
Ano Letivo 2018-19

Unidade Curricular CONCEÇÃO E GESTÃO DE RECURSOS LINGÜÍSTICOS E COMUNICATIVOS

Cursos LÍNGUAS E COMUNICAÇÃO (1.º ciclo)
LÍNGUAS, LITERATURAS E CULTURAS (1.º ciclo) (*)
- RAMO DE PORTUGUÊS E FRANCÊS (1.º Ciclo)
- RAMO DE PORTUGUÊS E ESPANHOL (1.º Ciclo)
- RAMO DE ESTUDOS PORTUGUESES E LUSÓFONOS (1.º Ciclo)

(*) Curso onde a unidade curricular é opcional

Unidade Orgânica Faculdade de Ciências Humanas e Sociais

Código da Unidade Curricular 15711016

Área Científica CIÊNCIAS DA COMUNICAÇÃO

Sigla

Línguas de Aprendizagem Português (algumas referências estão em inglês; o manual do principal software é fornecido em inglês e francês).

Modalidade de ensino Presencial

Docente Responsável Jorge Manuel Evangelista Baptista

DOCENTE	TIPO DE AULA	TURMAS	TOTAL HORAS DE CONTACTO (*)
Jorge Manuel Evangelista Baptista	OT; PL; TP	TP1; PL1; OT1	13TP; 26PL; 5OT

* Para turmas lecionadas conjuntamente, apenas é contabilizada a carga horária de uma delas.

ANO	PERÍODO DE FUNCIONAMENTO*	HORAS DE CONTACTO	HORAS TOTAIS DE TRABALHO	ECTS
2º	S2	13TP; 26PL; 5OT	140	5

* A-Anual; S-Semestral; Q-Quadrimestral; T-Trimestral

Precedências

Sem precedências

Conhecimentos Prévios recomendados

Conhecimentos básicos de informática na ótica do utilizador (gestão de ficheiros, utilização de editor de texto e de folha de cálculo).

Objetivos de aprendizagem (conhecimentos, aptidões e competências)

Capacidade de determinar necessidades de recursos linguísticos e comunicativos para diversas situações, de os localizar e obter ou, em alternativa, de conceber uma metodologia com vista à sua construção, avaliação e distribuição; capacidade de estruturar um projeto de construção de recursos em equipa; requisitos básicos para documentação de recursos e sua distribuição.

Conteúdos programáticos

Codificação de caracteres e conversão de ficheiros (Unicode, IsoLatin2). Expressões Regulares, máscaras lexicais e morfológicas. Máquinas de estados finitos: autómatos e transdutores. Segmentação de texto: tokenização e segmentação de frases. Dicionários electrónicos e para utilizadores humanos. Tratamento da morfologia flexional e derivacional, lema, estema e forma flexionada. Codificação da informação morfossintática. Processamento sintático: conceito de 'chunk', 'phrase' e de constituinte sintático, recursividade e regras de dependência. Filtros lexicais. Gramáticas locais. Interseção de transdutores e matrizes léxico-sintáticas: processamento de expressões idiomáticas, construção de grafo de referência e geração automática de transdutores. Completude lexical, medidas de associação e lexicografia. Análise de diferentes recursos linguísticos existentes. Identificação, extração, armazenamento, distribuição e gestão da informação linguística. Bases de dados e bases de conhecimento.

Metodologias de ensino (avaliação incluída)

A avaliação por frequência com pelo menos um elemento de avaliação presencial (50%) e o exame de época normal (50%). Critério de assiduidade: exigência de assistência a, pelo menos, 75% das horas de contacto para admissão a avaliação no exame normal. Todos os alunos regularmente inscritos estão automaticamente admitidos a exame de época de recurso. Em todos os elementos de avaliação (incluindo exames) será atribuída uma percentagem de 20% à correção linguística e discursiva e à estruturação textual.

Bibliografia principal

- Baptista, J. Conceção e Gestão de Recursos Linguísticos e Comunicativos (Manual). Faro: UALG/FCHS/DAH.
- Clark, A.; Fox, C.; Lappin, S. 2010. Handbook of Computational Linguistics. Wiley-Blackwell (ISBN: 9781405155816)
- Jurafsky, D; Martin, J. 2000. Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition. New Jersey: Prentice Hall.
- Manning, Ch.; Schütze, H. 1999. Foundations of Statistical Natural Language Processing. Cambridge, MA: MIT Press.
- Mitkov, R. (ed.) 2003. The Oxford Handbook of Computational Linguistics. Oxford Academic Press.
- Ranchhod, E. (Org.) 2001. Tratamento das Línguas por Computador. Uma introdução à Linguística Computacional e suas aplicações. Lisboa: Caminho.
- Paumier, S. 2016 [Março]. Unitex 3.1 - Manuel /Manual, IGM, Univ. Marne-la-Vallée, Paris. <http://www-igm.univ-mlv.fr/~unitex/>
- Sardinha, T. 2004. Linguística de Corpus. Rio de Janeiro : Manole.

