



Guía Docente

| Datos Identificativos | | | | | 2022/23 |
|-----------------------|---|--------------------|----------------------|----------|---------|
| Asignatura (*) | Modelado da Linguaxe | Código | 614544009 | | |
| Titulación | Máster Universitario en Intelixencia Artificial | | | | |
| Descritores | | | | | |
| Ciclo | Período | Curso | Tipo | Créditos | |
| Mestrado Oficial | 2º cuatrimestre | Primeiro | Optativa | 3 | |
| Idioma | Inglés | | | | |
| Modalidade docente | Presencial | | | | |
| Prerrequisitos | | | | | |
| Departamento | Ciencias da Computación e Tecnoloxías da Información | | | | |
| Coordinación | Vilares Calvo, David | Correo electrónico | david.vilares@udc.es | | |
| Profesorado | Vilares Calvo, David | Correo electrónico | david.vilares@udc.es | | |
| Web | campusvirtual.udc.es | | | | |
| Descrición xeral | <p>Fornecer coñecementos teóricos que permitan profundar no estudo de modelos lingüísticos: modelos de lingua e modelos semánticos distribucionais.</p> <p>Asociar o modelado lingüístico e os tipos de modelos con diferentes tarefas dentro da área das tecnoloxías lingüísticas e do procesamento da lingua natural.</p> <p>Avaliar diferentes aspectos dos modelos lingüísticos.</p> <p>Fornecer coñecemento práctico que permita poder adestrar novos modelos e usalos convenientemente en diferentes tarefas de procesamento da lingua natural.</p> | | | | |

Competencias do título

| Código | Competencias do título |
|--------|---|
| A2 | CE01 - Comprensión e dominio de técnicas para o procesamento léxico, sintáctico e semántico de textos en linguaxe natural |
| A3 | CE02 - Comprensión e dominio dos fundamentos e técnicas de procesamento de documentos enlazados, estruturados e non estruturados, e da representación do seu contido |
| A4 | CE03 - Comprensión e coñecemento das técnicas de representación e procesado de coñecemento mediante ontoloxías, grafos e RDF, así como das ferramentas asociadas ás mesmas |
| B1 | CG01 - Manter e extender os plantexamentos teóricos fundados para permitir a introducción e explotación de tecnoloxías novas e avanzadas no eido da Intelixencia Artificial |
| B3 | CG03 - Buscar e seleccionar a información útil necesaria para resolver problemas complexos, manexando con soltura as fontes bibliográficas do campo |
| B4 | CG04 - Elaborar axeitadamente e con certa orixinalidade composicións escritas ou argumentos motivados, redactar plans, proxectos de traballo, artigos científicos e formular hipóteses razoables no campo |
| B6 | CB01 - Poseer e comprender coñecementos que aporten unha base ou oportunidade de ser orixinais no desenvolvemento e/ou aplicación de ideas, a miúdo nun contexto de investigación |
| B7 | CB02 - Que os estudantes saiban aplicar os coñecementos adquiridos e posúan capacidade de resolución de problemas en entornos novos ou pouco coñecidos dentro de contextos máis amplos (ou multidisciplinares) relacionados coa su área de estudo |
| B10 | CB05 - Que os estudantes posúan as habilidades de aprendizaxe que lles permitan continuar estudando dun xeito que terá que ser en gran medida autodirixido ou autónomo |
| C2 | CT02 - Dominar a expresión e comprensión, de xeito oral e escrito, dun idioma estranxeiro |
| C3 | CT03 - Utilizar as ferramentas básicas das tecnoloxías da información e as comunicacións (TIC) necesarias para o exercicio da súa profesión e para a aprendizaxe ao longo da súa vida |
| C7 | CT07 - Desenvolver a capacidade de traballar en equipos interdisciplinares ou transdisciplinares, para ofrecer propostas que contribúan a un desenrolo sostible ambiental, económico, político e social |



| | |
|----|---|
| C8 | CT08 - Valorar a importancia que ten a investigación, a innovación e o desenvolvemento tecnolóxico no avance socioeconómico e cultural da sociedade |
|----|---|

