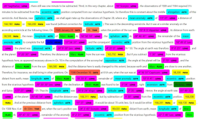
Refined Name Enitiy Recognition (NER) by A Customized SpaCy Model and Pattern Rules of RegEx

Moha Yeghaneh

January 13, 2020



28° Sagittarius LONG , there still was one minute to be subtracted. Third, in this very chapter, about 16° Scorpio LONG , the observations of 1589 and 1584 required 3½ minutes to be subtracted from the eccentric ASTR position computed from our vicarious hypothesis. So therefore this is constant about the middle elongation ASTR of this semicircle. And likewise, near aphelion ASTR , we shall again take up the observations of Chapter 28, where at a mean anomaly ASTR of 11° 37' LONG a distance of 166,180 PARA or 166,200 PARA was found (without correction for latitude ASTR). This was in the descending semicircle. But it was at a similar anomaly on the ascending semicircle at the following times. On 1585 January 24 DATE at 9h TIME , when the position of the sun was 15° 9' 5" Aquarius LONG , its distance from earth 98,590 PARA , the mean longitude ASTR of Mars PLAN 4s 16° 50' 10" LONG , the aphelion ASTR 4s 28° 49' 41" LONG , the remainder of the mean anomaly ASTR to complete the full circle GEOM 11° 56' 31" LONG , and the consequent eccentric ASTR position from the vicarious hypothesis 18° 49' 0" Leo LONG ; the planet was observed ASTR at 24° 9' 30" Leo LONG , latitude ASTR 4° 31' 0" LONG N *30. The angle at earth was therefore 9° 0' 25" LONG , and at the planet 5° 20' 30" LONG . Therefore, the distance of Mars PLAN from the sun was 165,792 PARA . But if you subtract 1' 30" LONG from the vicarious hypothesis here, as appeared necessary above in Ch. 18 in the computation of the acronychal opposition ASTR , the angle at the planet will be 5° 19' LONG , and the distance of Mars PLAN from the sun 166,580 PARA . And the distance here is easily changed to this extent, because earth and Mars PLAN are close to one another. Therefore, for insurance, we shall bring in other positions. On 1586 December 16 DATE at 6½h am, when the sun was at 4° 16' 51" Capricorn LONG , 98,200 PARA distant from earth, mean longitude ASTR of Mars PLAN 4s 18° 39' 9" LONG , remainder of mean anomaly ASTR 10° 9' 41" LONG , eccentric ASTR position from the vicarious hypothesis 20° 20' 30" Leo LONG ; the declination of Mars PLAN was found to be 3° 54' LONG , right ascension ASTR from Arcturus STAR and Spica 177° 27' LONG *31. Thus its longitude ASTR was 20° 6' 24" Virgo LONG , latitude ASTR 2° 35' LONG . Hence, the angle at earth was 85° 49' 33" LONG , at the planet 35° 45' 58" LONG . And the distance was 166,311 PARA , but by subtraction of 1' 30" LONG from the eccentric ASTR position, 166,208 PARA . And at the previous distance from aphelion ASTR , 11° 37' LONG , it would be about 70 units less. So it would be either 166,241 PARA or 166,138 PARA . On 1588 Nov. 6 at 8h 50m TIME am, when the sun's position was 24° 3' 43" Scorpio LONG , 98,630 PARA distant from earth, mean longitude ASTR of Mars PLAN 4s 20° 47' 35" LONG , remainder of the anomaly 8° 2' 51" LONG , eccentric ASTR position from the vicarious hypothesis 22° 7' 48" Leo LONG ; Mars

1 Introduction

2 Data Preparation

3 Data Annotation

4 Refined Named Entity Recognition Model

5 Explicit Observation Extraction

6 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**

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.**

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.

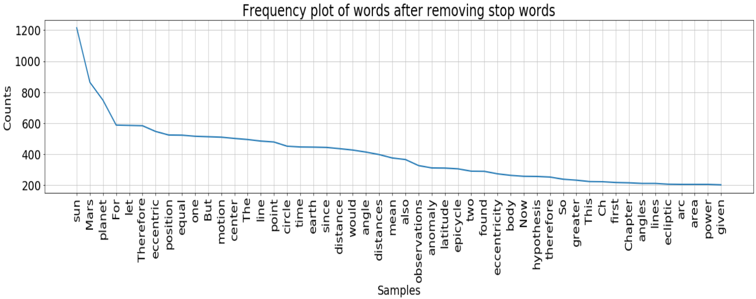
Purpose of this presentation

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 ...

