Knight Foundation School of Computing & Information Sciences

Spring 2021 Senior Design Project

ANAnSI

Student: Joshuan Jimenez

Mentor/ Project Managers: *Antonela Radas, Mustafa* **Product Owner:** *Dr. Mark Finlayson*

Instructor: *Dr. Masoud Sadjadi*, Florida International University

Problem

Qualitative temporal graphs that are derived from annotations on text| e.g., TimeML annotations explicitly reveal only partial orderings of events and times. For many purposes, however, a total ordering (i.e., a timeline) is more useful. We adapt prior work on solving point algebra problems to the task of extracting multiple timelines from timeML annotated texts and demonstrate, for the first time, an exact, end-to-end solution, which we call TLEX (TimeLine Extraction). In collaboration with MITRE, we must convert the ANAnSI, also known as Advanced Narrative Analytics System Infrastructure, prototype system output to the TimeML standards.

Current System

The purpose of the ANAnSI is to map the ANAnSI data structure to the TimeML structure for Timeline extraction.

The current system accepts a JSON file and starts extracting the individual information needed to create the ANAnSI data structure.

After creating the structure, the system will start the mapping process using an adaptor to create the TimeML object, which extracts the timeline and outputs the information.

Requirements

The system will allow users to pass JSON files in order to do timeline extraction.

After the system receives the files from the user it will create a ANAnSI data structure.

After the data structure is created the system will map the ANAnSI data structure to TimeML data structure.

Afterward the system will begin the process of timeline extraction.

Contribution

Before creating the ANAnSI data structure each team member was assigned various parts of the data structure to work on, as well as different methods in order for the parser and adaptor to work. The contributions done by me are as follows:

- Term Class
- Term Interface
- Junit Testing for Term
- Javadocs for Term Class
- parseTerm method for JsonParser class
- Main method of the JsonParser: parseJSon
- convertLinguisticEvent method for Adaptor class
- Revision and debugging of the Adaptor
- Validation of output to ensure correct conversions

Implementation

The backend of the library was done in Java 14, JUnit was used to test the all-possible cases that might give errors. Trello was implemented in order to work on an Agile Environment and setting up tasks for the Team. lastly SVN(Subversion) was used to upload the code and store it.

Screenshots

Figure 1: Sample Output of Links

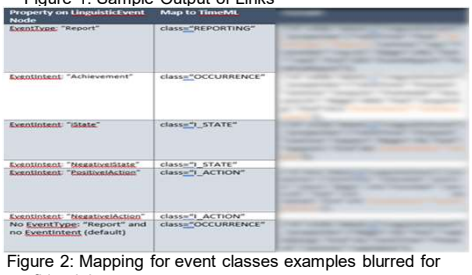

Figure 2: Mapping for event classes examples blurred for confidential reasons

Summary

The project is currently in development. This new Technique for timeline extraction was created using techniques for extracting Timelines from natural language texts that were Annotated with the TimeML markup language TLEX draws ideas from both NLP and the AI sub-field of temporal constraint problem solving.

Our ANAnSI team working alongside MITRE was in charge of implementing the Advanced Narrative Analytics System Infrastructure (ANAnSI) prototype which entails parsing json files to map the data structure to TimeML to extract timelines.

Take Away(Learn)

From this project, what I have learned is:

- Working in an AGILE Environment
- How SVN functions
- How to do and implement Junit testing
- Neo4J and how it functions as a database
- How to meet deadlines in order to move on to other aspects of the project
- Timeline extractions
- Natural Language Processing

Resources

1) Essential Scrum: A Practical Guide to the Most Popular Agile Process (please read only chapter 2 and chapter 4)
2) J. F. Allen (1983) Maintaining Knowledge about Temporal Intervals
3) Saurí et. al. (2006) TimeML Annotation Guidelines
4) Pustejovsky et. al. (2003) The TimeBank Corpus
5) TLEX: A Formally Correct Method for Extracting Exact Timelines from TimeML Temporal Graphs
6) ANANSI-TIMEML mapping schema
7) Neo4j courses

Acknowledgement

The material presented in this poster is based upon the work supported Dr. Mark Finlayson, Mustafa Ocal, Dr. Karine Megerdoomian. I'd like to thank Antonela Radas and The ANAnSI team for assistance and cooperation received throughout this process.