Tackling Semantic :
Semantic Analysis of Text Corpora using AI

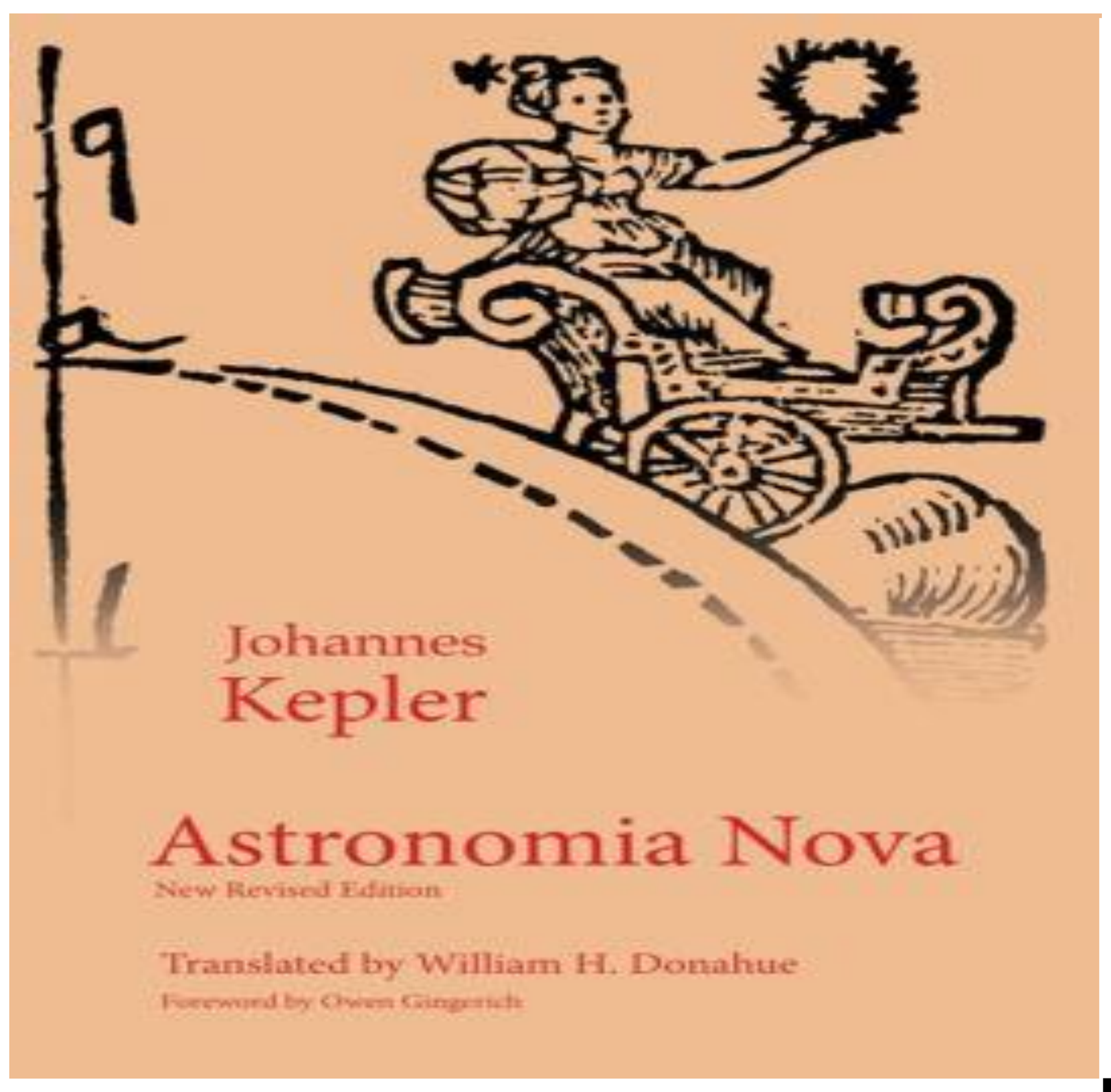
HU Berlin

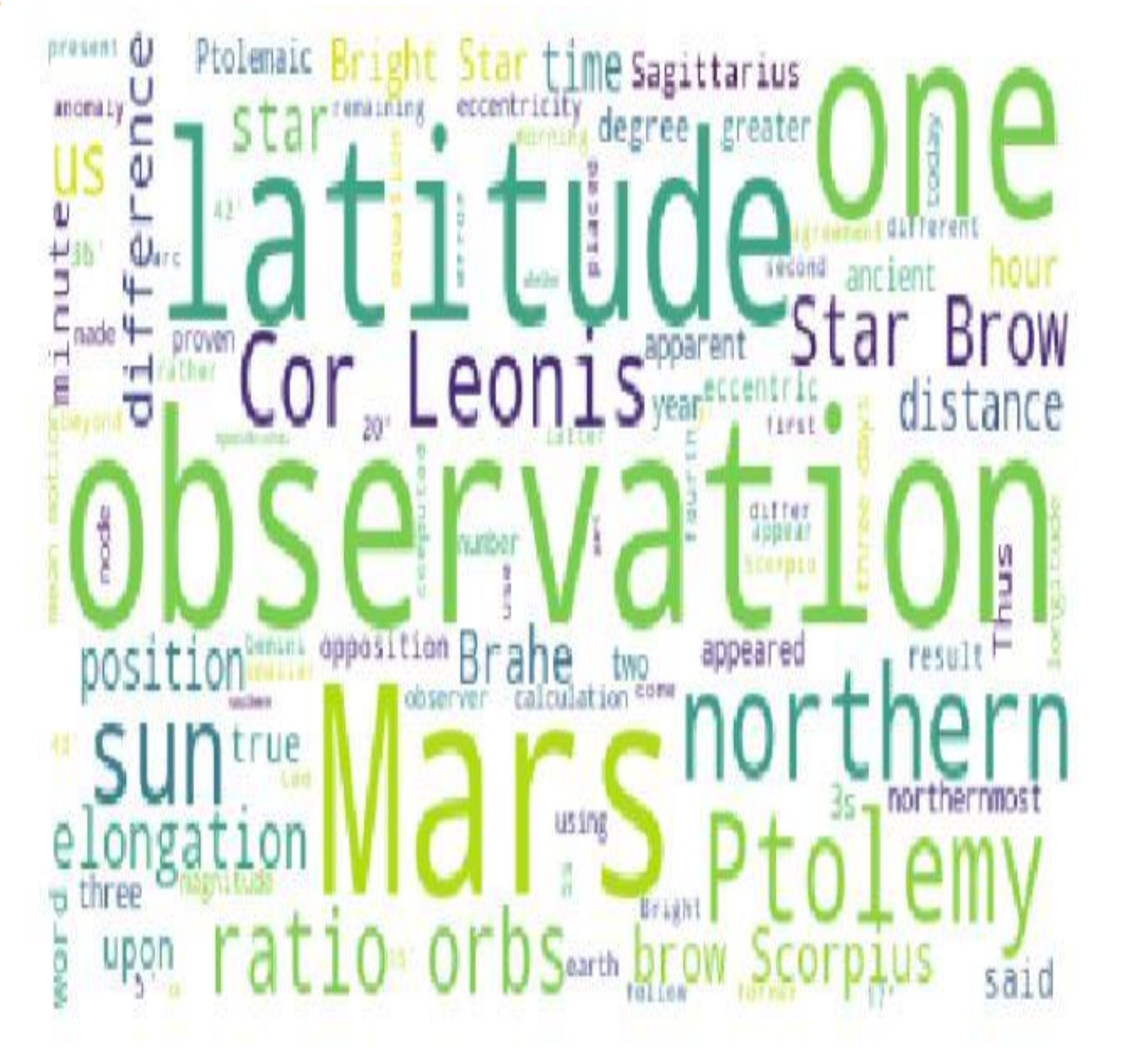
Analysis of Text Corpora

1) Kepler's book "Astronomia nova" Corpus
-
-
-

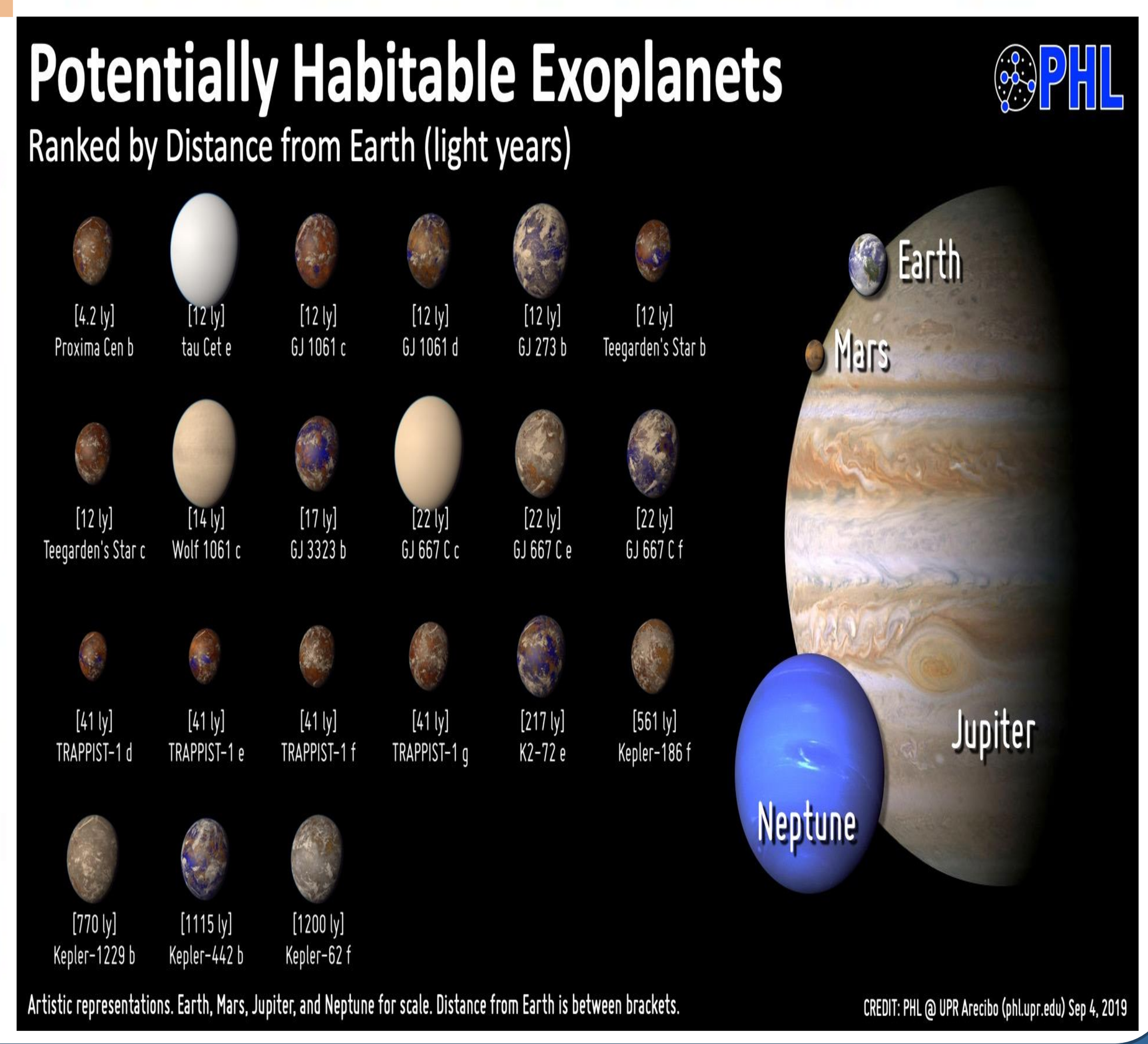
2) Exoplanet Publication Corpus [1]
-
-

3) Corona Virus Publication Corpus
-





y=0 top features		y=1 top features		y=2 top features		y=3 top features	
Weight?	Feature	Weight?	Feature	Weight?	Feature	Weight?	Feature
+3.054	present	+21.317	compare	+14.381	propose	+13.982	conclude
+2.851	<BIAS>	+12.742	compared	+9.543	proposed	+1.150	strong
+2.061	model	+2.287	predictions	+1.402	macho	+0.932	fully
+1.711	using	+1.631	spectra	+1.239	microlensing	+0.858	suggests
+1.207	use	+1.444	values	+1.106	theta_	+0.840	likely
+1.190	planet	+1.341	derived	+0.950	blg	+0.813	taken
+1.186	explore	+1.301	obtained	+0.936	better	+0.802	pulsation
+1.023	demonstrate	+1.240	results	+0.925	test	+0.799	presented
... 6401 more positive ...		+1.225	trappist	+0.923	coronagraphic	+0.696	nonequilibrium
... 1810 more negative ...		+1.144	observed	+0.901	statistical	+0.671	considered
-1.006	optically	+1.060	slow	+0.814	interception	+0.658	necessary
