



**Master « Mathématiques Appliquées, Statistique »**  
*parcours-type*

**Science des données pour la décision publique**

**Majeure Statistique et traitement des données (STD)**

**ANNÉE SCOLAIRE 2023 / 2024**



École nationale  
de la statistique  
et de l'analyse  
de l'information

délivré par



Campus de Ker Lann

Rue Blaise Pascal

BP 37203

35172 Bruz cedex

Tél. : 02 99 05 32 32

[www.ensai.fr](http://www.ensai.fr)

en cohabilitation avec



Faculté des sciences économiques

Campus centre

7 place Hoche

35065 Rennes cedex

Tél. : 02 23 23 35 45

[www.eco.univ-rennes1.fr](http://www.eco.univ-rennes1.fr)

<b>PRESENTATION GENERALE DES ENSEIGNEMENTS .....</b>	<b>4</b>
UE1 MACHINE LEARNING .....	10
<i>APPRENTISSAGE STATISTIQUE</i> .....	10
<i>APPRENTISSAGE PROFOND</i> .....	11
<i>WEBMINING ET TRAITEMENT DU LANGAGE</i> .....	12
UE2 DEVELOPPEMENT D'APPLICATION ET DATAOPS.....	13
<i>GENIE LOGICIEL</i> .....	13
<i>DEVELOPPEMENT WEB</i> .....	15
<i>DATAOPS</i> .....	16
UE3 BIG DATA .....	17
<i>TECHNOLOGIES NOSQL ET CLOUD</i> .....	17
<i>INDEXATION WEB</i> .....	18
<i>PUBLICATION DE DONNEES RESPECTUEUSES</i> .....	20
UE4 SYSTEME ET RESEAUX .....	21
<i>INITIATION A UNIX</i> .....	21
<i>RESEAUX ET SYSTEME D'EXPLOITATION</i> .....	22
<i>SYSTEMES REPARTIS</i> .....	23
<i>SECURITE DES DONNEES</i> .....	24
UE5 FORMATION ANALYSTE .....	25
<i>GESTION DE PROJET</i> .....	25
<i>ARCHITECTURE INFORMATIQUE</i> .....	26
<i>ENVIRONNEMENT INFORMATIQUE MINISTERIEL ET INTERMINISTERIEL</i> .....	27
UE6 FORMATION AVANCEE A L'EXAMEN D'ANALYSTE .....	28
<i>PREPARATION A L'EXAMEN ORAL</i> .....	28
<i>PREPARATION A L'EXAMEN ECRIT</i> .....	29
<i>ANGLAIS</i> .....	30
UE7 PROJET PROFESSIONNEL ET STAGES .....	31
<i>STAGE DE FIN D'ETUDES</i> .....	31

# Présentation générale des enseignements

---

## Objectifs

Les acteurs publics disposent de bases de données locales, nationales voire internationales, dont la multiplication a été favorisée par le développement des systèmes d'information ; ils ont donc besoin de statisticiens qualifiés au plus haut niveau pour les concevoir, les analyser et les exploiter. Le but de ce parcours-type du master « mathématiques appliquées -statistiques » est précisément de répondre à cette demande croissante en expertise statistique pour l'aide à la décision publique.

Ce master bénéficie du label European master in official statistics (EMOS) délivré par Eurostat.

## Quels métiers ? Quels secteurs d'activité ?

D'une manière générale, la formation reçue prépare les diplômés aux métiers scientifiques et techniques centrés sur l'ingénierie statistique, les systèmes d'information économique et sociale et les métiers informatiques liés à la statistique publique.

Elle répond tout d'abord aux besoins des services publics, administrations nationales ou collectivités locales. Les principaux ministères (économie, agriculture, développement durable, santé et action sociale, etc.), des établissements publics (notamment les caisses de sécurité sociale et les organismes gestionnaires de prestations), les organismes consulaires et les agences dépendant des collectivités locales sont susceptibles de recourir aux compétences des titulaires de ce master. Ils pourront aussi trouver des opportunités dans des organismes européens (Eurostat) et internationaux (OCDE, FMI, Banque mondiale).

Par ailleurs, les décideurs publics, notamment au niveau local, peuvent s'appuyer sur des cabinets de conseil spécialisés dans la décision publique : ces cabinets peuvent ainsi également offrir des débouchés aux titulaires du master de statistique publique.

Enfin, les attachés statisticiens de l'Insee peuvent compléter leur formation initiale par ce master.

## Conditions d'admission

La formation est ouverte aux élèves fonctionnaires de l'Ensaï admis à la formation complémentaire diplômante, ainsi qu'aux étudiants ayant suivi la première année du master ; la formation est également ouverte aux diplômés de niveau équivalent. La sélection se fait par examen du dossier et un entretien. La formation est également ouverte en formation continue.

Calendrier de candidature : d'avril à fin mai.

Pour tout renseignement sur l'admission : [admission-statpublique@ensai.fr](mailto:admission-statpublique@ensai.fr)

## **Responsables de la formation**

A l'université Rennes 1 : Isabelle Cadoret : [isabelle.cadoret-david@univ-rennes1.fr](mailto:isabelle.cadoret-david@univ-rennes1.fr)

A l'Ensai : Stéphane Legleye : [stephane.legleye@ensai.fr](mailto:stephane.legleye@ensai.fr)

Formation continue diplômante des attachés (FCDA) et scolarité du master : Aurélie Duchesne et Stéphane Legleye : [aurelie.duchesne@ensai.fr](mailto:aurelie.duchesne@ensai.fr) et [stephane.legleye@ensai.fr](mailto:stephane.legleye@ensai.fr).

