I. Use Case Description	
Use Case Name	Political Journalism Ontology
Use Case Identifier	OE2024-PJO-2
Source	RPI Ontology Engineering
Point of Contact	Avery Iorio (iorioa@rpi.edu) Kirk Olkowski (olkowk@rpi.edu) Nathaniel Adair (adairn@rpi.edu)
Creation / Revision Date	10/21/2024
Associated Documents	*Requirements documentation, traceability matrix if applicable*

II. Use Case Summary	
Goal	The goal is to provide users the ability to discover trends in how specific journalists and news outlets report on political issues and politicians.
Requirements	. We intend to evaluate the relative (rather than objective) bias without making claims about the factual nature of the content of the various articles. We plan to develop an ontology-enabled system that allows a comparison between these various outlets and journalists that accounts for a difference in terminology and content focus that results from media outlets appealing to different audiences. Users need to be able to connect various articles across media outlets to the same journalist or editor. The ontology serves the purpose of providing a way to link these articles and events that is robust to biased reporting. This is because different outlets will refer to the same issues, politicians, and events using different terminology while selectively including/emphasising various details. For example, a right leaning outlet may have a particular story about immigration with the *immigration* tag whereas a left leaning outlet may use the *human rights* tag. We need a semantic way of seeing if these articles are generally referring to the same event. Users need a way to connect politicians or prominent figures to pieces of legislation, political issues, other politicians, political parties, events, elections, and articles.
Scope	The current scope of this project restricts itself to only using the New York Times and Fox News for data. We intend to limit ourselves to relative bias between given articles or news outlets. This allows us to avoid the definition of a 'ground truth' opinion on the significance of any given event, and instead allows us to leave such value determinations to the user. This project will focus primarily on federal and gubernatorial elections and positions insofar as political elections and positions are concerned. Local elections and local positions will be ignored due to the assumption that there is far less measurable media "spin" regarding local politics. This project relies on the existing tags and terminology associated with each article as found on each publication's RSS feed. Sentiment analysis and the generation of tags from the body of articles are outside of the scope of this project. Simple keyword searching within an article's title may be necessary. Evaluating the provenance of the actual article content is beyond the scope of this system. Biased reporting may still be useful to a reader so long as the

	direction and relative degree of the bias is well understood.
Priority	*N/A*
Stakeholders	**Politicians and Political Campaign Strategists** - To understand how their policy platform is represented in media and compare how they are portrayed across outlets and between journalists. **Voters** - To track the political platforms of candidates and uncover disconnects and biases in how different media outlets represent specific issues, politicians, or events. **Media** - To evaluate potential journalists biases, issues of interest, and reporting patterns **Public Relations Team** - To gauge how a particular journalist or media outlet might cover a political issue and who might be exposed to the coverage.
Description	*A critical component of functioning democracies is a voting populace withs the information necessary to make an informed vote. The relaying of this information should naturally facilitate a rational citizen in voting in a manner that best suits their well-being and beliefs. Due in part to the advent of the digital age and the ad-view revenue model dominating media, there are perverse incentives that favour viewer engagement over unbiased and factual reporting.1* *This perverse media incentive structure naturally results in outlets finding a position on this bias/accuracy continuum that caters to a specific cross section of viewers to maximise engagement and build a viewer base. As a result outlets choose editors, journalists, stories, and language that best appeals to this base and is largely responsible for today's polarised media climate. Unbiased reporting is often conflated with factual reporting* This use case focuses on using a political journalism ontology to allow the comparison of media coverage across different outlets and journalists. The primary goal is for the user to understand the relative bias of a news outlet or journalist regarding a particular issue or politician. The desire for this functionality comes from the majority held belief [3] that U.S. media in general is biassed. The frequency with which outlets publish articles, the publishing time, the tags, and the publishing history of the journalist all carry information about how the outlet and journalist want the issue or associated people to be perceived by the public.

Actors / Interfaces	The **primary actor** is the user, namely a person looking for information regarding journalism and politics.
	
