INFORMATION RETRIEVAL AND TEXT MINING
DATA SCIENCE AND KNOWLEDGE ENGINEERING (DKE)
MSC IN ARTIFICIAL INTELLIGENCE, MAASTRICHT UNIVERSITY, NETHERLANDS

Text Mining Foundation Trilogy

S. Nogues Farres (i6267073)
May 26, 2021

More than 70 characters and 15 planets interact in Isaac Asimov's first three Foundation books, collectively known as the Foundation Trilogy. To extract, classify, and connect all of these entities, a rules-based Named Entity Recognition model is proposed. A thorough description of the methodology and the outcomes may be found on the following pages. [1]

NER, Text Mining

1. INTRODUCTION

In automated text or corpus processing, the extraction and automatic recognition of named entities (NER) is part of a syntactico-semantic analysis. The aim is to recognize and classify named entities in text into pre-defined groups such as person, organisation or location. NER can solve a variety of narrative problems, including role detection, summarizing or community analysis.

Fiction books have unique features that make these tasks more challenging [1]. NER is usually machine learning-based (ML) and thus, a training Dataset is required. This makes the ML approach not suitable for Sci-Fi NER. Rule-based algorithms, on the other hand, are more practical when there is a finite number of examples to be found, or a well-defined structured pattern with which to extract entities.

A thorough description of a Named Entity Recognition and Linking solution for the Foundation Trilogy books can be found in the following chapters.

[1] The source code can be found here: `https://github.com/serginogues/nlp_FoundationSeries/blob/master/README.md`

2. BACKGROUND

2.1. TOKENIZATION

Tokenization is the process of breaking down an expression, sentence, paragraph, or an entire text document into smaller units called tokens.

2.2. STEMMING AND LEMMATIZATION

The aim is to reduce a word's inflectional and sometimes derivationally related forms into a single base form (e.g. "went" would be reduced to "go").

Both stemmers and lemmatizers reduce the vocabulary size and increase the ambiguity of the text. However, depending on how the word was used within the text and its intended context, lemmatizers do a better job at preserving as much of the information content as possible.

2.3. PART-OF-SPEECH TAGGING (POS TAGGING)

Every token (word) in a document is given a Part-of-speech tag. The output of this process is a list of tuples where tuple is a word-tag pair. Some NLTK POS tagging examples are: CC, CD, EX, JJ, MD, NNP, PDT, or TO (see figure 1).

```
    Sentence = "Can you please buy me an Arizona Ice Tea ? It‘s \$ 0.99."
    POS tag list = [(‘Can’, ‘MD’), (‘you’, ‘PRP’), (‘please’, ‘VB’),
(‘buy’, ‘VB’), (‘me’, ‘PRP’), (‘an’, ‘DT’), (‘Arizona’, ‘NNP’), (‘Ice’,
‘NNP’), (‘Tea’, ‘NNP’), (‘?’, ‘.’), (‘It’, ‘PRP’), (“‘s”, ‘VBZ’), (‘\$’,
‘\$’), (‘0.99’, ‘CD’), (‘.’, ‘.’)]
```

Figure 1: POS tagging example

2.4. PARSING

The aim of this step is to extract precise, or dictionary-like, meaning from the text. Syntax analysis compares the text to formal grammar rules to determine its meaning. A parser finds the best tree for the given sentence.

One of the most popular parsing techniques is regexp parsing. To parse the input sentences and generate a parse tree, it applies a regular expression specified in the form of grammar on top of a POS-tagged string.

2.5. NAMED ENTITY RECOGNITION (NER)

The goal is to identify and cluster named entities into different categories that can be 'person', 'organization' or 'location' among others.

2.6. COREFERENCE RESOLUTION

Coreference resolution is the challenge of grouping noun phrases (NP), also known as mentions, into sets that belong to the same speech person. In text, character occurrences take three forms: proper nouns, nominals, and pronouns. Unifying these occurrences can be considered as a special case of the coreference resolution (CR) problem, called *alias resolution.*

Anaphora resolution (AR) is the most distinct case from coreference resolution (CR). Anaphora arises in a text when one element refers to another, defining the meaning of the second. These mentions do not refer to each other, but rather have an anaphoric relationship. Despite the fact that anaphora resolution differs from coreference resolution, in the vast majority of situations, the two are equivalent.

Here is an overview of different types of references (see [2] for a deep overview):

2.6.1. ANAPHORA AND CATAPHORA

These two are almost exactly the same. The key distinction is that anaphora occurs after the term to which it refers, while cataphora occurs before it. The word that comes before an anaphora is known as an antecedent, and the word that comes after a cataphora is known as a postcedent.

2.6.2. COREFERRING NOUN PHRASES

An anaphoric condition in which the second noun phrase refers to an earlier descriptive form of an expression.

3. APPROACH

The entire pipeline is written in Python, and the implemented free open-source NLP libraries are described in greater detail throughout this section.

3.1. DATASET

The dataset consists of the three books that comprise the Foundation Trilogy. They are available online and can be freely downloaded at:

`https://archive.org/details/AsimovTheFoundation_201705`

3.2. PREPROCESS

Until relevant information can be extracted, the text must be preprocessed.

The Foundation Trilogy has 1240496 characters. The text is first tokenized into paragraphs. Text spans with only upper case characters (e.g. 'FIRST INTERLUDE') or with less than 5 characters (e.g. '2 . ') are excluded and extra blank spaces between sentences are removed. This is achieved with the *re* package (Regular expression operations). A list of 7225 paragraphs is obtained.

The following preprocessing steps are done with *SpaCy*, a Python-based open-source library for Natural Language Processing. It provides a trained English pipeline optimized for CPU called *en_core_web_sm* [3]. The pretrained pipeline is the following:

1 A tok2Vec (token to vector) subnetwork to map tokens into vector representations.

2 A tagger model responsible for the part-of-speech (POS) tagging step. To predict the tag probabilities given the token vectors, the tagger model simply adds a linear layer with softmax activation.

3 A transition-based parser model [4]. Transition-based parsing is an approach to structured prediction where predicting the structure is delegated to a sequence of state transitions.

4 An atribute ruler that can be used to match patterns in text.

5 A lemmatizer that assigns base forms to tokens using rules based on part-of-speech tags, or lookup tables.

6 A pretrained NER agent that has been disabled for this project, because it has been trained with ordinary text and consequently cannot correctly identify and classify fiction named entities.

After this step, every token is labeled with a POS tag, a dependency tag and a lemmatized version of the token as shown in figure 2. 7225 paragraphs are preprocessed in 2 minutes.

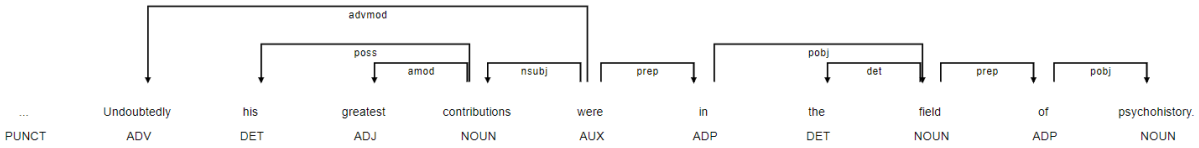

Figure 2: Parsed sentence after preprocessing

3.3. NAMED ENTITY RECOGNITION (NER)

Tokens labeled with the 'NPROP' POS tag (proper noun) [5] are considered entity candidates. This means that the POS tagging pretrained agent will propagate an error to the NER stage. This is analysed in the Results section. Candidates are classified as 'PER' (person) or 'LOC' (location) when certain conditions are satisfied. 7225 paragraphs are preprocessed in 5 minutes.

3.3.1. PERSON RULES

The 'PER' label is given to a token when one of the following conditions is satisfied:

1 The candidate is the subject of a 'person verb' included in figure 3. This means that is labeled with the 'nsubj' tag, must have a 'head' (root) labeled as 'VERB', and the verb must be a 'person verb'. Verbs are compared after being lemmatized with the NLTK lemmatizer. See example 5.

2 The candidate token is preceded by a 'honorific word' included in figure 4. See example 7.

3 The candidate is followed by the word 'who' or by ', who'. See example 6.

857 'PER' entities are found using rule 1, 229 using rule 2 and 58 with rule 3. A total of 1144 candidates are classified as 'PER' in 7225 paragraphs.

```
'said', 'sniffed', 'met', 'greet', 'walked', 'respond', 'talk',
'think', 'hear', 'wait', 'pause', 'write', 'smile', 'answer', 'wonder',
'reply', 'read', 'sit', 'muttered', 'fumble', 'ask', 'sigh', 'frowned',
'cry', 'chuckled', 'murmured', 'find', 'shoved', 'fastened'
```

Figure 3: Person verbs

```
'Dr.', 'Prof.', 'Mr.', 'Ms.', 'Msr.', 'Jr.', 'Sr.', 'Lord', 'Sir',
'Professor', 'Doctor', 'King', 'Commdor', 'Lady', 'Captain', 'Colonel',
'Miss', 'General', 'Mayor'
```

Figure 4: Honorific words

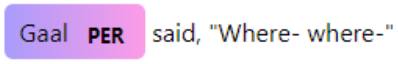

Figure 5: 'Gaal' labeled as 'PER' by using the 'person verbs' rule

3.3.2. LOCATION RULES

Locations have been found with the help of the SpaCy *Matcher*, which detects word spans under given patterns. A candidate is labeled with the 'LOC' tag if any of the two following patterns is matched:

1 See pattern 1 in figure 10 where 'NOUN' must be a word included in list 8. An example is shown in figure 11. The words are compared in lowercase.