## **Organisation des études**

Beaucoup d'enseignements sont associés à la réalisation de travaux de groupes et de travaux personnels sous forme de dossiers à rendre.

La professionnalisation est au cœur du master. La majorité des enseignements est dispensée par des acteurs du système statistique public.

L'anglais est obligatoire en formation initiale.

	Volume horaire				Crédits
	Cours	Ateliers	Projets	Total	
<b>UE1 Machine Learning</b>					
Apprentissage statistique	21	12		33	2,5
Apprentissage profond	12	12		24	2
Webmining et traitement du langage	9	12		21	1,5
<b>Total</b>	<b>42</b>	<b>36</b>		<b>78</b>	<b>5</b>
<b>UE2 Développement d'application et DataOps</b>					
Génie logiciel	24	24		48	3
Développement web	12	15		27	2
Ou Dataops	12	15		27	2
<b>Total</b>	<b>36</b>	<b>39</b>		<b>75</b>	<b>5</b>
<b>UE3 Big data</b>					
Technologies NoSQL et Cloud	18	12		30	2
Indexation web	9	6		15	1
Publication de données respectueuse de la vie privée	9	6		9	1
<b>Total</b>	<b>36</b>	<b>18</b>		<b>54</b>	<b>4</b>
<b>UE4 Système et réseaux</b>					
Initiation à Unix	9	6		15	0
Réseaux et systèmes d'exploitation	15	6		21	1,5
Systèmes Répartis	15	6		21	1,5
Sécurité des données	9	9		18	1
<b>Total</b>	<b>48</b>	<b>27</b>		<b>75</b>	<b>4</b>
<b>UE5 Formation analyste</b>					
Gestion de projet	24			24	2
Architecture informatique	12	12		24	2
Environnement informatique ministériel / interministériel	18			18	2
<b>Total</b>	<b>54</b>	<b>12</b>		<b>66</b>	<b>6</b>
<b>UE6 Formation avancée à l'examen d'analyste</b>					
Préparation à l'examen oral	18	9		27	2
Préparation à l'examen écrit	18	48		66	3
Anglais		30		30	1
<b>Total</b>	<b>0</b>	<b>87</b>	<b>0</b>	<b>123</b>	<b>6</b>
<b>UE7 Projet professionnel et stages</b>					
Stage de fin d'études					25
Stage d'application					5
Séminaires et projet professionnels	30			30	0
<b>Total</b>	<b>30</b>			<b>30</b>	<b>30</b>
<b>Sport</b>		30		30	0
<b>TOTAL</b>	<b>246</b>	<b>219</b>	<b>0</b>	<b>495</b>	<b>60</b>

Pour les élèves ne suivant pas le cursus d'analyste :

<b>UE5 Compléments de Machine Learning</b>				
Apprentissage par renforcement	15	0	15	1
Apprentissage à grande échelle	6	6	12	1
Systèmes de recommandation	6	6	12	1
Régularisation	9	6	15	1
Traitement du langage avancé	0	12	12	1
<b>Total</b>	<b>36</b>	<b>30</b>	<b>66</b>	<b>5</b>

<b>UE6 Projet</b>				
Projet méthodologique	9	18	27	1.5
Projet de fin d'étude	9	36	45	3.5
Data Challenge	12		12	0
Anglais	30		30	1
<b>Total</b>	<b>60</b>	<b>54</b>	<b>114</b>	<b>6</b>

## **Contrôle des connaissances**

L'évaluation des enseignements est précisée dans chaque présentation d'enseignement.

## **Règlement intérieur**

Le règlement intérieur de la scolarité est celui de l'Ensaï.

## APPRENTISSAGE STATISTIQUE

### Machine Learning

<i>Enseignant</i>	: Sébastien Da Veiga
<i>Nombre d'ECTS</i>	: 2,5
<i>Répartition des enseignements</i>	: 18h de cours et 12h d'ateliers
<i>Langue d'enseignement</i>	: Anglais (cours) et Français (ateliers)

#### Modalités d'évaluation :

- 1 compte-rendu de TP
- Examen final

#### Acquis d'apprentissage (objectifs) :

Comprendre les différents modèles de l'état de l'art (modèle linéaire: régression et classification, pénalisation, méthodes locales, SVM, forêt) ; Connaître les cas d'usage de ces modèles ; Savoir comparer empiriquement différents modèles pour une tâche donnée ; Savoir implémenter les méthodes étudiées en Python ou R

#### Principales notions abordées :

Supervised learning; Regression; Classification; Empirical risk minimization; Model evaluation; Cross validation; Functional approximation; Model complexity; Large scale optimization; Stochastic gradient descent; Regularization; RIDGE and LASSO; Support Vector Machine; Kernel trick; Ensemble methods; Aggregation and Boosting; Random forest.

