

**COURSE DATA****Data Subject**

Code	33847
Name	Documentary Management Systems
Cycle	Grade
ECTS Credits	6.0
Academic year	2022 - 2023

Study (s)

Degree	Center	Acad. year	Period
1007 - Degree in Information and Documentation	Faculty of Geography and History	3	Second term

Subject-matter

Degree	Subject-matter	Character
1007 - Degree in Information and Documentation	7 - Information technologies and digital editing	Obligatory

Coordination

Name	Department
CIRILO GIMENO, RAMON VICENTE	240 - Computer Science

SUMMARY

This course covers the functional and technological foundations of computer systems designed to store documents in digital format, specially text documents, and the base technologies that can be found in the implementation of this type of applications (Document Management Systems – DMS).

With this view, students will be presented the information elements fundamentals in the scope of document management, DMS processes for indexing and querying documents, the general functional architecture of a DMS, several information retrieval models, some algorithms for pre-processing, indexing, querying and matching texts, and assessing the overall performance of any DMS and queries processes and results.



PREVIOUS KNOWLEDGE

Relationship to other subjects of the same degree

There are no specified enrollment restrictions with other subjects of the curriculum.

Other requirements

It is recommended to have taken Informàtica 1, Informàtica 2, Sistemas de Representación de Información y Conocimiento, y Arquitectura de Información en la Web prior to taking this course.

OUTCOMES

1007 - Degree in Information and Documentation

- Capacity to write analytical reports and summaries with regard to management and organisation of information.
- Demonstrate organisational and planning skills.
- Know a foreign language.
- Have computer skills related to the field of study.
- Have skills for information management.
- Have problem-solving skills.
- Have decision-making capacity.
- Be able to work in a team and to integrate into multidisciplinary teams.
- Be able to apply critical reasoning to the analysis and assessment of alternatives.
- Show ethical commitment in the relationships with users and in information handling.
- Be able to learn independently.
- Show management and leadership skills.
- Show creativity.
- Show motivation for quality.
- Be sensitive to environmental issues, sustainability and human rights.
- Be able to search and retrieve information by methods that meet the expectations and needs of users in optimal conditions of cost and time.
- Be able to design information products and services in any field and by any means of dissemination (electronic edition) according to the information and training needs detected in a community of users.
- Be able to use and put into practice methods, techniques and computer tools (hardware or software) for the design, implementation, development and operation of information systems.



- Understand, design and apply models for data and information representation, and mechanisms for data extraction and exploitation and for information retrieval.
- Know, use and apply information and communication technologies applied to the storage, use, management, handling, distribution and exploitation of data, information and knowledge.
- Know, use and apply the computer and telecommunications tools that support the development of the set of skills that must be acquired in the training process.

LEARNING OUTCOMES

By the end of this course, students will be able to:

- To describe and to work with concepts of information, information needs and information retrieval in a technological context, and the different representation types of textual information.
- To understand the basic information retrieval models that are implemented in the most widely used Document Management Systems
- To define the structure and functional architecture of a Document Management System, user typology and integrating processes.
- To use the basic algorithms for text processing that a Document Management System implements for indexing, storing and retrieving contents.
- To use and to manage a Document Management System.
- To assess a Document Management System and the retrieval performance through a series of evaluation metrics.

DESCRIPTION OF CONTENTS

1. Introduction and basic concepts

Lecture 1. Introduction

Information needs

Concept of information and document

Information representation

Concept of Information Retrieval. Functional scheme.

2. Information Retrieval Models

Lecture 2. Information Retrieval Models

The IR ideal model

o Definition of precision and recall

o Behaviour of the ideal system

Classical IR models

o Boolean model

o Vector-space model

o Probabilistic model

More models



Lecture 3. Functional architecture of a DMS

Functional description

- o Document indexing
- o Full search
- o Selective dissemination of information
- o Feedback
- DMS structure
- Query capabilities
- o Types of queries (terms, phrases, proximity, fuzzy, wildcards)
- o Hits management (ranking, zoning, highlighting)
- o General (vocabulary navigation, query storage, thesaurus management)

3. Processes, data structures and algorithms

Lecture 4. Document management and retrieval

Document indexing

- o Indexes implementation
- Query operations
- o Boolean queries resolution
- o Vector-space similarity calculation
- o Sequential search
- Feedback implementation

Lecture 5. Text processing

Lexical analysis

Stopwords

Stemming

Thesaurus and clustering

4. Evaluation of DMS

Systems evaluation

- o Systems and components to be evaluated
- Query evaluation
- o Evaluation measurements
- The TREC conference as an example of evaluation

**WORKLOAD**

ACTIVITY	Hours	% To be attended
Theory classes	37,50	100
Laboratory practices	22,50	100
Attendance at events and external activities	2,00	0
Development of group work	10,00	0
Development of individual work	13,00	0
Study and independent work	20,00	0
Readings supplementary material	5,00	0
Preparation of evaluation activities	15,00	0
Preparing lectures	5,00	0
Preparation of practical classes and problem	20,00	0
TOTAL	150,00	

TEACHING METHODOLOGY**LECTURES:**

The lectures will be based on active lectures where every 20/25 minutes will be introduced in any activity that requires the involvement of students, so that 1) they can do an activity based on the content they have just learnt, 2) they recover the level of attention to the next block.

