2007

The WorkGroup – CS 501

Michael Wang, Dan Rassi,
Sam Phillips, Ray McGill,
and Krzysztof Findeisen

[PHASE II REPORT]

This document will describe all of the requirements for the eRulemaking Project. It will outline all the specifics associated with the development and project outline of this system as current as our second iteration.

Table of Contents

The Workgroup

Upson Hall
Cornell University
Ithaca, NY 14850

March 12, 2007

Tom Bruce
Legal Information Institute
Cornell Law School Myron Taylor Hall
Ithaca, NY 14853

Dear Mr. Bruce:

We are pleased to contact you again regarding the eRulemaking project we are developing for the LII. As part of the CS501 class we have divided this project into four major development parts; this letter comes to you between the second and third major phases of development. Attached you will find a summary of the work which we have been performing for the last three weeks as well as a chart showing current progress and future steps.

Sincerely,

Michael Wang, Dan Rassi, Sam Phillips, Ray McGill, and Krzysztof Findeisen
The Workgroup

1. Executive Summary

We have made significant strides in our eRulemaking project for the LII. The goal of the project is to create a report outlining the anticipated requirements that a complete eRulemaking system may have, and to provide a demonstrative prototype implementing a significant portion of those requirements. This goal reflects the fact that the requirements of the full project are not yet well known even to the LII and an important service which we will provide is working with many stakeholders to elucidate those requirements.

To make sure that the many stakeholders for this project are involved and informed we have nd currently are taking several proactive steps early in this project. We are using an iterative design approach to prototype functionality early to get feedback from all stakeholders. Additionally we have solicited input from the LII Law Student Annotators and read transcripts provided by Dr. Wagner about FTA workflow. We have attended a full Cornell eRulemaking Initiative (CERI) meeting where we spoke with NLP researchers about their requirements and standard formats.

We have implemented an early prototype system based on the open source Drupal content management system (CMS). This prototype demonstrates the feasibility of annotation functionality using AJAX technology. By producing this UI early in the project many design issues will be discovered as requirements are discussed in the context of the prototype UI.

The immediate future work will be in two main areas, the first will be a continuation of the iterative design of the UI, adding more functionality to the current web service and discussing with the stakeholders to get immediate feedback. The second major area will be the implementation of the back end, which is a major portion allowing XML in the ATLAS format to be imported and exported from our system. ATLAS is the standard which is used currently by the NLP group.

2. Introduction to the eRulemaking Project

2.1 Introduction

The project we're undertaking is to explore design and implementation considerations for an eRulemaking client program and to implement a prototype for the LII. It is important to note that a very significant part of this project is the investigation of what features and UI a fully completed eRulemaking program might have. However, the implementation and exact specification of such a full-featured client is outside the scope of the current project. This project represents the first phase of a multiphase development process to create the eventual system rulemakers will use; as such our most important deliverable will be a report summarizing the design considerations and a manual describing how to use and extend our implementation. Additionally we will implement a prototype system that will demonstrate a significant amount of functionality.

2.2 Document Summary

This document serves three distinct purposes:

1. Increase the visibility of the project by communicating back to the LII our understanding of their requirements.
2. Increase the visibility of the project by communicating back to the LII of our progress so far and future steps.
3. Serve as part of the documentation package for this project, to be used by future groups working with CERI

3. Project Detail Description

3.1 General Description

Our project is a subproject for CERI; quoting from [1]

> *"[We] propose to apply and develop a range of methods from the field of natural language processing (NLP) to create NLP tools to aid agency rule writers in:*
>
> - *Organization, analysis, and management of the sometimes overwhelming volume of comments, studies, and other supporting documents associated with a proposed rule; and*
> - *Analyzing proposed rules to flag possibly relevant mandates from the large number of statutes and Executive Orders that require studies, consultations, or certifications during rulemaking."*

The end-product of the CERI project will be a fully-featured computer service which accomplishes the above mentioned goals. The Workgroup will help accomplish goals related to the "consumer-facing" side of the project, allowing agency rule writers to easily organize, analyze and manage comments associated with a proposed rule.

3.2 Product Functions

The fully featured product as will be made available to Rulemakers will need to combine the current agencies' workflows along with a variant on the current LII workflow. The eRulemaking client system must allow users to add, modify and delete annotations on comments; functionality currently provided by Callisto to the LII student annotators. The system must also allow Rulemakers the ability to sort, filter and process comments based on their own specific agency workflow, an example of which is shown in §3.3.2

Our prototype implementation will have a subset of the functionality needed for the full system, the full functionality is not yet known, however the scope of this project as currently know is shown in §3.4.