#### Références bibliographiques :

- T. HASTIE, J. FRIEDMAN and R. TIBSHIRANI. *The elements of Statistical Learning: Data Mining, Inference and Prediction* (2nd ed.), 2009
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013), *An introduction to statistical learning*, New York: springer.
- Statistical learning with sparsity: the lasso and generalizations, T Hastie, R Tibshirani, M Wainwright – 2019 Link: <https://web.stanford.edu/~hastie/StatLearnSparsity/>

## APPRENTISSAGE PROFOND

### *Deep Learning*

*Enseignant* : Ikko Yamane (Ensaï)

*Nombre d'ECTS* : 1

*Répartition des enseignements* : 24h de cours

*Langue d'enseignement* : Français

### **Modalités d'évaluation :**

1 Quizz et 1 compte-rendu de TP

### **Acquis d'apprentissage (objectifs) :**

- identifier une tâche particulièrement adaptée pour un réseau de neurones et/ou un réseau de neurones profond
- identifier et implémenter un réseau de neurone approprié pour un tâche d'apprentissage supervisé donnée (architecture, fonction de coût, méthode d'optimisation)
- utiliser et spécialiser un réseau de neurones pré-entraîné

### **Principales notions abordées :**

Les réseaux de neurones profonds sont au cœur d'avancées rapides en traitement d'image et de la langue depuis les années 2010. Ce cours présente ces modèles, leur fonctionnement, ainsi que comment les utiliser.

- Principe des réseaux de neurones
- Propriétés des réseaux de neurones simples
- Descente de gradient
- Réseaux de neurones profonds
- Architectures particulières : réseaux à convolution ; réseaux récurrents.

### **Références bibliographiques :**

- Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep Learning. 2016

# WEBMINING ET TRAITEMENT DU LANGAGE

*Natural language processing and webmining*

*Enseignant* : Guillaume Gravier (Irisa)

*Nombre d'ECTS* : 1,5

*Répartition des enseignements* : Cours : 9h ☒ Atelier : 12h

*Langue d'enseignement* : Anglais

## Modalités d'évaluation :

Projet

## Acquis d'apprentissage (objectifs) :

- collecter des données, extraire de l'information et apparier des sources textuelles
- choisir une méthode de traitement automatique de la langue pour une tâche classique (classification, analyse de sentiment, détection > d'entités...)
- se repérer parmi le foisonnement des modèles d'étude de la langue

## Principales notions abordées :

1. What's natural language and its processing
2. The representation of words
3. The representation and classification of documents
4. Language modeling and contextual word embedding
5. Sentence-level tagging (token level tasks)
6. Sequence to sequence models and transformers
7. Overview of standard NLP tasks today

## Références bibliographiques :

- Daniel Jurafsky, James H. Martin. *Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition*, 2nd edition, Prentice-Hall, 2009. Draft of the 3rd edition partly available at <https://web.stanford.edu/~jurafsky/slp3>.
- Yoav Goldberg. *Neural Network Methods for Natural Language Processing*. 2017. An earlier draft is freely available online at <http://u.cs.biu.ac.il/~yogo/nnlp.pdf>.
- Kevin Gimpel's lectures (Toyota Technological Institute at Chicago and UChicago) on *Natural Language Processing* (<https://ttic.uchicago.edu/~kgimpel/teaching/31190-s18/index.html>) and on *Advanced Natural Language Processing* (<https://ttic.uchicago.edu/~kgimpel/teaching/31210-s19/index.html>).

## GENIE LOGICIEL

### Software engineering

<i>Enseignant</i>	: Djamel Eddine KHELLADI (Université Rennes 1) et Alban GRÉAU
<i>Nombre d'ECTS</i>	: 3
<i>Répartition des enseignements</i>	: Cours : 24h ☐ Atelier : 24h
<i>Langue d'enseignement</i>	: Français

### Modalités d'évaluation :

Un TP noté sera rendu

### Acquis d'apprentissage (objectifs) :

L'objectif de ce cours est d'introduire les moyens de concevoir des applications informatiques de qualité (répondant aux besoins, évolutives et faciles à maintenir).

Le cours présente une introduction aux modèles de conception classiques, base du génie logiciel autour des technologies objet, en proposant des applications pratiques au cours de travaux pratiques et en étudiant des patrons de conception développés en Java. Cet enseignement vise également à apprendre à développer et déployer un site Web dynamique en Java en utilisant Spring Boot. Il permet de se familiariser avec les architectures n-tiers et les serveurs d'applications et de bien maîtriser les principaux outils et langages avancés de développement des applications Web/JavaEE.

L'autre partie de ce cours se concentrera sur les méthodes de packaging et de livraison d'application pour se rapprocher des standards de l'entreprise. Pour java on le cours se concentrera sur Maven, python ne disposant pas d'outil de livraison car le langage est simplement interprété on parlera rapidement de pip et des venv. Enfin le cours s'attardera sur les technologies de conteneurisation avec l'exemple de Docker et d'orchestration avec Docker Compose et Kubernetes.

