

SYLLABUS

1. Information regarding the programme

| | |
|-------------------------------------|---|
| 1.1 Higher education institution | Babeş Bolyai University |
| 1.2 Faculty | Faculty of Mathematics and Computer Science |
| 1.3 Department | Department of Computer Science |
| 1.4 Field of study | Computer Science |
| 1.5 Study cycle | Bachelor |
| 1.6 Study programme / Qualification | Computer Science/ Applied Computational Intelligence |

2. Information regarding the discipline

| | | | | | | | |
|----------------------------|--|--------------|----------|-------------------------|----------|------------------------|-----------------|
| 2.1 Name of the discipline | Introduction to Natural Language Processing | | | | | | |
| 2.2 Course coordinator | Lecturer Ph.D. Lupea Mihaiela | | | | | | |
| 2.3 Seminar coordinator | Lecturer Ph.D. Lupea Mihaiela | | | | | | |
| 2.4. Year of study | 3 | 2.5 Semester | 2 | 2.6. Type of evaluation | C | 2.7 Type of discipline | optional |

3. Total estimated time (hours/semester of didactic activities)

| | | | | | |
|---|-----------|----------------------|-----------|------------------------|-----------|
| 3.1 Hours per week | 3 | Of which: 3.2 course | 2 | 3.3 seminar/laboratory | 1 |
| 3.4 Total hours in the curriculum | 36 | Of which: 3.5 course | 24 | 3.6 seminar/laboratory | 12 |
| Time allotment: | hours | | | | |
| Learning using manual, course support, bibliography, course notes | 30 | | | | |
| Additional documentation (in libraries, on electronic platforms, field documentation) | 40 | | | | |
| Preparation for seminars/labs, homework, papers, portfolios and essays | 49 | | | | |
| Tutorship | 10 | | | | |
| Evaluations | 10 | | | | |
| Other activities: | - | | | | |
| 3.7 Total individual study hours | 139 | | | | |
| 3.8 Total hours per semester | 175 | | | | |
| 3.9 Number of ECTS credits | 7 | | | | |

4. Prerequisites (if necessary)

| | |
|-------------------|---|
| 4.1. curriculum | Formal languages, Data structures, Graphs Algorithms |
| 4.2. competencies | Programming skills in a high level programming language |

5. Conditions (if necessary)

| | |
|--------------------------------------|--|
| 5.1. for the course | |
| 5.2. for the seminar /lab activities | <ul style="list-style-type: none"> Laboratory with computers; high level programming language environment (.NET or any Java environment a.s.o.) |

6. Specific competencies acquired

| | |
|----------------------------------|---|
| Professional competencies | <ul style="list-style-type: none"> • Assimilation of mathematical concepts and formal models to understand, verify and validate software systems ; • Advanced ability to approach, model and solve phenomena and problems from natural language and economy using fundamental knowledge from mathematics and computer science; • Ability to approach and solve complex problems using various techniques of computational intelligence; • Proficient use of methodologies and tools specific to programming languages and software systems. |
| Transversal competencies | <ul style="list-style-type: none"> • Etic and fair behavior, committment to professional deontology • Team work capabilities; able to fulfill different roles • Professional communication skills; concise and precise description, both oral and written, of professional results , negotiation abilities; • Antepreneurial skills; working with economical knowledge; continuous learning • Good English communication skills |

7. Objectives of the discipline (outcome of the acquired competencies)

| | |
|--|---|
| 7.1 General objective of the discipline | <ul style="list-style-type: none"> • To introduce the basic principles, domains and tasks in Natural Language Processing (NLP) • To understand the current state of the art in order to realize an overview of a specific domain in NLP and to implement a NLP tool. |
| 7.2 Specific objective of the discipline | <ul style="list-style-type: none"> • Apply and use formal models (logics, grammars, parsing), statistic models (HMM), artificial intelligence algorithms and techniques to solve different tasks at the syntactic level (POS-tagging, parsing, chunking), and semantic level (word sense disambiguation, document summarization, anaphora resolution) in Natural Language Processing domain. |

8. Content

| 8.1 Course | Teaching methods | Remarks |
|---|--|---------|
| Course 1. Natural Language Processing (NLP): stages, domains, applications. | Exposure: description, explanation, examples, debate, dialogue | |
| Course 2. -WordNet and RoWordNet - knowledge structure, semantic relations, lexical relations. - WordNetSimilarity tool and similarity measures for words. | Exposure: description, explanation, examples, debate, dialogue | |
| Course 3. Part of speech tagging. - tools for part-of-speech tagging in English. - tools for part-of-speech tagging in Romanian | Exposure: description, explanation, examples, debate, dialogue | |
| Course 4. Syntactic parsing - grammar rules for English - sentence level construction; - Cocke-Kasami-Yonger (CKY) algorithm; | Exposure: description, explanation, examples, debate, dialogue | |

