Refined Name Enitiy Recognition (NER) by A Customized SpaCy Model and Pattern Rules of RegEx

Moha Yeghaneh

January 12, 2020

The aim of this presentation is to show the initial result and direction of our research. Here is the main area that we have worked:

- **Data Preparation**

The aim of this presentation is to show the initial result and direction of our research. Here is the main area that we have worked:

- **Data Preparation**
- **Data Annotation**

The aim of this presentation is to show the initial result and direction of our research. Here is the main area that we have worked:

- **Data Preparation**
- **Data Annotation**
- **Data Investigation**

The aim of this presentation is to show the initial result and direction of our research. Here is the main area that we have worked:

- **Data Preparation**
- **Data Annotation**
- **Data Investigation**
- **Data Visualization**

The aim of this presentation is to show the initial result and direction of our research. Here is the main area that we have worked:

- **Data Preparation**
- **Data Annotation**
- **Data Investigation**
- **Data Visualization**
- **Refined Named Entity Recognition Model.**

The aim of this presentation is to show the initial result and direction of our research. Here is the main area that we have worked:

- **Data Preparation**
- **Data Annotation**
- **Data Investigation**
- **Data Visualization**
- **Refined Named Entity Recognition Model.**
- **Classification of Observational Sentences using NER**

The aim of this presentation is to show the initial result and direction of our research. Here is the main area that we have worked:

- **Data Preparation**
- **Data Annotation**
- **Data Investigation**
- **Data Visualization**
- **Refined Named Entity Recognition Model.**
- **Classification of Observational Sentences using NER**
- **Data Publishing.**

The aim of this presentation is to show the initial result and direction of our research. Here is the main area that we have worked:

- **Data Preparation**
- **Data Annotation**
- **Data Investigation**
- **Data Visualization**
- **Refined Named Entity Recognition Model.**
- **Classification of Observational Sentences using NER**
- **Data Publishing.**
- **Toward Relation Extraction!**

Difficulty and Challenges

- Noisy and inconsistent text data.
- Time consuming and tedious manual modification of annotation.
- Unavailability of training data and research paper in the domain.

Purpose of this Presentation

In this report we want to communicate what we have already done including:

- **Prepossessing** and preparation of text data for classification

Purpose of this Presentation

In this report we want to communicate what we have already done including:

- **Prepossessing** and preparation of text data for classification
- Information extraction in an informative and **interactive** way

Purpose of this Presentation

In this report we want to communicate what we have already done including:

- **Prepossessing** and preparation of text data for classification
- Information extraction in an informative and **interactive** way
- Introducing a refined named entity recognition (NER) model using deep learning

Purpose of this Presentation

In this report we want to communicate what we have already done including:

- **Prepossessing** and preparation of text data for classification
- Information extraction in an informative and **interactive** way
- Introducing a refined named entity recognition (NER) model using deep learning
- evaluation of model using gold standard data

Purpose of this Presentation

In this report we want to communicate what we have already done including:

- **Prepossessing** and preparation of text data for classification
- Information extraction in an informative and **interactive** way
- Introducing a refined named entity recognition (NER) model using deep learning
- evaluation of model using gold standard data
- Using machine learning and deep learning methods for text classification

