

G3	Kh n ng cài t và ánh giá c các thu t toán, mô hình và s d ng c các công c mã ngu n m	2.1, 2.3, 2.4, 2.5
G4	K n ng làm vi c nhóm, và thuy t trình b ng mi ng	2.4, 2.5, 3.1, 3.2

8. Chu n u ra c a h c ph n

M c tiêu	Chu n u ra h c ph n	Mô t (Sau khi h c xong môn h c này, ng i h c có th :)	Chu n u ra CDIO
G1	G1.1	Hi u c ý ngh a, t m quan tr ng và các thách th c c a truy tìm thông tin.	1.3
	G1.2	Trình bày c ki n trúc t ng quát c a m t h th ng truy tìm thông tin	1.3, 2.3
	G1.3	Hi u c các b c quan tr ng c a quá trình tí n x lý tài li u	1.3
	G1.4	Hi u c c u trúc ch m c ng c (inverted index)	1.3
	G1.5	Trình bày c các k thu t ph n h i và m r ng truy v n	1.3
	G1.6	Trình bày c cách ho t ng c a m t h th ng tìm ki m thông tin trên web	1.3
G2	G2.1	Hi u c ph ng pháp xây d ng ch m c ng c cho tài li u	2.1
	G2.2	Hi u c mô hình Boolean và mô hình không gian vector	2.1
	G2.3	Hi u c ph ng pháp ánh giá th c nghi m m t h th ng truy tìm thông tin	2.1, 2.3, 2.4, 2.5
	G2.4	Hi u c mô hình xác su t nh phân c l p và Okapi BM25 cho bài toán truy tìm thông tin	2.1
	G2.5	Hi u c mô hình ngôn ng query likelihood và các ph ng pháp làm tr n cho bài toán truy tìm thông tin	2.1
	G2.6	Hi u c các thu t toán phân tích liên k t (link analysis) nh Google PageRank, HITS	2.1
G3	G3.1	Cài t c các thu t toán, mô hình và s d ng c các công c mã ngu n m	2.1, 2.3, 2.4, 2.5
	G3.2	ánh giá c các thu t toán, mô hình s d ng các kho d li u chu n	2.1, 2.3, 2.4, 2.5
G4	G4.1	Làm vi c hi u qu trong m t nhóm	3.1, 2.4, 2.5
	G4.2	Trình bày tr c ám ông s d ng ph ng tí n trình chi u	3.2

9. Tài liệu học tập

- Sách, giáo trình chính:

+ [IIR] C. Manning, P. Raghavan and H. Schuetze, *Introduction to Information Retrieval*, Cambridge University Press, 2008. (<http://www-csli.stanford.edu/~hinrich/information-retrieval-book.html>)

- Sách (TLTK) tham khảo:

+ [SEIRP] W. B. Croft, D. Metzler, and T. Strohman, *Search Engines: Information Retrieval in Practice*, Pearson, London, England, 2009. (<http://www.search-engines-book.com/>)

+ [MIR] R. Baeza-Yates and B. Ribeiro-Neto, *Modern Information Retrieval: the concepts and technology behind search*, 2nd edition, ACM Press, 2010. (<http://www.mir2ed.org/>)

+ [IRIESE] S. Buettcher, C. L. A. Clarke and G. V. Cormack, *Information Retrieval: Implementing and Evaluating Search Engines*, The MIT Press, 2010. (<http://www.ir.uwaterloo.ca/book/>)

+ Các bài báo liên quan.

9. Nhiệm vụ của sinh viên

- SV không thể chỉ nhận điểm trong các nhiệm vụ sau đây sẽ bị coi là:

+ Điểm: 80%

+ Bài tập: 100%

10. Tiêu chí đánh giá thành phần và các hình thức đánh giá sinh viên

+ Thang điểm: 10

+ Kế hoạch kiểm tra như sau:

