

ONTOLOGÍAS, DICIONARIOS Y LEXICONES RESTRINGIDOS COMPUTACIONALES

Curso 2014/2015

(Código: 24409193)

1. PRESENTACIÓN

La asignatura Ontologías, diccionarios y lexicones restringidos computacionales ofrece una visión introductoria de la codificación computacional del significado, donde se estudian las características de estos tres formatos que tienen que ver con el tratamiento del léxico. Asimismo se analizan herramientas como los corpora y se inicia el tratamiento informático de los datos textuales.

La lexicología ha experimentado una rápida evolución, desde la simple concepción como nomenclatura hasta el tratamiento actual a través del análisis informático del corpus. Por ello el uso de las herramientas informáticas abre nuevas perspectivas al análisis lingüístico, ya que es posible disponer de una masa considerable de textos en formato electrónico, lo que permite la extracción de datos de manera fiable.

Asimismo la codificación del conocimiento a través del desarrollo de ontologías, abre amplias perspectivas para el manejo de facilidades computacionales en entornos como internet, que cada vez en mayor medida son el principal medio en el que se desenvuelve el conocimiento.

Los estudiantes deberán familiarizarse con el uso del software existente de extracción del léxico, la noción de frecuencia, las co-ocurrencias, el análisis estadístico de las extracciones, etc. con objeto de iniciarse en un tema complejo como es la creación y manipulación de un corpus. Asimismo deberán familiarizarse con los diferentes tipos de ontologías de orientación y/o soporte lingüístico, con su configuración y con sus aplicaciones.

2. CONTEXTUALIZACIÓN

Esta asignatura, que se imparte en el segundo semestre del Máster de las TIC en la enseñanza y el tratamiento de las lenguas, consta de 5 créditos ECTS y está encuadrada en el módulo temático de contenidos formativos propios con asignaturas fundamentales para la consecución de los objetivos del máster.

Hay otras asignaturas de estas mismas enseñanzas de posgrado que también emplean el análisis de corpus para sus diferentes objetivos, así como el desarrollo de la llamada red semántica. No obstante, en este curso se trata de ofrecer a los alumnos una perspectiva general de los distintos tipos de herramientas de análisis conceptual y textual.

Mientras que análisis de corpus es un elemento común a muchas de las asignaturas que componen este máster, el tratamiento y uso de los corpora, como se verá a lo largo del curso, difiere en cada una de ellas. Por otra parte, el estudio de la codificación conceptual, aunque se pueda tratar en otras asignaturas, en esta adquiere una relevancia importante.



3. REQUISITOS PREVIOS RECOMENDABLES

En esta asignatura se aplican los requisitos generales de admisión al máster. Además, es necesario estar en posesión de un nivel B1/B2 de inglés y se recomienda el conocimiento de una segunda lengua extranjera (francés / alemán) para poder leer bibliografía especializada y artículos científicos y técnicos.

El alumno deberá poseer conocimientos de informática a nivel de usuario y disponer de los medios técnicos y conocimientos para poder acceder al curso virtual, dado que esta asignatura sigue la metodología "a distancia" propia de la UNED.

4. RESULTADOS DE APRENDIZAJE

Al finalizar el curso el estudiante será capaz de obtener los siguientes resultados:

(i) Resultados relacionados con el aprendizaje

1. Conocerá las principales aplicaciones de la lingüística de corpus a la semántica computacional, como el desarrollo y uso de diccionarios electrónicos, lexicones computacionales y las ontologías.
2. Habrá desarrollado las destrezas analíticas necesarias para la gestión, extracción e interpretación de datos semánticos de un corpus y/o una ontología en formato digital.
3. Habrá obtenido los conocimientos lingüísticos necesarios para manejar de forma eficaz las herramientas computacionales para el análisis léxico y/o semántico.
4. Podrá desarrollar proyectos de investigación lingüística e interdisciplinar relacionados con el análisis léxico.

(ii) Resultados relacionados con las actitudes

1. Habrá desarrollado una actitud positiva hacia el valor del análisis de los datos empíricos en el avance del conocimiento..

