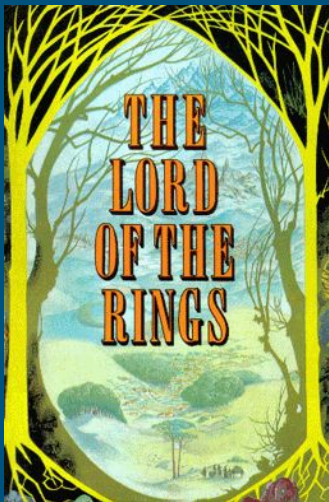




Textual Exploration of LOTR



Blaz



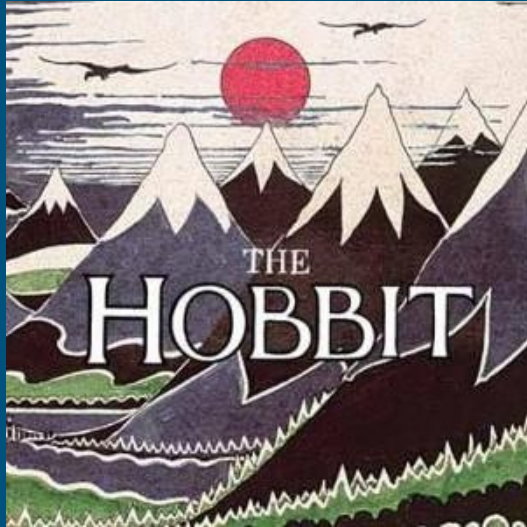
Agenda

- Background
- Corporuses
- Goal
- Sentiment Analysis
- Topic Modeling
- Conclusion





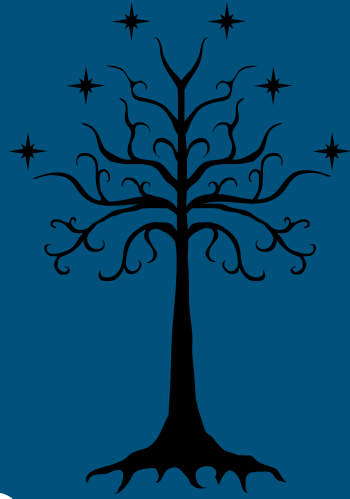
Background



Lord of the Rings

J.R.R. Tolkien





Corpuses

- The Silmarillion
- The Hobbit
- Fellowship of the Ring
- The Two Towers
- Return of the King





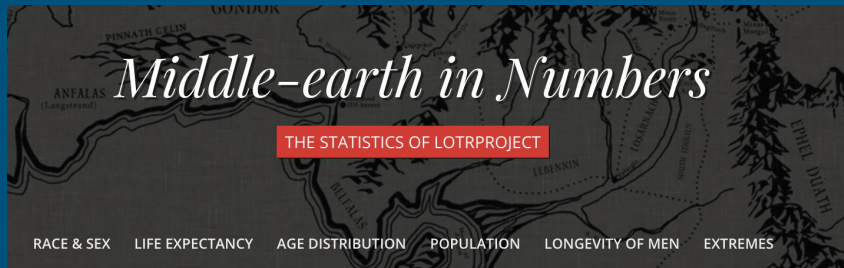
Goal

- Sentiment Analysis
- Topic Modeling
- Learn more about LOTR





Related Work



- Emil Johansson
 - LOTRproject.com
- Sentiment140
 - Jurafsky and Martin
 - SALP
- Machine Learning w/ Python
 - Müller and Guido