Hình thức KT	Nội dung	Thời gian	Công cụ KT	Chuẩn ra KT	Tỉ lệ (%)
Bài tập - thực hành					30
BT#1	Bài tập về tối ưu hóa và xây dựng chỉ số tài liệu	Tuần 4	Bài tập - thực hành	G2.1 G3.1	5
BT#2	Bài tập về mô hình boolean và mô hình không gian vector	Tuần 5	Bài tập - thực hành	G2.2 G3.1	5
BT#3	Bài tập về đánh giá mô hình, hình thức truy tìm thông tin	Tuần 7	Bài tập - thực hành	G2.3 G3.1 G3.2	5
BT#4	Bài tập về các thuật toán phân tích và mở rộng truy vấn	Tuần 9	Bài tập - thực hành	G1.5 G3.1	5
BT#5	Bài tập về các mô hình xác suất	Tuần 12	Bài tập - thực hành	G2.4 G3.1 G3.2	5
BT#6	Bài tập về các mô hình ngôn ngữ	Tuần 12	Bài tập - thực hành	G2.5 G3.1 G3.2	5

Ti u lu n - Báo cáo					20
	M i nhóm sinh viên 2-3 ng i ch n l trong các tài sau (có th thay i) tìm hi u và trình bày báo cáo: tài 1: Text classification tài 2: Text clustering tài 3: Text summarization tài 4: Question answering tài 5: Recommender systems tài 6: Learning to rank tài 7: XML retrieval tài 8: Multimedia information retrieval tài 9: Information extraction tài 10: Semantic search	Tu n 14,15	Báo cáo nhóm	G4.1 G4.2	
Thi cu i k					50
	- N i dung bao quát t t c các chu n u ra quan tr ng c a môn h c. - Th i gian làm bài 60-90 phút.		Thi t lu n	G1.1-G1.6 G2.1-G2.6	

11. N i dung chi ti t h c ph n

Tu n	N i dung	Chu n u ra h c ph n
1	Ch ng 1: T ng quan v truy tìm thông tin (4/0/8)	
	A/ Các n i dung và PPGD chính trên l p (4) N i dung GD lý thuy t + Các khái ni m c b n + Ki n trúc c a m t h th ng truy tìm thông tin + M t s v n và thách th c PPGD chính + Thuy t gi ng + Minh h a các h th ng truy tìm thông tin + Trình chi u Powerpoint	G1.1 G1.2
	B/ Các n i dung c n t h c nhà (8) + c l i v ki n trúc c a m t h th ng truy tìm thông tin + Tài li u o A. Singhal, <i>Modern Information Retrieval: A Brief Overview</i> , In IEEE Data Engineering Bulletin 24(4), pages 35 -43, 2001. o [SEIRP] ch ng 1. o [MIR] ch ng 1.	G1.2

	Chương 2: Tin x lý và xây dựng cơ sở dữ liệu (4/0/8)	
2	A/ Tóm tắt các ND và PPGD chính trên lớp (4) Nội dung GD lý thuyết <ul style="list-style-type: none"> + Tin x lý dữ liệu: tokenization, stemming, lemmatization, stop words, phrases + Ý nghĩa của việc xây dựng cơ sở dữ liệu + Cấu trúc cơ sở dữ liệu (inverted index) + Phương pháp xây dựng cơ sở dữ liệu PPGD chính <ul style="list-style-type: none"> + Thuyết giảng + Minh họa cấu trúc cơ sở dữ liệu + Trình chiếu Powerpoint 	G1.3 G1.4 G2.1
	B/ Các nội dung cần thực hành (8) <ul style="list-style-type: none"> + Các bước chính của quá trình tin x lý dữ liệu + Các cách xây dựng cấu trúc cơ sở dữ liệu + Làm các bài tập thực hành + Tài liệu <ul style="list-style-type: none"> o [IIR] chương 2 & 4. o [SEIRP] chương 4. o [IRIESE] chương 3 & 4. 	G1.3 G1.4 G2.1 G3.1
3-5	Chương 3: Mô hình Boolean và mô hình không gian vector (4/8/8)	
	A/ Các nội dung và PPGD chính trên lớp (4) Nội dung GD lý thuyết <ul style="list-style-type: none"> + Khái niệm mô hình truy tìm thông tin + Mô hình Boolean + Mô hình không gian vector: tính toán, chuẩn hóa, xếp hạng PPGD chính <ul style="list-style-type: none"> + Thuyết giảng + Minh họa cách tính toán cho các mô hình + Trình chiếu Powerpoint 	G2.2
	B/ Các nội dung cần thực hành (8) <ul style="list-style-type: none"> + Các ví dụ mô hình boolean và mô hình không gian vector + Làm các bài tập thực hành + Cài đặt phần tin x lý và xây dựng cơ sở dữ liệu + Cài đặt mô hình boolean và mô hình không gian vector + Tài liệu <ul style="list-style-type: none"> o [IIR] chương 1 & 6. o [MIR] chương 3. o S. Heinz and J. Zobel, <i>Efficient Single-Pass Index Construction for Text Databases</i>, JASIST, 2003. 	G2.2 G2.7 G4.1