5. CONTENIDOS DE LA ASIGNATURA

Temas

1. Introducción a la semántica computacional: tratamiento computacional del léxico, del significado y del conocimiento.
2. El almacenamiento computacional del léxico: diccionarios digitales y lexicones computacionales
3. El almacenamiento computacional del significado y del conocimiento: tesauros digitales y ontologías. Bases de datos: Wordnet y otros.
4. Ontolingüística: ontologías léxicas (WordNet, EuroWordNet, etc.), ontologías del lenguaje y ontologías lingüísticas.
5. Tratamiento lexicométrico:
 - a. Herramientas de análisis y extracción de datos de los corpora para la codificación del significado.
 - b. Funciones de búsqueda: búsqueda simple y booleana, expresiones regulares.



6. Herramientas de análisis y extracción de datos de los corpora
7. Funciones de búsqueda : búsqueda simple y booleana, expresiones regulares.
8. Tratamiento lexicométrico

6.EQUIPO DOCENTE

- [MARGARITA GODED RAMBAUD](#)

7.METODOLOGÍA

Como complementación a su aprendizaje, el alumno podrá consultar los enlaces de internet relacionados con la asignatura, que serán facilitados por el equipo docente.

La asignatura tiene una carga de 5 créditos ECTS, que equivalen a 125 horas de trabajo (1 crédito ECTS = 25 horas de trabajo), de las cuales, el 40 % se dedicará al trabajo de interacción con el equipo docente y, y el 60% al trabajo autónomo.

Dentro de estas dos modalidades, las actividades formativas se distribuyen de la siguiente manera:

INTERACCIÓN CON EL EQUIPO DOCENTE (40% = 50 horas = 2 créditos)

- Lectura de las unidades didácticas y/o materiales impresos (recomendados y complementarios).
- Lectura de las orientaciones de la asignatura para la realización de las actividades prácticas.
- Trabajo con las herramientas computacionales (software).
- Resolución de dudas en línea, a través del curso virtual o a través del teléfono.
- Revisión de actividades y trabajos propuestos.

TRABAJO AUTÓNOMO (60% = 75 horas = 3 créditos)

- Estudio de los temas de la asignatura.
- Participación en grupos de estudio.
- Interacción con los compañeros en el foro.
- Búsqueda de información y datos en bibliotecas, Internet u otras fuentes documentales.

Actividades prácticas de cada asignatura, tales como:

- Resolución de problemas y ejercicios.
- Análisis y comprensión de textos.
- Elaboración de informes, resúmenes, redacciones o trabajos diversos.
- Auto-evaluación de las actividades prácticas.
- Trabajo en grupo
- Discusión de temas y solución de problemas en grupo dentro del curso virtual en el foro de alumnos.
- Auto-evaluación del grupo.

8.BIBLIOGRAFÍA BÁSICA

Comentarios y anexos:

Aguado de Cea, G., Álvarez de Mon Rego, I., Pareja-Lora, A. 2004. "OntoTag's Linguistic Ontologies: Enhancing Higher Level



and Semantic Web Annotations". En Proceedings of the IV International Conference on Language Resources and Evaluation (LREC 2004), Vol. VI, pp. 1905 - 1908. European Language Resources Association (ELRA). Lisbon, Portugal.

Aguado de Cea, G., Álvarez de Mon Rego, I., Pareja-Lora, A. 2005. "Conocimiento lingüístico para la Web Semántica: los nombres de persona en el dominio del cine". *Perspectivas Interdisciplinarias de la Lingüística Aplicada*, Tomo II, pp. 369 - 380. Asociación Española de Lingüística Aplicada (AESLA), 05/2005.

Aguado de Cea, G., Álvarez de Mon Rego, I., Pareja-Lora, A. 2009. "Una visión interdisciplinar de la anotación semántica". *Terminología y Sociedad del Conocimiento*, pp. 219 - 254. Peter Lang Verlag, A.G.

Christiane Fellbaum (Ed.)1998. *WordNet: An electronic lexical database*. Cambridge, MA: MIT Press.

De Miguel, E. 2009. *Panorama de la lexicología*. Barcelona: Ariel.

García, A., Pareja-Lora, A., Pradana, D. (2008). "Reutilización de tesauros: el documentalista frente al reto de la web semántica". *El Profesional de la Información*, 17(1):8-21

Goded, M., A. Ibañez, and V. Hoste. 2014. *On the Architecture of Words: applications of meaning studies*. Madrid: UNED.

