



# **Digital Humanities and Computer Assisted Textual Analyses (CATA) Technologies**

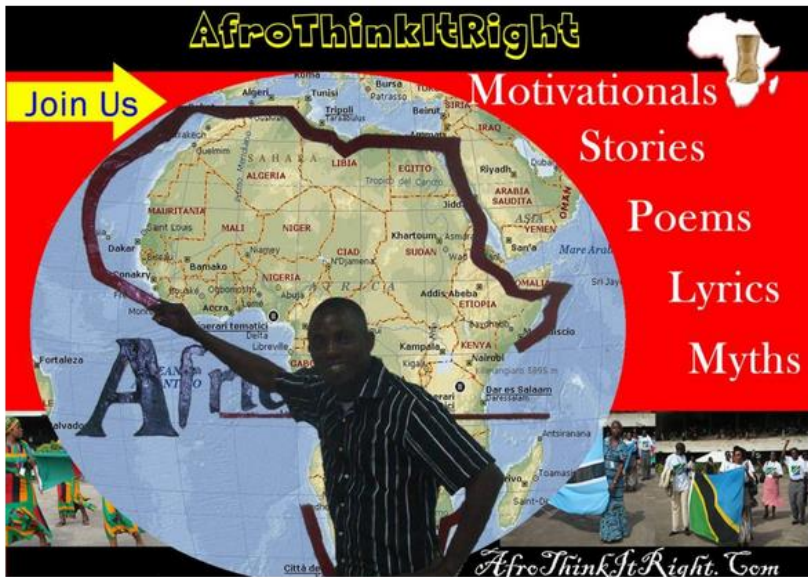
Augustine Farinola, PhD Researcher, Department of English  
Literature, University of Birmingham, United Kingdom.

[www.augustinefarinola.com](http://www.augustinefarinola.com); [aaf936@bham.ac.uk](mailto:aaf936@bham.ac.uk)

# AFRO-THINK-IT-RIGHT

A Web Portal For Young African Writers

## Home



Welcome to Afro-think-it-Right Platform

Our primary goal is to provide a medium of self-expression for millions of talented young writers out there, whose minds have become fertile grounds for ideas that enhance personal growth and fulfilment. Therefore, this website serves as a medium for these creative individuals to experiment their thoughts and gain new insights by being expose to views of like-minds.

### SCIENCE AND TECHNOLOGY



Science And Technology

### WHAT IS THE FUTURE OF AI TECHNOLOGY?



# PAINT ME BLACK

A Visual DH Project

- THE PROJECT
- BEADS
- CLOTHINGS
- HAIRSTYLES
- SCULPTURE
- THE AUTHOR
- AFRO-THINK-IT-RIGHT



## The Aesthetics and Rationality of African Fractals

- Sculpture
- Hairstyles
- Clothings
- Beads
- The Project



Project Description

Please follow & like us :)

78
 
 125
 356

Subscribe to get Research Updates

Email

**SUBSCRIBE**

What's on your mind?

Name

# CreativeTechies

[LOG IN](#) [REGISTER](#) [PYTHON](#)

## Log In

You are currently logged in as Sagaust. [Log out »](#)

### Recent Posts

[Blog Post Title](#)

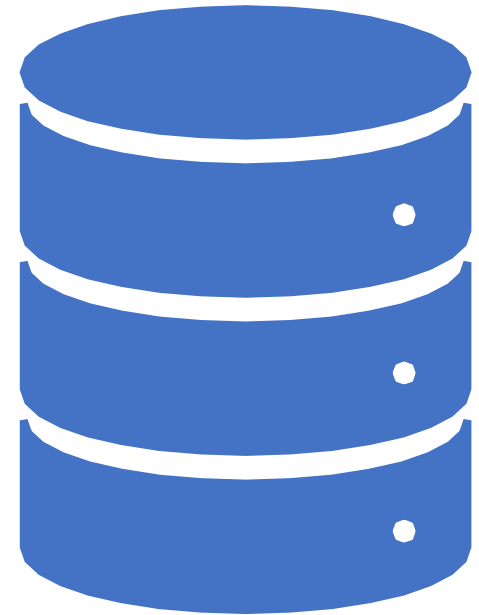
### OUTLINE

# TEXT AS DATA

- The large amount of data made available through online repositories, digital archives, and Corpora necessitate the development of digital technologies that would be used for analysis and visualization of textual data (Patrik Svensson 2012).
- According to Miriam Posner, humanists do not think of the material they work with as something that it exists in discrete, fungible units....In this digital age, humanists have begun to see that books, paintings, movies, fiction, paintings, or whatever we are studying, in the form of texts, images, or sounds, are all 'data'). Although the understanding of 'data' varies across disciplines (science, social science, arts, management, and engineering)

# TEXT AS DATA

- With the increased amount of data publicly available and the increased focus on unstructured text data, understanding how to clean, process, and analyse that text data is tremendously valuable.
- Humanists like tried to explain what data are and what we are to data (Mary Poovey et. al., *Raw Data Is an Oxymoron*, 2013).
- Retrieving, storing and analysing digital data is neither an easy nor quick task in the humanities (Gabriele Griffin and Matt Hayler, 2016)







online   SEM  

FREQUENCY      CONTEXT

to carry all business before um: demosthenes calls the people an evil beast: **plato**, a monster with many neac  
 and truth: this some of the Ethnick philosophers were not ignorant of; therefore **plato** forbid that any ceremo  
 fire: moses in his laws ordain'd most severe punishment for this crime: and **plato** extirpates it out of his Repu  
 'um quite away, and prostitute'um in common to their fellows, according to **plato's** law; &; whereas they ought  
 rest of the citizens whatever commodities of the illyrians they stood in need of: **plato** very much blames Mercha  
 city of athens first taken; which makes me wonder to find hunting commended by **plato** prince of the Academick  
 serve all the stratagems, ambuscades, and rules of war? yet the divine **plato** praises this art, teaches it to his scho  
 the Bloud, or the most useful part of the excrement of the meat: **plato** affirms it to be a deflux of the back-bone-n  
 the example of eudemus, in livia wife of drusus caesar: socates also in **plato** advises not to let physicians multiply  
 art of cookery, being the dressing and ordering of victuals: for which reason **plato** calls it the flatteress of Physid  
 have convinced by most strenuous arguments: nor shall we speak of the religion of **plato** or the rest of the philo  
 others by reason: others constitute infinite realities, as they call um, like **plato's** idea's; which some again deny, a  
 , virgil, livie, strabo, varro, seneca, cicero, aristotle and **plato**: in stead of preaching the gospel, and word of god,  
 up seven apparitions in dreams, and as many waking: so we read in **plato** and proculus of socrates, that he v  
 and for all their skill could never double the cape of good hope: where **plato** ended, our author begins, ubi  
 : [.] # whether therefore the Platonick triad, or the threefold principle mention'd by **plato** and his followe  
 favourable interpretation that infers an unity in the trinity, and trinity in unity from **plato's** principle: y  
 asserters of the Platonick trinity do acknowledge: [.] # whether orpheus, trismegist or **plato** had indee  
 trinity, as relating to the christian mystery, is much more obscurely express'd in **plato** himself, ther  
 there said of it, nor could it probably have been concluded to have been **plato's** meaning, if the apr

# CORPUS

## Code Repository for: Conjoint Analysis Tools for Developing Country Contexts

Version 1.1

Meyer, Alexander; Rosenzweig, Leah R., 2016, "Code Repository for: Conjoint Analysis Tools for Developing Country Contexts", <https://doi.org/10.7910/DVN/DURDD3>, Harvard Dataverse, V1

Cite Dataset

Learn about Data Citation Standards.

Dataset Metrics

624 Downloads

### Description

This data set contains the code repositories for the paper entitled: Conjoint Analysis Tools for Developing Country Contexts. Up to date versions of the repositories visit: <https://github.com/acmeyer/Conjoint-PDF-Generator-App> and <https://github.com/acmeyer/Conjoint-for-Qualtrics-Offline>. (2016-02-15)

### Subject

Social Sciences

### Keyword

conjoint analysis, offline, Qualtrics

Files Metadata Terms Versions

Search this dataset... Find

Filter by

File Type: All Access: All

1 to 10 of 133 Files

.ctags  
Plain Text - 31 B - Feb 16, 2016 - 12 Downloads  
MDS: 8274970ed8bfc7e84ceb2a5de904c5f

.gitignore  
Plain Text - 216 B - Feb 16, 2016 - 7 Downloads

# REPOSITORIES



Repository details



## HSRC Research Data Service

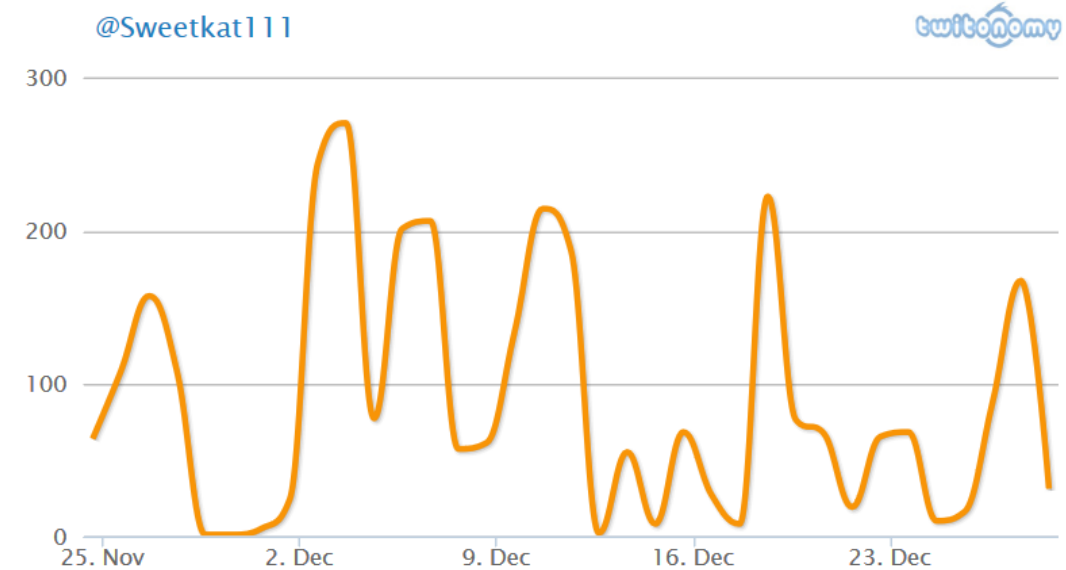
**General** Institutions Terms Standards

Name of repository	<b>HSRC Research Data Service</b>
Additional name(s)	Human Sciences Research Council Research Data Service
Repository URL	<a href="http://datacuration.hsrc.ac.za/">http://datacuration.hsrc.ac.za/</a>
Subject(s)	<a href="#">Empirical Social Research</a> <a href="#">Political Science</a> <a href="#">Economic and Social Policy</a> <a href="#">Social Sciences</a> <a href="#">Social and Behavioural Sciences</a> <a href="#">Humanities and Social Sciences</a> <a href="#">Economics</a>
Description	The HSRC Research Data Service provides a digital repository facility for the HSRC's research data in support of evidence based human and social development in South Africa and the broader region. It includes both quantitative and qualitative data. Access to data is dependent on ethical requirements for protecting research participants, as well as on legal agreements with the owners, funders or in the case of data owned by the HSRC, the requirements of the depositors of the data.
Contact	<a href="mailto:lloffer@hsrc.ac.za">lloffer@hsrc.ac.za</a> <a href="http://datacuration.hsrc.ac.za/contact/datahelp">http://datacuration.hsrc.ac.za/contact/datahelp</a>
Content type(s)	<a href="#">Standard office documents</a> <a href="#">Plain text</a> <a href="#">Scientific and statistical data formats</a> <a href="#">Databases</a> <a href="#">Images</a> <a href="#">Archived data</a>
Keyword(s)	<a href="#">South Africa</a> <a href="#">census data</a> <a href="#">microdata</a> <a href="#">community survey</a> <a href="#">unemployment</a> <a href="#">agrarian reform</a> <a href="#">poverty</a> <a href="#">HIV/AIDS</a> <a href="#">health system</a> <a href="#">mental health</a> <a href="#">child maltreatment prevention</a> <a href="#">alcohol prevention</a> <a href="#">schools</a>



