Ranking Philosophers using Wikipedia

Giuliano Martinelli 1915652
Mattia Pocci 1688770
Luigi Sigillo 1761017

Github Repo

Reference Paper: Ranking Guitarists

Apply Google PageRank to Wikipedia guitarists articles to rank them based on influence they had on each other

Key Points:

1) Collecting data about guitarists and their influences from Wikipedia.
2) Converting this data into a directed graph, where a node represents a guitarist, and an edge from A to B represents the influence from A to B.
3) Applying Google PageRank algorithm to rank the guitarists.

Goals:

1) Using a quantitative method to find the most influential guitarist

2) Estimation of influence from the guitarist community itself, instead of fans

Our Approach: Ranking Philosophers

We decided to apply the same approach to Philosophers. The concept of influences is crucial in philosophy and Wikipedia is a good dataset to retrieve this informations.

In the making, we decided to change some of the decisions made on the original paper, since we found better solutions in order to obtain more accurate results.

Resources

We have consulted these related papers:

- *Larry Page, Sergey Bryn, The PageRank Citation Ranking: Bringing Order to the Web, Stanford University, 1998*
  This paper describes PageRank, a method for rating Web pages objectively and mechanically, effectively measuring the human interest and attention devoted to them.

- *Kotaro Nakayama, Wikipedia as a Corpus for Knowledge Extraction, Osaka University, December 2010.*
  This paper analyzes the Wikipedia link structure, in order to exploit it for mining purposes.

- *Bhargav Srinivasa-Desikan, Natural Language Processing and Computational Linguistics: A practical guide to text analysis with Python, Gensim, spaCy, and Keras, 2018.*
  This paper summarizes some of the techniques we used in order to exploit Spacy Language Dependencies.

