

Business Intelligence

Infos pratiques

- > ECTS : 3.0
- > Nombre d'heures : 36.0
- > Niveau d'étude : BAC +4
- > Période de l'année : Enseignement huitième semestre
- > Méthodes d'enseignement : En présence
- > Forme d'enseignement : Cours magistral et Travaux dirigés
- > Composante : Sciences économiques, gestion, mathématiques et informatique

Objectifs

Ce cours aborde les techniques d'informatique décisionnelle permettant d'accompagner les décideurs dans leurs prises de décisions stratégiques. Ces techniques se basent en entrée sur des données provenant de données transactionnelles relatives à l'activité de l'entreprise mais également d'autres données pertinentes provenant d'autres sources (capteurs, fichiers de logs, données du net...). Ces dernières doivent être collectées efficacement parmi de larges collections de données.

La première partie du cours est une introduction aux techniques de la recherche d'information, qui vise à trouver automatiquement et rapidement les informations pertinentes à un besoin particulier, à partir des collections d'une grande masse de documents structurés. Il s'agit d'étudier des techniques visant à améliorer la qualité de la recherche (nettoyage, tokenisation et normalisation, filtrage), l'amélioration de la performance (l'indexation), et des modèles d'appariement.

La deuxième partie du cours est consacrée à la présentation des techniques d'intégration, de consultation et d'analyses de données en vue de prendre des décisions dans les organisations. Les nouvelles technologies des entrepôts de données, OLAP et Analyse de données représentent une façon de mettre les données à la disposition des personnes chargées de

prendre des décisions dans les organisations. Il s'agit dans ce cas de trois technologies complémentaires : l'entrepôt de données, la technologie OLAP et les langages de requêtes avancées pour l'analyse de données. Cette partie du cours vise à développer la créativité autour de ces trois technologies.

Approche pédagogique et plan de cours

Le cours suit une approche théorique-pratique avec des petits projets pour mettre en œuvre une architecture complète de BI

- Présentation des concepts de base de recherche d'information.
- Prétraitement des documents (tokenisation et normalisation, lemmatisation, filtrage).
- Les index inversés.
- Les modèles de recherche d'information.
- Évaluation de la recherche d'information.
- Les entrepôts de données
- Le BI et quelques outils du domaine
- OLAP
- SQL Analytique

Évaluation

Session 1 : Évaluation continue (cf. règle par défaut de la section « Modalités spécifiques » des M3C spécifiques)

Session 2 : Règle par défaut décrite dans la section « Modalités de contrôle et examens / Modalités spécifiques »

Compétences visées

- Connaître les principes de la Recherche d'Information
- Connaître les techniques de représentation et d'indexation de collection de documents
- Connaître les métriques de pertinence des résultats de la recherche et savoir les appliquer
- Comprendre l'intérêt de la mise en place d'une solution BI en entreprise
- Aborder des problématiques de traitement/intégration de données sur des exemples concrets
- Connaissance d'au moins un outil du domaine

Bibliographie

- Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, Introduction to Information Retrieval, Cambridge University Press. 2008.
- Charles L. A. Clarke, Gordon V. Cormack, and Stefan Büttcher Information Retrieval: Implementing and Evaluating Search Engines. MIT Press, 2010.
- Kimball, L. Reeves, M. Ross, W. Thorn Waite (Eyrolles,2000) « Concevoir et déployer un data warehouse, Guide du conduit de projet »
- Kimball (Vuibert, 2001) : 'Entrepôts de données, Guide pratique du concepteur de data warehouse'
- Berson, S. J. Smith (McGraw-Hill, 1997). 'Data Warehousing, Data Mining & OLAP'
- H. Inmon (Wiley 2002). 'Building the Data Warehouse'
- Ramakrishnan R. Gehrke J. (Mc Graw Hill 2003). Database Management Systems
- Christian VIGOUROUX 'Pentaho - Mise en place d'une solution Open Source de Business Intelligence' Broché – ENI – 2011
- Philippe Rigaux. Bases de données documentaires et distribuées. <http://b3d.bdpedia.fr/>
- Jimmy Lin and Chris Dyer. Data-Intensive Text Processing with MapReduce. <https://lintool.github.io/MapReduceAlgorithms/MapReduce-book-final.pdf>