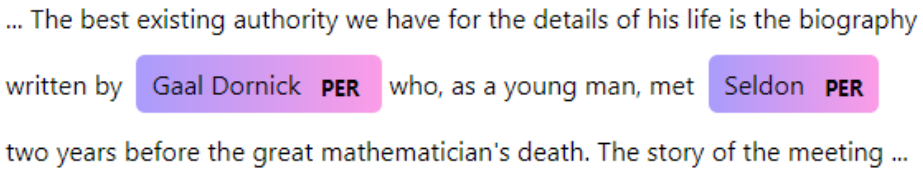

Figure 6: 'Gaal Dornick' found using the 'who' rule

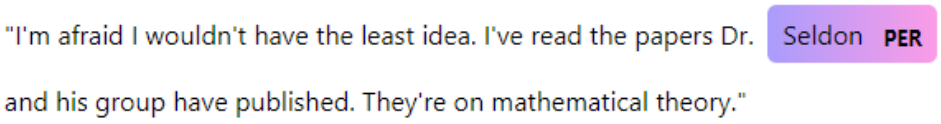

Figure 7: 'Seldon' found using the 'Dr.' honorific word

2 See pattern 2 in figure 10 where 'VERB' must be a word included in list 9. Verbs are compared after being lemmatized with the NLTK lemmatizer. An example is shown in figure 12.

18 'LOC' entities are found using pattern 1 and 35 with pattern 2. A total of 53 candidates are classified as 'LOC' in 7225 paragraphs.

```
    'planet', 'kingdom', 'world', 'region', 'location', 'republic',
'street', 'neighborhood', 'realm', 'sight'
```

Figure 8: Location names

```
    'go', 'travel', 'move', 'exiled', 'come'
```

Figure 9: 'LOC' associated verbs

```
Pattern 1: {'POS': 'NOUN'}, {'LOWER': 'of'}, {'POS': 'PROPN'}
Pattern 2: {'POS': 'VERB'}, {'LOWER': 'to'}, {'POS': 'PROPN'}
```

Figure 10: 'LOC' NER pattern rules

3.3.3. Unclassified candidates

If a candidate does not satisfy any 'PER' or 'LOC' rule, it is stored in a 'unclassified candidates' list which is used at the end of the NER process to label all the tokens whose name is the same as any of the already classified entities. Figure 13 shows the example of the entity 'Gaal' correctly labeled as 'PER' but initially unclassified.

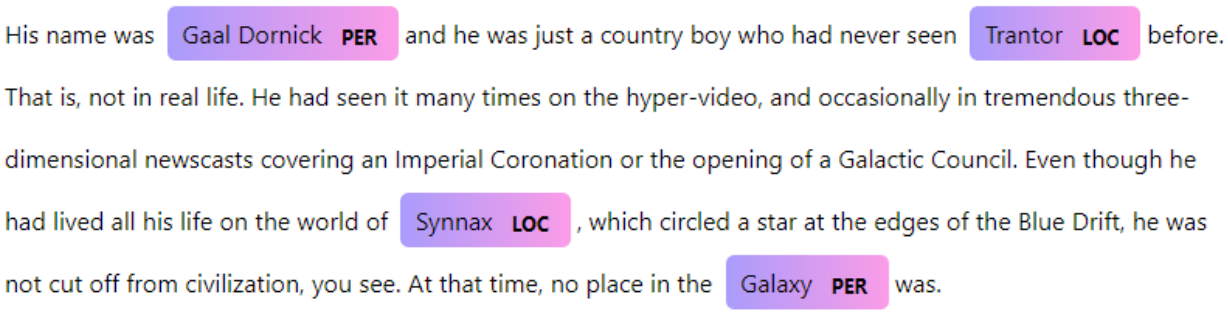

Figure 11: 'Synnax' labeled as 'LOC' by using pattern 1 (see 10)

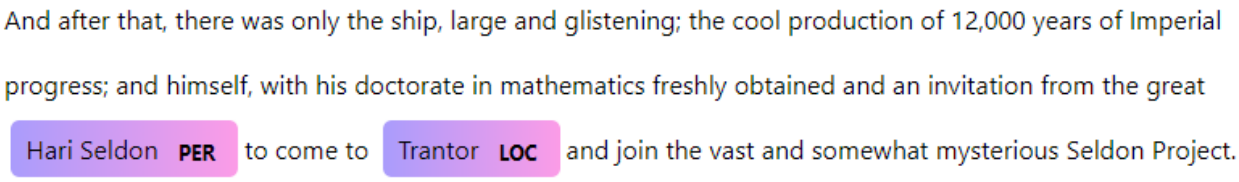

Figure 12: 'Trantor' labeled as 'LOC' by using pattern 2 (see 10)

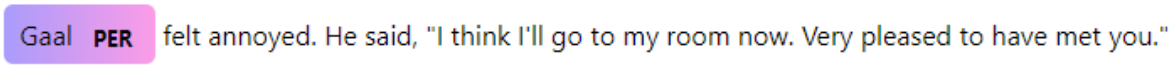

Figure 13: 'Gaal' correctly labeled as 'PER' but initially unclassified

3.3.4. NER TAGGING

NER model returns a list of tags corresponding to each parsed token of each paragraph. The beggining of a 'PER' span is labeled as 'B-PER' and the end as 'I-PER'. The same is done with locations where the start is labeled as 'B-LOC' and the end as 'I-LOC' (see example 14).

3.4. NER QUALITY

Given 50 samples, two annotators classify each token into 5 different categories: 'B-PER' (start of person entity), 'I-PER' (continuation of person entity), 'B-LOC' (start location), 'I-LOC' (continuation of location name) and 'O' (nothing). These annotations are then compared with the NER predictions. The full Dataset can be found in appendix A as well as the two manual annotations and the NER predictions. Appendix B contains the raw printed results from the validation.

Since the Dataset is composed of 50 paragraphs with a total of 1869 tokens, only tokens without the 'O' label or differently labeled in between annotators are considered. This means that, for instance, a token predicted as 'O' but manually labeled as 'B-LOC' by one of the

```
    Parsed Sentence:
    'What Gaal was waiting for after the disappointment of the Jump was
that first sight of Trantor.'

    NER output:
    'O', 'B-PER', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O',
'O', 'O', 'O', 'B-LOC', 'O'

    Paired word-tag:
    (What, 'O'), (Gaal, 'B-PER'), (was, 'O'), (waiting, 'O'), (for, 'O'),
(after, 'O'), (the, 'O'), (disappointment, 'O'), (of, 'O'), (the, 'O'),
(Jump, 'O'), (was, 'O'), (that, 'O'), (first, 'O'), (sight, 'O'), (of,
'O'), (Trantor, 'B-LOC'), (., 'O')
```

Figure 14: NER tagging example

annotators is taken into consideration (and viceversa). This reduces the number of compared annotations to 83.

By means of the SKLearn.metrics package the three vectors are evaluated in Precision, Recall, F1-Score, Confusion matrix and Kappa distance.

3.5. Coreference Resolution (CR) and Entity connections

Coreferences are solved using the pretrained coreference resolution agent Neuralcoref which can easily be coupled into the *en_core_web_sm* pipeline. Coreferences are solved in order to find events between two 'PER' or 'LOC' entities. An event between two entities appears in the form of 'nsubj' + 'VERB + 'dobj', where 'nsubj' (subject) and 'dobj' (direct object) must be any of an entity's references. Figure 15 shows an example.

"Other things more important than a message from Hari Seldon PER ? I think not." Fara PER was growing more pontifical than ever, and Hardin PER eyed him thoughtfully. What was he getting at?

```
CR cluster: Fara: ['Fara', 'him']
Event found: ['Fara', 'eyed', 'Hardin']
```

Figure 15: Event found with coreference resolution

The problem with this approach is that only a few entity interactions follow this pattern. Specifically, only 60 events can be found (see the complete list in Appendix C).

A new linking approach has been taken instead: two entities ('PER' or 'LOC') are consid-

ered to have an interaction when they appear within a 2 paragraphs range during the text (appendix C).