His Project Organization
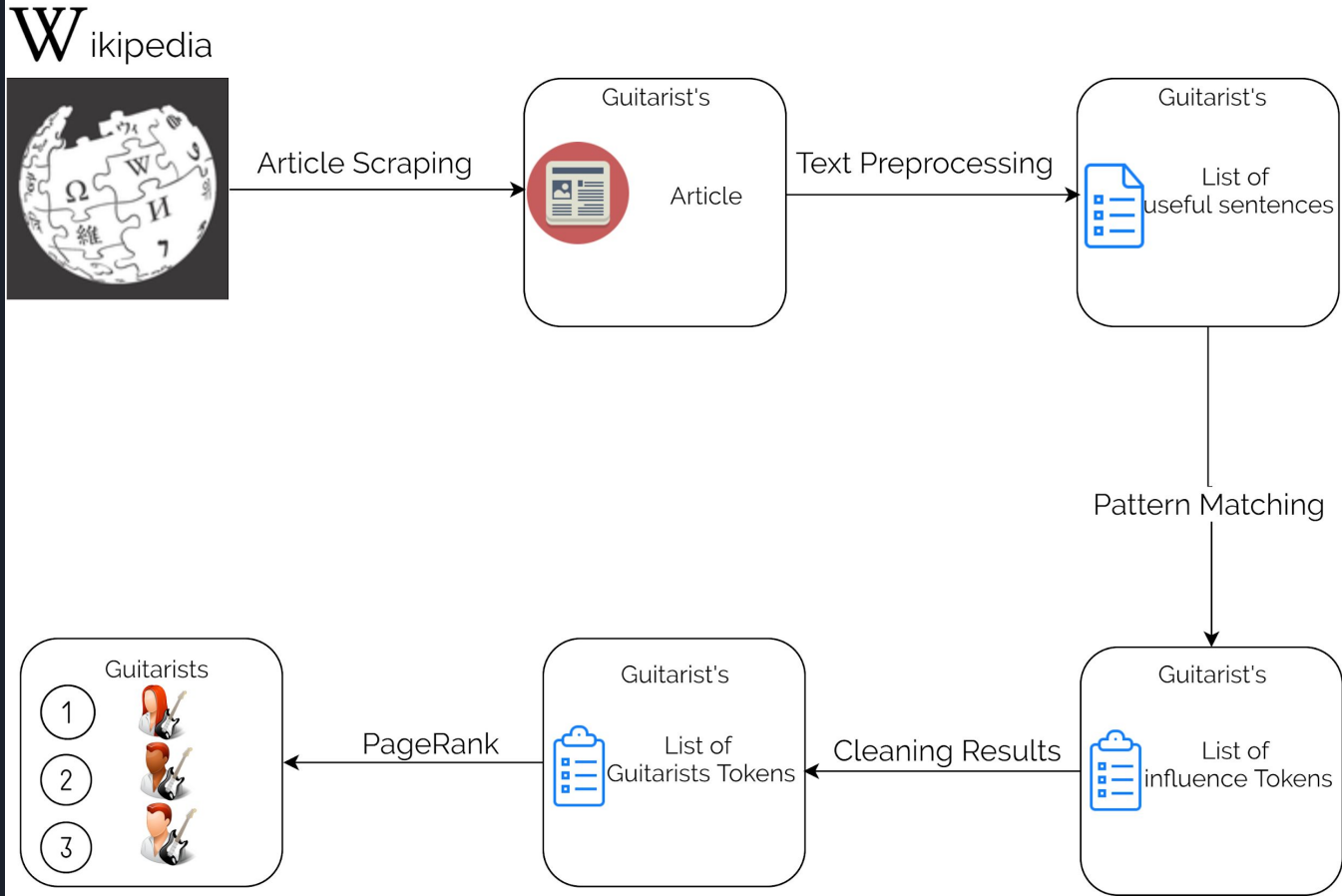
Our Project Organization
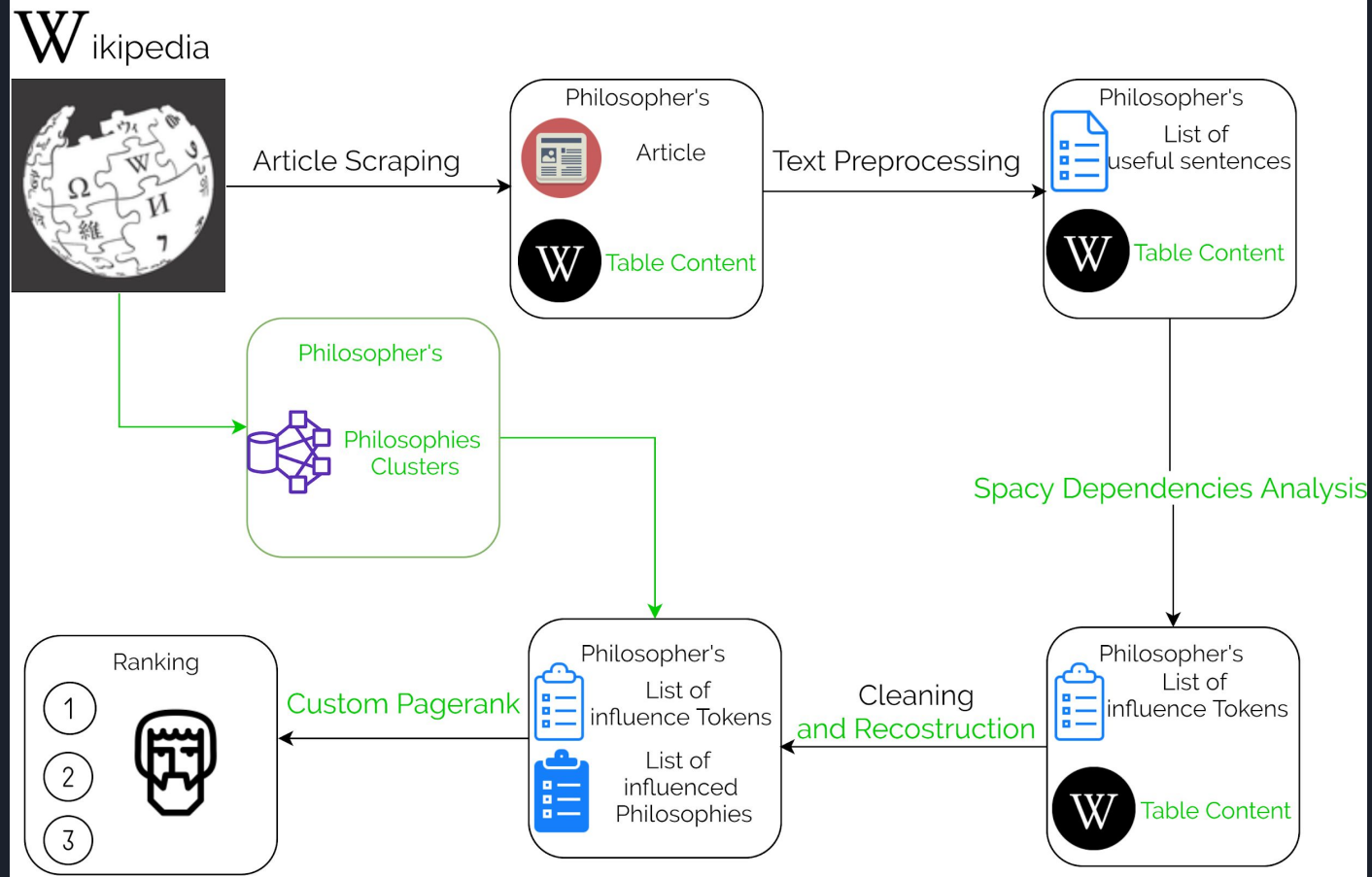
Motivation of the differences