-1.011	coronagraphic	+1.048	j_	+0.787	extra	+0.643	explanation
-1.020	events	+1.023	t_	+0.764	collect	+0.641	normal
-1.024	conclusions	+1.016	000	+0.737	experimental	+0.621	114762b
-1.051	derived	+1.007	literature	... 752 more positive ...		+0.604	difficult
-1.060	predicted	+1.007	frequencies	... 7459 more negative ...		... 555 more positive ...	
-1.069	spectra	+0.977	chromospheric	-0.732	different	... 7656 more negative ...	
-1.093	values	+0.972	existing	-0.742	spectra	-0.594	results
-1.099	support	+0.967	measurements	-0.746	surface	-0.601	new
-1.107	obtained	+0.967	compares	-0.757	planets	-0.643	apply
-1.110	microlensing	+0.951	spectrum	-0.799	stellar	-0.659	methods
-1.141	macho	... 1287 more positive ...		-0.819	temperature	-0.660	models
-1.143	trappist	... 6924 more negative ...		-0.895	evolution	-0.682	based
-1.158	pulsation	-0.944	use	-0.930	light	-0.816	discuss
-1.233	better	-0.954	investigate	-0.935	using	-0.882	use
-1.237	literature	-0.977	composition	-1.001	earth	-0.935	planet
-2.106	predictions	-1.070	using	-1.015	modeling	-0.984	planets
-8.006	proposed	-1.210	conclude	-1.086	compare	-1.118	stellar