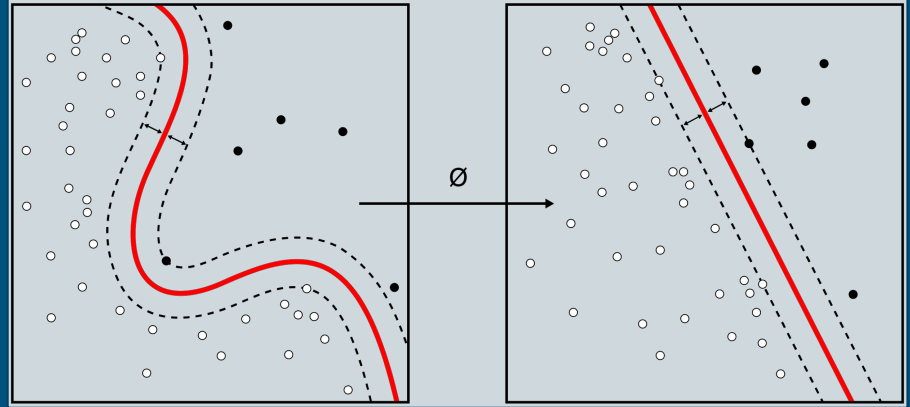


Sentiment Analysis



Sentiment140

1. Naive Bayes
2. Maximum Entropy
3. SVM



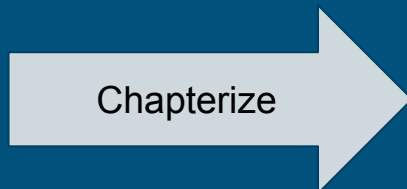
TwittRatr.com Lexicon



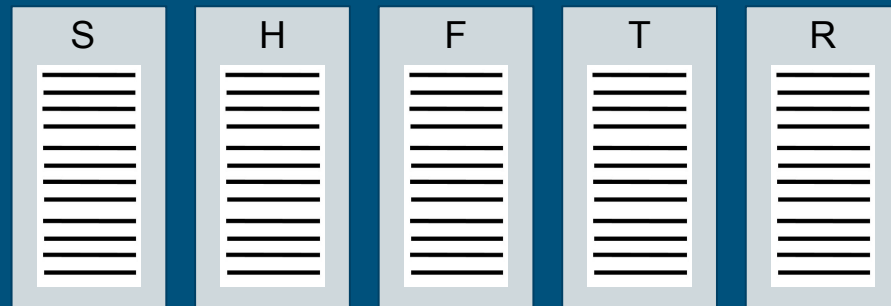
Sentiment Analysis

- Sentiment140 API
- Chapterize
- Pandas

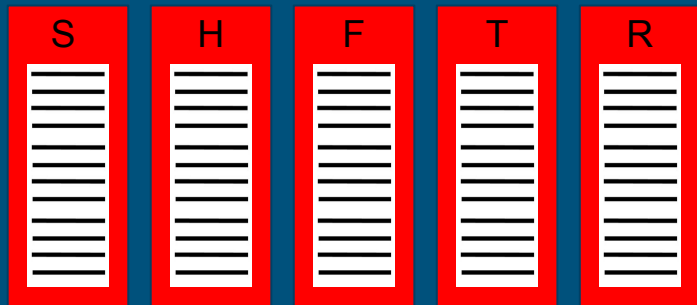
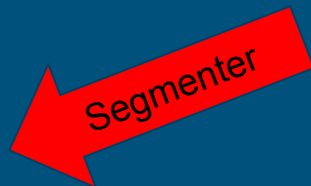




Chapter Objects



Page Objects



Sentiment140

Search by product or brand. Discover the Twitter sentiment.

Spanish



Requests

0 - Negative, 2 - Neutral, 4 - Positive

```
d = {"data": [{"text": 'the titanic was awesome'}, \
            {"text": "shrek is the worst movie in cognitive human history"}]}
```

```
{'text': 'the titanic was awesome', 'polarity': 4, 'meta': {'topic': 'movies', 'language': 'en'}}
{'text': 'shrek is the worst movie in cognitive human history', 'polarity': 0, 'meta': {'language': 'en'}}
```

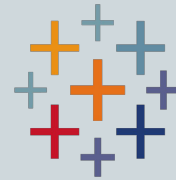
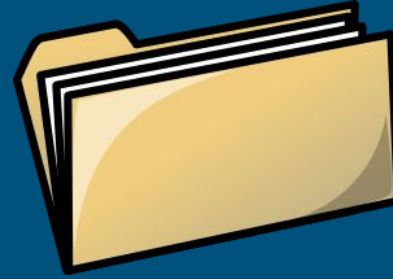
~1,300 Requests



Dataframes



CSVs

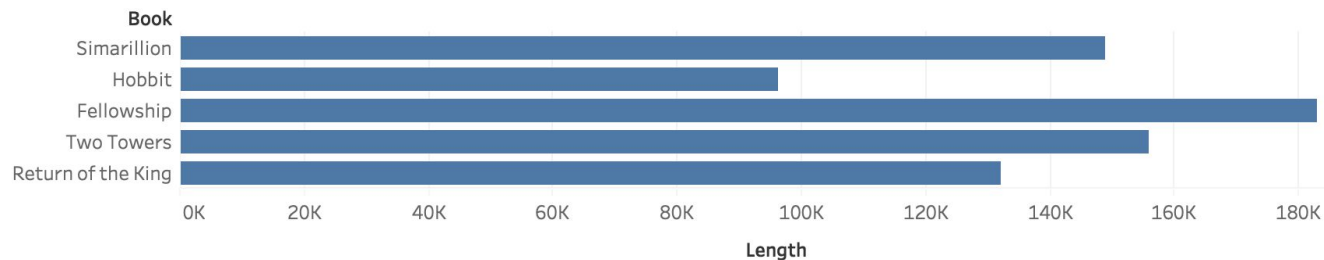


+ a b l e a u[®]