	Secondary actors include:
	- RSS Feeds
	- The New York Times
	- Fox News
	- Individual Journalists
	- Individual Articles
	- AllSides Media Bias Rankings
	- U.S. Government data archives
	  - Congressional Voting Records
	  - Election Records

Pre-conditions	Before a user can ask this system a competency question, there must be at least 2 media outlets with some difference in intended audience with a non trivial number of articles published over a reasonably long time. While the exact number of articles and timespan are left intentionally vague, the more important observation is that the quality of the answers is directly dependent on the quantity of media available. The lower the number of articles that have been ingested by the ontology the more likely the answers to the competency questions are going to capture week to week noise in media coverage as opposed to long term bias trends. Additionally, this system assumes some basic political knowledge on the part of the user to ask questions that can be objectively answered. For example, they must ask more nuanced questions than "What outlet publishes the most fake news?" or "Is Fox or CNN more correct about the economy?"
Post-conditions	There will be **no orphan entities:** e.g. a politician that is not connected to any articles, parties, issues, bills, events, etc is just an empty name that cannot be compared across outlets. There will be **no contradictions:** e.g. an outlet can't lean both left and right, a politician can't both support and oppose a bill, two candidates can't win the same election. The engine will make sure new information reflects the current state of established facts.
Triggers	Elections, political events, house and senate votes, speeches, and article publishing all cause the system to update the underlying political knowledge representation. **For Users:** Asking a query results in the engine pulling the necessary information about media outlets and articles
Performance Requirements	Results should be returned in a reasonable amount of time but it does not have to be exceptionally performant. (i.e. it is fine if it takes a few minutes for a difficult query.
Assumptions	
Open Issues	

III. Usage Scenarios

Provide at least two usage scenarios that flesh out the requirements outlined in the summary, including identification of requirements specific to any envisioned ontology or semantically-driven service or application. Scenarios should be described as narrative, with supporting diagrams as appropriate. In an Agile process, every user story relevant to the use case should be included and elaborated/rolled up into one or more usage scenarios, with a clear mapping from the user story to the scenario it is integrated in or mapped to.

A 54-year-old man who is concerned that his tap water has recently started tasting funny. After ordering an at-home test kit, he found that it has unacceptably high levels of lead. He does not have lead pipes and tested this himself. He took water samples from his own house and various friend's houses over the city and brought them to a lab which confirmed the high levels of lead. The city refuses to do anything as the recent water safety tests show acceptable lead levels. He wants to contact a reporter who has previously written articles about contaminants in consumer products to try to pressure the local city council into investigating. He wants help choosing a reporter to reach out to that has reported seriously on environmental issues and currently works for a paper that is syndicated where he lives.

A 38 year old woman with children lives in rural Pennsylvania. She is an avid Fox News viewer but recently saw an article on CNN about the incumbent Dem Rep in her district being pro-family. She is

wondering if news coverage regarding this candidate is bipartisan or if the reporter working for CNN is simply trying to make the candidate look good.

IV. Basic Flow of Events

Narrative: *Often referred to as the primary scenario or course of events, the basic flow defines the process/data/work flow that would be followed if the use case were to follow its main plot from start to end. Error states or alternate states that might occur as a matter of course in fulfilling the use case should be included under Alternate Flow of Events, below. The basic flow should provide any reviewer a quick overview of how an implementation is intended to work. A summary paragraph should be included that provides such an overview (which can include lists, conversational analysis that captures stakeholder interview information, etc.), followed by more detail expressed via the table structure.*

In cases where the user scenarios are sufficiently different from one another, it may be helpful to describe the flow for each scenario independently, and then merge them together in a composite flow.
User can ask a series of questions

Basic / Normal Flow of Events			
Step	**Actor (Person)**	**Actor (System)**	**Description**
1	User		Launches application
2		App	Loads the user interface
3	User		Inputs query into input text box
4		App	Processes query
5		App	Render query in terms of ontology, and query resulting data
6		App	Return answer in triples, and attendant links to actual news articles
7	User		Closes application