Huang, Chu-Ren, N. Calzolari, A.Gangemi, A.Oltramari, L. Prevot. (editors). 2010. *Ontology and the Lexicon*. Cambridge University Press.

Lipka, L. 1990. *An Outline of English Lexicology*, Tubinga, Niemeyer.

Mitkov, R. 2003. *Oxford Handbooks in Linguistics*. Oxford University Press

Ooi,Vincent B. Y.1998. *Computer Corpus lexicography*. Edinburgh University Press. Edinburgh.

Aguado de Cea, G., Álvarez de Mon Rego, I., Pareja-Lora, A. 2003. "Primeras aproximaciones a la anotación lingüístico-ontológica de documentos de la Web Semántica: OntoTag". *Inteligencia Artificial*, Vol. 2003, n. 18, pp. 37 - 50.

Pareja-Lora, A. 2013. "An Ontology-Driven Methodology to Reuse, Link and Merge Terminological and Language Resources". En Proceedings of the 10th International Conference on Terminology and Artificial Intelligence – Terminology for a networked society: Recent advances in multilingual knowledge-based resources, pp. 189-196. Paris (Villetaneuse), Francia. 28-30 de octubre de 2013 [https://lipn.univ-paris13.fr/tia2013/Proceedings/actesTIA2013.pdf#page=187].

Pareja-Lora, A. 2014. "The Pragmatic Level of OntoLingAnnot's Ontologies and Their Use in Pragmatic Annotation for Language Teaching". En E. Bárcena et al. (editores), *Languages for Specific Purposes in the Digital Era*. Series: Educational Linguistics, Vol. 19, pp.323-344. Suiza: Springer International Publishing Switzerland. DOI: 10.1007/978-3-319-02222-2_15 [http://link.springer.com/chapter/10.1007/978-3-319-02222-2_15].

Pareja-Lora, A., Aguado de Cea, G. 2010. "Modelling Discourse-related Terminology in OntoLingAnnot's Ontologies". En Proceedings of the Workshop on Establishing and using ontologies as a basis for terminological and knowledge engineering resources. TKE 2010: Presenting terminology and knowledge engineering resources online: models and challenges, pp. 547-574. Ass. for Terminology and Knowledge Transfer (Gesellschaft für Terminologie und Wissenstransfer, GTW). Dublín, Irlanda. 11/08/2010 [http://oa.upm.es/6249/].

Pareja-Lora, A., Blume, M., Lust, B. 2013. "Transforming the Data Transcription and Analysis Tool Metadata and Labels into a Linguistic Linked Open Data Cloud Resource". En Proceedings of the 2nd Workshop on Linked Data in Linguistics (LDL-2013): Representing and linking lexicons, terminologies and other language data, pp. 34-43. ACL Anthology – Association for Computational Linguistics (ACL) [http://aclweb.org/anthology/W/W13/W13-5506.pdf].

Pareja-Lora, A., María Liakata, Stefanie Dipper (editores). 2013. En Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse (LAW VII & ID). Association for Computational Linguistics, pp. 1-256. Sofia, Bulgaria. 8-9 de agosto de 2013. ISBN 978-1-937284-58-9 [http://www.aclweb.org/anthology/W13-23].

WebProtégé (2014) <http://webprotege.stanford.edu/>.



NeOn Toolkit (2014) http://neon-toolkit.org/wiki/Main_Page.

Nierenburg, S & V. Raskin. 2005. *Ontological Semantics*. MIT Press.

Schalley, A. C., D. Zaefferer eds, 2007. *Ontoquistics*. Berlin. New York: Mouton De Gruyter

9. BIBLIOGRAFÍA COMPLEMENTARIA

10. RECURSOS DE APOYO AL ESTUDIO

Como recursos de apoyo, la asignatura contará con un curso virtual, en el que el alumno se comunicará con el profesor y completará los contenidos de la asignatura y resolverá sus dudas.

La plataforma virtual permite también la comunicación entre los propios alumnos que estudian a distancia, lo que fomenta el aprendizaje colaborativo y la consecución de objetivos comunes y compartidos por diversos alumnos.