### Tweet history

Download: [Chart](#) [Tweets](#)



### Users most retweeted

	@SweetKat111	74		@Brn2Wander1	72
	@OrenIShiKiddo	69		@BlueSea1964	64
	@vanessacvenus	60		@silver_fox9	59
	@rosedixontx	59		@GA_peach3102	58
	@Boyd_2650	58		@PoliticalIntent	57



Tweet

### Followers

Download & Print

- 滴答 @doperfectly**  
 0 tweets 211 following 6 followers 0 listed May 2013
- John @John47070623**  
 111,358 tweets 4,991 following 3,689 followers 107 listed June 2016  
 Minnesota/Wisconsin  
 CNN=fake news! #DRAINTHESWAMP #MAGA #Hillary4prison2017 #Trump2020 Blocked by @GeorgeTakei
- fred haerer @fred0007**  
 1,143 tweets 2,167 following 215 followers 4 listed February 2009 Forest Hills, MI
- Journey Home Project/ Founder Charlie Daniels @TJHproject**  
 3,395 tweets 3,085 following 4,384 followers 11 listed July 2014 Tennessee, USA  
 We assist not-for-profit causes that provide veterans with rehabilitation, reintegration and opportunities as they transition to civilian life.
- ZEN @supermary2020**  
 2 tweets 65 following 0 followers 0 listed November 2019  
 Bless nature!
- Sofa King Anti-Socialist @jocase192**  
 6,447 tweets 7,206 following 7,194 followers 22 listed August 2016 Florida, USA  
 Conservative commentator. "government is not the solution to our problem; government is the problem" Ronald Reagan
- Sam @lihaiting1993**  
 1,201 tweets 426 following 47 followers 0 listed July 2017

### Following

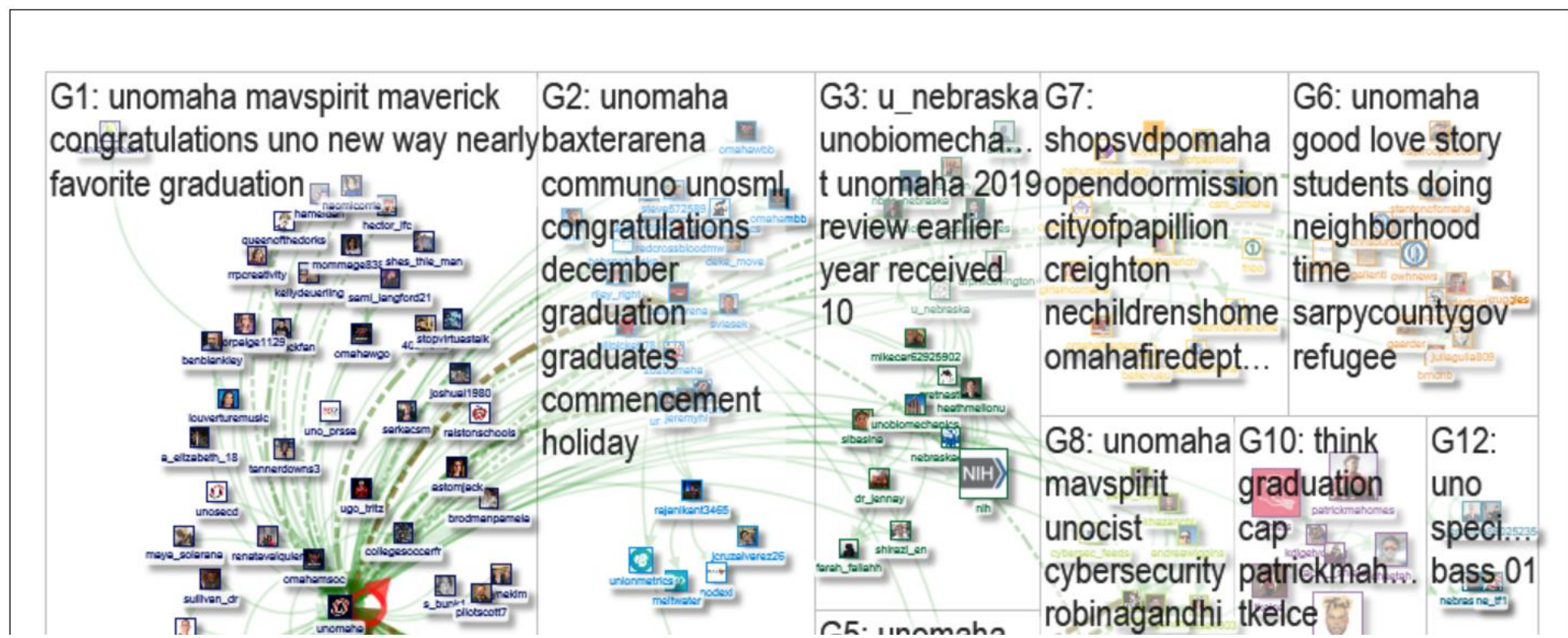
Download & Print

- Raphael jake @Raphaeljake1**  
 3 tweets 416 following 13 followers 0 listed January 2019

# NODE GRAPH GALLERY

A project from the  socialmedia  
 RESEARCH FOUNDATION

[FAQ](#) | [Problem?](#)





# Understand and activate the audiences that matter

[SIGN UP FOR FREE](#)

No credit card needed

By using this website you agree to our [cookie policy](#) ×





Dashboard

Overview >

All Tweets >

Hashtags >

Influencers >

Media >

URL >

@Sweetkat111

Search

Save This Tracker



### Twitter Analytics for @Sweetkat111



28/12/2019 - 28/12/2019

Total Tweets ⓘ

97

Total Contributors ⓘ

75

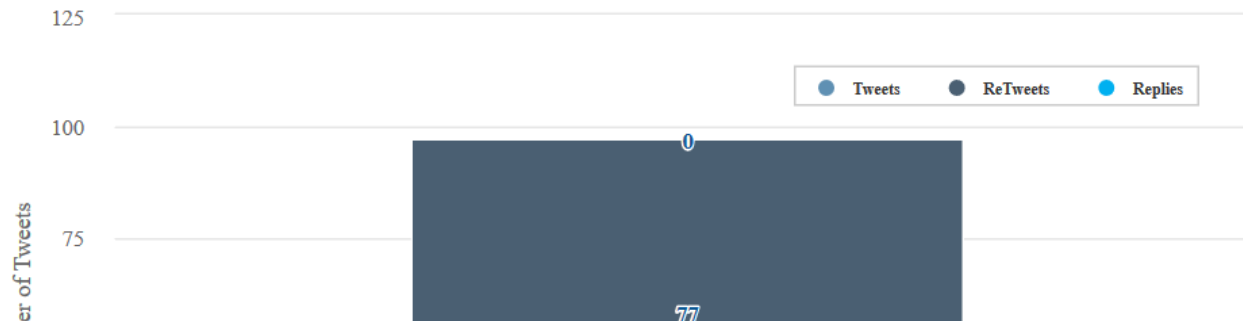
Potential Impressions ⓘ

1520622

Potential Reach ⓘ

895400

### Tweet Timeline ⓘ



### Tweet Timeline Statistics

Total Tweets 97

Av. Tweets per day

Tweets 0 (0.00%)





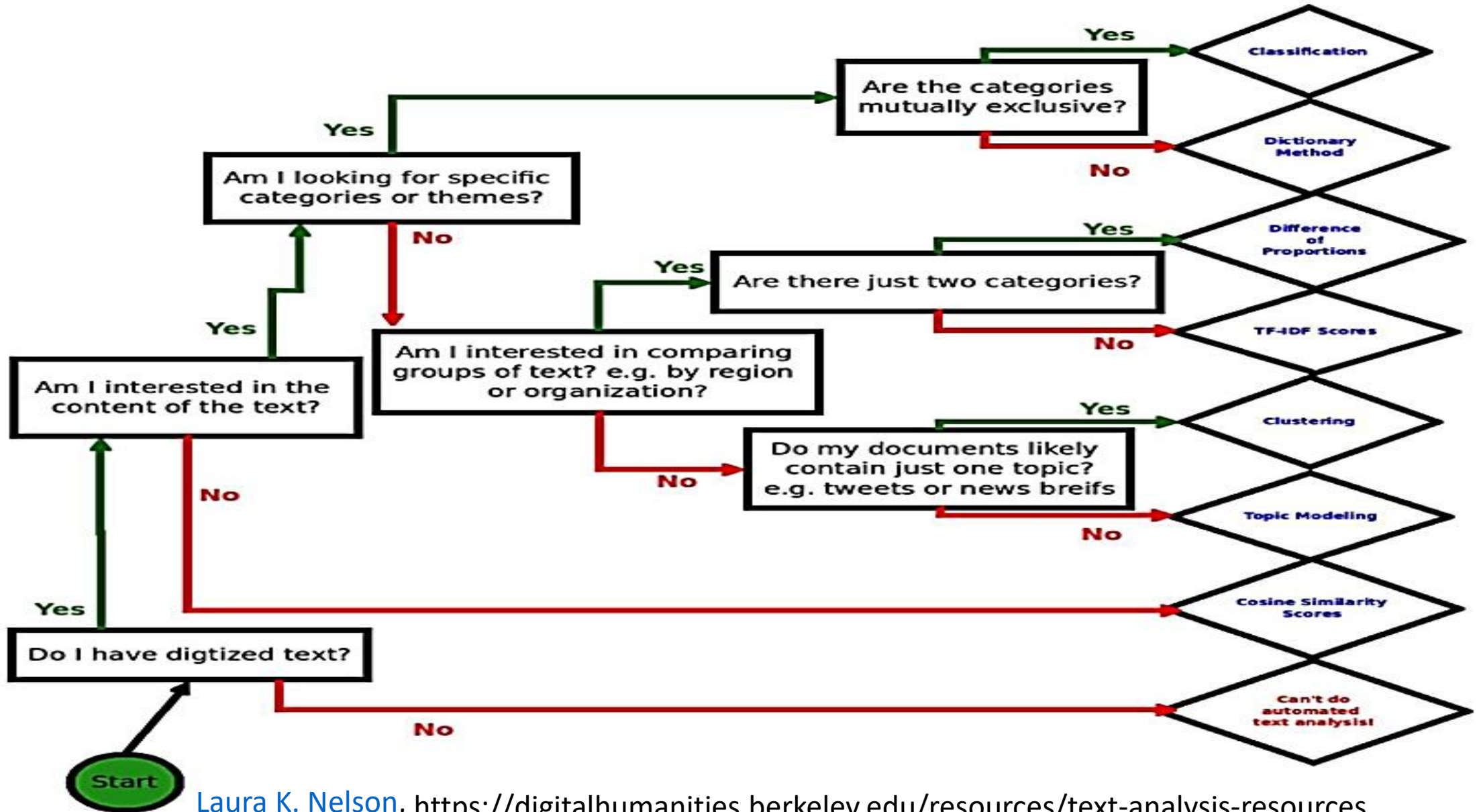
# **SOCIAL MEDIA AND DARK WEB MONITORING**

**What Can Echosec Do For Your Business?**



```
1 import os
2 import sys
3 from tweepy import API
4 from tweepy import OAuthHandler
5
6 def get_twitter_auth():
7     ''' Setup Twitter authentication.
8
9     Return: tweepy.OAuthHandler object
10    '''
11    try:
12        consumer_key = os.environ['TWITTER_CONSUMER_KEY']
13        consumer_secret = os.environ['TWITTER_CONSUMER_SECRET']
14        access_token = os.environ['TWITTER_ACCESS_TOKEN']
15        access_secret = os.environ['TWITTER_ACCESS_SECRET']
16    except KeyError:
17        sys.stderr.write("TWITTER_* environment variables not set\n")
18        sys.exit(1)
19    auth = OAuthHandler(consumer_key, consumer_secret)
20    auth.set_access_token(access_token, access_secret)
21
22 def get_twitter_client():
23     ''' Setup Twitter API client.
24
25     Return: tweepy.API object
26    '''
27    auth = get_twitter_auth()
28    client = API(auth)
29    return client
30
31 client
32 cleint
33 get_twitter_client
```

