Ritsumeikan University
College of Information Science and Engineering
Digital Governance Systems Laboratory

RITSUMEIKAN UNIVERSITY

Dynamic VAA Profile Generation Using Topic Extraction and Sentiment Analysis on Politician Tweets

Presented by: Lin Chung-Hsi, 26002304790
Course: PBL5, Professor Uwe Serdült

Date: June 11th, 2025

Table of Contents

I Recalling Progress

II Data Collection

III Preprocessing

IV Topic Extraction

V Sentiment Analysis

VI Model Diagram

VII Reference

RITSUMEIKAN UNIVERSITY

I – Recall Progress

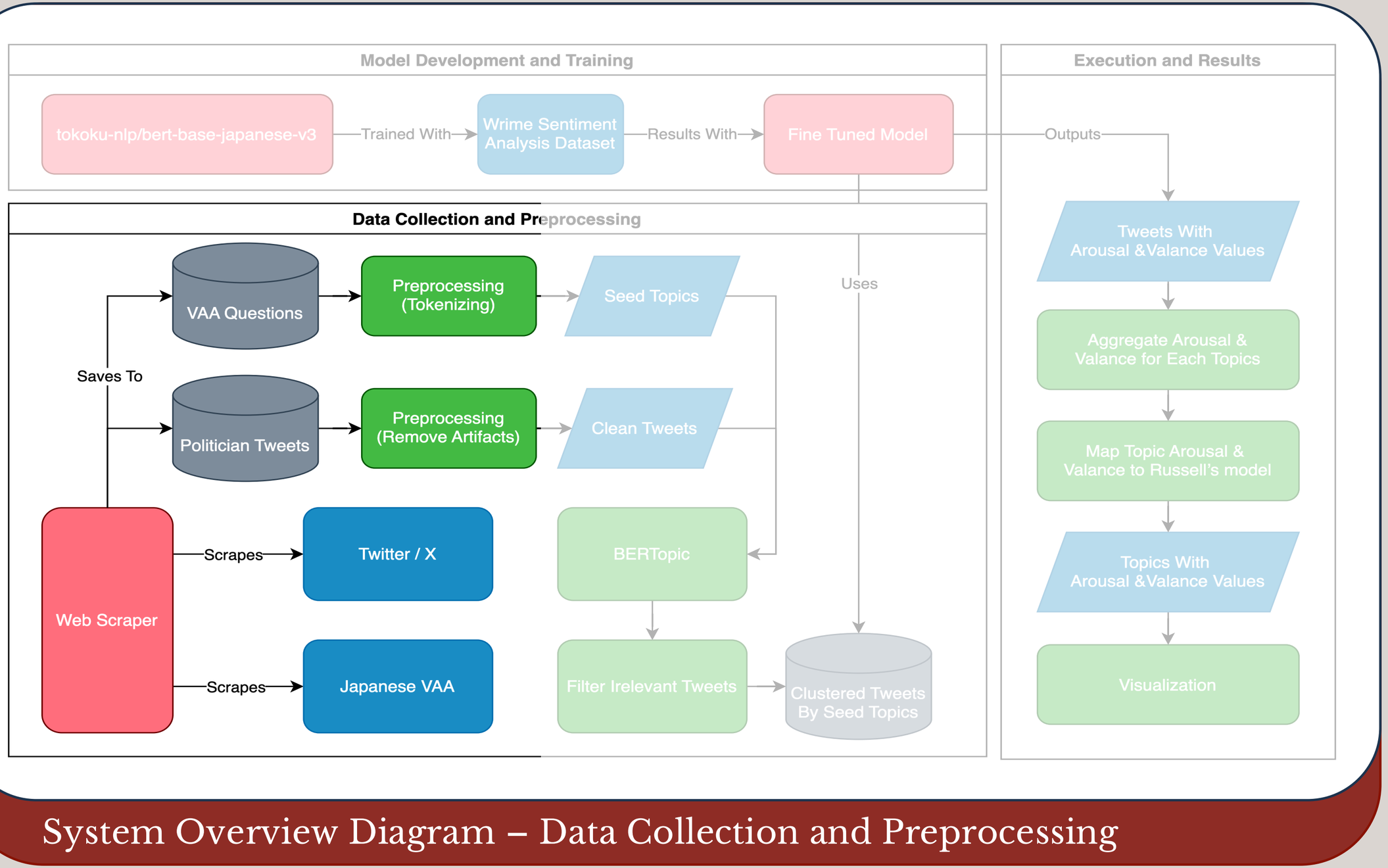
RITSUMEIKAN UNIVERSITY

Proposed Solution in Literature

A system that generates Dynamic VAA profiles from performing sentiment analysis on politician social media posts.

Major Processes/Steps in the Project

- Data Collection from Twitter & VAA. Tweets, VAA questions, and Candidates' sentiments.
- With VAA questions as cluster seed, use BERTopic to classify tweets.
- With sentiment dataset, train a model to output Arousal & Valance value.
- Aggregate values for each topic and map Arousal & Valance space to Likert scale.
- Compare resulting sentiment with scraped VAA data.



System Overview Diagram – Data Collection and Preprocessing

II – Data Collection

Target Election and Candidates

- 東京都知事選挙2024
