

UNBIASED

Brian Hao
Ricky Chan
Erica Hendricks-Elliston
Phillip Foo

Hunter College CSCI 499 Capstone
December 18, 2018



What is **UNBIASED**?

- A one-stop news application
- Simple and intuitive to use
- Aggregate articles from multiple top news sources
- For any story, provide the user with access to its related articles



DEMO

<https://csci499-unbiased.herokuapp.com/>



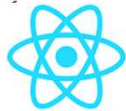
Ideas For Future Development

- User login/accounts to save news display preferences
- Allow commenting / discussion on articles / events
- Aggregate from even more news sources to select from
- More filtering options (Trending, Most Viewed, Most Commented, ect.)
- Allow users to search specific article topics (Ex. all related articles about bitcoin)
- Add categories and allow displaying by categories (Politics, Sports, Gaming, ect.)
- Improved relational algorithm (natural language processing)

Tools/Technologies

Frontend :

ReactJS



;



:SS

Database :

Neo4j



Hosted on:

Heroku

API: **News API**

Algorithms: Porter Stemmer, Sørensen–Dice



Thank you!

We'd be happy to answer any questions.

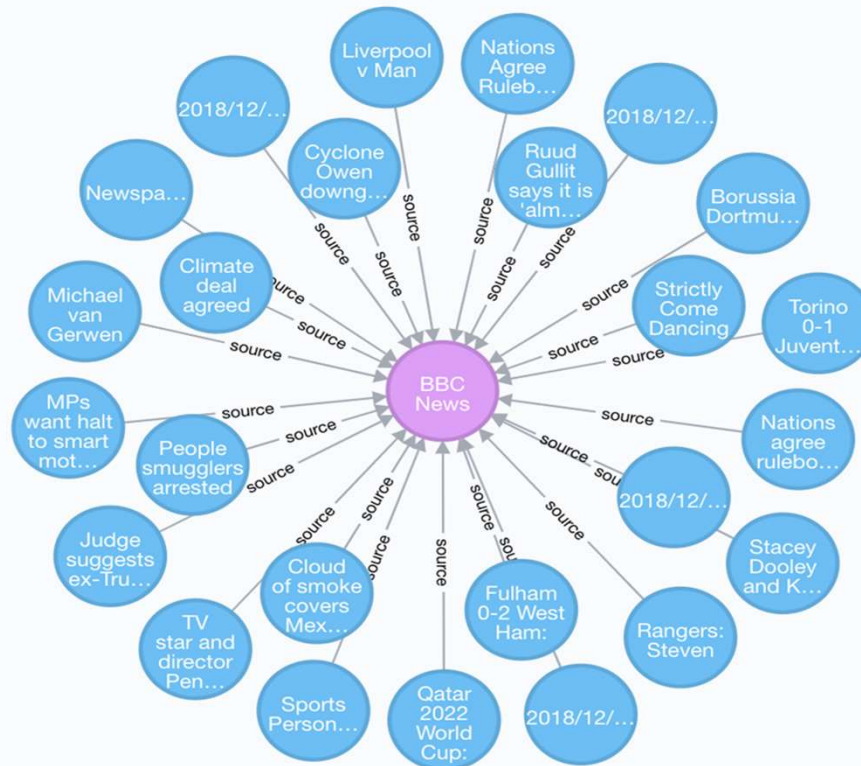


What is Neo4j ?

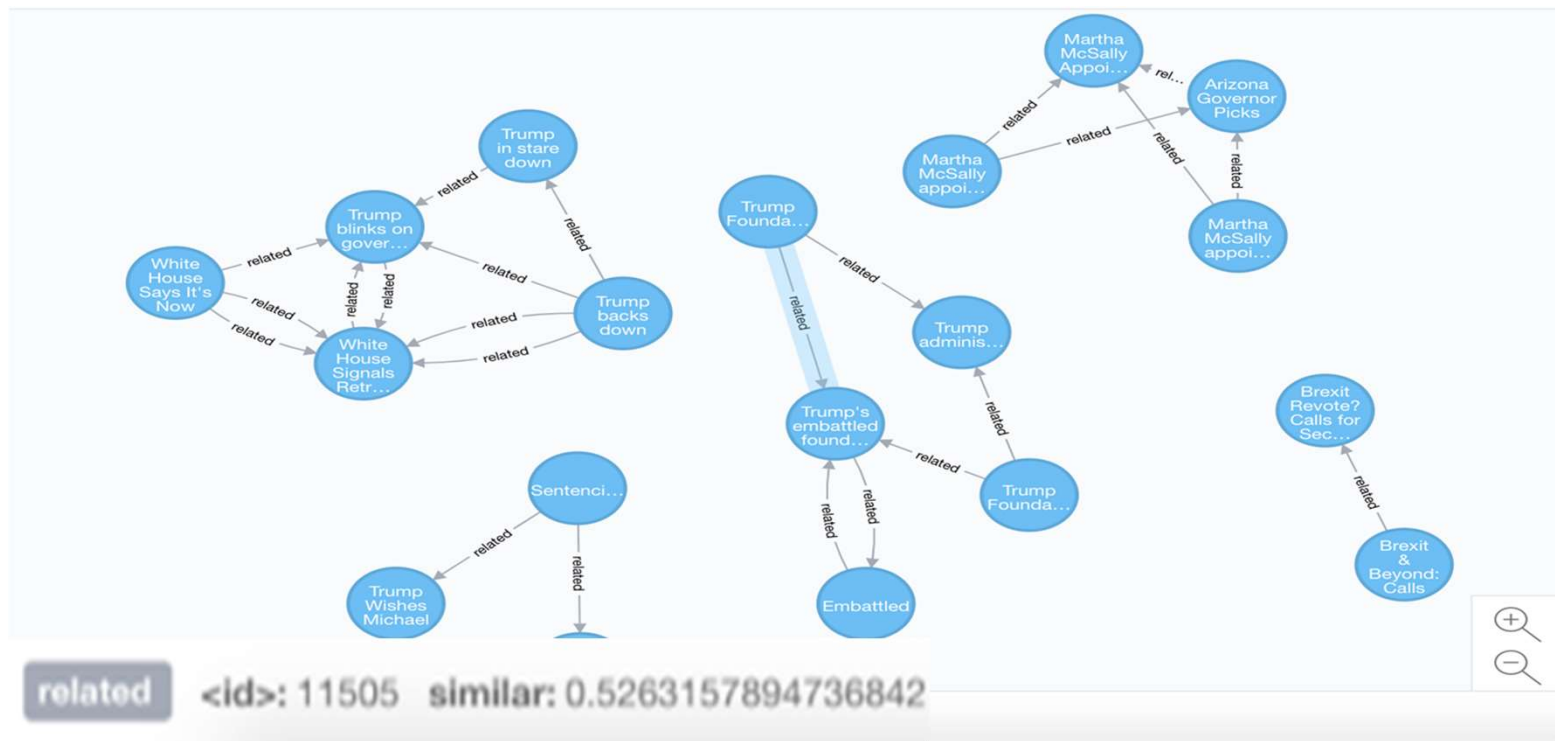
- RDMS
- Graph DB
 - Nodes , edges, and properties
- Utilizes Cypher Query Language (CQL)
- Allows relationships to explicitly be defined