# WEB SCRAPING AND SOCIAL MEDIA MINING



Laura K. Nelson, <https://digitalhumanities.berkeley.edu/resources/text-analysis-resources>



# AUTOMATED TEXT ANALYSIS

- Classification
- Dictionary Method
- Difference of Propositions
- TF-IDF Scores
- Clustering
- Topic Modelling



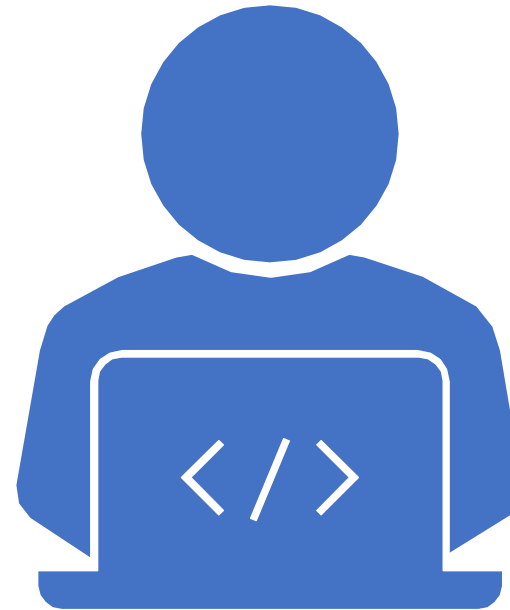
# Textual analysis

- It is a method of data analysis that closely examines either the content and meaning of texts or their structure and discourse (Sharon Lockyer, 2008:885)

Texts	Analysis
a. Standardized. The text used must be the same across all the tools being analysed to ensure that comparisons between tools are possible.	a. Standardized. The methods used should be the same across all texts, to the point where it should be (almost automatic).
b. Open. The texts must be easily accessible and (ideally) free so that others can acquire them should they wish to carry out their own analysis.	b. Replicable. The methods used should be available to anyone who wants them as recipes including both instructions and scripts.
c. Varied. More than one type of text should be included and these should be from a range of sources generally of interest within and across the humanities	c. Varied. More than one type of analysis should be included to ensure that a wide user base across the humanities will be interested in the results.

# Computer-Assisted Textual Analysis Technologies

- According to Rockwell and Sinclair, the foundation of computer-assisted interpretation is formalization.
  - Rockwell, Geoffrey and Sinclair, Stefan, *Hermeneutica: Computer-Assisted Interpretation in the Humanities*, The MIT Press, Cambridge, 2016
- For T.E Clement, CATA Technologies provide new ways of studying texts in the humanities. They extract, count, stem, parse, name, disambiguate, collocate, contextualize, and visualize words as list, line graph, or histogram
  - (Clement, T. E. “A Thing Not Beginning and Not Ending”: Using Digital Tools to Distant-Read Gertrude Stein’s *The Making of Americans*’. *Literary and Linguistic Computing*, vol. 23, no. 3, Sept. 2008, pp. 361–81.)





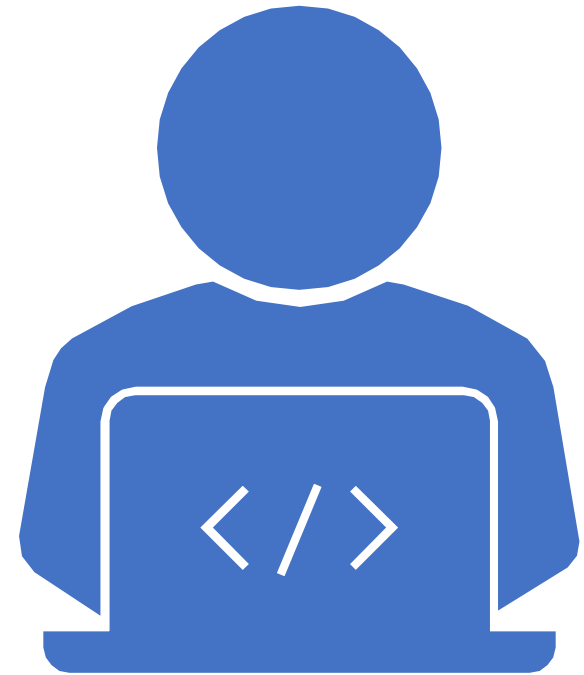
## Computer-Assisted Textual Analysis Technologies

- **Text analysis systems can search large texts quickly.**
- **Text analysis systems can conduct complex searches.**
- **Text analysis systems can present the results in ways that suit the study of texts.**



# Textual Analysis for all disciplines in the Humanities

- Rockwell and Sinclair explained that corpus linguistics and humanists are deploying computer-assisted textual analysis tools in different ways.
- linguistic tools are often designed to parse a sentence and generate a formal representation of meaning,
- the humanist tools aim at the interpretation of human expressions and not in language (18).
- In other words, even when the same corpora are being used by both humanists and linguists, their methods and goals differ (19).
  - Rockwell, Geoffrey and Sinclair, Stefan, *Hermeneutica: Computer-Assisted Interpretation in the Humanities*, The MIT Press, Cambridge, 2016





## Techniques used for the textual interpretation

- protocol analysis,
- procedural task analysis,
- proximity analysis
- Textual entailment recognition
- Correspondence Analysis
- Diachronic Analysis
- Distant Reading

# Techniques used for the textual interpretation

- Keyword based analysis
- Interpretive Phenomenological Analysis
- Narrative Analysis
- Map analysis
- Conversational analysis
- Semiotic



# DH Scholars and CATA

Franco Moretti:  
'Distant Reading'

- *Graphs, maps, trees: abstract models for a literary history* (2005),
- *Atlas of the European Novel, 1800–1900*

Matthew Jockers:  
'Macroanalysis'

The Stanford  
Literary Lab (Lit Lab)

- hypothesis-testing
- computational modelling
- quantitative analysis





# DISTANT READING

a technique that allows us to extract meaning and think along with machine by searching, sorting, grouping, and comparing texts within large textual data

According to M.G. Phillips et al, distant reading is a term that 'embraces many practices including data mining, aggregation, text analysis, and the visual representation of these practices in the forms of charts and graphs' (Phillips et. al. 1731).

# DH TOOLS FOR TEXTUAL ANALYSIS

**Voyant:** A web-based, dashboard-style tool that allows users to upload a corpora and visualize patterns in various ways.



**TAPoR Portal:** It is not a digital humanities tool, but a portal to find curated lists of text analysis tools commonly used in the field.



**Methodica:** A Collection of Digital Text Methods. Source: <http://www.methodi.ca>



**Python or R:** A high level programming language commonly used in text analysis tools in the digital humanities.

# VOYANT

see through your text

Add Texts 🔍 🔌 ?

Type in one or more URLs on separate lines or paste in a full text.

*Voyant Tools is a web-based reading and analysis environment for digital texts.*



# Word Cloud Function



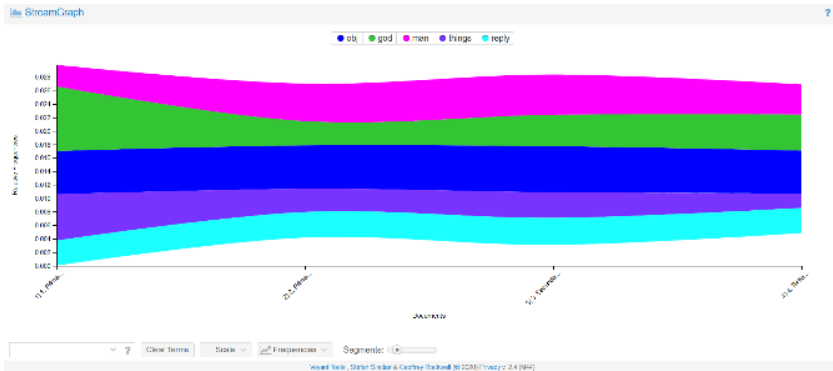


Document	Left	Term	Right
1) 1. Pri...	73. On the Things That Belong to the Seventh Day 74. On All the Seven Days in Common...	man	75. Of Man Who Is Composed of a Spiritual and a Corporeal Substance: and in the First P...
1) 1. Pri...	Things That Belong to the Seventh Day 74. On All the Seven Days in Common TREATISE...	man	Who Is Composed of a Spiritual and a Corporeal Substance: and in the First Place, Conco...
1) 1. Pri...	of the First Man's Body 82. The Production of the Woman 93. The End or Term of the Prod...	man	94. Of the State and Condition of the First Man as Regards His Intellect 95. Of Things Pert...
1) 1. Pri...	Woman 93. The End or Term of the Production of Man 94. Of the State and Condition of th...	man	as Regards His Intellect 95. Of Things Pertaining to the First Man's Will—Namely, Grace a...
1) 1. Pri...	Intellect 95. Of Things Pertaining to the First Man's Will—Namely, Grace and Righteousnes...	man	in the State of Innocence 97. Of the Preservation of the Individual in the Primitive State 98...
1) 1. Pri...	109. The Ordering of the Bad Angels 110. How Angels Act on Bodies 111. The Action of th...	man	112. The Mission of the Angels 113. Of the Guardianship of the Good Angels 114. Of the A...
1) 1. Pri...	Demons 115. Of the Action of the Corporeal Creature 116. On Fate 117. Of Things Perlaini...	man	118. Of the Production of Man from Man As to the Soul 119. Of the Propagation of Man As...
1) 1. Pri...	the Corporeal Creature 116. On Fate 117. Of Things Pertaining to the Action of Man 118. ...	man	from Man As to the Soul 119. Of the Propagation of Man As to the Body _____...
1) 1. Pri...	Creature 116. On Fate 117. Of Things Pertaining to the Action of Man 118. Of the Producti...	man	As to the Soul 119. Of the Propagation of Man As to the Body _____...
1) 1. Pri...	Action of Man 118. Of the Production of Man from Man As to the Soul 119. Of the Propaga...	man	As to the Body _____ PROLOGUE Because the Master of Catholic ...
1) 1. Pri...	Doctrine Is Required? Objection 1: It seems that, besides philosophical science, we have ...	man	should not seek to know what is above reason: "Seek not the things that are too high for t...
1) 1. Pri...	that there should be a knowledge revealed by God besides philosophical science built up ...	man	is directed to God, as to an end that surpasses the grasp of his reason: "The eye hath not ...
1) 1. Pri...	men who are to direct their thoughts and actions to the end. Hence it was necessary for th...	man	that certain truths which exceed human reason should be made known to him by divine re...
1) 1. Pri...	by divine revelation. Even as regards those truths about God which human reason could h...	man	should be taught by a divine revelation; because the truth about God such as reason coul...
1) 1. Pri...	learned through revelation. Reply Obj. 1: Although those things which are beyond man's k...	man	through his reason, nevertheless, once they are revealed by God, they must be accepted ...
1) 1. Pri...	be accepted by faith. Hence the sacred text continues, "For many things are shown to the...	man	" (Ecclesi. 3:25). And in this, the sacred science consists. Reply Obj. 2: Sciences are differ...
1) 1. Pri...	not indeed, in its material aspect, but as regards the precise formality under which it is an ...	man	, ass, stone agree in the one precise formality of being colored; and color is the formal obj...