-10.899	conclude	-1.216	planet	-1.597	models	-1.245	present
-11.718	compared	-2.089	present	-1.912	present	-1.578	using
-12.273	propose	-2.091	model	-1.917	model	-2.369	model
-19.718	compare	-3.357	<BIAS>	-3.698	<BIAS>	-4.037	<BIAS>



Structural Corpora

1) Multi Index Dataframe

2) Sentence Structure with Spacy

a) Customized Entities

b) Dependency Parsing

Customized Named Entity Recognition (NER) [2]

Key Idea. Hybrid approach; Pattern based & deep learning to provide NER Model by developing spaCy.

```
12]: df.head()
```

	Chapter	ParaIndex	SentIndex	Paragraph	Sentence	Entities	Label
0	10	218	967	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.	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.	[[1580 November 12, DATE], [10h 50m,1 they, TIME], [Mars, PLAN], [8° 36′ 50″, LONG], [diurnal, ASTR]]	Obs
1	10	221	977	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, 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.	On 1582 December 28 at 11h 30m, they set Mars down at 16° 47′ Cancer by observation ^6.	[[1582 December 28, DATE], [11h 30m, TIME], [Mars, PLAN], [16° 47′ Cancer, LONG], [observation, ASTR]]	Obs
2	10	222	984	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.	On 1585 January 31 at 12h 0m, Mars was placed at 21° 18′ 11″ Leo ^8.	[[1585 January 31, DATE], [12h 0m, TIME], [Mars, PLAN], [21° 18′ 11″ Leo, LONG]]	Obs
3	10	223	992	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.	On 1587 March 7 at 19h 10m they deduced the position of Mars from the observations, which was 25° 10′ 20″ Virgo.	[[1587 March 7, DATE], [19h 10m, TIME], [Mars, PLAN], [25° 10′ 20″ Virgo., LONG]]	Obs
4	10	224	997	On 1589 April 15 at 12h 5m they established the position of Mars very carefully at 3° 58′ 21″ Scorpio, and corrected for longitudinal parallax so as to make it 3° 57′ 11″ ^10. There remain 1h 30m until the designated moment of opposition, which, for a diurnal motion of 22′, bring the planet back 1′ 22″, so as to be at 3° 55′ 49″. They took the value of 3° 58′ 10″. The former is closer to the sun's mean position.	On 1589 April 15 at 12h 5m they established the position of Mars very carefully at 3° 58′ 21″ Scorpio, and corrected for longitudinal parallax so as to make it 3° 57′ 11″ ^10.	[[1589 April 15, DATE], [12h 5m, TIME], [Mars, PLAN], [3° 58′ 21″ Scorpio, LONG], [longitudinal, ASTR], [parallax, ASTR], [3° 57′ 11″, LONG]]	Obs

