

FIȘA DISCIPLINEI

(Metode de prelucrare a limbajului natural)

Anul universitar 2025-2026

1. Date despre program

| | |
|--|--|
| 1.1. Instituția de învățământ superior | Universitatea Babeș-Bolyai |
| 1.2. Facultatea | Facultatea de Matematică și Informatică |
| 1.3. Departamentul | Departamentul de Matematică și Informatică al Liniei Maghiare |
| 1.4. Domeniul de studii | Informatică |
| 1.5. Ciclul de studii | Masterat |
| 1.6. Programul de studii / Calificarea | Adatelemzés és modellezés / Analiza datelor și modelare Data analysis and modelling |
| 1.7. Forma de învățământ | IF |

2. Date despre disciplină

| | | | | | | | |
|---|---|----------------|---|------------------------|-------------------|--------------------------|--------------------------|
| 2.1. Denumirea disciplinei | Természetesnyelv-feldolgozási módszerek / Metode de prelucrare a limbajului natural / Methods of natural language processing | | | | Codul disciplinei | MMM8081 | |
| 2.2. Titularul activităților de curs | conf. dr. Bodó Zalán-Péter | | | | | | |
| 2.3. Titularul activităților de seminar | conf. dr. Bodó Zalán-Péter | | | | | | |
| 2.4. Anul de studiu | 1 | 2.5. Semestrul | 2 | 2.6. Tipul de evaluare | E | 2.7. Regimul disciplinei | obligatorie-fundamentală |

3. Timpul total estimat (ore pe semestru al activităților didactice)

| | | | | | |
|---|----|---------------------|-----|--------------------------------|-------|
| 3.1. Număr de ore pe săptămână | 5 | din care: 3.2. curs | 2 | 3.3. seminar/laborator/proiect | 1+2+0 |
| 3.4. Total ore din planul de învățământ | 70 | din care: 3.5. curs | 28 | 3.6 seminar/laborator/proiect | 42 |
| Distribuția fondului de timp pentru studiul individual (SI) și activități de autoinstruire (AI) | | | | | ore |
| Studiul după manual, suport de curs, bibliografie și notițe (AI) | | | | | 50 |
| Documentare suplimentară în bibliotecă, pe platformele electronice de specialitate și pe teren | | | | | 30 |
| Pregătire seminare/ laboratoare/ proiecte, teme, referate, portofolii și eseuri | | | | | 40 |
| Tutoriat (consiliere profesională) | | | | | 6 |
| Examinări | | | | | 4 |
| Alte activități | | | | | 0 |
| 3.7. Total ore studiu individual (SI) și activități de autoinstruire (AI) | | | 130 | | |
| 3.8. Total ore pe semestru | | | 200 | | |
| 3.9. Numărul de credite | | | 8 | | |

4. Precondiții (acolo unde este cazul)

| | |
|--------------------|--|
| 4.1. de curriculum | Nu sunt |
| 4.2. de competențe | Competențe de bază de dezvoltare software. |

5. Condiții (acolo unde este cazul)

| | |
|---|---|
| 5.1. de desfășurare a cursului | <ul style="list-style-type: none"> • Sală dotată cu videoproiector • Tablă pentru explicații detaliate, pentru rezolvarea problemelor |
| 5.2. de desfășurare a seminarului/laboratorului | <ul style="list-style-type: none"> • Calculatoare, videoproiector, tablă |

6. Competențele specifice acumulate

| | |
|--|---|
| Competențe profesionale/esențiale | <ul style="list-style-type: none"> • Cunoașterea conceptelor și a algoritmilor de lingvistică computațională. • Utilizarea și proiectarea de corpusuri. • Analiza și dezvoltarea de algoritmi. |
| Competențe transversale | <ul style="list-style-type: none"> • Studiu individual • Metode de lucru, competențe metodologice • Gândire critică și reflecție |

7. Obiectivele disciplinei (reieșind din grila competențelor acumulate)

| | |
|--|--|
| 7.1 Obiectivul general al disciplinei | <ul style="list-style-type: none"> • Scopul general al acestui curs este de a oferi cunoștințe teoretice și practice privind metodele de prelucrare a textelor scrise într-un limbaj natural (de exemplu engleză, română, maghiară). |
| 7.2 Obiectivele specifice | <ul style="list-style-type: none"> • Cunoașterea și aplicarea conceptelor și metodelor de prelucrare a limbajului natural: <ul style="list-style-type: none"> ○ Corpusuri lingvistice ○ Modele bazate pe cuvinte și pe n-grame ○ Analize sintactice ○ Modele Markov ○ Probleme de lingvistică computațională: dezambiguizarea cuvintelor, rezolvarea coreferențelor, extragerea informației, analiza sentimentelor, traducerea automată, etc. |