1) **Article Retrieval**
We try to scrape the entire wikipedia dump, but the result of the scraping were lacking of many philosophers, so we proceeded using Wikipedia API to retrieve articles starting from Wikipedia list of philosophers.

2) **Biography table content**
For most of the philosophers, the biography table includes a list of influences and a list of influenced people. We exploited this already available structure to gather more information.

3) **Regular Expressions vs Spacy Dependency Analysis**
Our first attempt with RegEx pattern matching showed a very low accuracy. We concluded that using the NLP framework called Spacy to analyse in-sentence dependencies was the best solution.

4) **Philosophies Clusters**
After some attempts we found out that sentences with form "P influenced Western Philosophy" where not taken into account, and this was producing some inaccurate results. Our solution to this problem was to build different clusters, one for each philosophy current.

Socrates' biography table

Differences: Article Retrieval

- **Wikipedia Dump vs Lists**
  We try to scrape the entire wikipedia dump, but the result of the scraping were lacking of many philosophers, so we proceeded using Wikipedia API to retrieve articles starting from Wikipedia list of philosophers.

- **Biography table content**
  For most of the philosophers, the biography table includes a list of influences and a list of influenced people. We exploited this already available structure to gather more information.

Influences	[hide]
Prodicus, Anaxagoras, Archelaus, Diotima, Damon	
Influenced	[hide]
Virtually all subsequent Western philosophy, but Plato and Xenophon in particular	