V. Alternate Flow of Events

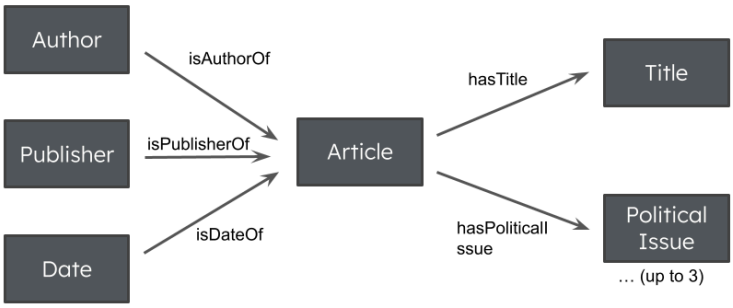
Narrative: *The alternate flow defines the process/data/work flow that would be followed if the use case enters an error or alternate state from the basic flow defined, above. A summary paragraph should be included that provides an overview of each alternate flow, followed by more detail expressed via the table structure.*

Alternate Flow of Events			
Step	**Actor (Person)**	**Actor (System)**	**Description**
1	User		Launches application
2		App	Loads the user interface
3	User		Inputs query
		App	Fails to represent query in terms of the ontology
4		App	Unable to process query, reports ‘Query unrepresentable’ error to user.

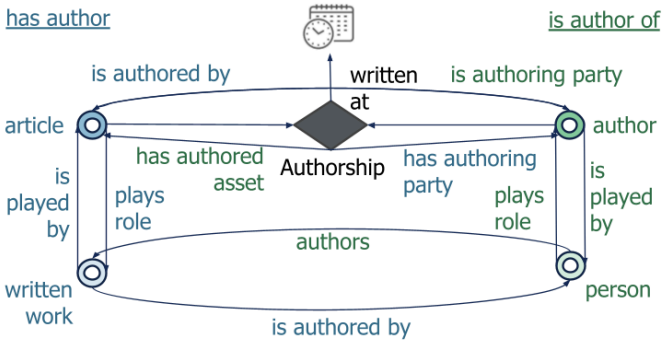
VI. Use Case and Activity Diagram(s)

Provide the primary use case diagram, including actors, and a high-level activity diagram to show the flow of primary events that include/surround the use case. Subordinate diagrams that map the flow for each usage scenario should be included as appropriate

