

A Research Tool for Insights in Academic Fields



Contents

- Problem Statement
- Data
- Exploration
- Data Modelling
- Findings
- Further Development

Problem Statement

Academic Research Tool

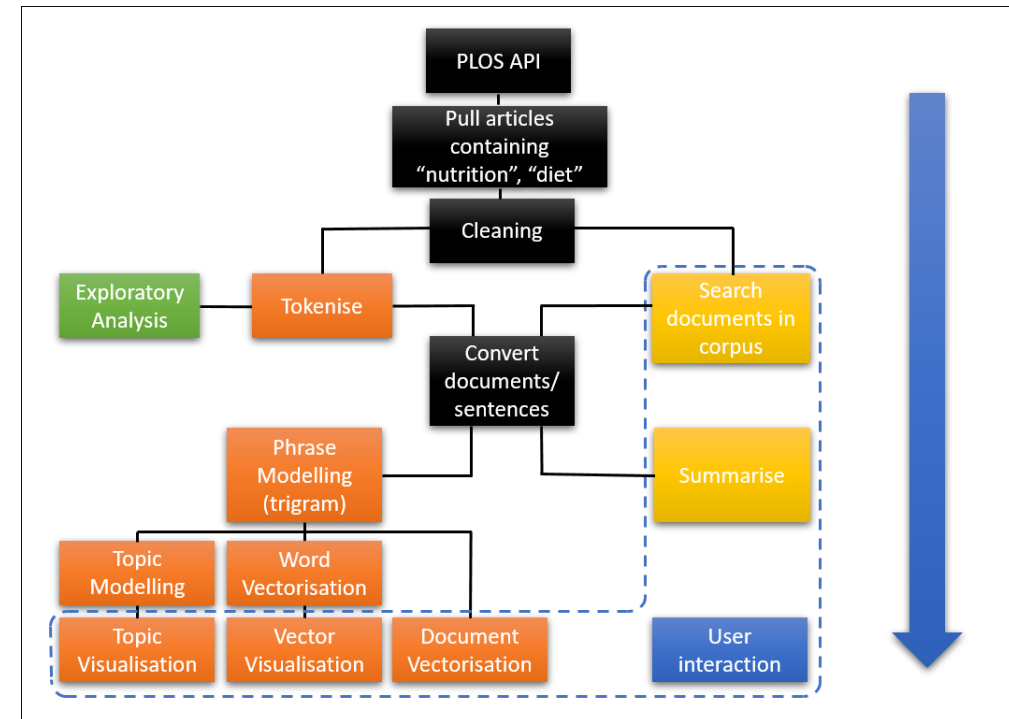




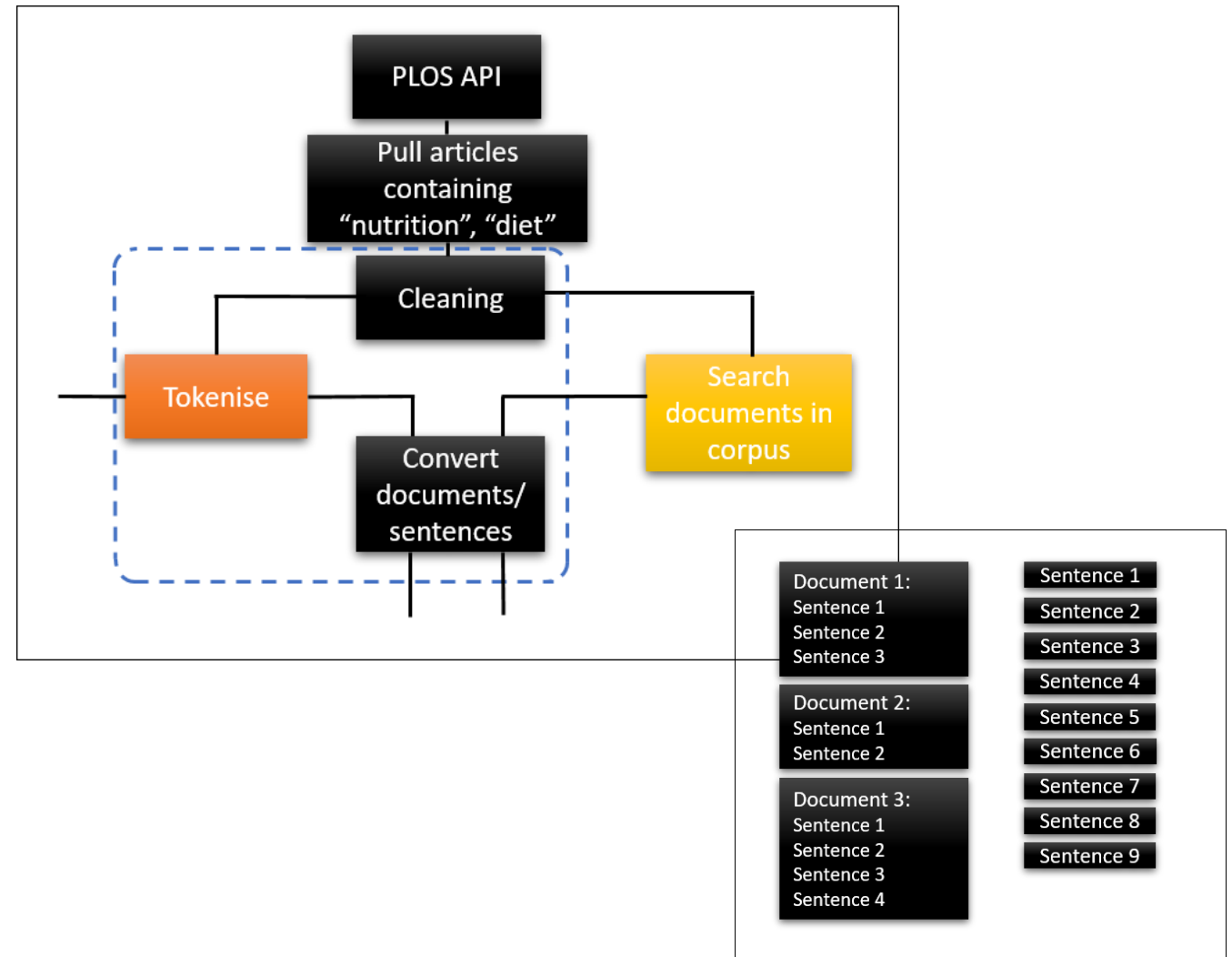
- Opened 2003
- Non-profit open access
- Public Library Of Science
- www.plos.org

