FITXA IDENTIFICATIVA

DADES DE L'ASSIGNATURA

Codi: 46578
Nom: Processament del llenguatge natural
Cicle: Màster Universitari Oficial
Crèdits ECTS: 6
Curs acadèmic: 2026-27

TITULACIONS

Titulació	Centre	Curs	Periode
2262 - Màster Universitari en Ciència de Dades	Escola Tècnica Superior d'Enginyeria	1	Segon quadrimestre

MATÈRIES

Titulació	Matèria	Caràcter
2262 - Màster Universitari en Ciència de Dades	Processament del llenguatge natural	OBLIGATÒRIA

COORDINACIÓ

VIVES GILABERT YOLANDA

RESUM

Actualment, gran part de les dades disponibles per a l'anàlisi estan formats per informació no estructurada en forma de textos en llenguatge natural. Entre aquesta informació trobem pàgines web (Wikipedia, periòdics digitals, blogs) o xarxes socials (Facebook, Twitter). Poder analitzar aquests textos, mitjançant algorismes de Processament de Llenguatge Natural (PLN), resulta molt útil perquè les organitzacions puguen prendre millors decisions.

Els algorismes d'aprenentatge automàtic no són capaços d'entendre text o caràcters, per la qual cosa el PLN realitza tot el pre-processament necessari per a convertir aquestes dades en forma de text en un format comprensible per les màquines (números) i així poder realitzar tot tipus d'anàlisi posterior per obtindré la informació rellevant. Entre les aplicacions més comunes del PLN es troben la classificació de textos, cerca i extracció d'informació, sumarització, traducció automàtica o sistemes de resposta automàtica, entre altres.

Es tracta d'una assignatura teòrico-pràctica, on hi haurà sessions de teoria, sessions pràctiques i sessions mixtes. Les classes de teoria s'impartiran en castell

ns mixtes. Les classes de teoria s'impartiran en castell

CONEIXEMENTS PREVIS

RELACIÓ AMB ALTRES ASSIGNATURES DE LA MATEIXA TITULACIÓ

No s'ha especificat restriccions de matrícula amb altres assignatures del pla d'estudis.

ALTRES TIPUS DE REQUISITS

Es recomana haver superat lassignatura dAprenentatge Màquina (primer quadrimestre).

COMPETÈNCIES / RESULTATS D' APRENENTATGE

2262 - Màster Universitari en Ciència de Dades

Capacitat per a treballar en equip per a arribar a solucions de problemes interdisciplinaris usant tècniques d'anàlisi de dades.

Extraure coneixement de conjunts de dades en diferents formats.

Modelar la dependència entre una variable resposta i diverses variables explicatives, en conjunts de dades complexes, per mitjà de tècniques d'aprenentatge màquina, interpretant els resultats obtinguts.

Posseir i comprendre coneixements que aportin una base o oportunitat de ser originals en el desenvolupament i / o aplicació d'idees, sovint en un context de recerca.

Que els estudiants posseïsquen les habilitats d'aprenentatge que els permeten continuar estudiant d'una forma que haurà de ser en gran manera autodirigida o autònoma.

Que els estudiants sàpiguen comunicar les conclusions (i els coneixements i les raons últimes que les sustenten) a públics especialitzats i no especialitzats d'una manera clara i sense ambigüitats.

Ser capaços d'accedir a ferramentes d'informació (bibliogràfiques i d'ocupació) i utilitzar-les apropiadament.

Ser capaços d'assumir la responsabilitat del seu propi desenvolupament professional i de la seua especialització en un o més camps d'estudi, aplicant els coneixements adquirits en la identificació d'eixides professionals i jaciments d'ocupació.

Ser capaços de valorar la necessitat de completar la seua formació tècnica, científica, en llengües, en informàtica, en literatura, en ètica, social i humana en general, i d'organitzar el seu propi autoaprenentatge amb un alt grau d'autonomia

Usar les tècniques de processament de llenguatge natural per analitzar textos extraient-ne coneixement útil.