Three different lists are obtained from this stage: 'PER'-'PER' links, 'LOC'-'PER' links and events (see figure 16).

```
'PER'-'PER' link: ['Ebling Mis', 'Bayta', 40]
'LOC'-'PER' link: ['Trantor', 'Hari Seldon', 7217]
event: ['Turbor', 'followed', 'Semic', 6992]
```

Figure 16: Entity links examples, where 40, 7217 or 6992 is the sentence number

3.6. NORMALIZATION

Entity names are normalized before visualization by using the hybrid similarity measure Monge-Elkan [6]. The input bags are split into tokens and pairwise measured. Matches are given by the Jaro measure, which is ideal to compare short strings, such as first and last names. The best matching tokens are compared to get the monge-elkan score.

A threshold of 0.81 gives the best results (figure 17). In addition, entities not belonging to any cluster and sharing the same surname are unified (figure 18). There are 110 entities before normalization and 89 after normalization (see appendix D).

```
['Salvor Hardin', 'Hardin'],
['Ebling Mis', 'Mis'],
['Fara', 'Far Star'],
['Channis', 'Bail Channis'],
['Korell', 'Forell'],
['Flan Pritcher', 'Han Pritcher'],
['Foundation', 'Second Foundation']
```

Figure 17: String normalization with Monge-Elkan

```
['Riose', 'Bel Riose'],
['Jorane Sutt', 'Sutt'],
['Speaker', 'First Speaker'],
['Anthor', 'Pelleas Anthor'],
['Seldon', 'Hari Seldon'],
['Hober Mallow', 'Mallow']
```

Figure 18: String normalization by surname

3.7. VISUALIZATION

Two interactive renders are used to visualize entity links. A character Network is used to visualize 'PER'-'PER' links and a Geo-mapping render for the 'LOC'-'PER' links. (see Results Section).

3.7.1. PER-PER CHARACTER NETWORK

The input of this stage is a list of triples in the form of (entity1, entity2, weight). To visualise the connections between person named entities (NE) a Character Network is drawn with the help of two open-source libraries: *Networkx* and *Bokeh* [7].

3.7.2. PER-LOC GEO-MAPPING

The input is a list of ('LOC' name, 'PER' name, num sentence) triples which is converted to a list of ('LOC','PER') unique tuples. This is converted to a list of 'PER' entities per 'LOC' entity that is added as a hover to its corresponding geographically predefined plotted scatters in a static image (a Galaxy image).

4. RESULTS

During the following sections the NER quality (3.4) results are evaluated and the interactive renders (3.7) shown. Precision, Recall, F1-Score, Confusion matrix and Kappa distances are shown in figures 22, 23 and 24 for Annotator 1 vs Predicted, Annotator 2 vs Predicted and Annotator 1 vs Annotator 2 respectively. Figure 21 shows Kappa distances between the classifier and the annotators. Figures 26 and 27 show the visualizations.

4.1. NER QUALITY AND POS TAG ERROR

In order to understand the quality results, the discordance between the classifier and the annotators is analysed in the following examples.

Figure 19 shows a propagated SpaCy POS tagging error, where the tagged sentence is extracted from the Validation Dataset (see appendix A). Both annotators labeled the token 'Suit' as 'B-PER', but it was predicted as 'O'. The problem is that the pretrained POS tagging agent classified the token as 'NOUN' instead of 'PROPN' and thus, it was discarded by the NER model. One possible answer to the logic behind the POS tag decision is that the token 'Suit' has a first Upper case character because it is the first word of the sentence. In addition, the word 'suit' can refer to the 'outfit' word.

In the sample sentence "And now, my boy, what if Smyrno decides to attack the Foundation for its own part and thus gains all that power?", the entity "Smyrno" has been classified as 'LOC' by both annotators but as 'O' by the NER model. Both annotators have read the book and know that Smyrno is a planet (could be considered as Organization 'ORG' in this short context), but the classifier didn't find any occurrence matching any of the predefined 'LOC' rules in 10. There is only one 'Smyrno' ocurrence in the book that could be matched by the NER rules. This example is shown in figure 20, where a sequence of errors ocurred: The POS

```
Sentence:
Suit said cynically, "Very nicely put. So, to get back...

POS tag:
Suit => NOUN => NN => noun
said => VERB => VBD => verb
cynically => ADV => RB => adverb
```

Figure 19: POS tag error example

agent labeled the token 'Kingdom' as 'PROPN', so it was not matched in pattern 1 from 10 and eventually, none of the 'Smyrno' ocurrences was predicted as 'LOC'. The pretrained POS tagger considered the word 'Kingdom' to be an entity because the first character is uppercase.

```
Sentence:
...his own exploits as battalion head during the recent war between
Anacreon and the neighboring newly proclaimed Kingdom of Smyrno.

POS tag:
Kingdom => PROPN => NNP => proper noun
of => ADP => IN => adposition
Smyrno => PROPN => NNP => proper noun
```

Figure 20: NER Rules error example

Another example of the Rule-Based NER model limitations is shown in the sentence "The stars were thick here near the Galactic center.", where "Galactic center" was labeled as 'B-LOC' 'I-LOC' by Annotator 2 but as 'B-PER' 'O' by the classifier, since the word 'center' was not POS tagged as 'PROPN' but as 'NOUN' and eventually was not considered as a NER candidate.

The sentence "What can Setdon do about it?" shows a typing error from the book where "Setdon" should be "Seldon". Since this is probably the only ocurrence for the candidate 'Setdon', it is discarded and labeled as 'O' by the NER model because it never matches any rule. Annotator 1 and 2 labeled this token as 'B-PER'.

Although the NER model evidences some weaknesses due to its rule limitations, it proved to be robust on entity recognition and classification with the Foundation Trilogy as shown in tables 22, 23, 24 and 21. This can be seen as well in figures 26 and 27 (see the Github code source at page 1 to properly interact with the Network graph and the Galaxy map).

4.2. Interactive Renders

Two different visualizers are rendered: A network graph and a map. The network graph 26 has been exported with Gephi, but another version created with Networkx and Bokeh can be found in the source code (see the Readme). In the network graph, the size of the nodes and

	Predicted	Annotator 1	Annotator 2
Predicted	1	0.953	0.868
Annotator 1	0.953	1	0.943
Annotator 2	0.868	0.943	1

Figure 21: Kappa distances

	precision	recall	f1-score	nº
B-PER	0.88	0.72	0.79	50
I-PER	0.86	0.55	0.67	11
B-LOC	1.00	0.64	0.78	11
I-LOC	0	0	0	0
micro avg	0.89	0.68	0.77	72
macro avg	0.68	0.48	0.56	72
weighted avg	0.89	0.68	0.77	72

		True			
		B-PER	I-PER	B-LOC	I-LOC
Pred	B-PER	36	0	0	0
	I-PER	0	6	0	0
	B-LOC	1	0	7	0
	I-LOC	0	0	0	0

Figure 22: NER predictions vs Annotator 1

	precision	recall	f1-score	nº
B-PER	0.83	0.65	0.73	52
I-PER	0.86	0.43	0.57	14
B-LOC	1.00	0.64	0.78	11
I-LOC	0	0	0	1
micro avg	0.85	0.60	0.71	78
macro avg	0.67	0.43	0.52	78
weighted avg	0.85	0.60	0.70	78

		True			
		B-PER	I-PER	B-LOC	I-LOC
Pred	B-PER	34	0	0	0
	I-PER	2	6	0	0
	B-LOC	1	0	7	0
	I-LOC	0	0	0	0

Figure 23: NER predictions vs Annotator 2

the labels is based on the number of interactions with other entities/nodes. Figure 25 shows the centrality degree of the top 10 entities. It is based on the fraction of nodes it is connected

	precision	recall	f1-score	nº
B-PER	0.96	0.92	0.94	52
I-PER	1	0.79	0.88	14
B-LOC	1	1	1	11
I-LOC	0	0	0	1
micro avg	0.97	0.9	0.93	78
macro avg	0.74	0.68	0.71	78
weighted avg	0.96	0.90	0.93	78

		True			
		B-PER	I-PER	B-LOC	I-LOC
Pred	B-PER	48	0	0	0
	I-PER	2	11	0	0
	B-LOC	0	0	11	0
	I-LOC	0	0	0	0

Figure 24: Annotator 1 vs Annotator 2

to. Communities are displayed by colors, showing the entities that interacted most between them.

```
('Foundation', 0.6091954022988506),
 ('Seldon', 0.3563218390804598),
 ('Mule', 0.3448275862068966),
 ('Hardin', 0.25287356321839083),
 ('Fie', 0.24137931034482757),
 ('Emperor', 0.1839080459770115),
 ('Second Empire', 0.1724137931034483),
 ('Arcadia', 0.1724137931034483),
 ('Toran', 0.16091954022988506),
 ('Darell', 0.16091954022988506)]
```

Figure 25: Network centrality degree from top 10 nodes

5. FUTURE RESEARCH

A machine Learning agent can be trained to perform NER and CR on the Foundation Trilogy, which would probably improve the linking performance. Furthermore, more advanced visualizations could be plotted such as an Event Timeline.