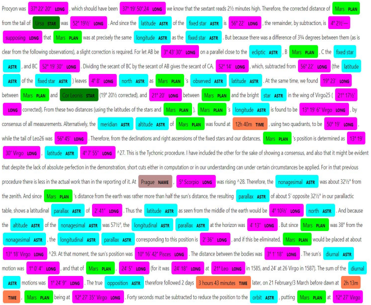
../../../Documents/Former_projec

- 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!

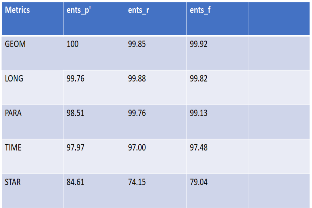
.



Procyon was 37° 22′ 20″ LONG , which should have been 37° 19′ 50″,24 LONG we know that the sextant reads 2½ minutes high. Therefore, the corrected distance of Mars PLAN from the tail of Ursa STAR was 52° 19½′ LONG . And since the latitude ASTR of the fixed star ASTR is 56° 22′ LONG , the remainder, by subtraction, is 4° 2½′— supposing LONG that Mars PLAN was at precisely the same longitude ASTR as the fixed star ASTR . But because there was a difference of 3¾ degrees between them (as is clear from the following observations), a slight correction is required. For let AB be 3° 43′ 30″ LONG on a parallel close to the ecliptic ASTR , B Mars PLAN , C the fixed star ASTR , and BC 52° 19′ 30″ LONG . Dividing the secant of BC by the secant of AB gives the secant of CA, 52° 14′ LONG , which, subtracted from 56° 22′ LONG (the latitude ASTR of the fixed star ASTR) leaves 4° 8′ LONG north ASTR as Mars PLAN 's observed ASTR latitude ASTR . At the same time, we found 19° 23′ LONG between Mars PLAN and Cor Leonis STAR (19° 20½ corrected), and 21° 20′ LONG between Mars PLAN and the bright star ASTR in the wing of Virgo25 (21° 17½′ LONG corrected). From these two distances (using the latitudes of the stars and Mars PLAN), Mars PLAN 's longitude ASTR is found to be 13° 19′ 6″ Virgo LONG , by consensus of all measurements. Alternatively, the meridian ASTR altitude ASTR of Mars PLAN was found at 12h 40m TIME , using two quadrants, to be 50° 19′ LONG , while the tail of Leo26 was 56° 45′ LONG . Therefore, from the declinations and right ascensions of the fixed stars and our distances, Mars PLAN 's position is determined as 13° 19′ 30″ Virgo LONG , latitude ASTR 4° 7′ 55″ LONG ^27. This is the Tychonic procedure. I have included the other for the sake of showing a consensus, and also that it might be evident that despite the lack of absolute perfection in the demonstration, short cuts either in computation or in our understanding can under certain circumstances be applied. For in that previous procedure there is less in the actual work than in the reporting of it. At Prague NAME , 5° Scorpio LONG was rising ^28. Therefore, the nonagesimal ASTR was about 32½° from the zenith. And since Mars PLAN 's distance from the earth was rather more than half the sun's distance, the resulting parallax ASTR of about 5′ opposite 32½° in our parallactic table, shows a latitudinal parallax ASTR of 2′ 41″ LONG . Thus the latitude ASTR as seen from the middle of the earth would be 4° 10⅔′ LONG north ASTR . And because the altitude ASTR of the nonagesimal ASTR was 57½°, the longitudinal ASTR parallax ASTR at the horizon was 4′ 13″ LONG . But since Mars PLAN was 38° from the nonagesimal ASTR , the longitudinal ASTR parallax ASTR corresponding to this position is 2′ 36″ LONG , and if this be eliminated, Mars PLAN would be placed at about 13° 18′ Virgo LONG ^29. At that moment, the sun's position was 10° 16′ 42″ Pisces LONG . The distance between the bodies was 3° 1′ 18″ LONG . The sun's diurnal ASTR motion was 1° 0′ 4″ LONG , and that of Mars PLAN , 24′ 5″ LONG (for it was 24′ 18″ LONG at 21° Leo LONG in 1585, and 24′ at 26 Virgo in 1587). The sum of the diurnal ASTR motions was 1° 24′ 9″ LONG . The true opposition ASTR therefore followed 2 days 3 hours 43 minutes TIME later, on 21 February/3 March before dawn at 2h 13m TIME , Mars PLAN being at 12° 27′ 35″ Virgo LONG . Forty seconds must be subtracted to reduce the position to the orbit ASTR , putting Mars PLAN at 12° 27′ Virgo

