Matthew Van Soelen, Tom Orth, Justin Pabon, and Kathleen Burke
CSC 335-01 and LNG 371/HON 270
Professor Degood and Dr. Steele
Stage IIa

- **Problem statement**

    Within the Trentoniana Collection and the connected Internet Archive, there is a lack of an easy to use interface to favorite or bookmark transcripts. Additionally, there seems to be no way to bulk download transcripts that you like. Each transcript entry seems to have a lack of information associated with it, which makes it hard for the user to tell if they found the right transcript.

- **Objective of the module**
    The objective of the module we will design is to allow for users to favorite transcripts that they search for. The module would also provide a cleaner interface to allow for easier use and clearer search results. It would provide a better user experience overall while searching for transcripts. It would help users be able to efficiently conduct research and access necessary files.

- **Description of the desired end product, and the part you will develop for this class.**

    The end product would be a web application. It would support 2 user types. First, a general user that can view transcripts, download them and favorite them to refer to later. The second user would be a librarian or other administrator who could update existing transcripts as needed and add new ones. The search functionality would be able to filter by keyword in title, keyword in the transcript, keyword in summary, and by participant names in the transcript entry.
    Each transcript would have an audio file, title, transcript text, transcript file, summary and a list of participants in the transcript to ensure that each entry has descriptive information.

- **Description of the importance and need for the module, and how it addresses the problem.**

    There is an importance and need for this module to ensure that there is an easy way to retrieve these transcripts. There is a rich history with the Trentonia transcripts and there needs to be a module that provides an informative way to understand these transcript entries. Our module addresses this problem by providing an easy to use interface as well as a more detailed entry description.

- **Plan for how you will research the problem domain and obtain the data needed.**

We plan to research the domain area in two ways. The first way is by exploring the website provided to us during the first collaborative meeting: trentonlib.org/trentoniana/audio-visual/. Secondly, we will explore the domain area by discussing what occurs in LNG-371/HON 270 since the course is directly related to the transcripts. We will obtain the data through the transcripts provided by Dr. Steele to the LNG/HON students so that we can store this into our database.

- **Other similar systems/ approaches that exist, and how your module is different or will add to the existing system.**

The current similar system that exists is the internet archive that is linked on the trentonian section of the trenton public library website. Our module will be different in the sense that it is currently only meant to hold data in the specified format that the transcripts have. The internet archive stores many different types of data and information that can bloat the search results and make it hard to find the right things. As a result of this, our module will be more user friendly and less complicated since we are only supporting one type of data.

- **Possible other applications of the system (how it could be modified and reused.)**

Since this system needs to support authentication, the user and librarian authentication services could be used to authenticate other applications for the organization. Additionally, if the participants of the transcript are important to other sources of data, the participants data, which will be stored in its own table, could be used by some other application if we exposed it through an API endpoint.

- **Performance –specify how and to what extent you will address this.**

We will address performance by optimizing all our queries as we study more and more about databases. We anticipate that our queries to start off may be inefficient in some areas but once we progress into the semester and the project, we expect to be able to upgrade all the ones that can be made more efficient.

- **Security –specify how and to what extent you will provide security features**

Given the nature of the data, the only private information that will need to be encrypted is the password for the users. This is due to the transcripts being publicly available. We will ensure proper rights and access to data by authenticating all our routes so that no one can access information or perform actions that they do not have the privilege to do.

- **Backup and recovery –specify how and to what extent you will implement this.**

We can write scripts that will support both backup and recovery. Postgres has support for data backup and recovery, as described here: https://www.postgresql.org/docs/9.1/backup.html. Additionally, the upload files can be zipped and stored separately and be unzipped and moved to the proper directory during recovery.