3.3 User Characteristics

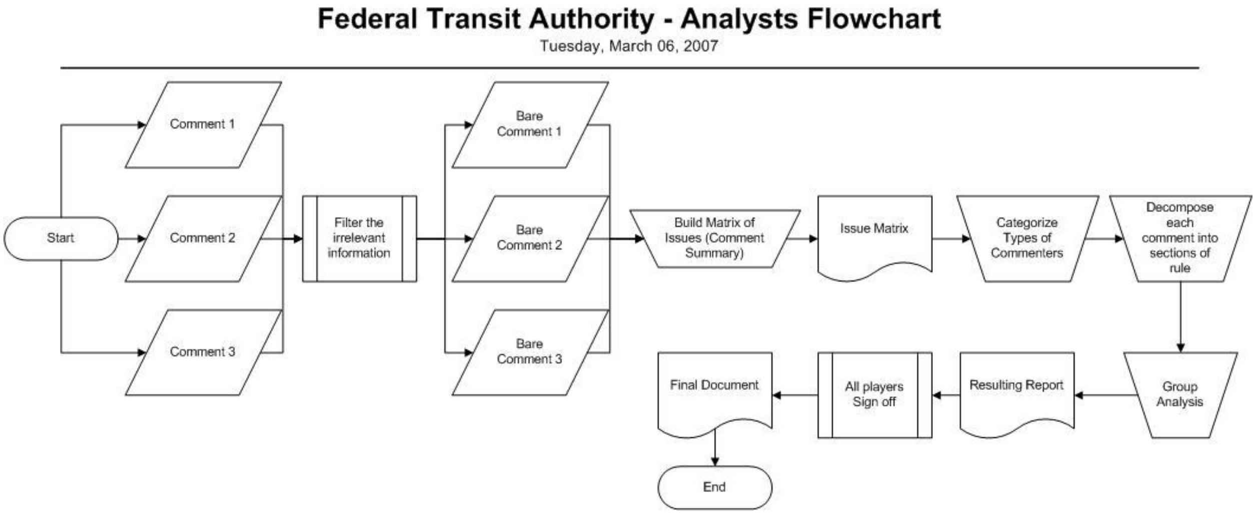
The users of this system will be varied. The system will first be used internally to CERI to transition away from Callisto, so the first users will probably be law student annotators who will use the system to annotate static, relatively small sets of comments.

The final users of the system will be Agency Rulemakers who will use the system to train NLP software to automatically classify new comments; they will also use the system as a tool to augment their current workflow of evaluating comments.

Intermediate users will be different groups within CERI who will use this project as a framework for expansion. As an example, the NLP group may use the system to create an interface the student annotators can use to examine inter-annotator agreement. Many other unforeseen expansions to the system are likely to arise in the future as well in the medium term. As it is expanded and refined to create usability for both internal CERI uses and in anticipation of eventual deployment in many different agencies, accommodating these expansions is part of the scope of this project. Also, the technical teams who will want to extend our prototype are also users of the system.

3.4 Federal Transit Authority Analyst Workflow

Analyst Flowchart:


Federal Transit Authority - Analysts Flowchart
Tuesday, March 06, 2007
Start
Comment 1
Comment 2
Comment 3
Filter the irrelevant information
Bare Comment 1
Bare Comment 2
Bare Comment 3
Build Matrix of Issues (Comment Summary)
Issue Matrix
Categorize Types of Commenters
Decompose each comment into sections of rule
Group Analysis
Resulting Report
All players Sign off
Final Document
End

Description:

This flow chart shows the common analyst workflow. The users are the Analysts. They start at the left side 'start' unit and end at the bottom side 'end' unit. The process includes:

(a) The analysts receive lots of comments by e-mail and filter the irrelevant information to get the bare comments.

(b) Building an issue matrix (comment summary) to read them.

(c) Categorizing the types of commenter and organize them into sections of regulations.

(d) Doing group analysis and get results report.

(e) They sign off and archive the documents of final results.

3.5 Scope

Since this project is the first phase of a multiphase development process the scope of this particular project is not entirely defined. The best way to describe the limited scope of this project as compared to the scope of the full featured project is to describe a set of minimal requirements which will be implemented by the prototype system. It is our understanding from discussions with LII that they require that our delivered product shall:

1. **Features for Annotators**

1.1. **Immediate Features**

1.1.1. Allow users to add and delete annotations to comments from a set of issue tags

1.1.2. Provide a UI which works well for up to 50 issue tags per regulation

1.1.3. Allow a subset of users to modify the issue tags, adding and collapsing issue tags (as Regulation Administrators).

1.1.4. Allow users to choose different comments from the full comment set, the UI shall support >300 comments easily

1.1.5. Comment access will be asynchronous, that is, each comment may only be edited by one user at a time.

1.1.6. Comments will support only one set of annotations, that is annotations will be per-comment and not per-user

1.2. **Optional Features**

1.2.1. Allow annotators to feed changes in annotation back into the NLP software to see how classification changes.

1.2.2. Allow multiple regulations each with multiple comments and users who are allowed access to those regulations, the UI should support >20 regulations.

2. **Features for Agency Rule Makers**

2.1. **Immediate Features**

2.1.1. Support filtering comments by issue tag

2.1.2. Support viewing summaries of comments which match filters

2.1.3. Support viewing comment text with annotation highlighting

2.1.4. Support Firefox 1.5 and Internet Explorer versions > 5.5

2.1.4.1. Some “look and feel” features may not work on some browsers, but functionality won’t degrade

2.2. **Optional Features**

2.2.1. Support issue hierarchies

2.2.2. Support “flags” non-issue related full comment metadata

2.2.3. Support individual creation of flags

2.2.4. Support automatic workflow metadata (last read by, annotation / flag history, etc)

3. **Features for Developers Extending Our System**

3.1. **Immediate Features**

3.1.1. Support import and export of XML annotation data in the ATLAS format

3.1.2. Design system with low coupling between the front-end, middleware and back-end

3.1.3. Provide documentation of technical functionality for programmers to use to extend system.

3.1.4. Provide documentation of requirements / design process for other groups to learn from our design process.

3.2. **Optional Features**

3.2.1. Work with “inter-annotator agreement” code produced by NLP group

3.2.2. Provide simple setup-and-install scripts for deployment at new sites.

3.6 General problems and risks

The risks in our system are as follows:

1. There is a time-limit imposed by the CS501 class and the end of the semester, care must be taken to ensure we stay on track to accomplish our goals.
2. Some optional functionality depends on the NLP group in CERI, who may not have sufficiently robust software in place for automatic interactions.
3. Our group is new to collaborative web development, adjustments to collaborative web paradigms may take more time that we anticipated.
4. Relatively little is known about potential users of the system at Federal Agencies, so care must be taken to ensure the final system will be useful and useable by agency Rulemakers.

