

Fouille de Données

HMIN326M

Konstantin Todorov - Pascal Poncelet

Konstantin.Todorov@lirmm.fr - Pascal.Poncelet@lirmm.fr
<http://www.lirmm.fr/~poncelet>
<http://www.lirmm.fr/users/utilisateurs-lirmm/konstantin-todorov>



Les informations sont disponibles sous Moodle

- Nom : Fouille de Données
- Code : HMIN326M

HMIN326M - Fouille de données

Accueil > Cours > Faculté des Sciences > Informatique > Département Informatique > M2 > S3 > HMIN326M - Fouille de données

Active le mode lecture

Votre progression



Forum des nouvelles

Planning des cours

Planning prévisionnel

Cours

Vous trouverez ici les copies des transparents qui sont présentés pendant les cours.
 The copy of the slides that are presented during the lectures.

Cours 1 : Le processus d'extraction de connaissances

Cours 2 : Extraction de règles d'association et de motifs séquentiels



Les informations sont disponibles sous Moodle

- Contenu des cours, des TD, des TP
- Devoir à rendre via Moodle
- Informations diverses via Moodle
- Utilisation du forum pour prévenir des changements



Les intervenants

- Konstantin Todorov, MCF UM
- Pascal Poncelet, Prof UM
- Dino Ienco, CR Irstea
- Cette année un autre intervenant pour les projets



Le planning prévisionnel

- Les cours ont lieu le mercredi matin

16/09/2020	9h45	Intro + KD	3h	P. Poncelet
23/09/2020	9h45	Règles d'Association	3h	P. Poncelet
30/09/2020	9h45	Recherche d'Information	3h	K. Todorov
07/10/2020	9h45	Clustering	3h	D. Ienco
14/10/2020	9h45	Classification	3h	K. Todorov
21/10/2020	9h45	Fouille de Texte, NER / TP	3h	K. Todorov
28/10/2020		VACANCES FDS		
04/11/2020	8h00	Ensemble Learning + TP	4,5h	D. Ienco
11/11/2020		FERIE		
18/11/2020	9h45	TP Projet	3h	K. Todorov
25/11/2020	9h45	TP Projet	3h	K. Todorov
02/12/2020	9h45	TP Projet	3h	K. Todorov
09/12/2020	9h45	TP Projet	3h	K. Todorov
16/12/2020	8h	Soutenances	6h	P. Poncelet, K. Todorov

Attention aux horaires – le plus possible modifications reportées sur PROSE
Information éventuellement via Moodle



Le programme

- Le processus d'extraction de connaissances
- Recherche de motifs
- Recherche d'information
- Classification et Clustering
- Fouille de textes
- Ensemble Learning
- Plateforme : WEKA ou Scikit-Learn



6



Un projet

- Cette année il y a un projet sur de la fouille de données textuelles
- Thématique : détection d'opinions
- Les données sont fournies pour faciliter les traitements et se focaliser sur la partie fouille
- Un challenge pour comparer les approches ?
- Evaluation : document expliquant les prétraitements, choix des représentations, choix des descripteurs, choix des classifieurs, analyse des résultats, codes ...



7



Evaluation

- 2 notes :
 - 1 note individuelle : article à lire et réponse à des questions
 - 1 note commune (Projet avec soutenance)
 - Attention en fonction de la présentation, les notes peuvent être différentes entre les membres du groupe
- **Note finale : 100% CCI**
- Pour le projet : remise du code, des données, des résultats et du rapport



Contactez les enseignants

- Les adresses mails :
 - Pascal.Poncelet@umontpellier.fr
 - Konstantin.Todorov@umontpellier.fr
 - Dino.Ienco@irstea.fr
- S'il vous plaît mettre dans le sujet :
- [HLIN326]
- Utilisez vos adresses institutionnelles (SPAM)



9



La plateforme utilisée

- WEKA
- Plateforme libre service qui dispose de nombreux algorithmes
- Java
- Besoin d'autres logiciels :
- Ils seront indiqués ...



10



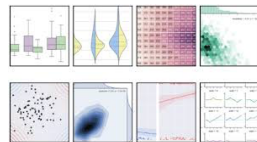
Scikit-Learn



- Possibilité d'utiliser

Scikit-Learn pour le projet

- Scikit learn : bibliothèque libre Python destinée à l'apprentissage automatique
- Pandas, SciPy : pour manipuler les données
- Seaborn pour visualiser les statistiques sur les données



11