

INF344 – Données du Web

Présentation générale

Antoine Amarilli



Responsable du cours : Antoine Amarilli

- Maître de conférences à Télécom Paris
- Équipe DIG
- <https://a3nm.net/>

Modalités d'enseignement à distance

- Supports de cours, TP et rendus sur **Moodle** :
<https://moodle.r2.enst.fr/moodle/course/view.php?id=87>
(identifiants Télécom Paris)
- Discussion sur **Mattermost** : <https://mattermost.a3nm.net/>
(compte à créer avec le lien d'invitation reçu par mail)
- Cours et visio pour les TP sur **BigBlueButton** :
<https://digimedia1.r2.enst.fr/b/ant-4xn-jpy>
(pas besoin de compte; horaires habituels)

Modalités à ajuster en fonction de comment ça se passe, n'hésitez pas à nous faire des retours!

Structure générale du cours

Introduction à la **gestion de données du Web** en allant des **données brutes** vers les **connaissances structurées**

- Réseau, HTTP, HTML, CSS
- Sécurité des données
- Technologies côté client et côté serveur
- Traitement parallèle de données
- Traitement du langage
- Extraction d'information
- Crawling et scraping
- Désambiguation
- Analyse des opinions
- Web sémantique

Différents intervenants pour chacune de ces thématiques

Modalités d'évaluation

- Chaque intervenant vous fera réaliser un **TP**
- Le premier TP est en **temps limité**
- Pour les autres : échéance de rendu à la fin de la semaine
- Encadrement sur les TP sur **Mattermost** et par visio
 - pendant les horaires de séance de TP
 - pendant les “office hours” (cours INF731)
- Votre note pour INF344 est la moyenne des **6 meilleurs TP** (sur les 8)
- Il n’y a **pas de rattrapage** pour les TP ou pour le cours en général

Rendu des TP : **exclusivement sur Moodle**

Mise en garde sur le plagiat

- La validation du cours est **individuelle**, pas **collective**
- Discuter des sujets de TP est autorisé, mais le code que vous rendez doit avoir été **intégralement écrit** par vous
- Si vous utilisez des sources externes (fragments de code, etc.), il faut **clairement l'indiquer**
- Il est **strictement interdit** :
 - De **partager** votre code avec d'autres
 - De **mettre en ligne** votre code (pendant la durée du cours)
 - De **rendre** du code adapté de quelqu'un d'autre

Attention : quand des rendus sont plagiés, tous les rendus concernés reçoivent la note de 0.

(Chaque année, certains d'entre vous ne valident pas à cause de rendus plagiés...)

SONDAGE : Plagiat

Pour les TP notés du cours INF344, je peux...



- **A:** Demander son code à un camarade et m'en inspirer
- **B:** Partager mon code avec les autres
- **C:** Mettre mon code sur Github
- **D:** Rien de tout ça

SONDAGE : Plagiat

Pour les TP notés du cours INF344, je peux...



- **A:** Demander son code à un camarade et m'en inspirer
- **B:** Partager mon code avec les autres
- **C:** Mettre mon code sur Github
- **D: Rien de tout ça**