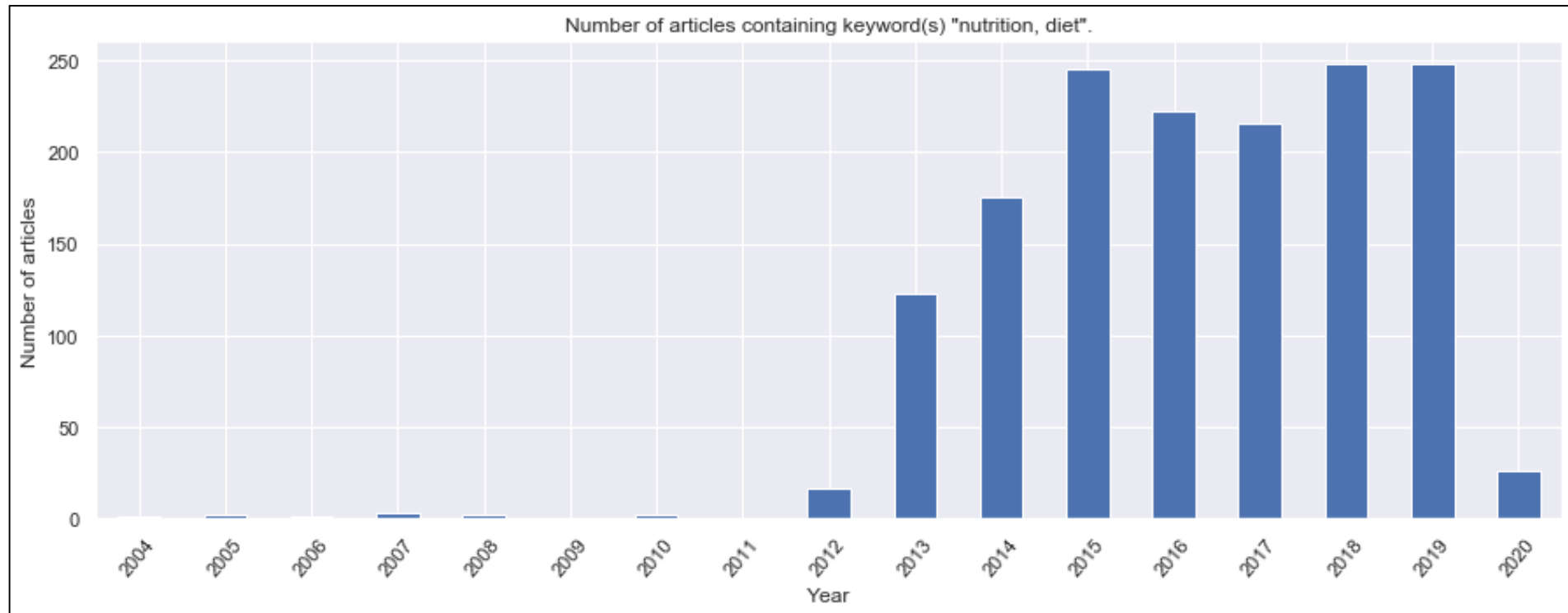


- “nutrition” and “diet”
- Titles, Conclusions, Date
- Articles – 1,531

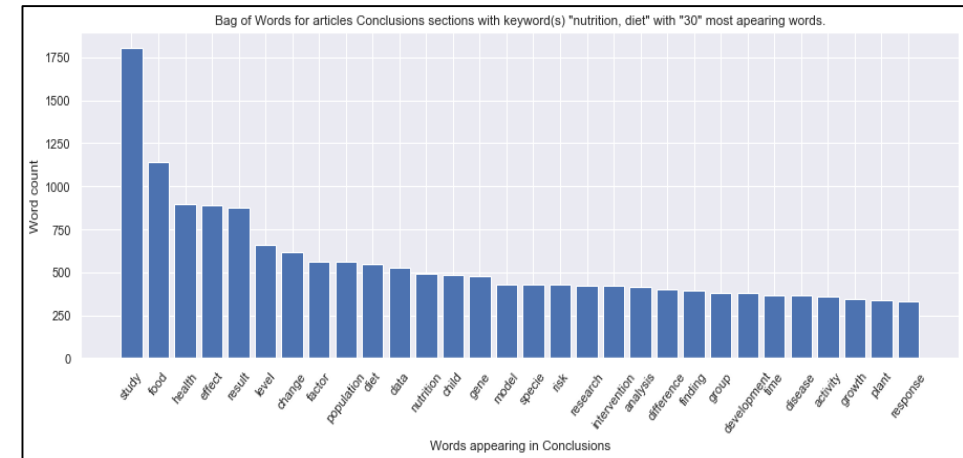


- Text sources
- {}, [], (), no spaces
- Tokenisation
 - Punctuation
 - Lowercase
 - Lemmatisation