### Principales notions abordées :

- I. Le génie logiciel
  1. Introduction au génie logiciel et bonne pratique de conception.
  2. Architecture logicielle et modèle en couche, Exemple sur GWT.
  3. Principaux patrons de conception, principe et mise en œuvre en Java.
  4. Le test logiciel et l'ingénierie des langages.
  5. L'ingénierie dirigée par les modèles.
- II. Programmation Client Serveur (JavaEE)
  1. Architectures distribuées et plate-forme JavaEE
    - i. Le framework SpringBoot
    - ii. Architecture : composants, services et communications
    - iii. Les problématiques des applications serveurs
  2. La persistance avec JPA
    - i. Problématique du "mapping" objet-relationnel
    - ii. Les outils de mapping : JPA, Hibernate
    - iii. Le mapping
    - iv. L'entity-manager
- III. Packaging et livraison
  1. Maven
  2. Docker
  3. Docker Compose
  4. Kubernetes

## Références bibliographiques :

- I. SOMMERVILLE, *Le Génie logiciel*, Addison Wesley-France, 1988
- B. BEIZER, *Software Testing Techniques*, Second Edition, Van Norstrand, 1990
  
- B.W. BOEHM, *Software Engineering Economics*, Prentice-Hall, 1981
- E. GAMMA, R. HELM, R. JOHNSON, J. VLISSIDES, *Design patterns, catalogue de modèles de conception réutilisables*, Vuibert, 2007
- GOUIGOUX, J. P. (2021). Docker : Concepts fondamentaux et déploiement d'applications distribuées. Éditions ENI

## DEVELOPPEMENT WEB

### Software engineering

<i>Enseignant</i>	: Olivier Chantrel
<i>Nombre d'ECTS</i>	: 2
<i>Répartition des enseignements</i>	: Cours : 12h et Atelier : 15h
<i>Langue d'enseignement</i>	: Français

### Modalités d'évaluation :

Les étudiants sont évalués sur la base du rapport d'étude et d'une soutenance devant un jury, incluant une démonstration de leur réalisation

### Acquis d'apprentissage (objectifs) :

L'objectif de cet enseignement est d'effectuer un projet de bout en bout. Ce projet commence par une modélisation utilisant les méthodes et techniques vues en génie logiciel et se termine par une implémentation en JavaEE.

Ce cours vise à donner aux étudiants une vision détaillée du web d'aujourd'hui en présentant les technologies historiques du web (html, xml) et les technologies plus récentes du web 2.0 (css, javascript, HTML5, Ajax, Php, MySql).

### Principales notions abordées :

- Encadrement en début de projet afin de préciser les besoins et pour déterminer l'architecture générale du programme.
- Encadrement distant sur des questions techniques ponctuelles
- Encadrement technique lors de la phase d'implémentation
- Soutenance des projets
- L'historique du web / XML et ses applications (html, xml, dtd, web services etc.)
- Le web actuel (css, javascript, HTML5, Ajax, MySql, Php)

### Références bibliographiques :

- L. Roland, « Structurez vos données avec XML », 2014
- L Van Lancker, « AJAX - Développez pour le Web 2.0 - Entrez dans le code : JavaScript, XML, DOM, XMLHttpRequest2... (2ième édition) », 2015
- C Pierre de Geyer & E Daspect, « PHP5 avancé », 2012

## DATAOPS

### *Dataops*

<i>Enseignant</i>	: Christophe Blefari
<i>Nombre d'ECTS</i>	: 2
<i>Répartition des enseignements</i>	: Cours : 27h
<i>Langue d'enseignement</i>	: Anglais

### **Modalités d'évaluation :**

Written exam and short projet.

### **Acquis d'apprentissage (objectifs) :**

This classe aims to introduce to the DataOps skills set and philosophy. DataOps isa new concept that appeared around 2018. Because companies started to treat data as a product we needed to approach data in a more agile way. In order to simplify data usage for all, engineers need to treat data like an asset. This is DataOps.

During this classe students will see and create self-service data platform through concepts like Infrastructure as Code B IaC — Terraform, Ansible), CI/CD and Software Engineering abstraction.

This lecture will also help you navigate through the whole data tools ecosystem and will give you the decision framework to choose a tool over another.

### **Principales notions abordées :**

- Why DataOps, MLOps and DevOps?
- AWS Cloud platform, a reminder
- Infrastructure as Code
- CI/CD concepts
- Docker 101 and how it changed applications deployment
- Kubernetes light introduction
- Data pipelines and data observability

### **Références bibliographiques :**

- DataOps 101 — DataTalks.Club podcast with Lars Albertsson ; <https://datatalks.club/podcast/s02e11-dataops.html>
- DataOps Manifesto ; <https://dataopsmanifesto.org/fr/>
- Getting DataOps Right — O'Reilly Media, Inc. ; <https://www.oreilly.com/library/view/getting-dataops-right/9781492031765/>

## TECHNOLOGIES NOSQL ET CLOUD

### *NoSQL technologies and cloud*

*Enseignant* : Shadi Ibrahim

*Nombre d'ECTS* : 2

*Répartition des enseignements* : Cours : 30h

*Langue d'enseignement* : Français

### **Modalités d'évaluation :**

Examen écrit

### **Acquis d'apprentissage (objectifs) :**

Cet enseignement vise à donner aux étudiants les connaissances de base sur les architectures distribuées spécialisées dans le stockage et le traitement du Big Data. Les hypothèses, concepts and algorithmes seront détaillés pour chacune d'entre elles.

L'objectif du cours est double. Il permettra aux élèves d'avoir une connaissance des systèmes sur Cloud disponibles actuellement et de pointer les directions futures de ces architectures et de savoir s'orienter dans la jungle des technologies NoSql pour trouver la solution qui répondra le plus à leurs besoins. Ces bases de données se distinguent des approches classiques relationnelles. Ces approches abandonnent la représentation matricielle de l'information ainsi que le langage SQL au profit d'une plus grande simplicité, d'une meilleure performance et d'une meilleure scalabilité.

