

Tárgyleírás angol nyelvű képzés tárgya esetén

Tárgy neve: Natural Language Processing and Language-based Foundation Models

Tárgyfelelős neve: Dr. habil Lőrincz András

Tárgyfelelős tudományos fokozata: CSC

Tárgyfelelős MAB szerinti akkreditációs státusza: AT

Az oktatás célja angolul / Aim of the subject:

Knowledge

- They have comprehensive and up-to-date knowledge of general mathematical and computing principles
- Possess the knowledge of specific tools and methods of Natural Language Processing and Language-based Foundation Models

Abilities:

- They are able to apply their mathematical, computer science and informatics skills in a novel way in order to solve tasks in IT research and development.
- They are able to formalize complex IT tasks, to identify and study their theoretical and practical background and then to solve them.

Attitude:

- They follow professional and technological developments in their IT field.
- They are committed to lifelong learning and are open to acquiring new IT competencies.

Autonomy, responsibility:

- They take responsibility for their professional decisions made in their IT-related activities.
- They undertake to meet deadlines and to have deadlines met.
- They bear responsibility for their own work as well as for the work of their colleagues they work together with in a project.
- Regarding mission critical IT systems, they can be entrusted with developing and operational responsibilities that are in accordance with their professional competencies.

Az oktatás tartalma angolul / Major topics:

During the semester, student learn concepts of natural language processing and applications towards human-machine interaction. Topics in the course are as follows:

- Introduction to NLP
- Computational Linguistics: syntax, parsing, semantics, POS tagging, tokenization
- Databases: WordNet, Wikipedia, ConceptNet
- Basic methods: bag of words, word sense disambiguation, latent semantic analysis
- Overview of applications: document retrieval, knowledge extraction and summarization, sentiment analysis, chatbots
- Introduction to Word Vectors
- Word Vectors and Word Senses
- Linguistic Structure: Dependency Parsing

- N-Grams, Recurrent Neural Networks and Language Models
- Machine Translation, Seq2Seq and Attention
- Contextual Word Representations (using Transformers): BERT, GPT
- Efficient Attention Mechanisms
- State-of-the-Art Foundational Large Language Models (LLMs)
- Training and Fine-Tuning Methods of LLMs (Adapters, LoRA, RLHF)
- Prompt Engineering, LLM Tooling
- Conversational Agents
- LLMs in Speech-to-Text Conversion
- Multimodal Foundation Models (Training Methods, Modality Fusion)
- Applied Multimodal Foundational Models in Language Grounding
- Summary of Ethical and Social Issues in NLP and Multimodal Foundation Models
- Paper reading seminars on SOTA Technologies
- Applied mini-project in NLP and Multimodal technologies

A számonkérés és értékelés rendszere angolul / Requirements and evaluation:

Mixed assessment

Irodalom / Literature:

Recommended:

1. Jacob Eisenstein: Introduction to Natural Language Processing, MIT. 2019, ISBN 9780262042840
2. Computational linguistics, Stanford Encyclopedia, 2014
<https://plato.stanford.edu/entries/computational-linguistics/>
3. Chris Manning: Natural Language Processing with Deep Learning, CS224n Stanford Course, 2020. <http://web.stanford.edu/class/cs224n/>
4. Katerina Fragkiadaki: Language Grounding to Vision and Control, CMU course 10-808, 2017.
<https://katefvision.github.io/LanguageGrounding/>
5. D. Jurafsky, J. H. Martin: Speech and Language Processing 3rd ed. draft, (2023).
<https://web.stanford.edu/~jurafsky/slp3/>
6. Cem Akkus, et.al: Multimodal Deep Learning, arXiv preprint arXiv:2301.04856 (2023).
<https://arxiv.org/pdf/2301.04856.pdf>

KITÖLTÉSI ÚTMUTATÓ (végleges tárgyleírásból törölendő):

1. **Tárgy neve:** az excel táblázattal teljes mértékben megegyezően
2. **Tárgyfelelős neve:** az excel táblázattal teljes mértékben megegyezően
3. **Tárgyfelelős tudományos fokozata (ld. MAB):** PhD, DLA, CSc, DSc, akadémikus
4. **Tárgyfelelős MAB szerinti akkreditációs státusza, vagyis a felsőoktatási intézményhez (FOI) tartozás státusza (ld. MAB):** AT/AR/AE/V

az egyes betűk jelentésének magyarázata:

A: akkreditációs célból az adott FOI-nak nyilatkozatot tett oktató, aki az Nftv. 26. §-ának (3) bekezdése szerint kizárólag az adott felsőoktatási intézményt jelölte meg annak, amelyben figyelembe veendő a működési feltételek vizsgálatában;

V: vendégoktató, aki más FOI-nek írt alá, vagy sehol sem tett „kizárólagossági” nyilatkozatot;
T: teljes munkaidőben, határozott vagy határozatlan idejű munkaviszonyban, közalkalmazotti jogviszonyban, ill. ezekkel azonos elbírálás alá eső jogviszonyban;
R: részmunkaidőben, határozott vagy határozatlan idejű munkaviszonyban, közalkalmazotti jogviszonyban, ill. ezekkel azonos elbírálás alá eső jogviszonyban;
E: egyéb módon, pl. megbízási szerződéssel alkalmazott, vagy prof. emeritus

5. **Az oktatás célja angolul: [Az előírt szakmai kompetenciáknak, kompetencia-elemeknek (tudás, képesség, attitűd, autonómia, felelősség) tömör leírása]**
Részben a szak KKK szerinti kompetencia mondatai, részben a tárgyhoz kapcsolódó konkrét kompetenciák.
6. **Az oktatás tartalma angolul: [Az elsajátítandó ismeretanyag tömör leírása]**
7. **A számonkérés és értékelés rendszere angolul: az excel táblázatban szereplő követelménnyel szinkronban**
8. **Idegen nyelven történő indítás esetén az adott idegen nyelvű irodalom: [2-5 kötelező és/vagy ajánlott irodalom]**