| Resultados da aprendizaxe | | | |
|---|------------------------|---|--------------------------|
| Resultados de aprendizaxe | Competencias do título | | |
| Saber utilizar as técnicas e métodos de tratamento da linguaxe natural para resolver problemas reais de análise de textos en linguaxe natural. | AM1 AM3 | BM1 BM3 BM4 BM7 BM10 | CM2 CM3 CM7 |
| Coñecer, comprender e analizar técnicas de aprendizaxe profunda aplicadas ao procesamento da linguaxe natural | AM1 AM2 | BM1 BM3 BM6 BM7 BM10 | CM2 CM3 CM7 CM8 |
| Saber utilizar técnicas e métodos de aprendizaxe profunda para resolver problemas prácticos de procesamento da linguaxe natural | AM1 AM2 | BM1 BM3 BM4 BM6 BM7 BM10 | CM2 CM3 CM7 CM8 |
| Coñecer e comprender os problemas ambientais que supón o custo computacional das técnicas de deep learning cando se aplican á análise de textos | AM1 AM2 | BM1 BM6 | CM2 CM8 |

| Contidos | |
|-------------------------------------|--|
| Temas | Subtemas |
| Modelos de lingua | Modelos de lingua baseados en n-gramas Modelos de lingua baseados en redes neuronais |
| Modelos de semántica distribucional | Hipótese lingüística sobre o significado distribucional Modelos distribucionais clásicos Modelos neuronais de significado estático (word embeddings) Modelos neuronais de significado dinámico-contextual Modelos composiciónais |
| Modelos de lingua | Modelos de lingua baseados en n-gramas Modelos de lingua baseados en redes neuronais |
| Etiquetado secuencial | Uso e adaptación de modelos para o etiquetado secuencial |

| Planificación | | | | |
|----------------------------|--------------------------------------|-------------------|---|--------------|
| Metodoloxías / probas | Competencias | Horas presenciais | Horas non presenciais / traballo autónomo | Horas totais |
| Sesión maxistral | A2 A3 A4 B1 B3 B6 B7 B10 C2 C8 | 10 | 10 | 20 |
| Prácticas de laboratorio | A2 A3 B3 B4 B6 B7 B10 C2 C3 C7 C8 | 5 | 17 | 22 |
| Solución de problemas | A2 A3 B3 B4 B6 B7 B10 C2 C8 | 6 | 15 | 21 |
| Proba de resposta múltiple | A2 A3 B1 B6 B7 B10 C2 | 0 | 1 | 1 |



| | | | | |
|------------------------|-----------------------------|---|---|----|
| Proba obxectiva | A2 A3 B1 B6 B7 B10 C2 C3 | 2 | 8 | 10 |
| 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 alumno disporá previamente de copias das transparencias e o profesor promoverá unha actitude activa, formulando preguntas que permitan aclarar aspectos concretos e deixando preguntas abertas para a reflexión do alumno. |
| Prácticas de laboratorio | Clases prácticas co uso do ordenador, que permitan ao alumno 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 de resposta múltiple | Cuestionarios breves que se habilitarán tras algunhas sesións teóricas para comprobar a asimilación de contidos. |
| Proba obxectiva | Valorarase o dominio dos coñecementos teóricos e operativos da materia. |

| Atención personalizada | |
|--|---|
| Metodoloxías | Descrición |
| Prácticas de laboratorio Solución de problemas Proba obxectiva Sesión maxistral | O desenvolvemento, tanto das clases maxistras como das de resolución de problemas e os laboratorios de prácticas, realizarase atendendo ao progreso dos alumnos nas capacidades de comprensión e asimilación dos contidos impartidos. O avance xeral da clase compaxinarase cunha atención específica a aqueles alumnos que presenten maiores dificultades na tarefa da aprendizaxe e cun apoio adicional a aqueles que presenten maior desenvolvemento e desexen ampliar coñecementos. 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 profesor utilizaráas como unha interacción que lle permita extraer conclusións respecto ao grao de asimilación da materia por parte dos alumnos. |

| Avaliación | | | |
|----------------------------|--------------------------------------|--|---------------|
| Metodoloxías | Competencias | Descrición | Cualificación |
| Prácticas de laboratorio | A2 A3 B3 B4 B6 B7 B10 C2 C3 C7 C8 | As entregas das prácticas deben realizarse dentro do prazo establecido no campus virtual e deben seguir as especificacións indicadas no enunciado tanto para a súa presentación como o seu defensa. | 50 |
| Proba obxectiva | A2 A3 B1 B6 B7 B10 C2 C3 | Realización obrigatoria. Avaliarase o dominio dos coñecementos teóricos e operativos da materia. | 45 |
| Proba de resposta múltiple | A2 A3 B1 B6 B7 B10 C2 | Pequenos cuestionarios de avaliación continua que se propoñeran ao final dalgunhas sesións teóricas e onde se preguntará de maneira simple por algúns dos conceptos explicados en dita sesión. Avisarase con antelación. | 5 |

| Observacións avaliación |
|-------------------------|
| |



Cada alumno deberá alcanzar un mínimo de 40% da máxima nota das partes "Prácticas de Laboratorio" e "Proba obxectiva", e en calquera caso a suma das tres partes debe ser maior a 5 para superar a materia. De non cumprirse algún dos requisitos anteriores, a cualificación da convocatoria establecerase de acordo coa menor puntuación obtida.