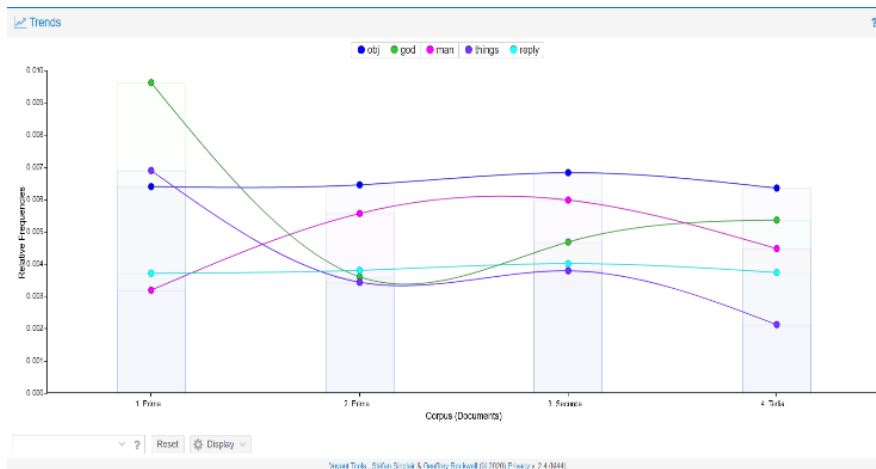
## Keyword in Context (KWIC) Concordance Function

# Textual Visualization Function

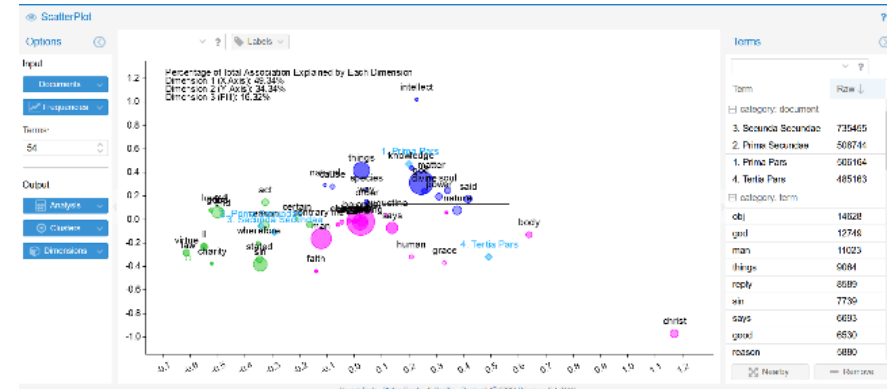
StreamGraph: A visualization that depicts the change of the frequency of words in a corpus (or within a single document)



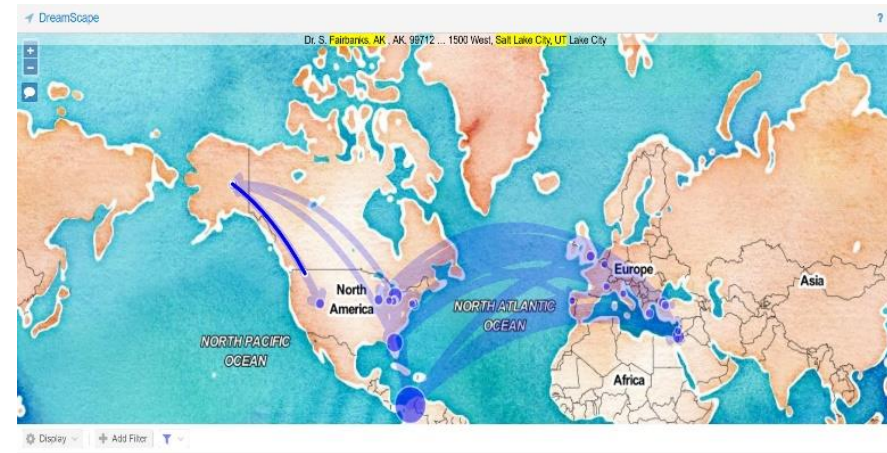
Trends: A line graph depicting the distribution of a word's occurrence across a corpus or document.



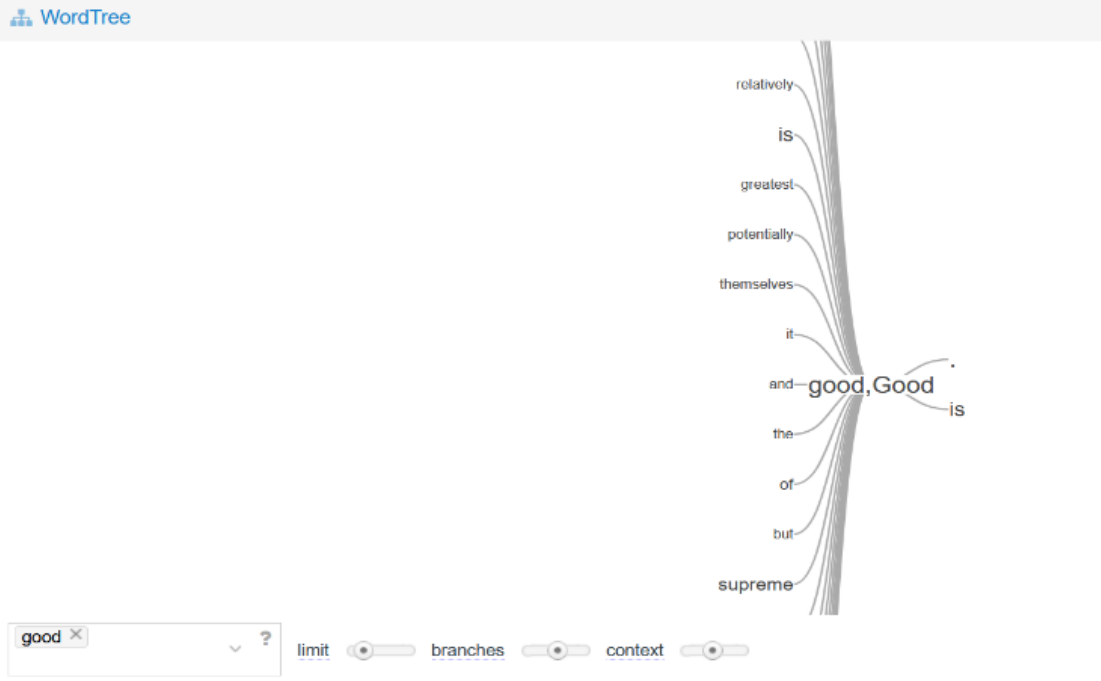
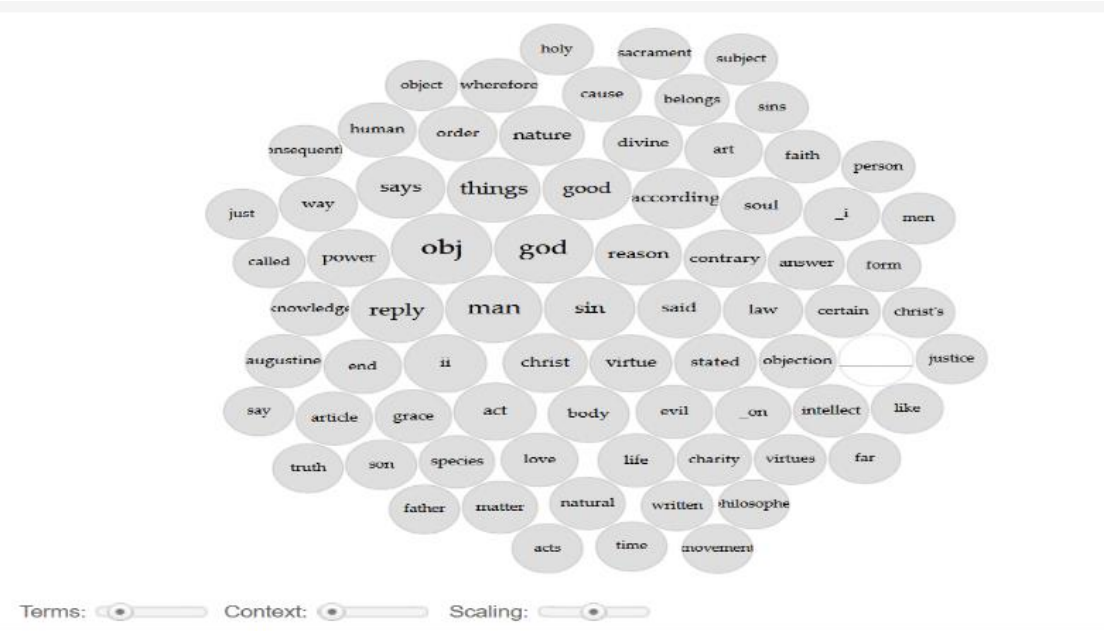
ScatterPlot: A graph visualization of how words cluster in a corpus using document similarity, correspondence analysis or principal component analysis



Dreamscape Tool: An experimental tool for exploring geospatial aspects of texts.



# Textual Visualization Function



TermsBerry Tool: tool provides (words that occur in proximity). Word Tree Tool: A tool that allows the user to explore how a way of exploring high frequency terms and their collocates words are used in phrases

# Tables

**Document Terms**

#	Term	Count	Relative	Trend
1	god	4867	9,815	
4	christ	4318	8,800	
1	things	3491	6,897	
3	obj	5023	6,830	
2	obj	3283	6,453	
1	obj	3239	6,399	
4	obj	3083	6,355	
3	man	4400	5,583	
2	man	2833	5,569	
4	god	2603	5,365	
2	good	2729	5,364	
2	sin	2656	5,221	
2	reason	2484	4,883	
3	god	3445	4,684	
3	sin	3382	4,599	
2	law	2317	4,554	
4	man	2174	4,481	
4	sacrament	2098	4,320	

**Document Terms Tool:** A table view of document term frequencies

**Collocates**

Term	Collocate	Count (context)
reply	obj	8403
god	man	1070
man	god	1059
says	contrary	947
says	_on	916
sin	obj	875
things	god	868
god	obj	846
god	things	840
obj	says	832
says	obj	832
good	evil	827
god	god	825
sin	sin	782
obj	god	686
man	obj	675
good	good	657
says	ii	632

**Corpus Collocates Tool:** A table view of which terms appear more frequently in proximity to keywords across the entire corpus

**Phrases**

Term	Count	Length	Trend
for it is written ex 12 48 if any stranger be willing to dwell among you and to keep the phrase of the lord all his males shall first be circumcised and then shall he celebrate...	2	48	
this ebook is for the use of anyone anywhere at no cost and with almost no restrictions whatsoever you may copy it give it away or re use it under the terms of the proje...	8	45	
_on the contrary the apostle says col 2 16 17 let no man judge you in meat or in drink or in respect of a festival day or of the new moon or of the sabbaths which are a s...	2	44	
jer 18 7 i will suddenly speak against a nation and against a kingdom to root out and to pull down and to destroy it if that nation against which i have spoken shall repe...	2	43	
augustine says de doct: christ 1 28 since one cannot do good to all we ought to consider those chiefly who by reason of place time or any other circumstance by a kind o...	2	41	
if a man while governing the people in his church under the bishop and leading a secular life is inspired by the holy ghost to cease to work out his salvation in a monast...	2	39	
vit qui i can hoc nequequam the holy synod orders that any man who has been degraded from the episcopal dignity to the monastic life and a place of repentance shoul...	2	39	
christ on his right hand in the heavenly places above all principaty and power and virtue and dominion and every name that is named not only in this world but also in l...	2	37	
the apostle says rom 5 15 if by the offense of one many died much more the grace of god and the gift by the grace of one man jesus christ hath abounded unto many b...	2	37	
eph 4 13 until we all meet into the unity of faith and of the knowledge of the son of god unto a perfect man unto the measure of the age of the fulness of christ	2	36	
if no one imposes this burden upon us we must devote ourselves to the research and contemplation of truth but if it be imposed on us we must bear it because charity d...	2	36	
to me the least of all the saints is given this grace to enlighten all men that they may see what is the dispensation of the mystery which hath been hidden from eternity in...	2	35	
what means this way of addressing him save that though he knew that the son of god was to come yet he did not think that he had come in the weakness of the flesh	2	35	
isaias wishing to be of profit to his neighbor by means of the active life desired the office of preaching whereas jeremias who was fain to hold fast to the love of his creator	2	34	
the apostle says 1 tim 5 6 if any man have not care of his own and especially of those of his house he hath denied the faith and is worse than an infidel	2	34	
according to rom 3 24 25 being justified freely by his grace through the redemption that is in christ jesus whom god hath proposed to be a propitiation through faith in his...	2	33	
let him that hath understanding beware lest he withhold his knowledge let him that hath abundance of wealth watch lest he slacken his merciful bounty let him who is a s...	2	33	
_on the contrary our lord said john 3 5 unless a man be born again of water and the holy ghost he cannot enter into the kingdom of god _i answer that	2	32	

