

Titulació	Tipus	Curs
2504392 Intel·ligència Artificial / Artificial Intelligence	OB	2

Professor/a de contacte

Nom: Alicia Fornes Bisquerra

Correu electrònic: alicia.fornes@uab.cat

Equip docent

Pau Torras Coloma

Idiomes dels grups

Podeu consultar aquesta informació al [final](#) del document.

Prerequisits

No hi ha requisits previs oficials, però es recomana haver cursat les assignatures de Fonaments de Programació I i II, Fonaments de Matemàtiques I i II, Probabilitat i Estadística, Enginyeria de Dades i Fonaments d'Aprenentatge Automàtic.

Objectius

Aquest curs proporciona una descripció general de les tècniques fonamentals per al processament del llenguatge natural (Natural Language Processing, NLP), cobrint els enfocaments clàssics per al processament i anàlisi de text, el llenguatge i el modelatge de seqüències i la representació de text, mostrant aplicació als problemes habituals de NLP. El curs també cobreix una introducció a l'aplicació de tècniques recents d'aprenentatge profund a la NLP. El contingut del curs s'anirà ampliant en cursos opcionals posteriors, on s'aprofundiran els enfocaments basats en l'aprenentatge profund, juntament amb altres temes més avançats com l'anàlisi semàntica, la generació del llenguatge o el processament de la parla.

Al final d'aquest curs, els estudiants podran:

- Comprendre els conceptes i tècniques fonamentals utilitzats a NLP.
- Implementar i avaluar diverses tècniques de NLP utilitzant Python i llibreries populars de NLP.
- Aplicar mètodes de NLP a problemes del món real i interpretar-ne els resultats.

Competències

- Analitzar i resoldre problemes de manera efectiva, i generar propostes innovadores i creatives per aconseguir els objectius.
- Conèixer i aplicar les tècniques del processament del llenguatge natural per a l'explotació de dades de naturalesa lingüística i per a la creació i avaluació dels components dels sistemes d'IA basats en el llenguatge.
- Conèixer, comprendre, utilitzar i aplicar adequadament els fonaments matemàtics necessaris per desenvolupar sistemes de raonament, aprenentatge i manipulació de grans volums de dades.
- Desenvolupar pensament crític per analitzar de manera fonamentada i argumentada alternatives i propostes tant pròpies com alienes.
- Elaborar estratègies per formular i solucionar diferents problemes d'aprenentatge de manera científica, creativa, crítica i sistemàtica, coneixent les capacitats i limitacions dels diferents mètodes i eines existents.
- Identificar, comprendre i aplicar els conceptes i tècniques fonamentals de representació del coneixement, raonament i aprenentatge computacional més adequats per a la solució de problemes d'intel·ligència artificial.
- Introduir canvis en els mètodes i els processos de l'àmbit de coneixement per donar respostes innovadores a les necessitats i demandes de la societat.
- Que els estudiants sàpiguen aplicar els coneixements propis a la seva feina o vocació d'una manera professional i tinguin les competències que se solen demostrar per mitjà de l'elaboració i la defensa d'arguments i la resolució de problemes dins de la seva àrea d'estudi.
- Treballar cooperativament per aconseguir objectius comuns, assumint la pròpia responsabilitat i respectant el rol dels diferents membres de l'equip.

