UNIVERSITÀ DEGLI STUDI DI BARI ALDO MORO

DIPARTIMENTO DI RICERCA E INNOVAZIONE UMANISTICA

COURSE OF STUDY PATRIMONIO DIGITALE. MUSEI, ARCHIVI, BIBLIOTECHE

ACADEMIC YEAR *2024/25*

ACADEMIC SUBJECT ***Elaborazione e gestione di documenti digitali***

General information			
Year of the course	*II*		
Academic calendar (starting and ending date)	*II semester (24.02.2025-16.05.2025)*		
Credits (CFU/ETCS):	9		
SSD	*INF/01*		
Language	*Italian*		
Mode of attendance	*Optional*		
Professor/ Lecturer			
Name and Surname	Stefano Ferilli		
E-mail	stefano.ferilli@uniba.it		
Telephone	+39 080 544 2293		
Department and address	*Informatica – Via E. Orabona, 4 – 70125 Bari (BA)*		
Virtual room	*Dipartimento di Informatica, floor V, room 519*		
Office Hours (and modalities: e.g., by appointment, on line, etc.)	10-12am Wednesday + meetings by appointment		
Work schedule			
Hours			
Total	Lectures	Hands-on (laboratory, workshops, working groups, seminars, field trips)	Out-of-class study hours/ Self-study hours
225	*63*	*0*	*162*
CFU/ETCS			
9	*9*	*0*	*0*
Learning Objectives	• Knowledge of the history, motivations, objectives, open issues, branches, approaches and techniques of Digital Document Processing and Management • Ability to identify the appropriate approaches, techniques and tools of Digital Document Processing and Management to be applied to given problems • Ability to properly set up components for solving Digital Document Processing and Management tasks • Ability to evaluate the performance of components for Digital Document Processing and Management tasks		
Course prerequisites	*Basics of Computer Science, algorithms and data structures.*		
Teaching strategie	***Lectures***		
Expected learning outcomes in terms of			
Knowledge and understanding on:	Foundations and main tasks, approaches, methods, techniques and tools for digital document processing and management. Outstanding algorithms from the literature.		

Applying knowledge and understanding on:	The students will be able to choose the proper approaches, methods, techniques and tools for digital document processing and management to apply to specific problems, to properly set up the techniques for fruitful application, and to set up evaluation experiments.

Soft skills	• *Making informed judgments and choices*
	  ○ The students will be able to compare the features, pros and cons of different approaches, methods, techniques and tools for digital document processing and management, and to choose those that are appropriate to tackle specific problems.
	  ○ They will also be able evaluate the experimental outcomes and to trace them to the features of the evaluated technique.
	• *Communicating knowledge and understanding*
	  ○ The students will be able to work in team, bringing to bear their knowledge of digital document processing and management in order to carry out fruitful cooperation with other kinds of expertise from other members of the team.
	• *Capacities to continue learning*
	  ○ The students will be provided with all the historical and methodological foundations that will allow them to understand the latest developments in digital document processing and management and to stay up-to-date with advances in the field.

Syllabus	
Content knowledge	***1. Documents*** ***1. Definitions*** ***2. Formats*** ***1. Plain text*** ***2. Images: Color Spaces, Raster Graphics Vector Graphics*** ***3. Layout-based Formats*** ***4. Content-oriented Formats*** ***2. Document Analysis*** ***1. Image Processing*** ***1. Color Representation*** ***2. Color Depth Reduction*** ***3. Content Processing*** ***4. Edge Detection*** ***2. Document Image Analysis*** ***1. Document Structures*** ***2. Pre-processing for Digitized Documents*** ***3. Segmentation*** ***4. Document Image Understanding*** ***3. Content Processing*** ***1. Natural Language Processing*** ***1. Tokenization*** ***2. Language Recognition*** ***3. Stopword Removal*** ***4. Stemming*** ***5. Part-of-Speech Tagging*** ***6. Word Sense Disambiguation*** ***7. Parsing*** ***2. Information Management*** ***1. Keyword Extraction*** ***2. Text Categorization*** ***3. Information Extraction***
Texts and readings	*S. Ferilli. Automatic Digital Document Processing and Management - Problems,*

	Algorithms and Techniques. Advances in Pattern Recognition series, ISBN 978-0-85729-197-4, eISBN 978-0-85729-198-1, ISSN 1617-7916, Springer, London, 2011.
Notes, additional materials	***Slides, Papers, Software if appropriate***
Repository	*http://lacam.di.uniba.it/~ferilli/ufficiale/egdd.html*
Assessment	
Assessment methods	*Oral test, possibly discussing a case study previously agreed upon*

Assessment criteria	• *Knowledge and understanding*
	  ○ Knowledge of definitions, approaches, techniques and tools
	• *Applying knowledge and understanding*
	  ○ Ability to correctly apply the definitions, approaches, techniques and tools
	• *Autonomy of judgment*
	  ○ Ability to identify the suitable approaches, techniques and tools for specific cases
	• *Communication skills*
	  ○ Ability to use correctly and appropriately technical terminology
	  ○ Ability to clearly explain the concepts and motivations of the choices
	• *Capacities to continue learning*
	  ○ Ability to identify the proper background useful to approach a new topic
	  ○ Ability to carry out suitable literature research

Final exam and grading criteria	*Clarity, appropriateness, correctness, balance and completeness of answers to oral questions.* *Ability to provide insight or advice on specific issues based on the acquired knowledge.*
Further information	
	.