### **Principales notions abordées :**

- Système de fichiers distribué avec HDFS
- Traitement distribué avec Hadoop MapReduce
- Virtualisation et conteneurisation

### **Références bibliographiques :**

- G. PLOUIN, Cloud computing et SaaS, Editions Dunod
- Le livre blanc du Cloud, du SaaS et des Managed Services pour les partenaires IT et télécoms. Edition 2013
- R. HENNION, H. TOURNIER, E. BOURGEOIS, Cloud computing : Décider - Concevoir - Piloter - Améliorer, Editions Eyrolles, 2012

# INDEXATION WEB

## Web Datamining

<i>Enseignant</i>	: Lara Perinetti
<i>Nombre d'ECTS</i>	: 1
<i>Répartition des enseignements</i>	: 9h de cours + 6h d'ateliers
<i>Langue d'enseignement</i>	: Français

### Modalités d'évaluation :

Projet par groupe d'élèves.

### Acquis d'apprentissage (objectifs) :

A l'issue de ce cours, les élèves devront savoir collecter des informations issues du web, connaître la notion d'Information Retrieval, savoir constituer des corpus, et les organiser à des fins d'analyse exploratoires. Ils devront maîtriser également l'algorithme qui permet de hiérarchiser les pages web (pagerank) et les techniques de classification de documents textuels.

Par ailleurs, ils devront avoir acquis les notions d'opinion mining (classification de textes, analyses de sentiments, évaluation de modèles).

Toutes les applications seront traitées en R.

### Principales notions abordées :

Partie 1 - Information Retrieval : Preprocessing, Extraction and PageRank

*Mots clés : twitter, R, pagerank, corpus, term-document matrix, Information retrieval, tf-idf, stemming, Regex, kmeans*

Partie théorique (3h)

- Information Retrieval
  - o Concepts & Définitions
  - o Term Document Matrix
  - o Tf-idf, Cosine Index, jaccard Index
  - o Stemming
- Web Search : Google
  - o Google et le Page Rank
  - o Pages Jaunes (Notion de tri alpha)
  - o Notion de graphes et de vecteurs propres

Partie pratique (9h)

- TP1 : Introduction à R pour le Web Mining (3h)
  - o Installation de bibliothèques de textmining disponible dans R
  - o Collecter les informations issues du WEB : Twitter, Wikipedia
  - o Pre-processing : Stemmatisation, Lemmatisation,
  - o Parsing HTML, XML,
  - o Tokenization
  - o Introduction à la term-document matrix
- TP2 : Similarité de documents (Applications aux recherches utilisateurs sur le site pagesjaunes.fr (3h)
  - o Indices de similarité : Tf, tf-idf Jaccard, Cosine
  - o Distance de Damerau, Distance de jaro
  - o Liens entre les recherches, Notion de graphe de recherche
- TP3 : Ordonnement des résultats d'une recherche (3h)

- o PageRank
- o Détecter les mots clés
- o Intro à la classification des docs sur mots clés

Partie 2 - Opinion Mining : Textmining, analyse de sentiments, classification et évaluation des modèles.

*Mots clés : Facebook, R, opinion mining, corpus, sentiment analysis, annotation syntaxique.*

Partie théorique (4h)

- Introduction
  - o Quelles applications dans quels domaines d'activités
- État de l'art (opinion mining, sentiment analysis, affective computing) o Quels descripteurs pour quels types de données ?
  - Textuelles
  - Audio
  - Images
- o Sélection automatique de descripteurs (réduction de l'espace de recherche)
  - o Quels algorithmes de classification dans quels cas ?
- Constitution du corpus
  - o Réflexions générales sur la qualité des données et son impact
  - o Annotation manuelle et automatique (schéma d'annotation, calcul d'un score d'agrément inter-annotateur,)
  - o Répartition des données dans les classes
- Pre-processing (texte)
  - o Quelle granularité pour mes données (mot, phrases, paragraphes)
  - o Annotation syntaxique et sémantique (exemples de POS, WordNet-Affect, etc)
- Évaluation
  - o Quelles mesures utiliser pour mesurer la qualité d'un modèle (rappel, précision, f-score, ROC, indices de confiance a 0.95)
- Les produits du marché (exemples)
  - o Produit de la société TEMIS (cartouche sentiments)
  - o Produit de la société Sinequa

Partie pratique (8h)

- TP1 : classification de la valence d'un texte littéraire (critiques de cinéma)
- TP2 : classification de la valence de textes issus de réseaux sociaux (twitter, facebook)
- TP3 : Fusion de modèles (à partir des modèles créés dans le TP2)
- TP4 (optionnel) : Constructions de modèles à partir d'indices multimodaux (texte + audio)

## Références bibliographiques :

Les \* indiquent les lectures fortement conseillées.