observations ^2. Suppose, however, that on 1580 November 12 DATE at 10h 50m TIME Mars PLAN was observed at 8° 37′ Gemini,3 COOR and the motion over five days was 1° 55′ COOR . Since at the given time the sun stood at 0° 45′ 36″ Sagittarius COOR , and its motion over five days is 5° 5′ COOR , the sum of the two motions will come to 7° 0′ COOR . But the sun is 7° 51′ 24″ COOR removed from opposition to Mars PLAN . Of this, seven degrees exactly are traversed in 5 days, or 120 hours TIME . So, according to the same ratio, the remaining 51′ 24″ COOR will be traversed in 14 hours 41 minutes TIME . Therefore, the moment of opposition was November 18 DATE at 1h 31m TIME . Its position was 6° 28′ Gemini COOR on the ecliptic. Now this is 20° away from 16½° Taurus COOR . I want to know how much longer this makes the arc on the orbit extended from the node to the arc of latitude through 6° 28′ Gemini COOR .
So I turn to Philip Lansberg NAME 's trigonometry ^4. I mention him out of honor and gratitude, for he has supplied me in abundance with the finest axes, best adapted for building astronomical foundations, from nearby, and at little expense of time. Without him, these would have had to be sought from afar with great deal of trouble and toil, and the handles would not have been so well fitted.
From Lansberge, then, the tangent of the 20° side multiplied by the secant of the angle of inclination, 1° 50′ COOR , the last five digits being dropped,5 gives an increase of only 18½ of the smallest units, to which corresponds about 35 seconds. Mars PLAN therefore, standing opposite 6° 28′ Gemini COOR , is 35″ further along on its orbit. It should therefore be placed at 6° 28′ 35″ Gemini COOR , a tiny and quite unnecessary correction. The latitude is 1° 40′ north COOR . On the night following 1582 December 28 DATE , at 11h 30m TIME , Mars PLAN was observed at 16° 47′ Cancer,6 COOR while the true position of the sun was 17° 13′ 45″ Capricorn. COOR The moment of opposition had therefore passed. Now the sun's diurnal motion was 61′ 18″ COOR , that of Mars PLAN 24′, and their sum, 85′ 18″ COOR . At this moment, the distance between the stars was 26′ 45″ COOR . Therefore, as 1° 25′ 18″ COOR is to 24 hours TIME , so is 26′ 45″ COOR to 7 hours 32 minutes TIME . Subtracting this from 11 hours 30 minutes TIME gives December 28 DATE at 3h 58m TIME after noon as the moment of true opposition. Its position on the ecliptic was 16° 54′ 32″ Cancer COOR , and by reduction to the orbit (a 50″ correction), 16° 55½′ Cancer COOR The latitude was 4° 6′ north COOR , as given by Brahe's table of oppositions. For among the observations I find various latitudes: on the night following December 26 DATE , 4° 6′ COOR or 4° 2′ COOR , while on the night following December 29 DATE , 4° 8′ COOR or 4° 6½ COOR . On 1585 January 31 DATE at 12h 0m TIME , Mars PLAN was observed at 21° 18′ 11″ Leo COOR ^7. The sun was at 22° 21′ 31″Aquarius. COOR The true opposition had therefore passed. The distance was 1° 3′ 20″ COOR . The sun's diurnal motion was 61′ 16″ COOR , that of Mars PLAN 24′ 15″ COOR , and their sum 85′ 31″ COOR . Now as 1° 25′ 31″ COOR is to 24 hours TIME , so is 1° 3′ 20″ COOR to 17 hours 46 minutes TIME , to which correspond about 18′ of Mars PLAN 's motion. Therefore the time was January 30 DATE at 19h 14m TIME , and Mars PLAN 's ecliptic position was 21° 36′ 10″ Leo. COOR For reduction, some very small quantity is subtracted, because Mars PLAN is then beyond the limit. Therefore, the extension of the arc on the orbit from the following node is directed westward. But because Mars PLAN was only 4 or 5 degrees from the node, the subtraction is rendered quite imperceptible. The latitude, on the authority of the Tychonic table, was 4° 32′ 10″ north COOR . For the observation on January 31 DATE at 12h TIME gave 4° 31′ COOR .