- **Technologies and database concepts the team will need to learn, and a plan for learning these.**
    There are three categories for technology that we will need to familiarize with:
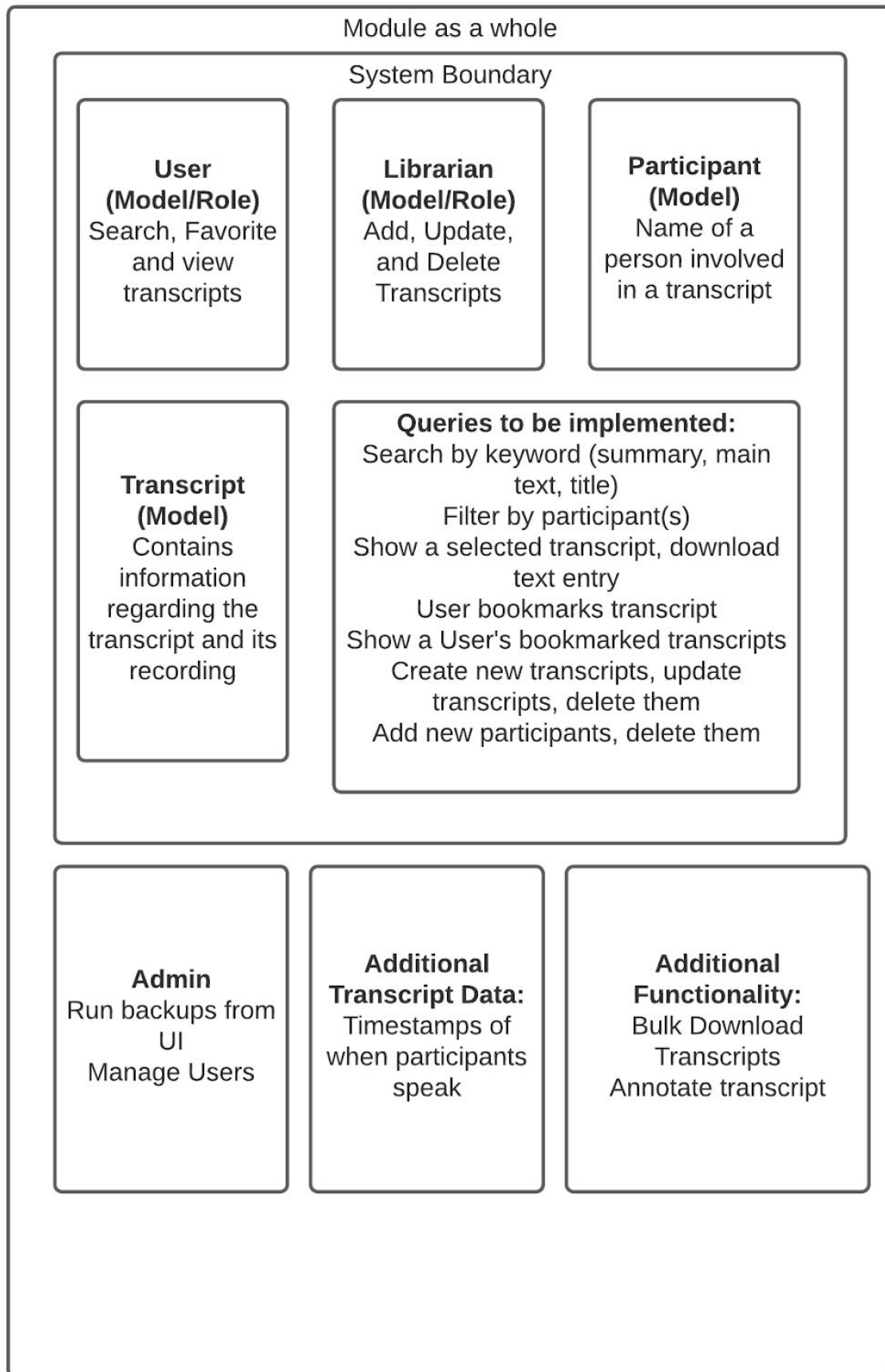   - Database
        - We will be using postgres: https://www.postgresql.org/docs/
    - Back-End
        - We will be using Python and Flask as an application server layer. We also would need a library to interact with the database. We will explore SQLAlchemy and psycopg2.
        - For Python: https://www.udacity.com/course/introduction-to-python--ud1110
        - For Flask: https://flask.palletsprojects.com/en/1.1.x/tutorial/
        - SQLAlchemy: https://flask-sqlalchemy.palletsprojects.com/en/2.x/quickstart/
        - Psycopg2: https://www.postgresqltutorial.com/postgresql-python/
    - Front-End
        - We will be using HTML, CSS and the built in templating engine for Flask. We will learn these through w3schools.com

    There is also a number of database concepts we will have to learn
    - Queries
    - Relationships
        - We expect to have 2 N:N relationships
            - N:N between a User and a Transcript
            - N:N between a Transcript and a Participant
    - Optimizing Queries
    - Optimizing database structure for efficiency and scale

    For these concepts, the 315 students will learn this through the class and ensure that all group members have at least a surface level understanding of the concept, depending on their role in the team.

- A diagrammatic representation of the system boundary that specifies what data you will model and which queries you will implement. (next page)

Module as a whole

System Boundary

User (Model/Role)
Search, Favorite and view transcripts

Librarian (Model/Role)
Add, Update, and Delete Transcripts

Participant (Model)
Name of a person involved in a transcript

Transcript (Model)
Contains information regarding the transcript and its recording

Queries to be implemented:
Search by keyword (summary, main text, title)
Filter by participant(s)
Show a selected transcript, download text entry
User bookmarks transcript
Show a User's bookmarked transcripts
Create new transcripts, update transcripts, delete them
Add new participants, delete them

Admin
Run backups from UI
Manage Users

Additional Transcript Data:
Timestamps of when participants speak

Additional Functionality:
Bulk Download Transcripts
Annotate transcript

- 1-page quad chart (See next page)

TCNJ Providing a User Friendly Module for Transcript Data
Thomas Orth, Matthew Van Solen, Justin Pabon, Kathleen Burke

Need

- Current website is hard to use
- Too much other data stored current solutions

Approach

- Design a simpler interface
- Allow for users to bookmark favorites
- Provide more information about each entry to allow for easier searches

Benefit

- Increased User Experience
- More robust handling of data
- Easily obtain the necessary trans -cript

Competition

- The interface would be superior
- The entries would be more informative
- The search results would not encumbered by unrelated data

02/10/2021