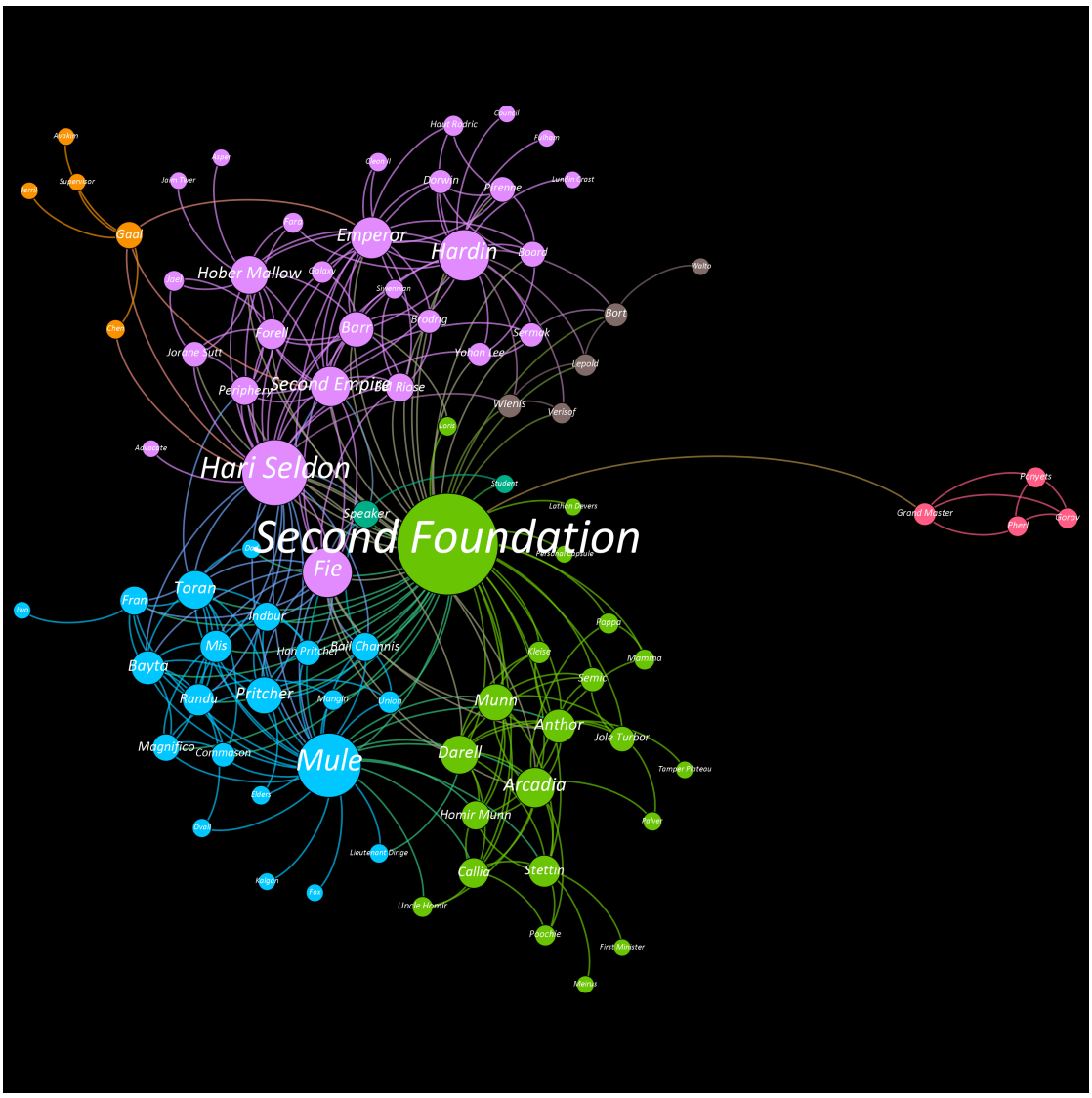

Figure 26: Network graph The Foundation Trilogy

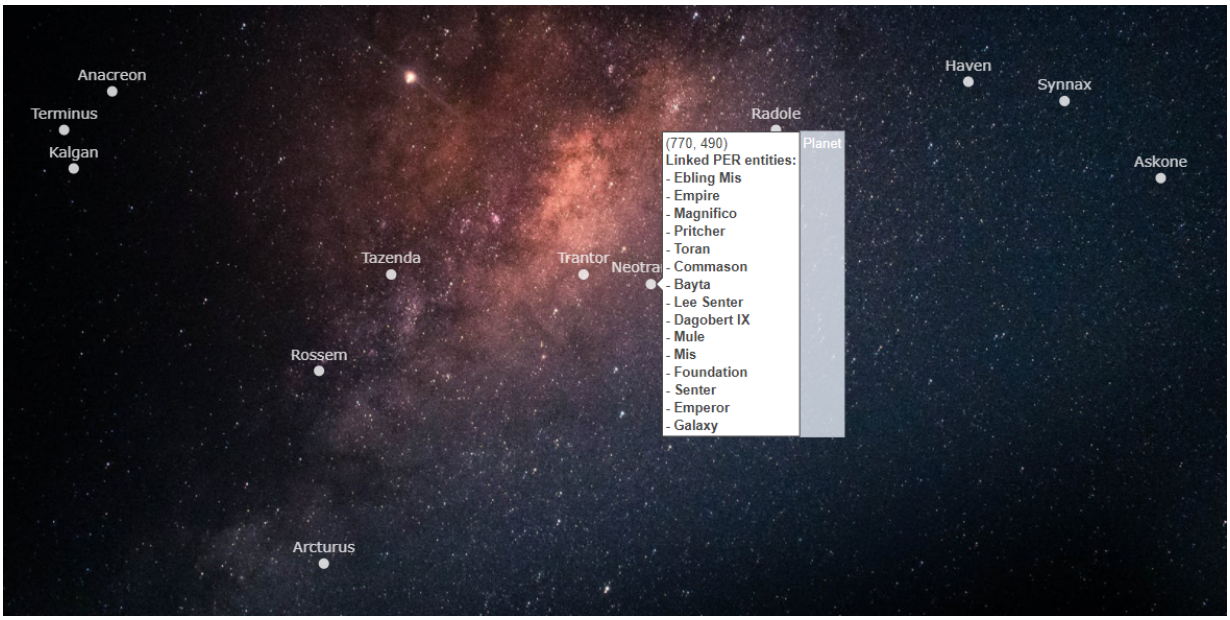

Figure 27: Geo-mapping The Foundation Trilogy

REFERENCES

[1] V. Labatut and X. Bost, "Extraction and Analysis of Fictional Character Networks: A Survey," *ACM Computing Surveys*, vol. 52, no. 5, 2019. [Online]. Available: https://dl.acm.org/doi/10.1145/3344548

[2] R. Sukthanker, S. Poria, E. Cambria, and R. Thirunavukarasu, "Anaphora and coreference resolution: A review," 2018. [Online]. Available: https://arxiv.org/pdf/1805.11824.pdf

[3] "Spacy trained pipelines." [Online]. Available: https://spacy.io/models/en#en_core_web_sm

[4] M. Honnibal, "Parsing english in 500 lines of python," 2013. [Online]. Available: https://explosion.ai/blog/parsing-english-in-python

[5] "Universal pos tags." [Online]. Available: https://universaldependencies.org/docs/u/pos/

[6] F. Naumann, "Similarity measures." [Online]. Available: https://hpi.de/fileadmin/user_upload/fachgebiete/naumann/folien/SS13/DPDC/DPDC_12_Similarity.pdf

[7] M. Walsh, "Make an interactive network visualization," 2021. [Online]. Available: https://melaniewalsh.github.io/Intro-Cultural-Analytics/Network-Analysis/Making-Network-Viz-with-Bokeh.html

A. APPENDIX

0) "You see, you're beginning to understand. And now, my boy, what if Smyrno decides to attack the Foundation for its own part and thus gains all that power? How long do you suppose we could escape becoming a vassal power? How long would you hold your throne?"

1) Gorov's smile was sardonic, "And you've been to theological school as well. You're all right, Ponyets. I'm glad they sent you. But the Grand Master doesn't love my soul exclusively. Has he mentioned a ransom?"

2) He proceeded. As each item was mentioned, new symbols sprang to life at his touch, and melted into the basic function which expanded and changed.

3) "What is it?"

4) The first sight of Trantor's sun was that of a hard, white speck all but lost in a myriad such, and recognizable only because it was pointed out by the ship's guide. The stars were thick here near the Galactic center. But with each Jump, it shone more brightly, drowning out the rest, paling them and thinning them out.

5) Gaal had followed after, clutching at the sleeve of the white uniform with the Spaceship-and-Sun of the Empire on it.

6) When Hardin denied owning the Journal, he was perhaps technically correct, but no more. Hardin had been the leading spirit in the drive to incorporate Terminus into an autonomous municipality-he had been elected its first mayor-so it was not surprising that, though not a single share of Journal stock was in his name, some sixty percent was controlled by him in more devious fashions.