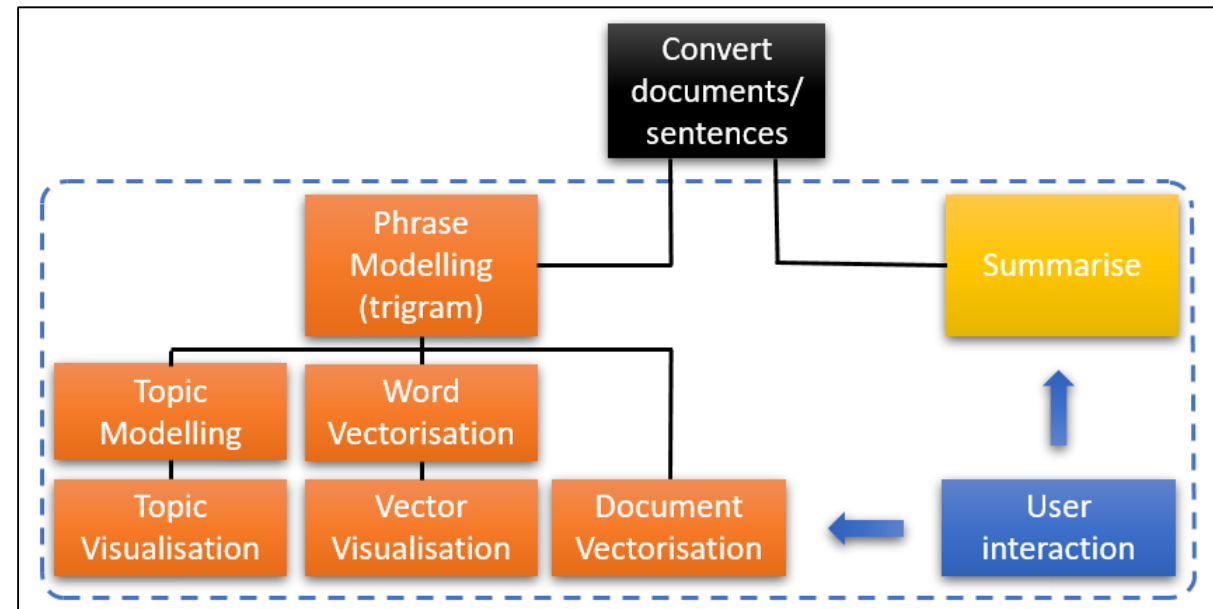




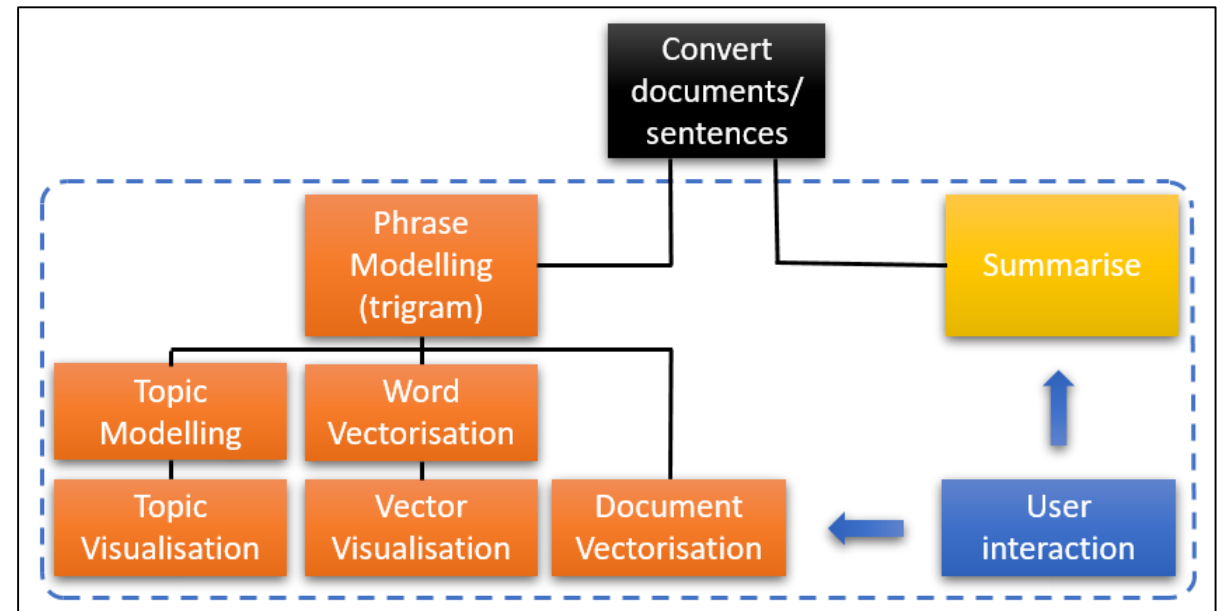
Academic Research Tool



- Gensim
- Phrase Modelling
- Topic Modelling (LDA)
 - Visualisation (pyLDavis)
- Word Vectorisation (Word2Vec)
 - Similarity
 - Difference
 - Visualisation (t-SNE)