  - 小池百合子 Koike Yuriko - 2,918,015 (42.8%)
  - 石丸伸二 Ishimaru Shinji - 1,658,363 (24.3%)
  - 蓮舫 Renhou - 1,283,262 (18.8%)
  - 田母神俊雄 Tamogami Toshio - 267,699 (3.9%)

Web Scraping

- Collecting all tweets from the candidates' official twitter account between the election date and one year prior.
- Scraping results with 450 ~ 500 tweets per candidates.

II – Data Collection

Target VAA - Tokyo MX with 投票マッチング

sentiment.txt

```
2, -2, 2, 2, 2, 1, 2, 0, 2, 2, 2, 2, 2, -2, 0, 2, 2, 1, 1, 2
-1, 2, 1, -2, , 1, 1, 2, 1, 0, 1, -2, -1, 1, 2, 2, 2, 2, 1, 1, 1
2, 0, -2, -2, -2, 2, 0, 0, 2, 2, 1, 2, 2, 2, 1, 2, 2, 1, 1, 0
0, 2, -2, -2, -2, 1, -2, 0, 1, 1, 2, -2, 1, 2, 2, -2, -2, -2, 2, 0
```

touhyoumatching.txt

高校授業料の「実質無償化」
所得制限のない高校授業料の「実質無償化」を今後も継続すべきですか？
これまで都立高校は国の支援により無償化され、私立高校については国の支援に上乗せして都が助成してきましたが、年収910万円未満が目安となっていました。2024年度からは都が所得制限を撤廃し、所得要件により国の支援が得られない場合は都が支援することにより「実質無償化」となります。保護者が都内に住んでいれば、都外の私立高校に通っても対象となりますが、都外から都内へ通学するケースは対象外です。,

公立小中学校の給食費補助
公立小中学校の給食費は東京都が全額負担すべきですか？
都内の公立小中学校の給食費を区市町村が支援する場合に都が半額を補助することになったことを受け、23区は全ての自治体で、そのほかの地域は一部の自治体で無償化が実現しました。しかし、財政的な事情から無償化や補助に踏み切れない自治体もあり、半額の補助では足りないという意見もあります。なお、学校給食法には、給食費は保護者の負担と明記されています。,

神宮外苑の再開発
樹木の伐採を含めた神宮外苑の再開発は現行の計画通り進めるべきですか？
明治神宮外苑の再開発では、老朽化が進む神宮球場と秩父宮ラグビー場の建て替えや芝生広場の整備、高層ビルの建設などが主に計画されています。再開発エリアでは樹