4. Functional Requirements

4.1 Introduction

To succinctly capture the functional requirements of the prototype system we have created a set of flowcharts that show the flow of the system, they reflect our understanding of the LII's requirements.

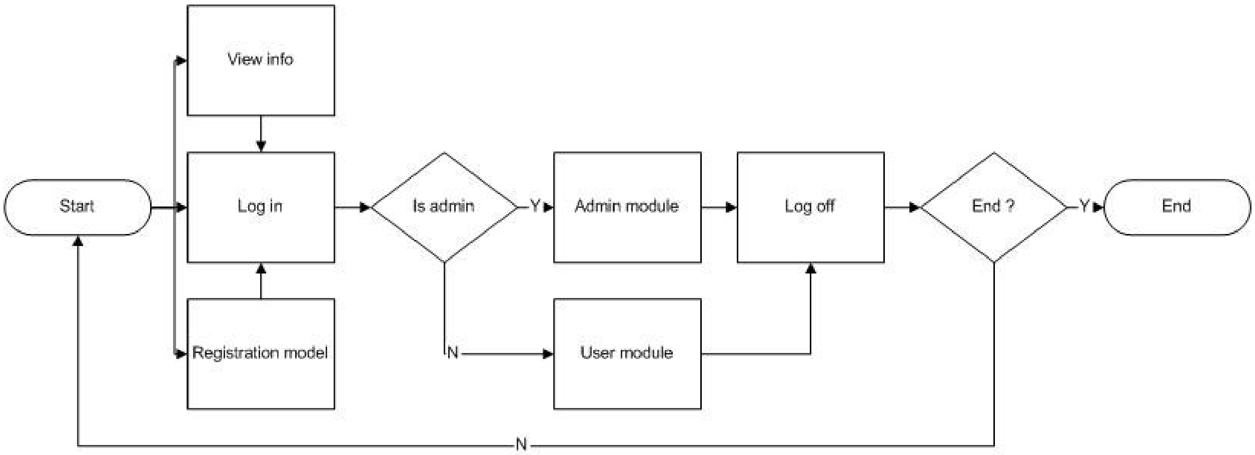
4.2 Flowchart Description

A flowchart is a graphic and schematic representation of the logic or steps in a program or system. It represents how a program or activity moves through various processes or program routines. It uses symbols to represent the activities, and uses arrows to represent the direction of activity through the processes. Flowcharts can define the behavior of a single program or a system (a combination of programs).

In the E-rulemaking project, we use flowcharts to define and show the functions and main components of the whole system. Each flowchart represents a view of one specific module. The six flowcharts are: main module flowchart, registration module flowchart, admin module flowchart, user module flowchart, view module flowchart and annotation module flowchart.

Please see the details below.

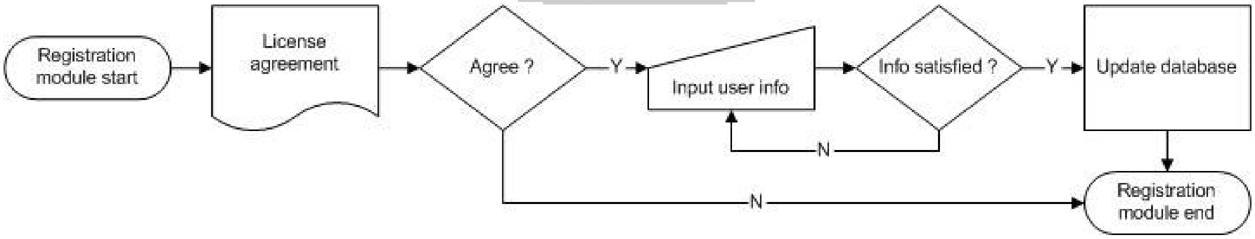
4.3 Main Module Flowchart



Description:

This is the top level flowchart that describes a global process of the e-rulemaking system. Assume that you are a user or administrator. You will start from the left side 'start' unit. You can first read the relevant information of the system (e.g. announcements), and then sign into the system. If you are a new user or administrator, the system will let you register. The registration module provides this service. After login, the system will see if you are a user or administrator. The system will then lead you to the correct modules to start further work, since a user may have several roles, you can start again in this module with the logic circulation. Before you leave, you must sign off and affirm. This flow ends at the right side 'end' unit.

