

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

January 14, 2020

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

Introduction

Data Preparation

Data Annotation

Refined Named Entity Recognition Model

Explicit Observation Extraction

Data Publishing

1 Introduction

2 Data Preparation

3 Data Annotation

4 Refined Named Entity Recognition Model

5 Explicit Observation Extraction

6 Data Publishing

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

Introduction

Data
Preparation

Data
Annotation

Refined
Named Entity
Recognition
Model

Explicit
Observation
Extraction

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

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

Introduction

Data
Preparation

Data
Annotation

Refined
Named Entity
Recognition
Model

Explicit
Observation
Extraction

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**

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

Introduction

Data Preparation

Data Annotation

Refined Named Entity Recognition Model

Explicit Observation Extraction

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**

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

Introduction

Data
Preparation

Data
Annotation

Refined
Named Entity
Recognition
Model

Explicit
Observation
Extraction

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

Introduction

Data Preparation

Data Annotation

Refined Named Entity Recognition Model

Explicit Observation Extraction

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.**

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

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

- Noisy and inconsistent text data.

Introduction

Data Preparation

Data Annotation

Refined Named Entity Recognition Model

Explicit Observation Extraction

Data Publishing

Difficulty and Challenges

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

Introduction

Data Preparation

Data Annotation

Refined Named Entity Recognition Model

Explicit Observation Extraction

Data Publishing

- Noisy and inconsistent text data.
- Time consuming and tedious manual modification of annotation.

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

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

Introduction

Data
Preparation

Data
Annotation

Refined
Named Entity
Recognition
Model

Explicit
Observation
Extraction

Data
Publishing

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

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

Introduction

Data Preparation

Data Annotation

Refined Named Entity Recognition Model

Explicit Observation Extraction

Data Publishing

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

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

Introduction

Data Preparation

Data Annotation

Refined Named Entity Recognition Model

Explicit Observation Extraction

Data Publishing

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

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

Introduction

Data Preparation

Data Annotation

Refined Named Entity Recognition Model

Explicit Observation Extraction

Data Publishing

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

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

Introduction

Data Preparation

Data Annotation

Refined Named Entity Recognition Model

Explicit Observation Extraction

Data Publishing

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

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

Introduction

Data Preparation

Data Annotation

Refined Named Entity Recognition Model

Explicit Observation Extraction

Data Publishing

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.