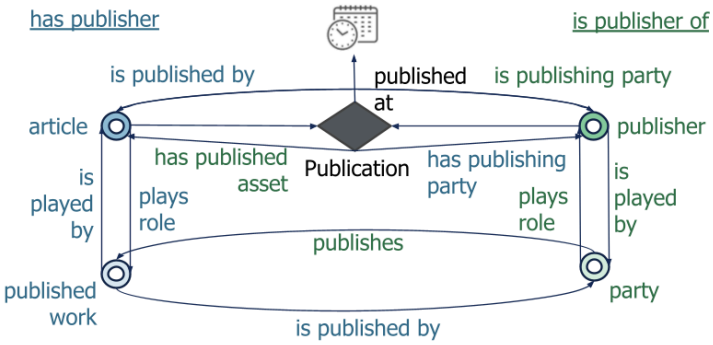
CONCEPTUAL DIAGRAM *#1*



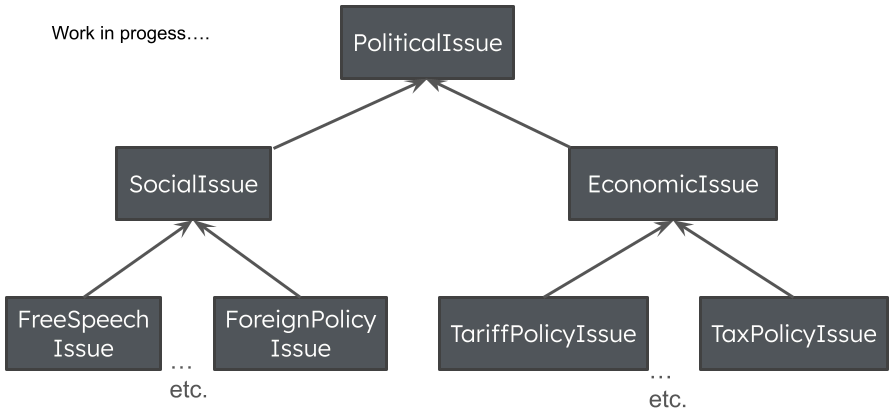
CONCEPTUAL DIAGRAM *#2*



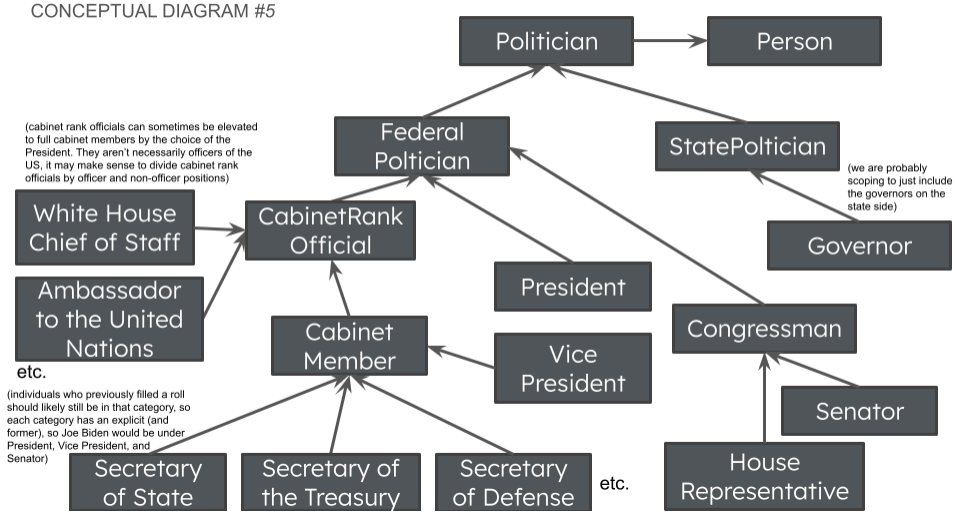
CONCEPTUAL DIAGRAM *#3*



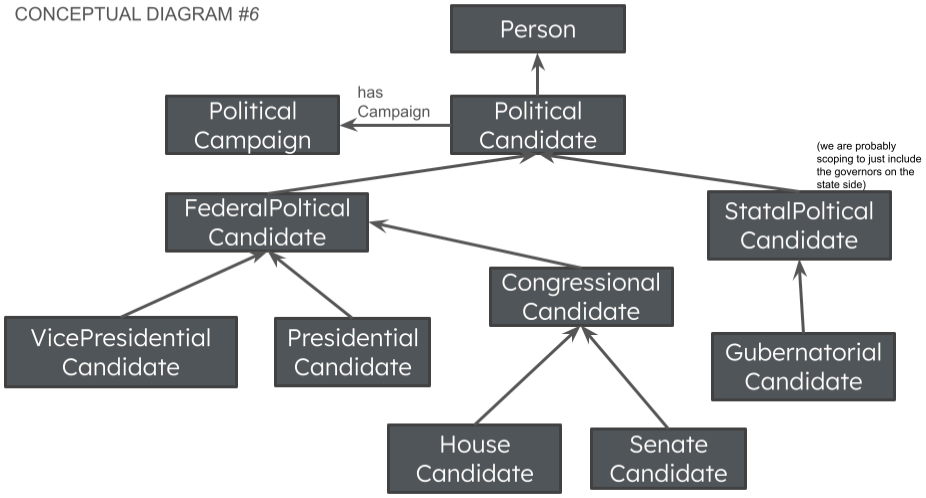
CONCEPTUAL DIAGRAM *#4*

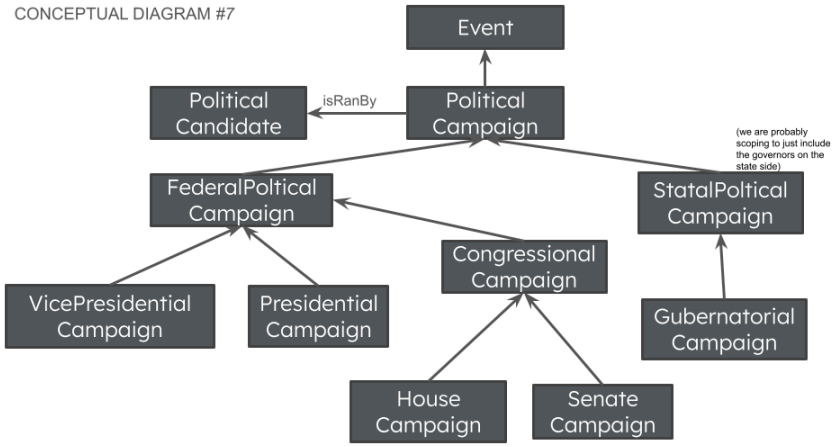


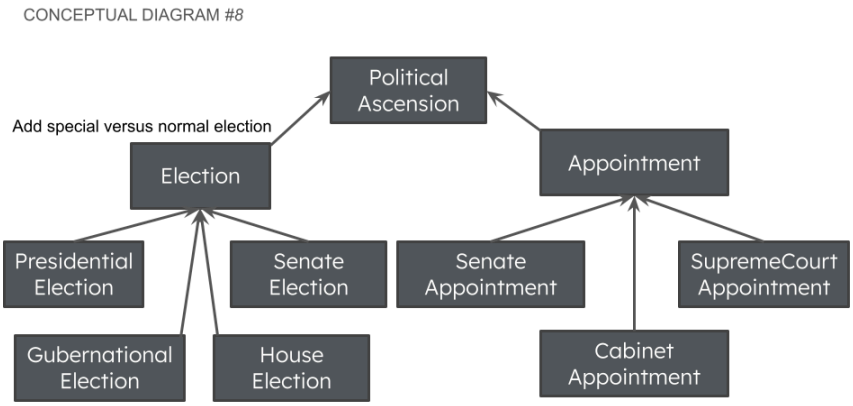
CONCEPTUAL DIAGRAM *#5*



CONCEPTUAL DIAGRAM *#6*







VII. Competency Questions

Provide at least 2 competency questions that you will ask of the vocabulary/ontology/knowledge base to implement this use case, including example answers to the questions.

Describe at least one way you expect to use the semantics and/or provenance to propose an answer to the questions. Include an initial description of why the semantics and/or provenance representation and reasoning provides an advantage over other obvious approaches to the problem. (optional – depending on the use case and need for supporting business case).

Question 1: Which journalist who writes on issues in the Upstate New York area writes the most about environmental issues and conservation?

Answer 1: [Journalist X] has written [Z] articles about environmental and conservation issues concerning the Upstate New York area since YYYY.

Reason 1: We will need to look through the tags of various articles from disparate sources to compare which are focusing on environmental issues and conservation. We will also see if these articles concern areas which pertain to Upstate New York, either specifically or inclusively. Additionally, we need to determine which authorial additions are considered journalists (i.e. editors, etc.)

Question 2: Which journalists wrote articles about Senator Bernie Sanders for both the New York Times and Fox News from 2016–2020?

Answer 2: [Journalist X], [Journalist Y], and [Journalist Z] have written articles about Bernie Sanders for both the New York Times and Fox News from 2016–2020.

Reason 2: We will need to create two lists looking at various articles during 2016–2020 with a tag referencing Bernie Sanders from the New York Times and Fox News respectively. We will need to see which journalist wrote each article and which of these journalists are present on both lists.

Question 3: Has the New York Times or Fox news published more articles about Democratic candidates from the Northeast in relation to the economy in 2023? What about in 2018?