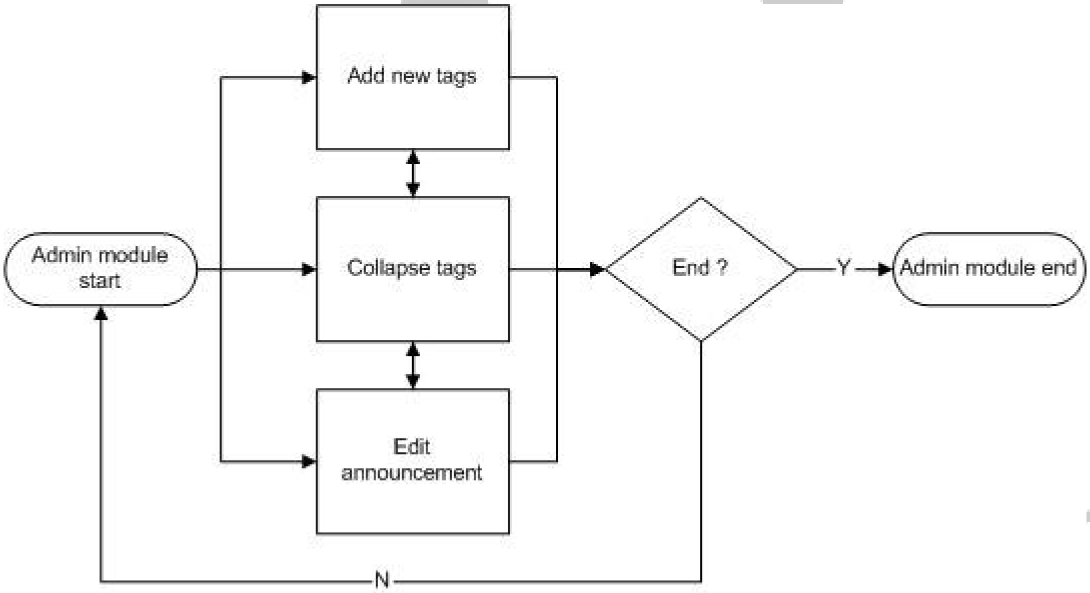
4.4 Registration Module Flowchart



Description:

This module is one of the sub-modules of main module. It provides registration function. You will start at the left side 'registration module start' unit to read the license agreement first and choose if you accept it or not. If accept, you will input some registration information under the system instructions. After that, the system will save this information and update database if the data are verified. Otherwise, you have to check the information you entered and try submit again. This flow ends at the right side 'registration module end' unit.

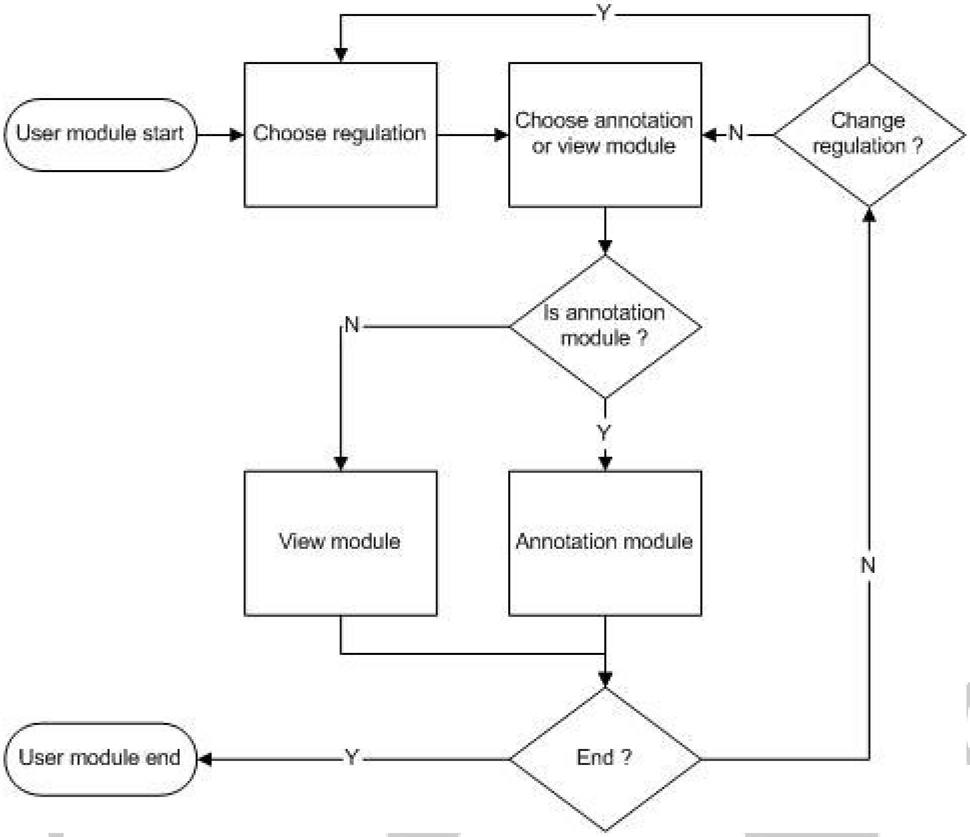
4.5 Admin Module Flowcharts



Description:

This module provides function for administrators. You will start at the left side 'admin module start' unit. There are three parallel functions: adding new tags, collapsing tags and editing announcements. The administrator may complete many of these task in one session using the logical circulation. This flow ends at the right side 'admin module end' unit.

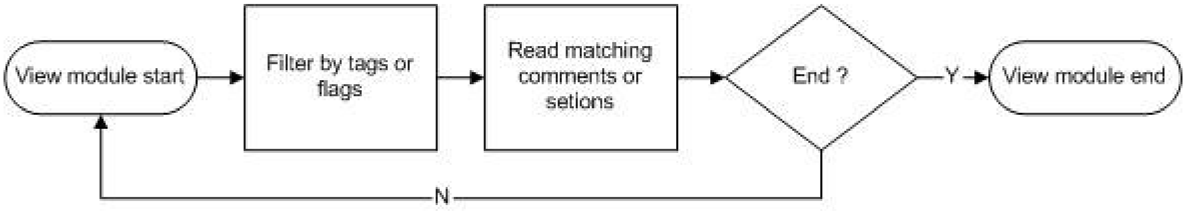
4.6 User Module Flowchart



Description:

This module provides functionality for common users. You will start at the top-left side 'user module start' unit. Following the clients' requirements, we let you start from a regulation. After you choose one, you can simply decide which you want to process next, viewing or annotating. The system will allow you to choose and enter to view module or annotation module. You can do it again and change regulation and modules with the logical circulation. This flow ends at the bottom-left 'user module end' unit.