DESCRIPCIÓ DE CONTINGUTS

1. Introducció al Processament de Llenguatge Natural

1.1. Què és el PLN
1.2. La importància del text
1.3. Aproximacions històriques al PLN
1.4. Aplicacions i flux de treball

2. Ús i captura de text

2.1. Cadenes de text en Python
2.2. Expressions regulars
2.3. Càrrega de text
2.4. Captura de contingut web (web scraping)
2.5. Reconeixement Òptic de Caràcters (OCR)

3. Preprocessament de text

3.1. Divisió de text
3.2. Neteja i normalització del text
3.3. Anàlisi morfològic
3.4. Anàlisi semàntic
3.5. Anàlisi gramatical

4. Extracció de característiques

4.1. Característiques simples
4.2. Model Bag of Words
4.3. Model TF-IDF
4.4. Vectors de paraula (word embeddings)
4.5. Vectors de document

5. Aplicacions del PLN

5.1. Classificació
5.2. Extracció de la informació
5.3. Mineria de text
5.4. Recerca de informació
5.5. Models seqüencials

6. Aprenentatge profund en PLN

6.1 Xarxes neuronals recurrents en PLN
6.2 Atenció en aprenentatge profund
6.3 Transformers
6.4 Models i aplicacions

VOLUM DE TREBALL (HORES)

ACTIVITATS PRESENCIALS

Activitat	Hores
Teoria-Pràctiques	4,00
Teoria	28,00
Laboratori	28,00
Total hores	**60,00**

ACTIVITATS NO PRESENCIALS

Activitat	Hores
Assistència a altres activitats	2,00
Elaboració de treballs individuals o en grup	18,00
Estudi i treball autònom	20,00
Preparació de classes	20,00
Preparació d'activitats d'avaluació	15,00
Resolució de casos pràctics	15,00
Total hores	**90,00**

METODOLOGIA DOCENT

Activitats teòriques.

- Desenvolupament expositiu de la matèria amb la participació de l'estudiant en la resolució de qüestions puntuals.
- Realització de qüestionaris individuals d'avaluació.

Activitats pràctiques.

- Aprenentatge mitjançant la resolució de problemes, exercicis i casos d'estudi a través dels quals s'adquireixen competències sobre els diferents aspectes de la matèria.

Treballs en laboratori i/o aula d'ordinadors.

- Aprenentatge mitjançant la realització d'activitats desenvolupades de forma individual o en grups reduïts i portades a terme en aules d'ordinadors.

es d'ordinadors.

AVALUACIÓ

Prova objectiva, consistent en un o diversos exàmens que constaran tant de qüestions teòrico-pràctiques com de problemes (45%). Prova recuperable en segona convocatòria.

Avaluació de les activitats pràctiques a partir de l¿elaboració de treballs/memòries, exposicions orals i eines d¿e-learning de la Universitat (45%). Prova recuperable en segona convocatòria.

Avaluació basada en la participació i el grau d¿implicació de l¿alumne en el procés d¿ensenyament-aprenentatge, tenint en compte l¿assistència regular a les activitats presencials previstes i la resolució periòdica de qüestions i problemes proposats (10%). Prova no recuperable.

Per a aprovar l¿assignatura és necessari que la mitjana de les tres proves siga superior a 5, tant en primera com en segona convocatòria.

BIBLIOGRAFIA

- Sohom Ghosh, Dwight Gunning. Natural Language Processing Fundamentals. Packt Publishing, 2019
- Akshay Kulkarni, Adarsha Shivananda. Natural Language Processing Recipes: Unlocking Text Data with Machine Learning and Deep Learning using Python. Apress, 2019 (disponible e-libro)
- Dipanjan Sarkar. Text Analytics with Python: A Practitioner's Guide to Natural Language Processing. Apress 2019 (disponible e-libro)
- Steven Bird, Ewan Klein, Edward Loper. Natural Language Processing with Python. OReally Media, 2009
- Jacob Eisentein. Natural Language Processing. 2018 (disponible bajo licencia CC-BY-NC-ND)
- Sowmya Vajjala, Bodhisattwa Majumder, Anuj Gupta, Harshit Surana. Practical Natural Language Processing. OReally Media, 2020.
- Build a Large Language Model (From Scratch). Sebastian Raschka. Editorial Manning. Septiembre 2024.