

Fouille de données

NGUYỄN Thị Minh Huyền ©2016

huyenntm@hus.edu.vn

Plan

1. Introduction

2. Plan du cours

Plan

1. Introduction

2. Plan du cours

Responsable du cours

- NGUYEN Thi Minh Huyen
- Département de Mathématiques, de Mécanique et d'Informatique
Université des Sciences (Université nationale de Hanoi)
- Domaine de recherche : Traitement automatique des langues
⇒ Méthodes d'analyse linguistique basées sur les corpus

Introduction du cours

- Fouille de données : techniques et méthodes pour l'extraction automatique d'informations ou de connaissances nouvelles ou cachées dans des collections de données (qui peuvent être de très grande taille et rassembler différents types de média).
- Fouille de textes : Recherche d'information

Plan

1. Introduction

2. Plan du cours

Plan du cours

Cours

- Introduction à la fouille de données (*Data Mining*) et à l'analyse de données textuelles (*Text Mining*)
- Rappel de probabilités
- Exploration des données : cas d'une et de deux dimensions
- Exploration des données multidimensionnelles - Apprentissage non supervisé
- Exploration des données multidimensionnelles - Apprentissage supervisé
- Analyse de données textuelles

Plan du cours

TPs

- Exercices avec un logiciel de fouille de données à faire en binôme
- Exercices individuels
- Projet final en binôme
- Données

<http://archive.ics.uci.edu/ml/datasets.html>

Note finale : Exercices ($7 * 0,5$) + Projet final (6,5)