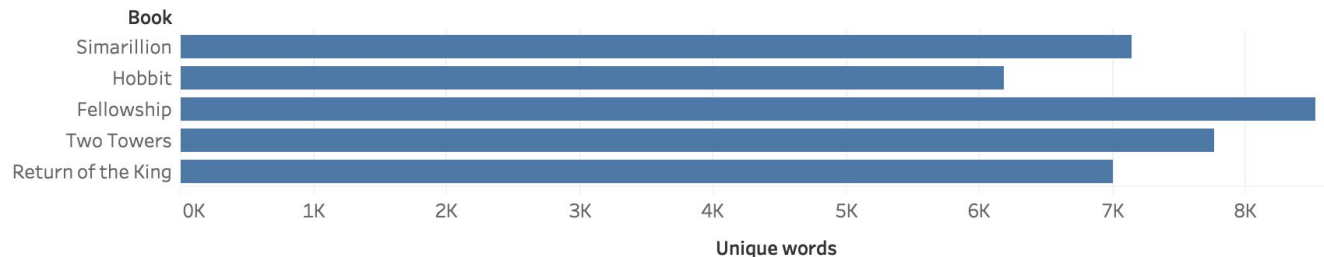
Corpus Information

Lengths



Length for each Book.

Unique Words



Sum of Unique words for each Book.



Translation issues

- Character names
 - Sméagol (Gollum)
 - Éowyn
- Orcs

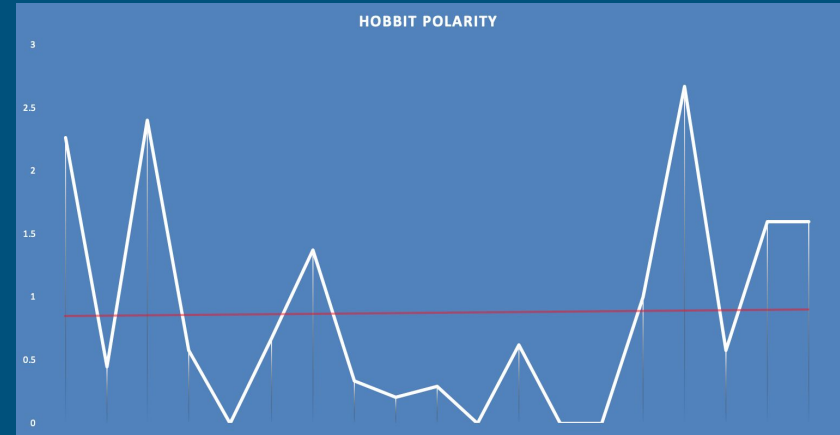




Sentiment averages

| Book | AVG Polarity |
|--------------|--------------|
| Silmarillion | 1.80 |
| Hobbit | 0.87 |
| Fellowship | 1.22 |
| Two Towers | 1.05 |
| Return | 1.06 |