- Web DataMining, Exploring Hyperlinks, Contents, and Usage Data, Bing Liu, Springer (Chapitre 6 à 13)
- (\*) - Information Retrieval, <http://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf> (chapitres 1-3) (\*)
- **package tm in R**, <http://cran.r-project.org/web/packages/tm/vignettes/tm.pdf> (\*)
- Infrastructure of Textmining with R, <http://www.jstatsoft.org/v25/i05/paper>

## PUBLICATION DE DONNEES RESPECTUEUSES

### *Privacy preserving data publishing*

<i>Enseignant</i>	: Tristan Allard
<i>Nombre d'ECTS</i>	: 1
<i>Répartition des enseignements</i>	: Cours : 9h de séminaire
<i>Langue d'enseignement</i>	: Français

### Modalités d'évaluation :

Rapport

### Acquis d'apprentissage (objectifs) :

« Les données personnelles sont le nouveau pétrole d'Internet et la nouvelle monnaie du monde numérique » a déclaré M. Kouneva, commissaire européen à la protection des consommateurs en mars 2009. La valeur de l'analyse massive des données personnelles pour les industriels, les scientifiques et la société en général est largement reconnue aujourd'hui. Cependant, leur caractère personnel et potentiellement sensible est un obstacle majeur à leur partage à grande échelle. L'objectif des modèles et algorithmes de publication de données respectueuse de la vie privée est précisément d'offrir des garanties fortes de respect de la vie privée tout en autorisant un partage de qualité à des fins d'analyse. La tâche est loin d'être triviale comme l'ont démontré plusieurs scandales de ré-identification. L'objectif de ce cours est de présenter aux étudiants les principaux paradigmes et techniques de publication de données respectueuse de la vie privée.

L'accent sera particulièrement mis sur un modèle prééminent aujourd'hui : la differential privacy.

### Principales notions abordées :

Introduction : motivation, défis, survol

Paradigmes : non-informatif, *differential privacy*

Publication interactive: modèles type *differential privacy*, mécanismes principaux de perturbation interactive (e.g., Laplace)

Perturbation locale : le mécanismes des réponses randomisés pour satisfaire la *differential privacy*

Publication centralisée : mécanismes de génération de données synthétiques satisfaisant la *differential privacy*, survol des modèles basés sur le partitionnement (e.g., *k*-anonymat, *l*-diversité) et des mécanismes principaux pour les satisfaire (e.g., algorithme de Mondrian)

Conclusion : les pratiques « dans le monde réel », questions ouvertes

### Références bibliographiques :

- B.-C. Chen, D. Kifer, K. LeFevre, et A. Machanavajjhala, Privacy-Preserving Data Publishing, Found. Trends databases, vol. 2, no 1-2, p. 1-167, 2009.

- C. Dwork et A. Roth, The Algorithmic Foundations of Differential Privacy, Found. Trends Theor. Comput. Sci., vol. 9, no 3-4, p. 211-407, 2014.

- B. C. M. Fung, K. Wang, R. Chen, et P. S. Yu, Privacy-preserving data publishing : A survey of recent developments, ACM Comput. Surv., vol. 42, no 4, p. 14:1-14:53, 2010.

## INITIATION A UNIX

### *Networks end Systems*

<i>Enseignant</i>	: Sébastien Le Corre
<i>Nombre d'ECTS</i>	: 0
<i>Répartition des enseignements</i>	: 15h de cours + 6h d'ateliers
<i>Langue d'enseignement</i>	: Français

#### **Modalités d'évaluation :**

Aucune

#### **Acquis d'apprentissage (objectifs) :**

Il s'agit d'un atelier intense pendant lequel les étudiants vont installer une version récente de Linux et apprendre à manipuler ce système d'exploitation afin de l'utiliser tout au long de l'année.

Linux est en particulier central pour utiliser et développer les technologies Big Data.

#### **Principales notions abordées :**

1. Présentation d'UNIX (Linux en particulier)
2. Découverte pratique (GUI, Ligne de commande, Bash)
3. Gestion de la machine (installation & configuration de logiciels)
4. Aperçu des versions "Cloud" (spécificités "rapides")
5. Scripting (Bash, Base de Awk & Sed)

#### **Références bibliographiques :**

- Bash : <https://tldp.org/LDP/abs/html/index.html>
- AWS EC2 : [https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EC2\\_GetStarted.html](https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EC2_GetStarted.html)
- GCP Compute : <https://cloud.google.com/compute/docs/create-linux-vm-instance>
- O'Reilly - Learning Modern Linux par ichael Hausenblas

(<https://www.oreilly.com/library/view/learning-modern-linux/9781098108939/>)

# RESEAUX ET SYSTEME D'EXPLOITATION

## Computer Networks and Operating System

<i>Enseignant</i>	: Thomas Correge
<i>Nombre d'ECTS</i>	: 1.5
<i>Répartition des enseignements</i>	: 15h de cours + 6h d'ateliers
<i>Langue d'enseignement</i>	: Français

### Modalités d'évaluation :

Examen écrit

### Acquis d'apprentissage (objectifs) :

This course aims to provide students with an understanding of the core principles of technologies constituting the foundation of the IT world: operating systems and computer networks.

In the first part, we will study the way an operating system organizes and facilitates the interaction of its key resources such as processor, memory, and file system in a multi-tasking and multi-user context.

The second part will focus on networks and will address various topics, such as network topology and technologies, Ethernet, ADSL, LAN, WAN, VLAN, Internet, Wifi and secure Wifi, TCP/IP layers, major protocols (DNS, SMTP...), network devices, architecture designs (dimensioning, redundancy, segmentation, DMZ.).

Implications for the security of the Information System will also be touched when addressing these topics, in order to raise awareness about inherent security risks and relevant countermeasures.