**Phrases Tool:** A table that shows repeating sequences of words organized by frequency of repetition or number of words in each repeated phrase

**Terms**

Term	Count	Trend
1 obj	14628	
2 god	12749	
3 man	11023	
4 things	9064	
5 reply	8589	
6 sin	7730	
7 says	6693	
8 good	6530	
9 reason	5880	
10 according	5714	
11 nature	5295	
12 christ	5056	
13 said	4693	
14 i	4456	
15 act	4351	
16 virtue	4310	
17 power	4306	
18 stated	4277	

**Corpus Terms Tool:** A table view of term frequencies in the entire corpus

**Documents**

Title	Words	Types	Ratio	Words/Sentence
1. Prima Pars	506...	8,425	2%	24.7
2. Prima Secundae	508...	9,904	2%	25.2
3. Secunda Secundae	735...	13,...	2%	25.8
4. Tertia Pars	485...	10,...	2%	26.9

**Documents:** A table of the documents in the corpus, including functionality for modifying the corpus





# Textual Formalization

**Summary**

This corpus has 4 documents with 2,235,526 total words and 19,140 unique word forms. Created about 10 hours ago.

**Document Length:**

- Longest: 3. Aquinas\_Summa Theologi... (735455); 2. Aquinas\_Summa Theologi... (508744)
- Shortest: 4. Aquinas\_Summa Theologi... (485163); 1. Aquinas\_Summa Theologi... (506164)

**Vocabulary Density:**

- Highest: 4. Aquinas\_Summa Theologi... (0.021); 2. Aquinas\_Summa Theologi... (0.019)
- Lowest: 1. Aquinas\_Summa Theologi... (0.017); 3. Aquinas\_Summa Theologi... (0.018)

**Average Words Per Sentence:**

- Highest: 4. Aquinas\_Summa Theologi... (26.9); 3. Aquinas\_Summa Theologi... (25.8)
- Lowest: 1. Aquinas\_Summa Theologi... (24.7); 2. Aquinas\_Summa Theologi... (25.2)

Most frequent words in the corpus: **ob** (14628); **god** (12749); **man** (11023); **things** (9064); **reply** (8589)

Distinctive words (compared to the rest of the corpus):

1. Aquinas_Summa Theologi...	710	actually	65	notional	61	subsisting	109	procession	210
2. Aquinas_Summa Theologi...	746	venial	282	volution	37	ceremonial	154	judicial	127
3. Aquinas_Summa Theologi...	1,112	religions	616	void	511	possibilities	87	reviling	83
4. Aquinas_Summa Theologi...	644	sanctify	179	hypostasis	176	suppositum	143	mael	66

**Summary Tool:** A tool that provides a simple, textual overview of the current corpus, including information about words and documents

**Phrases**

Term	Count	Length	Trend
for it is written ex 12 46 if any stranger be willing to dwell among you and to keep the phrase of the lord at his mace shall first be circumcised and then shall he celebrate...	2	48	
this ebook is for the use of anyone anywhere at no cost and with almost no restrictions whatsoever you may copy it give it away or re-use it under the terms of the proe...	8	45	
on the contrary the apostle says col 2 16 7 let no man judge you in meat or in drink or in respect of a festival day or of the new moon or of the sabbaths which are a s...	2	44	
let 19 7 8 will sincerely speak against a nation and against a kingdom to root out and to pull down and to destroy it if that nation against which i have spoken shall free...	2	43	
augustine says de civit crist 1 26 since one cannot do good to all we ought to consider those chiefly who by reason of place time or any other circumstance by a kind o...	2	41	
if a man while governing the people in his church under the bishop and leading a secular life is inspired by the holy ghost to desire to work out his salvation in a monast...	2	39	
all qui can hoc requiretur the holy spirit orders that any man who has been degraded from the episcopal dignity to the monastic life and a piece of repentance about...	2	39	
christ on his right hand in the heavenly places above all principality and power and virtue and dominion and every name that is named not only in this world but also in t...	2	37	
the apostle says rom 5 15 if by the offense of one manny died much more the grace of god and the gift by the grace of one man jesus christ hath abounded unto many bu...	2	37	
epi 4 13 until we all meet in the unity of faith and of the knowledge of the son of god into a perfect man unto the measure of the age of the fulness of christi...	2	38	
if one express his trust upon us we must decide ourselves to the research and contemplation of truth but if it be imposed on us we must bear it because charity d...	2	38	
so me the seed of all the saints is given this grace to enlighten all men that they may see what is the disposition of the mystery which hath been hidden from eternity in...	2	35	
what means the way of addressing him says that though he knew that the son of god was to come yet he did not think that he had come in the weakness of the flesh...	2	35	
leave wishing to be of profit to his neighbor by means of the active life desired the office of preaching whereas jonas who was fain to hold fast to the love of his creator...	2	34	
the apostle says 1 tim 5 8 if any man have not care of his own and especially of those of his house he hath denied the faith and is worse than an infidel...	2	34	
according to rom 3 25 25 being justified freely by his grace through the redemption that is in christ jesus whom god hath prepared to be a propitiation through faith in hi...	2	33	
let him that hath understanding beware lest he without his knowledge let him that hath abundance of wealth watch lest he slacken his merciful bounty let him who is a s...	2	33	
on the contrary our lord said john 3 5 unless a man be born again of water and the holy ghost he cannot enter into the kingdom of god i answer that	2	32	

**Phrases Tool:** A table that shows repeating sequences of words organized by frequency of repetition or number of words in each repeated phrase

**Topics**

Topic

person procession persons procession providence father trinity form relation number

god reason things infinite good evidence goodness form seen time

god divine reply creatures essence way sense art

cause things knowledge action stated natural means effects just greater

bodies demons heavenly angelic art good augustine miracles held sensitive

man certain human objection make active philosopher reason things

power divine work matter nature saint angels way according certain

angels order angel mon orders ob corporeal says man god

said reply called word divine regards existing better substance art

say reason love said private augustine word form taken

angels nature heaven angel natural day love created beatitude good

intellect knowledge soul things understand act intelligible power understands known

act body obj corporeal light said spiritual end says augustine

man image state god innocence order government paradise divine soul

soul body power powers intellect form intellectual obj reason sensitive

angel angels evil movement good world form said time nature

according things obj contrary substance order proper agent follows subject

grace natural answer obj second creaturas belong

on particular

Search Terms Topics run 100 iterations (100 iterations done) adapted from David Mimno's pt LDA

**Topic tool:** A tool that provides a rudimentary way of generating term clusters from a document or corpus and then seeing how each topic (term cluster) is distributed across the document or corpus

**Correlations**

Term 1	Term 2	Correlation...	Significanc...
obj	man	0.6903461	0.30965385
contrary	man	0.03323988	0.78169996
stated	man	0.025437523	0.93686256
on	man	0.024702208	0.75354314
j	man	0.024525820	0.7582012
faith	man	0.024310667	0.7632295
according	man	0.022836939	0.89482373
virtue	man	0.022259941	0.8659394
species	man	0.020800117	0.7889946
reason	man	0.020275462	0.9122958
life	man	0.018851128	0.82936287
natural	man	0.018123966	0.81344837
grace	man	0.017833043	0.8343481
evil	man	0.016389013	0.8520661
answer	man	0.016340835	0.84162635
article	man	0.016176614	0.8452834
augustine	man	0.01583744	0.85055935

man 100 minimum coverage (%100) Scale

**Correlations tool:** A tool that enables an exploration of the extent to which term frequencies vary in sync (terms whose frequencies rise and fall together or inversely)

**Veliza**

Hello, I'm Veliza, and I'm here to talk to you about your texts (you may know my sister Eliza she's a famous psychotherapist, I'm just learning to talk about text documents, but please, let me know about any anxieties you're feeling about your texts. Type a message in the box below and hit enter. Or, if you're feeling playful, hit the from text section in the lower right-hand corner to fetch a sentence from the corpus.

That's quite interesting.

Don't any other reasons come to mind?

Are they always derived from various forms are predicated of one another?

Thinly, because teachers judge in the abstract and of future events, whereas those who sit in judgment judge of things present, towards which they are affected by love, hatred, or some

**Veliza tool:** An experimental tool that enables the user to have a (limited) natural language exchange (in English) based on the corpus.

**Reader**

**1. Prima Pars**

Project Gutenberg's Summa Theologica, Part I (Prima Pars), by Thomas Aquinas

This ebook is for the use of anyone anywhere at no cost and with almost no restrictions whatsoever. You may copy it, give it away or re-use it under the terms of the Project Gutenberg License included with this ebook or online at www.gutenberg.org

Title: Summa Theologica, Part I (Prima Pars)

From the Complete American Edition

Author: Thomas Aquinas

Translator: Fathers of the English Dominican Province

Release Date: January 26, 2006 [EBook #17611]

Language: English

\*\*\* START OF THIS PROJECT GUTENBERG EBOOK SUMMA THEOLOGICA, PART I \*\*\*

Produced by Sandra K. Perry, with corrections and supplementation by David McClamrock

ST. THOMAS AQUINAS

SUMMA THEOLOGICA

PART I ("Prima Pars")

Translated by

Fathers of the English Dominican Province

BENZIGER BROTHERS

NEW YORK

**Reader Tool:** A tool that provides a way of reading documents in the corpus, text is fetched as needed.

# Voyant Tools

a free, user-friendly, extensively documented, suite of web-based environment, and a collection of tools built on existing techniques for textual analysis in literary studies.

used to visualize texts (using Cirrus, Bubblelines, Knots, Mandala tools); to generate trees (using Word Tree Tool),