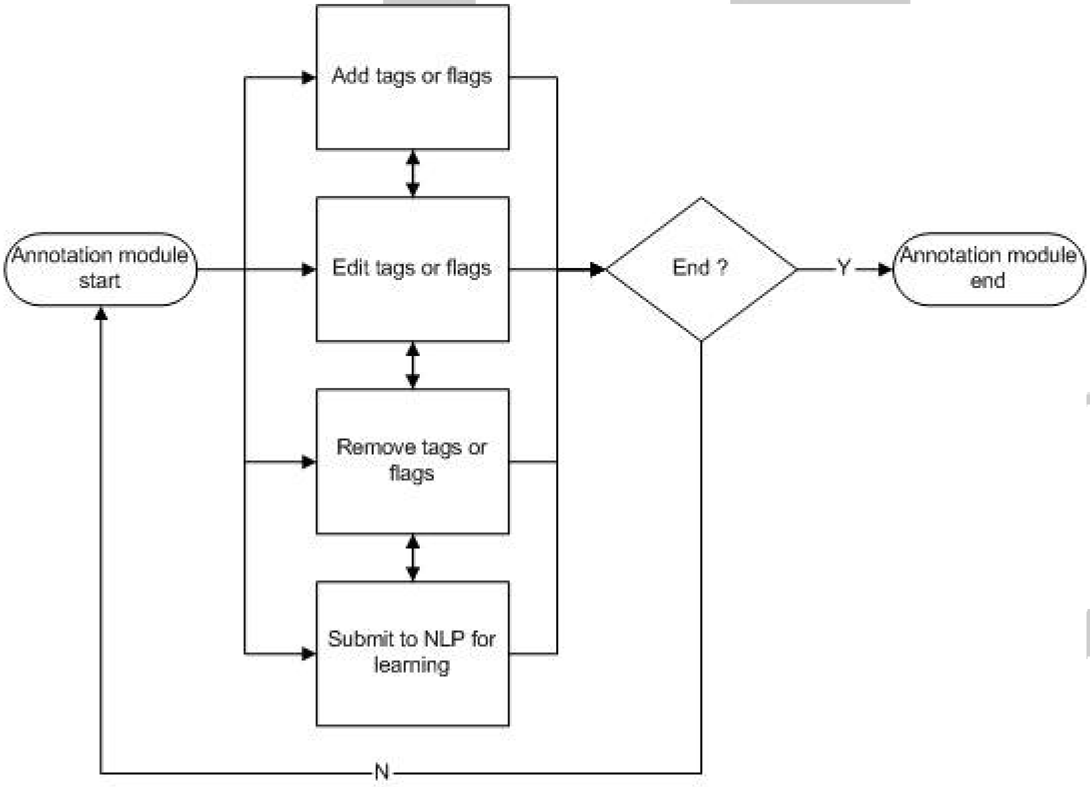
4.7 View module flowchart



Description:

This is a read only module for users. You will start at the left side 'view module start' unit. This module provides a filter to users. You can search tags or flags and then view them in the view page. You cannot modify tags in this module. And you can do it again in this module with the logic circulation. This flow ends at the right side 'view module end' unit.

4.8 Annotation Module Flowchart



Description:

This module provides function for users. You will start at the left side 'annotation module start' unit. There are four parallel functions: adding new tags and flags, editing tags and flags, removing tags and flags and submitting annotations for further learning. Many of these actions can be completed in one session using the logical circulation. This flow ends at the right side 'annotation module end' unit.

4.9 Technical Issues

There are a number of technical issues that may hinder either our development process or the usage of the system after it is developed.

Cross-browser compatibility will be a problem during development of the web application front-end due to our heavy JavaScript usage combined with notorious JavaScript standards incompatibilities across browsers. We hope that we will be able to minimize this problem and make the issue transparent to users by taking time to insure that our application works for major browsers that annotators are likely to use.

Our database system will possess some finite capacity and processing speed that may not function as expected under extreme circumstances, for example, a very large comment, a large number of annotations for one comment, or a very large number of comments. This problem can be addressed as the issue arises by scaling up the database system to match the particular user's needs. Similarly, our web server can only satisfy a finite load of users and page requests.

Another issue that may arise is the problem of keeping our code base extensible. We need to constantly keep in mind the ideas of modularity and loose coupling and provide interfaces between our code modules that allow a future user to swap an underlying implementation (database system, xml annotation output format, etc) without having to modify dependencies in the software except the interface implementation.

5. Non-Functional Requirements

These requirements are attributes that are indirectly related to the desired functionality of the system. They do not specifically outline technical behavior of the system, but more so the non-technical additions to the system that make this project a successful one.

5.1 Documentation

The LII eRulemaking Initiative must have detailed and fully understandable documentation to aid the next group of developers in understanding the implementation, installation, operation, and maintenance of all the components of this framework-system. This documentation will allow future developers to modularly add features and enhancements to the existing framework-system. Thus the structure and framework of the system is a crucial aspect of this project and must be constantly adjusted to support the possible future additions by other developers.