- Novel
- One Adventure
- Apart from Trilogy

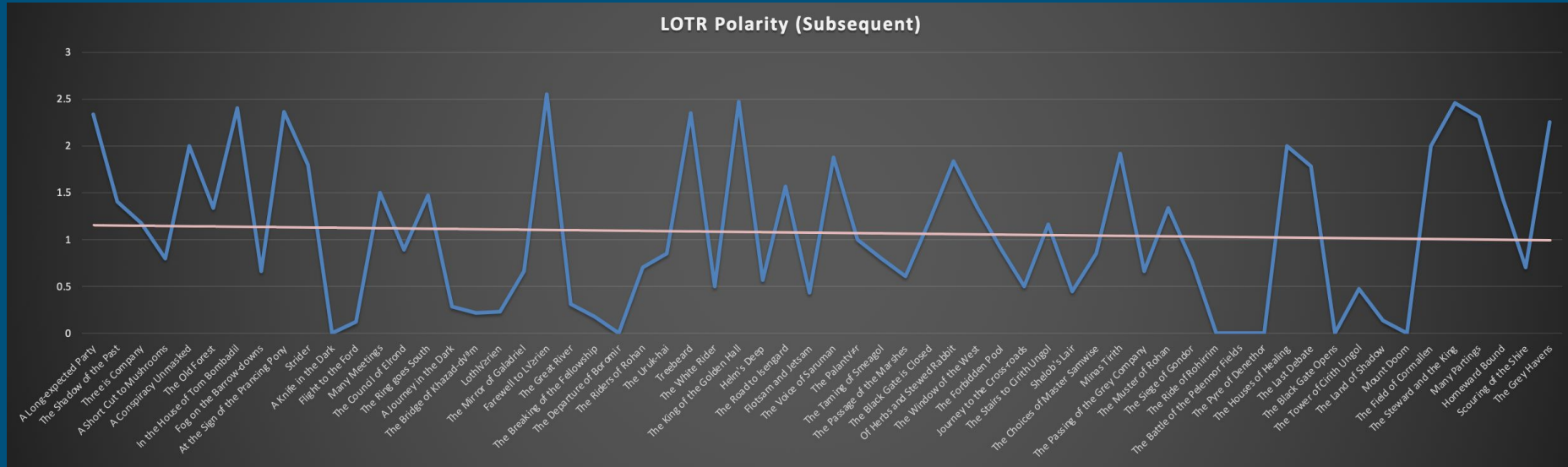


- Generally between Neutral and Negative





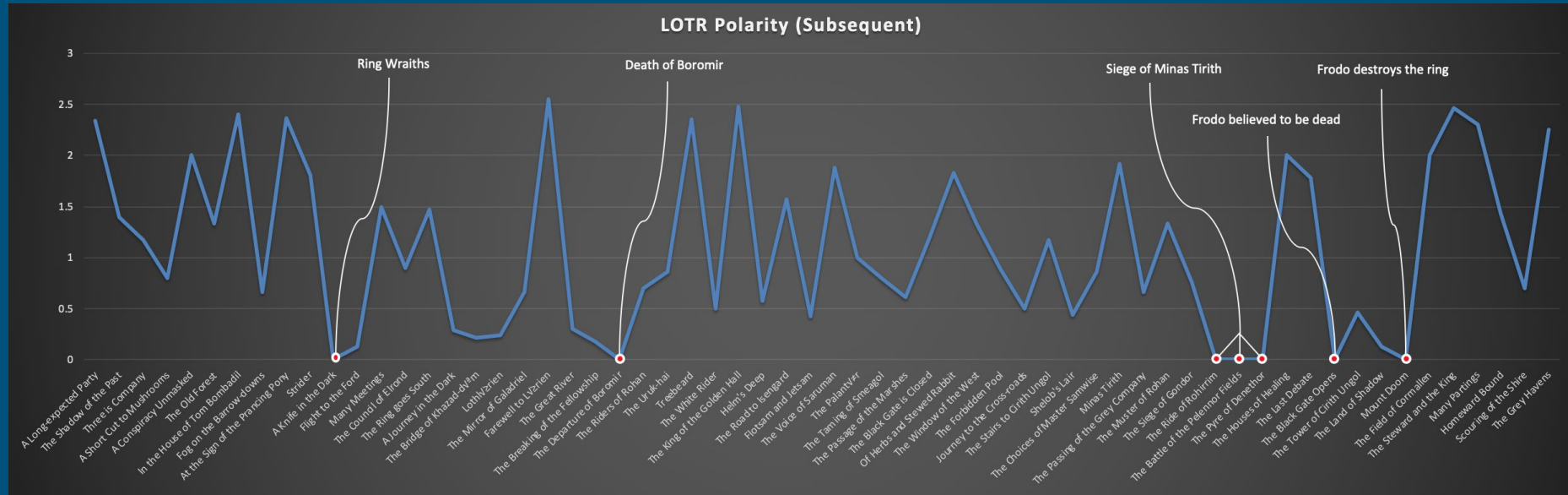
Visuals



- Removed The Fellowship prologue
- Added Trendline



Visuals



- Full-Negative Chapters



Chapters with Polarity of 0

- A Knife in the Dark
- The Departure of Boromir
- The Ride of Rohirrim
- The Battle of the Pelennor Fields
- The Pyre of Denethor
- The Black Gate Opens
- Mount Doom



