[Relation Extraction from Clinical Narratives Using Pre-trained Language Models](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7153059/)

Electronic Health Records contain a large amount of free textual data recorded by healthcare providers during patient care, such as medical records, abstracts of extracts, laboratory reports, and pathology reports. Compared to structured data, textual data usually records detailed information about clinical events and the exchange of information between teams in a hospital environment. Much imperative persistent data regularly exists in unstructured arrange as it were, and manual extraction of such data is costly and time devouring. Relation extraction is one of the main tasks of extracting information, the purpose of which is to identify the semantic connection between the concepts mentioned in the document. But there was a limitation that severely restrict the portability and generalizability of RE to other novel resources. However, despites BERT’s success in the open domain and biomedical literature, researcher aimed to investigate BERT to clinical RE tasks by comparing two different implementations of the BERT model for clinical RE, evaluating BERT models across different clinical RE tasks and assessing four BERT models trained from the open domain, biomedical literature, and clinical text on the clinical RE task**.** In BERT-based relational extraction all documents went through a pre-processing procedure that includes basic steps such as sentence boundary detection and tokenization, which were done using the CLAMP Toolkit. Fine-tuned BERT uses the semantic type of an entity to replace the entity itself. FC-BERT Model utilized the BERT model to generate vectors for all words in sequence; in parallel to the BERT model, the BIO tag sequence of the sentence was represented in an embedding layer Consequently, researcher have studied deep learning methods that use pre-trained language models to extract relationships from clinical stories. This fills the gap in the application of new deep learning methods with pre-trained language models for clinical RE tasks. The results show that the FT-BERT pre-trained language model received F1 scores of 0.9409 and 0.7679 for n2c2 and i2b2 datasets, respectively, which surpassed existing technology in both tasks, which proved the advantages of using BERT mode

[Lithium NLP: A System for Rich Information Extraction from Noisy User Generated Text on Social Media](https://arxiv.org/pdf/1707.04244.pdf)

Nowadays social networks have become one of the main means of communication and content production. As a result, industrial systems with the ability to process rich user content from social networking platforms have many practical applications. NLP on social media can be very complicated and complex for various reasons such as Noisy unnormalized data, Multi-lingual content, Large scale datasets, Rich set of information. In this article author introduced the lithium NLP system that solves these problems. It is a resource-limited, high-performance and language-independent system for extracting information from noisy user text (for example, text available on social networks). It can extract a rich set of information, including entities, themes, topic tags, and emotions. Lithium NLP currently supports several languages, including several languages. Lithium NLP extracts a wide range of information from the text, including entities, themes, topic tags, and emotions. Author discussed several practical applications of the systems currently integrated in lithium products and also compared system to existing commercial and academic NLP systems in terms of performance, information retrieved, and supported languages. They have proven that in some cases lithium NLP can be compared with the latest commercial NLP systems. In Offline Resource Generation phase (Mention-Entity Co-occurrence, Entity-Entity Co-occurrence, Entity Importance, Topic Parents, Topic Hashtags) generated several dictionaries that capture language models, probabilities and relations across entities and topics, by leveraging various multi-lingual data sources. The Text Processing phase (Language Detection, Text Normalization, Sentence Breaking, Tokenization, Entity Extraction, EDL, Topic Projection, Hashtag Recommendation, Sentiment Analysis) of the system processes the input text document through several stages and the information (entities, topics etc.) extracted at every stage is added as a semantic annotation to the text. The Lithium NLP system provides a REST API via which client applications can send a text document as request and receive the

annotated text as JSON response.

[A Detailed Analysis of Core NLP for Information extraction](https://poseidon01.ssrn.com/delivery.php?ID=641117074119116065092029083093091024016073027027075062101005081022078126102115125011096100063045053098009088085098104019101028059021009023036093087108016095092089125060082063067114002101020000119008082016124069081114093083125006084104108105092070030108&EXT=pdf)

Bootstrapping is a statistical method used to extract large amounts of information from annotated seeds and objects (text data). For guidance, we need a lot of training data, which usually extracts the relevant data characteristics in annotated form. The guide program builds effective entities from the body of the subject area, applying various training rules. Several algorithms, such as the Basilisk algorithm, were developed to automatically or semi-automatically obtain a semantic dictionary, including extracting information, analyzing anaphora, answering questions, and embedding suggested phrases. Although there are some semantic dictionaries, such as word networks and these resources do not contain a specialized dictionary for specific fields. To perform the function of extracting information, a Natural Language Tag (NLTK) is required, which consists of three parts: sentence tokenizer, word tokenizer and part of speech tagging. Detecting or retrieving an object is done using this token generator, but it can also be done using managed technology. The bootstrap algorithm uses two ideas: extracting data from templates used to output various semantic categories, and exploring several bootstrap algorithms to extract semantic categories. In this article, the result generated by the guidance algorithm is a token, and the Basilisk algorithm is used to improve the relationship between tokens. The Basilisk algorithm provides a concise and clear summary of gold with high reliability and accuracy. By improving the coil algorithm and combining it with other bootstrap algorithms, several semantic categories can be extracted in advance to improve the results of the marking mode, which will be used in the processing of biomedical natural language, maintain clinical lists, provide voice assistance and other applications. The field helps children and machine translation to cope, we translate the semantic features of one language into another language

*[All links are hyperlinked in research paper name]*