Es importante que el alumno utilice el curso virtual, en el que encontrará los materiales para trabajar en cada unidad, enlaces Web recomendados, además de otros recursos de apoyo como ejercicios interactivos multimedia, las autoevaluaciones para cada unidad, tutoriales y diverso material que sirve de ayuda para el estudio de la asignatura.

Por otra parte, es aconsejable la participación del alumno en los distintos foros de comunicación. En dichos foros se aclaran numerosas dudas sobre el contenido y el funcionamiento de la asignatura, además de servir también como instrumento de aprendizaje y de interacción en la lengua objeto de estudio.

La UNED cuenta con una biblioteca extensa y rica en bibliografía, a la cual se puede acceder también de forma virtual. Asimismo, en la UNED existe la posibilidad de la comunicación a través de la videoconferencia o la conferencia en línea, herramienta que podrá usarse para mantener un contacto directo con el grupo de alumnos.

11. TUTORIZACIÓN Y SEGUIMIENTO

Los profesores de la Sede Central (equipo docente) están a disposición del alumno a través del curso virtual, y también por teléfono o personalmente (para esto último deberán concertar una cita) en los siguientes horarios, direcciones y teléfonos:

Dra. Margarita Goded Rambaud

Tel. 91398 8181

Despacho 622, planta 6

e-mail: margarita.goded@flog.uned.es

Horario de atención al alumno:



- Martes: de 11:00 a 14:00 horas.
- Miércoles: de 10:00a 15:00 horas .
- Jueves: de 10:00a 14:00 horas.

Dr. Antonio Pareja Lora

Tlf.: 91394 7578 [NO ACONSEJADO: mejor correo electrónico

e -mail: aplor@ucm.es

Horario de atención al alumno:

Viernes de 17:30 a 19:30.

12.EVALUACIÓN DE LOS APRENDIZAJES

Para llevar a cabo la evaluación, se definirán los criterios y los indicadores de progresión, en función de los objetivos previstos y en relación con los contenidos que han sido objeto del aprendizaje.

Se realizará una evaluación continua a lo largo de todo el proceso de aprendizaje con objeto de describir, interpretar y comparar el progreso de los alumnos en relación con los objetivos establecidos previamente.

Se detectarán las necesidades y carencias de los alumnos, orientándoles a lo largo de todo el proceso de aprendizaje y proporcionándoles pistas para la reorientación de los elementos formativos.

La evaluación continua es una etapa más del aprendizaje, por lo que el alumno tendrá en todo momento una visión global de su progresión.

La valoración del progreso de los alumnos se realizará mediante pruebas de evaluación continua a través de trabajos: informes, ensayos, resúmenes, presentaciones, e-portafolio, webquest, contribuciones al foro, participación, etc.

Las correcciones proporcionarán información sobre el proceso de aprendizaje permitiendo así realizar los cambios oportunos tanto al estudiante como al profesor.

Para poder obtener la nota final de aprobado, el estudiante deberá entregar todas las actividades obligatorias. La omisión de presentación de alguna de las actividades obligatorias implica la imposibilidad de superar la asignatura. Sin embargo, las calificaciones obtenidas por las actividades presentadas en los plazos indicados tendrán validez durante todo el año académico. Por lo tanto, aquellos alumnos que no hayan presentado todas las actividades, podrán completarlas en la convocatoria de septiembre.

Se considerará que un alumno ha superado la asignatura cuando haya obtenido una nota igual o mayor a 5 puntos, como resultado de la media ponderada de las actividades planteadas (siempre que se hayan presentado todas ellas).

Los resultados obtenidos por el alumno se calificarán de 0 a 10, con expresión de un decimal, al que podrá añadirse su correspondiente calificación cualitativa, tal y como establece el artículo 5 del Real Decreto 1125/2003, de 5 de septiembre (BOE 18/09/2003).

13.COLABORADORES DOCENTES



14. Igualdad entre los sexos

En coherencia con el valor asumido para promover la igualdad entre los sexos, todas las denominaciones que en esta guía hacen referencia a miembros de la comunidad universitaria y se efectúan en género masculino, se entenderán hechas indistintamente en género femenino.

