



The Markdown

May 3rd, 2023

Roses are red, violets are blue, I wrote this quote just for you!



Recent Research Papers:

MultiLegalSBD: A Multilingual Legal Sentence Boundary Detection Dataset

2023-05-02 05:52:03+00:00

['Tobias Brugger', 'Matthias Stürmer', 'Joel Niklaus']

Sentence Boundary Detection (SBD) is a crucial aspect of Natural Language Processing (NLP), with even a single incorrect split sentence leading to compromised quality of downstream tasks. Legal documents pose a massive challenge for NLP algorithms, given their complex structures and specific terminologies. In a recent paper titled MultiLegalSBD: A Multilingual Legal Sentence Boundary Detection Dataset, researchers endeavored to create a diverse multilingual legal dataset containing over 130,000 annotated sentences in six different languages. The main aim was to assess the degree of difficulty involved in SBD for the legal domain and to train models for the same. Results indicated that the current SBD models failed to perform adequately on multilingual legal datasets. To bridge the gap, monolingual and multilingual models were trained and tested using CRF, BiLSTM-CRF, and transformer models. The researchers demonstrated state-of-the-art performance by achieving highly accurate results. Furthermore, the researchers discovered that multilingual models outperformed all previous benchmarks in the zero-shot setting, as seen in the Portuguese test set. In the interest of enabling further growth and development in this domain, the team made their dataset, models, and code available for the wider community to access and use. This development in the field of Natural Language Processing could lead to the accurate and efficient analysis of legal documents for a variety of applications.

[Read more](#)

Lessons Learned in ATCO2: 5000 hours of Air Traffic Control Communications for Robust Automatic Speech Recognition and Understanding

Top Github Repos:

1- [mlc-llm](#)

2- [FigmaChain](#)

3- [ReplitLM](#)

4- [datacomp](#)

5- [Discord-Chatbot-Gpt4Free](#)

6- [PandaLM](#)

7- [NakedAttention](#)

8- [langchain-cohere-qdrant-doc-retrieval](#)

9- [unlimiformer](#)

10- [znn_chatglm](#)

Top News:

1. How ChatGPT and Other LLMs Work—and Where They Could Go Next

Large language models like AI chatbots seem to be everywhere. If you understand them better, you can use them better.

[Read more](#)

2. NSA Cybersecurity Director Says ‘Buckle Up’ for Generative AI

The security issues raised by ChatGPT and similar tech are just beginning to emerge, but Rob Joyce says it's time to prepare for what comes next.

[Read more](#)

3. Tesla’s Magnet Mystery Shows Elon Musk Is Willing to Compromise

The EV maker said it will expunge rare earths from its vehicles' motors, suggesting Tesla engineers will have to get creative.

[Read more](#)

4. Senator Takes the First Step Towards Federal AI Regulation

Colorado Senator and failed Democratic presidential candidate Michael Bennet introduced a new bill Friday to create an “AI Task Force” of government experts who could identify risks and reduce potential civil liberties and privacy drawbacks from the disruptiv...

[Read more](#)

Placeholder text until AI conclusion writer is implemented.



® Ty Church, Somewhere 2023
[Unsubscribe instantly](#)