Explicit Observation Extraction [2]

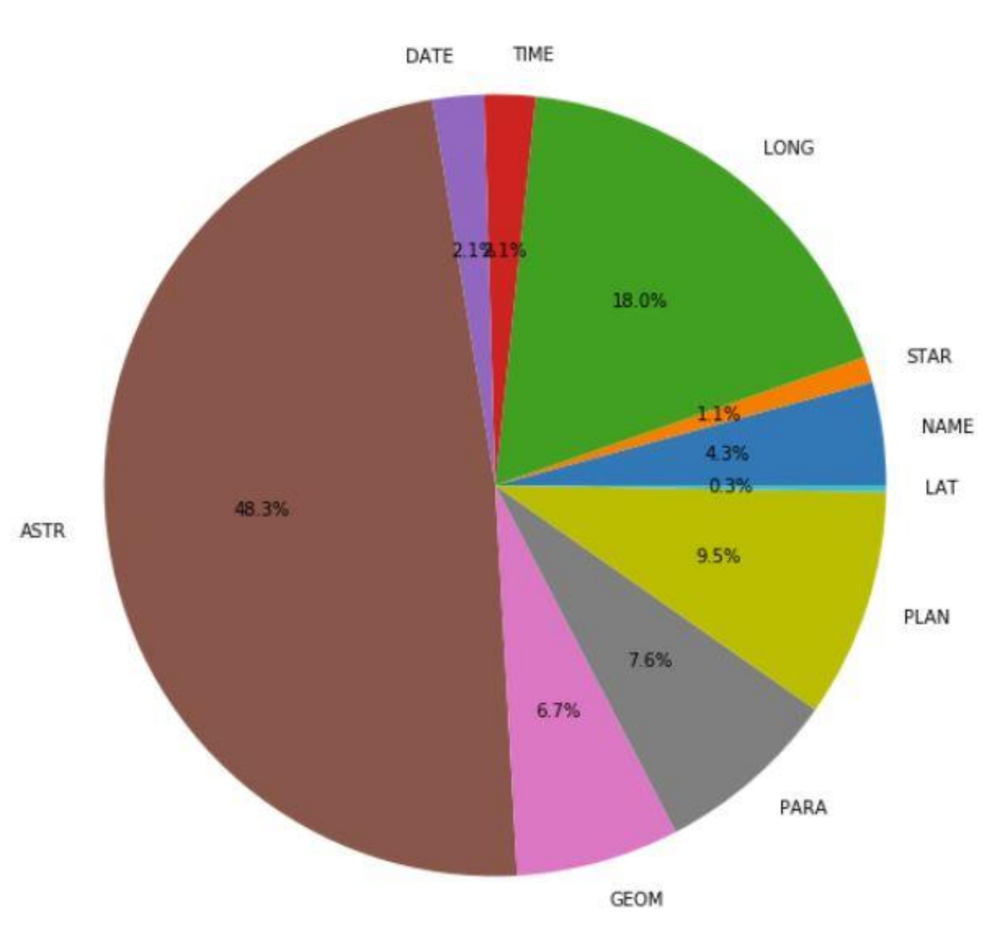
Key Idea. Extract Semantic from the corpus using customized NER model.

x 60

Finally, on 1595 July 22 DATE at 2h 40m TIME am, when the sun was at 7° 59′ 52″ Leo COOR , 101,487 PARA distant from earth, Mars PLAN 's mean longitude 11s 14° 9′ 5″ COOR , and anomaly 164° 48′ 55″ COOR , and consequent eccentric position from the vicarious hypothesis 17° 16′ 36″ Pisces COOR : the apparent position of Mars PLAN , from the most select observations, was 4° 11′ 10″ Taurus COOR , lat.