- Document Vectorisation (Doc2Vec)
- Summarisation
- User interaction



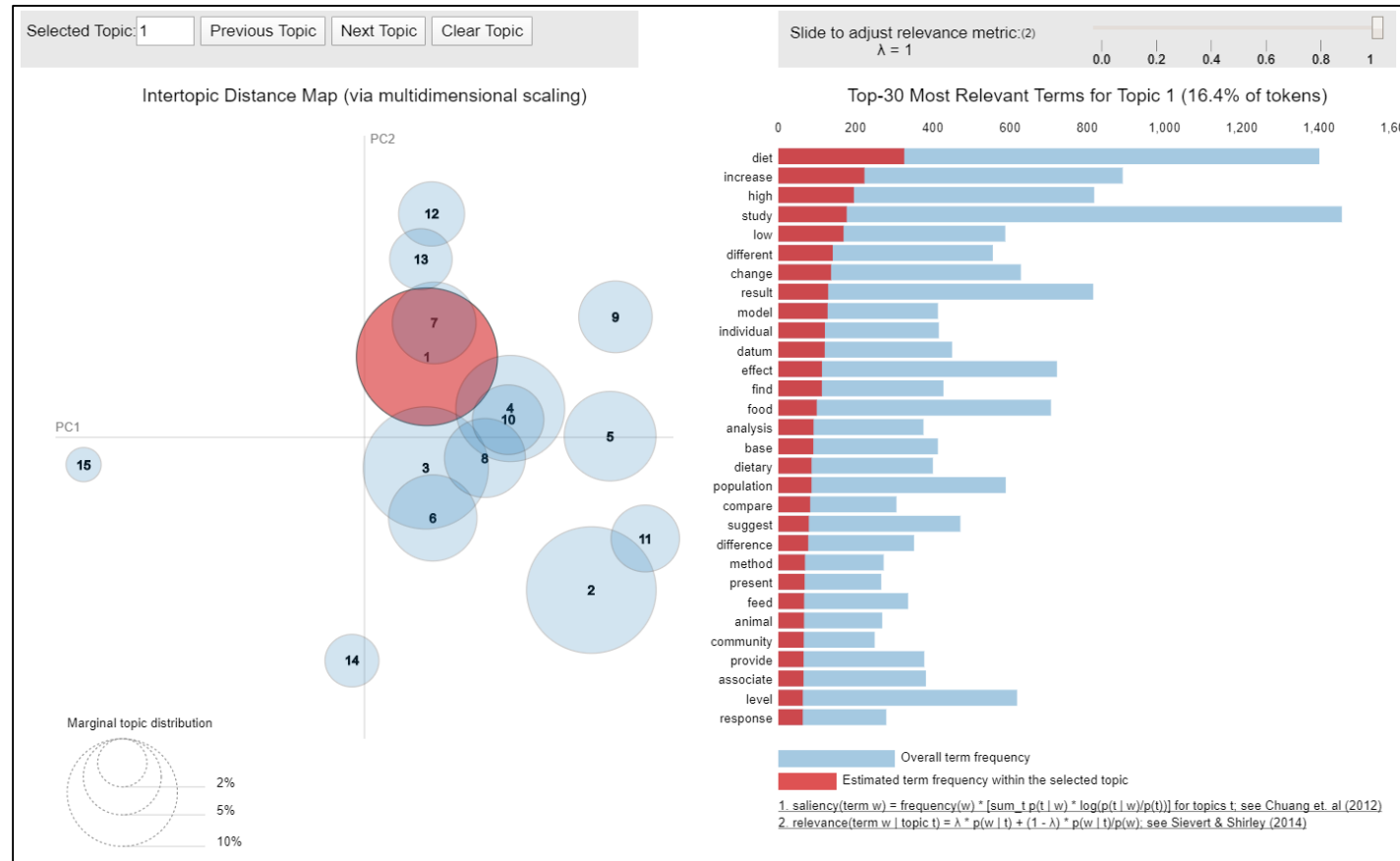
Topic Model

Academic Research Tool

Topic	Top 10 relevant words									
No.	0	1	2	3	4	5	6	7	8	9
topic0	food	result	nutritional	change	increase	health	datum	low	include	provide
topic1	result	level	population	change	increase	find	low	suggest	patient	research
topic2	high	food	increase	result	use	important	low	protein	change	level
topic3	increase	low	result	population	level	high	health	child	food	different
topic4	change	different	high	find	result	group	specific	protein	use	population
topic5	child	high	result	feed	protein	research	food	datum	need	female
topic6	dietary	level	increase	result	high	datum	model	provide	change	individual
topic7	food	different	change	woman	result	increase	dietary	high	group	population
topic8	result	different	low	food	intervention	control	increase	change	reduce	high
topic9	increase	high	change	nutritional	obesity	population	level	improve	base	result
topic10	increase	intervention	change	high	base	health	result	different	cell	suggest
topic11	plant	change	high	food	result	child	specie	time	low	increase
topic12	high	result	increase	level	factor	different	suggest	improve	find	population
topic13	population	level	community	increase	model	result	low	health	child	associate
topic14	increase	high	result	patient	specie	low	reduce	protein	find	gene