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

Introduction

Data Preparation

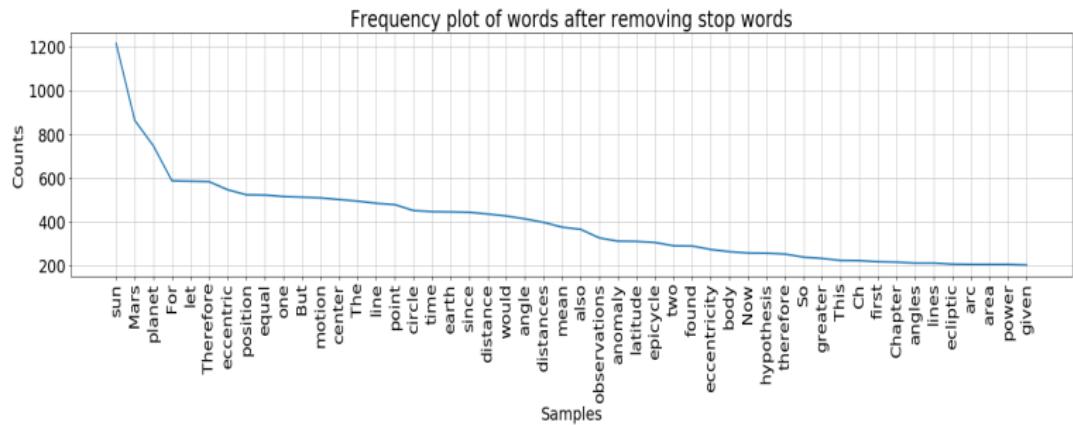
Data Annotation

Refined Named Entity Recognition Model

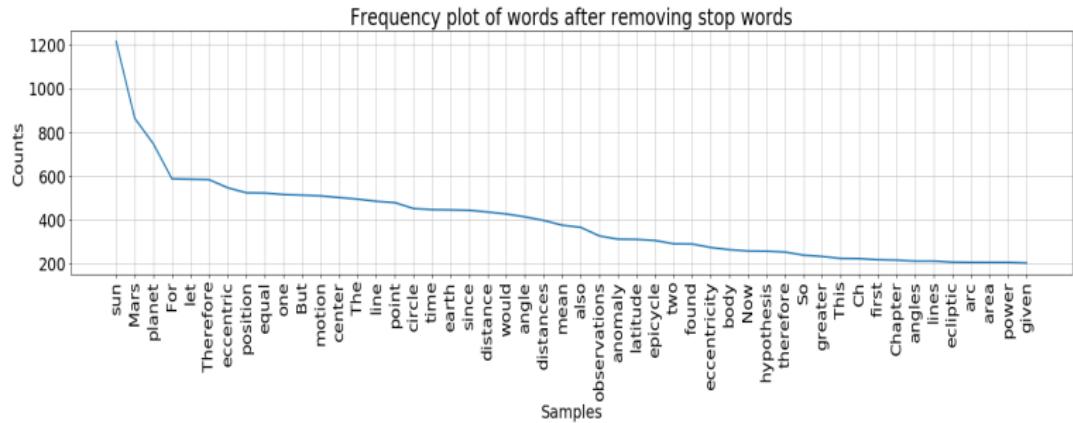
Explicit Observation Extraction

Data Publishing

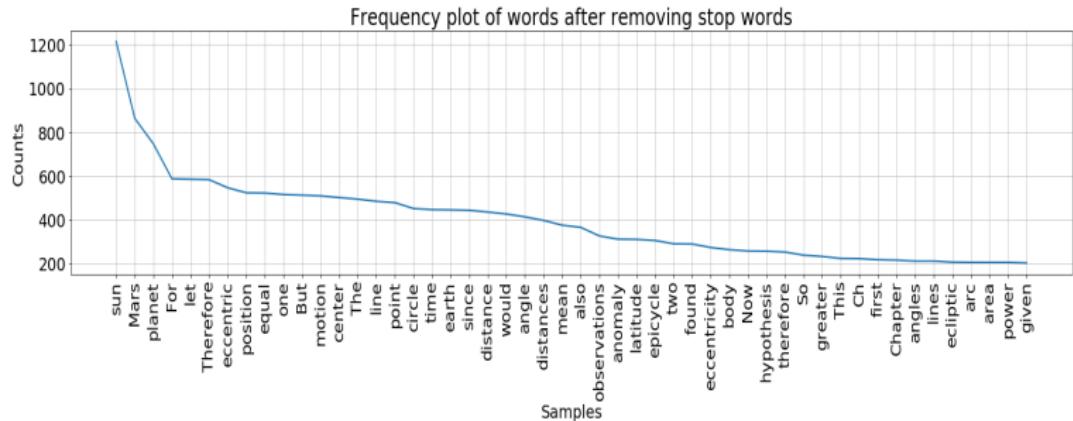
- Corpus has 70 chapters including 1605 paragraphs, 6699 sentences



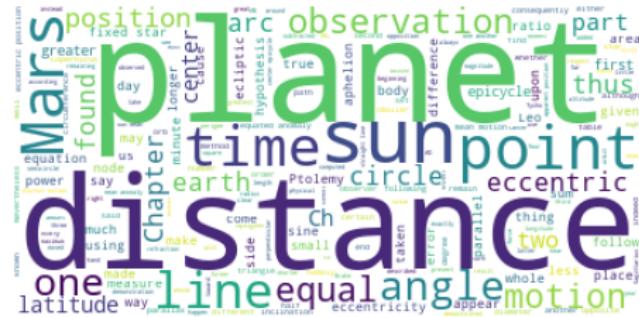
- Corpus has 70 chapters including 1605 paragraphs, 6699 sentences
- Corpus includes 169231 tokens (roughly speaking; words) and 9513 unique tokens



