M3G training school
Creative natural language generation
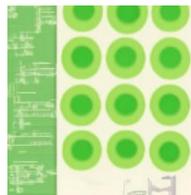
September 19-23, 2022
Ca' Foscari University of Venice

Mihailo Škorić

Title of the project

Perplexity time-series in generated language evaluation

Goal of the project

Explore the value of GPT-computed perplexity time series representations of texts and their effectiveness in evaluation and classification of said texts.

Expected outputs

Language independent (LI) system for transformation of a text into the perplexity time-series representation.

Stacked classifier trained on annotated array of said perplexity time series (e.g., for detection of humorous sentences).

Cosimo Palma

Title of the project

Leveraging Nooj for Natural Language Reasoning: towards automatic knowledge discovery

Goal of the project

Deriving new sentences from a set of phrases by using NooJ syntax and semantic tools to distill it into logical representation

Expected outputs

Automatic solution of a natural-language-written logic grid

Hugh Ormond

Title of the project

Phonetic Aware Language Models for Aesthetic Language Generation Tasks

Goal of the project

Currently, state of the art language models create embeddings from natural, written text, and are able to do an impressive job of capturing grammar structures and semantic elements of languages as they are used, but seem to not be entirely aware of the way spoken languages are pronounced.

This project would investigate ways to encourage models to learn ideas based on phonemes, accent etc., such as by creating a representation that already embeds this information, or selecting data where the task would require picking up on phonetic trends.

Expected outputs

A model like this may show improvements for "performative" language generation tasks, such as generation of songs, poetry or political speeches, when compared to models using embeddings trained mostly on prose where features such as rhyme or meter are not important.

Erinn Van der Sande

Title of the project

A model for predicting a readers emotion

Goal of the project

Look at the feasibility of a model that can predict the expected emotional response of a single/group of reader to a generated sequence of events in the Sjuzhet/Plot of a story. Use it in the ordering of a Fabula and emotional sentiment of a text.

Expected outputs

A model that can predict the emotional response to a certain event in the Sjuzhet/Plot of a story.

Teemu Pöyhönen

Title of the project

Interactive Storytelling via Creative NLG

Goal of the project

Having a system that understands player input & generates creative narration based on the input. For example, tabletop (D&D etc.) game masters could use such a tool for preparation (create/suggest story twists, beats, etc.)

Expected outputs

A pipeline for text-based adventures, or a more specific system, e.g. dialog.

Lin de HUYBRECHT

Title of the project

Tongue Twister Generator

Goal of the project

create a tool that generates personal tongue twister exercises to help people with their speech difficulties

Expected outputs

A minimum viable product of the tongue twister generator that is ready to be tested

Cezary Klamra

Title of the project

Creative text style transfer

Goal of the project

An idea close to my previous work would be to apply creative NLG methods to text style transfer task (for example: rephrasing the text so that it's more humorous/polite/formal or non-offensive). If I was to go for something crazy I would try to examine whether embodied/enactive systems could be utilised in creative NLG. However, I do not really have an idea how that could be done. Maybe in conversational setting?

Expected outputs

Better understanding of creative NLG

Ana-Maria Bucur

Title of the project

Generating Synthetic Data for Suicide Risk Assessment

Goal of the project

Text used for suicide risk assessment is usually gathered from social media platforms or from specific repositories with data donated by relatives of people who commited suicide. Given the sensitivity of this kind of data, it is not straightforward to obtain aceess to such datasets; most of them are private, not shareable with other researchers. The goal of this project is to generate publicly available synthetic data for suicide risk assessment that can be used for reasearch.

Expected outputs

A publicly available synthetic dataset for suicide risk detection.

Nicola Cirillo

Title of the project

Automatic term generation

Goal of the project

The mission aims at developing an algorithm that automatically generates concise and easy to remember terms from their definitions.

Expected outputs

A corpus of terms with their definitions; a tool to automatically generate terms.