

## CHƯƠNG TRÌNH HỌC PHẦN

### 1. Thông tin chung về học phần

Tên học phần: Kho dữ liệu và khai phá dữ liệu

Mã học phần:

Số tín chỉ: 3

Học phần tiên quyết: Cơ sở dữ liệu, Hệ quản trị Cơ sở dữ liệu, Xác xuất thống kê

Đào tạo trình độ: Đại học

Giảng dạy cho ngành: Công nghệ Thông tin

Bộ môn quản lý: Hệ thống Thông tin

Phân bổ tiết giảng của học phần:

- Nghe giảng lý thuyết: 30 tiết
- Làm bài tập trên lớp: 5 tiết
- Thảo luận: 5 tiết
- Thực hành: 5 tiết (yêu cầu phòng máy có cài đặt SQL Server 2005 Standard)
- Tự nghiên cứu: 90 tiết

### 2. Mô tả tóm tắt học phần

Học phần trang bị cho người học những kiến thức về vấn đề kho dữ liệu và khai phá dữ liệu, rút trích tri thức từ các dữ liệu thô ban đầu. Các phương pháp tiền xử lý dữ liệu, tổ chức và xử lý dữ liệu đa chiều và kỹ thuật OLAP, các phương pháp khai phá dữ liệu như luật kết hợp, phân cụm, phân lớp và dự đoán...

### 3. Chủ đề và chuẩn đầu ra của học phần

#### 3.1. Danh mục chủ đề của học phần

1. Các phương pháp tiền xử lý dữ liệu
2. Kho dữ liệu và kỹ thuật OLAP trong khai phá dữ liệu
3. Khai phá các mẫu phổ biến
4. Phân lớp dữ liệu
5. Phân cụm dữ liệu
6. Khai phá dữ liệu với Hệ quản trị Cơ sở dữ liệu SQL Server

#### 3.2. Chuẩn đầu ra của quá trình dạy - học từng chủ đề của học phần

**Chủ đề 1:** Các phương pháp tiền xử lý dữ liệu

Nội dung	Mức độ
<b>Kiến thức</b> 1. Làm sạch dữ liệu 2. Tích hợp và chuyển đổi dữ liệu 3. Rút gọn dữ liệu 4. Rời rạc hóa dữ liệu	2

<b>Thái độ</b> 1. Kết quả khai phá dữ liệu phụ thuộc vào chất lượng và sự đầy đủ của dữ liệu do đó việc làm sạch, tích hợp, chuyển đổi dữ liệu là cần thiết 2. Khai phá dữ liệu xử lý trên lượng dữ liệu rất lớn. Việc rút gọn dữ liệu và rời rạc hóa làm giảm dung lượng dữ liệu nhưng ít ảnh hưởng đến chất lượng của kết quả khai phá dữ liệu	
<b>Kỹ năng</b> 1. Thực hiện một số biện pháp tiền xử lý dữ liệu	3

### Chủ đề 2: Kho dữ liệu và các kỹ thuật OLAP trong khai phá dữ liệu

Nội dung	Mức độ
<b>Kiến thức</b> 1. Các mô hình dữ liệu đa chiều 2. Các thao tác OLAP trong CSDL đa chiều 3. Kiến trúc nhà kho dữ liệu	2
<b>Thái độ</b> Trong khai phá dữ liệu và phân tích OLAP, chỉ sử dụng các thao tác chỉ đọc. Đề chú ý đến hiệu năng của hệ thống, dữ liệu cần phải được lưu trữ thích hợp	
<b>Kỹ năng</b> 1. Xây dựng lược đồ kho dữ liệu ứng với một chủ đề cụ thể	3

### Chủ đề 3: Khai phá các mẫu phổ biến

Nội dung	Mức độ
<b>Kiến thức</b> 1. Thuật toán Apriori 2. Thuật toán FP-GROW 3. Các dạng khác nhau của luật kết hợp 4. Các biện pháp cải thiện hiệu suất thuật toán Apriori	2
<b>Thái độ</b> 1. Một trong những ứng dụng của luật kết hợp là bài toán “giỏ hàng siêu thị”, dự đoán mặt hàng khách hàng sẽ mua dựa vào các mặt hàng đã có trong giỏ hàng. Ngoài ra luật kết hợp cũng có thể giải quyết được bài toán phân lớp	
<b>Kỹ năng</b> 1. Ứng dụng việc khai phá các mẫu phổ biến trong thực tế 2. Cài đặt được 2 thuật toán Apriori và FP-GROW	3

