CURSO	:	Introducción a la lingüística computacional
SIGLA	:	LET082E
MODULOS	:	5
DISCIPLINA	:	Lingüística
PROFESOR	:	César Antonio Aguilar
CATEGORÍA	:	Profesor asistente

I. DESCRIPCION

Este curso ofrece una introducción general a la lingüística computacional (LC), la cual es una interdisciplina surgida en los años 60 (por lo menos) que relaciona, en primer lugar, a la lingüística y las ciencias computacionales, así como otras áreas tales como la filosofía, la estadística o la psicología. El impacto de la LC se puede ver tanto en la creación de herramientas capaces de analizar y generar lenguaje natural, hasta el desarrollo de teorías formales que ayudan a modelar las lenguas humanas en términos de algoritmos. Así, en el curso se dará una descripción sobre la fundación de la LC, los problemas que aborda, sus principales líneas de investigación, así como sus aplicaciones más relevantes.

II. OBJETIVOS

Generales

Conocer de forma panorámica todo el conjunto conformado por teorías, procesos y herramientas que procesan lenguaje natural, y que configura el campo de la LC contemporánea, ponderando sobre todo el análisis de textos.

Específicos

- Conocer los fundamentos teóricos de las LC como interdisciplina.
- Conocer las principales tareas y aplicaciones de la LC.
- Conocer los principales métodos tanto lingüísticos como estadísticos aplicados por la LC en sus tareas y procesos.
- Conocer algunas herramientas computacionales para el tratamiento de fuentes textuales, específicamente en los niveles léxico, sintáctico y semántico.

III. CONTENIDOS

- Introducción: definición de la LC, relación entre LC con la lingüística teórica y aplicada, distinción entre lingüística computacional e ingeniería lingüística, distinción entre tecnologías del habla y del texto.
- Nociones básicas: lenguaje natural vs lenguaje artificial, gramáticas formales, gramáticas libres de contexto, *n*-gramas, corpus lingüísticos, etiquetado lingüístico, *chunking* y *parsing*, bases de conocimiento léxicas.
- Técnicas de análisis probabilístico.
- Aplicaciones: recuperación y extracción de información, reconocimiento de entidades nombradas, resumen automático, clasificación de textos, detección de plagio, minería de textos, WEB semántica.

IV. METODOLOGIA

- Clases expositivas
- Trabajo en grupo por parte de los estudiantes.
- Revisión de lecturas específicas para cada contenido del curso.

V. EVALUACION

- Aplicación de 4 controles temáticos: 20%.
- Desarrollo de 4 trabajos parciales, para el desarrollo de una cadena de procesamiento para un corpus textual en inglés: 20%
- Reporte final del resultado obtenido tras desarrollar la cadena de procesamiento previamente descrita: 60%

VI. BIBLIOGRAFIA

Básica

Bird, Steven, Klein, Ewan y Loper, Edward (2010): *Natural Language Processing with Python. Analyzing Text with the Natural Language Toolkit*, Sebastropol, Cal., O'Reilly Media.

Bolshakov, Igor y Gelbukh, Alexander (2004): *Computational Linguistics: Models, Resources, Applications*, México, IPN-UNAM-FCE. Sitio electrónico: www.gelbukh.com/clbook/Computational-Linguistics.pdf.

Galicia, Sofía y Gelbukh, Alexander (2007): *Investigaciones en análisis sintáctico para el español*, México, IPN. Sitio electrónico: www.gelbukh.com/libro-investigaciones/LibroSint.pdf.

Gelbukh, Alexander y Sidorov, Grigori (2010): *Procesamiento automático del español con enfoque en recursos léxicos grandes*, México, IPN. Sitio electrónico: www.gelbukh.com/libro-procesamiento-2/procesamiento-automatico.pdf.

Jurafsky, Daniel y Martin, James (2010): *Speech and Language Processing*, Second Edition, New Jersey, Prentice Hall.

Complementaria

Clark, Alexander, Fox, Chris and Lappin, Shalom (eds.) (2010): *The Handbook of Computational Linguistics and Natural Language Processing*, Oxford, UK, Wiley-Blackwell.

Cole, Ronald A., *et al*. (1997): *Survey of the State of the Art in Human Language Technology*, Cambridge, Cambridge University Press.

Manning, Christopher y Schütze, Hinrich, *Foundations of Statistical Natural Language*, Cambridge, Mass., MIT Press, 2000.

Mitkov, Ruslan (ed.), *The Oxford Handbook of Computational Linguistics*, Oxford, Oxford University Press, 2003.

VII. SITIOS ELECTRÓNICOS

Jurafsky, Daniel y Kay, Martin (2011): *History of Computational Linguistics*, Stanford, Ca., Stanford University: www.stanford.edu/class/linguist289/.

Martin, James (2011): *Course of Natural Language Processing*, Boulder, Co., University of Colorado: www.cs.colorado.edu/~martin/csci5832/.

Bird, Steven (2012): *Natural Language Processing with Python*, University of Melbourne, Australia: www.nltk.org.