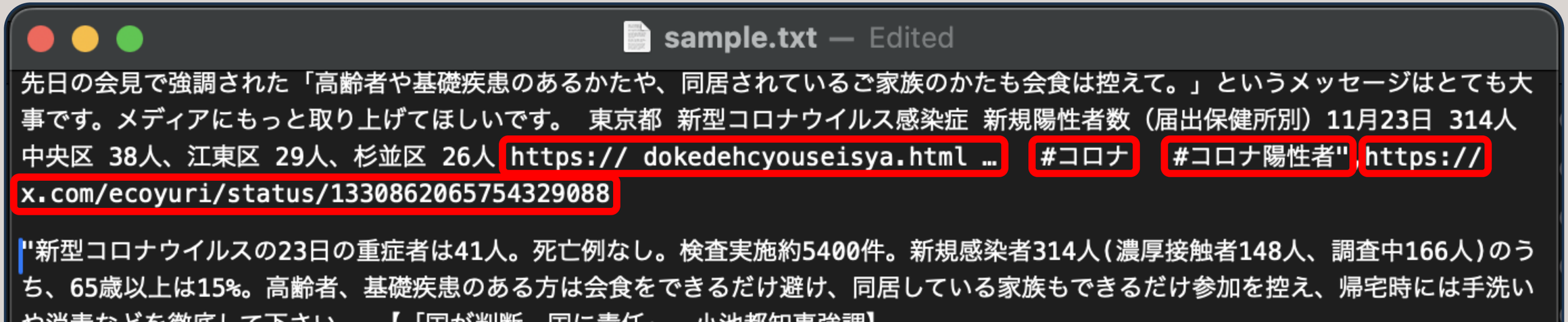
III – Preprocessing

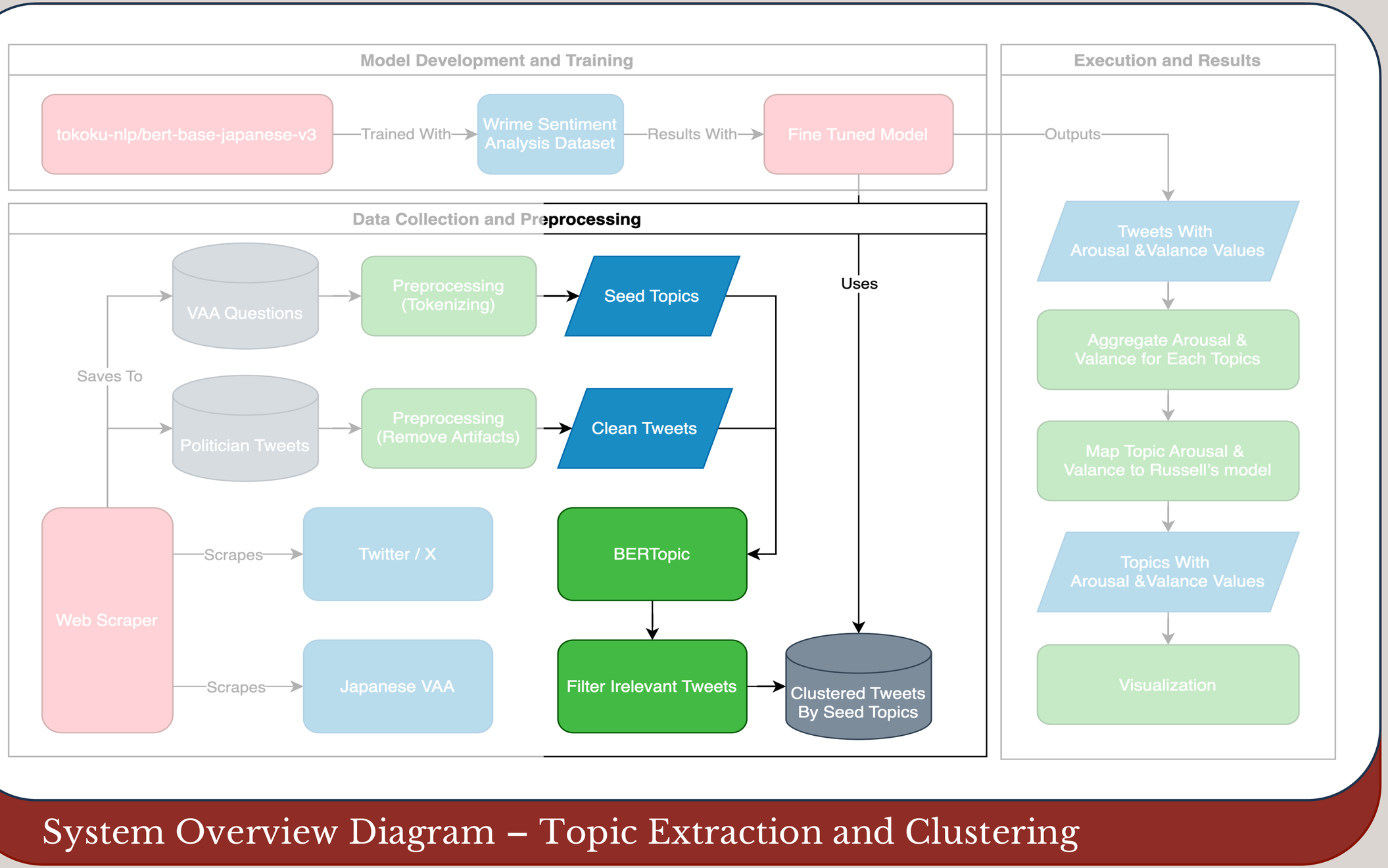
Politician Tweets

- Removing Artifacts
  - Emojis
  - 顔文字(kaomoji)
  - Hyperlinks
  - Hashtags
- English, languages not Japanese

VAA Questions

- Removing Artifacts
  - Any non-Japanese Characters
  - Stop words
  - Grammar Specific word
  - Punctuations
- Tokenization (For semi-supervised)