- 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



- You can find here the word cloud of the whole book and by chapters It can give us some simple initial and simple intuition



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

Introduction

Data Preparation

Data Annotation

Refined Named Entity Recognition Model

Explicit Observation Extraction

Data Publishing

	Sentence	SentIndex	Chapter		Paragraph	ParaIndex	ASO	Entities	CNER	L
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]	
168	On 1582 December 28 at 11h 30m, they set Mars down at 16° 47' Cancer by observation ^6.	977	10		On 1582 December 28 at 11h 30m, they set Mars down at 16° 47' Cancer by observation ^6. The moment of opposition assigned by Tycho comes 46 minutes later, during which the planet retrogressed less than one minute. Tycho therefore puts it at 16° 46' 16" Cancer ^7. On an inserted sheet here, an attempt was made to correct for a refraction of two minutes. This was, I think, first trial of the theory of refraction then being developed. Nevertheless, followed the observed value unchanged, thus declining to consider the planet as something which could alter its position. Nor was there any need for correction, since it was in Cancer, beyond the reach of refraction, and was in mid-sky where, in Cancer, there is no longitudinal parallax.	221	{'act': 'set', 'subject': 'they', 'obj': 'Mars'}	[[1582 December 28, DATE], [11h 30m, TIME], [Mars, PLAN], [16° 47' Cancer, LONG], [observation, ASTR]]	[1, 1]	
175	On 1585 January 31 at 12h 0m, Mars was placed at 21° 18' 11" Leo ^8.	984	10		On 1585 January 31 at 12h 0m, Mars was placed at 21° 18' 11" Leo ^8. The diurnal motion, by comparison of observations, was 24° 15'. The moment of opposition followed at 19h 35m, 7 hours and 35 minutes later. To this period belongs 7° 41' of diurnal motion westward. Therefore, at the designated moment, it would have been at 21° 10' 30" Leo, which is what was accepted. There is no mention of parallel. Nothing had to be done about refraction, because Mars was high and at mid-sky. I therefore find the bit of advice in the table about refraction (properly) ignored.	222	{'act': 'placed'}	[[1585 January 31, DATE], [12h 0m, TIME], [Mars, PLAN], [21° 18' 11" Leo, LONG]]	[1, 1]	

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

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

Introduction

Data Preparation

Data Annotation

Refined Named Entity Recognition Model

Explicit Observation Extraction

Data Publishing

■ Data is annotated entity by entity using regex pattern.

Dragon was **17° 25' 30" south**, which should have been **17° 25' 30" south**. We know that the sunset ends 270 minutes high. Therefore, the corrected distance of **1800 miles** from the tail of **Regulus** was **1800 miles**. And since the **Latitude ATTR** of the **Regulus ATTR** is **17° 25' 30"**, the remainder by subtraction is **15° 30' 30"**. However, **Regulus** was at precisely the same **Longitude ATTR** as the **Regulus ATTR**. But because there was a difference of 5% degrees between them (as is clear from the following observations), a digit correction is required. For let AB be **17° 25' 30"** on a parallel close to the **Ursa Major ATTR** & **Regulus ATTR**. C the **Regulus ATTR** and D the **Ursa Major ATTR**. Dividing the arc of CD by the arc of AB gives the secant of C, **sec(AB)**, which, substituted from **17° 25' 30"** the **Latitude ATTR** of the **Regulus ATTR** leaves **17° 25' 30" south**. Now, if **17° 25' 30"** is subtracted from **1800 miles**, we find **17° 25' 30"**. At the same time, we find **17° 25' 30"** between **Regulus** and **Regulus** **18° 20' 20" corrected**, and **Regulus** between **Regulus** and the bright star **Alcyone** in the wing of **Vega**. **17° 25' 30" south** is connected. From these two distances (using the latitudes of the stars and **Regulus** **18° 20' 20" corrected**) **Regulus** is found to be **17° 25' 30" south** **1800 miles** by consensus of all measurements. Alternatively, the **Latitude ATTR** **Latitude ATTR** of **Regulus** was found to be **17° 25' 30" south** **1800 miles**, using two quadrants, to be **17° 25' 30" south**, while the tail of **Ursa** was **17° 25' 30"**. Therefore, from the latitudes and right ascensions of the fixed stars and our distances, **Regulus**' position is determined as **17° 25' 30" south**.