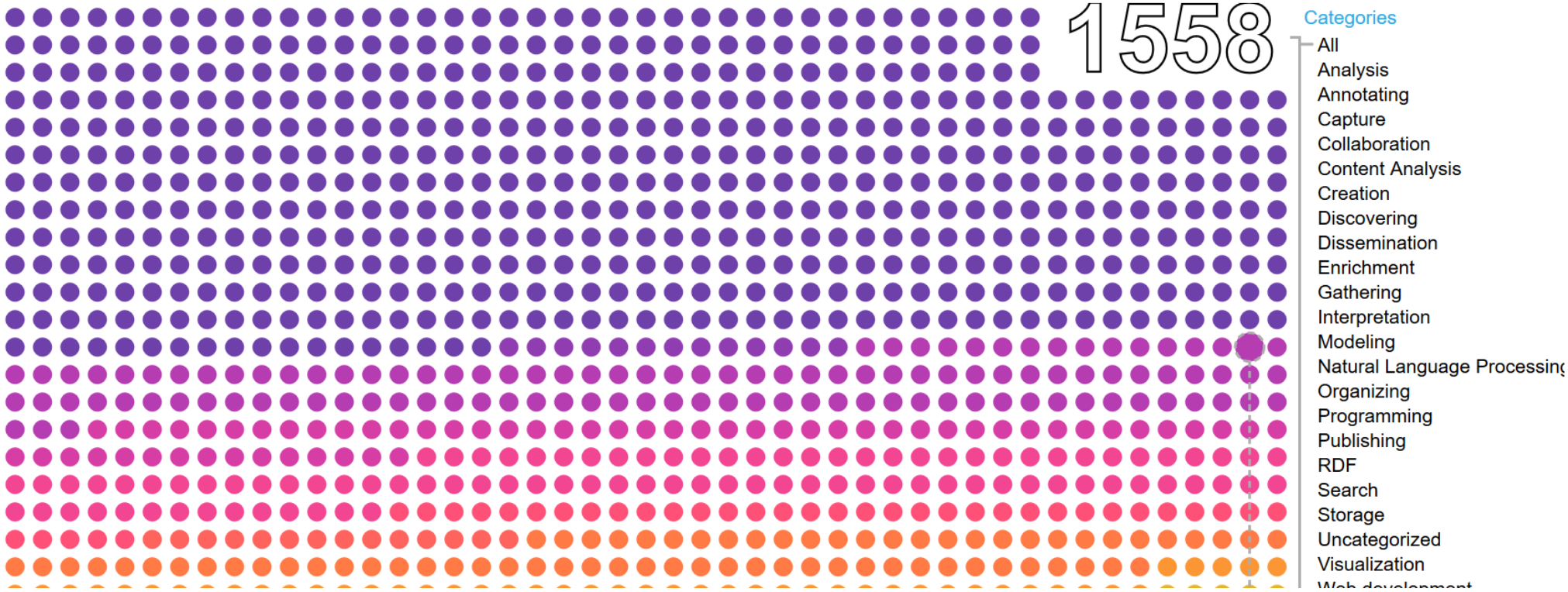
graphs (using ScatterPlot, Collocates Graph, StreamGraph, Trends-line graph tools),

tables (using Corpus Collocates, Keywords, Document Terms, Corpus Terms, Documents, and Phrases tools);

to quantitatively analyse texts by performing Word Count (using Summary Tools), Keywords in Context (Contexts tool), Correlation (Correlations tool) and Topic Modelling (Topic tool);

It has a unique tool called 'Veliza' – an experimental tool which can be used for Natural Language Exchange in English (Stéfan et al., 2012)

Browse the TAPoR collection:

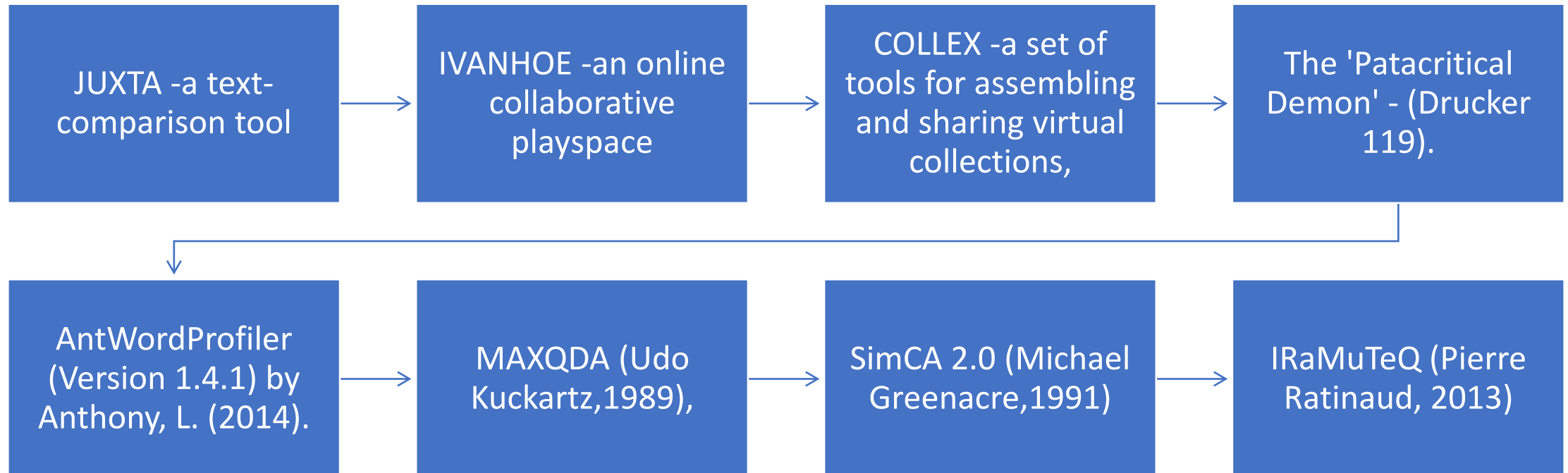


# TOOLS

Rockwell, Geoffrey.Uszkalo, Kirsten C. Sinclair,Stéfan.  
TAPoR 3: Discover research tools for studing texts Source:  
<http://www.tapor.ca>

- Presently, there are over 1,500 technological tools – as collated in TAPoR 3 collection (Rockwell et al., 2020) being used by scholars in the humanities in order to make meaning out of the enormous textual data available online and offline.

# Other Tools like Voyant





Learn how to identify the locations named in a corpus and generate a map of the results.

### USER LOGIN

Username \*

Password \*

[Request new password](#)

Log in

### LATEST RECIPES

[Sentiment Analysis: SVM vs. Naive Bayes](#)

This recipe will compare two machine learning approaches to see which is more likely to give an accurate analysis of sentiment...



## INTRODUCTION

---

This recipe will compare two machine learning approaches to see which is more likely to give an accurate analysis of sentiment. Both approaches analyse a corpora of positive and negative [Movie Review data](#) by training and thereafter testing to get an accuracy score. The techniques are [Support Vector Machines \(SVM\)](#) and [Naive Bayes](#). This recipe is based on Jinman Zhang's iPython Notebook found on [TAPoR Github repository](#).

## INGREDIENTS

---

- Python 3
- [sklearn](#)
- [NumPy](#)
- Corpora
  - [Polarity dataset v1.1 \(2.2Mb\)](#)
  - [Polarity dataset v0.9 \(2.8Mb\)](#)
- Provenance: Jinman Zhang's [Notebook](#)

## INTRODUCTION

---

This recipe will show you how to prepare Voronoi diagrams, one way of showing relationships between words in a text and a search term. In order to do this, we will employ the use of word embeddings. These represent individual words in a text as real-valued vectors in a confined vector space. This recipe is based on Kynan Ly's cookbook as seen on this [notebook](#).

## INGREDIENTS

---

- Python 3
- [NLTK](#)
- [Scipy](#)
- [Matplotlib](#)
- [Regular Expressions \(Regex\)](#)
- [Shapely](#)

<http://methodica.ca/recipes/voronoi-diagram-word-embedding>

# Creating a Basic Web Scraper with Python

---

## INTRODUCTION

---

This utility is for creating a simple web scraper with Python.

## INGREDIENTS

---

- Python 3
- A notebook editor such as Jupyter
- Example code from The Art of Literary Text Analysis (forthcoming)

## STEPS

---

- Create a text file listing the url's to scrape from, with each url on a new line
- Create a list of urls to scrape by reading a string from the file and splitting it on the newline character
- Import the urllib.request library
- Try and scrape a page that you know works

<http://methodi.ca/content/creating-basic-web-scraper-python>

# Tokenization using NLTK

---

## INTRODUCTION

---

Tokenization is the process of splitting a sentence or a chunk of text into its constituent parts. These “tokens” may be the letters, punctuation, words, or sentences. They could even be a combination of all these elements. This recipe was adapted from a [Python Notebook](#) written by Kynan Lee.

## INGREDIENTS

---

- Python 3
- The Natural Language Toolkit (NLTK) module
- [NLTK tokenization libraries](#): word\_tokenize, TweetTokenizer, MWETokenizer, sent\_tokenize.
- A text to analyze.
- Sample tokenization code on [TAPoR](#).

<http://methodica.ca/recipes/tokenization-using-nltk>

# Study Vocabulary Change over Time in an Existing Corpus

---

## INTRODUCTION

---

This recipe uses Voyant Tools to explore vocabulary change over time in a corpus of 7 Canadian Throne Speeches (from 2006 to 2013).

## INGREDIENTS

---

- A corpus of seven Canadian Throne speeches spanning over seven years (the speeches were compiled from a [Canadian parliamentary site](#) – only the filenames were modified to better indicate the year and session of the speech)
- A general-purpose text analysis suite like [Voyant Tools](#)

### Questions

1. What terms occur consistently across the seven years of the corpus?
2. What document is most unusual?

<http://methodi.ca/recipes/study-vocabulary-change-over-time-existing-corpus>



# Text Classification by Topic

---

## INTRODUCTION

---

This recipe will show you how to classify text into general topics. We will use supervised machine learning model called Support Vector Machines ([SVM](#)) and [the 20 Newsgroups](#) subject matter data set as the topic classifier. This recipe is based on [Jinman Zhang's Cookbook](#).

## INGREDIENTS

---

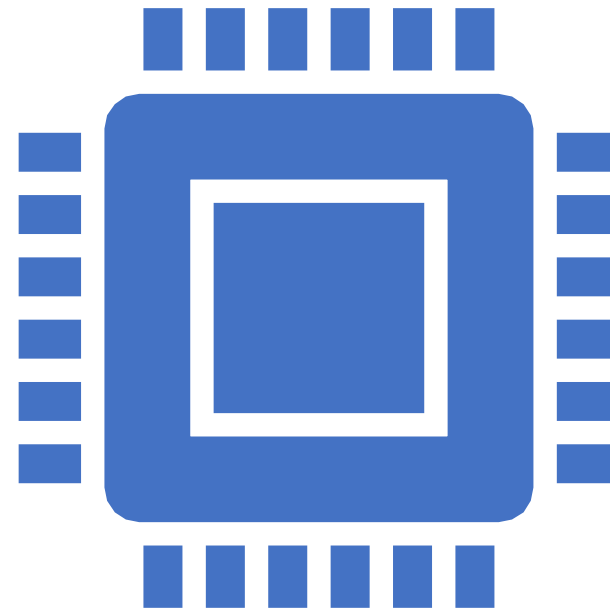
- Python 3
- [20 Newsgroups Dataset](#)
- [sklearn](#)
- [NumPy](#)
- Provenance: [Jinman Zhang's Notebook](#)

## STEPS

<http://methodi.ca/recipes/text-classification-topic>

# Methods, Tools, and Techniques

- **Index Text Files with Python for Rapid Information Retrieval**
- <http://methodi.ca/content/index-text-files-python-rapid-information-retrieval>
- **Counting Word Frequencies with Python(From the Programming Historian)**
- <http://methodi.ca/content/counting-word-frequencies-pythonfrom-programming-historian>
- **Dictionary Based Sentiment Analysis in Python**
- <http://methodi.ca/content/dictionary-based-sentiment-analysis-python>
- **Simple Sentiment Analysis Using Python**
- <http://methodi.ca/recipes/simple-sentiment-analysis-using-python>



# Methods, Tools, and Techniques

- **Multiple Correspondence Analysis and Content Analysis**
- <http://methodi.ca/recipes/multiple-correspondence-analysis-and-content-analysis>
- **Twitter Analysis and Text Research (TATR)**
- <http://methodi.ca/content/twitter-analysis-and-text-research-tatr>
- **Graphing Data in Python**
- <http://methodi.ca/recipes/graphing-data-python>
- **Analyze a Text's Parts of Speech with Python**
- <http://methodi.ca/recipes/analyze-texts-parts-speech-python>
- **Analyzing Repeating Phrases (ngrams) with Python**
- <http://methodi.ca/recipes/analyzing-repeating-phrases-ngrams-python>

# Methods, Tools, and Techniques

- **TATR: Graphing Twitter Data**
- <http://methodi.ca/recipes/tatr-graphing-twitter-data>
- **Build a Social Network Map from a Text – Voyant**
- <http://methodi.ca/recipes/build-social-network-map-text>
- **Compare Texts to Verify Authorship**
- <http://methodi.ca/recipes/compare-texts-verify-authorship>
- **Index Text Files with Python for Rapid Information Retrieval**
- <http://methodi.ca/recipes/index-text-files-python-rapid-information-retrieval>
- **TATR: Panda and CSV of Tweets**
- <http://methodi.ca/recipes/tatr-panda-and-csv-tweets>

# Methods, Tools, and Techniques

- **Scrape Comments from a YouTube Video**
- <http://methodi.ca/recipes/scrape-comments-youtube-video>
- **Explore A Text in a Foreign Language –Voyant**
- <http://methodi.ca/recipes/explore-text-foreign-language>
- **Exploring Word Sense and Tense in a Simple Text - Voyant**
- <http://methodi.ca/recipes/exploring-word-sense-and-tense-simple-text>
- **Identify Simple Themes within a Text – Voayant**
- <http://methodi.ca/recipes/identify-simple-themes-within-text>
- **Explore Themes Within a Text**
- <http://methodi.ca/recipes/explore-themes-within-text>
- **Analyze Blog Discourse – Voyant**
- <http://methodi.ca/recipes/analyze-blog-discourse>



# What is Python all about?

It is a general purpose programming language



It is incredibly powerful.