7) So it also happened, that when Ebling Mis decided to allow Indbur to honor him with an audience, he did not wait for the usual rigid line of command to pass his request up and the favored reply down, but, having thrown the less disreputable of his two formal jackets over his shoulders and pounded an odd hat of impossible design on one side of his head, and lit a forbidden cigar into the bargain, he barged past two ineffectually bleating guards and into the mayor's palace.

8) The officer called after him, "Trantor would only be gray blur anyway, Kid. Why don't you take a space-tour once you hit Trantor. They're cheap."

9) "But is that why I'm being investigated?"

10) "Why? What would he do in case they were?"

11) Q. (theatrically) Do you realize, Dr. Seldon, that you are speaking of an Empire that has stood for twelve thousand years, through all the vicissitudes of the generations, and which has behind it the good wishes and love of a quadrillion human beings?

12) "It may. But the Board of Trustees does not. I am the Emperor's representative on Terminus, Hardin, and have full powers in this respect."

13) "The thanks of a weak one are of but little value," he muttered, "but you have them, for truly, in this past week, little but scraps have come my way - and for all my body is small, yet is my appetite unseemly great."

14) Added Sutt: "He does not realize, moreover, that building armaments would mean withdrawing men - valuable men - from the Encyclopedia. That cannot be done, come what may."

15) Said Lundin Crast, his long nose wrinkling angrily: "If you're proposing the militarization of the Foundation, I won't hear a word of it. It would mark our open entrance into the field of politics. We, Mr. Mayor, are a scientific foundation and nothing else."

16) The inquisitor said, "Has anyone told you of such destruction; set a date?" And, as the young man hesitated, he went on, "You have been followed, doctor. We were at the airport when you arrived; on the observation tower when you waited for your appointment; and, of course, we were able to overhear your conversation with Dr. Seldon."

17) Mallow nodded, "Of course. They'll declare war, eventually, though I'm betting it'll take another pair of years."

18) Hardin flicked open the curiously carved silver lid of the cigar box that had once belonged to Jord Fara of the old Board of Trustees in the long-dead days of the Encyclopedists. It was a genuine Empire product from Santanni, though the cigars it now contained were home-grown. One by one, with grave solemnity, the four of the deputation accepted cigars and lit up in ritualistic fashion.

19) "That, neither."

20) "Wait. I have a right to a lawyer. I demand my rights as an Imperial citizen."

21) "You shall have them."

22) Ducem Barr's nostrils flared, and the tendons of his old right hand jerked; but he said nothing.

23) Gaal was not certain whether the sun shone, or, for that matter, whether it was day or night. He was ashamed to ask. All the planet seemed to live beneath metal. The meal of which he had just partaken

had been labelled luncheon, but there were many planets which lived a standard timescale that took no account of the perhaps inconvenient alternation of day and night. The rate of planetary turnings differed, and he did not know that of Trantor.

24) Suit said cynically, "Very nicely put. So, to get back to the original point of discussion, what are your terms? What do you require to exchange your ideas for mine?"

25) "That in a month we celebrate our fiftieth anniversary." Fara had a trick of uttering the most obvious platitudes with great profundity.

26) "Forget that. That's an old story." His eyes traveled round the circle. "I'm referring to the people.

27) Linge Chen said, "I will speak," and the other Commissioners sat back in their chairs, prepared to listen. A silence formed about Chen into which he might drop his words.

28) Hardin answered, half in reverie: "Yes, I never completed my studies, though. I got tired of theory. I wanted to be a psychological engineer, but we lacked the facilities, so I did the next best thing - I went into politics. It's practically the same thing."

29) He drew forth a chair, dusted it carefully with the iridescent square of fabric attached to the top of his white stick, and seated himself. Devers glanced towards the mate to the chair, but Brodrig said lazily, "You will stand in the presence of a Peer of the Realm."

30) Gaal closed his mouth; he had been gaping; and said, "It certainly seems so." He started for them automatically, then stopped.

31) The old patrician retreated noiselessly with a slow bow that was part of the ceremonious legacy left by the aristocracy of the last century's better days.

32) The trader frowned hopelessly. "I don't know. It can't really work like that; not just like magic. Psychohistory or not, they're terribly strong, and we're weak. What can Setdon do about it?"

33) And what would that mean?

34) "Tell me," he said, "will your only activity be that of preparing this encyclopedia you speak of?" "It will."

35) "That so? Recently?"

36) "I am instructed that this probability is 77.2%."

37) "He's a queer fellow," said Bayta, with amusement, and Toran agreed indifferently. The clown was close enough now to be seen clearly. His thin face drew together in front into a nose of generous planes and fleshy tip that seemed all but prehensile. His long, lean limbs and spidery body, accentuated by his costume, moved easily and with grace, but with just a suggestion of having been thrown together at random.

38) Seldon thought a moment, and the last minute began to die. Fie said, "I accept exile."

39) "No exactly. At least, I've always wanted to visit Trantor but I came here primarily for a job." "Oh?"

40) "And even if by some chance the Mule did not establish a dynasty, he would still establish a distorted new Empire upheld by his personal power only. It would die with his death; the Galaxy would be left where it was before he came, except that there would no longer be Foundations around which a real and healthy Second Empire could coalesce. It would mean thousands of years of barbarism. It would mean no end in sight."

41) "Uh-huh," grunted Hardin dryly.

42) Jerril waved his arm indifferently in farewell.

43) "Injudicious, yes," said the Askonian, curtly. "So much so, that your comrade is likely to lose life in payment."

44) Haut Rodric bounced it and stared. "What is it? Steel?"

45) "But could you have arranged-"

46) "Y'know, the brain gets calcified when you get as old as I am. What are you trying to do?"

47) "Violence is the last refuge-"

48) Mallow continued:

49) Arcadia looked across at the kind gray eyes of the woman and felt her lips quivering. One part of her brain was telling her that here were people from Trantor, with whom she could go, who could help her remain on that planet until she could decide what next to do, where next to go. And another part of her brain, much the louder, was telling her in jumbled incoherence that she did not remember her mother, that she was weary to death of fighting the universe, that she wanted only to curl into a little hall with strong, gentle arms about her, that if her mother had lived, she might ... she might-

y_pred = ['O', 'B-PER', 'B-PER', 'B-PER', 'B-PER', 'I-PER', 'B-LOC', 'O', 'O', 'B-PER', 'B-PER', 'B-PER', 'O', 'B-LOC', 'B-PER', 'I-PER', 'B-PER', 'O', 'O', 'B-LOC', 'O', 'B-PER', 'B-PER', 'B-PER',

```
'O', 'O', 'B-LOC', 'B-PER', 'O', 'B-PER', 'I-PER', 'B-PER', 'O', 'O', 'O', 'O', 'O', 'O', 'B-PER', 'B-PER',
'B-PER', 'O', 'O', 'B-PER', 'O', 'O', 'O', 'B-PER', 'I-PER', 'B-PER', 'B-LOC', 'O', 'B-PER', 'B-PER',
'I-PER', 'B-PER', 'B-PER', 'B-PER', 'O', 'B-PER', 'B-PER', 'O', 'O', 'O', 'B-PER', 'B-PER', 'O', 'B-PER',
'B-PER', 'B-LOC', 'B-PER', 'B-PER', 'B-PER', 'B-PER', 'I-PER', 'O', 'B-PER', 'B-PER', 'B-PER', 'I-PER',
'B-PER', 'B-PER', 'B-LOC']

    y_true1 = ['B-LOC', 'B-LOC', 'B-PER', 'B-PER', 'B-PER', 'I-PER', 'B-LOC', 'B-LOC', 'O', 'B-PER',
'B-PER', 'B-PER', 'B-PER', 'B-LOC', 'B-PER', 'I-PER', 'B-PER', 'B-LOC', 'O', 'B-LOC', 'O', 'B-PER',
'O', 'B-PER', 'I-PER', 'I-PER', 'B-LOC', 'B-PER', 'B-PER', 'B-PER', 'I-PER', 'B-PER', 'B-PER', 'I-PER',
'B-PER', 'I-PER', 'B-PER', 'O', 'B-PER', 'B-PER', 'B-PER', 'B-PER', 'I-PER', 'B-PER', 'B-PER', 'O',
'O', 'B-PER', 'I-PER', 'B-PER', 'B-LOC', 'B-PER', 'B-PER', 'B-PER', 'I-PER', 'B-PER', 'B-PER', 'B-PER',
'B-PER', 'B-PER', 'B-PER', 'B-PER', 'B-PER', 'B-PER', 'B-PER', 'B-PER', 'B-PER', 'B-PER', 'B-PER', 'B-LOC',
'B-PER', 'O', 'O', 'O', 'O', 'B-PER', 'B-PER', 'B-PER', 'B-PER', 'I-PER', 'B-PER', 'B-PER', 'B-LOC']

    y_true2 = ['B-LOC', 'B-LOC', 'B-PER', 'B-PER', 'B-PER', 'I-PER', 'B-LOC', 'B-LOC', 'I-LOC', 'B-PER',
'B-PER', 'B-PER', 'B-PER', 'B-LOC', 'B-PER', 'I-PER', 'B-PER', 'B-LOC', 'B-PER', 'B-LOC', 'B-PER', 'I-PER',
'O', 'B-PER', 'I-PER', 'I-PER', 'B-LOC', 'B-PER', 'B-PER', 'B-PER', 'I-PER', 'B-PER', 'B-PER', 'I-PER',
'B-PER', 'I-PER', 'B-PER', 'B-PER', 'I-PER', 'B-PER', 'B-PER', 'B-PER', 'I-PER', 'B-PER', 'B-PER', 'B-PER',
'I-PER', 'B-PER', 'I-PER', 'B-PER', 'B-LOC', 'B-PER', 'B-PER', 'B-PER', 'I-PER', 'B-PER', 'B-PER', 'B-PER',
'B-PER', 'B-PER', 'B-PER', 'B-PER', 'B-PER', 'B-PER', 'B-PER', 'B-PER', 'B-PER', 'B-PER', 'B-PER', 'B-LOC',
'B-PER', 'O', 'O', 'O', 'O', 'B-PER', 'B-PER', 'B-PER', 'B-PER', 'I-PER', 'B-PER', 'B-PER', 'B-LOC']
```