Socrates' biography table

Differences: Spacy, What are dependencies?
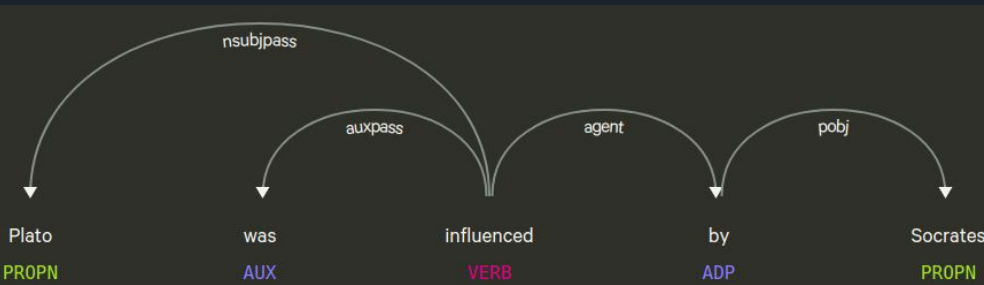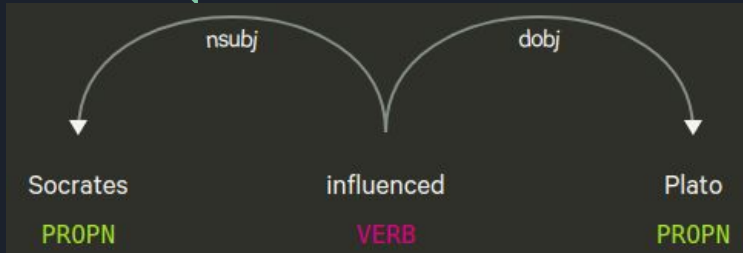
- Dependency Parsing gives for each word an NLP token that has different fields.
- An example of Parsing is shown on the side. Spacy builds a tree on the phrase highlighting the logical relation of the words contained.
- We decided to apply pattern matching on those dependencies instead than directly on the phrase, resulting an improvement on the overall accuracy.

spaCy

Differences: Clusters based on philosophies

- Since a cluster like "Western Philosophy" included lots of philosophers, we had to add an edge (namely an influence relation) between who influenced and every philosopher in the cluster.
- Given that we could not extrapolate the degree of influence on every single philosopher in the cluster, we made an assumption: The edges of this kind will have a lower weight than a direct influence .
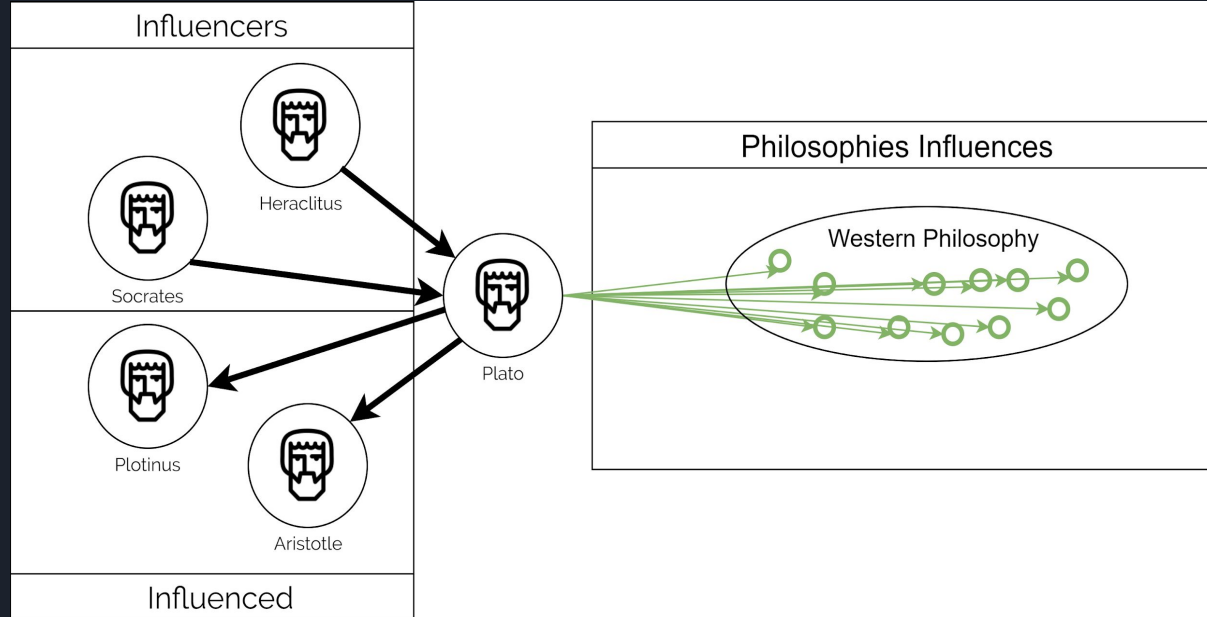
Results: Using Biography Table Content

