

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

January 15, 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**

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**
- **Customized 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**
- **Customized 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**
- **Customized 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 annotation workflow.

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 annotation workflow.
- Unavailability of training data.

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 annotation workflow.
- Unavailability of training data.
- Limited 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

communicate the workflow and ideas:

- **Preprocessing** 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

communicate the workflow and ideas:

- **Preprocessing** 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

communicate the workflow and ideas:

- **Preprocessing** and preparation of text data for classification.
- **Information Extraction** in an informative and interactive way.
- **Customized Named Entity Recognition (CNER)** model using state of the art deep learning models.

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

communicate the workflow and ideas:

- **Preprocessing** and preparation of text data for classification.
- **Information Extraction** in an informative and interactive way.
- **Customized Named Entity Recognition (CNER)** model using state of the art deep learning models.
- **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

communicate the workflow and ideas:

- **Preprocessing** and preparation of text data for classification.
- **Information Extraction** in an informative and interactive way.
- **Customized Named Entity Recognition (CNER)** model using state of the art deep learning models.
- **Evaluation** of model using gold standard data.
- **Explicit Observation Extraction** from Text Corpus.

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

communicate the workflow and ideas:

- **Preprocessing** and preparation of text data for classification.
- **Information Extraction** in an informative and interactive way.
- **Customized Named Entity Recognition (CNER)** model using state of the art deep learning models.
- **Evaluation** of model using gold standard data.
- **Explicit Observation Extraction** from Text Corpus.
- Toward **Relation Extraction** and Causal Inference.

