

Degree	Type	Year
Data Engineering	OB	2

Contact

Name: Joaquin Borges Ayats

Email: joaquim.borges@uab.cat

Teachers

Joaquin Borges Ayats

Adrià Figuerola Torrell

Teaching groups languages

You can view this information at the [end](#) of this document.

Prerequisites

There are no prerequisites. However, students should be familiar with the most basic concepts of fundamental linear algebra, mathematical analysis and probability theory.

Objectives and Contextualisation

To study the mathematical theory of information, for the discrete case, based on C.E. Shannon's papers in 1948. To study sources of data, source coding, data compression and channel coding. To study error-detecting and correcting codes for an efficient data transmission or data storage.

Competences

- Generate innovative and competitive proposals in professional activity and research.
- Students must be capable of collecting and interpreting relevant data (usually within their area of study) in order to make statements that reflect social, scientific or ethical relevant issues.
- Transmit data with efficiency, precision and security.

- Work cooperatively in complex and uncertain environments and with limited resources in a multidisciplinary context, assuming and respecting the role of the different members of the group.

Learning Outcomes

1. Analyse and evaluate the advantages and disadvantages of a lossy, a lossless and a quasi-lossless compression.
2. Decide on the most suitable type of coding for the characteristics of the signal and the transmission channel.
3. Formulate methods for information compression and encoding for error correction.
4. Generate innovative and competitive proposals in professional activity and research.
5. Students must be capable of collecting and interpreting relevant data (usually within their area of study) in order to make statements that reflect social, scientific or ethical relevant issues.
6. Work cooperatively in complex and uncertain environments and with limited resources in a multidisciplinary context, assuming and respecting the role of the different members of the group.

Content

- 1.- Basic concepts. Discrete memoryless sources.
 - 1.1.- The problems of communication.
 - 1.2.- Measure of information.
 - 1.3.- Shannon's model of discrete memoryless source.
 - 1.4.- Entropy function.
 - 1.5.- Mutual information.
 - 1.6.- Discrete memoryless channels. Capacity.
- 2.- Source coding.
 - 2.1.- Introduction and objectives.
 - 2.2.- Constant length codes.
 - 2.3.- Variable length codes. Unique decipherability.
 - 2.4.- Shannon's bounds.
 - 2.5.- Construction of optimal codes.
- 3.- Data compression.
 - 3.1.- Types of compression. Measures of compression.
 - 3.2.- Compression techniques.
 - 3.3.- Statistical methods.
 - 3.4.- Dictionary techniques.
 - 3.5.- Compression of sound and images.
- 4.- Discrete memoryless channels.

- 4.1.- Models for channels.
- 4.2.- Calculation of channel capacity.
- 4.3.- Decoding rules.
- 4.4.- The fundamental theorem,
- 5.- Coding theory I: linear codes.
- 5.1.- Block codes. Minimum distance decoding.
- 5.2.- Introduction to finite fields.
- 5.3.- Linear codes. Generator matrices.
- 5.4.- Equivalent codes. Systematic encoding.
- 5.5.- Dual codes. Parity-check matrices.
- 5.6.- Decoding. Standard array and syndrome.
- 5.7.- Some families of important linear codes.
- 6.- Coding theory II: cyclic codes.
- 6.1.- Cyclic codes as ideals of polynomial rings.
- 6.2.- Generator and parity-check polynomials.
- 6.3.- Systematic encoding with cyclic codes.

Activities and Methodology

Title	Hours	ECTS	Learning Outcomes
Type: Directed			
Practice sessions	12	0.48	1, 2, 3, 4, 5, 6
Problem sessions	12	0.48	1, 2, 5
Theory lectures	26	1.04	1, 2, 5
Type: Supervised			
Tutoring and consultations	17	0.68	1, 2, 3, 5
Type: Autonomous			
Independent study	25	1	1, 2, 3, 4, 5
Preparing exercises and practice	25	1	1, 2, 3, 4, 5
Preparing the final test	25	1	1, 2, 3, 5

Theoretical content will be taught through lectures, although students will be encouraged to actively participate in the resolution of examples. During problem sessions, a list of exercises will be resolved. Students are encouraged to solve the problems on their own in advance. Students will also be encouraged to present their own solutions in class. In the practical sessions, subjects related to the theoretical content of the course will be developed. Campus Virtual will be used for communication between lecturers and students (material, updates, announcements, etc.).

Annotation: Within the schedule set by the centre or degree programme, 15 minutes of one class will be reserved for students to evaluate their lecturers and their courses or modules through questionnaires.

Assessment

Continous Assessment Activities

Title	Weighting	Hours	ECTS	Learning Outcomes
Final test	60%	3	0.12	1, 2, 5
Practice	25%	2	0.08	1, 2, 3, 4, 5, 6
Tests based on exercise resolutions in exercise-base classes	15%	3	0.12	1, 2, 3, 5

Continuous-assessment dates will be published on Campus Virtual. Specific programming may change when necessary. Any such modification will always be communicated to students through Campus Virtual, which is the usual communication platform between lecturers and students.