Training

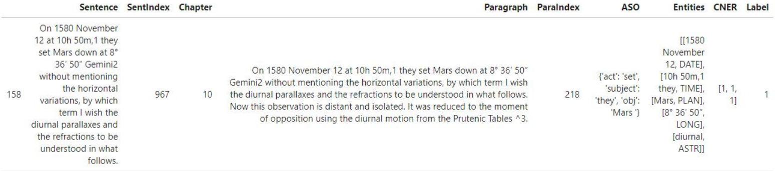
- A model word representation Bert with100 iteration and batch size 16 has been used for NER classifcation
- The evaluation result per entity and overall is calculated by comparing with gold stadnard format as follows:



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	

An example form the book

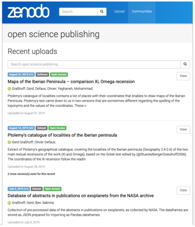
- you can see here an example of explicit observation extraction.
- moreover, you can find how the text has been structured.



	Sentence	SentIndex	Chapter	Paragraph	ParaIndex	ASO	Entities	CNER	Label
158	On 1580 November 12 at 10h 50m,1 they set Mars down at 8° 36′ 50″ Gemini2 without mentioning the horizontal variations, by which term I wish the diurnal parallaxes and the refractions to be understood in what follows.	967	10	On 1580 November 12 at 10h 50m,1 they set Mars down at 8° 36′ 50″ Gemini2 without mentioning the horizontal variations, by which term I wish the diurnal parallaxes and the refractions to be understood in what follows. Now this observation is distant and isolated. It was reduced to the moment of opposition using the diurnal motion from the Prutenic Tables ^3.	218	{'act': 'set', 'subject': 'they', 'obj': 'Mars '}	[[1580 November 12, DATE], [10h 50m,1 they, TIME], [Mars, PLAN], [8° 36′ 50″, LONG], [diurnal, ASTR]]	[1, 1, 1]	1

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



open science publishing

Recent uploads

Search open science publishing

August 31, 2019 (v1) Software Open Access

Maps of the Iberian Peninsula – comparison Xi, Omega recension

Graßhoff, Gerd; Defaux, Olivier; Yeghaneh, Mohammad;

Ptolemy's catalogue of localities contains a list of places with their coordinates that enables to draw maps of the Iberian Peninsula. Ptolemy's text came down to us in two versions that are sometimes different regarding the spelling of the toponyms and the values of the coordinates. These v

Uploaded on August 31, 2019

August 14, 2019 (v3) Dataset Open Access

Ptolemy's catalogue of localities of the Iberian peninsula

Gerd Graßhoff; Olivier Defaux;

Extract of Ptolemy's geographical catalogue, covering the localities of the Iberian peninsula (Geography 2.4-2.6) of the two main textual recensions of the work (Xi and Omega), based on the Greek text edited by (@StueckelbergerGrasshoff2006). The coordinates of the Xi recension follow the readin

Uploaded on August 28, 2019

2 more version(s) exist for this record

July 5, 2019 (v1) Dataset Open Access

Database of abstracts in publications on exoplanets from the NASA archive

Graßhoff, Gerd; Bier, Sabrina;

Collection of pre-processed data of the abstracts in publications on exoplanets, as collected by NASA. The dataframes are stored as JSON prepared for importing as Pandas dataframes.

Uploaded on July 5, 2019

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