27. This is the typical procedure I have included the order 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 the previous procedure there is less to do than in the reporting of E. A. [Tragedy NAME] **17° 25' 30" south**, we bring **1800**. Therefore, the **Longitude ATTR** was about **102°** from the zenith. And since **Regulus**'s distance from the earth was rather more than half the sun's distance, the resulting **Latitude ATTR** of about 32° is opposite 32°^{1/2} in our astrolabe table, shows a latitudinal **Latitude ATTR** of **17° 25' 30"**. Thus the **Latitude ATTR** as seen from the middle of the earth would be **17° 25' 30" south**. And because the **Latitude ATTR** of **Regulus** was **17° 25' 30"**, the **Longitude ATTR** **Longitude ATTR** at the horizon was **17° 25' 30"**. But since **Regulus** was **38°** from the **Longitude ATTR**, the **Longitude ATTR** **Longitude ATTR** corresponding to this position is **17° 25' 30"**, and if this is eliminated, **Regulus** could be placed at about **17° 25' 30" south**. 28. At that moment, the sun's position was **17° 25' 30" south**. The distance between the bodies was **17° 25' 30"**. The sun's **Ursa Major ATTR** moves west **17° 25' 30" south**, and that is **17° 25' 30" south**. (For it was **17° 25' 30" south** in 1858, and 24 of **Vega** in 1867). The sum of the **Ursa Major ATTR** motions was **35° 50' 30" south**. The true **Ursa Major ATTR** therefore followed 2 days 43 minutes **TIME** later, on 21 February/1 March before dawn at **17° 25' 30" south** being at **17° 25' 30" south**. Forty seconds must be subtracted to reduce the position to the **Ursa Major ATTR**, putting **Regulus** at **17° 25' 30" south**.

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

Introduction

Data Preparation

Data Annotation

Refined Named Entity Recognition Model

Explicit Observation Extraction

Data Publishing

- Data is annotated entity by entity using regex pattern.
- The result of each step is saved as jsonl fomat

Dragon was 10° 30' 30'' above which should have been 10° 30' 30.00''. We know that the sun's radius is 216 minutes high. Therefore, the corrected distance of 10° 30' 30'' from the tail of the dragon was 10° 30' 30.00''. And since the altitude **ATN** of the **Red star** **ATN** is 10° 30' 30'', the remainder by subtraction is 10° 30' 30.00'' minus 10° 30' 30'' = 0''. That is precisely the same longitude **ATN** as the **Red star** **ATN**. But because there was a difference of 3/5 degrees between them (as is clear from the following observations), a digit correction is required. For let AB be 10° 30' 30.00'' on a parallel close to the **Ursa Major** & **Ursa Minor**, C the **Red star** **ATN** and D, the **Red star** **ATN**. Dividing the arc of CD by the arc of AB gives the secant of C, $\sec(CD)$, which, substituted from $\sec(CD) = \frac{AB}{CD}$ the secant **ATN** of the **Red star** **ATN** leaves $CD = AB \cdot \sec(CD)$, or $CD = AB \cdot \sec(10^{\circ} 30' 30'')$. At the same time, we found $CD = 10^{\circ} 30' 30.00''$ between **Red star** and **Red star** **ATN** 19' 20'' corrected, and **Red star** **ATN** between **Red star** **ATN** and the bright star **Regulus** in the wing of **Vega**, $10^{\circ} 30' 30.00''$ corrected. From these two distances (using the latitudes of the stars and **Red star** **ATN**) **Regulus** **ATN** is found to be $10^{\circ} 30' 30.00''$ by consensus of all measurements. Alternatively, the **Red star** **ATN** **Altitude** **ATN** of **Red star** **ATN** is found to be $10^{\circ} 30' 30.00''$ **TAE**, using two quadrants, to be $10^{\circ} 30' 30.00''$. While the tail of **Ursa** was $10^{\circ} 30' 30''$. Therefore, from the latitudes and right ascensions of the fixed stars and our distances, **Regulus** **ATN** position is determined as $10^{\circ} 30' 30.00''$.