Only processed text Pagerank

Philosopher	Rank	#Influenced	#Influencers
Friedrich Nietzsche	0.01695682301781985	19	0
Joseph Maréchal	0.013912066103688604	4	0
Karl Rahner	0.01298324617554614	4	0
Gorgias	0.009086894584104166	4	0
Henry David Thoreau	0.008748517527482735	9	0
Ralph Waldo Emerson	0.008190490226735344	1	0
Niccolò Machiavelli	0.007954178114939144	6	0
Thucydides	0.007363254280333316	1	0
Wilhelm Dilthey	0.006405159027511502	9	0
Plato	0.006089153665907449	8	1
Edmund Husserl	0.005962646530872782	7	0
Georg Wilhelm Friedrich Hegel	0.005575252419420598	4	2
Plotinus	0.0051229381091 37774	2	0
Voltaire	0.00511490283316811	7	0
Maurice Merleau-Ponty	0.005026016491553915	3	0
Ammonius Saccas	0.00495670027540215	1	0
Mary Wollstonecraft	0.004900535717203327	2	0
Immanuel Kant	0.004621187292755193	9	2
Joseph Butler	0.004574205204558209	5	0
Max Stirner	0.004571506832867443	5	0

Processed Text + Biography Table Content

Philosopher	Rank	#Influenced	#Influencers
Plato	0.023768886355211988	9	9
Immanuel Kant	0.020332166741262408	20	16
Aristotle	0.017906698900770603	4	2
Heraclitus	0.017070844212399502	16	3
Anaximander	0.016476528409080177	5	0
Georg Wilhelm Friedrich Hegel	0.015079128477618324	32	13
Friedrich Nietzsche	0.013072622167738586	28	11
Parmenides	0.011730985871019358	7	2
Plotinus	0.010805183776088495	12	4
Confucius	0.010584369770989371	3	0
Mencius	0.009609210097035312	1	1
Socrates	0.009453581690851824	3	1
Baruch Spinoza	0.008769434260147072	24	12
Niccolò Machiavelli	0.008633043592530056	19	6
David Hume	0.008268684701638423	20	11
John Locke	0.008223025070522357	7	7
René Descartes	0.008199688665028453	9	14
Karl Popper	0.007415363259352624	19	0
Avicenna	0.007186447672458926	5	0
Porphyry (philosopher)	0.006399903100483041	3	0

Results: Using Philosophies information

We choose to lower importance of the philosophy currents, because the ancient Greeks philosophers would take a unproportioned boost

Cluster importance = 0.01

Philosopher	Rank	#Influenced	#Influencers
Plato	0.029511039306611152	9	9
Heraclitus	0.02593571105873722	16	3
Aristotle	0.023244316987223353	4	2
Anaximander	0.022131254866566325	5	0
Immanuel Kant	0.019284901106381674	20	16
Confucius	0.016300948348405158	3	0
Parmenides	0.015012187311323515	7	2
Socrates	0.01477418359059414	3	1
Friedrich Nietzsche	0.014212252410572244	28	11
Georg Wilhelm Friedrich Hegel	0.013467626284281655	32	13
Mencius	0.01213571409751495	1	1
Plotinus	0.011768492629341315	12	4
Pythagoras	0.00967819445334169	3	1
René Descartes	0.00944178924941778	9	14
David Hume	0.009285330894875538	20	11
Gorgias	0.009266634465392241	5	3
Niccolò Machiavelli	0.009102557451412582	19	6
Anaximenes of Miletus	0.00874015614755261	1	0
Karl Popper	0.00795845826330670	19	0

