

Gramàtiques i Diccionaris Electrònics per a la Traducció Automàtica i Motors de Cerca**2013/2014**

Codi: 43024

Crèdits: 10

Titulació	Tipus	Curs	Semestre
4313807 Tractament de la informació i comunicació multilingüe	OT	0	1

Professor de contacte

Nom: Sandrine Fuentes Crespo

Correu electrònic: Sandrine.Fuentes@uab.cat

Utilització d'idiomes

Llengua vehicular majoritària: espanyol (spa)

Prerequisits

No hi ha prerequisits específics per aquest mòdul.

Objectius

Aquest mòdul té com a objectiu familiaritzar als estudiants amb l'elaboració de recursos lingüístics per al processament automàtic del llenguatge natural. Es presentaran en particular els criteris per a la recopilació de grans corpora lingüístics, la macro- i micro-estructura d'un diccionari electrònic, així com les bases teòriques per a la creació de gramàtiques electròniques. Així mateix es presentaran eines per a l'elaboració i aplicació de tals recursos.

Competències

- Tractament de la informació i comunicació multilingüe
- Analitzar i gestionar el multilingüisme en contextos i aplicacions diferents.
- Buscar informació en la literatura científica fent servir els canals apropiats i integrar aquesta informació per plantejar i contextualitzar un tema de recerca
- Que els estudiants sàpiguin aplicar els coneixements adquirits i la seva capacitat de resolució de problemes en entorns nous o poc coneguts dins de contextos més amplis (o multidisciplinaris) relacionats amb la seva àrea d'estudi.
- Que els estudiants tinguin les habilitats d'aprenentatge que els permetin continuar estudiant, en gran manera, amb treball autònom a autodirigit
- Seleccionar i aplicar una metodologia científica adequada a l'objecte d'estudi
- Utilitzar i avaluar programari específic per al processament automàtic del llenguatge (escrit) i desenvolupar nous recursos lingüístics (especialitat PLN).
- Utilitzar les tecnologies de la informació i de la comunicació i les eines específiques per a l'anàlisi i l'obtenció de corpus.

Resultats d'aprenentatge

1. Aplicar amb rigor i de manera sistemàtica la metodologia científica adoptada
2. Aplicar els coneixements necessaris per a la gestió de l'intercanvi de documents digitals multilingües (adaptació d'alfabets, convencions de símbols i formats, etc.).
3. Aplicar els coneixements teòrics adquirits per al disseny de nous recursos lingüístics.
4. Completar la informació rebuda en el mòdul mitjançant la recerca de bibliografia complementària.
5. Elaborar un corpus en funció de diferents objectius d'investigació o aplicació.

6. Emprar les eines informàtiques actuals per a la gestió de referències bibliogràfiques
7. Enriquir els recursos existents en una plataforma destinada al processament de corpus amb gramàtiques i bases de dades lèxiques.
8. Identificar la metodologia més adequada per a la resolució d'un problema lingüístic
9. Seleccionar la bibliografia científica més rellevant i actualitzada relacionada amb l'objecte d'estudi.
10. Usar de manera apropiada les tecnologies lingüístiques per a la elaboració de corpus i el seu posterior processament.
11. Utilitzar les eines de lliure accés per al tractament de dades lingüístiques a gran escala.
12. Utilitzar una plataforma de processament de corpus lingüístic.

Continguts

I. Elaboració de recursos electrònics per al processament automàtic del llenguatge natural

- Corpus lingüístics (definició, classificació, disseny i fases de construcció)
- Els diccionaris electrònics (concepte, tipus, finalitats, avantatges i inconvenients...)
- Els diccionaris electrònics orientats al PLN: macroestructura i microestructura
- Gramàtiques electròniques

II. Ús de la plataforma lingüística NooJ

- Textos i corpus
- Descripció del vocabulari
- Construcció de gramàtiques
- Gramàtiques locals
- Anàlisi estructural i transformacional

Metodologia

- Classes expositives
- Classes de resolució de problemes/casos/exercicis
- Elaboració de treballs
- Estudi personal

Activitats formatives

Títol	Hores	ECTS	Resultats d'aprenentatge
Tipus: Dirigides			
ELABORACIÓ DE CORPUS LINGUISTICS	30	1,2	5, 10
ÚS DE LA PLATAFORMA LINGUISTICA NOOJ I DESENVOLUPAMENT DE NOUS RECURSOS LINGUISTICS	45	1,8	7, 12
Tipus: Supervisades			
APLICACIÓ DE LA METODOLOGIA	25	1	1, 8
CREACIÓ I DESENVOLUPAMENT DE RECURSOS	25	1	3, 11
Tipus: Autònomes			
CERCA BIBLIOGRÀFICA	10	0,4	6, 9
GESTIÓ DE RECURSOS	15	0,6	3, 4, 11

Avaluació

Els estudiants hauran de lliurar diferents tipus d'exercicis al llarg del semestre per tal de posar en pràctica els coneixements adquirits i desenvolupar les competències indicades.

Activitats d'avaluació

Títol	Pes	Hores	ECTS	Resultats d'aprenentatge
ENTREGA DE TREBALL	40%	40	1,6	1, 3, 4, 5, 6, 8, 9, 10, 11
PROVA PRÀCTICA	60%	60	2,4	2, 7, 12

Bibliografia

Adolphs, S. & Lin, P. (2010) «Corpus linguistics». In J. Simpson (Ed.), The Routledge handbook of applied linguistics. Oxford: Routledge.

Blanco, X. (1999) Lexicographie bilingue français-espagnol et classes d'objets. Bellaterra, Universitat Autònoma de Barcelona.

Buvet, P.-A., Grezka, A. (2009) «Les dictionnaires électroniques du modèle des classes d'objets». Langages 176, pp. 63-79.

Caravedo, R. (1999) Lingüística del corpus. Cuestiones teórico-metodológicas aplicadas al español. Salamanca: Ediciones de la Universidad de Salamanca.

Fuchs, C. (1993) Linguistique et Traitements Automatiques des Langues. Paris: Hachette.

Gross, G. (2013) Manual de análisis lingüístico. Aproximación sintáctico-semántica al léxico. Barcelona: Editorial UOC.

Gross, G. (2004) «Réflexions sur le traitement automatique des langues», Actes de JADT 2004, Vol. 1. 545-556.

Gross G. (1999) «Élaboration d'un dictionnaire électronique». Bulletin de la Société de Linguistique de Paris, Tome (XCIV/1). Peeters : 113-138.

Gross, G. (1992) «Forme d'un dictionnaire électronique». A. Clas y H. Safar (dirs.) L'environnement traductionnel de l'an 2001. Presses de l'Université de Québec. Montréal. pp. 255-271.

Habert, B., Nazarenko, A., & Salem, A. (1997) Les linguistiques de corpus. Paris: Armand Colin.

Parodi, G. (2008) «Lingüística de corpus: Una introducció al àmbit». Revista de Lingüística Teórica y Aplicada, 46 (1), pp. 93-119.

Silberstein, M. (2003) NooJ manual. [disponible en <http://www.nooj4nlp.net>].

VidalBeneyto, José (dir.) (1991). Las industrias de la lengua, equipo de traductores coordinado por Manuel Alvar Ezquerro (Biblioteca del Libro, V), Salamanca - Madrid, Fundación Germán Sánchez Ruipérez, Madrid: Pirámide.

Vučković K., Bekavac B., Silberstein M. (dirs.) (2012) Selected Papers from the 2011 International NooJ Conference. Cambridge Scholars Publishing: Newcastle.