Common words in Full-Negative chapters



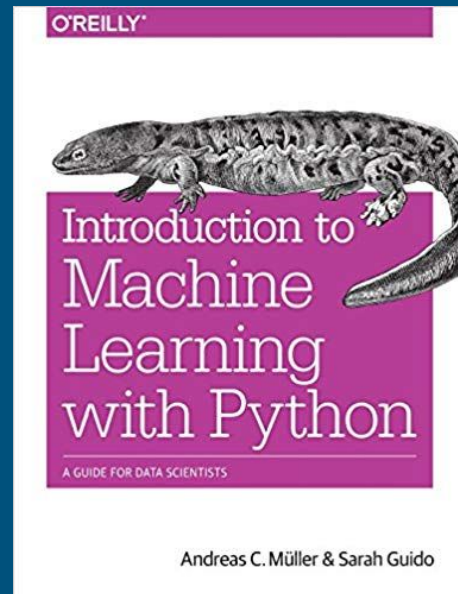
| Chapter | Words (In order of frequency) |
|-----------------------------------|------------------------------------|
| A Knife in the Dark | Last, us, away, left, road |
| The Departure of Boromir | Orcs, away, long, boat, wind |
| The Ride of the Rohirrim | Wild, king, road, like, away |
| The Battle of the Pelennor Fields | King, like, fell, black, city |
| The Pyre of Denethor | Lord, stood, city, door, away |
| The Black Gate Opens | Mordor, black, sauron, last, away |
| Mount Doom | Mountain, master, away, dark, last |

Topic Modeling



Topic Modeling

- LDA (Latent Dirichlet Allocation)
 - Finds groups of words (features) that appear together frequently
 - Loosely define these words as 'topics'
 - Unlabeled data





Steps

- Tokenize
- Lemmatization
- Stop words/'Tolkien' Stop words
 - Remove words that will dominate the text
- Vectorize
- Fit to LDA model



LDA Examples

| topic 0 | topic 1 | topic 2 | topic 3 | topic 4 |
|-------------|-------------|----------|------------|---------|
| between | war | funny | show | didn |
| young | world | worst | series | saw |
| family | us | comedy | episode | am |
| real | our | thing | tv | thought |
| performance | american | guy | episodes | years |
| beautiful | documentary | re | shows | book |
| work | history | stupid | season | watched |
| each | new | actually | new | now |
| both | own | nothing | television | dvd |
| director | point | want | years | got |

| topic 5 | topic 6 | topic 7 | topic 8 | topic 9 |
|----------|-----------|-------------|--------------|---------|
| horror | kids | cast | performance | house |
| action | action | role | role | woman |
| effects | animation | john | john | gets |
| budget | game | version | actor | killer |
| nothing | fun | novel | oscar | girl |
| original | disney | both | cast | wife |
| director | children | director | plays | horror |
| minutes | 10 | played | jack | young |
| pretty | kid | performance | joe | goes |
| doesn | old | mr | performances | around |

- Supervised learning
- Random state
- Feature Weights

LDA - LOTR

The Black Gate Opens

```
lda.components .shape: (10, 1039)
```

```
topic 0
-----
black
day
army
fear
forth
left
tall
south
mind
nazgyl
```

```
topic 1
-----
mordor
far
cried
pass
north
wish
banner
grey
saw
silence
```

```
topic 2
-----
hills
long
tower
messenger
set
laughed
minas
land
march
remained
```

```
topic 3
-----
away
west
heard
did
soon
red
east
began
anduin
enemies
```

```
topic 4
-----
enemy
ores
destroyed
maybe
assault
way
scouts
blew
watched
time
```

```
topic 5
-----
stood
captains
imrahil
king
head
hideous
wind
leave
air
eastward
```