Cluster importance = 0.05

Philosopher	Rank	#Influenced	#Influencers
Plato	0.028269111635187315	9	9
Heraclitus	0.026073735214490707	16	3
Aristotle	0.023479208861731663	4	2
Immanuel Kant	0.019940629919449452	20	16
Anaximander	0.018918978755500728	5	0
Socrates	0.015958600361062936	3	1
Friedrich Nietzsche	0.01390278432972968	28	11
Parmenides	0.013829591529717345	7	2
Georg Wilhelm Friedrich Hegel	0.012397805765709828	32	13
Plotinus	0.01066843964749936	12	4
Gorgias	0.010534826656383783	5	3
Confucius	0.01050816095174532	3	0
Pythagoras	0.010340968894952686	3	1
René Descartes	0.010248288532378318	9	14
David Hume	0.010070586111792925	20	11
Theophrastus	0.008721450894434149	2	2
Niccolò Machiavelli	0.00869522736735554	19	6
Antisthenes	0.008669916329436811	0	2
Karl Popper	0.008538377928140873	19	0
Anaximenes of Miletus	0.008516996538594562	1	0

Cluster importance = 1

Philosopher	Rank	#Influenced	#Influencers
Heraclitus	0.0438997341932021	16	3
Antisthenes	0.038259753695967624	0	2
Theophrastus	0.03385332848090129	2	2
Aristotle	0.025928440604278843	4	2
Plato	0.025137773546423017	9	9
Immanuel Kant	0.02272122279466714	20	16
Thomas Hobbes	0.02193086690139938	0	9
Socrates	0.021447020260104265	3	1
Albertus Magnus	0.019952836260057717	4	8
David Hume	0.014952401772742025	20	11
René Descartes	0.014660267967962326	9	14
Augustine of Hippo	0.014532615619224476	3	10
Thomas Aquinas	0.014230844885580195	1	13
Pythagoras	0.014161845611165447	3	1
Gorgias	0.013976511789632015	5	3
Ludwig Wittgenstein	0.013738091374414431	12	12
Cicero	0.013702642297205043	0	3
Karl Popper	0.012820459112676867	19	0
Moritz Schlick	0.012123062421422248	3	7
Gottlob Frege	0.011513172987398021	6	6

Results and final considerations

Philosopher PageRank	Leiter Reports	Listverse	Live Mirror
Plato	Plato	Aristotle	Aristotle
Heraclitus	Aristotle	Plato	Kant
Aristotle	Kant	Paul of Tarsus	Locke
Kant	Hume	Descartes	Epicurus
Anaximander	Descartes	Confucius	Zeno of Citium
Socrates	Socrates	Aquinas	Plato
Nietzsche	Wittgenstein	Avicenna	Confucius
Parmenides	Locke	Zeno of Citium	Hume
Hegel	Aquinas	Epicurus	Descartes
Plotinus	Hegel	Locke	Socrates
Gorgias	Leibniz		
Confucius	Spinoza		
Pythagoras	Mill		
Descartes	Hobbes		
Hume	Augustine		
Theophrastus	Marx		
Machiavelli	Nietzsche		
Antisthenes	Kierkegaard		
Popper	Rousseau		

- Since there is not an objective ground truth that can rank philosophers, we compared our result with different online resources.
- As we can see, after tuning the value of the weight of the philosophy currents, we have found that this is a good result.
- Interesting statements can be made on the presence of Anaximander and Heraclitus, Pagerank has made them more influential!
- In the end, we can say we succeeded to apply Pagerank on Wikipedia Philosophers articles, producing some interestin results.
- Improvements to the project could be to work out better patterns on word dependencies and to find a more reliable ground truth.

Thanks for the attention

iano Martinelli 1915652
Mattia Pocci 1688770
Luigi Sigillo 1761017

shutterstock.com · 1256481601