27. This is the typical procedure I have included the order 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 its understanding can under certain circumstances be applied. For in the previous procedure there is less to do than in the repeating of E. **ATN** **Regulus** **ATN** **TAE** using 1/28. Therefore, the **Regulus** **ATN** was about 102° from the zenith. And since **Regulus** **ATN**'s distance from the earth was rather more than half the sun's distance, the resulting **Regulus** **ATN** of about 32° (approx 32°) in our realistic table shows a latitudinal **ATN** of $10^{\circ} 30' 30''$. Thus the **Regulus** **ATN** as seen from the middle of the earth would be $10^{\circ} 30' 30''$ **Regulus** **ATN**. And because the **Altitude** **ATN** of **Regulus** **ATN** was 50°, the **Longitude** **ATN** **Regulus** **ATN** at the horizon was $10^{\circ} 30' 30''$. But since **Regulus** **ATN** was 38° from the **Regulus** **ATN**, the **Longitude** **ATN** **Regulus** **ATN** corresponding to this position is $10^{\circ} 30' 30''$, and if this is eliminated, **Regulus** **ATN** could be placed at about $10^{\circ} 30' 30''$. 28. At that moment, the sun's position was $10^{\circ} 30' 30.00''$ **Regulus** **ATN**. The distance between the bodies was $10^{\circ} 30' 30.00''$. The sun's **Uranus** **ATN** motion was $10^{\circ} 30' 30.00''$. The true **Uranus** **ATN**, therefore followed 2 days 43 minutes **TAE** later, on 21 February/March before dawn at $10^{\circ} 30' 30.00''$ **Regulus** **ATN** **Regulus** **ATN** being at $10^{\circ} 30' 30.00''$ **Regulus** **ATN**. Forty seconds must be subtracted to reduce the position to the **Uranus** **ATN**, putting **Uranus** **ATN** $10^{\circ} 30' 30.00''$.

Refined Name Entity Recognition (NER) by A Customized Spacy Model and Pattern Rules of RegEx

Introduction

Data Preparation

Data Annotation

Refined Named Entity Recognition Model

Explicit Observation Extraction

Data Publishing

- 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...)

