

CSCE 771: Computer Processing of Natural Language

Lecture 21: Project Milestone #2 Presentations

PROF. BIPLAV SRIVASTAVA, AI INSTITUTE

31ST OCTOBER, 2024

Carolinian Creed: “I will practice personal and academic integrity.”

Organization of Lecture 21

- Opening Segment
 - Review of Lecture 20

- Main Lecture



Main Section

- Project Updated #2; 5 mins per student

- Concluding Segment
 - About Next Lecture – Lecture 22

Recap of Lecture 20

We looked at

- Topic Analysis
- LSA
- LDA
- Topic Classification

Main Segment

Course Project

Discussion: Course Project

Theme: Analyze quality of official information available for elections in 2024 [in a state]

- Take information available from
 - Official site: State Election Commissions
 - Respected non-profits: League of Women Voters
- Analyze information
 - State-level: Analyze quality of questions, answers, answers-to-questions
 - Comparatively: above along all states (being done by students)
- Benchmark and report
 - Compare analysis with LLM
 - Prepare report

- Process and analyze using NLP
 - Extract entities
 - Assess quality – metrics
 - Content – *Englishness*
 - Content – *Domain* -- election
 - ... other NLP tasks
 - Analyze and communicate overall

Major dates for project check

- Sep 10: written – project outline
- Oct 8: in class
- **Oct 31: in class // LLM**
- Dec 5: in class // Comparative

Obtaining Election Data

Here are a few things to do:

A) **Official data** backed by laws: state election commission

a) Find the state's election commission

b) Find the Q/As they provide. They may be as FAQs or on different web pages.

c) Collect the Q/A programmatically

B) **Secondary data** sources: non-profit

a) Find Q/As from Vote 411 which is supported by the non-profit: LWV.

For reference, for SC,

A) Official - <https://scvotes.gov/voters/voter-faq/>

B) Secondary - <https://www.vote411.org/south-carolina>

For extraction, one or more approaches:

- Manually annotating
- BeautifulSoup,
- Tika
- or other open source libraries.

<Student Name>

CSCE 771 Fall 2024: Milestone#1

1. State Selected:
2. Election data sites:
 - Official site (e.g., State Election Commission) url
 - Secondary site (e.g., League of Women Voters) url
1. Report how data collected and Q/A statistics
1. Take on NLP methods you will use and why for Q/A analysis
 1. State-level (right)
 2. Comparatively: above along states being done by peers

Initial analysis of questions (Q)

*

Initial analysis of answers (A)

*

Initial analysis of an answer (a_i)
for a question (q_i)

*

Election Q/A for Your State

- Format in .json; name file as “**xy**_qa.json”, where **xy** is the two-character US state acronym
- Fixed attributes in .json
 - state: **xy**
 - num_questions: **a**, where **a** is the number of questions
 - num_answers: **b**, where **b** is the number of answers
 - contributor: student name
- **questions**: List of Q/As with attributes for each it:
 - **q** // question
 - **a** // answer
 - **s** // source url from where the information is taken
 - **t** // time when the information is taken – UTC format
- Store it in your github repo; put in sub-dir like “project/data”
- Instructor will keep it in common place inside course github repo and share.

Election Q/As for Multiple States

- Instructor will keep it common place inside course github repo and share.
- You will be able to access Q/As of all states from common location
 - To compare data across all states

<Student Name>

CSCE 771 Fall 2024: Milestone#2

1. State Selected:

1. Optionally: do for SC
2. Optionally: compare your state and SC

2. Election data sites:

- Official site (e.g., State Election Commission) url
- Secondary site (e.g., League of Women Voters) url

1. LLM (s) used:

1. Tasks used for:
2. Comments: <highlight effort, if any>

Analysis of questions (Q)

*LLM-specific comments

Analysis of answers (A)

*LLM-specific comments

Analysis of an answer (a_i) for a question (q_i)

*LLM-specific comments

Discussion – a Paper Based on All Data?

- Contributions
 - Analysis of current situation, perspective on gaps and opportunities with NLP
 - Dataset
- Logistics
 - Target venue
 - People
 - Timeline

Concluding Segment

Lecture 20: Concluding Comments

- LLM usage for election FAQs

About Next Lecture – Lecture 22

Lecture 22 – Sentiment Analysis

- Sentiment Analysis
- Methods
 - Lexicon-based Methods
 - Learning-based Methods
- Usability considerations - Ethical Issues

18	Oct 22 (Tu)	Entity extraction, linking
19	Oct 24 (Th)	Events extraction, spatio-temporal analysis
20	Oct 29 (Tu)	Topic Analysis
21	Oct 31 (Th)	PROJ REVIEW
	Nov 5 (Tu)	
22	Nov 7 (Th)	NLP Task: Sentiment
23	Nov 12 (Tu)	NLP Task: Summarization
24	Nov 14 (Th)	Conversation Agents
25	Nov 19 (Tu)	Ethical Concerns with NLP, Trusted AI and Societal Impact
26	Nov 21 (Th)	Working with LLMs for NLP Tasks - programming, Quiz