Subject assessment (out of 10 points) will be carried out as follows:

- Two individual partial tests, 6 points (3 points each). As part of continuous assessment, the first test will take place during lectures; the second will take place on the date specified by coordination. These individual tests will consist mostly of exercises in the style of those worked on during the course; a smaller part will consist of more theoretical questions.
- Exercise resolutions, 1.5 points. As part of continuous assessment, activities must be carried out or exercises must be solved by applying the given methods.
- Compulsory practicals, 2.5 points. As part of the continuous assessment, computer implementations will be carried out and will be evaluated. In case of absence in a session without a previous and adequate justification, the contents of that session will be evaluated as 0.
- Final exam, 6 points. Those who have not passed the subject through the individual partial tests will have the option to take final exam as a re-assessment grade to compensate the individual partial tests. There is therefore no separate re-assessment for partial tests; this exam covers material from the entire course. It will consist mainly of exercises in the style of those worked on during the course; a smaller part will consist of more theoretical question.

For each of the individual partial tests a minimum qualification of 0.5 points is required. There are no other minimum qualifications in each part. In order to pass, the different parts must add up to at least 5 points. No single assessment is offered.

Notwithstanding other disciplinary measures deemed appropriate, and in accordance with the academic regulations in force, assessment activities will receive a zero whenever a student commits academic irregularities that may alter such assessment. Assessment activities graded in this way and by this procedure

will not be re-assessable. Irregularities contemplated in this procedure include, among others: the total or partial copying of an evaluation activity; allowing others to copy; presenting group work that has not been done entirely by the members of the group; presenting any materials prepared by a third party as one's own work, even if these materials are translations or adaptations, including work that is not original or exclusively that of the student; having communication devices (such as mobile phones, smart watches, etc.) accessible during theoretical-practical assessment tests (individual exams).

In this subject, the use of Artificial Intelligence (AI) technologies is not allowed in any of its phases. Any work that includes fragments generated with AI will be considered a lack of academic honesty and may lead to a partial or total penalty in the grade of the activity, or greater sanctions in serious cases.

An overall grade of 5 or higher is required to pass the subject. A "non-assessable" grade cannot be assigned to students who have participated in any of the individual partial tests or the final exam. No special treatment will be given to students who have completed the course in the previous academic year, except that the practice grade previously obtained can be assigned to this course gradebook. In order to pass the course with honours, the final grade must be a 9.0 or higher. Because the number of students with this distinction cannot exceed 5% of the number of students enrolled in the course, this distinction will be awarded to whoever has the highest final grade.

In the case of exercise resolution, a review may be requested after the date of the activity. For all other assessment activities, a place, date and time of review will be indicated allowing students to review the activity with the lecturer. If students do not take part in this review, no further opportunity will be made available.

To consult the academic regulations approved by the Governing Council of the UAB, please follow this link: http://webs2002.uab.es/afers_academics/info_ac/0041.htm

Bibliography

- *L. Huguet i J. Rifà. *Comunicación Digital*. Ed. Masson, 1991.
- *D. Salomon: *Data compression - The Complete Reference*, 4th Edition. Springer 2007.
- *W.C. Huffman, V. Pless. *Fundamentals of Error-Correcting Codes*. Cambridge University Press, 2003.
- *R.B. Ash. *Information Theory*. John Wiley and Sons Inc, 1965.
- *G. Alvarez. *Teoría matemática de la información*. Ediciones ICE, 1981.
- *T.C. Bell, J.G. Cleary i I.H. Witten. *Text Compression*. Prentice Hall, 1990.
- *F.J. MacWilliams and N.J.A. Sloane. *The theory of error-correcting codes*. North-Holland, Amsterdam, 1977.

Software

For practices it will be used SageMath. <https://www.sagemath.org/> SageMath is an open source mathematical software system licensed under the GPL. It is based on different open source packages: NumPy, SciPy, matplotlib, Sympy, Maxima, GAP, FLINT, R and others. Its combined potential can be accessed through a common Python-based language or directly through interfaces. Since version 9.0 in January 2020, SageMath has been using Python 3.

Groups and Languages

Please note that this information is provisional until 30 November 2025. You can check it through this [link](#). To consult the language you will need to enter the CODE of the subject.

Name	Group	Language	Semester	Turn
(PAUL) Classroom practices	811	Catalan	second semester	morning-mixed
(PAUL) Classroom practices	812	Catalan	second semester	morning-mixed
(PLAB) Practical laboratories	811	Catalan	second semester	morning-mixed
(PLAB) Practical laboratories	812	Catalan	second semester	morning-mixed
(PLAB) Practical laboratories	813	Catalan	second semester	morning-mixed
(TE) Theory	81	Catalan	second semester	morning-mixed