B. APPENDIX

```
Annotator 1 vs predicted:
        [[36  0  0  0]
         [ 0  6  0  0]
         [ 1  0  7  0]
         [ 0  0  0  0]]
                     precision    recall  f1-score   support

              B-PER       0.88      0.72      0.79        50
              I-PER       0.86      0.55      0.67        11
              B-LOC       1.00      0.64      0.78        11
              I-LOC       0.00      0.00      0.00         0

          micro avg       0.89      0.68      0.77        72
          macro avg       0.68      0.48      0.56        72
       weighted avg       0.89      0.68      0.77        72

       Kappa: 0.9535315985130112
Annotator 2 vs predicted:
        [[34  0  0  0]
         [ 2  6  0  0]
         [ 1  0  7  0]
         [ 0  0  0  0]]
                     precision    recall  f1-score   support

              B-PER       0.83      0.65      0.73        52
              I-PER       0.86      0.43      0.57        14
              B-LOC       1.00      0.64      0.78        11
              I-LOC       0.00      0.00      0.00         1

          micro avg       0.85      0.60      0.71        78
          macro avg       0.67      0.43      0.52        78
```

```
        weighted avg       0.85      0.60      0.70        78

        Kappa: 0.8681898066783831
Annotator 1 vs Annotator 2:
        [[48  0  0  0]
         [ 2 11  0  0]
         [ 0  0 11  0]
         [ 0  0  0  0]]
                     precision    recall  f1-score   support

              B-PER       0.96      0.92      0.94        52
              I-PER       1.00      0.79      0.88        14
              B-LOC       1.00      1.00      1.00        11
              I-LOC       0.00      0.00      0.00         1

           micro avg       0.97      0.90      0.93        78
           macro avg       0.74      0.68      0.71        78
        weighted avg       0.96      0.90      0.93        78

        Kappa: 0.9428571428571428
```

C. APPENDIX

'PER' entity links before normalization (ent1,ent2, number of interactions):

```
        [['Mule', 'Foundation', 150], ['Bayta', 'Toran', 129], ['Galaxy', 'Foundation', 124], ['Mule',
'Second Foundation', 105], ['Seldon', 'Foundation', 94], ['Magnifico', 'Bayta', 85], ['Darell', 'Anthor',
83], ['Channis', 'Mule', 80], ['Mule', 'Pritcher', 75], ['Mule', 'Toran', 71], ['Pritcher', 'Channis',
70], ['Foundation', 'Empire', 69], ['Mule', 'Bayta', 69], ['Mule', 'Galaxy', 68], ['Toran', 'Magnifico',
66], ['Empire', 'Galaxy', 65], ['Galaxy', 'Seldon', 61], ['Galaxy', 'Second Foundation', 59], ['Channis',
'Second Foundation', 59], ['Foundation', 'Bayta', 53], ['Mule', 'Seldon', 52], ['Hari Seldon', 'Foundation',
51], ['Foundation', 'Mallow', 49], ['Hardin', 'Wienis', 48], ['Mule', 'Magnifico', 48], ['Arcadia',
'Darell', 48], ['Pirenne', 'Hardin', 46], ['Seldon', 'Gaal', 45], ['Hardin', 'Foundation', 44], ['Seldon',
'Second Foundation', 43], ['Galaxy', 'Hari Seldon', 42], ['Foundation', 'Toran', 42], ['Munn', 'Darell',
42], ['Seldon', 'Empire', 40], ['Randu', 'Mule', 40], ['Ebling Mis', 'Bayta', 40], ['Ebling Mis', 'Mule',
40], ['Seldon', 'Hari Seldon', 39], ['Mallow', 'Sutt', 38], ['Fie', 'Mule', 38], ['Foundation', 'Second
Foundation', 38], ['Anthor', 'Second Foundation', 37], ['Pritcher', 'Second Foundation', 36], ['Jael',
'Mallow', 35], ['Fie', 'Foundation', 35], ['Ebling Mis', 'Magnifico', 34], ['Randu', 'Foundation', 33],
['Toran', 'Ebling Mis', 33], ['Munn', 'Anthor', 33], ['Sermak', 'Hardin', 31], ['Barr', 'Emperor', 31],
['Brodrig', 'Emperor', 31], ['Mis', 'Mule', 31], ['Arcadia', 'Callia', 31], ['Arcadia', 'Pappa', 31],
['Lee', 'Hardin', 30], ['Lepold', 'Wienis', 30], ['Riose', 'Barr', 30], ['Darell', 'Second Foundation',
30], ['Board', 'Hardin', 29], ['Wienis', 'Foundation', 29], ['Fie', 'Second Foundation', 29], ['Ebling
Mis', 'Foundation', 28], ['Mamma', 'Pappa', 28], ['Pritcher', 'Foundation', 27], ['Darell', 'Turbor',
27], ['Dorwin', 'Hardin', 26], ['Stettin', 'Foundation', 26], ['Emperor', 'Empire', 25], ['Hober Mallow',
'Foundation', 25], ['Mis', 'Bayta', 25], ['Pritcher', 'Bayta', 25], ['Arcadia', 'Foundation', 25], ['Foundation',
'Salvor Hardin', 24], ['Korell', 'Foundation', 24], ['Barr', 'Foundation', 24], ['Mis', 'Indbur', 24],
['Fie', 'Seldon', 23], ['Lepold', 'Foundation', 23], ['Korell', 'Mallow', 23], ['Riose', 'Emperor',
23], ['Galaxy', 'Bayta', 23], ['Indbur', 'Foundation', 23], ['Arcadia', 'Munn', 23], ['Arcadia', 'Mule',
23], ['Callia', 'Stettin', 23], ['Galaxy', 'Mallow', 22], ['Arcadia', 'Anthor', 22], ['Anthor', 'Turbor',
22], ['Darell', 'Semic', 22], ['Arcadia', 'Galaxy', 22], ['Arcadia', 'Mamma', 22], ['Hardin', 'Emperor',
21], ['Periphery', 'Galaxy', 21], ['Verisof', 'Hardin', 21], ['Second Foundation', 'Hari Seldon', 21],
['Ebling Mis', 'Second Foundation', 21], ['Emperor', 'Galaxy', 20], ['Second Empire', 'Seldon', 20],
['Mis', 'Magnifico', 20], ['Mis', 'Toran', 20], ['First Speaker', 'Mule', 20], ['Darell', 'Galaxy',
20], ['Munn', 'Second Foundation', 20], ['Empire', 'Hari Seldon', 19], ['Fie', 'Galaxy', 19], ['Board',
'Pirenne', 19], ['Barr', 'Empire', 19], ['Seldon', 'Bayta', 19], ['Darell', 'Kleise', 19], ['Turbor',
```