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

Introduction

Data Preparation

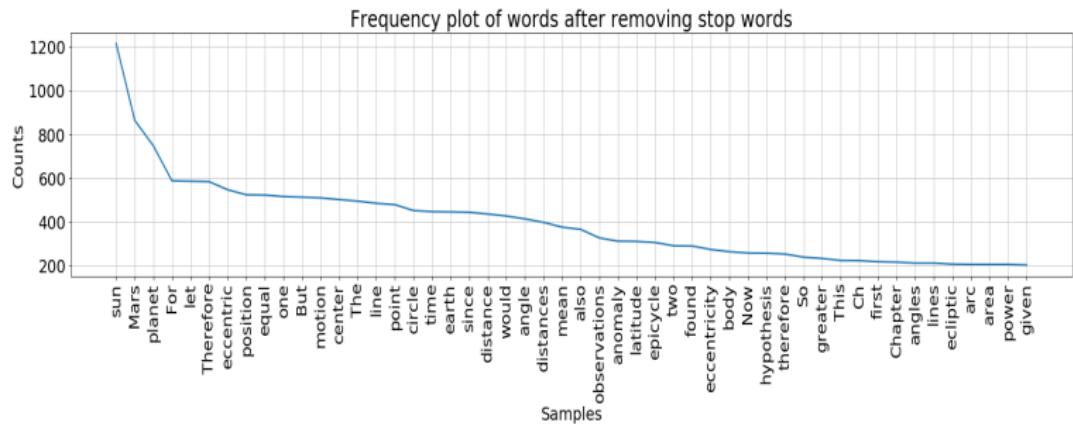
Data Annotation

Refined Named Entity Recognition Model

Explicit Observation Extraction

Data Publishing

- Corpus has 70 chapters including 1605 paragraphs, 6699 sentences



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

Introduction

Data Preparation

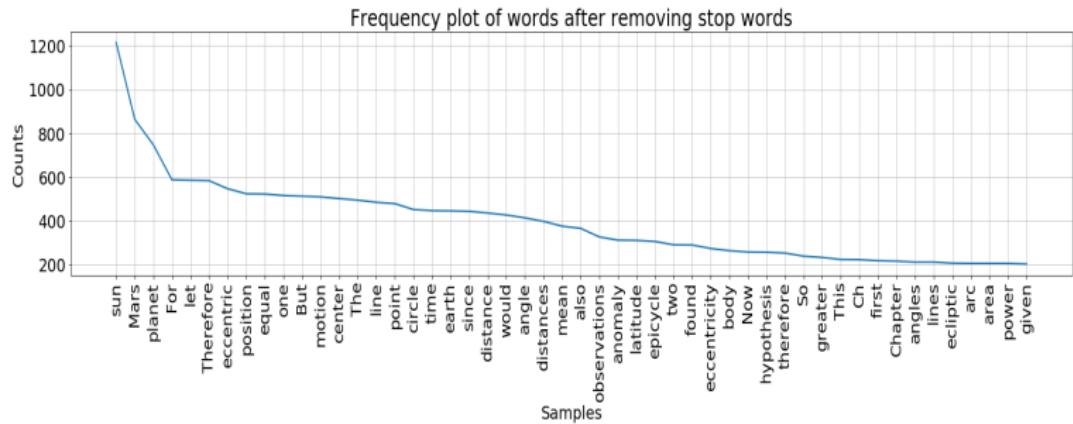
Data Annotation

Refined Named Entity Recognition Model

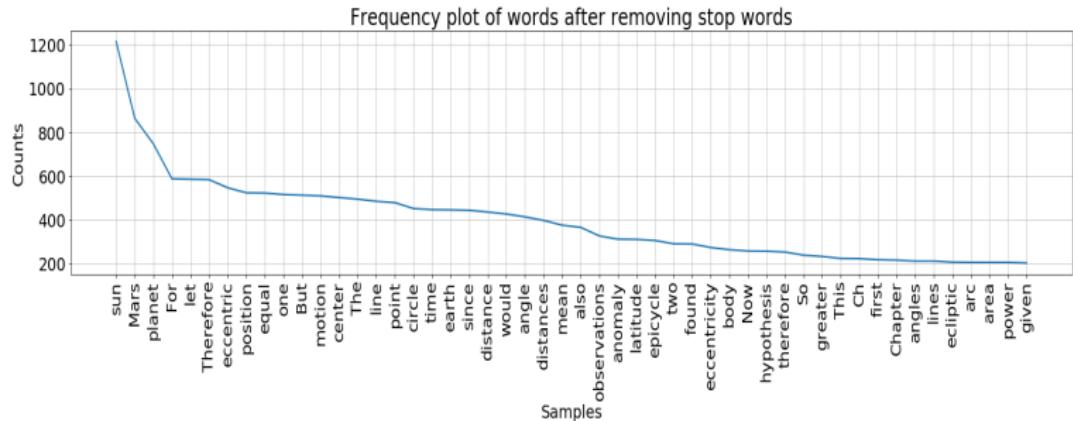
Explicit Observation Extraction

Data Publishing

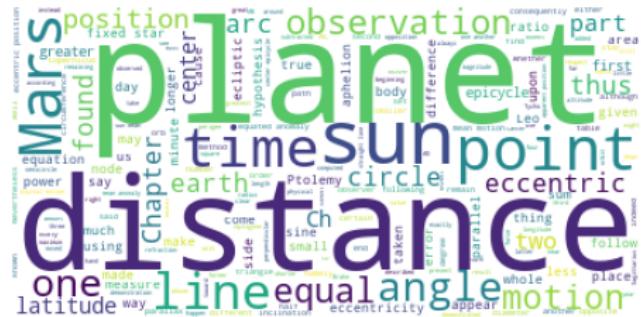
- Corpus has 70 chapters including 1605 paragraph, 6699 sentences
- Corpus includes 169231 tokens (roughly speaking; words) and 9513 unique tokens



- 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



- Word cloud of the whole book and by chapters as initial 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

■ Couple of sentences with different attribute that has been add by our analysis and can be used later for classification, relation extraction and...

| | | Sentence | SentIndex | Chapter | Paragraph | ParalIndex | ASO | Entities | CNER | Label |
|-----|--|----------|-----------|---------|--|------------|--|--|-----------|-------|
| 158 | On 1580 November 12 at 10h 50m, I 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, I 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 TIME], [Mars, PLAN], [8° 36' 50", LONG], [diurnal, ASTR]] | [1, 1, 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 regressed 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, he 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", LONG], [observation, ASTR]] | [1, 1, 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 parallax. 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, 1] | 1 |

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" long**, which should have been **17° 26' 30" long**. We know that the sunset ends 2½ minutes high. Therefore, the corrected distance of **1800 miles** from the tail of **Regulus** was **17° 25' 30" long**. And since the **Latitude ATTR** of the **Red star ATTR** is **17° 25' 30" long**, the remainder by subtraction is **15° 45' 30" long**, meaning **15° 45' 30" long** was at precisely the same **Longitude ATTR** as the **Red star ATTR**. 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 **17° 25' 30" long** on a parallel close to the **Ursa Major ATTR** & **17° 26' 30" long** on a parallel close to the **Ursa Minor ATTR**, and let C be **17° 25' 30" long**. Dividing the arc of CB by the arc of AB gives the secant of **C**, **sec 17° 26' 30"**, which, subtracted from **sec 17° 25' 30"**, the **Latitude ATTR** of the **Red star ATTR** leaves **1° 20' 30" right ATTR** & **1° 20' 30" left ATTR**. At the same time, we find **17° 25' 30" long** between **Regulus** and **Antares**, **1° 20' 30" corrected**, and **17° 26' 30" uncorrected** between **Regulus** and the bright **star Alpha** in the wing of **Vega**, **1° 20' 30" long** connected. From these two distances (using the latitudes of the stars and **Regulus** & **Alpha**) **Regulus** is found to be **17° 25' 30" long**, **1° 20' 30" long** by consensus of all measurements. Alternatively, the **Latitude ATTR** **Latitude ATTR** of **Regulus** was found to be **17° 26' 30" long**, using two quadrants, to be **17° 25' 30" long**, while the tail of **Castor** was **17° 25' 30" long**. Therefore, from the declinations and right ascensions of the fixed stars and our distances, **Regulus**' position is determined as **17° 25' 30" long**.