8. Conținuturi

| | | |
|---|---|------------|
| 8.1 Curs | Metode de predare | Observații |
| 1. Limbaje naturale și formale, istoria și conceptele de bază ale prelucrării limbajului natural. | expunere interactivă, explicație, conversație, demonstrație didactică | |
| 2. Metode de bază de prelucrare a textului. Distanța de editare. | expunere interactivă, explicație, conversație, demonstrație didactică | |

| | | |
|---|---|--|
| 3. Clasificarea/categorizarea documentelor. Algoritmul Naive Bayes | expunere interactivă, explicație, conversație, demonstrație didactică | |
| 4. Metode de evaluare | expunere interactivă, explicație, conversație, demonstrație didactică | |
| 5. Entropia și codurile corectoare de erori în învățarea automată și în prelucrarea limbajului natural. | expunere interactivă, explicație, conversație, demonstrație didactică | |
| 6. Modele lingvistice, modele n-gram. | expunere interactivă, explicație, conversație, demonstrație didactică | |
| 7. Recunoașterea știrilor false | expunere interactivă, explicație, conversație, demonstrație didactică | |
| 8. Analiza sentimentelor, extragerea opiniilor | expunere interactivă, explicație, conversație, demonstrație didactică | |
| 9. Reprezentarea cuvintelor, propozițiilor, documentelor în prelucrarea limbajului natural și în regăsirea informației I. | expunere interactivă, explicație, conversație, demonstrație didactică | |
| 10. Reprezentarea cuvintelor, propozițiilor, documentelor în prelucrarea limbajului natural și în regăsirea informației II. | expunere interactivă, explicație, conversație, demonstrație didactică | |
| 11. Analize sintactice | expunere interactivă, explicație, conversație, demonstrație didactică | |
| 12. Sumarizarea automată a documentelor | expunere interactivă, explicație, conversație, demonstrație didactică | |
| 13. Metode de extragere a informațiilor | expunere interactivă, explicație, conversație, demonstrație didactică | |
| 14. Rezumarea conținutului disciplinei | expunere interactivă, explicație, conversație, demonstrație didactică | |

Bibliografie:

- [1] JURAFSKY M., MARTIN J.H. Speech and Language Processing. 3rd ed. (draft), 2024 (<https://web.stanford.edu/~jurafsky/slp3/>).
- [2] TĂȚAR D. Inteligență artificială. Aplicații în prelucrarea limbajului natural. Editura Albastră, Cluj Napoca, 2003.
- [3] MANNING C.D., SCHÜTZE H. Foundations of statistical language processing. MIT Press, Cambridge, 1999.
- [4] TIKK DOMONKOS. Szövegbányászat. Typotex, Budapest, 2007.
- [5] Andriy Burkov. The Hundred-Page Language Models Book: hands-on with PyTorch. True Positive Inc., 2025.

| | | |
|---|-------------------|------------|
| 8.2 Seminar / laborator: | Metode de predare | Observații |
| În cadrul seminariilor și laboratoarelor citim și discutăm articole împreună cu studenții, discutăm proiectele obligatorii și problemele legate de acestea, studenții prezintă proiectele obligatorii (vezi mai jos), iar apoi, în a doua parte a | | |

| | | |
|---|-------------------|--|
| semestrului, în cadrul acestor cursuri se fac prezentările proiectelor de grup (aplicație NLP + prezentare). Teme: | | |
| Proiecte de grup (pe baza unei teme stabilite de comun acord cu profesorul) | muncă în grup | |
| 1. Analiza sentimentelor: crearea propriului dicționar și aplicarea acestuia la analiza emoțională a textelor | muncă individuală | |
| 2. Wikification: <i>wikificarea</i> automată a unui text | muncă individuală | |
| 3. Vectori de înglobare pentru cuvinte: construirea și antrenarea propriului model neuronal | muncă individuală | |
| Bibliografie: [1]-[5]+ [6] JACKSON P., MOULINIER I. Natural Language Processing for Online Applications. Text Retrieval, Extraction and Categorization. John Benjamins Publishing Company, 2002. [7] ABNEY S. Semisupervised learning for computational linguistics. Chapman & Hall/CRC, 2008. [8] FELDMAN R., SANGER J. The Text Mining Handbook. Advanced Approaches in Analyzing Unstructured Data. Cambridge University Press, 2007. | | |




9. Coroborarea conținuturilor disciplinei cu așteptările reprezentanților comunității epistemice, asociațiilor profesionale și angajatori reprezentativi din domeniul aferent programului

- Acest curs este bazat pe cursul Coursera “Natural Language Processing” de Dan Jurafsky și Christopher Manning: <https://web.stanford.edu/~jurafsky/NLPCourseraSlides.html>
- Există o suprapunere semnificativă între conținutul acestui cursului și cel al cărților „Speech and Language Processing” și „Foundations of statistical language processing” (vezi în bibliografie), considerate lucrările de baze ale NLP.

10. Evaluare

| Tip activitate | 10.1 Criterii de evaluare | 10.2 Metode de evaluare | 10.3 Pondere din nota finală |
|---|--|-------------------------|------------------------------|
| 10.4 Curs | Examen scris la sfârșitul semestrului din materialul predat pe parcursul semestrului | Examen scris | 40% |
| 10.5 Seminar/laborator | Proiecte individuale | Evaluarea proiectelor | 30% |
| | Proiecte de grup (2-3 studenți) pe baza unei teme stabilite de comun acord cu profesorul | Evaluarea proiectelor | 30 |
| 10.6 Standard minim de performanță | | | |
| <ul style="list-style-type: none"> • Minim 50% din puncte la fiecare criteriu. | | | |

11. Etichete ODD (Obiective de Dezvoltare Durabilă / Sustainable Development Goals)

| | | | | | | | | |
|---|---|--|--|--|--|--|--|--|
|  | Eticheta generală pentru Dezvoltare durabilă | | | | | | | |
|  |  | | | | | | | |

Data completării:
2025.04.30

Semnătura titularului de curs
conf. dr. Bodó Zalán-Péter

Semnătura titularului de seminar
conf. dr. Bodó Zalán-Péter

Data avizării în departament:
2025.04.30

Semnătura directorului de departament
Conf. dr. András Szilárd