

NLP Aplicado

Viviana Erica Cotik (Profesora adjunta, Depto de Computación, FCEN-UBA)

Programa:

El objetivo de la materia es brindar el conocimiento necesario como para realizar y comprender tareas de procesamiento del lenguaje natural (NLP). En esta materia no se planea estudiar a fondo los modelos de lenguaje, ni el NLP clásico. Se introducirán y se darán referencias, pero el objetivo es abordar temas fundamentales para el entrenamiento, uso, usabilidad, construcción de recursos y evaluación de modelos, que no se abordan tradicionalmente en otros cursos.

Temario:

- Introducción al NLP y sus tareas.
- Conceptos básicos de modelos de lenguajes, pre-entrenamiento, fine-tuning y prompting. Transformers y large language models.
- Evaluación de métodos y métricas
- Construcción y búsqueda de recursos lingüísticos: Colección y corrección de datos. Diseño de tareas de anotación, paradigmas de anotación. Evaluación de anotación. Data statements.
- Embeddings contextualizados, almacenamiento de datos. Bases de datos vectoriales. Recuperación de información.
- Sesgos. Ética.
- Interpretabilidad.
- Se podría sumar algo de aplicaciones y recursos en BioNLP (procesamiento del lenguaje natural en biomedicina)

Bibliografía:

- Jurafsky D, Martin JH. An introduction to natural language processing, computational linguistics, and speech recognition. Prentice Hall. 2024.
- “Neural Network Methods in Natural Language Processing (Synthesis Lectures on Human Language Technologies)”. Yoav Goldberg. Morgan & Claypool Publishers. 2017.
- “Introduction to Natural Language Processing”. Jacob Eisenstein. The MIT Press. 2019. “Foundations of Statistical Natural Language Processing”. Chris Manning , Hinrich Schütze. The MIT Press. 1999
- Tom Mitchell, "Machine Learning", McGraw Hill, 1997.
- Hastie, Tibshirani, Friedman, "The Elements of Statistical Learning", Springer, 2001.
- Data statements for natural language processing: Toward mitigating system bias and enabling better science. Emily M. Bender, Batya Friedman. Transactions of the Association for Computational Linguistics, 6:587–60E. 2018
- “Inspecting Algorithms for Bias”. MIT Technology Review. 2017. “How We Analyzed the COMPAS Recidivism Algorithm.” J. Larson, S. Mattu, L. Kirchner, J. Angwin. ProPublica, 2016.
- “Fairness and Machine Learning”. MIT Press. Solon Barocas, Moritz Hardt, Arvind Narayanan. 2023.

- “Discriminating Data”. Wendy Hui Kyong Chun. Charlas.
<https://www.youtube.com/watch?v=vnseu29xZHg>,
<https://www.youtube.com/watch?v=AVMTGBecxfk>
- "Causal Inference for Statistics, Social, and Biomedical Sciences", Imbens, Rubin. Cambridge University Press, 2015.
- “Fairness tutorial: CVPR2022 Fairness Tutorial”. Conference on Computer Vision and Pattern Recognition. Koyejo, Russakovsky.