'Second Foundation', 19], ['Anthor', 'Semic', 19], ['Hardin', 'Empire', 18], ['Emperor', 'Foundation', 18], ['Hardin', 'Hari Seldon', 18], ['Periphery', 'Foundation', 18], ['Grand Master', 'Pherl', 18], ['Seldon', 'Mallow', 18], ['Galaxy', 'Toran', 18], ['Pritcher', 'Toran', 18], ['Fie', 'Pritcher', 18], ['Student', 'First Speaker', 18], ['Poochie', 'Callia', 18], ['Callia', 'Foundation', 18], ['Second Empire', 'Empire', 17], ['Sermak', 'Lee', 17], ['Mule', 'Union', 17], ['Darell', 'Pelleas Anthor', 17], ['Darell', 'Foundation', 17], ['Munn', 'Mule', 17], ['Periphery', 'Empire', 16], ['Hardin', 'Lepold', 16], ['Grand Master', 'Ponyets', 16], ['Foundation', 'Sutt', 16], ['Second Empire', 'Foundation', 16], ['Randu', 'Toran', 16], ['Randu', 'Seldon', 16], ['Mis', 'Ebling Mis', 16], ['Mis', 'Pritcher', 16], ['Arcadia', 'Second Foundation', 16], ['Anthor', 'Galaxy', 16], ['Pirenne', 'Haut Rodric', 15], ['Sermak', 'Foundation', 15], ['Second Empire', 'Galaxy', 15], ['Riose', 'Brodrig', 15], ['Barr', 'Brodrig', 15], ['Randu', 'Fran', 15], ['Fran', 'Foundation', 15], ['Randu', 'Ebling Mis', 15], ['Mule', 'Indbur', 15], ['First Speaker', 'Second Foundation', 15], ['Darell', 'Mule', 15], ['Mule', 'Stettin', 15], ['Fie', 'Anthor', 15], ['Haut Rodric', 'Hardin', 14], ['Hari Seldon', 'Emperor', 14], ['Korell', 'Sutt', 14], ['Fran', 'Toran', 14], ['Pritcher', 'Galaxy', 14], ['Mis', 'Foundation', 14], ['First Speaker', 'Galaxy', 14], ['Munn', 'Foundation', 14], ['Darell', 'Fie', 14], ['Galaxy', 'Hardin', 13], ['Emperor', 'Pirenne', 13], ['Grand Master', 'Gorov', 13], ['Hober Mallow', 'Mallow', 13], ['Riose', 'Foundation', 13], ['Barr', 'Forell', 13], ['Magnifico', 'Foundation', 13], ['Bayta', 'Second Foundation', 13], ['Mis', 'Second Foundation', 13], ['Galaxy', 'Channis', 13], ['Fie', 'Channis', 13], ['Turbor', 'Munn', 13], ['Arcadia', 'Homir', 13], ['Gaal', 'Hari Seldon', 12], ['Fara', 'Hardin', 12], ['Ducem Barr', 'Foundation', 12], ['Fie', 'Toran', 12], ['Mule', 'Empire', 12], ['Ebling Mis', 'Indbur', 12], ['Toran', 'Second Foundation', 12], ['First Speaker', 'Seldon', 12], ['Arcadia', 'Homir Munn', 12], ['Anthor', 'Kleise', 12], ['Pappa', 'Foundation', 12], ['Seldon', 'Emperor', 11], ['Emperor', 'Board', 11], ['Fara', 'Hari Seldon', 11], ['Sermak', 'Bort', 11], ['Bort', 'Hardin', 11], ['Hober Mallow', 'Salvor Hardin', 11], ['Mallow', 'Emperor', 11], ['Cleon II', 'Brodrig', 11], ['Dad', 'Toran', 11], ['Fran', 'Bayta', 11], ['Commason', 'Mule', 11], ['Mule', 'Flan Pritcher', 11], ['First Speaker', 'Channis', 11], ['Student', 'Speaker', 11], ['Mule', 'Homir Munn', 11], ['Munn', 'Stettin', 11], ['Callia', 'Mule', 11], ['Galaxy', 'Gaal', 10], ['Chen', 'Seldon', 10], ['End', 'Galaxy', 10], ['Periphery', 'Hardin', 10], ['Fie', 'Hari Seldon', 10], ['Bort', 'Foundation', 10], ['Barr', 'Mallow', 10], ['Ducem Barr', 'Empire', 10], ['Randu', 'Bayta', 10], ['Indbur', 'Seldon', 10], ['Ebling Mis', 'Galaxy', 10], ['Mis', 'Seldon', 10], ['Hari Seldon', 'Mule', 10], ['Channis', 'Elders', 10], ['Munn', 'Galaxy', 10], ['Stettin', 'Second Foundation', 10], ['Homir Munn', 'Foundation', 10], ['Homir', 'Foundation', 10], ['Callia', 'Munn', 10], ['Arcadia', 'Stettin', 10], ['Hardin', 'Seldon', 9], ['Wienis', 'Verisof', 9], ['Sutt', 'Seldon', 9], ['Second Empire', 'Hari Seldon', 9], ['Mallow', 'Empire', 9], ['Galaxy', 'Sutt', 9], ['Jael', 'Sutt', 9], ['Bel Riose', 'Empire', 9], ['Forell', 'Empire', 9], ['Cleon II', 'Emperor', 9], ['Mule', 'Han Pritcher', 9], ['Ebling Mis', 'Pritcher', 9], ['Randu', 'Mangin', 9], ['Randu', 'Indbur', 9], ['Commason', 'Foundation', 9], ['Empire', 'Second Foundation', 9], ['Pritcher', 'Union', 9], ['Bail Channis', 'Second Foundation', 9], ['Darell', 'Homir Munn', 9], ['Mule', 'Anthor', 9], ['Munn', 'Fie', 9], ['Arcadia', 'Uncle Homir', 9], ['Arcadia', 'Fie', 9], ['Gaal', 'Empire', 8], ['Fie', 'Emperor', 8], ['Board', 'Foundation', 8], ['Yohan Lee', 'Hardin', 8], ['Lee', 'Foundation', 8], ['Salvor Hardin', 'Seldon', 8], ['Gorov', 'Ponyets', 8], ['Loris', 'Foundation', 8], ['Korell', 'Galaxy', 8], ['Jaim Twer', 'Mallow', 8], ['Mallow', 'Far Star', 8], ['Periphery', 'Mallow', 8], ['Mallow', 'Jorane Sutt', 8], ['Korell', 'Empire', 8], ['Hober Mallow', 'Seldon', 8], ['Riose', 'Empire', 8], ['Riose', 'Galaxy', 8], ['Riose', 'Siwennian', 8], ['Riose', 'Hari Seldon', 8], ['Ducem Barr', 'Bel Riose', 8], ['Bel Riose', 'Foundation', 8], ['Galaxy', 'Brodrig', 8], ['Barr', 'Galaxy', 8], ['Flan Pritcher', 'Foundation', 8], ['Magnifico', 'Pritcher', 8], ['Second Empire', 'Second Foundation', 8], ['Galaxy', 'Union', 8], ['Mule', 'Bail Channis', 8], ['Speaker', 'First Speaker', 8], ['Kleise', 'Second Foundation', 8], ['Munn', 'Semic', 8], ['Homir Munn', 'Second Foundation', 8], ['Stettin', 'First Minister', 8], ['Stettin', 'Homir Munn', 8], ['Supervisor', 'Gaal', 7], ['Avakim', 'Gaal', 7], ['Chen', 'Gaal', 7], ['Fulham', 'Hardin', 7], ['Salvor Hardin', 'Hari Seldon', 7], ['Verisof', 'Foundation', 7], ['Wienis', 'Salvor Hardin', 7], ['Salvor Hardin', 'Korell', 7], ['Hardin', 'Mallow', 7], ['Riose', 'Ducem Barr', 7], ['Fie', 'Empire', 7], ['Brodrig', 'Foundation', 7], ['Lathan Devers', 'Foundation', 7], ['Bayta', 'Fie', 7], ['Fran', 'Seldon', 7], ['Randu', 'Galaxy', 7], ['Fie', 'Indbur', 7], ['Fie', 'Flan Pritcher', 7], ['Mis', 'Hari Seldon', 7], ['Mule', 'Mangin', 7], ['Ebling Mis', 'Seldon', 7], ['Bayta', 'Commason', 7], ['Union', 'Foundation', 7], ['Bail Channis', 'Channis', 7], ['Bail Channis', 'Pritcher', 7], ['Fie', 'First Speaker', 7], ['Darell', 'Seldon', 7], ['Speaker', 'Second Foundation', 7], ['Darell', 'Homir', 7], ['Semic', 'Second Foundation', 7], ['Mule', 'Uncle Homir', 7], ['Stettin', 'Meirus', 7], ['Stettin', 'Homir', 7], ['Callia', 'Darell', 7], ['Emperor', 'Gaal', 6], ['Advocate', 'Seldon', 6], ['Hardin', 'Salvor Hardin', 6], ['Pirenne', 'Foundation', 6], ['Board', 'Dorwin', 6], ['Dorwin', 'Emperor', 6], ['Lee', 'Hari Seldon', 6], ['Council', 'Hardin', 6], ['Empire', 'Salvor Hardin', 6], ['Sutt', 'Hari Seldon', 6], ['Periphery', 'Korell', 6], ['Periphery', 'Seldon', 6], ['Hober Mallow',