Answer 3: In 2023, Fox news published more articles about Democratic Candidates from the Northeast and the economy while in 2018 Fox news published more articles about the same topic.

Reason 3: We will reason through the political journalist ontology to understand which articles concern Democratic Candidates, determining both the relevant party affiliations of party candidates during the relevant time slices, as well as limiting ourselves to articles which additionally cover economic policy.

How do we evaluate and understand political affiliation? Especially track it as it changes over time.

How do we evaluate candidacy as relative to a given time frame? Do we consider it based on their candidacy for an election in the given time frame? or based on which the candidacy was declared? or just based on how they are referred to in the article? (which I think is a nontrivial NLP problem)

VIII. Resources

In order to support the capabilities described in this Use Case, a set of resources must be available and/or configured. These resources include the set of actors listed above, with additional detail, and any other ancillary systems, sensors, or services that are relevant to the problem/use case.

Knowledge Bases, Repositories, or other Data Sources

Data	Type	Characteristics	Description	Owner	Source	Access Policies & Usage
(dataset or repository name)	*(remote, local/in situ, etc.)*	*e.g. – no cloud cover*	*Short description of the dataset, possibly including rationale of the usage characteristics*		*Source (possibly a system, or remote site) for discovery and access*	
NYT US RSS Feed			Description of the latest NYT articles on US news with tags	*NYT*	*https://rss.nytimes.com/services/xml/rss/nyt/US.xml*	

NYT Politics RSS Feed			Description of the latest NYT articles on politics with tags	*NYT*	*https://rss.nytimes.com/services/xml/rss/nyt/Politics.xml*	
Fox News US RSS Feed			Description of the latest Fox articles on US news with tags	*Fox News*	*https://moxie.foxnews.com/google-publisher/us.xml*	
Fox News Politics RSS Feed			Description of the latest Fox articles on politics with tags	*Fox News*	*https://moxie.foxnews.com/google-publisher/politics.xml*	

External Ontologies, Vocabularies, or other Model Services

Resource	Language	Description	Owner	Source	Describes/Uses	Access Policies & Usage
(ontology, vocabulary, or model name)	*(ontology language and syntactic form, e.g., RDFS - N3)*	*If the service is one that runs a given ontology or model-based application at a given frequency, state that in addition to the basic description*		*Source (link to the registry or directly to the ontology, vocabulary, or model where that model is maintained, if available)*	*List of one or more data sources described by and/or used by the model*	
dbpedia classes				https://mappings.dbpedia.org/server/ontology/classes/		
Ontology of Election		Nigerian Election Ontology	Ayorinde, I. T. Akinkunmi, B. O. Adenuga, A. O.	https://www.researchgate.net/publication/334272639_A_Formalised_Ontology_of_Election		
POWER		Portuguese Election Ontology	S. Moreira, D. Batista, P. Carvalho, F. Couto, and M. Silva	https://www.researchgate.net/publication/252067745_POWER_Politics_Ontology_for_Web_Entity_Retrieval		
oeGOV	ontology	US Government Ontology	TopQuadrant	http://www.oegov.us/		

Other Resources, Service, or Triggers *(e.g., event notification services, application services, etc.)*

Resource	Type	Description	Owner	Source	Access Policies & Usage
(sensor or external service name)		*Include a description of the resource as well as availability, if applicable*	*Primary owner of the service*	***Application or service URL****; if subscription based, include subscription and any subscription owner*	

IX. References and Bibliography

List all reference documents – policy documents, regulations, standards, de-facto standards, glossaries, dictionaries and thesauri, taxonomies, and any other reference materials considered relevant to the use case

Footnotes
1. https://www.allsides.com/blog/why-allsides-doesn-t-rate-media-accuracy
2. https://www.pewresearch.org/short-reads/2022/07/13/u-s-journalists-differ-from-the-public-in-their-views-of-bothsidesism-in-journalism/

X. Notes

There is always some piece of information that is required that has no other place to go. This is the place for that information.