System Overview Diagram – Topic Extraction and Clustering

IV – Topic Extraction

RITSUMEIKAN UNIVERSITY

Zero Shot Topic Modeling

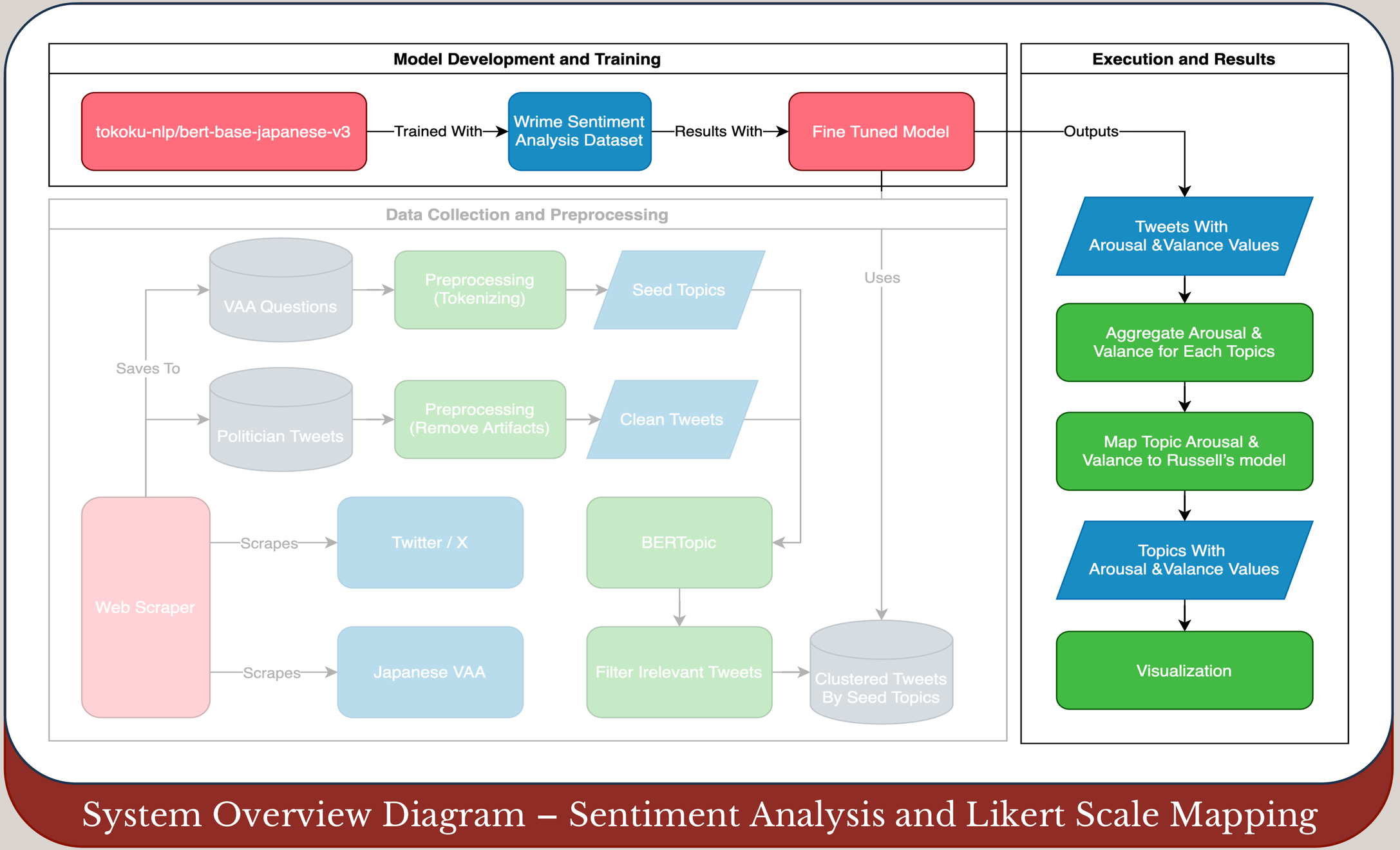
- Directly Compare Tweet to VAA Questions.
- Guarantees to cluster into the questions.
- Requires predefined topics (VAA Questions).

Semi-supervised Topic Modeling

- Let's BERTopic extract topic seed words from the VAA Questions.
- Doesn’t retain Semantics.
- The only effective method without predefined topics.

Filtering

- Discarding tweets that are dissimilar to topics with low cosine similarity score.
- The score threshold is arbitrary and non-determined.
- If implementing semi-supervised learning, ignore none seed topics.



System Overview Diagram – Sentiment Analysis and Likert Scale Mapping

V – Sentiment Analysis

Japanese Sentiment Dataset

- WRIME – 41400 annotated sentences.
  - Writer's subjective annotation.
  - 3 annotators' annotation.
  - Average of 50 participants' annotation.

Conversion between Spaces – No Literature

- Using a constant matrix to convert from. Plutchik's to Russell's emotion model.