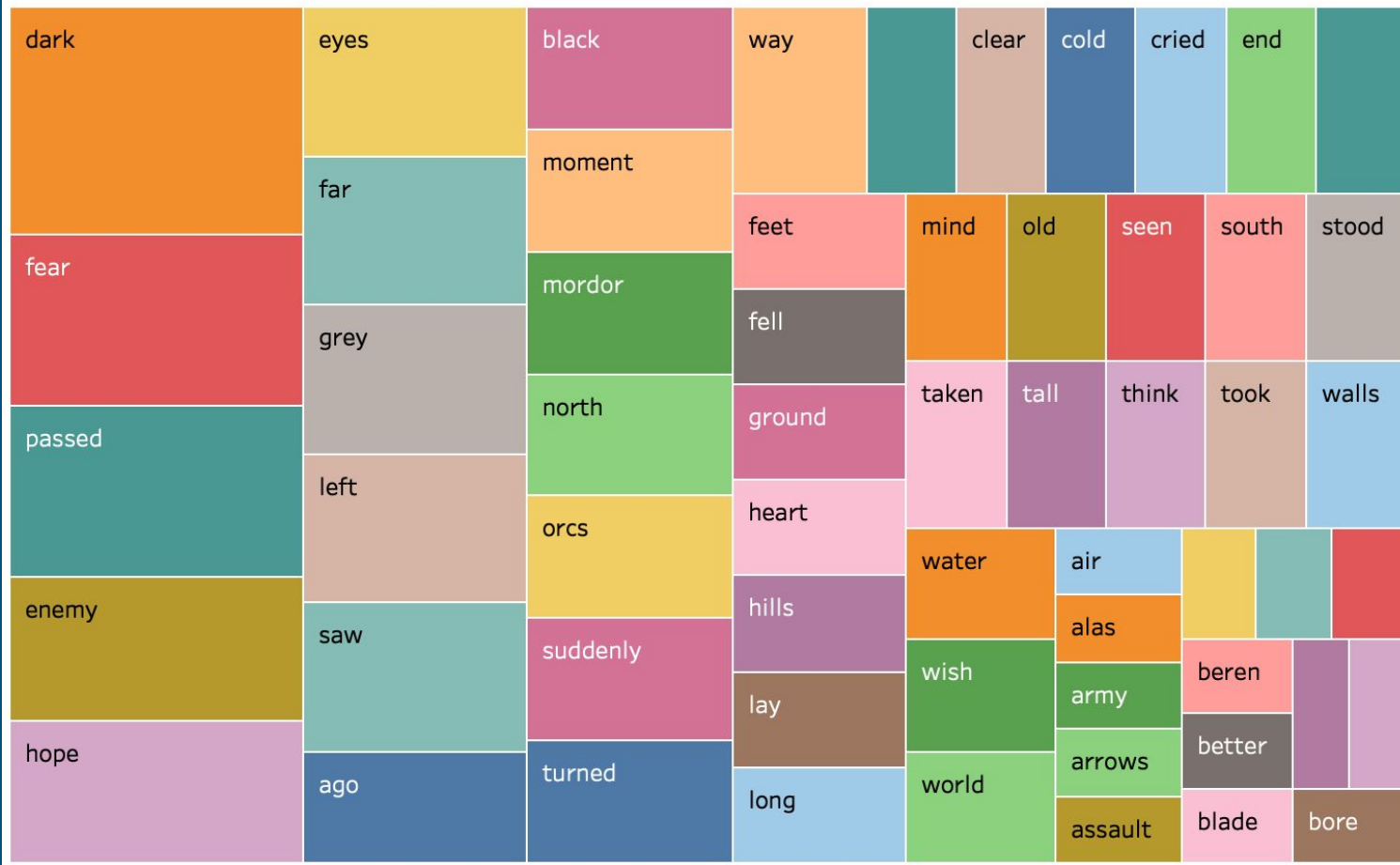
```
topic 6
-----
evil
end
thought
road
host
thou
broken
lord
trumpets
forward
```

```
topic 7
-----
sauron
hope
lay
dark
heralds
terms
guard
rohan
war
blade
```

```
topic 8
-----
gondor
company
ring
mouth
tirith
gone
fair
sun
let
city
```

```
topic 9
-----
like
little
old
eyes
morannon
sword
white
master
rode
morgul
```

Common Features in Full-Negative Chapter Topic Modeling



Positive Chapter Word Tiles





Negative Chapter Word Tiles



Conclusions

Conclusions

- Latent Dirichlet Allocation
 - Weighted Features
 - High-Fantasy
- Recurrent Neural Networks
 - Summarization
- LOTR

References

Go, Alec, M, et al. *Twitter Sentiment Classification using Distant Supervision*. Retrieved from <https://cs.stanford.edu/people/alecmgo/papers/TwitterDistantSupervision09.pdf>

Guido, S. (2016). *Introduction to Machine Learning with Python: A Guide for Data Scientists*. Retrieved from <https://www.safaribooksonline.com/library/view/introduction-to-machine/9781449369880/preface01.html#idm140338259230048>