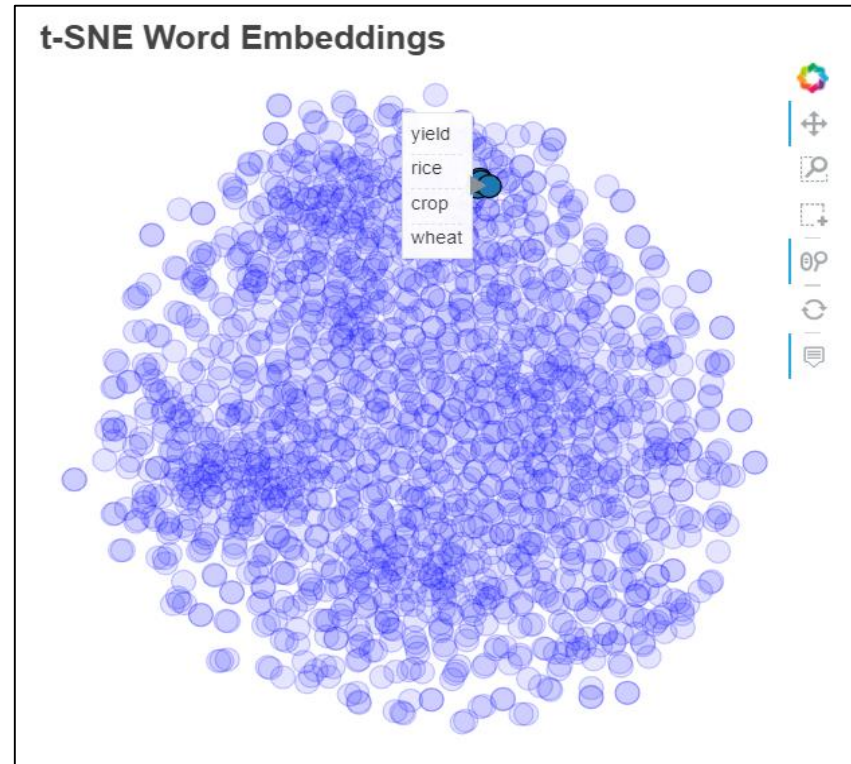
Topic Visualisation

Academic Research Tool



Similar Words							
carbohydrate		fat		protein		death	
synthesis	0.4663	cholesterol	0.4715	plasma	0.5040	clinical	0.3634
energy	0.3804	supplement	0.4500	acid	0.4908	mortality	0.3618
insulin	0.3765	fatty_acid	0.4492	fatty_acid	0.4818	life	0.3578
intake	0.3724	liver	0.4358	expression	0.4619	cardiovascular	0.3465
metabolism	0.3644	body_weight	0.4187	enzyme	0.4599	neonatal	0.3399
fat	0.3584	dietary	0.4169	amino_acid	0.4148	peptide	0.3289
transport	0.3567	total	0.4115	source	0.4142	perform	0.3208
ingest	0.3562	DM	0.4010	metabolism	0.4005	screening	0.3174
reveal	0.3558	weight_gain	0.3916	degrade	0.3971	newborn	0.3135
high_fat	0.3532	dairy	0.3898	induce	0.3935	birth	0.3126