	Chapter	ParaIndex	SentIndex	Paragraph	Sentence	Entities	Label
60	51	1136	4873	Finally, on 1595 July 22 at 2h 40m am, when the sun was at 7° 59′ 52″ Leo, 101,487 distant from earth, Mars's mean longitude 11s 14° 9′ 5″, and anomaly 164° 48′ 55″, and consequent eccentric position from the vicarious hypothesis 17° 16′ 36″ Pisces: the apparent position of Mars, from the most select observations, was 4° 11′ 10″ Taurus, lat. 2° 30′ S ^37. Thus we twice have Mars in the most opportune position, in quadrature with the sun, while the positions of earth and Mars are also distant by a quadrant.\n	Finally, on 1595 July 22 at 2h 40m am, when the sun was at 7° 59′ 52″ Leo, 101,487 distant from earth, Mars's mean longitude 11s 14° 9′ 5″, and anomaly 164° 48′ 55″, and consequent eccentric position from the vicarious hypothesis 17° 16′ 36″ Pisces: the apparent position of Mars, from the most select observations, was 4° 11′ 10″ Taurus, lat.	[(1595 July 22, DATE), (2h 40m, TIME), (7° 59′ 52″ Leo, COOR), (101,487, PARA), (Mars, PLAN), (longitude, ASTR), (14° 9′ 5″, COOR), (164° 48′ 55″, COOR), (eccentric, ASTR), (17° 16′ 36″ Pisces, COOR), (Mars, PLAN), (4° 11′ 10″ Taurus, COOR)]	Obs

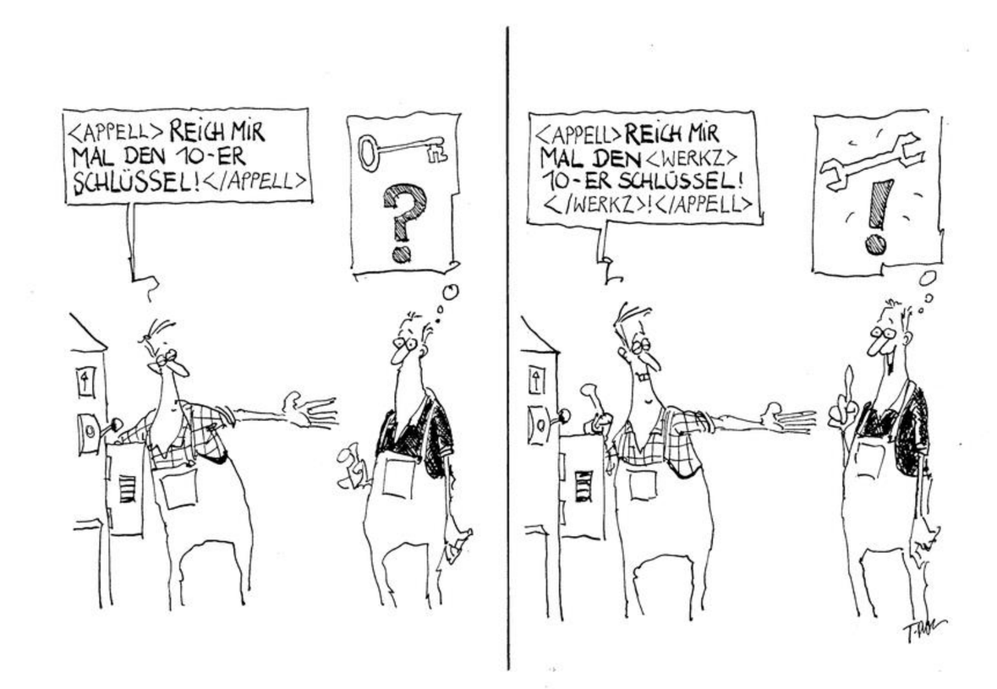
```
['NAME', 'STAR', 'LONG', 'TIME', 'DATE', 'ASTR', 'GEOM', 'PARA', 'PLAN', 'LAT'] [421, 105, 1780, 208, 211, 4773, 666, 754, 934, 27]
```