6-7	<p>Ch 4: ánh giá h th ng truy tìm thông tin (4/4/8)</p>	
	<p>A/ Các n i dung và PPGD chính trên l p (4)</p> <p>N i dung GD lý thuy t</p> <ul style="list-style-type: none"> + T m quan tr ng c a vi c ánh giá th c nghi m m t h th ng truy tìm thông tin + Các b s u t p ánh giá chu n + Các ch s ánh giá chu n: Precision, Recall, MAP, ... + So sánh 2 h th ng truy tìm thông tin + G i ý các tài báo cáo nhóm cho sinh viên. <p>PPGD chính</p> <ul style="list-style-type: none"> + Thuy t gi ng + Minh h a k thu t + Trình chi u Powerpoint + Th o lu n nhóm 	G2.3
	<p>B/ Các n i dung c n t h c nhà (8)</p> <ul style="list-style-type: none"> + Xem thêm v ph ng pháp ánh giá h th ng truy tìm thông tin + Download công c ánh giá h th ng truy tìm thông tin + Download các kho d li u th nghi m chu n + c tài li u h ng đ n s đ ng công c ánh giá v i các kho d li u chu n + Th nghi m ánh giá mô hình + Báo cáo k t qu (n p file k t qu) + Tài li u <ul style="list-style-type: none"> o [IIR] ch ng 8. o [SEIRP] ch ng 8. o [IRIESE] ch ng 12. 	G2.3 G3.1 G3.2 G4.1
8-9	<p>Ch 5: Ph n h i và m r ng truy v n (4/4/8)</p>	
	<p>A/ Các n i dung và PPGD chính trên l p (4)</p> <p>N i dung GD lý thuy t</p> <ul style="list-style-type: none"> + S đ ng thông tin ph n h i c a ng i dùng c i thi n chính xác c a h th ng + Các k thu t m r ng truy v n + T ng k t vi c ch n tài báo cáo nhóm c a SV. <p>PPGD chính</p> <ul style="list-style-type: none"> + Thuy t gi ng + Minh h a k thu t + Trình chi u Powerpoint + Th o lu n nhóm 	G1.5
	<p>B/ Các n i dung c n t h c nhà (8)</p> <ul style="list-style-type: none"> + Xem thêm v ph n h i và m r ng truy v n 	G1.5 G3.1

