UNIVERSITÀ DEGLI STUDI DI PALERMO

DEPARTMENT	Matematica e Informatica
ACADEMIC YEAR	2023/2024
MASTER'S DEGREE (MSC)	DATA, ALGORITHMS AND MACHINE INTELLIGENCE
SUBJECT	INFORMATION THEORY AND DATA COMPRESSION
TYPE OF EDUCATIONAL ACTIVITY	B
AMBIT	50341-Discipline Informatiche
CODE	22450
SCIENTIFIC SECTOR(S)	INF/01
HEAD PROFESSOR(S)	SCIORTINO MARINELLA Professore Ordinario Univ. di PALERMO
OTHER PROFESSOR(S)	
CREDITS	6
INDIVIDUAL STUDY (Hrs)	94
COURSE ACTIVITY (Hrs)	56
PROPAEDEUTICAL SUBJECTS	
MUTUALIZATION	
YEAR	1
TERM (SEMESTER)	2° semester
ATTENDANCE	Not mandatory
EVALUATION	Out of 30
TEACHER OFFICE HOURS	**SCIORTINO MARINELLA** Tuesday 13:00 14:00 Ufficio n. 201, II piano del DMI Friday 12:00 13:30 Ufficio n. 201, II piano del DMI

DOCENTE: Prof.ssa MARINELLA SCIORTINO

PREREQUISITES	Ability to use a programming language. Basic notions of theory of algorithms and data structures.
LEARNING OUTCOMES	Knowledge and understanding Learning the main concepts of information theory and the main data compression methods, with particular reference to the Mathematical Theory of Communication Shannon C., in which the formalizations of the concepts of information and code plays a central role. Knowledge of the main data compression techniques. Acquisition of advanced tools for reading basic aspects of the specialized literature of the discipline. Ability to use the technical language of the discipline. Applying knowledge and understanding Ability to recognize and organize autonomously the basis notions of information theory. Capacity to implement and apply the main methods of data compression. Ability to use the knowledge gained (in particular, the data compression methodologies) in specific applicative fields. Making judgments Being able to assess the overall relevance of topics of the discipline, and to link the theoretical aspects of information theory with the practical aspects of data compression. Students will be guided to learn critically and responsibly the topics that will be proposed to them in the classroom and laboratory. Each student will also have opportunity to enrich his own independent judgment by carrying out a study project which will form part of the assessment test. Communication skills Ability to clearly and rigorously expose general topics of the information theory also to a non-expert audience, showing how theoretical methods and results relate to specific application fields. Through the scheduled laboratory activities, the course will tend to develop in students the interaction and the ability to work in group, to discuss the issues in order to find solutions based on the knowledge acquired during the course. The acquisition of communication skills will be achieved through the active participation of the student to laboratory activities as well as the exposition of the results on topics or issues proposed by the teacher. Learning skills Students will develop their update capability by consulting the specialized scientific publications. The material proposed in the classroom and in the laboratory will develop the learning skills of the students who will be able to "interrogate" in an integrated way their knowledge-skills in respect of the addressed issues. Students will also be stimulated to a deeper understanding and criticism of compression methods already known to them.
ASSESSMENT METHODS	The final verification aims to estimate whether the student has acquired the knowledge and the understanding about the Information Theory and the main techniques uses in data compression; the competence to apply this knowledge and understanding in practical contexts; if the student owns the technical language. The final verification consists of a project in which students in groups will have to show that they can apply the topics studied in the course and an oral examination in which they will describe the work they have done. During the oral examination the students should describe their project and answer up to three questions based on all the contents of the course. The final evaluation will be scaled according to the following conditions: a) does not possess an acceptable knowledge of the contents of the presented topics (not sufficient); b) sufficient base knowledge of the contents of the course and of the technical language; sufficient ability to deepen and study the assigned project (18-20); c) discrete knowledge of the main contents of the course, discrete property of language, sufficient ability to independently apply the acquired knowledge; sufficient ability to deepen and study the assigned project (21-23); d) discrete knowledge the contents of the course, discrete property of language, discrete ability to independently apply the competence to solve the proposed problems; discrete ability to deepen and study the assigned project (24-26); e) good mastery of the contents of the course, good property of language, good competence in problem-solving; good ability to deepen and study the assigned project (27-29); f) optimal knowledge of the contents of the course, optimal property of language, very good analytic abilities and competence in problem solving; very good ability to deepen and study the assigned project (30-30 cum laude).
EDUCATIONAL OBJECTIVES	Learning the main concepts of information theory and the most common techniques of coding and lossless compression of information. Ability to use the specific technical language of the discipline. Ability to use the knowledge gained (in particular, the lossless data compression methodologies) in specific applicative fields. Update capability by consulting the scientific specialized results and publications.
TEACHING METHODS	

	Teaching is organized by classroom lectures and laboratory lectures, where it is expected that both the knowledge is transferred and the acquisition of practical skills is verified by using computer. In order to facilitate participation, collaborative learning methodologies are implemented.
SUGGESTED BIBLIOGRAPHY	Testi di riferimento/Reference textbooks: Cover T.M. and Thomas J.A., Elements of Information Theory, Wiley-Interscience, 2006 (per gli argomenti di teoria dell'Informazione/for topics on Information Theory) ISBN 978-0471241959 Sayood K., Introduction to Data Compression, Morgan Kauffman, 2017 (Per argomenti su compressione dati lossless/for topics on lossless data compression) ISBN 978-0124157965 Salomon D., Motta G. Handbook of Data Compression, Fifth Edition, Springer 2010 (Per argomenti su compressione dati lossless/for topics on lossless data compression) ISBN 978-1-84882-902-2 Testi di consultazione/Consultation textbooks: Shannon, C.E. (July 1948). "A Mathematical Theory of Communication". Bell System Technical Journal 27: 379–423. DOI:10.1002/j.1538-7305.1948.tb01338 Nelson M. and Gailly, J.-L., The Data Compression Book, John Wiley & Sons, 1998. ISBN 978-1558514348

SYLLABUS

Hrs	Frontal teaching
4	Presentation of the course. Introduction to the Shannon Information Theory. Entropy of the source as measure of the information produced by the source. Entropy and Compression.
6	Codes. Uniquely Decodable Codes. Prefix codes. Kraft-McMillan Inequality. Sardinas-Patterson Algorithm.
6	Huffman Coding Algorithm. Optimality of Huffman Codes. Adaptive Version of the algorithm.
6	Arithmetic Coding. Techniques for coding integers.
5	Dictionary based compression methods. Static dictionary and optimal parsing problem. Dynamic dictionary. Lempel and Ziv algorithm. LZ77 compression algorithm.
2	Variants of LZ77 algorithm. LZ78 Compressor.
4	Context-Based Compression. Burrows- Wheeler Transform (BWT). Reversibility of the BWT. Mathematical properties of BWT. BWT and suffix array. Algorithm for computing the suffix array of a text.
5	Move-to front coding. Bwt-based compression and indexing methods.
2	Kolmogorov complexity. Smallest grammar Problem.

Hrs	Workshops
4	Implementation of Huffman coding, arithmetic coding and techniques for coding integers.
5	Implementation of dictionary-based compression algorithms.
4	Implementations of BWT-based compressors and indexing.
3	Compression and search engines. Using Compression Benchmarks.