




5.00 credits

30.0 h

Q1

Teacher(s)	Fouss François ;Vande Kerckhove Corentin ;Vande Kerckhove Corentin (compensates Fouss François) ;
Language :	French
Place of the course	Mons
Main themes	<p>Web mining is the application of techniques and models to search, collect, clean, analyse, classify and recommend information/data from the web. These techniques are used in particular in search engines, which play a central role in the connected information society, and in social networks. The objective of this course is to learn how to master these techniques and models, so that they can be used/applied in real-life situations.</p> <p>The main topics of this course are:</p> <ul style="list-style-type: none"> • Searching for information on the Web <ul style="list-style-type: none"> • Basic concepts • Web data collection, cleaning and analysis • Text mining: Analysis of textual data from the Web <ul style="list-style-type: none"> • Basic elements of text analysis (corpus, bag of words, etc.) • Term extraction and document representation (word embedding) • Categorisation of documents • Analysis • Link analysis : Content analysis based on the network/graph structure of web data <ul style="list-style-type: none"> • Basic elements of network/graph structure • Methodology of network/graph analysis <ul style="list-style-type: none"> • Identification of cohesive subgroups • Notions of similarity and distance • Identification of prestigious nodes • Identification of central nodes • Prediction of new links • etc
Learning outcomes	<p>At the end of this learning unit, the student is able to :</p> <p><u>Learning Outcomes (LO) at the end of the learning unit</u></p> <p>At the end of this learning unit, the student is able to:</p> <p>1</p> <ul style="list-style-type: none"> • Understand how the main Web data extraction tools work; • Understand how the main algorithms for classifying, analysing and exploiting information from the Web (texts or links) work and use these algorithms; • Make the right decisions in the process of searching and/or analysing information on the Web.
Bibliography	<ul style="list-style-type: none"> • MCILWRAITH D., MARMANIS H., BABENKO D. 2nd ed, Algorithms of the Intelligent Web, Manning Publications, 2016. • LANGVILLE A., MEYER C., Google's PageRank and Beyond : The Science of Search Engine Rankings, Princeton University Press, 2012. • FOUSS F., SAERENS M., SHIMBO M., Algorithms and Models for Network Data and Link Analysis, Cambridge University Press, 2016. • AMINI M.-R., GAUSSIER E., Recherche d'information : Applications, modèles et algorithmes, Eyrolles, 2013. • MANNING C. D., RAGHAVAN P., SCHÜTZE H., Introduction to Information Retrieval, Cambridge University Press, 2008. • MARTIN A., CHARTIER M., ANDRIEU O., Techniques de référencement web : Audit et suivi SEO, Eyrolles, 2016.
Faculty or entity in charge	CLSM

Programmes containing this learning unit (UE)				
Program title	Acronym	Credits	Prerequisite	Learning outcomes
Master [120] in Data Science : Statistic	DATS2M	5		
Master [120] : Business Engineering	INGE2M	5		
Master [120] : Business Engineering	INGM2M	5		
Master [120] in Management (with work-linked-training)	GESA2M	5		