5.1.1 Design Document

The design document will serve as an adequate manual for the system implementation to the LII eRulemaking Initiative. This document will give an overview and detailed explanation of the framework-system which allows the viewer to understand and easily navigate through each of the components available on the current system. This design document, however will not be intended for use by anyone other than internal CERI users simply because this system is not the final product, but more so a working framework from which to build more advanced implementations of this system. Diagrams and screen-shots will be provided with this manual where necessary to facilitate this understanding. Upon completion of reading the document, the reader will have sufficient knowledge to use all the functions of the system and even learn more about how exactly some of the components work. Hopefully this document will not only provide future developers with an incentive to document their design considerations, but also serve as a basis from which to build a working manual for the final system.

5.1.2 Source Code Documentation

The source code documentation for this project will include PHP class and function definitions, as well as comments interleaved into the source code. This documentation will allow the reader to fully understand the implementation of this system. This will provide future developers with adequate information for configuring old components of the system and give them an insider's view as to how exactly the system works. Each PHP class or file will have a unified source header with an overview of the purpose, environment context, global fields/methods/functions, and author of the class. Additionally any long blocks of code will be clarified with inline comments to describe processes within methods. See figure below:

----

crud.php [created: 8/2/07]

Class Definition

Purpose: CRUD work of the database.

Environment context: to be used with any pages requesting db access

Global Fields: db_user, db_pass, db_host

Methods: connect_to_db,

Functions: retrieve_data, update_data

Authors: Ray McGill

Edited by: Sam Phillips

5.1.3 User / Admin Manual

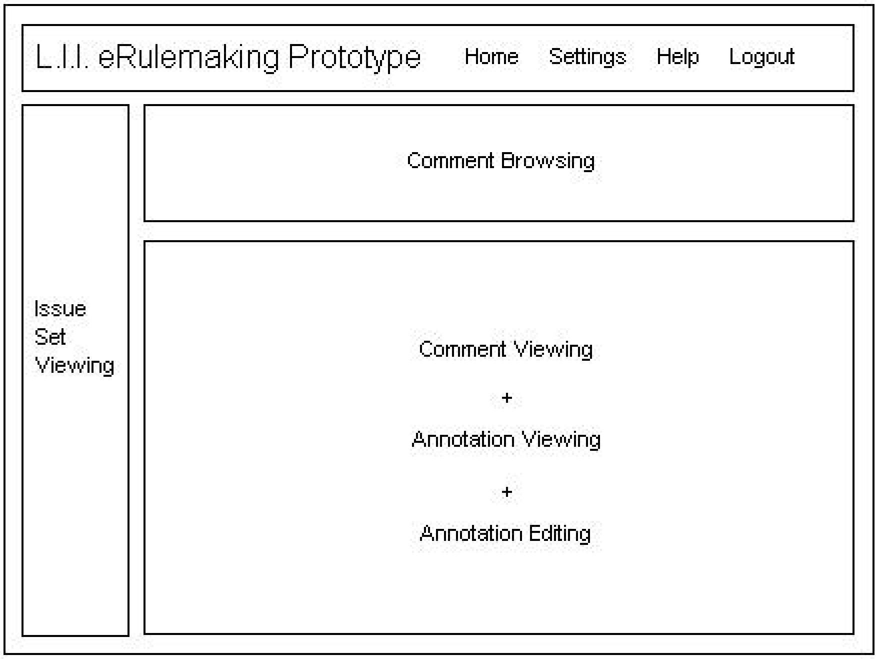
Manual-type documentation for feature explanation will be provided on the system in the form of a help page that is tailored depending on the user profile type, i.e. admin, user, etc. The documentation will outline all the features available for that particular user profile and exactly how to best utilize these features. The purpose of this specification is to simply serve as a manual for any users new to the system. This document/page can aid in explaining features in a more user-centric mindset to developers to help in the development process for the next possible phases of this framework-system. This document is outside of the scope of this project, although the framework will be designed to incorporate such a help system.

6 User Interface and Usability Requirements

6.1 User Interface

Our user interface is different for the three types of users that we allow, rule administrators, system administrators, and regular annotator users. For all types of users, there will be two important pages accessible. One of these is the main landing page that will include a welcome message, possible announcements, and links to all current rules (proposed regulations) open on the server. The second important page is the annotation page, which is accessed by clicking on the link to a particular rule on the main page. This page is the workspace where users of the system will spend most of their time and will provide the following features:

- Links: The ability to navigate to the help page, back to the main page, to the account settings page, or to logout.
- Comment Browsing: The ability to browse through all comments made regarding the particular rule chosen on the main page.
- Issue Set Viewing: The ability to view a list of all of the issues that apply to the rule selected on the main page.
- Comment Viewing: The ability to display a comment, selected by the Comment Browsing feature.
- Annotation Viewing: The ability to view the current annotations for the displayed comment. The annotations will be displayed in-line with the comment by highlighting the annotated text and displaying the particular issue that the annotation captures.
- Annotation Editing: The ability to add and remove annotations for the displayed comment. Adding will be accomplished by highlighting the text to annotate and clicking on the appropriate issue in the Issue Set Viewing component. Removing will be accomplished by selecting an annotation and clicking a "Delete annotation" button.



Above: A simplified mock-up of our user interface, as described in the preceding section.