Resultats d'aprenentatge

1. Analitzar i resoldre problemes de manera efectiva, i generar propostes innovadores i creatives per aconseguir els objectius.
2. Analitzar una situació i identificar-ne els punts de millora.
3. Conèixer i aplicar tècniques fonamentals de modelatge del llenguatge natural i la parla.
4. Conèixer, aplicar i adaptar metodologies d'avaluació i anàlisi de sistemes de processament del llenguatge natural.
5. Conèixer, comprendre i utilitzar representacions algebraiques d'alfabets, paraules i llenguatges mitjançant llenguatges formals tals com autòmats i gramàtiques.
6. Conèixer, comprendre, utilitzar i aplicar els fonaments matemàtics necessaris per al processament del llenguatge natural.
7. Desenvolupar pensament crític per analitzar de manera fonamentada i argumentada alternatives i propostes tant pròpies com alienes.
8. Elaborar solucions per a projectes específics de processament del llenguatge natural.
9. Entendre els conceptes de biaix i variància, i poder utilitzar mètodes de preparació de dades i tècniques de regularització per obtenir solucions generalitzables a partir de les dades disponibles.
10. Identificar situacions que necessiten un canvi o millora.
11. Que els estudiants sàpiguen aplicar els coneixements propis a la seva feina o vocació d'una manera professional i tinguin les competències que se solen demostrar per mitjà de l'elaboració i la defensa d'arguments i la resolució de problemes dins de la seva àrea d'estudi.
12. Treballar cooperativament per aconseguir objectius comuns, assumint la pròpia responsabilitat i respectant el rol dels diferents membres de l'equip.

Continguts

1. Introducció a la lingüística i NLP
2. Processament de text bàsic
3. Anàlisi sintàctica
4. Modelatge del llenguatge

5. Etiquetatge de seqüències
6. Incrustacions de text (text embeddings)
7. Aprenentatge profund per al processament del llenguatge

Activitats formatives i Metodologia

Títol	Hores	ECTS	Resultats d'aprenentatge
Tipus: Dirigides			
Classes de teoria	25	1	2, 7, 11, 3, 5, 9, 4, 6
Sessions de problemes	16	0,64	2, 1, 7, 8, 11, 3, 5, 4, 6, 10
Sessions de projecte	4	0,16	2, 1, 7, 8, 11, 3, 4, 6, 12
Tipus: Supervisades			
Treballar en el projecte	50	2	2, 1, 8, 11, 3, 6, 12
Tipus: Autònomes			
Estudi individual	24	0,96	11, 3, 5, 9, 4, 6
Resolució de problemes (individual)	25	1	2, 1, 7, 8, 11, 3, 5, 9, 4, 6, 10

Hi haurà tres tipus d'activitats docents: classes teòriques, resolució d'exercicis pràctics de forma individual (problemes) i desenvolupament d'un projecte en petits grups de 2-3 alumnes.

1. Classes de teoria: Presentació dels continguts teòrics de l'assignatura. Per a cadascun dels temes estudiats s'exposen els principals conceptes teòrics i formulació matemàtica, així com les solucions algorísmiques corresponents.
2. Sessions de laboratori: Les sessions de laboratori tenen com a objectiu facilitar la interacció i reforçar la comprensió dels temes tractats a les classes de teoria. Durant les sessions de laboratori abordarem dos tipus d'activitats: la resolució d'exercicis pràctics (problemes) i el seguiment i la presentació de projectes.

2.1. Problemes: S'utilitzarà un conjunt de problemes per treballar proporcionats a quaderns Jupyter que exemplifiquen els detalls de codificació dels conceptes exposats durant les classes de teoria. El treball dels problemes s'iniciarà a classe i haurà de ser completat a casa. Es requerirà que els estudiants facin enviaments regulars de la seva feina, que comprendrà la carpeta de problemes (portfoli).

2.2. Projecte: Es farà un projecte durant el semestre, on els alumnes hauran de resoldre un problema específic de certa complexitat. El projecte es resoldrà en petits grups de 2-3 alumnes, on cada membre del grup haurà d'aportar-ne una part i ajuntar-la amb la resta per obtenir la solució final. Aquests grups de treball s'han de mantenir fins al final del semestre i s'han d'autogestionar quant a distribució de rols, planificació del treball, assignació de tasques, gestió dels recursos disponibles, conflictes, etc. Per desenvolupar el projecte, els grups treballaran de forma autònoma, mentre que les sessions pràctiques es faran servir (1) perquè el professor presenti el tema del projecte i discuteixi possibles enfocaments, (2) per al seguiment de l'estat del projecte i (3) perquè els equips presentin els resultats finals.