LECTURES PREPARATION:

Students will have to prepare the lecture content, following the plan of the course. To do this they will use the literature suggested by the teacher as well as materials provided or/and any other directions given by the lecturer.

PREPARATION OF PRACTICAL WORK:

To better assimilate the contents of the lectures, practical sessions will be conducted in the laboratories. Attendance at practical sessions is mandatory and will be verified by the lecturer in charge of the session. Those students that (for a justified reason) cannot attend the practical sessions should contact the lecturer before the beginning of the first session. The results of these activities must be submitted to the lecturer in charge of the group during the course and in the terms established by the teacher. Students are expected to do/prepare some of these activities at home. Attendance to the practical sessions is mandatory.

TEAM WORK:

A set of problems will be proposed that should be solved in teams of 3 to 6 persons. Each member of the group will be graded both the joint mark of the group as the individual mark of each member.



TUTORIALS:

a) Scheduled tutorials:

Students will work in small groups on some of the more complex concepts introduced in the lectures. They will be provided with a range of activities / problems that will be solved with the help of the lecturer

b) Unscheduled tutorials:

There will be some tutoring hours per week where the lecturer will assist students to clarify concepts or doubts that have arisen during the lectures.

COMPLEMENTARY ACTIVITIES:

There will be a seminar which will address in detail some of the topics discussed throughout the course. The seminar duration will be 2 hours (in-person).

EVALUATION

The evaluation of the subject will follow a scheme of continuous assessment. The following aspects will be considered:

1. Written tests: there will be only one final written test of theoretical - practical character. The minimum score a student must achieve to pass the course is 5 out of 10. The grade obtained in this test represent 50% of the final grade.
2. Preparation of lectures, activities and problems that arise in the context of the lectures will have 15% of the final grade. All activities are mandatory for completion of the course evaluation and are not recoverable in the second call.
3. Practical work: The grade in this section represent 30% of the final grade. All practical work is required for completion of the course evaluation and is not recoverable in the second call.
4. Team work: In the evaluation of team work will be graded both the joint mark of the group as the individual mark of each member. The grade in this section represent 5% of the final grade. All supervised work in a team is required to conduct the evaluation of the subject and is not recoverable in the second call.
5. The composition of the final grade will follow, in summary, the following table:
 - Exam: 50%
 - Preparation of lectures: 15 %
 - Practical work: 30 %
 - Teamwork: 5 %

The evaluation will follow the same scheme in the first and second calls. The marks obtained in the activities and works will be computed in both calls, and these activities will not be recoverable in the second call.



This assessment starts from the premise that teaching at the University of Valencia is, by definition, on-campus lecture delivery method. In this sense, the student should be aware that attendance at both the theoretical and practical lectures is essential for proper monitoring of the contents of the course. The student must also consider the possibility to enroll part time when it is unable to attend all courses (60 credits). However, there is an exception for those students that justify it and request it. They have the possibility of being assessed without attending to all or part of the lectures. For these cases, students should proceed as follows:

- At the beginning of the course, student should inform to lecturer responsible for the course, the incidence that makes her/him unable to attend the class. This must be adequately justified in documentary form.
- The lectures in charge, in the light of this information, will decide the possibility of evaluation without full or partial assistance to the lectures.

Students who are in this situation must submit for evaluation all work required by the lecturer (not necessarily the same to those required for the course) and may also be called to defend them orally to the lecturer, and conduct a knowledge test. The weight of the final grade work will be 50% and the test the remaining 50% knowledge.

REFERENCES

Basic

- Modern Information Retrieval. R. Baeza-Yates, B. Ribeiro-Neto. Addison-Wesley. 1999.
- Introduction to Information Retrieval. C.D. Manning, P. Raghavan, H. Schütze. 2008. <http://nlp.stanford.edu/IR-book/information-retrieval-book.html>
- Information Retrieval. Data structures and algorithms. W.B. Frakes, R. Baeza-Yates. Prentice-Hall. 1992
- Information Retrieval. C. J. van Rijsbergen. 1979. <http://www.dcs.gla.ac.uk/Keith/Preface.html>

Additional

- Bases de Datos Documentales. E. Abadal, L. Codina. 2005.
- "Text Information Retrieval Systems" C.T. Meadow. Academic Press. 1992.
- "Automated Information Retrieval. Theory and Methods" V. Frants, J. Shapiro, V. Voiskunskii. Academic Press. 1997.
- "Information Retrieval Systems. Theory and Implementation" G. Kowalsky. Kluwer Academic Publishers. 1997.
- Information Storage and Retrieval. R. Korfhage. Wiley. 1997.
- Information Retrieval Resources: <http://nlp.stanford.edu/IR-book/information-retrieval.html>