| | | |
|---|--|--|
| Course 5. Statistical parsing - Probabilistic Context-Free Grammars (PCFG); - Probabilistic CKY (Cocke-Kasami-Yonger) parsing of PCFGs. | Exposure: description, explanation, examples, debate, dialogue | |
| Course 6. Hidden Markov Model - Markov chains, Hidden Markov Model(HMM); - three canonical problems associated with HMM - the forward algorithm; the Viterbi algorithm. | Exposure: description, explanation, examples, debate, dialogue | |
| Course 7. Word Sense Disambiguation - dictionary and graph based approaches. | Exposure: description, explanation, examples, debate, dialogue | |
| Course 8. Document summarization - approaches based on clustering, graphs. - tools for summarization. | Exposure: description, explanation, examples, debate, dialogue | |
| Course 9. Textual entailment | Exposure: description, explanation, examples, debate, dialogue | |
| Course 10. Anaphora resolution - Mitkov's algorithm - tools for co-reference resolution | Exposure: description, explanation, examples, debate, dialogue | |
| Course 11. Sentiment analysis | Exposure: description, explanation, examples, debate, dialogue | |
| Course 12. Students' presentations of the practical project. | Exposure: description, explanation, examples, debate, dialogue | |

Bibliography

1. J.ALLEN : Natural language understanding, Benjamin/Cummings Publisher, 2nd ed., 1995.
2. E. CHARNIAK: Statistical language learning, MIT press, 1996.
3. B.CARPENTER: ALE: The attribute logic engine. User's guide. Carnegie Mellon University, 1994.
4. D.FEHRER et al: Description logics for natural language processing. In Proc. of the 1994 Description Logic Workshop (DL'94), 1994.
5. H. HELBIG: Knowledge Representation and the Semantics of Natural Language, Springer, 2006.
6. D.JURAFSKY, J.MARTIN: Speech and language processing, Prentice Hall, 2000.
7. C.MANNING, H.SCHUTZE: Foundation of statistical natural language processing, MIT, 1999.
8. R. MITKOV(ed): The Oxford Handbook of Computational Linguistics, Oxford University Press, 2003.
9. D.TATAR: Inteligenta artificiala: demonstrare automata de teoreme, prelucrarea limbajului natural, Editura Albastra, Microinformatica, 2001.
10. D. TATAR: Inteligenta artificiala. Aplicatii in prelucrarea limbajului natural, Editura Albastra, Microinformatica, 2003, ISBN 973-650-100-01.

| 8.2 Seminar / laboratory | Teaching methods | Remarks |
|--|-------------------------------------|--|
| 1. Working with WordNet, Romanian WordNet and WordNetSimilarity. | Explanation, dialogue, case studies | The seminar/lab is structured as 2 hours classes every second week |
| 2. Students' presentations of a NLP domain and a corresponding tool. | Dialogue, debate | |
| 3. Working with dedicated parsers and taggers | Explanation, | |

| | | |
|---|-------------------------------------|--|
| (Stanford, CST tools, Racai tools) | dialogue, case studies | |
| 4. Students' presentations of a NLP domain and a corresponding tool. | Dialogue, debate | |
| 5. Working with dedicated tools for information extraction and summarization | Explanation, dialogue, case studies | |
| 6. Working with dedicated tools for anaphora and co-reference resolution | Explanation, dialogue, case studies | |
| 7. Students' presentations of the practical projects. | Dialogue, debate | |
| Bibliography 1. Rada Mihalcea: www.cs.unt.edu/~rada/downloads.html 2. Resurse lingvistice in limba romana: www.racai.ro | | |

9. Corroborating the content of the discipline with the expectations of the epistemic community, professional associations and representative employers within the field of the program

- The course respects the IEEE and ACM Curricula Recommendations for Computer Science studies;
- The course exists in the studying program of all major universities in Romania and abroad;
- The optimization of the search on Web, the interfaces in natural language and the recent aspects of text mining need a good understanding of Natural Language Processing.

10. Evaluation

| Type of activity | 10.1 Evaluation criteria | 10.2 Evaluation methods | 10.3 Share in the grade (%) |
|---|---|--|-----------------------------|
| 10.4 Course | - know the theoretical concepts of the domain; - apply the course methods, algorithms in problem solving | Written exam | 40% |
| 10.5 Seminar/lab activities | - know to write an overview of a specific domain | Presentation of a NLP domain and a corresponding tool. | 20% |
| | - be able to implement course algorithms | Practical project - implementation of a NLP tool based on the studied methods. | 40% |
| 10.6 Minimum performance standards | | | |
| ➤ At least grade 5 (from a scale of 1 to 10) at all four evaluation stages. | | | |

Date

Signature of course coordinator

Signature of seminar coordinator

Lecturer Ph.D. Lupea Mihaiela

Lecturer Ph.D. Lupea Mihaiela

Date of approval

Signature of the head of department

.....

Prof. PhD. Andreica Anca