Les activitats anteriors es complementaran amb un sistema de tutories i consultes fora de l'horari de classes.

Tota la informació de l'assignatura i els documents relacionats que necessitin els alumnes estaran disponibles al campus virtual (cv.uab.cat).

Dins l'horari marcat pel centre o titulació, es reservaran 15 minuts d'una classe per a que els alumnes avaluin els professors i els cursos o mòduls a través de qüestionaris.

Nota: es reservaran 15 minuts d'una classe, dins del calendari establert pel centre/titulació, per a la complementació per part de l'alumnat de les enquestes d'avaluació de l'actuació del professorat i d'avaluació de l'assignatura/mòdul.

Avaluació

Activitats d'avaluació continuada

Títol	Pes	Hores	ECTS	Resultats d'aprenentatge
Exàmens	40	4	0,16	2, 1, 7, 8, 11, 3, 5, 9, 4, 6, 10
Projecte	40	2	0,08	2, 1, 7, 8, 3, 5, 9, 4, 6, 10, 12
Resolució de problemes	20	0	0	2, 1, 8, 11, 3, 9, 4

L'avaluació es fa de forma contínua, per tant no s'ofereix la possibilitat d'avaluació única. Per avaluar el nivell d'aprenentatge dels alumnes s'estableix una fórmula que combina l'adquisició de coneixements, la capacitat de resolució de problemes i la capacitat de treball en equip, així com la presentació dels resultats obtinguts.

Nota final

La nota final es calcula de la manera següent i en funció de les diferents activitats que es realitzin:

$$\text{Nota final} = 0,4 * \text{Nota de teoria} + 0,2 * \text{Nota de problemes} + 0,4 * \text{Nota de projecte}$$

Aquesta fórmula s'aplicarà sempre que la nota de teoria i del projecte sigui superior a 5. No hi ha restricció a la nota dels problemes. Si en fer el càlcul de la fórmula s'obté ≥ 5 però l'alumne no arriba al mínim exigít en alguna de les activitats d'avaluació, aleshores s'atorgarà una qualificació final de 4,5.

Nota de teoria

La nota de teoria té com a objectiu avaluar les capacitats individuals de l'alumne quant als continguts teòrics de l'assignatura. Això es fa de manera contínua durant el curs a través de dos exàmens parcials:

$$\text{Nota de Teoria} = 0.5 * \text{Nota Examen 1} + 0.5 * \text{Nota Examen 2}$$

L'examen parcial (Examen 1) es fa a meitat del quadrimestre i serveix per eliminar part de la matèria si s'aprova. L'examen final (Examen 2) es fa en finalitzar el quadrimestre i serveix per eliminar la resta d'assignatura si s'aprova.

Per aprovar la part de teoria es requerirà que les notes dels exàmens parcials 1 i 2 siguin ambdós superiors a 4,5 i la mitjana superior a 5,0.

Examen de recuperació: En cas que la nota de teoria no assoleixi el nivell adequat per aprovar, els alumnes podran fer un examen de recuperació, destinat a recuperar la part suspesa (parcial 1, 2 o ambdues) del procés d'avaluació contínua.

Nota de problemes

L'objectiu dels problemes és que l'alumne es familiaritzi amb l'aplicació pràctica dels conceptes teòrics. El lliurament regular dels problemes resolts s'utilitzarà com a evidència d'aquest treball.

Per obtenir la qualificació dels problemes cal que més del 50% es lliurin durant el semestre. En cas contrari, la nota dels problemes serà 0.

A cadascun dels dos exàmens parcials hi haurà preguntes sobre els problemes d'aquesta part de l'assignatura. La nota final dels problemes serà la combinació del dossier de problemes i de les preguntes de l'examen.

Nota de problemes = $0.5 * \text{Avaluació del portfoli} + 0.5 * \text{Preguntes d'examen}$

Nota del projecte

El projecte té un pes essencial a la nota global de l'assignatura. El desenvolupament del projecte requereix que els alumnes treballin en grup i dissenyin una solució integral al repte definit. A més, els estudiants han de demostrar les seves habilitats de treball en equip i presentar els resultats a classe.

El projecte s'avalua a través del lliurament, una presentació oral que els estudiants realitzaran a classe i un procés d'avaluació individual. La participació dels estudiants a les tres activitats (preparació del lliurable, presentació i avaluació individual) és necessària per obtenir la qualificació dels projectes. La qualificació es calcula de la manera següent:

Nota del Projecte = $0.6 * \text{Avaluació Lliurament} + 0.3 * \text{Avaluació Presentació} + 0.1 * \text{Avaluació individual}$

Si en fer el càlcul anterior s'obté ≥ 5 però l'estudiant no va participar en alguna de les activitats (lliurables, presentació, avaluació individual), aleshores s'atorgarà una qualificació final de 4.5 al projecte corresponent.

En cas que es presenti el lliurament, però la nota final del projecte no arribi al mínim de 5, hi haurà recuperació del projecte. En cas de no presentar el lliurament o considerar-lo copiat, no hi haurà recuperació i l'assignatura serà considerada suspesa. La nota màxima del projecte que es pot obtenir en cas de recuperació és de 7.

Notes importants

Sense perjudici d'altres mesures disciplinàries que es considerin oportunes, i de conformitat amb la normativa acadèmica vigent, les activitats d'avaluació se suspendran amb zero (0) quan un alumne incorri en irregularitats acadèmiques que puguin alterar aquesta avaluació (per exemple, plagiar, copiar, deixar copiar, ...). Les activitats d'avaluació qualificades d'aquesta manera i per aquest procediment no seran recuperables. Si necessiteu superar alguna d'aquestes activitats d'avaluació per aprovar l'assignatura, se suspendrà directament aquesta assignatura, sense possibilitat de recuperar-la al mateix curs.

En cas que l'estudiant no lliuri la solució de problemes, no assisteixi a les sessions de presentació de projectes durant les sessions de laboratori i no faci cap examen, la qualificació corresponent serà "no avaluable". En un altre cas, els "no shows" compten com a 0 per al càlcul de la mitjana ponderada.

Per aprovar l'assignatura amb matrícula dhonor, la nota final obtinguda haurà de ser igual o superior a 9 punts. Com que el nombre d'alumnes amb aquesta distinció no pot excedir el 5% del total d'alumnes matriculats al curs, s'atorga a qui obtingui la nota final més alta. En cas d'empat, cal tenir en compte els resultats dels exàmens parcials.

Bibliografia

- D. Jurafsky, JH Martín. *Procesamiento del habla y el lenguaje* . Tercera edicion. 2021 < <https://web.stanford.edu/~jurafsky/slp3/> >
- J. Eisenstein. *Procesamiento del lenguaje natural* . 2018. Prensa del MIT
- H. Lane, C. Howard, HM Hapke. *Procesamiento del lenguaje natural en acción* . 2019. Publicaciones de Manning

- Kenny, Dorothy, ed. *Traducción automática para todos* . Prensa científica lingüística, 2022. < <https://langsci-press.org/catalog/book/342> > _ –
- Rowe, Bruce M. y Diane P. Levine. *Una breve introducción a la lingüística* . Routledge, 2018.

Programari

Pels problemes i projectes del curs utilitzarem Python, juntament amb algunes llibreries de Python per a NLP que s'especificaran durant el curs.

Llista d'idiomes

Nom	Grup	Idioma	Semestre	Torn
(PAUL) Pràctiques d'aula	711	Anglès	segon quadrimestre	tarda
(PLAB) Pràctiques de laboratori	711	Anglès	segon quadrimestre	tarda
(PLAB) Pràctiques de laboratori	712	Anglès	segon quadrimestre	tarda
(TE) Teoria	71	Anglès	segon quadrimestre	tarda