
Ano Letivo 2020-21

Unidade Curricular CONCEÇÃO E GESTÃO DE RECURSOS LINGUÍSTICOS E COMUNICATIVOS

Cursos LÍNGUAS E COMUNICAÇÃO (1.º ciclo)

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 eletrónicos e para utilizadores humanos. Tratamento da morfologia flexional e derivacional, lema, estema e forma flexionada. Codificação da informação morfosintática. 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 dois elementos de avaliação presencial (testes: 50%+50%). Nota de aprovação: 9,5/20+critério de assiduidade. 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. 2019. Conceção e Gestão de Recursos Linguísticos e Comunicativos (Manual) (3a edição). Faro: UALG/FCHS/DAH. (policopiado)
- Clark, A.; Fox, C.; Lappin, S. 2010. Handbook of Computational Linguistics. Wiley-Blackwell (ISBN: 9781405155816)
- Jurafsky, D; Martin, J. 2008. Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition. New Jersey: Prentice Hall. (ASIN: B00XDGGF5M; <https://web.stanford.edu/~jurafsky/NLPCourseraSlides.html>)
- Manning, Ch.; Schütze, H. 1999. Foundations of Statistical Natural Language Processing. Cambridge, MA: MIT Press. [Cap.5]
- Ranchhod, E. (Org.) 2001. Tratamento das Línguas por Computador. Uma introdução à Linguística Computacional e suas aplicações. Lisboa: Caminho.
- Paumier, S. 2016 [Março]. Unitex 3.2 - Manuel /Manual, IGM, Univ. Marne-la-Vallée, Paris.
<https://unitexgramlab.org/releases/3.2/man/Unitex-GramLab-3.2-usermanual-en.pdf>

Academic Year 2020-21

Course unit DESIGN AND MANAGEMENT OF LANGUAGE AND COMMUNICATION RESOURCES

Courses LANGUAGES AND COMMUNICATION (1st Cycle)

Faculty / School FACULTY OF HUMAN AND SOCIAL SCIENCES

Main Scientific Area

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 them, or, alternatively, to devise a methodology for their construction, evaluation and distribution; ability to structure a linguistic resources construction project within a team; basic requirements for documentation of and distribution of linguistic resources.

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 resources. Databases and knowledge bases.

Teaching methodologies (including evaluation)

Two elements of face-to-face evaluation ("tests"): 50%+50%. Approval grade: 9,5/20+Attendance criterion. Attendance criteria: requirement of attendance with at least 75% of the contact hours for admission to the normal examination season. All students regularly enrolled are automatically admitted to the appeal examination season . All elements of assessment (including examinations) will be assigned a percentage of 20% to linguistic and discursive correction and textual structuring.

Main Bibliography

Baptista, J. 2019. Conceção e Gestão de Recursos Linguísticos e Comunicativos (Manual) (3a edição). Faro: UALG/FCCHS/DAH. (policopiado)

Clark, A.; Fox, C.; Lappin, S. 2010. Handbook of Computational Linguistics. Wiley-Blackwell (ISBN: 9781405155816)

Jurafsky, D; Martin, J. 2008. Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition. New Jersey: Prentice Hall. (ASIN: B00XDGGF5M; <https://web.stanford.edu/~jurafsky/NLPCourseSlides.html>)

Manning, Ch.; Schütze, H. 1999. Foundations of Statistical Natural Language Processing. Cambridge, MA: MIT Press. [Cap.5]

Ranchhod, E. (Org.) 2001. Tratamento das Línguas por Computador. Uma introdução à Linguística Computacional e suas aplicações. Lisboa: Caminho.

Paumier, S. 2016 [Março]. Unitex 3.2 - Manuel /Manual, IGM, Univ. Marne-la-Vallée, Paris. <https://unitexgramlab.org/releases/3.2/man/Unitex-GramLab-3.2-usermanual-en.pdf>