Dragon was **10° 30' 30'' south**, which should have been **10° 30' 30'' south**. We know that the sun's nadir is 21° minutes high. Therefore, the corrected distance of **10° 30' 30''** from the tail of **IC** is **10° 30' 30''**. And since the **Latitude ATTR** of the **Scutellate ARCT** is **10° 30' 30''**, the remainder by subtraction is **10° 30' 30''**.
Remember, **IC** was at precisely the same **Longitude ATTR** as the **Scutellate ARCT**. But because there was a difference of 3½ degrees between them (as is clear from the following observations), a digit correction is required. For let AB be **10° 30' 30''** on a parallel close to the **Scutellate ARCT** & **10° 30' 30''** C the **Scutellate ARCT** and IC **10° 14' 52''**. Dividing the arc of IC by the arc of AB gives the secant of **CA**, **sec 10° 30' 30''**, which, subtracted from **sec 10° 30' 30''**, the **Scutellate ARCT** of the **Scutellate ARCT** leaves **10° 30' 30''** left. **AB** is **10° 30' 30''** distant. **AB** is **10° 30' 30''**. At the same time, we found **10° 30' 30''** between **IC** & **IC** & **10° 14' 52''** 19° 20' corrected, and **10° 30' 30''** between **IC** & **IC** and the bright **star ATTR** in the wing of **Vega**, **10° 30' 30''** **south** corrected. From these two distances (using the latitudes of the stars and **IC**), **10° 30' 30''** & **10° 30' 30''** **Scutellate ARCT** is found to be **10° 30' 30''** **south** by consensus of all measurements. Alternatively, the **Scutellate ARCT** **Latitude ATTR** of **10° 30' 30''** was found to be **10° 40' 10''** **TAME**, using two quadrants, to be **10° 14' 52''**, while the tail of **IC** was **10° 30' 30''**. Therefore, from the declinations and right ascensions of the fixed stars and our distances, **Scutellate ARCT**'s position is determined as **10° 30' 30''**.

Scutellate ARCT **Latitude ATTR** is **10° 30' 30''**. This is the typical procedure I have included the order for the sake of showing a consensus, and also that it might help the reader that despite the lack of absolute perfection in the demonstration, short cuts either in computation or in its understanding can under certain circumstances be applied. For in the previous procedure there is less to do than in the repeating of R. A. (**Scutellate ARCT**) **Latitude ATTR** using **IC**. Therefore, the **Scutellate ARCT** was about 10° from the zenith. And since **IC**'s distance from the earth was rather more than half the sun's distance, the resulting **Scutellate ARCT** of about 10° opposite 32½° in our altitude table shows a latitudinal **Latitude ATTR** of **10° 30' 30''**. Thus the **Scutellate ARCT** as seen from the middle of the earth would be **10° 30' 30''** **south**. And because the **Latitude ATTR** of **Scutellate ARCT** was 50°, the **Scutellate ARCT** **Scutellate ARCT** at the horizon was **10° 30' 30''**. But since **IC** was 38° from the **Scutellate ARCT**, the **Scutellate ARCT** **Scutellate ARCT** corresponding to this position is **10° 30' 30''**, and if this is eliminated, **Scutellate ARCT** could be placed at about 10° 10' **south**. 25. At that moment, the sun's position was **12° 30' 45''** **Pisces ATTR**. The distance between the bodies was **10° 30' 30''**. The sun's **Scutellate ARCT** moves west **10° 30' 30''**, and that is **10° 30' 30''**. (For it was **10° 30' 30''** on **15° 15' 15''** **Scutellate ARCT** in 1958, and 24 of **26** **Vega** in 1967). The sum of the **Scutellate ARCT** motions was **10° 30' 30''**. The true **Scutellate ARCT** therefore followed 2 days **43** minutes **TAME** later, on **21 February** (March before a **Scutellate ARCT** **Scutellate ARCT** being at **10° 30' 30''**). Forty seconds must be subtracted to reduce the position to the **Scutellate ARCT**, putting **Scutellate ARCT** = **10° 30' 30''**.

- 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!

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

Introduction

Data Preparation

Data Annotation

Refined Named Entity Recognition Model

Explicit Observation Extraction

Data Publishing

- A model word representation Bert with 100 iteration and batch size 16 has been used for NER classification

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	

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

Introduction

Data Preparation

Data Annotation

Refined Named Entity Recognition Model

Explicit Observation Extraction

Data Publishing

- A model word representation Bert with 100 iteration and batch size 16 has been used for NER classification
- The evaluation result per entity and overall is calculated by comparing with gold standard 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	

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

Introduction

Data Preparation

Data Annotation

Refined Named Entity Recognition Model

Explicit Observation Extraction

Data Publishing

- you can see here an example of explicit observation extraction.