Word Algebra				
life + negative	child - food	nutrition - food	growth - food	health - nutrition
death	undernutrition	breastfeed	height	colony
year	stunting	organ	organ	nutritional_intervention
young	stunt	clearly	damage	overweight
10	height	nutrition_intervention	deposition	behavioral
live	nutritional_intervention	underlying	inflammation	indicator



- Proof of concept
 - Academic consensus
 - Research Tool
- 1 API – data
- Pro's
 - Word/Doc Vectorisations
 - Summarise
 - Interactions
- Con's
 - Topic modelling

- Data

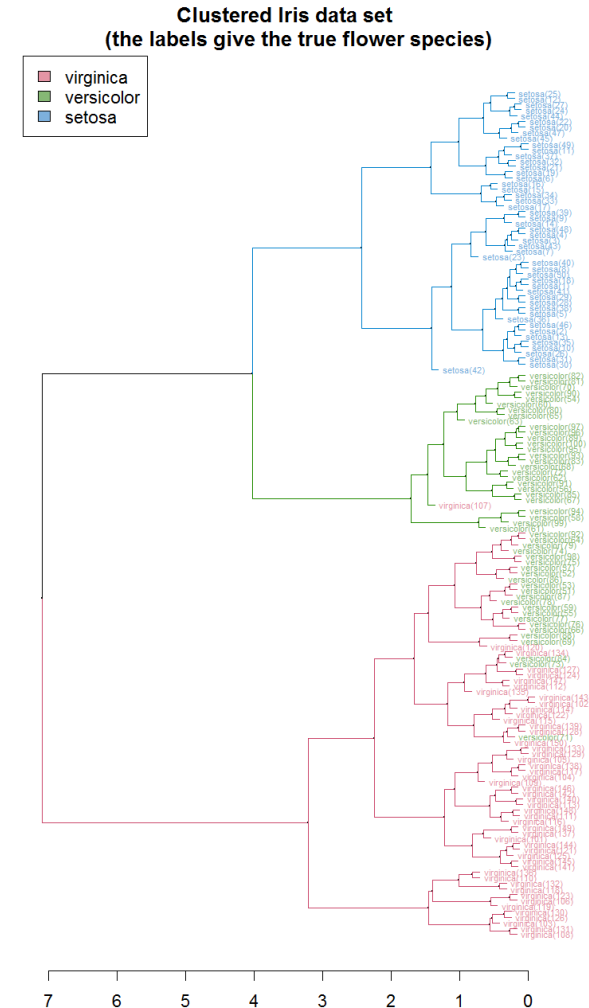
- MIT Libraries – 27 APIs
- JSTOR > 20,000 articles (1,500)
- 15,000 Articles < 2012
- Not 'Conclusions' (Abstract)
- Academic field 'History'

