

Primeira Escola de Verão da Linguateca

*organizada pela **Linguateca** (principalmente o seu pólo do Porto) e tendo como anfitrião o **Centro de Linguística da Universidade do Porto***

Área: Processamento computacional da língua portuguesa (escrita)

Local: Universidade do Porto

Data: 10 a 14 de Julho de 2006

Público alvo:

- alunos de doutoramento e mestrado nas áreas de processamento de linguagem natural, linguística computacional, linguística com corpora, recolha, extracção e gestão de informação, terminologia, lexicografia, tradução, inteligência artificial, ...;
- investigadores em empresas e no mundo académico que trabalhem ou tencionem vir a trabalhar em processamento computacional da língua portuguesa;
- todos os interessados nestas áreas que queiram apostar numa formação mais interdisciplinar.

URL com informação detalhada e actualizada: <http://www.linguateca.pt/EscolaVerao2006/>

1. Corpo docente e organização

Corpo docente

Ana Frankenberg-Garcia, ISLA

Belinda Maia, Universidade do Porto

Diana Santos, SINTEF

Eckhard Bick, University of Southern Denmark

Elisabete Ranchhod, Universidade de Lisboa

José João Dias de Almeida, Universidade do Minho

Lúcia Rino, Universidade Federal de São Carlos

Mário J. Silva, Universidade de Lisboa

Rachel Aires, Universidade de Goiânia, *a confirmar presença*

Monitores (doutorandos)

Alberto Simões, Universidade do Minho

Anabela Barreiro, Universidade do Porto

Cristina Mota, Instituto Superior Técnico

Luís Sarmento, Universidade do Porto

Marcirio Chaves, Universidade de Lisboa

Nuno Seco, Universidade de Coimbra

Susana Afonso, Universidade de Manchester

Organização local

Belinda Maia (responsável), Universidade do Porto

Ana Sofia Pinto, pólo do Porto da Linguateca

António Silva, pólo do Porto da Linguateca

2. Objectivos

A Primeira Escola de Verão da Linguateca tem como principais objectivos:

1. estimular a cooperação entre os investigadores do processamento computacional da língua portuguesa com formações de base diferentes, aproximando mutuamente as comunidades provenientes das áreas de letras, em particular de linguística, e as comunidades formadas pelas áreas da engenharia, em particular a informática;
2. fomentar a discussão sobre tópicos actuais e sobre vertentes aplicacionais do processamento computacional da língua portuguesa;
3. criar um espaço de formação especializada em processamento computacional da língua portuguesa complementar ao actualmente oferecido no ensino superior em Portugal;
4. criar um ambiente propício à criação de ideias e ao estabelecimento de contactos e parcerias dentro e fora da comunidade;
5. utilizar a massa crítica da Linguateca para divulgar para o exterior o próprio saber-fazer dentro da Linguateca.

Esta Escola de Verão será abrangente pela diversidade de áreas que aborda, mas ao mesmo tempo é orientada para aplicações práticas actuais, em expansão no momento presente. O fio condutor da escola é o tema "Questões actuais no processamento computacional da língua portuguesa".

3. Organização

Durante uma semana serão realizadas várias sessões (ou módulos) que tentarão ser ao mesmo tempo teóricas e práticas, enquadrando a teoria em aplicações práticas (veja-se lista em baixo). Cada uma destas sessões será sempre leccionada por pelo menos duas pessoas, sendo uma da área da linguística e outra da área da engenharia, de forma a não privilegiar nenhuma das áreas de origem e fomentando o debate e a integração dos vários saberes, quer entre o pessoal docente, quer entre a audiência. A maioria dos módulos incluirá uma componente prática, de forma a estimular o aprofundamento de ideias e experiências, estando para isso previstos espaços apetrechados para o efeito na Faculdade de Letras da Universidade do Porto.

Cada módulo terá a duração aproximada de um dia, a dividir pelos dois (ou mais) professores, que organizarão o tempo da sessão entre si, incluindo a componente prática. Haverá também sessões sobre métodos e prática da avaliação na área, que será a única "transversal". Note-se que, ao contrário de outras escolas de Verão, não haverá módulos em paralelo, esperando-se que os alunos assistam a todas as sessões.

Um primeiro esboço de horário é o seguinte:

| Horário | 2ª feira, 10 | 3ª feira, 11 | 4ª feira, 12 | 5ª feira, 13 | 6ª feira, 14 |
|----------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| 9h30-11h00 | Abertura | Módulo 2 | Módulo 3 | Módulo 4 | Módulo 5 |
| 11h30-13h00 | Módulo 1 | Módulo 2 | Módulo 3 | Módulo 4 | Módulo 5 |
| 14h30-16h00 | Módulo 1 | Módulo 2 | Módulo 3 | Módulo 4 | Módulo 5 |
| 16h30-18h00 | Módulo 1 | Círculo | Avaliação | Avaliação | Discussão |

No período denominado “Círculo” propomos ouvir uma apresentação curta do trabalho de todos os participantes (5 a 10 minutos cada um), enquanto que em “Discussão” queremos ouvir a apreciação geral dos alunos e seus desejos e sugestões para futuras Escolas de Verão nesta área.

Segue-se uma descrição resumida de cada módulo.

4. Programa

Abertura: Questões actuais no processamento computacional da língua portuguesa

Diana Santos

Uma panorâmica do que existe e do esforço que há em PLN em Portugal e no Brasil, juntamente com uma tentativa de previsão de quais os assuntos mais importantes a resolver na próxima década.

Módulo 1: Tradução automática: a tecnologia e a sua aplicação

Belinda Maia e Diana Santos/José João Almeida

Problemas associados à tradução do ponto de vista linguístico e cultural. Técnicas e Métodos de TA. A aplicação da TA no dia-a-dia. Avaliação de TA

Módulo 2: Análise sintáctica computacional: problemas e métodos

Eckhard Bick e Elisabete Ranchhod

A análise sintáctica do português, questões linguísticas e dificuldades. Sistemas automáticos de análise sintáctica para o português. Extracção de padrões sintácticos de corpora. A questão das expressões multi-palavras e a sua importância no processamento das línguas. Aplicação em sistemas de resposta automática a perguntas (RAP) e outras tarefas.

Módulo 3: Desenho, construção e utilização de corpora (e sua aplicação em ensino e extracção de conhecimento)

Ana Frankenberg-Garcia, Belinda Maia, Diana Santos e José João Almeida

A teoria e as técnicas de construção de corpora. Ferramentas e métodos de pesquisa. O uso de corpora no ensino e na construção de recursos de conhecimento.

Módulo 4: Terminologia, tesauros e ontologias (e a sua aplicação em tradução, recolha e extracção de informação)

Belinda Maia e Mário J. Silva

Questões associadas à Terminologia. Dificuldades e desafios. Os tesauros e ontologias, sua construção e utilização, em aplicações de recolha e extracção de informação. A Web semântica e os motores de busca actuais.

Módulo 5: Entrando em linha de conta com os utilizadores

Ana Frankenberg-Garcia e Rachel Aires/Diana Santos

Estudo de utilizadores e adaptação de sistemas às necessidades de utilizadores, no contexto de aplicações de linguagem natural ou linguística. Desenho de interfaces, estudo de diários (logs), casos concretos de desajuste entre a funcionalidade e o funcionamento.

Avaliação

Diana Santos e Lucia Rino

Questões principais; métodos e resultados. O paradigma da avaliação conjunta. Casos concretos, relacionados com os módulos da escola. Avaliação de sumarização em português, de recolha de informação, de análise sintáctica e morfológica.