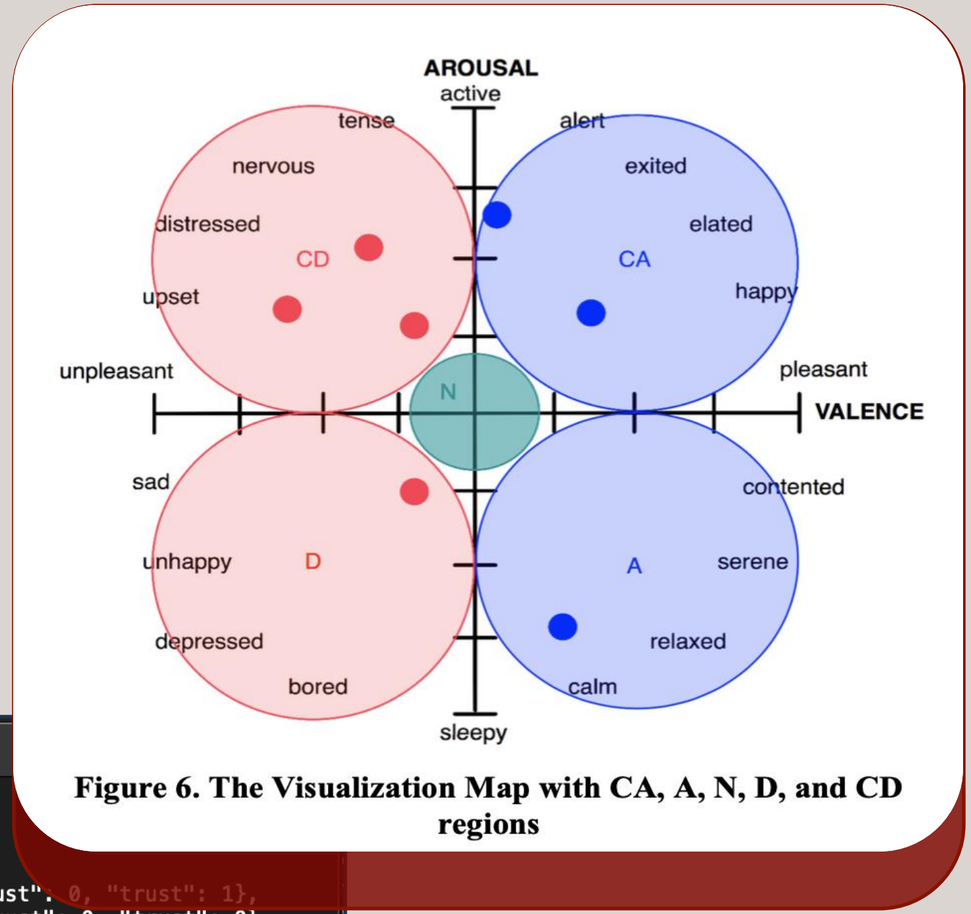
