PROJECT CONCEPT

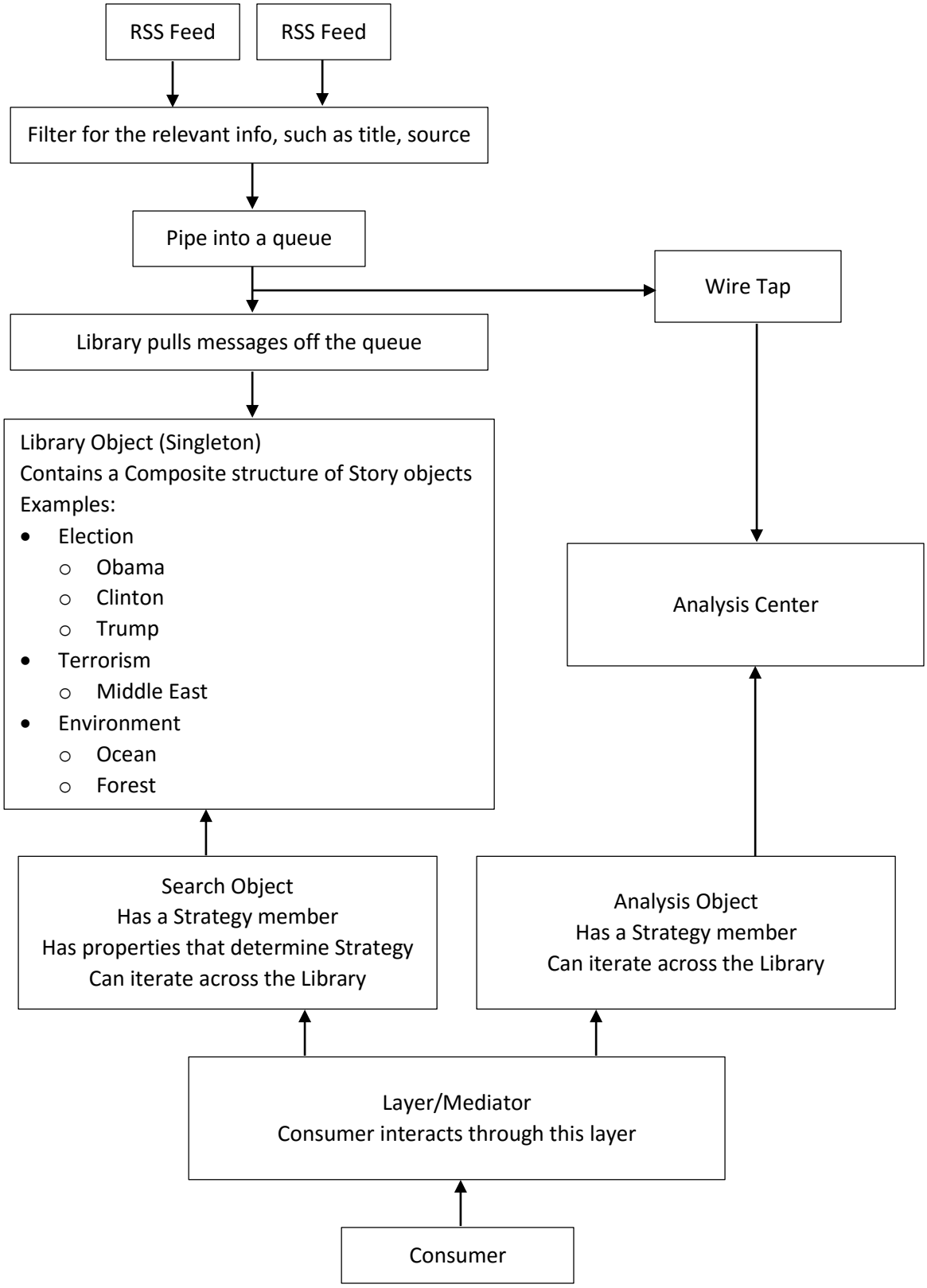
My project will accumulate stories from several RSS feeds and organize them in a directory structure for various uses, including searching and analysis of common topics.

The stories will be accumulated using Camel and sent to a queue [MESSAGE BROKER, ENDPOINT, MESSAGE, MESSAGE PROCESSOR, MESSAGE CHANNEL]. A central consumer, called a Library [SINGLETON], will read messages from this queue and create Story objects. These story objects will then be filtered [CONTENT-BASED ROUTER] and stored in a directory structure [COMPOSITE] under the Library based on their topic. The Library will have a way to identify and discard redundant messages. My initial thought for filtering these stories involves hardcoding certain keywords that all fall under a certain topic, and then filtering all the message in that topic into further subdirectories with different keywords. For example, I could filter for stories about the election with words like 'election', 'polls', 'primaries', etc. I could then further filter these stories by candidate names or some other criteria. The ability to iterate over the Library will be implemented as well [ITERATOR].

I will also write a Search class that will represent searches that can be performed on the Library. This could be as simple as searching for a particular string, but it could also be more complicated, such as searching for popular or longer stories. The Search class will be able to search the Library in different ways by specifying a search strategy [STRATEGY]. The Library can be designed with certain state variables that can be used to help regulate certain search strategies, such as a 'most popular' story.

I would potentially like to analyze the data in the Library for trends among certain stories, tracking the most popular type of story, and other things. Any methods for analysis would likely use the Template pattern [TEMPLATE], as well as particular strategies for analysis. I don't know a tremendous amount about big data, so I'm not going to be able to go for anything too complicated, but my goal would be to make this project extensible so that I could revisit it and apply big data knowledge if I learn more about that later on.

An Analysis class might be helpful in this regard. Consumers could then request statistics information from the current Library, which could be sent to another queue/topic. I'm not so sure about this last part, but it seems like an interesting extension. Another way to handle analysis might be to watch all messages that are being accumulated and sent to the Library [WIRE TAP] and copy all messages to another location for analysis by some centralized AnalysisCenter. Redundant messages may even be helpful in this situation for analysis. (This latter idea is the way I drew it in the diagram, but I'm not actually sure of the best method by which to implement the analysis.) A consumer will be able to interact with the Library through a Layer or Mediator [LAYER/MEDIATOR] (though this might be overkill).


RSS Feed
RSS Feed
Filter for the relevant info, such as title, source
Pipe into a queue
Wire Tap
Library pulls messages off the queue
Library Object (Singleton)
Contains a Composite structure of Story objects
Examples:
Election
Obama
Clinton
Trump
Terrorism
Middle East
Environment
Ocean
Forest
Analysis Center
Search Object
Has a Strategy member
Has properties that determine Strategy
Can iterate across the Library
Analysis Object
Has a Strategy member
Can iterate across the Library
Layer/Mediator
Consumer interacts through this layer
Consumer