### Chủ đề 4: Phân lớp dữ liệu

Nội dung	Mức độ
<b>Kiến thức</b> 1. Các vấn đề đối với phân lớp và dự đoán 2. Phân lớp bằng phương pháp qui nạp trên cây quyết định 3. Phân lớp Bayes 4. Phân lớp bằng mạng Neural 5. Các mô hình hồi qui	2
<b>Thái độ</b> Phân lớp là một tác vụ cần thiết trong khai phá dữ liệu. Một số ví dụ về phân	

lớp như: Phân lớp khách hàng, phân lớp văn bản...	
<b>Kỹ năng</b> <ol style="list-style-type: none"> <li>1. Ứng dụng bài toán phân lớp trong thực tế</li> <li>2. Cài đặt được thuật toán ID3</li> <li>3. Xây dựng mô hình Bayes, mạng Neural cho bài toán phân lớp dữ liệu</li> </ol>	3

#### Chủ đề 5: Phân cụm dữ liệu

Nội dung	Mức độ
<b>Kiến thức</b> <ol style="list-style-type: none"> <li>1. Các dạng dữ liệu trong phân tích cụm</li> <li>2. Một số phương pháp phân cụm</li> <li>3. Thuật toán K-Mean</li> <li>4. Thuật toán K-Medoid</li> </ol>	2
<b>Thái độ</b> Phân tích cụm có nhiều ứng dụng rộng rãi, bao gồm nghiên cứu thị trường, nhận dạng mẫu, phân tích dữ liệu và xử lý ảnh. Trong kinh doanh, phân tích cụm có thể giúp các nhân viên tiếp thị khám phá sự khác nhau giữa các nhóm khách hàng dựa trên thông tin khách hàng và các đặc trưng của các nhóm khách hàng dựa trên các mẫu mua hàng. Trong sinh học, nó có thể được sử dụng để phân loại thực vật và động vật, các mẫu gen với các chức năng tương tự nhau.	
<b>Kỹ năng</b> <ol style="list-style-type: none"> <li>1. Cài đặt được các thuật toán K-Mean, K-Medoid</li> <li>2. Sử dụng phân cụm trong chuẩn hóa dữ liệu</li> </ol>	3

#### Chủ đề 6: Khai phá dữ liệu với Hệ quản trị CSDL SQL Server

Nội dung	Mức độ
<b>Kiến thức</b> <ol style="list-style-type: none"> <li>1. Giới thiệu tổng quan về khai phá dữ liệu trong SQL Server</li> <li>2. Ngôn ngữ DMX</li> <li>3. Xây dựng và sử dụng mô hình khai phá dữ liệu trong SQL Server</li> </ol>	2
<b>Thái độ</b> Hệ quản trị cơ sở dữ liệu SQL Server tích hợp sẵn các công cụ để triển khai một ứng dụng Data Mining gồm các thuật toán về Data mining và các dịch vụ: Analysis, Integration, Reporting, ngôn ngữ DMX. Thông qua SQL Server có thể triển khai các ứng dụng Data mining trên dữ liệu sẵn có của các ứng dụng.	
<b>Kỹ năng</b> <ol style="list-style-type: none"> <li>1. Xây dựng các mô hình khai phá dữ liệu trên cơ sở dữ liệu quan hệ kho dữ liệu</li> </ol>	3

#### 4. Phân bổ thời gian chi tiết

Chủ đề	Phân bổ số tiết cho hình thức dạy - học				
	Lên lớp			Thực hành, thực tập	Tự nghiên cứu
	Lý thuyết	Bài tập	Thảo luận		
Vấn đề 1	4		1		10

Vấn đề 2	4		1		10	
Vấn đề 3	6	2	1		18	
Vấn đề 4	6	2	1		18	
Vấn đề 4	5	1	1		14	
Vấn đề 6	5			5	20	

## 5. Tài liệu

TT	Tên tác giả	Tên tài liệu	Năm xuất bản	Nhà xuất bản	Địa chỉ khai thác tài liệu
1	Huỳnh Tuấn Anh	Bài giảng Data Warehouse And Data Mining		Lưu hành nội bộ	Liên hệ giáo viên
2	Jiawei Han	Data Mining Concepts And Techniques	2006	Morgan Kaufmann	Thư viện
3	Vincent Rainardi	Building a Data Warehouse With Examples in SQL Server	2008	Apress	Thư viện
4	ZhaoHui Tang, Jamie MacLennan	Data Mining with SQL Server 2005	2005	Wiley	Thư viện

## 6. Đánh giá kết quả học tập

TT	Các chỉ tiêu đánh giá	Phương pháp đánh giá	Trọng số (%)
1	Tham gia học trên lớp: chuẩn bị bài tốt, tích cực thảo luận...	Quan sát, điểm danh	10
2	Tự nghiên cứu: hoàn thành nhiệm vụ giảng viên giao trong tuần, bài tập nhóm/tháng/học kỳ...	Chăm báo cáo, bài tập...	10
3	Hoạt động nhóm	Trình bày báo cáo	10
4	Kiểm tra giữa kỳ	Viết	10
5	Kiểm tra đánh giá cuối kỳ	Viết	10
6	Thi kết thúc học phần	Viết, vấn đáp, tiểu luận....	50

**TRƯỞNG KHOA**

**TRƯỞNG BỘ MÔN**