### Principales notions abordées :

1. Operation Systems overview	1. Introduction
2. Operation Systems overview	2. Host-network layer
3. Processes	3. Internet layer
4. Inter-process communication	4. Transport layer
5. Memory management	5. Application layer
6. Processes scheduling	
7. File systems	6. Architecture review
8. Disk management systems (RAID)	
9. Virtualization	

### Références bibliographiques :

- Modern Operating Systems. Andrew Tanenbaum. Pearson Education. 4th edition (2014). ISBN-13: 978 0133591620 ISBN-10: 013359162X
- Computer networks. Andrew Tanenbaum & David Wetherall. Pearson. 5th edition (2010). ISBN-13: 978 0132126953 ISBN-10: 0132126958

## SYSTEMES REPARTIS

### Networks

<i>Enseignant</i>	: David Gross-Amblard
<i>Nombre d'ECTS</i>	: 1.5
<i>Répartition des enseignements</i>	: Cours : 15h, Ateliers : 6h
<i>Langue d'enseignement</i>	: Français

### Modalités d'évaluation :

Un projet de mise en œuvre de système décentralisé, type épidémique

### Acquis d'apprentissage (objectifs) :

Cet enseignement vise à donner aux étudiants les connaissances de base sur les architectures distribuées, réparties sur différents sites partout dans le monde. Les trois architectures réparties à grande échelles les plus courantes seront présentées : grilles, systèmes peer-to-peer, et cloud. Les hypothèses, concepts and algorithmes seront détaillés pour chacune d'entre elles. L'objectif est d'avoir une connaissance des systèmes répartis disponibles actuellement et de pointer les directions futures de ces architectures.

### Principales notions abordées :

1. Introduction aux architectures distribuées
2. Les concepts fondateurs (synchronisation, exclusion mutuelle, etc.)
3. Les approches centralisées et semi-centralisées (cloud, grilles, etc.)
4. Les approches décentralisées (systèmes P2P - structurés, non-structurés et hybrides)
5. Application aux systèmes de partage de fichiers et aux protocoles épidémiques

### Références bibliographiques :

- Andrew S. Tanenbaum et Maarten Van Steen. Distributed Systems: Principles and Paradigms. Pearson New International Edition (2013)
- Kenneth Birman : Guide to Reliable Distributed Systems. Springer Verlag (2012)
- Fabrice Le Fessant et Jean-Marie Thomas. Le peer-to-peer : Comprendre et utiliser. Eyrolles (2011)
- Andrew S. Tanenbaum. Systèmes d'exploitation : Systèmes centralisés, systèmes distribués. Dunod (1999)

## SECURITE DES DONNEES

### Data Security

<i>Enseignant</i>	: David Gross-Amblard
<i>Nombre d'ECTS</i>	: 1
<i>Répartition des enseignements</i>	: Cours : 9h, Ateliers : 9h
<i>Langue d'enseignement</i>	: Français

### Modalités d'évaluation :

Examen écrit

### Acquis d'apprentissage (objectifs) :

La sécurité informatique fait actuellement l'objet d'une actualité particulièrement dynamique : attaques spectaculaires (virus, intrusion, ...), commerce électronique, évolutions de législations...

L'objet de ce cours est de présenter les grands principes de la sécurité informatique et les techniques de protection des données.

L'usage de la cryptographie est l'un des outils de protection contre la divulgation, la modification ou l'accès illégitime à des données ou moyens.

Les techniques cryptographiques qui permettent d'assurer les services de confidentialité, d'intégrité, de signature ou d'authentification.

Finalement, des systèmes utilisant ces techniques seront schématiquement décrits.

### Principales notions abordées :

1. Introduction à la sécurité
  - 1.1. Besoins
  - 1.2. Menaces
2. Cryptographie
  - 2.1. Définitions générales
  - 2.2. Cryptographies à clés secrètes
  - 2.3. Cryptographies à clés publiques
  - 2.4. Protocoles cryptographiques
3. Systèmes utilisateurs
  - 3.1. Applications Web
  - 3.2. Carte bancaire
  - 3.3 Application réseaux

### Références bibliographiques :

- Schneier, Cryptographie appliquée, Thomson Publishing, 1997
- Stinson, Cryptographie : Théorie et pratique, Vuibert 2003
- Menezes, Van Oorschot, Vanstone, Handbook of applied Cryptography, CRC Press, 1997 (version actualisée en ligne)
- Vergnaud, Exercice et problèmes de la cryptographie, Dunod, 2012.
- Singh, Histoire des codes secrets, JC Lattes, 1999

UE5 Formation analyste

## GESTION DE PROJET

*Titre en anglais*

<i>Enseignant</i>	: Xavier Brossard
<i>Nombre d'ECTS</i>	: 2
<i>Répartition des enseignements</i>	: Cours : 24h
<i>Langue d'enseignement</i>	: Français

### **Modalités d'évaluation :**

À définir

### **Acquis d'apprentissage (objectifs) :**

À définir

### **Principales notions abordées :**

À définir

### **Références bibliographiques :**

À définir

## ARCHITECTURE INFORMATIQUE

*Titre en anglais*

*Enseignant* : DUFFOUR Nicolas

*Nombre d'ECTS* : 2

*Répartition des enseignements* : Cours : 24h

*Langue d'enseignement* : Français

### **Modalités d'évaluation :**

L'évaluation de la partie théorique du cours est réalisée par interrogation orale à chaque début de session.