	<ul style="list-style-type: none"> + Cài đặt các thuật toán phân tích và mã nguyên thủy + Tài liệu <ul style="list-style-type: none"> o [IIR] chương 9. o [MIR] chương 5. 	<p>G3.2 G4.1</p>
10	<p>Chương 6: Mô hình xác suất (4/0/8)</p>	
	<p>A/ Các nội dung và PPGD chính trên lớp (4)</p> <p>Nội dung GD lý thuyết</p> <ul style="list-style-type: none"> + Mô hình xác suất nh phân cấp + Mô hình Okapi BM25 <p>PPGD chính</p> <ul style="list-style-type: none"> + Thuyết giảng + Minh họa cách tính toán cho mô hình + Trình chiếu Powerpoint 	<p>G2.4</p>
	<p>B/ Các nội dung content học nhà (8)</p> <ul style="list-style-type: none"> + cli v mô hình xác suất và cách tính trọng số Okapi BM25 + Làm các bài tập giao + Tài liệu: <ul style="list-style-type: none"> o [IIR] chương 11. o [MIR] chương 3. o [IRIESE] chương 8. 	<p>G2.4 G3.1 G3.2 G4.1</p>
11-12	<p>Chương 7: Mô hình ngôn ngữ (4/4/8)</p>	
	<p>A/ Các nội dung và PPGD chính trên lớp (4)</p> <p>Nội dung GD lý thuyết</p> <ul style="list-style-type: none"> + Mô hình ngôn ngữ query likelihood + Các phương pháp làm trơn (smoothing) <p>PPGD chính</p> <ul style="list-style-type: none"> + Thuyết giảng + Minh họa cách tính toán cho mô hình + Trình chiếu Powerpoint 	<p>G2.5</p>
	<p>B/ Các nội dung content học nhà (8)</p> <ul style="list-style-type: none"> + cli v mô hình ngôn ngữ query likelihood và các phương pháp làm trơn + Làm các bài tập giao + Cài đặt và đánh giá thực nghiệm các mô hình xác suất + Cài đặt và đánh giá thực nghiệm các mô hình ngôn ngữ + Tài liệu: <ul style="list-style-type: none"> o [IIR] chương 12. o [MIR] chương 3. o [IRIESE] chương 9. 	<p>G2.5 G3.1 G3.2 G4.1</p>

	Chương 8: Hình thức tìm kiếm thông tin trên web (4/0/8)	
13	A/ Các nội dung và PPGD chính trên lớp (4) Nội dung GD lý thuyết + Hình thức truy tìm thông tin thông thường và hình thức truy tìm thông tin trên web + Kỹ thuật cơ bản hình thức truy tìm thông tin trên web + Các vấn đề về hình thức truy tìm thông tin trên web + Các thuật toán tìm kiếm thông tin trên web PPGD chính + Thuyết giảng + Minh họa kỹ thuật + Trình chiếu Powerpoint + Thảo luận nhóm	G1.6 G2.6
	B/ Các nội dung content học nhà (8) + Xem lại các thuật toán tìm kiếm thông tin trên web + Làm các bài tập case giao + Cài đặt các thuật toán thu thập và xử lý thông tin tài liệu + Tài liệu o [IIR] chương 19-21. o [MIR] chương 11. o [IRIESE] chương 15.	G1.6 G2.6
	Chương 9: Các chiến lược (8/0/16)	
14-15	A/ Các nội dung và PPGD chính trên lớp (4) Nội dung chính trên lớp + Các nhóm báo cáo các chiến lược (tổng 8) PPGD chính + Trình chiếu Powerpoint + Thảo luận nhóm	G4.1 G4.2
	B/ Các nội dung content học nhà (8) + Tìm kiếm tài liệu, trích dẫn, viết báo cáo và trình bày kết quả chiến lược (giảng viên sẽ cung cấp các tài liệu cần thiết) + Tài liệu o Các bài báo, sách liên quan về các chiến lược.	G4.1 G4.2

14. Các hoạt động

- + Các bài tập thực hành phải được thực hiện chính bản thân sinh viên. Nếu phát hiện có sao chép thì xử lý các sinh viên có liên quan bằng hình thức đánh giá 0 (không) điểm và cấm thi cử tiếp theo.
- + Không có thi học kỳ. Nếu phát hiện thì sẽ bị kỷ luật theo quy định của Khoa và nhà trường.

15. Ngày phê duyệt: ngày tháng năm

16. Cơ sở phê duyệt

Trưởng khoa

Trưởng BM

Nhóm biên soạn

TS. Ông Thanh Dũng

ThS. Nguyễn Thành Sơn

ThS. Nguyễn Thanh Tuấn

17. Tiến trình cập nhật CCT

Lần 1: Nội Dung Cập nhật CCT lần 1: ngày/tháng/năm	<nguyên cập nhật ký và ghi rõ họ tên) Trưởng Bộ môn:
Lần 2: Nội Dung Cập nhật CCT lần 2: ngày/tháng/năm	<nguyên cập nhật ký và ghi rõ họ tên) Trưởng Bộ môn: