



Guía Docente

| Datos Identificativos | | | | | 2024/25 |
|-----------------------|--|--------------------|----------------------|----------|---------|
| Asignatura (*) | Linguaxe Natural e Minaría de Textos | Código | 614G02043 | | |
| Titulación | Grao en Ciencia e Enxeñaría de Datos | | | | |
| Descritores | | | | | |
| Ciclo | Período | Curso | Tipo | Créditos | |
| Grao | 2º cuatrimestre | Cuarto | Optativa | 6 | |
| Idioma | Castelán | | | | |
| Modalidade docente | Presencial | | | | |
| Prerrequisitos | | | | | |
| Departamento | Ciencias da Computación e Tecnoloxías da Información | | | | |
| Coordinación | Vilares Calvo, David | Correo electrónico | david.vilares@udc.es | | |
| Profesorado | Gómez Rodríguez, Carlos | Correo electrónico | carlos.gomez@udc.es | | |
| | Vilares Calvo, David | | david.vilares@udc.es | | |
| Web | campusvirtual.udc.es | | | | |
| Descrición xeral | <p>O procesamento da linguaxe natural (PLN) é a área da intelixencia artificial que se ocupa do estudo e desenvolvemento de modelos computacionais que sexan capaces de procesar e comprender as particularidades da linguaxe natural coa mesma eficacia que os humanos.</p> <p>Nesta materia, introducirase ao alumnado nos fundamentos básicos e as técnicas de aprendizaxe automática asociadas ao PLN, que se utilizan como punto de partida para o desenvolvemento de numerosas tecnoloxías da linguaxe e a explotación automática de textos.</p> <p>O alumnado familiarizarase cos algoritmos e técnicas para representar a información latente presente en textos escritos en forma de árbores e grafos, con técnicas para representar palabras de forma que capturen de forma eficiente o seu significado, coa implementación de modelos utilizando técnicas de aprendizaxe automática e a súa aplicación a diversos problemas relacionados co PLN que resulten útiles para usuarios/usuarios non especializados, así como as técnicas existentes para que as tecnoloxías lingüísticas poidan aplicarse a diversas linguas, incluso aquelas para as que se dispoña de recursos escasos, anotados ou non.</p> | | | | |

Competencias / Resultados do título

| Código | Competencias / Resultados do título |
|--------|--|
| A28 | CE28 - Comprensión e dominio dos fundamentos e técnicas para o procesado de datos escritos, tanto en linguaxe formal como en linguaxe natural. |
| B2 | CB2 - Que os estudantes saiban aplicar os seus coñecementos ao seu traballo ou vocación dunha forma profesional e posúan as competencias que adoitan demostrarse por medio da elaboración e defensa de argumentos e a resolución de problemas dentro da súa área de estudo |
| B3 | CB3 - Que os estudantes teñan a capacidade de reunir e interpretar datos relevantes (normalmente dentro da súa área de estudo) para emitir xuízos que inclúan unha reflexión sobre temas relevantes de índole social, científica ou ética |
| B4 | CB4 - Que os estudantes poidan transmitir información, ideas, problemas e solucións a un público tanto especializado como non especializado |
| B7 | CG2 - Elaborar adecuadamente e con certa orixinalidade composicións escritas ou argumentos motivados, redactar plans, proxectos de traballo, artigos científicos e formular hipóteses razoables. |
| B8 | CG3 - Ser capaz de manter e estender formulacións teóricas fundadas para permitir a introdución e explotación de tecnoloxías novas e avanzadas no campo. |
| B9 | CG4 - Capacidade para abordar con éxito todas as etapas dun proxecto de datos: exploración previa dos datos, preprocesado, análise, visualización e comunicación de resultados. |
| B10 | CG5 - Ser capaz de traballar en equipo, especialmente de carácter multidisciplinar, e ser hábiles na xestión do tempo, persoas e toma de decisións. |



| Resultados da aprendizaxe | | |
|--|-------------------------------------|---|
| Resultados de aprendizaxe | Competencias / Resultados do título | |
| Coñecer, comprender e analizar técnicas de procesamento da linguaxe natural para o procesamento e a desambiguación a nivel sintáctico e semántico. | A28 | B2 B3 B4 B7 B8 B9 B10 |
| Saber utilizar as técnicas e métodos de procesamento da linguaxe natural para resolver problemas reais de minaría de textos. | A28 | B2 B3 B4 B7 B8 B9 B10 |
| Coñecer e comprender os problemas que supón o multilingüismo nas fontes de datos e técnicas para resolvelos. | A28 | B2 B8 B9 B10 |
| Coñecer e analizar paradigmas informáticos emerxentes con potencial para mellorar o paralelismo na minaría de textos. | A28 | B2 B4 B7 B8 |

| Contidos | |
|--|---|
| Temas | Subtemas |
| Análise sintáctica de constituíntes para a minaría de textos | Sintaxe dos constituíntes Análise estatística dos constituíntes con programación dinámica Análise de constituíntes de redución e desprazamento con redes neuronais Análise de constituíntes descontinuos Análise de constituíntes secuencia a secuencia |
| Análise de dependencias para a minaría de textos | Sintaxe de dependencia Criterios de anotación e dependencias universais Análise de dependencias baseada en transicións Análise de dependencias a partir de gráficos Non proxectividade |
| Semántica | Análise de dependencias semánticas Vectores densos usando SVD Vectores densos mediante a predición de palabras: skip-gram e CBOW Propiedades dos vectores densos Agrupación de Brown |
| Computación cos sentidos das palabras | Sentidos da palabra Relacións entre os sentidos Bases de datos de relacións léxicas Desambiguación do significado das palabras |
| Aplicacións prácticas da minaría de textos | - |



| | |
|----------------------------------|---|
| Tratamento da lingua multilingüe | Procesamento de linguas ricas en morfoloxía Procesamento da linguaxe non segmentado Procesamento da linguaxe con poucos recursos Tramitación translingüe |
| Tecnoloxías emerxentes | - |

| Planificación | | | | |
|--------------------------|---------------------------|---|-------------------------|--------------|
| Metodoloxías / probas | Competencias / Resultados | Horas lectivas (presenciais e virtuais) | Horas traballo autónomo | Horas totais |
| Sesión maxistral | A28 B3 B8 B9 | 21 | 21 | 42 |
| Prácticas de laboratorio | A28 B2 B4 B7 B9 B10 | 10 | 30 | 40 |
| Solución de problemas | A28 B2 B3 B8 | 8 | 30 | 38 |
| Proba obxectiva | A28 B2 B3 B8 | 2 | 10 | 12 |
| Traballos tutelados | A28 B3 B4 B7 B10 | 1 | 16 | 17 |
| Atención personalizada | | 1 | 0 | 1 |

*Os datos que aparecen na táboa de planificación son de carácter orientativo, considerando a heteroxeneidade do alumnado

| Metodoloxías | |
|--------------------------|--|
| Metodoloxías | Descrición |
| Sesión maxistral | Clases teóricas, nas que se expón o contido de cada tema. O alumnado disporá previamente de copias das transparencias e o profesorado promoverá unha actitude activa, formulando preguntas que permitan aclarar aspectos concretos e deixando preguntas abertas para a reflexión do estudantado. |
| Prácticas de laboratorio | Clases prácticas co uso do ordenador, que permitan ao alumnado familiarizarse dende un punto de vista práctico coas cuestións expostas nas clases teóricas. |
| Solución de problemas | Aprendizaxe baseada en problemas, seminarios, estudos de casos e proxectos. |
| Proba obxectiva | Valorarase o dominio dos coñecementos teóricos e operativos da materia. |
| Traballos tutelados | Traballo no que o alumnado consultará fontes de información actualizadas para familiarizarse con aspectos de investigación da materia. |

| Atención personalizada | |
|---|---|
| Metodoloxías | Descrición |
| Proba obxectiva Sesión maxistral Prácticas de laboratorio Traballos tutelados Solución de problemas | <p>O desenvolvemento, tanto das clases maxistras como das de resolución de problemas e os laboratorios de prácticas, realizarase atendendo ao progreso do alumnado nas capacidades de comprensión e asimilación dos contidos impartidos. O avance xeral da clase compaxinarase cunha atención específica a aqueles/as alumnos/as que presenten maiores dificultades na tarefa da aprendizaxe e cun apoio adicional a aqueles que presenten maior desvoltura e desexen ampliar coñecementos. No traballo tutelado prestarase unha atención personalizada ao alumnado para orientalo no seu traballo autónomo.</p> <p>No que respecta ás tutorías individuais, dado o seu carácter personalizado, non deben dedicarse a estender os contidos con novos conceptos, senón a aclarar os conceptos xa expostos. O profesorado utilizaráas como unha interacción que lle permita extraer conclusións respecto ao grao de asimilación da materia por parte do alumnado.</p> |

| Avaliación | | | |
|--------------|---------------------------|------------|---------------|
| Metodoloxías | Competencias / Resultados | Descrición | Cualificación |
| | | | |



| | | | |
|--------------------------|---------------------|---|----|
| Proba obxectiva | A28 B2 B3 B8 | Actuación obrigatoria. Valorarase o dominio dos coñecementos teóricos e operativos da materia. | 50 |
| Prácticas de laboratorio | A28 B2 B4 B7 B9 B10 | As entregas das prácticas deberán realizarse no prazo establecido no campus virtual e deberán seguir as especificacións sinaladas no escrito tanto para a súa presentación como para a súa defensa. | 40 |
| Traballos tutelados | A28 B3 B4 B7 B10 | Avaliarase o dominio do alumnado para comprender e assimilar o traballo de investigación. | 10 |

Observacións avaliación

O/A alumno/a deberá acadar polo menos o 40% da nota máxima das partes teóricas e prácticas, sendo en todo caso a suma de tres partes superior a 5 para superar a materia. De non cumprirse algún dos requisitos anteriores, a cualificación da convocatoria establecerase en función da nota máis baixa obtida.

No caso de non acadar o mínimo de teoría ou práctica, o/a alumno/a disporá dunha segunda oportunidade na que só estará obrigado/a a entregar dita parte. Os traballos tutelados considéranse como avaliación continua e non se entregarán na segunda oportunidade.

As entregas das prácticas deberán realizarse no prazo establecido no campus virtual e deberán seguir as especificacións sinaladas no escrito tanto para a súa presentación como para a súa defensa.

Quen concorra á proba obxectiva no período oficial de avaliación terá a condición de ?Presentado/a?.

Todos os aspectos relacionados con dispensa académica, dedicación ao estudo, permanencia e fraude académica rexeranse de acordo coa normativa académica vixente da UDC.

Fontes de información

| | |
|------------------------------------|--|
| Bibliografía básica | <ul style="list-style-type: none"> - Manning, Chris., & Schütze, Hinrich (1999). Foundations of statistical natural language processing. MIT Press - Jacob Eisenstein (2019). Introduction to Natural Language Processing. MIT Press - Jurafsky, Dan. & Martin, J. H. (2022). Speech and Language Processing (3rd ed. draft). Disponible en: https://web.stanford.edu/~jurafsky/slp3/ - Kochmar, Ekaterina (2022). Getting Started with Natural Language Processing. O'Reilly - Bengfort, Benjamin & Bilbro, Rebeca & Tony Ojeda (2018). Applied Text Analysis with Python. O'Reilly |
| Bibliografía complementaria | <ul style="list-style-type: none"> - Chollet, François. (2018). Keras: The python deep learning library. Astrophysics Source Code Library - Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze (2008). Introduction to Information Retrieval. Cambridge University Press, Cambridge - Kübler, Sandra., McDonald, Ronald., & Nivre, Joakim (2009). Dependency Parsing. Synthesis lectures on human language technologies. Morgan Claypool - Goldberg, Yoav (2017). Neural network methods for natural language processing. Synthesis lectures on human language technologies. Morgan Claypool |

Recomendacións

Materias que se recomenda ter cursado previamente

Procesamento da Linguaxe Escrita/614G02029

Aprendizaxe Automática III/614G02026

Aprendizaxe Automática I/614G02019

Aprendizaxe Automática II/614G02021

Materias que se recomenda cursar simultaneamente

Materias que continúan o temario

Observacións



Fomentarase o desenvolvemento dunha cidadanía crítica, aberta e respectuosa coa diversidade na nosa sociedade, salientando a igualdade de dereitos do alumnado sen discriminación por cuestión de xénero ou condición sexual. Empregarase unha linguaxe inclusiva no material e no desenvolvemento das sesións. Traballarase para identificar e modificar prexuízos e actitudes sexistas e influirase na contorna para modificalos e fomentar valores de respecto e igualdade. Facilitarase a plena integración do alumnado que por razóns físicas, sensoriais, psíquicas ou socioculturais, experimenten dificultades a un acceso axeitado, igualitario e proveitoso á vida universitaria

(*A Guía docente é o documento onde se visualiza a proposta académica da UDC. Este documento é público e non se pode modificar, salvo casos excepcionais baixo a revisión do órgano competente dacordo coa normativa vixente que establece o proceso de elaboración de guías