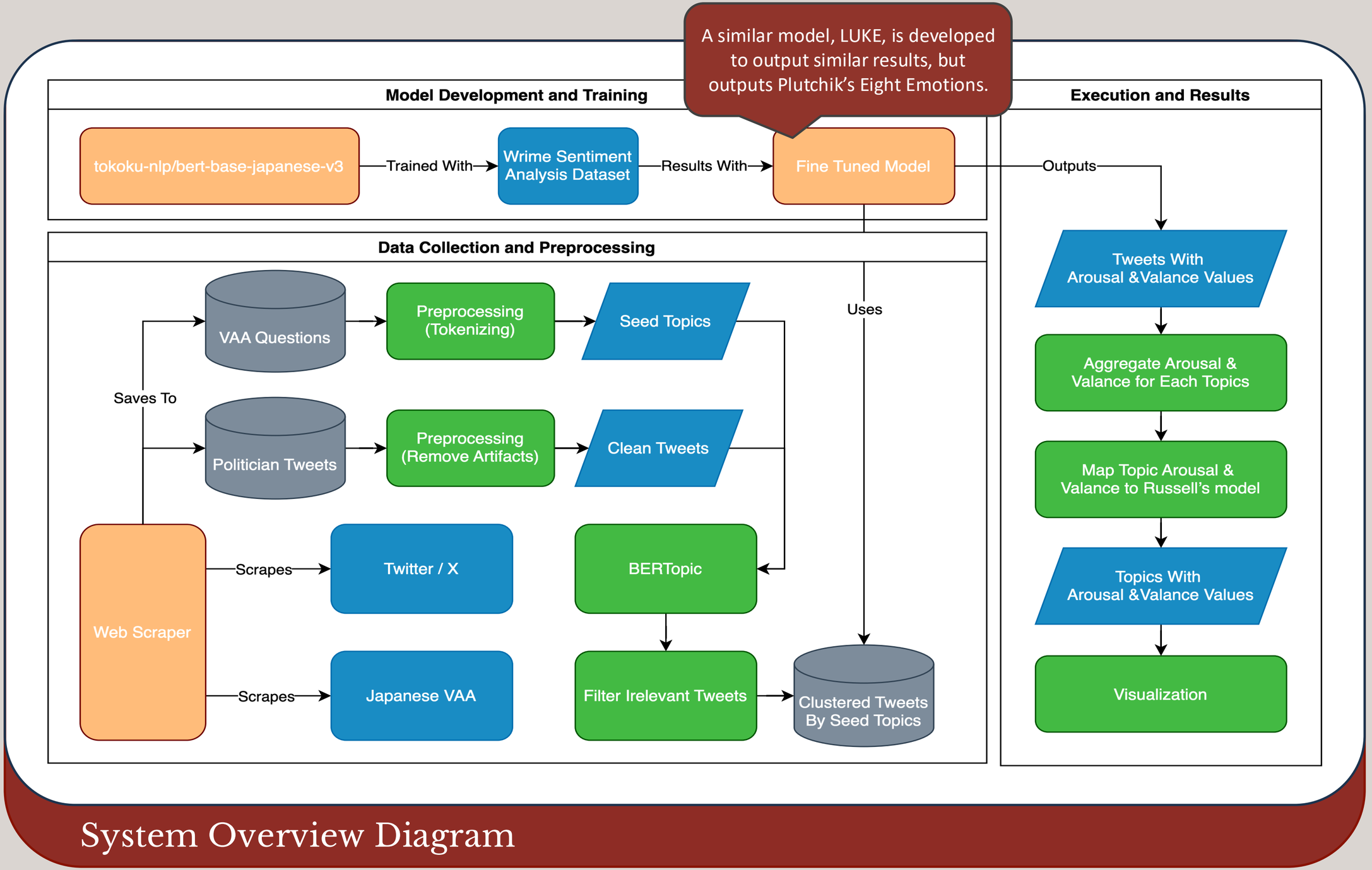