Neo4j - Article to Source Relationship Visualization



Neo4j - Article to Article Relationship Visualization





Determining Relationships

- Remove special characters
- Remove stop/common words
- Use the Porter-Stemmer algorithm
- Use Sørensen–Dice algorithm to return a coefficient
- **Two articles share a relationship if ...**
 - **the coefficient=0.2 AND have at least one word in common**
 - **coefficient=0.7**



Sørensen–Dice algorithm

- Returns a value that is the ratio of how many words the two sentence have in common to the total number of words between the two sentences.

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|}$$



Removing Special Characters and Stop Words

- What is a special character?
 - Example: periods, colons, parenthesis, exclamation points etc...
- Why remove special characters?
 - Algorithms interpret a word with a special character appended to it differently than the same word without it!
 - Example: "cat:" is not the same as "cat"
- What is a stop word?
 - Useless words that search engines are programmed to ignore when querying
 - Example: "the" "a" "an" "in"
- Why remove stop words?
 - Adds no value to the semantics of the sentence
 - Negatively affects the algorithm that returns the similarity score

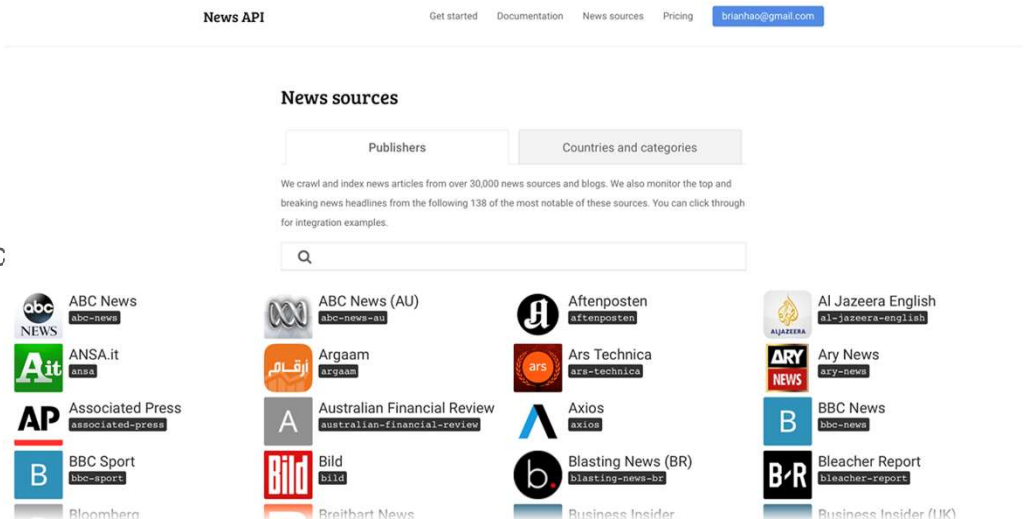


Porter-Stemmer Algorithm

- What is the porter-stemmer algorithm
 - "... is a process for removing the commoner morphological and inflexional endings from words in English. "
 - Martin Porter
 - Normalizes words in a sentence
- Why use Porter-Stemmer?
 - Sørensen–Dice algorithm interprets two words differently even if contextually they mean the same thing
 - Example: "hopefully" is not the same as "hopeful"

What Is The News API

- Aggregate of articles from over hundreds of news sources
- Return all information on a selected article (title, source, date ... etc)
- Query by source, top headlines, country of origin





Contributions