

CS-431

**Introduction to natural language processing**

Chappelier Jean-Cédric, Rajman Martin

<b>Cursus</b>	<b>Sem.</b>	<b>Type</b>
Computer science	MA1, MA3	Opt.
Cybersecurity	MA1, MA3	Opt.
Data Science	MA1, MA3	Opt.
Data science minor	H	Opt.
Digital Humanities	MA1, MA3	Opt.
SC master EPFL	MA1, MA3	Opt.
UNIL - Sciences forensiques	H	Opt.

Language of teaching	English
Credits	4
Session	Winter
Semester	Fall
Exam	Written
Workload	120h
Weeks	14
<b>Hours</b>	<b>4 weekly</b>
Courses	2 weekly
Exercises	2 weekly
<b>Number of positions</b>	

**Summary**

The objective of this course is to present the main models, formalisms and algorithms necessary for the development of applications in the field of natural language information processing. The concepts introduced during the lectures will be applied during practical sessions.

**Content**

Several models and algorithms for automated textual data processing will be described: (1) morpho-lexical level: electronic lexica, spelling checkers, ...; (2) syntactic level: regular, context-free, stochastic grammars, parsing algorithms, ...; (3) semantic level: models and formalisms for the representation of meaning, ...

Several application domains will be presented: Linguistic engineering, Information Retrieval, Text mining (automated knowledge extraction), Textual Data Analysis (automated document classification, visualization of textual data).

**Keywords**

Natural Language Processing; Computational Linguistics; Part-of-Speech tagging; Parsing

**Learning Outcomes**

By the end of the course, the student must be able to:

- Compose key NLP elements to develop higher level processing chains
- Assess / Evaluate NLP based systems
- Choose appropriate solutions for solving typical NLP subproblems (tokenizing, tagging, parsing)
- Describe the typical problems and processing layers in NLP
- Analyze NLP problems to decompose them in adequate independent components

**Teaching methods**

Ex cathedra ; practical work on computer

**Expected student activities**

attend lectures and practical sessions, answer quizzes.

**Assessment methods**

4 quiz during semester 25%, final exam 75%

**Resources**

**Virtual desktop infrastructure (VDI)**

No

**Bibliography**

1. M. Rajman editor, "*Speech and Language Engineering*", EPFL Press, 2006.
2. Daniel Jurafsky and James H. Martin, "*Speech and Language Processing*", Prentice Hall, 2008 (2nd edition)
3. Christopher D. Manning and Hinrich Schütze, "*Foundations of Statistical Natural Language Processing*", MIT Press, 2000
4. Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, "*Introduction to Information Retrieval*", Cambridge University Press, 2008
5. Nitin Indurkha and Fred J. Damerau editors, "*Handbook of Natural Language Processing*", CRC Press, 2010 (2nd edition)

**Ressources en bibliothèque**

- [Handbook of Natural Language Processing / Indurkha](#)
- [Introduction to Information Retrieval / Manning](#)
- [Foundations of Statistical Natural Language Processing / Manning](#)
- [Speech and Language Engineering / Rajman](#)
- [Speech and Language Processing / Jurafsky](#)

**Websites**

- <http://coling.epfl.ch>