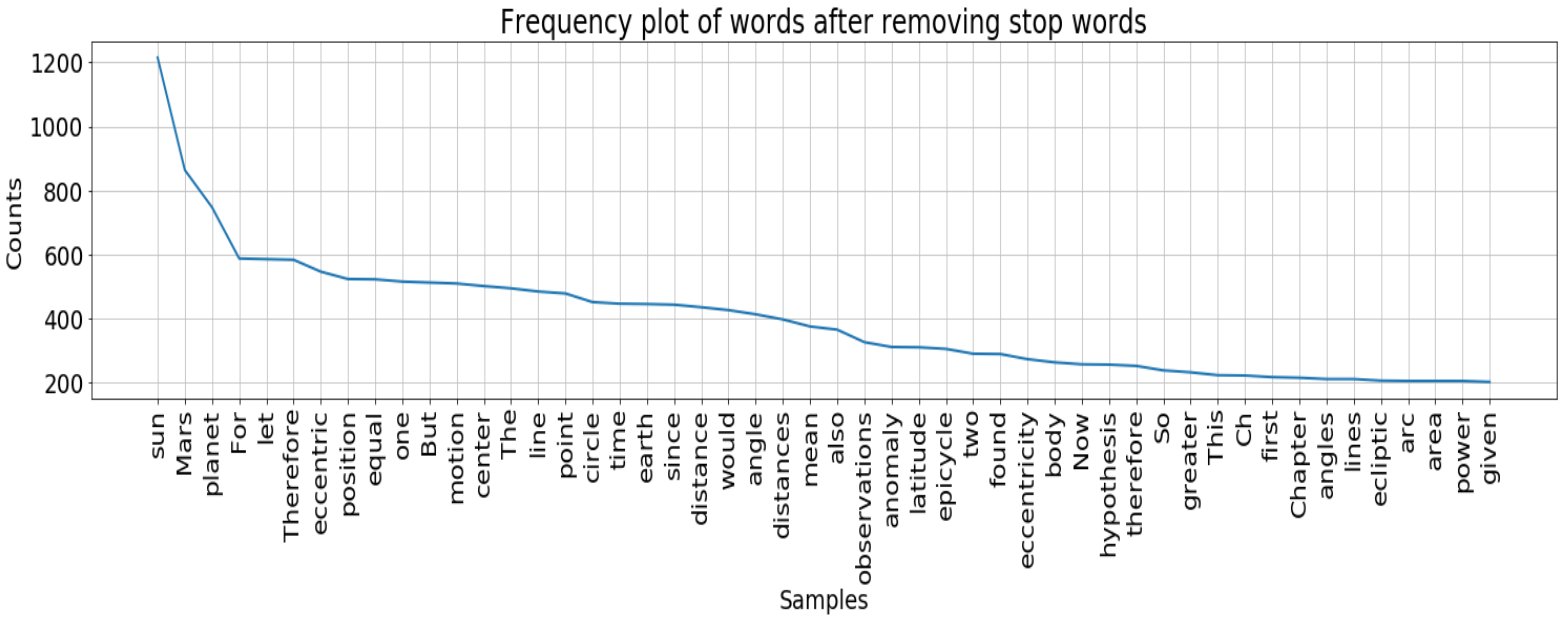
Purpose of this Presentation

In this report we want to communicate what we have already done including:

- **Prepossessing** and preparation of text data for classification
- Information extraction in an informative and **interactive** way
- Introducing a refined named entity recognition (NER) model using deep learning
- evaluation of model using gold standard data
- Using machine learning and deep learning methods for text classification
- Proposing some ideas for the future work toward relation extraction and causal inference

Feeling the Data through Some Statistics!

- Corpus has 70 chapters including 1605 paragraph, 6699 sentences
- Corpus includes 169231 tokens (roughly speaking; words) and 9513 unique tokens
- lexical_diversity which shows lexical richness is 1.2



Feeling the integrated data by NER and more attribute as dataframe

here you can see first 10 sentences with different attribute that has been add by our and can be used later for classification,relation extraction ...

`../../../Documents/Former_projec`

- You can find here the word cloud of the whole book and by chapter
- It can give us some simple initial and simple intution which can be used for the further text analysis

NER workflow

- Data is annotated entity by entity using regex pattern.
- The result of each step is saved as jsonl fomat
- After troubleshooting (false tokenization,double punctuation...)
- Annotated data is merged and now the training data is ready!

.

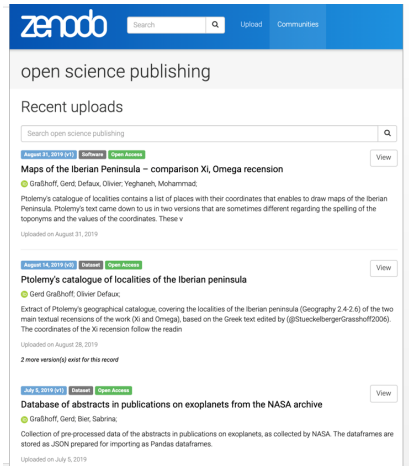
Training

- A Bert model with100 iteration and batch size 16 has been used for NER classifcation
- The evaulation result per entity and overall is calculated by comparing with gold stadnard format as follows:

Metrics	ents_p'	ents_r	ents_f	
GEOM	100	99.85	99.92	
LONG	99.76	99.88	99.82	
PARA	98.51	99.76	99.13	
TIME	97.97	97.00	97.48	
STAR	84.61	74.15	79.04	

CitableClass

- We have a developed a usefull framework citableclass that we can used in order to publich and use the data.
- Any user can have access to data with notebooks using DOI number



Thank you for your time and feedbacks :)

Many thanks

Ma, Y.; Zhou, G.; Wang, S.; Zhao, H.; Jung, W. SignFi: Sign Language Recognition Using WiFi. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 2018, 2, 23.

Muller, Machine Learning and AI for the sciences – Towards Understanding