183

On 1587 March 7 at 19h 10m they deduced the position of Mars from the observations, which was 25° 10' 20" Virgo.

On 1587 March 7 at 19h 10m they deduced the position of Mars from the observations, which was 25° 10' 20" Virgo. This they kept in the table, but changed the time to 17h 22m. The difference of 1h 48m multiplied by a diurnal motion of 24' gives the same number of minutes and seconds (that is, 1' 48"), no more. It therefore should have been 25° 8' 32" Virgo, which also approaches nearer the point opposite the sun. The difference is of practically no importance ^9.

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

Introduction

Data Preparation

Data Annotation

Refined Named Entity Recognition Model

Explicit Observation Extraction

Data Publishing

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

183

On 1587 March 7 at 19h 10m they deduced the position of Mars from the observations, which was 25° 10' 20" Virgo.

On 1587 March 7 at 19h 10m they deduced the position of Mars from the observations, which was 25° 10' 20" Virgo. This they kept in the table, but changed the time to 17h 22m. The difference of 1h 48m multiplied by a diurnal motion of 24' gives the same number of minutes and seconds (that is, 1' 48"), no more. It therefore should have been 25° 8' 32" Virgo, which also approaches nearer the point opposite the sun. The difference is of practically no importance ^9.

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

Introduction

Data Preparation

Data Annotation

Refined Named Entity Recognition Model

Explicit Observation Extraction

Data Publishing

- We have developed a useful framework citableclass that we can use in order to publish and use the data.

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

Introduction

Data Preparation

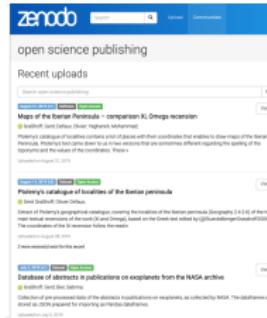
Data Annotation

Refined Named Entity Recognition Model

Explicit Observation Extraction

Data Publishing

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



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

Introduction

Data Preparation

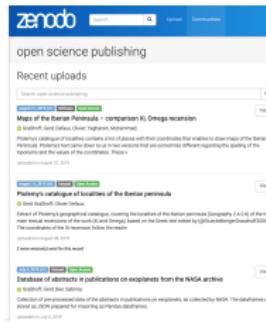
Data Annotation

Refined Named Entity Recognition Model

Explicit Observation Extraction

Data Publishing

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



- how can we have reusable notebooks files(env, requ...), there is some answers like but it seems not perfect

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

Introduction

Data Preparation

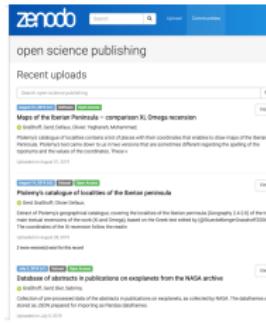
Data Annotation

Refined Named Entity Recognition Model

Explicit Observation Extraction

Data Publishing

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



- how can we have reusable notebooks files(env, requ...), there is some answers like but it seems not perfect
- find the best environment to cooperate with (colab,binder,

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

Introduction

Data Preparation

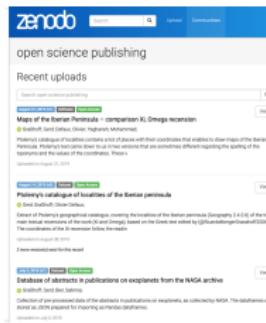
Data Annotation

Refined Named Entity Recognition Model

Explicit Observation Extraction

Data Publishing

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



- how can we have reusable notebooks files(env, requ...), there is some answers like but it seems not perfect
- find the best environment to cooperate with (colab,binder,
- find the best way to present the result in way that non-technical readers could be able to go through the text and computational analysis (voila)

-  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