Rule administrators will be allowed access to a special page that provides additional functionality, such as the ability to add and remove rules, issues and users. It will also allow them to modify the content of the main page, such as the welcome message and announcements. System administrators will be allowed the same access as rule administrators, but system administrators will additionally have full access to the Drupal configuration, which will allow them full flexibility in modifying and extending the content of the site.

7 First Iteration Report and Second Iteration Goals

7.1 First Iteration Report

As part of our first iteration we accomplished three main tasks, the first was obtaining information about the project from many stakeholders, the second was the implementation of a early prototype system, the last is significant visibility project, such as this report and our presentation to Dr. Arms and Mr. Bruce on 3/7/2007.

7.2 Information From Stakeholders

Early meetings with Tom Bruce were productive introducing the task and discussing possible sets of requirements, informal use cases and possible user interfaces. At a meeting attended by Dr. Wagner she suggested that early involvement of information she had about FTA workflows and discussions with student annotators may help clarify requirements and increase visibility to those stakeholders.

Our meeting the student annotators, Ms. Craig and Ms. Klimpel was very instructive. They gave good feedback about gripes and positive features with Callisto, additionally they gave early constructive feedback on many of the UI considerations which we had discussed in earlier weeks. We have taken their early feedback into consideration in designing the alpha version of the program.

We had the opportunity to attend a full CERI meeting on Friday, March 9th. There we learned about current progress of the NLP group in the classification problem and met many of the stakeholders for this project. Additionally we had the opportunity to speak with Mr. Simmons and Mr. Yoon, PhD students working on the NLP task. They will be important users of the system as well, and discussions with them clarified technical details regarding front / backend communication.

7.3 Prototype Implementation

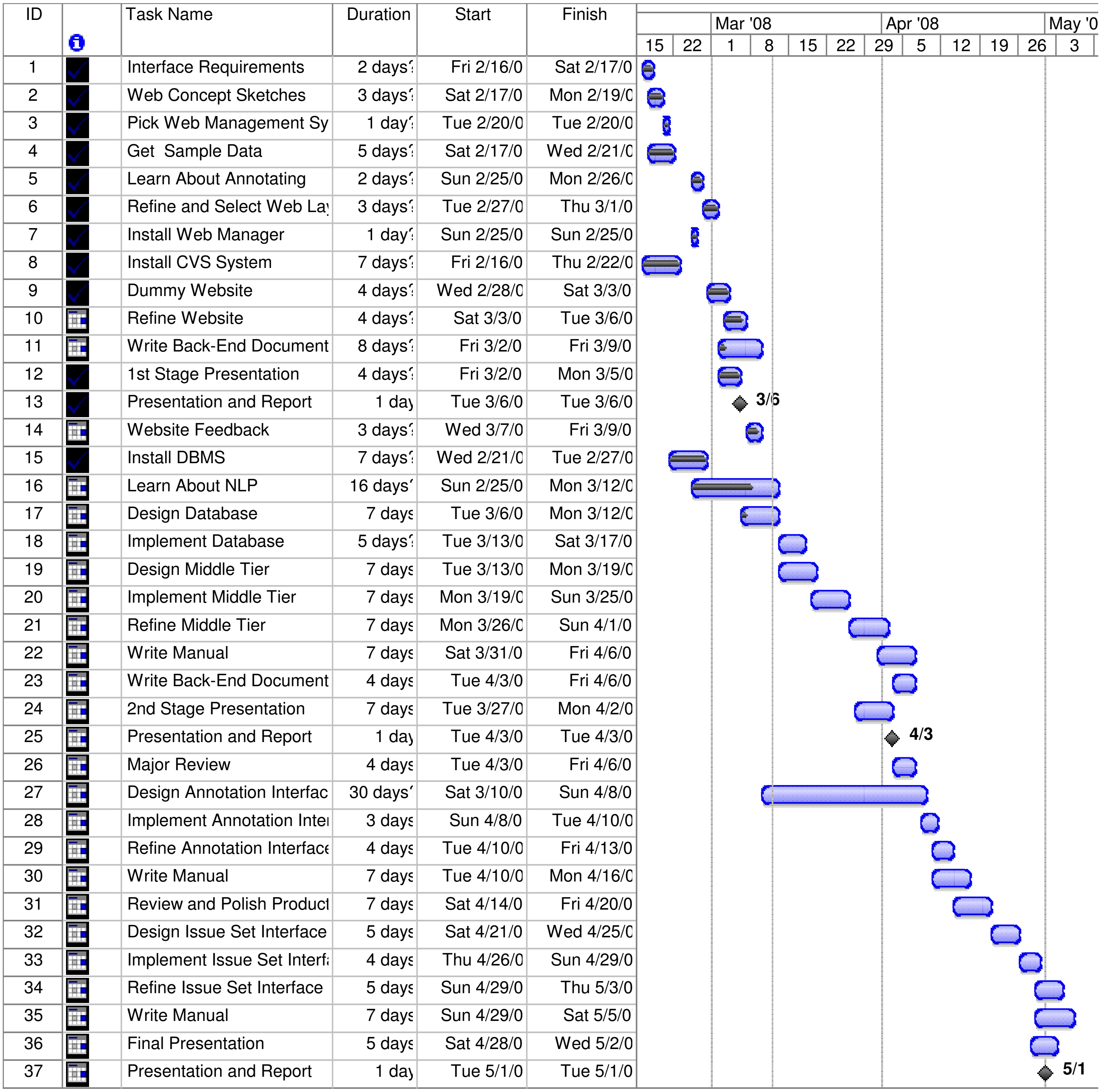
The first iteration of our software mostly lies in our Drupal installation, which we are still configuring and becoming familiar with. We had issues getting our Drupal site on the LII server when it was necessary, so we had a team member serve an initial Drupal setup on his personal Linux computer in order for the team to start experimenting and developing. We backed this installation with a mySQL database system that serves as the main database to store our comments and annotations. We recently obtained Drupal support on the LII server and are in the process of migrating our development over. We were able to experiment with JavaScript and PHP and achieve some menu function that will be used in further iterations. We are also researching and making progress on the technical problem of annotation by highlighting. This problem is tricky because while we want to achieve cross-browser compatibility, different browsers implement highlighting information functionality in very different ways.

