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.

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, Textual Data Analysis (automated document classification, visualization of textual data).

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
Flipped classroom (reviews and supervised "hands-on" in class); practical work on computer

Expected student activities
attend lectures and practical sessions, answer quizzes.

Assessment methods
4 quiz during semester 16%, final exam 84%

Resources

Bibliography
3. Christopher D. Manning and Hinrich Schütze, "Foundations of Statistical Natural Language Processing", MIT Press, 2000

Ressources en bibliothèque
• Handbook of Natural Language Processing / Indurkhya
• Introduction to Information Retrieval / Manning
• Foundations of Statistical Natural Language Processing / Manning
• Speech and Language Engineering / Rajman
• Speech and Language Processing / Jurafsky

Websites
• https://coling.epfl.ch/

Moodle Link
• https://go.epfl.ch/CS-431