Academic Year 2018-19

Course unit DESIGN AND MANAGEMENT OF LANGUAGE AND COMMUNICATION RESOURCES

Courses LANGUAGES AND COMMUNICATION (1st Cycle)

LANGUAGES, LITERATURES AND CULTURES (*)

- RAMO DE PORTUGUÊS E FRANCÊS (1.º Ciclo)
- RAMO DE PORTUGUÊS E ESPANHOL (1.º Ciclo)
- RAMO DE ESTUDOS PORTUGUESES E LUSÓFONOS (1.º Ciclo)

(*) Optional course unit for this course

Faculty / School Faculdade de Ciências Humanas e Sociais

Main Scientific Area CIÊNCIAS DA COMUNICAÇÃO

Acronym

Language of instruction Portuguese (some references are in English; the main software manual is provided in English and French).

Teaching/Learning modality Classroom (i.e. face-to-face) learning/teaching.

Coordinating teacher Jorge Manuel Evangelista Baptista

Teaching staff	Type	Classes	Hours (*)
Jorge Manuel Evangelista Baptista	OT; PL; TP	TP1; PL1; OT1	13TP; 26PL; 5OT

* For classes taught jointly, it is only accounted the workload of one.

Contact hours

T	TP	PL	TC	S	E	OT	O	Total
0	13	26	0	0	0	5	0	140

T - Theoretical; TP - Theoretical and practical ; PL - Practical and laboratorial; TC - Field Work; S - Seminar; E - Training; OT - Tutorial; O - Other

Pre-requisites

no pre-requisites

Prior knowledge and skills

Basic computer skills (file management, use of a text editor and of spreadsheets).

The students intended learning outcomes (knowledge, skills and competences)

Ability to determine needs of linguistic and communicative resources for different situations; to locate and obtain or, alternatively, to devise a methodology for the construction, evaluation and distribution; ability to structure a construction project team resources; basic requirements for documentation of resources and their distribution.

Syllabus

Character encoding and file conversion (Unicode, IsoLatin2). Regular Expressions, lexical and morphological masks. Finite state machines: automata and transducers. Text segmentation: tokenization and segmentation phrases. Machine- (or electronic) and Human users' dictionaries. Processing inflectional and derivational morphology, lemma, stem and inflected form. Codification of morphosyntactic information. Syntactic processing: concepts of 'chunk', 'phrase' and syntactical constituent, recursion and dependency rules. Lexical filters. Local grammars. Intersecting transducers and lexical-syntactical matrices: processing idioms, reference graph construction and automatic generation of transducers. Lexical completeness, association measures and lexicography. Analysis of different existing language resources. Identification, extraction, storage, distribution and management of linguistic information. Databases and knowledge bases.

Teaching methodologies (including evaluation)

Evaluation by test with at least one in-class assessment (50%) and an exam during época normal (50%). Attendance criteria: in order to be admitted to exam during época normal, students are required to attend, at least, 75% of the contact hours. All enrolled students are automatically admitted to exam during época de recurso. In all evaluation moments (including exams) 20% of the total grade will be attributed to linguistic, discursive and textual structuring correction.

Main Bibliography

- Baptista, J. Conceção e Gestão de Recursos Linguísticos e Comunicativos (Manual). Faro: UALG/FCHS/DAH.
- Clark, A.; Fox, C.; Lappin, S. 2010. Handbook of Computational Linguistics. Wiley-Blackwell (ISBN: 9781405155816)
- Jurafsky, D; Martin, J. 2000. Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition. New Jersey: Prentice Hall.
- Manning, Ch.; Schütze, H. 1999. Foundations of Statistical Natural Language Processing. Cambridge, MA: MIT Press.
- Mitkov, R. (ed.) 2003. The Oxford Handbook of Computational Linguistics. Oxford Academic Press.
- Ranchhod, E. (Org.) 2001. Tratamento das Línguas por Computador. Uma introdução à Linguística Computacional e suas aplicações. Lisboa: Caminho.
- Paumier, S. 2016 [Março]. Unitex 3.1 - Manuel /Manual, IGM, Univ. Marne-la-Vallée, Paris. <http://www-igm.univ-mlv.fr/~unitex/>
- Sardinha, T. 2004. Linguística de Corpus. Rio de Janeiro : Manole.