It has a very large and growing set of external modules (or blocks of code) written by other programmers



It is an **interpreted** language.

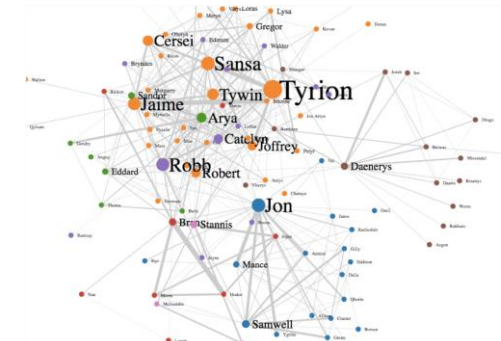
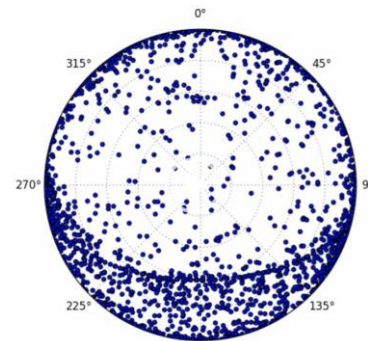
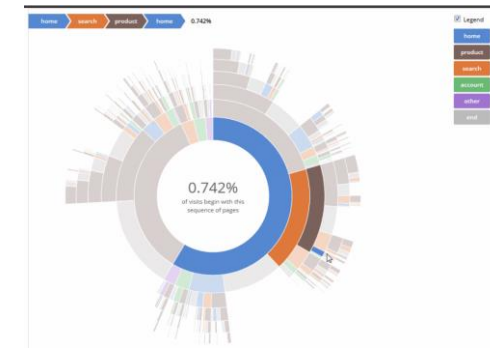
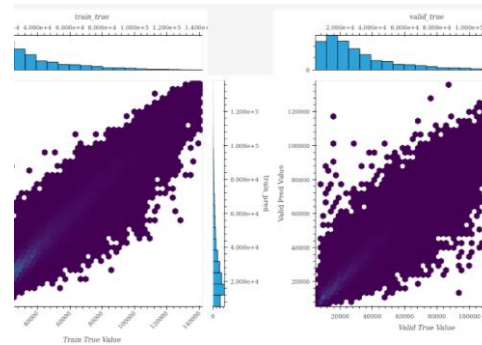
## Twitter API with Python: Part 1 – Streaming Live Tweets

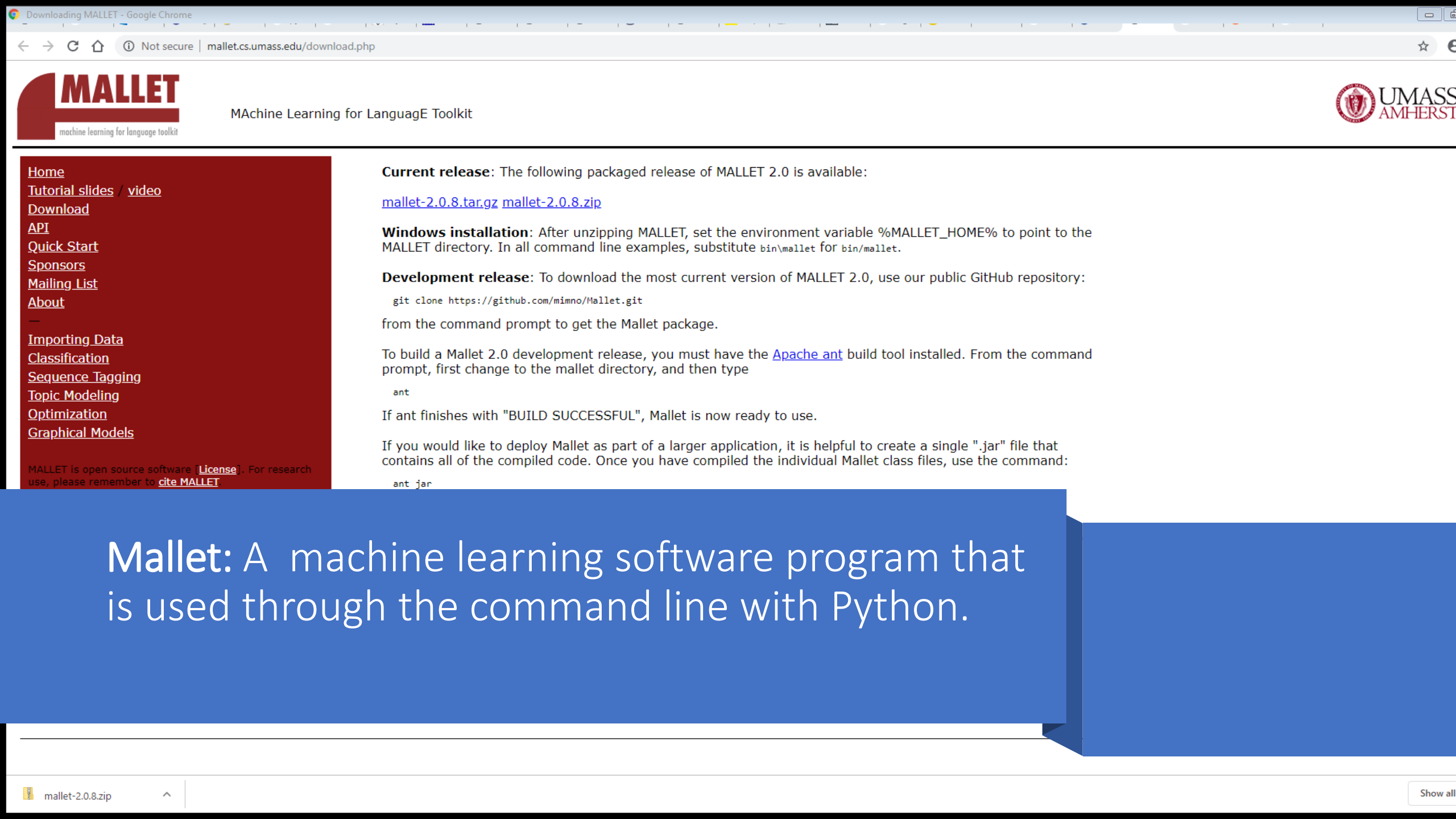
```
20 from tweepy import Stream
21
22 import twitter_credentials
23
24 class StdOutListener(StreamListener):
25
26     def on_data(self, data):
27         print(data)
28         return True
29
30     def on_error(self, status):
31         print(status)
32
33
34 if __name__ == "__main__":
35
36     listener = StdOutListener()
37     auth = OAuthHandler(twitter_credentials.CONSUMER_KEY, twitter_credentials.CONSUMER_SECRET)
38     auth.set_access_token(twitter_credentials.ACCESS_TOKEN, twitter_credentials.ACCESS_TOKEN_SECRET)
39
40     stream = Stream(auth, listener)
41
42     stream.filter(track=['donald trump', 'hillary clinton', 'barack obama', 'bernie sanders'])
```

# Data Mining and Analysis

# Data Visualization

- convert raw data into pictorial representation
- represent the quantitative information in a graphical form.





MALLET  
Machine Learning for Language Toolkit



- [Home](#)
- [Tutorial slides / video](#)
- [Download](#)
- [API](#)
- [Quick Start](#)
- [Sponsors](#)
- [Mailing List](#)
- [About](#)
- 
- [Importing Data](#)
- [Classification](#)
- [Sequence Tagging](#)
- [Topic Modeling](#)
- [Optimization](#)
- [Graphical Models](#)

MALLET is open source software [[License](#)]. For research use, please remember to [cite MALLET](#).

**Current release:** The following packaged release of MALLET 2.0 is available:

[mallet-2.0.8.tar.gz](#) [mallet-2.0.8.zip](#)

**Windows installation:** After unzipping MALLET, set the environment variable %MALLET\_HOME% to point to the MALLET directory. In all command line examples, substitute bin\mallet for bin/mallet.

**Development release:** To download the most current version of MALLET 2.0, use our public GitHub repository:

```
git clone https://github.com/mimno/Mallet.git
```

from the command prompt to get the Mallet package.

To build a Mallet 2.0 development release, you must have the [Apache ant](#) build tool installed. From the command prompt, first change to the mallet directory, and then type

```
ant
```

If ant finishes with "BUILD SUCCESSFUL", Mallet is now ready to use.

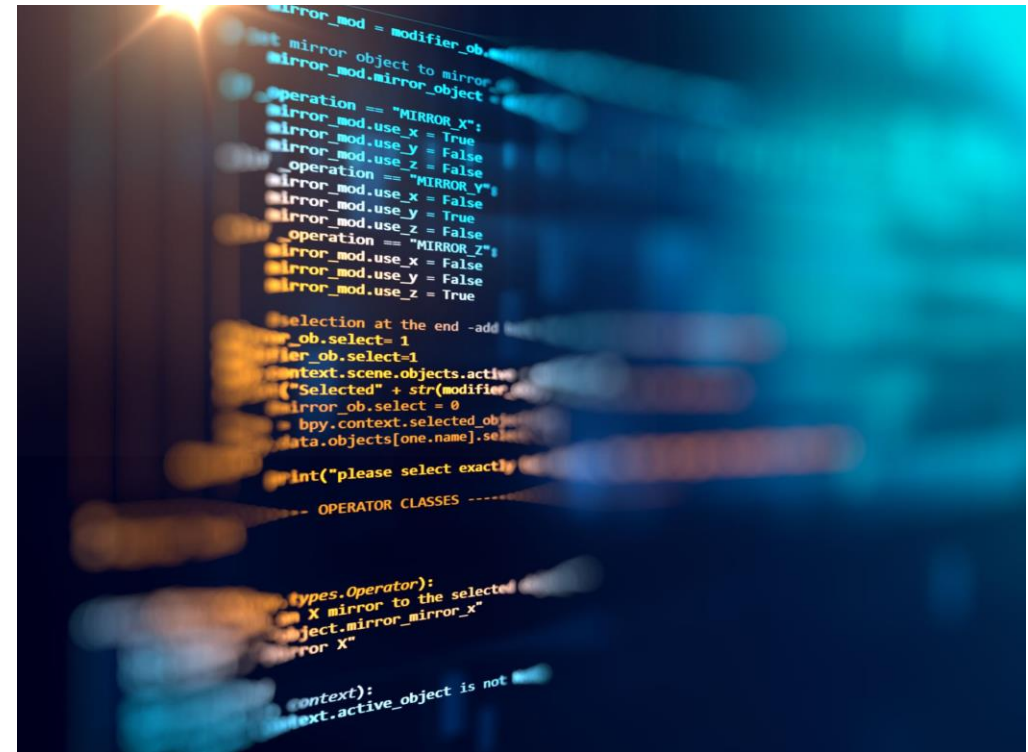
If you would like to deploy Mallet as part of a larger application, it is helpful to create a single ".jar" file that contains all of the compiled code. Once you have compiled the individual Mallet class files, use the command:

```
ant jar
```