- Data Exploration

- Geographical
- Journals
- Authors – bias
- Funds
- Explore after Phrase Modelling

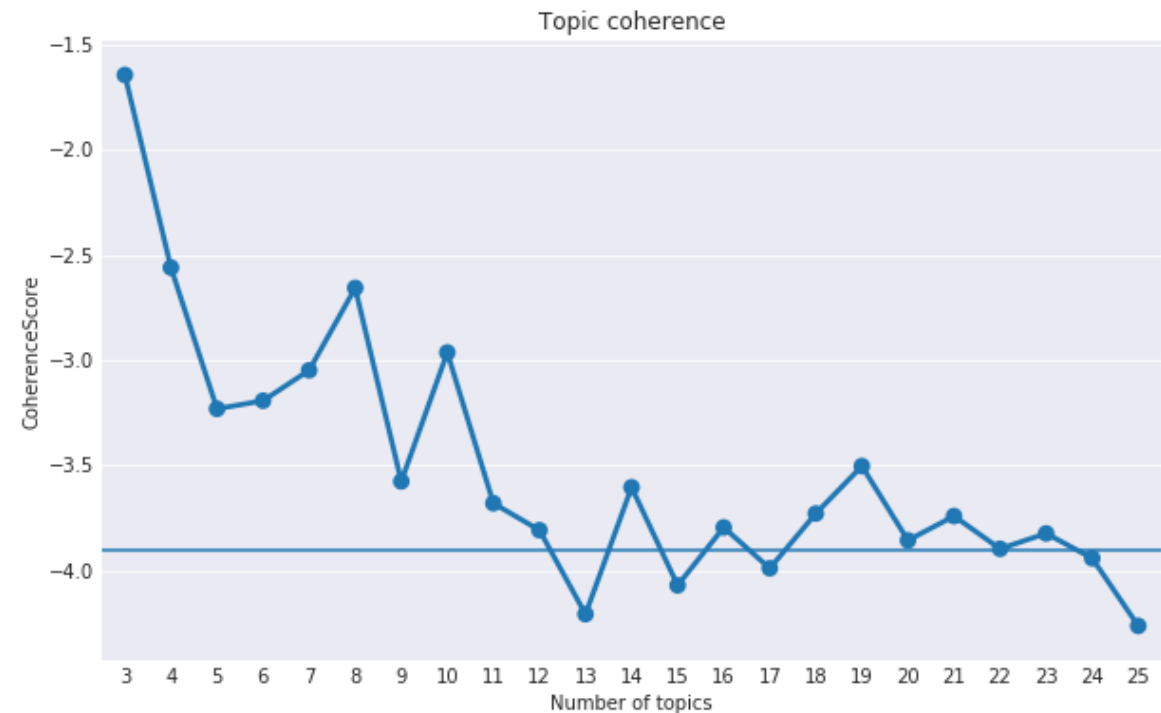
Hierarchical Vectorisation Topic Labelling

- Word Vectors
- Average per Level
- Similarity
- Topic Label



Optimum Topic Number

- Medium article
- Coherence Score / Topic
- Range of Topics
- 12 – 14
- Production - Webpage



A Research Tool for Insights in Academic Fields

Thank you!



Jacques Poolman