'Korell', 6], ['Fie', 'Mallow', 6], ['Mallow', 'Asper', 6], ['Barr', 'Siwennian', 6], ['Hober Mallow', 'Forell', 6], ['Riose', 'Fie', 6], ['Barr', 'Seldon', 6], ['Pritcher', 'Han Pritcher', 6], ['Pritcher', 'Seldon', 6], ['Mule', 'Kalgan', 6], ['Han Pritcher', 'Foundation', 6], ['Ebling Mis', 'Hari Seldon', 6], ['Fran', 'Iwo', 6], ['Mule', 'Fran', 6], ['Mule', 'Ovall', 6], ['Mule', 'Fox', 6], ['Magnifico', 'Commason', 6], ['Ebling Mis', 'Fie', 6], ['Mule', 'Second Empire', 6], ['Union', 'Channis', 6], ['Union', 'Second Foundation', 6], ['Channis', 'Seldon', 6], ['Student', 'Galaxy', 6], ['Speaker', 'Second Empire', 6], ['Anthor', 'Pelleas Anthor', 6], ['Munn', 'Homir', 6], ['Munn', 'Pelleas Anthor', 6], ['Munn', 'Homir Munn', 6], ['Callia', 'Uncle Homir', 6], ['Arcadia', 'Palver', 6], ['Mule', 'Lieutenant Dirige', 6], ['Anthor', 'Foundation', 6], ['Stettin', 'Darell', 6], ['Semic', 'Turbor', 6], ['Jerril', 'Gaal', 5], ['Chen', 'Galaxy', 5], ['Galaxy', 'Salvor Hardin', 5], ['Emperor', 'Haut Rodric', 5], ['Hardin', 'Lundin Crast', 5], ['Dorwin', 'Empire', 5], ['Dorwin', 'Pirenne', 5], ['Board', 'Lee', 5], ['Yohan Lee', 'Lee', 5], ['Empire', 'Sermak', 5], ['Wienis', 'Hari Seldon', 5], ['Lepold', 'Salvor Hardin', 5], ['Walto', 'Bort', 5], ['Bort', 'Lepold', 5], ['Grand Master', 'Foundation', 5], ['Gorov', 'Pherl', 5], ['Ponyets', 'Pherl', 5], ['Periphery', 'Sutt', 5], ['Korell', 'Seldon', 5], ['Hober Mallow', 'Empire', 5], ['Salvor Hardin', 'Mallow', 5], ['Jael', 'Korell', 5], ['Ducem Barr', 'Barr', 5], ['Ducem Barr', 'Fie', 5], ['Bel Riose', 'Galaxy', 5], ['Barr', 'Loris', 5], ['Dad', 'Foundation', 5], ['Fie', 'Fran', 5], ['Toran', 'Seldon', 5], ['Galaxy', 'Fran', 5], ['Indbur', 'Han Pritcher', 5], ['Personal Capsule', 'Foundation', 5], ['Galaxy', 'Magnifico', 5], ['Han Pritcher', 'Toran', 5], ['Pritcher', 'Indbur', 5], ['Randu', 'Ovall Gri', 5], ['Mangin', 'Foundation', 5], ['Randu', 'Ovall', 5], ['Periphery', 'Toran', 5], ['Toran', 'Commason', 5], ['Channis', 'Foundation', 5], ['Pritcher', 'Elders', 5], ['Arcadia', 'Seldon', 5], ['Student', 'Second Foundation', 5], ['Darell', 'Jole Turbor', 5], ['Munn', 'Kleise', 5], ['Anthor', 'Tamper Plateau', 5], ['Munn', 'Seldon', 5], ['Stettin', 'Galaxy', 5], ['Stettin', 'Poochie', 5], ['Arcadia', 'Poochie', 5], ['Callia', 'Homir Munn', 5], ['Mamma', 'Foundation', 5], ['Darell', 'Lieutenant Dirige', 5], ['Turbor', 'Palver', 5]]

List of tuple events (subject, verb, object, sentence index):

[['Gaal', 'emptying', 'Avakim', 187], ['Chen', 'watched', 'Gaal', 223], ['Pirenne', 'replied', 'Board', 458], ['Fara', 'eyed', 'Hardin', 580], ['Commissioners', 'maneuvered', 'Seldon', 722], ['Fara', 'reminded', 'Hardin', 724], ['Hardin', 'facing', 'Pirenne', 789], ['Foundation', 'established', 'Seldon', 858], ['Foundation', 'appointed', 'Seldon', 1061], ['Foundation', 'saved', 'Hardin', 1103], ['Wienis', 'bluff', 'Hardin', 1232], ['Lefkin', 'found', 'Fie', 1282], ['Riose', 'concerns', 'Brodrig', 2622], ['Riose', 'seized', 'Barr', 2991], ['Toran', 'can', 'Randu', 3249], ['Fran', 'muttered', 'Whew', 3277], ['Foundation', 'established', 'Seldon', 3279], ['Kalgan', 'left', 'Toran', 3620], ['Foundation', 'established', 'Seldon', 3650], ['Indbur', 'shook', 'Mis', 3690], ['Foundation', 'beat', 'Mule', 3719], ['Foundation', 'beat', 'Gri', 3772], ['Ovall', 'exploded', 'Mule', 3787], ['Mule', 'figure', 'Seldon', 4160], ['Fox', 'tossed', 'Fie', 4174], ['Foundation', 'conquered', 'Mule', 4222], ['Foundation', 'whispered', 'Bayta', 4271], ['Toran', 'cried', 'Pritcher', 4589], ['Mule', 'beat', 'Foundation', 4748], ['Empire', 'establish', 'Mule', 4751], ['Mis', 'see', 'Toran', 4769], ['Mule', 'joined', 'Neotrantor', 4799], ['Mule', 'fought', 'Pritcher', 4903], ['Foundation', 'fear', 'Channis', 4920], ['Pritcher', 'buried', 'Mule', 4973], ['Channis', 'reached', 'Pritcher', 5093], ['Pritcher', 'found', 'Channis', 5144], ['Fie', 'faced', 'Mule', 5543], ['Channis', 'seized', 'Mule', 5555], ['Foundation', 'founded', 'Seldon', 5562], ['Foundation', 'defeated', 'Fie', 5694], ['Homir', 'bother', 'Foundation', 6128], ['Callia', 'whimpered', 'Poochie', 6169], ['Meirus', 'questioned', 'Foundation', 6176], ['Stettin', 'repeated', 'Foundation', 6177], ['Callia', 'turned', 'Stettin', 6186], ['Foundation', 'conquered', 'Mule', 6241], ['Kalgan', 'stop', 'Foundation', 6285], ['Foundation', 'found', 'Pritcher', 6329], ['Homir', 'leaving', 'Arcadia', 6378], ['Terminus', 'left', 'Arcadia', 6563], ['Darell', 'asked', 'Mule', 6672], ['Foundation', 'save', 'Foundation', 6715], ['Galaxy', 'conquered', 'Mule', 6955], ['Turbor', 'followed', 'Semic', 6992], ['Galaxy', 'conquered', 'Mule', 7015], ['Foundation', 'established', 'Munn', 7032], ['Munn', 'waved', 'Darell', 7125], ['Anthor', 'admitted', 'Fie', 7166], ['Foundation', 'left', 'Seldon', 7217]]

D. APPENDIX

Normalized 'PER' entity list:

[['Salvor Hardin', 'Hardin'],

```
['Forell', 'Korell'],
['Bail Channis', 'Channis'],
['Ebling Mis', 'Mis'],
['Fara', 'Far Star'],
['Flan Pritcher', 'Han Pritcher'],
['Second Foundation', 'Foundation'],
['Dorwin'],
['Gaal'],
['Grand Master'],
['Emperor'],
['Ponyets'],
['Jaim Twer'],
['Commason'],
['Wienis'],
['Iwo'],
['First Minister'],
['Palver'],
['Dad'],
['Board'],
['Sermak'],
['Galaxy'],
['Avakim'],
['Siwennian'],
['End'],
['Periphery'],
['Pherl'],
['Fulham'],
['Jael'],
['Gorov'],
['Stettin'],
['Bort'],
['Pirenne'],
['Tamper Plateau'],
['Verisof'],
['Asper'],
['Chen'],
['Student'],
['Poochie'],
['Bayta'],
['Council'],
['Cleon II'],
['Meirus'],
['Mule'],
['Mangin'],
['Haut Rodric'],
['Pappa'],
['Darell'],
['Kalgan'],
['Lieutenant Dirige'],
['Fox'],
['Toran'],
['Pritcher'],
['Elders'],
['Semic'],
['Supervisor'],
['Brodrig'],
['Fie'],
['Indbur'],
['Union'],
```

```
['Jerril'],
['Callia'],
['Magnifico'],
['Mamma'],
['Lundin Crast'],
['Advocate'],
['Personal Capsule'],
['Lathan Devers'],
['Kleise'],
['Lepold'],
['Arcadia'],
['Randu'],
['Walto'],
['Loris'],
['Fran'],
['Lee', 'Yohan Lee'],
['Riose', 'Bel Riose'],
['Anthor', 'Pelleas Anthor'],
['Munn', 'Homir Munn'],
['Second Empire', 'Empire'],
['Ovall Gri', 'Ovall'],
['Turbor', 'Jole Turbor'],
['Sutt', 'Jorane Sutt'],
['First Speaker', 'Speaker'],
['Hari Seldon', 'Seldon'],
['Mallow', 'Hober Mallow'],
['Homir Munn', 'Homir'],
['Uncle Homir', 'Homir'],
['Barr', 'Ducem Barr']]
```