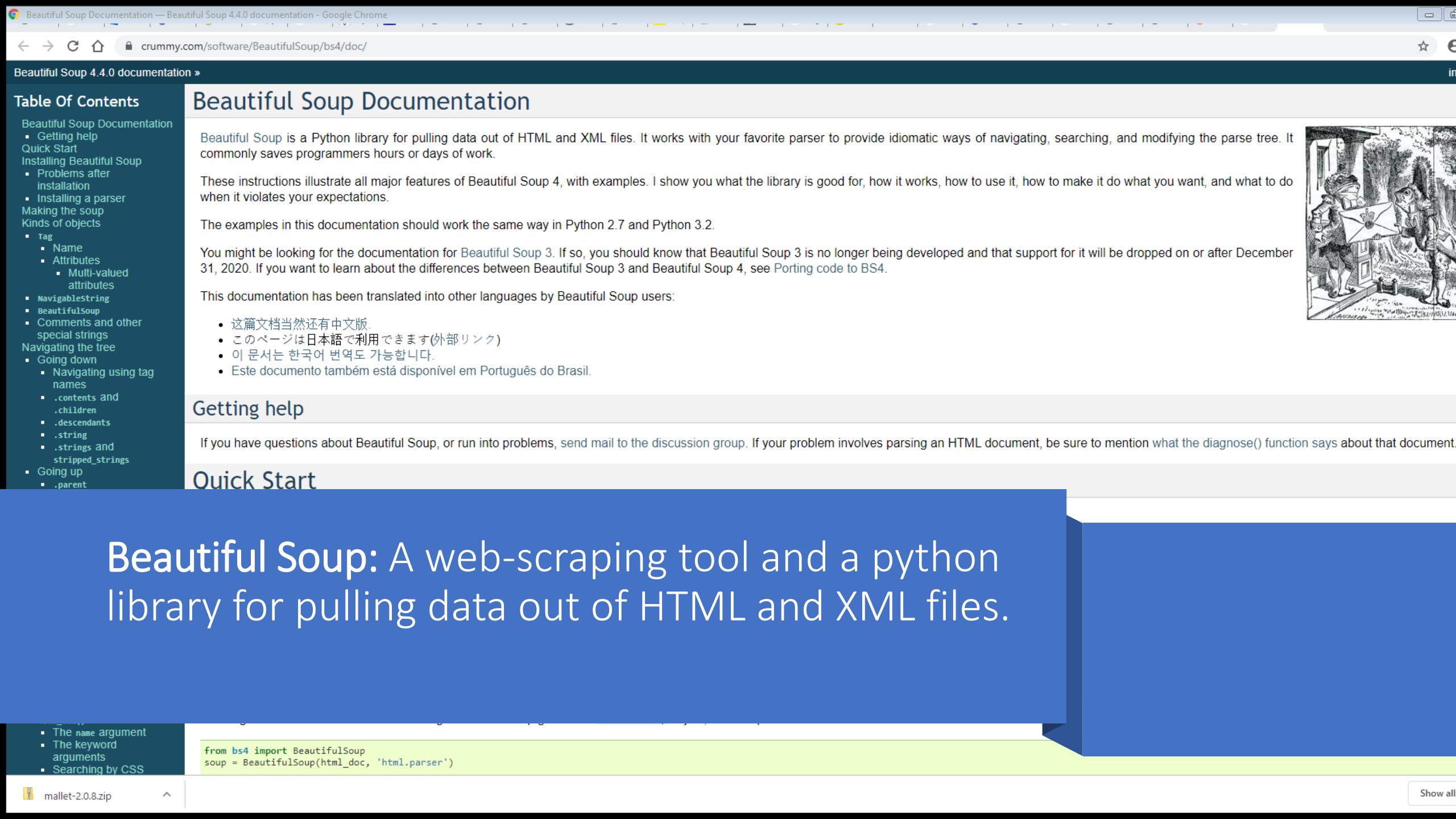
**Mallet:** A machine learning software program that is used through the command line with Python.

# NLTK

- NLTK is a leading platform for building Python programs to work with human language data. It provides easy-to-use interfaces to over 50 corpora and lexical resources such as WordNet, along with a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning.







## Table Of Contents

- Beautiful Soup Documentation
  - Getting help
- Quick Start
- Installing BeautifulSoup
  - Problems after installation
  - Installing a parser
- Making the soup
- Kinds of objects
  - Tag
    - Name
    - Attributes
      - Multi-valued attributes
  - NavigableString
  - BeautifulSoup
  - Comments and other special strings
- Navigating the tree
  - Going down
    - Navigating using tag names
    - .contents and .children
    - .descendants
    - .string
    - .strings and stripped\_strings
  - Going up
    - .parent

# Beautiful Soup Documentation

Beautiful Soup is a Python library for pulling data out of HTML and XML files. It works with your favorite parser to provide idiomatic ways of navigating, searching, and modifying the parse tree. It commonly saves programmers hours or days of work.

These instructions illustrate all major features of BeautifulSoup 4, with examples. I show you what the library is good for, how it works, how to use it, how to make it do what you want, and what to do when it violates your expectations.

The examples in this documentation should work the same way in Python 2.7 and Python 3.2.

You might be looking for the documentation for BeautifulSoup 3. If so, you should know that BeautifulSoup 3 is no longer being developed and that support for it will be dropped on or after December 31, 2020. If you want to learn about the differences between BeautifulSoup 3 and BeautifulSoup 4, see [Porting code to BS4](#).

This documentation has been translated into other languages by BeautifulSoup users:

- 这篇文档当然还有中文版
- このページは日本語で利用できます(外部リンク)
- 이 문서는 한국어 번역도 가능합니다.
- Este documento também está disponível em Português do Brasil.

## Getting help

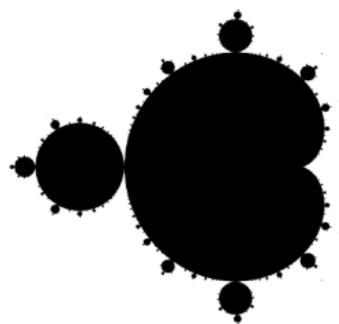
If you have questions about BeautifulSoup, or run into problems, send mail to the discussion group. If your problem involves parsing an HTML document, be sure to mention what the `diagnose()` function says about that document.

## Quick Start

Beautiful Soup: A web-scraping tool and a python library for pulling data out of HTML and XML files.

```
from bs4 import BeautifulSoup
soup = BeautifulSoup(html_doc, 'html.parser')
```

- The name argument
- The keyword arguments
- Searching by CSS



# TextBlob



Star 6,690

TextBlob is a Python (2 and 3) library for processing textual data. It

# TextBlob: Simplified Text Processing

Release v0.15.2. ([Changelog](#))

*TextBlob* is a Python (2 and 3) library for processing textual data. It provides a simple API for diving into common natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more.

```
from textblob import TextBlob
```

```
text = '''
```

```
The titular threat of The Blob has always struck me as the ultimate movie monster: an insatiably hungry, amoeba-like mass able to penetrate virtually any safeguard, capable of--as a doomed doctor chillingly
```

**TextBlob:** A Python (2 and 3) library for processing textual data. It provides a simple API for diving into common natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more.

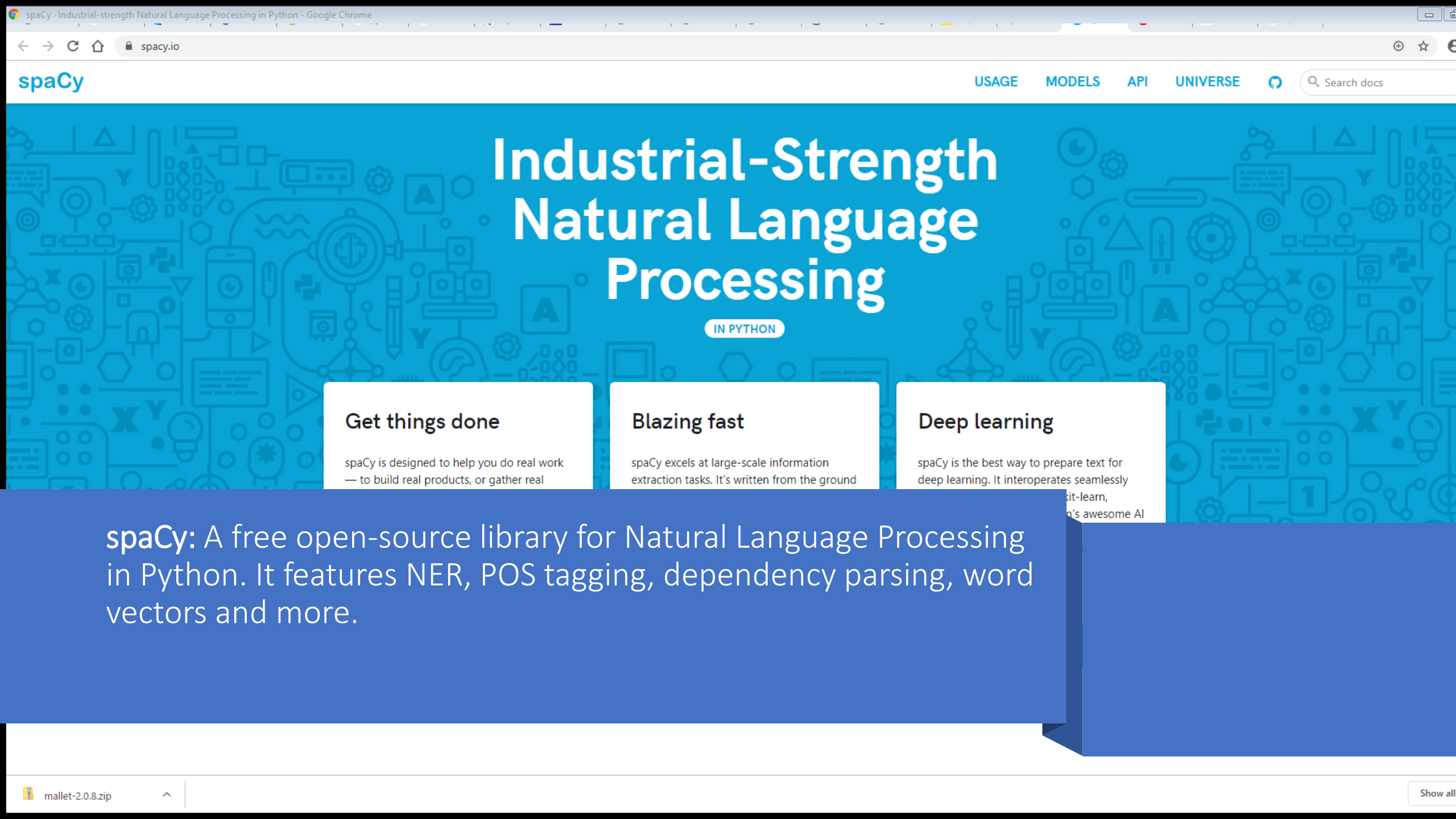
[TextBlob @ PyPI](#)

[TextBlob @ GitHub](#)

```
blob.tags
```

```
# [('the', 'DT'), ('titular', 'JJ'), ('threat', 'NN'), ('of', 'IN'), ...]
```

Fork me on GitHub



# Industrial-Strength Natural Language Processing

IN PYTHON

**Get things done**  
spaCy is designed to help you do real work — to build real products, or gather real

**Blazing fast**  
spaCy excels at large-scale information extraction tasks. It's written from the ground

**Deep learning**  
spaCy is the best way to prepare text for deep learning. It interoperates seamlessly with TensorFlow, Keras, PyTorch, and more. It's awesome AI

**spaCy:** A free open-source library for Natural Language Processing in Python. It features NER, POS tagging, dependency parsing, word vectors and more.

“Digital Humanities is not just about building things but also include those who theorize about building, those who design so that others might build, those who supervise building, and those who are working to re-build systems”

Stephen Ramsay





THANK YOU!