

Teacher(s)	Verhaeghe Hélène ;
Language :	English > French-friendly
Place of the course	Louvain-la-Neuve
Main themes	An important task in data mining is the discovery of patterns in data. Patterns are recurring structures in data; they can provide interpretable explanations for observations in data, can help to gain a better understanding in the structure of data, can be used to build better models, and can be used to solve other computational tasks (such as the construction of database indexes or data compression). Patterns can be found in many different forms of data, including data from supermarkets, insurance companies, scientific experiments, social networks, software projects, and so on. This course will provide an in-depth introduction to pattern mining. After an introduction to the basics of pattern mining, it will provide an in-depth discussion of a number of advanced pattern mining techniques. Topics that will be discussed are: • Categories of pattern mining tasks, including pattern and pattern set mining, supervised and unsupervised pattern mining, dataset types, and pattern scoring functions; • Algorithms for solving different pattern mining tasks; • Data structures for making pattern mining more efficient; • The implementation of pattern mining algorithms; • Mathematical foundations for the different categories of pattern mining tasks; • Complexity classes relevant to pattern mining; • Applications of pattern mining, with a special focus on the application of pattern mining techniques in software engineering.
Learning outcomes	At the end of this learning unit, the student is able to : Given the learning outcomes of the "Master in Computer Science and Engineering" program, this course contributes to the development, acquisition and evaluation of the following learning outcomes: • INFO 1 • INFO 2.1-4 • INFO 4.2-4 • INFO 5.5 • INFO 6.5 Given the learning outcomes of the "Master [120] in Computer Science" program, this course contributes to the development, acquisition and evaluation of the following learning outcomes: • SINF 1.M4, 1.M3 • SINF 2.1-4 • SINF 4.2-4 • SINF 5.5 • SINF 6.5 Students completing this course successfully will be able to • Identify the most appropriate pattern mining task for a given data set; • Explain the advantages and disadvantages of pattern mining algorithms in relation to the problem to be solved; • Identify appropriate approaches for evaluating the quality of patterns and apply them in various situations; • Determine the computational complexity of pattern mining problems; • Develop new pattern mining algorithms for new applications.
Evaluation methods	The final grade is determined by 3 projects during the quadrimester, and an exam organized during the exam session. The final grade is computed as follows: the grades of the 3 projects combined are worth 30% of the final grade (10% for each project), and the grade of the individual exam is worth 70%.
Teaching methods	• Lectures • Exercise sessions, during which exercises will be done that prepare for the exam and projects • 3 projects
Content	• Frequent itemset mining: algorithms, data structures; • Constraint-based itemset mining: algorithms, data structures;

	• Patterns in sequences, trees, graphs: algorithms, data structures, complexity classes; • Pattern mining in supervised data: scoring functions, algorithms; • Pattern set mining in supervised data: scoring functions, models (decision trees, boosting), algorithms • Pattern set mining in unsupervised data: scoring functions (minimum description length principle, maximum entropy), algorithms • Applications of pattern mining: software repositories, traces, log files, cheminformatics, bioinformatics, industrial applications
Inline resources	https://moodle.uclouvain.be/course/view.php?id=3069
Bibliography	"Data Mining: Concepts and Techniques" (4th Edition, 2022) by Jiawei Han, Jian Pei, and Hanghang Tong
Other infos	The programming language used in this course is Python. We assume the students have followed an introductory course on Python, such as LEPL1401, LINFO1101, or LSINC1101. The use of generative AI is: • authorized without the need to disclose it when it is used to help correct the orthography or grammatical mistakes in texts (ex: reports, slides, comments in code) • not authorised for any other use case, except if explicitly stated in the statement of the work
Faculty or entity in charge	INFO

Programmes containing this learning unit (UE)				
Program title	Acronym	Credits	Prerequisite	Learning outcomes
Master [120] in Data Science : Statistic	DATS2M	5		
Master [120] in Chemical and Materials Engineering	KIMA2M	5		
Master [120] in Civil Engineering	GCE2M	5		
Master [120] in Biomedical Engineering	GBIO2M	5		
Master [120] in Mechanical Engineering	MECA2M	5		
Master [120] in Electrical Engineering	ELEC2M	5		
Master [120] in Physical Engineering	FYAP2M	5		
Master [120] in Computer Science and Engineering	INFO2M	5		
Master [120] in Computer Science	SINF2M	5		
Master [120] in Electro- mechanical Engineering	ELME2M	5		
Master [120] in Mathematical Engineering	MAP2M	5		
Master [120] in Data Science Engineering	DATE2M	5		
Master [120] in Data Science: Information Technology	DATI2M	5		
Master [120] in Energy Engineering	NRGY2M	5		