Figure 6. The Visualization Map with CA, A, N, D, and CD regions

wrime_sample.txt — Edited


```
{
    "sentence": "ぼけっとしてたらこんな時間。チャリあるから食べにでたいのに…",
    "user_id": "1",
    "datetime": "2012/07/31 23:48",
    "writer": { "joy": 0, "sadness": 1, "anticipation": 2, "surprise": 1, "anger": 1, "fear": 0, "disgust": 0, "trust": 1},
    "reader1": { "joy": 0, "sadness": 2, "anticipation": 0, "surprise": 0, "anger": 0, "fear": 0, "disgust": 0, "trust": 0},
    "reader2": { "joy": 0, "sadness": 2, "anticipation": 0, "surprise": 1, "anger": 0, "fear": 0, "disgust": 0, "trust": 0},
    "reader3": { "joy": 0, "sadness": 2, "anticipation": 0, "surprise": 0, "anger": 0, "fear": 1, "disgust": 1, "trust": 0},
    "avg_readers": { "joy": 0, "sadness": 2, "anticipation": 0, "surprise": 0, "anger": 0, "fear": 0, "disgust": 0, "trust": 0}
}
```



System Overview Diagram

Academic Reference

Ravenda, F., Bahrainian, S. A., Raballo, A., Mira, A., & Crestani, F. (2024). A self-supervised seed-driven approach to topic modelling and clustering. Journal of Intelligent Information Systems, 1-21.

Terán, L., Kakenova, U., & Portmann, E. (2017, March). Analyzing and integrating dynamic profiles on voting advice applications. In Proceedings of the 10th International Conference on Theory and Practice of Electronic Governance (pp. 62-69).

Non-Academic Reference

投票マッチング
https://votematches.go2senkyo.com/tokyo_governor_2024

WRIME
https://huggingface.co/datasets/shunk031/wrime

BERT
https://huggingface.co/tohoku-nlp/bert-base-japanese-v3

BERTopic
https://maartengr.github.io/BERTopic/index.html

NHK選挙Web
https://www.nhk.or.jp/senkyo/database/local/shutoken/20336/skh54664.html