Future Direction: Semantic Objects

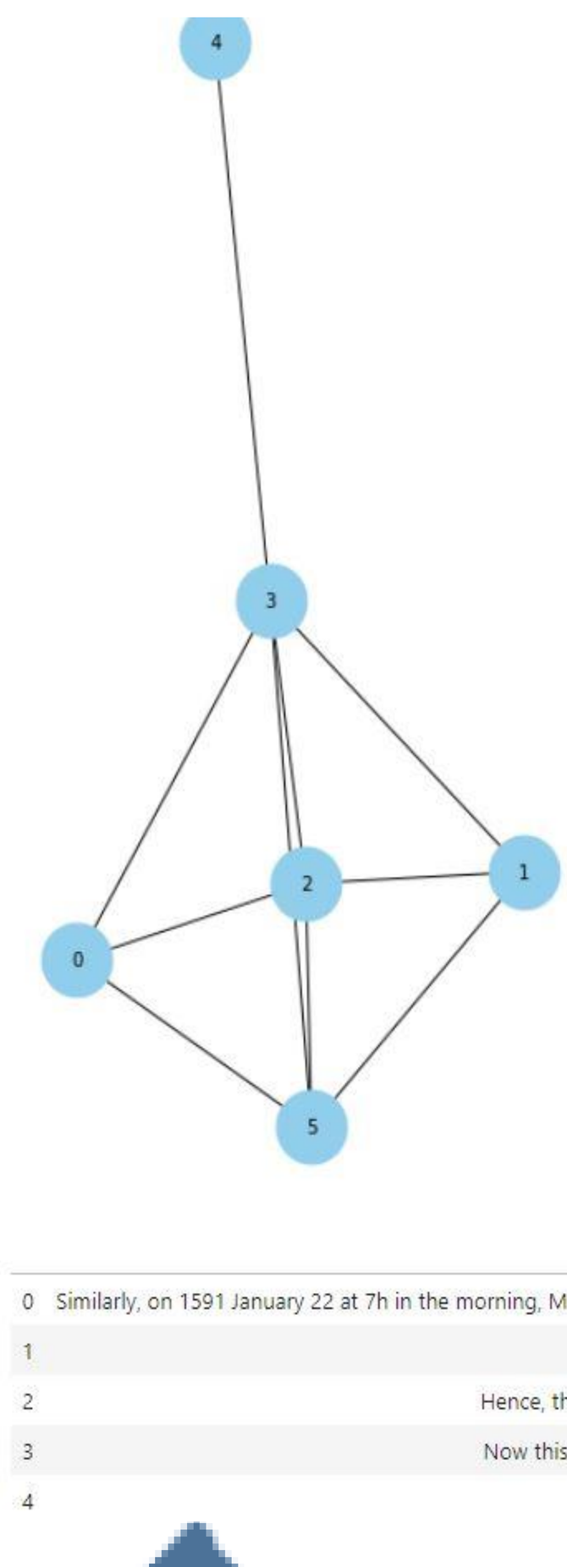
- *Tree structure*

Idea. Integrate sentences using xml format to have semantic annotation for extracted observations as a first use case



- *Semantic Relation Extraction*

Idea. Extract structure of argumentation components using semantic objects



	Sentence
0	Similarly, on 1591 January 22 at 7h in the morning, Mars was 34° 32′ 45″ from Spica with declination 17° 25′ south, at an altitude of 16° ^8.
1	Therefore, after correction for horizontal variations, the declination was 17° 30′.
2	Hence, the right ascension was 230° 23′ 12″, longitude 22° 33′ Scorpio, latitude 1° 0′ 30″ north.
3	Now this time differs from ours by 1 day 19 hours, and the diurnal motion, from Magini, is 33′.
4	Therefore, 59′ are required for the intervening time.

References

[1] Graßhoff, Gerd; Bier, Sabrina (2019): Database of abstracts in publications on exoplanets from the NASA archive. Zenodo. Dataset. https://doi.org/10.5281/zenodo.3269732

[2] Yeghaneh, Mohammad; https://github.com/myeghaneh

[3] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). " Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1135-1144).