L'évaluation de la préparation à l'examen écrit stricto sensu est réalisé par examen blanc et plusieurs exercices pratiques basés sur les examens des années précédentes.

La préparation à l'examen orale est réalisée par des exercices corrigés en cours.

### **Acquis d'apprentissage (objectifs) :**

Connaître les principes d'architecture informatique à l'état de l'art.

Savoir répondre aux questions architectures et technique de l'épreuve écrite de l'examen.

Savoir répondre aux questions techniques de l'épreuve orale de l'examen.

### **Principales notions abordées :**

Architecture : physique, technique, applicative, logique

Sécurité

Transformation numérique : cloud, big data, ia

### **Références bibliographiques :**

Mon cours est basé sur un support powerpoint qui a comme référence bibliographique les documents du DOD, de l'OPEN GROUP, de l'ISO, de SQL Ingénierie, de Octo, etc

# ENVIRONNEMENT INFORMATIQUE MINISTERIEL ET INTERMINISTERIEL

*Titre en anglais*

<i>Enseignant</i>	: Sophie Ronfle
<i>Nombre d'ECTS</i>	: 2
<i>Répartition des enseignements</i>	: Cours : 18h
<i>Langue d'enseignement</i>	: Français

**Modalités d'évaluation :**

**Acquis d'apprentissage (objectifs) :**

**Principales notions abordées :**

**Références bibliographiques :**

## PREPARATION A L'EXAMEN ORAL

*Titre en anglais*

<i>Enseignant</i>	: À définir
<i>Nombre d'ECTS</i>	: 2
<i>Répartition des enseignements</i>	: Cours : 27h
<i>Langue d'enseignement</i>	: français

**Modalités d'évaluation :**

À définir

**Acquis d'apprentissage (objectifs) :**

À définir

**Principales notions abordées :**

À définir

**Références bibliographiques :**

À définir

## PREPARATION A L'EXAMEN ECRIT

*Titre en anglais*

<i>Enseignant</i>	: Sonia Kouame et Xavier Brossard
<i>Nombre d'ECTS</i>	: 3
<i>Répartition des enseignements</i>	: Cours : 66h
<i>Langue d'enseignement</i>	: français

### **Modalités d'évaluation :**

Examens blancs

### **Acquis d'apprentissage (objectifs) :**

Appréhender les particularités de l'épreuve écrite et savoir utiliser l'UML dans ce cadre spécifique.

### **Principales notions abordées :**

Diagramme de communication  
Diagramme de classes  
Diagramme de cas d'utilisation  
Diagramme d'activités  
Diagramme d'états-transitions  
Dictionnaire de données  
TD /exercices pratiques  
Rédaction de cas d'utilisation  
Commentaires des différents diagrammes

### **Références bibliographiques :**

## ANGLAIS

### English

<i>Enseignant</i>	: Emily Burmeister et Todd Donahue
<i>Nombre d'ECTS</i>	: 1
<i>Répartition des enseignements</i>	: Cours : 24h
<i>Langue d'enseignement</i>	: Anglais

### Modalités d'évaluation :

La note finale se calcule ainsi :

- Participation active (en cours et aux séances de l'aide au projet) : 15 %
- Production écrite (*abstract*) : 15 %
- Présentation orale (conception d'enquêtes ou autre) au S1 : 20%
- Projet scientifique : 50% (rapport 25% et soutenance 25%)

### Acquis d'apprentissage (objectifs) :

Cette formation obligatoire cible le développement et l'approfondissement des compétences en expression écrite et production orale nécessaires pour affronter le monde professionnel et scientifique en anglais. Elle cible également le développement et le perfectionnement des compétences grammaticales et lexicales.

### Principales notions abordées :

Les cours seront organisés par groupe de niveau et amèneront les élèves à travailler l'expression orale (discussions/débats, présentation professionnelles...) afin de favoriser le développement de l'aisance et à pratiquer l'expression écrite dans les contextes professionnels et scientifiques.

Les premiers cours seront dédiés au développement des compétences en expression orale et écrite afin de préparer une présentation scientifique et un résumé scientifique en anglais.

Ensuite les cours seront consacrés à l'aide aux projets statistiques ou informatiques ainsi qu'à la préparation des soutenances orales en anglais.

### Références bibliographiques :

- <https://ec.europa.eu/eurostat/home>
- <https://www.oecd.org/france/>
- <https://www.imf.org/external/index.htm>
- <http://www.worldbank.org/>
- <https://www.insee.fr/en>
- Autres références sur l'écriture scientifique et professionnel en anglais seront données en cours

## STAGE DE FIN D'ETUDES

### *Internship*

<i>Enseignant</i>	: Divers maîtres de stage
<i>Nombre d'ECTS</i>	: 30

Pour les fonctionnaires en poste ou suivant le parcours STD (formation initiale ou continue), le stage est effectué dans le service d'affectation. Pour les élèves non-fonctionnaires et fonctionnaires en formation initiale des parcours ES et MSP, le stage est effectué dans une entité dont l'activité entre dans le champ de compétences du master. Le sujet de stage sera soumis à l'approbation du responsable du master. Les étudiants devront rédiger également un rapport de stage qui fera l'objet d'une soutenance. Les modalités et les attendus de ce stage sont détaillés dans le guide des stages.