



**SCC5908 INTRODUÇÃO AO
PROCESSAMENTO DE LÍNGUA
NATURAL**

Thiago A. S. Pardo

UM POUCO DA HISTÓRIA

- Até recentemente: tópicos
 - Aplicações, principalmente, com alguma fundamentação

- Foco desta disciplina
 - Fundamentação
 - Modelos formais e métodos

- Em princípio, oferecida em todo 1º semestre do ano
 - Esta é a 2ª vez

ESTA DISCIPLINA

- Um recorte do mundo de PLN
- Predominantemente teórica
- Cada tópico abordado poderia ser estendido e se tornar uma nova disciplina por si só

3

EMENTA

- Visão geral da área de Processamento de Línguas Naturais: recursos, ferramentas e aplicações; abordagens e paradigmas. Conceitos básicos: córpus, treebanks, leis de Zipf, Luhn, entropia e complexidade das línguas, níveis de tratamento das línguas naturais. Técnicas de interpretação e geração de língua natural. Introdução à morfologia e morfossintaxe, técnicas de etiquetagem morfossintática. Modelos e técnicas de análise sintática. Modelos formais de representação semântica: semântica lexical e sentencial. Discurso e pragmática: conceitos, modelos e métodos de análise. Prática com toolkits e ferramentas de Processamento de Línguas Naturais.

4

CENÁRIO

- Cenário confortável para o aluno atual
 - [Processamento Automático de Línguas Naturais](#) na UNESP/Araraquara: tópicos, foco mais linguístico
 - [Tópicos de PLN](#) no ICMC-USP
 - [Introdução ao PLN](#) no ICMC-USP
 - [A face linguística do PLN](#) no DL-UFSCar: foco mais linguístico

5

CENÁRIO

- Cenário confortável para o aluno atual
 - [Seminários de LHT](#) no DL-UFSCar
 - [Sumarização automática](#) no DL-UFSCar
 - [Semântica lexical](#) no DL-UFSCar
 - [Introdução à IA](#) no ICMC-USP
 - [Introdução a AM](#) no ICMC-USP
 - Etc.

6

COMPROMISSOS

○ Alunos

- Linguistas e informatas
 - Esforço dos dois lados
 - Linguistas: estatística e modelos mais computacionais
 - Informatas: base lingüística e modelos teóricos
- Ouvintes
 - Mesmos compromissos dos demais, mas podem se abster das provas ☺

○ Semanalmente, 12 créditos

- 3 horas de aula
- 9 horas de estudo fora de sala de aula

7

AVALIAÇÃO

○ Provas periódicas

- Início/fim de algumas aulas, abordando assunto das aulas e estudos fora de sala de aula
 - Nota final = média das 70% melhores notas

○ 70% de presença

- 10 aulas, com 4 faltas permitidas (usem sabiamente)

8

MATERIAL

- CoteiaWIKI
 - wiki.icmc.usp.br
- ACL (*Association for Computational Linguistics*) e NILC
 - <http://aclweb.org/>
 - <http://www.nilc.icmc.usp.br>
- Bibliografia **básica**
 - Allen, J. (1995). *Natural Language Understanding*. 2a. edição. Addison Wesley. 654p.
 - Jurafsky, D. and Martin, J.H. (2000). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*. 1a. edição. Prentice Hall. 934p.
 - Manning, C.D. and Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. 1a. edição. The MIT Press. 680p.
 - Mitkov, R. (2005). *The Oxford Handbook of Computational Linguistics*. Oxford University Press, 806p.
 - Reiter, E. and Dale, R. (2000). *Building Natural Language Generation Systems*. Cambridge University Press.

9

PLANEJAMENTO

- Previsão
 - Dias 21/04 e 23/06: não há aulas, mas haverá leituras
 - Eventualmente, seminários e palestras do NILC podem integrar a disciplina

<i>Data</i>	<i>Conteúdo previsto</i>
17/03	
24/03	
31/03	P1
07/04	
14/04	P2
21/04	Não há aula
28/04	
05/05	P3
12/05	
19/05	P4
26/05	
02/06	P5
09/06	
16/06	P6
23/06	Não há aula
30/06	P7

10

ALUNOS

- Quem são?
- Por que estão cursando a disciplina?
- O que esperam da disciplina?

11

EXPECTATIVA

- Postura de pós-graduação
 - Pontualidade
 - Iniciativa
 - Participação
 - Compromisso
- Ouvintes
 - Mesmos compromissos dos demais, mas podem se abster das provas ☺

12