No caso de non obter o mínimo nas "Prácticas de laboratorio" ou "Proba obxectiva", o alumno terá unha segunda oportunidade na que só se lle esixirá a entrega desa parte.

Non se gardarán cualificacións entre anos académicos.

As entregas das prácticas deben realizarse dentro do prazo establecido no campus virtual, e seguirán as especificacións indicadas no enunciado tanto para a presentación como para a defensa.

Terá a condición de ?Presentado? quen entregue todas as prácticas obrigatorias ou se presente á proba obxectiva no período oficial de avaliación.

No caso de realización fraudulenta de exercicios ou probas, aplicarase a Normativa de avaliación do rendemento académico dos estudantes e de revisión de cualificacións. En aplicación da normativa correspondente sobre plaxio, a copia total ou parcial dalgún exercicio de prácticas ou de teoría suporá o suspenso nas dúas oportunidades do curso, coa cualificación de 0,0 en ambos os casos.

Fontes de información

Bibliografía básica

Jurafsky, Daniel & James H. Martin (2021). ?N-gram Language Models.? Speech and Language Processing, Capítulo 3. <https://web.stanford.edu/~jurafsky/slp3/>

Jurafsky, Daniel & James H. Martin (2021). ?Vector Semantics and Embeddings.? Speech and Language Processing, Capítulo 6. <https://web.stanford.edu/~jurafsky/slp3/>

Jurafsky, Daniel & James H. Martin (2021). ?Neural Networks and Neural Language Models.? Speech and Language Processing, Capítulo 7. <https://web.stanford.edu/~jurafsky/slp3/>

Jurafsky, Daniel & James H. Martin (2021). ?Sequence Labeling for Parts of Speech and Named Entities.? Speech and Language Processing, Capítulo 8. <https://web.stanford.edu/~jurafsky/slp3/>



| | |
|------------------------------------|--|
| Bibliografía complementaria | <p>Baroni, Marco, Raffaella Bernardi & Roberto Zamparelli (2014). "Frege in space: A program for compositional distributional semantics." <i>Linguistic Issues in Language Technologies</i> 9(6): 5-110.</p> <p>Baroni, Marco, Georgiana Dinu & Germán Kruszewski (2014). "Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors." In <i>Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)</i>, pp. 238-247, Baltimore, Maryland. Association for Computational Linguistics.</p> <p>Church, Kenneth Ward, Zeyu Chen & Yanjun Ma (2021). "Emerging trends: A gentle introduction to fine-tuning." <i>Natural Language Engineering</i>, 27: 763-778.</p> <p>Devlin, Jacob, Ming-Wei Chang, Kenton Lee & Kristina Toutanova (2018). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." In <i>Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)</i>, pages 4171-4186, Minneapolis, Minnesota. Association for Computational Linguistics.</p> <p>Erk, Katrin (2012). "Vector space models of word meaning and phrase meaning: A survey." <i>Language and Linguistics Compass</i> 6.10: 635-653.</p> <p>Hirschberg, Julia & Christopher D. Manning (2015). "Advances in natural language processing." <i>Science</i> 349.6245: 261-266.</p> <p>Linzen, Tal (2016). "Issues in evaluating semantic spaces using word analogies." In <i>Proceedings of the 1st Workshop on Evaluating Vector-Space Representations for NLP</i>, pp. 13-18, Berlin, Germany. Association for Computational Linguistics.</p> <p>Mikolov, Tomas, Wen-tau Yih & Geoffrey Zweig (2013). "Linguistic Regularities in Continuous Space Word Representations." In <i>Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies</i>, pp. 746-751, Atlanta, Georgia. Association for Computational Linguistics.</p> <p>Taher Pilehvar, Mohammad & Jose Camacho-Collados (2021). <i>Embeddings in Natural Language Processing: Theory and Advances in Vector Representations of Meaning</i>. Morgan & Claypool (Synthesis Lectures on Human Language Technologies, volume 47).</p> |
|------------------------------------|--|

Recomendacións

Materias que se recomenda ter cursado previamente

Comprensión da Linguaxe Natural/614544008

Aprendizaxe Automática I/614544012

Materias que se recomenda cursar simultaneamente

Aprendizaxe Profunda/614544013

Aprendizaxe Automática II/614544014

Materias que continúan o temario

Minería de Textos/614544011

Intelixencia Web e Tecnoloxías Semánticas/614544010

Observacións

(*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