## Introduction to natural language processing

Chappelier Jean-Cédric, Rajman Martin

### Cursus

<table>
<thead>
<tr>
<th>Cursus</th>
<th>Sem.</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cybersecurity</td>
<td>MA1, MA3</td>
<td>Opt.</td>
</tr>
<tr>
<td>Data Science</td>
<td>MA1, MA3</td>
<td>Opt.</td>
</tr>
<tr>
<td>Humanités digitales</td>
<td>MA1, MA3</td>
<td>Opt.</td>
</tr>
<tr>
<td>Informatique</td>
<td>MA1, MA3</td>
<td>Opt.</td>
</tr>
<tr>
<td>SC master EPFL</td>
<td>MA1, MA3</td>
<td>Opt.</td>
</tr>
<tr>
<td>UNIL - Sciences forensiques</td>
<td>H</td>
<td>Opt.</td>
</tr>
</tbody>
</table>

### 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 independant 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
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
• Speech and Language Processing / Jurafsky
• Speech and Language Engineering / Rajman
• Foundations of Statistical Natural Language Processing / Manning

Websites
• http://coling.epfl.ch