chủ tịch hội đồng
Committee chair

Nhóm xây dựng đề cương
Course preparation group

Dr. Nguyen Kiem Hieu
Ass. Prof. Than Quang Khoat
Ass. Prof. Nguyen Kim Anh
Ass. Prof. Le Thanh Huong

8. QUÁ TRÌNH CẬP NHẬT - UPDATE INFORMATION

ST T No	Nội dung điều chỉnh Update content	Ngày tháng được phê duyet Date approved	Áp dụng từ kỳ/ khóa Applicable from	Ghi chú Note
1			
2			