7.4 Future Directions

In our next iteration, we expect to have a beta-level annotation highlighting system in place backed by our database. This will allow us to solicit feedback about the basic functionality of the application without having to develop a fully refined interface. The highlighting data will be collected by the client's browser JavaScript and sent to the server, where the data will persist in the database. The data will be sent back to the client whenever a new comment is loaded and the corresponding highlighting will be shown. We also hope to have a basic database schema designed for the hierarchy of rules and comments that we will need to organize. Another significant task in the next iteration will be converting annotation data to and from the ATLAS XML standard that is used currently by Callisto and the CERI NLP group.

7.5 Schedule

The Gaant chart shown below is the schedule which we have developed for this project. As is shown there we are generally on track with development, the anticipated time is shown in the blue bars, the black bar inside is the progress on that task so far. As is clear from the tasks, between now and phase III is a very important period where we will focus on the design the backend of the system. The last phase is iterative refinement of the complete system and creation of the various forms of documentation.



ID	ⓘ	Task Name	Duration	Start	Finish
1	✓	Interface Requirements	2 days?	Fri 2/16/0	Sat 2/17/0
2	✓	Web Concept Sketches	3 days?	Sat 2/17/0	Mon 2/19/0
3	✓	Pick Web Management Sy	1 day?	Tue 2/20/0	Tue 2/20/0
4	✓	Get Sample Data	5 days?	Sat 2/17/0	Wed 2/21/0
5	✓	Learn About Annotating	2 days?	Sun 2/25/0	Mon 2/26/0
6	✓	Refine and Select Web La	3 days?	Tue 2/27/0	Thu 3/1/0
7	✓	Install Web Manager	1 day?	Sun 2/25/0	Sun 2/25/0
8	✓	Install CVS System	7 days?	Fri 2/16/0	Thu 2/22/0
9	✓	Dummy Website	4 days?	Wed 2/28/0	Sat 3/3/0
10	▦	Refine Website	4 days?	Sat 3/3/0	Tue 3/6/0
11	▦	Write Back-End Document	8 days?	Fri 3/2/0	Fri 3/9/0
12	✓	1st Stage Presentation	4 days?	Fri 3/2/0	Mon 3/5/0
13	✓	Presentation and Report	1 day	Tue 3/6/0	Tue 3/6/0
14	▦	Website Feedback	3 days?	Wed 3/7/0	Fri 3/9/0
15	✓	Install DBMS	7 days?	Wed 2/21/0	Tue 2/27/0
16	▦	Learn About NLP	16 days'	Sun 2/25/0	Mon 3/12/0
17	▦	Design Database	7 days	Tue 3/6/0	Mon 3/12/0
18	▦	Implement Database	5 days?	Tue 3/13/0	Sat 3/17/0
19	▦	Design Middle Tier	7 days	Tue 3/13/0	Mon 3/19/0
20	▦	Implement Middle Tier	7 days	Mon 3/19/0	Sun 3/25/0
21	▦	Refine Middle Tier	7 days	Mon 3/26/0	Sun 4/1/0
22	▦	Write Manual	7 days	Sat 3/31/0	Fri 4/6/0
23	▦	Write Back-End Document	4 days	Tue 4/3/0	Fri 4/6/0
24	▦	2nd Stage Presentation	7 days	Tue 3/27/0	Mon 4/2/0
25	▦	Presentation and Report	1 day	Tue 4/3/0	Tue 4/3/0
26	▦	Major Review	4 days	Tue 4/3/0	Fri 4/6/0
27	▦	Design Annotation Interfac	30 days'	Sat 3/10/0	Sun 4/8/0
28	▦	Implement Annotation Inte	3 days	Sun 4/8/0	Tue 4/10/0
29	▦	Refine Annotation Interface	4 days	Tue 4/10/0	Fri 4/13/0
30	▦	Write Manual	7 days	Tue 4/10/0	Mon 4/16/0
31	▦	Review and Polish Product	7 days	Sat 4/14/0	Fri 4/20/0
32	▦	Design Issue Set Interface	5 days	Sat 4/21/0	Wed 4/25/0
33	▦	Implement Issue Set Interf	4 days	Thu 4/26/0	Sun 4/29/0
34	▦	Refine Issue Set Interface	5 days	Sun 4/29/0	Thu 5/3/0
35	▦	Write Manual	7 days	Sun 4/29/0	Sat 5/5/0
36	▦	Final Presentation	5 days	Sat 4/28/0	Wed 5/2/0
37	▦	Presentation and Report	1 day	Tue 5/1/0	Tue 5/1/0

8 References

[1] C. Cardie, C. Farina, & T. Bruce. Using Natural Language Processing to Improve eRulemaking. In Proceedings of 2006 International Conference on Digital Government Research, San Diego, 2006.