27. This is the typical procedure I have included the order for the sake of showing a consensus, and also that it might be the 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 be the actual work than in the repeating of E. A. [Trigonom. ATTR] **Latitude ATTR** using 1/26. 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 32½° in our astrolabe table shows a latitudinal **Latitude ATTR** of **17° 25' 30" long**. Thus the **Latitude ATTR** as seen from the middle of the earth would be **17° 25' 30" right ATTR**. And because the **Latitude ATTR** of **Regulus** was 50°N, the **Longitude ATTR** **Latitude ATTR** at the horizon was **17° 25' 30" long**. But since **Regulus** was 38° from the **Longitude ATTR**, the **Longitude ATTR** **Latitude ATTR** corresponding to this position is **17° 25' 30" long**, and if this is eliminated, **Regulus** could be placed at about **17° 25' 30" long**. 28. At that moment, the sun's position was **17° 25' 30" right ATTR**. The distance between the bodies was **17° 25' 30" long**. The sun's **Longitude ATTR** motions were **17° 25' 30" long**. The true **Longitude ATTR** therefore followed 2 days 43 minutes **TIME** later, on 21 February/March before dawn at **17° 25' 30" long**, **17° 25' 30" long** being at **17° 25' 30" right ATTR**. Forty seconds must be subtracted to reduce the position to the **Longitude ATTR**, putting **Regulus** at **17° 25' 30" long**.

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

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** and **IC + 3½°** = **10° 20' 30''** corrected, and **10° 30' 30''** between **IC** and the bright **star ARCT** in the wing of **Vega**.
IC is **10° 30' 30''** corrected. From these two distances (using the latitudes of the stars and **IC**) **10° 30' 30''** **Scutellate ARCT** is found to be **10° 30' 30''** **Scutellate ARCT** by consensus of all measurements. Alternatively, the **Scutellate ARCT** **Latitude ATTR** of **10° 30' 30''** was found to be **10° 40' 30''** **TAU**, using two quadrants, to be **10° 30' 30''** **Scutellate ARCT**, while the tail of **IC** was **10° 30' 30''**. Therefore, from the latitudes and right ascensions of the fixed stars and our distances, **Scutellate ARCT**'s position is determined as **10° 30' 30''**.

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 reworking of E. AR **(Trig-star NAME)** **(Scutellate ARCT)** using **10°**. 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 32½° in our unrealistic 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''** **Scutellate ARCT**. And because the **Latitude ATTR** of **Scutellate ARCT** was 50°, the **Scutellate ARCT** **Scutellate ARCT** at the horizon = **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° 30' 30''**.
28. At that moment, the sun's position was **12° 45' 42''** **Pisces** **1958**. The distance between the bodies was **10° 30' 30''**. The sun's **Scutellate ARCT** moved west **10° 30' 30''**, and that is **Scutellate ARCT** (for it was **10° 30' 30''** **Scutellate ARCT** in 1958, and 24 of 28 **Vega** in 1967). The sum of the **Scutellate ARCT** motions was **10° 30' 30''**. The true **Scutellate ARCT** therefore followed 2.6 hours 43 minutes **TAU** line, on 21 February/1 March before dawn at **10° 30' 30''** **Scutellate ARCT** being at **10° 30' 30''** **Scutellate ARCT**. Forty seconds must be subtracted to reduce the position to the **Scutellate ARCT**, putting **Scutellate ARCT** = **10° 30' 30''**.

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...)
- Annotated data is merged and now the training data is ready!

