

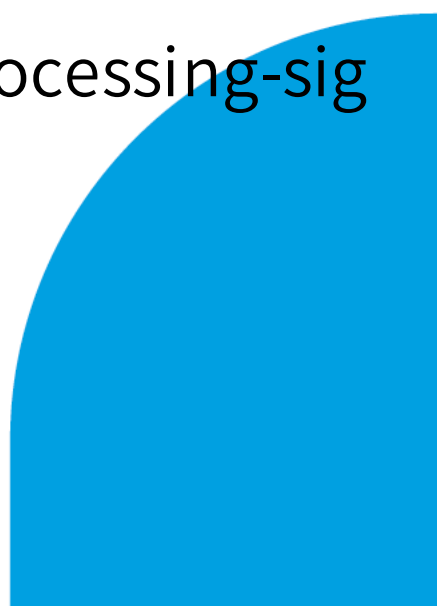
Natural Language Processing (NLP) SIG

Monday 13 December 2021

netherlands
eScience center

What is the NLP SIG?

- Engineers applying or interested in Natural Language Processing
- Meetings every four weeks, discussing NLP developments
- Mailing list: eHumanities@esciencecenter.nl
- Teams channel: Natural Language Processing SIG
- Github page: <https://github.com/NLeSC/natural-language-processing-sig>



Agenda

- NLP project updates
- NLP SIG planning
- Conference summary: Empirical Methods in Natural Language Processing (EMNLP) 2021

News: proposal for eScience Text Mining
Technology Week



eScience projects using NLP (1/3)

- Inside the Filter Bubble (Dafne & Jisk & Erik)
- Understanding visually grounded spoken language via multi-tasking (Patrick & Christiaan)
- StatCheck (Dafne, Erik, Rena)
- SSI: Text Analysis with Cognitive Mapping (Erik, Sven)
- SSI: Recognising symbolism in Turkish television drama (Florian, Berend, Christiaan, Jisk)
- SSI: Morphological Parser for Inflectional Languages Using Deep Learning (Dafne, Jisk)

eScience projects using NLP (2/3)

- SSI: Different kinds of laughter: A machine learning approach (Patrick, Florian, Jisk)
- Navigating Stories in Times of Transition: The COVID-19 Pandemic as a Use Case (Erik, Team Cat)
- The eye of the beholder: Transparent pipelines for assessing online information quality (interested: Cunliang)
- ImpFic: Impact and Fiction (Willem)
- Edo's project on political text analysis



eScience projects using NLP (3/3)

Open Data Infrastructure for Social Science and Economic Innovations (ODISSEI):

- Interacting factors in school choice: modelling consequences for school segregation
- Markets for resilience or “disaster capitalism”? A multi-method critical discourse analysis
- The robot or the brain? Building a classifier for visual news frames of Artificial Intelligence
- Tracking and visualizing states as fossil-fuel owners
- Transparency in the Netherlands’ non-profit sector

Three months of eScience engineer work available per project



NLP SIG Planning

Tech leads have asked for more SIG activities that are relevant for a larger audience. Options, some proposed by attendees:

- Coordinate meetings with other groups, like the RSE meetings
- Invite people from other SIGs, like Machine Learning SIG
- Organize events at conferences, e.g. DHBenelux and CLARIAH
- Invite external speakers, for example from projects
- Broaden focus to social sciences



Conference summary: EMNLP 2021 (1/5)



7-11 November 2021, Punta Cana, Dominican Republic, <https://2021.emnlp.org>

Keynotes

- Ido Dagan: *Where Next? Towards Multi-text Consumption via Three Interesting Research Lines* (among others: interactive NLP)
- Evelina Fedorenko: *Language is separate from thought: Evidence from fMRI*
- Steven Bird: *Language Technology for All?*

Papers (1,806)

- <https://aclanthology.org/events/emnlp-2021/>

Conference summary: EMNLP 2021 (2/5)

Best papers: <https://2021.emnlp.org/blog/>

- a. Liu et al.: Visually Grounded Reasoning across Languages and Cultures
- b. Yang et al.: CHoRaL: Collecting Humor Reaction Labels from Millions of Social Media Users
- c. Bara et al.: MindCraft: Theory of Mind Modeling for Situated Dialogue in Collaborative Tasks
- d. Zhang et al.: SituatedQA: Incorporating Extra-Linguistic Contexts into QA
- e. Lei: When Attention Meets Fast Recurrence: Training Language Models with Reduced Compute
- f. Branco et al.: Shortcutted Commonsense: Data Spuriousness in Deep Learning of Commonsense Reasoning
- g. Lhoest et al.: Datasets: A Community Library for Natural Language Processing



	2018	2019	2020	2021
0	neural	learning	language	language
1	translation	neural	learning	learning
2	learning	language	translation	models
3	machine	text	neural	generation
4	language	translation	models	text
5	using	machine	text	knowledge
6	wmt	generation	task	neural
7	word	using	machine	translation
8	classification	knowledge	generation	model
9	models	classification	knowledge	question
10	task	semantic	using	data
11	text	model	wmt	dialogue
12	networks	extraction	data	via
13	model	models	model	using
14	deep	word	extraction	classification
15	semantic	question	covid	machine
16	embeddings	detection	semantic	multilingual
17	generation	deep	detection	detection
18	sentiment	analysis	bert	answering
19	analysis	entity	graph	analysis

Conference summary: EMNLP 2021 (5/5)

Subject and author statistics:

- <https://2021.emnlp.org/blog/2021-11-05-interactive-graph/>

More information on Twitter and Medium