Dragon was **10° 20' 30'' south**, which should have been **10° 19' 30'' south**. We know that the sunset ends 2½ minutes high. Therefore, the corrected distance of **10° 19' 30''** from the tail of **Regulus** was **10° 18' 30''**. And since the **Latitude AT&T** of the **Regulus** is **10° 18' 30''**, the remainder by subtraction is **1' 30''**. Assuming **Regulus** that **Regulus** was at precisely the same **Longitude AT&T** as the **Regulus** star, **Regulus** would be **1' 30''** away. But because there was a difference of 5½ degrees between them ($\pi/8$ is clear from the following observation), a digit correction is required. For let AB be **10° 19' 30''** on a parallel close to the **Ursa Major** & **Ursa Minor** C the **Regulus** **AT&T** and IC **10° 18' 30''**. Dividing the arc of IC by the arc of AB gives the secant of CA, **sec(1.5°)**, which, substituted from **10° 19' 30''** the **Regulus** **AT&T** of the **Regulus** leaves **10° 18' 30''** with **AT&T** of **10° 18' 30''** different **1' 30''** **Regulus**. At the same time, we found **1' 30''** between **Regulus** and **Regulus** **AT&T** 18' 20'' corrected, and **Regulus** between **Regulus** and the bright **Regulus** in the wing of **Vega**, **Regulus**. **Regulus** connected. From these two distances (using the latitudes of the stars and **Regulus** **AT&T** **Regulus**) **Regulus** **AT&T** is found to be **10° 19' 30'' south**, by consensus of all measurements. Alternatively, the **Regulus** **AT&T** **Latitude AT&T** of **Regulus** **AT&T** is found to be **10° 19' 30''** **Regulus**, using two quadrants, to be **10° 19' 30''**, while the tail of **Regulus** was **10° 18' 30''**. Therefore, from the declinations and right ascensions of the fixed stars and our distances, **Regulus** position is determined as **10° 19' 30''**.

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 lies in the actual work than in the reporting of R. A. [**Regulus**] [**Regulus**] **AT&T** we using 1/8. Therefore, the **Regulus** **AT&T** was about 12½° from the zenith. And since **Regulus** **AT&T**'s distance from the earth was rather more than half the sun's distance, the resulting **Regulus** **AT&T** of about 32½° in our celestial table shows a latitudinal **Regulus** **AT&T** of **10° 19' 30''**. Thus the **Regulus** **AT&T** as seen from the middle of the earth would be **10° 19' 30''** **Regulus**. And because the **Latitude AT&T** of **Regulus** **AT&T** was 57°, the **Longitude AT&T** **Regulus** **AT&T** at the horizon was **171° 30'**. But since **Regulus** was 38° from the **Regulus** **AT&T**, the **Longitude AT&T** **Regulus** **AT&T** corresponding to this position is **171° 30'**, and if this is eliminated, **Regulus** **AT&T** could be placed at about **171° 30' Regulus**. 28. At that moment, the sun's position was **12° 45' 30''** **Pisces** **AT&T**. The distance between the bodies was **1' 30''**. The sun's **Regulus** **AT&T** motion was **1' 30''** **AT&T**, and that is **1' 30''** **Regulus** **AT&T**. If it was **1' 30''** **Regulus** **AT&T** (1858) and 24 of **Regulus** in 1867, the sum of the **Regulus** **AT&T** motions was **1' 30''** **AT&T**. The true **Regulus** **AT&T** therefore followed 2 days 43 minutes **Regulus** **AT&T** line, on 21 February/1 March before dawn at **Regulus** **AT&T** **Regulus** **AT&T** being at **10° 19' 30'' Regulus**. Forty seconds must be subtracted to reduce the position to the **Regulus** **AT&T**, putting **Regulus** **AT&T** **Regulus** **AT&T**.

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

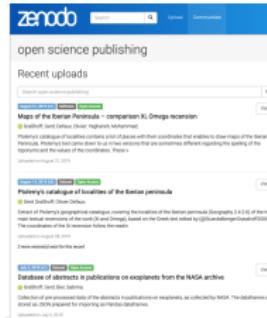
Data Annotation

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

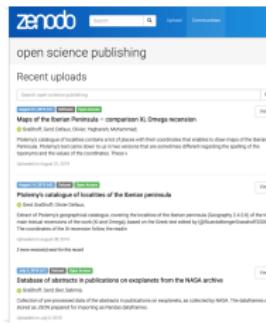
Data Annotation

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

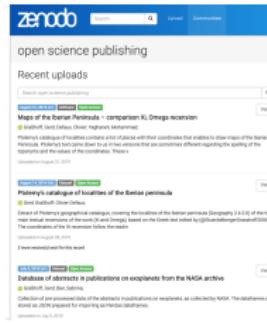
Data Annotation

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

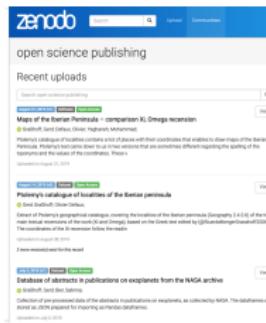
Data Annotation

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