# Chapter I

Applicants to the Graduate School at Auburn University are currently served through the Graduate Web Application and Admission Process system, or GWAAP. This web-based system is out of date, difficult to use, lacking in aesthetic quality, and in short a poor ambassador for the graduate programs of Auburn. Furthermore, the interface for faculty use of GWAAP is equally poor. CSSE faculty members evaluating applicants do so primarily through a physical paper trail that is collated and passed around for each applicant. The GWAAP system for faculty is a thin GUI around the GWAAP applicant database, offering little to ease the several pain points of evaluation. There is a need for a more robust solution for both groups.

Applicants will benefit from a new system in several ways. By collecting and managing data in one place, there is no need for the multiple usernames, PINs, passwords, and reference numbers that currently exist in the GWAAP system. Having a single point of reference also helps applicants by streamlining the assembly of their application and providing clearer, more immediately feedback. Applying a modern layer of polish to an online system also helps present a more desirable front for the university, which is especially important considering the CSSE department in particular.

The benefits of a new system for faculty members are greater still. One of the primary benefits is the elimination of a great deal of physical documents being passed through the department. It is more difficult for all involved parties to manage multiple paper copies of transcripts, resumes, letters of intent, and standardized test scores when the same documents can be both uploaded and used in purely digital forms. A new system can also allow a digital forum for faculty discussion on applicants. By opening up applications to easy, digital commentary and voting, more eyes are available to evaluate applicants and they can do so faster. Increased automation is another benefit that can be leveraged in a new system. Much of the boilerplate communication between the CSSE department and potential students is still done by hand. By turning control of these important, but mindless, tasks over to the system, the rate of error and the delay in communication can both be decreased. Finally, a streamlined application process for applicants means less time spent by faculty members responding to confused emails by potential students.

From this problem, a set of use cases was derived that a theoretical replacement system should be able to fulfill. These use cases are included as Appendix A. Details on the specific use cases and how they were implemented in this system are included in Chapter 3.

# Chapter II

The requirements for a new GWAAP system necessitated a web-based solution. Because the system deals with applicants’ private data, that data must be able to be secured. Additionally, there is a need for a definitive version of the material that makes up each application. The easiest way to ensure that an authoritative version of the data exists is to have control over it. Furthermore, this data may need to be accessed simultaneously by multiple users. These things point to a RDBMS as the method of data storage for the system. Next, some method of interacting with the database must be available to users. Providing SQL access directly to the data is a poor choice for reasons too numerous to list here. A desktop application that communicates with the database is an option, but is undesirable for at least two reasons. One, it would require that users download and install a desktop client prior to submitting their application. Two, it would ignore an extremely robust infrastructure that already exists for client-server applications: specifically, web browsers and the Internet. Web browsers provide very well-defined ways of interacting with a server running an authoritative copy of information at a remote location. Furthermore, the ability to use HTML and CSS markup for the application greatly speeds development time. It effectively provides a huge, stable, well-supported library of GUI widgets for building the frontend of the application.

Although a web application is the clear choice, there are roughly three levels of granularity in implementation that can be targeted. The first is a raw programming language, like PHP. The second is a Content Management System, or CMS, like Wordpress. The third is a web application framework, like ASP.NET. This is a spectrum of course, and the distinctions between these levels can be fuzzy. It could be argued that one system could be moved up or down the spectrum given certain circumstances, but the general idea is that different implementation methods provide different basic levels of functionality. There are additional concerns beyond programming granularity as well. One is the need to interact with a legacy system, i.e. the existing GWAAP implementation in use at the Graduate School). A second concern is the need for a secure system, meaning resistant to unauthorized access (accidental or malicious). A third is stability and control over the final system, since it will serve not only as the primary repository for official documents, but also as a platform for higher-level activities—most notably, faculty collaboration and evaluation of applicants.

Of the three implementation levels, a raw programming language provides the most flexibility. PHP is offered as an example because it is historically targeted at programming for web applications, with an extensive standard library for web development as well as thorough documentation. Other common choices include Python, Java, and Ruby, although nearly any language can be used as most applications using the Internet backbone boil down to an ordered series of string inputs and outputs (at the application level, at least). The downsides of a system programmed in this way from the ground up are centered on the need to reinvent the wheel for a great deal of functionality that isn’t specific to the system itself. For instance, user authentication in web applications is a complex and problematic area that has a very high degree of importance. However, it is also a problem that is very similar no matter what system is being implemented. By virtue of being a stateless protocol, HTTP has no (good) way of setting up a user session if the identity of the user is needed for interaction with the system. For instance, applicants and faculty members need to interact with the system in different ways, but it is infeasible for the server to open, authenticate, and maintain a socket connection with every user that connects to the system. Instead, HTTP requests are usually handled by the principle of REpresentational State Transfer, or REST. In a RESTful system, well-defined HTTP verbs are sent as requests against URIs that the server exposes to all users. This means that every user, at all times, is technically making each request anonymously. If some content on the server needs to be hidden behind authentication, those authentication details must be sent along with every request so that the server can authenticate the requester and determine if they have valid access permission to the resource exposed through the URI. This presents a number of problems. First, different URIs may require different access credentials. For example, the URI that Google exposes for checking webmail probably requires a different set of credentials (ex. username and password) from the ones an online banking website requires. It is undesirable to send all available credentials for a user to each server that the client makes a request of. At the same time, users expect a session-based interface even with web applications. Just as entering a username and password once for a terminal session authenticates a user until they explicitly log out, so do users expect a single login to authenticate them to a web service until they are finished making requests. (The alternative would be requiring a username and password field on every request made to the server, which is both unsafe and obnoxious.)

To bridge the gap between user expectations of sessions with web applications and the server’s necessity for stateless transactions, cookies are used. A modern session cookie usually includes nothing but a hashed or random string either derived from login info or issued by the server upon first authentication of the user. This cookie is used as user credentials on subsequent requests rather than the username and password strings. In addition to improving security, the server can use information about the resources being requested along with the requests themselves to provide the illusion of statefulness to the client.

This comes full circle to the problem of implementing web applications in “pure” programming languages. Every bit of the above functionality must be implemented along with the rest of the system. Not only is this time-consuming, error-prone, and difficult, it is a problem that has been solved already. In the context of developing web applications, using a pure language is comparable to programming using assembly language. While there is certainly still a need for it in certain cases, the power of already-existing systems is such there would have to be an extreme need for such a raw level of control to make the headache worthwhile.

In contrast, Content Management Systems form the other end of this spectrum. The WordPress CMS (related to but not the same as the WordPress blogging service) is arguably the best-known CMS, and implements not only every bit of authentication functionality but much more as well. WordPress, along with other CMS offerings like Joomla and Drupal, abstract away as much of the implementation details as possible and attempt to leave users with fully-functional systems with the absolute minimum of knowledge or setup required. (How little knowledge? Many web hosts now offer “One-Click Wordpress Install” to users who want to get the system running using all default settings.) There are benefits to this approach. First, third-party developers often offer plugins or themes that can change the look or functionality of a CMS install with little mess for the end user. Second, CMS providers tend to be very good at the things they do best—it would be difficult to design a blogging service from scratch that would operate as well as WordPress does.

Unfortunately, while a CMS is extremely well-suited for non-technical users or users who desire common functionality, they can be poor choices for niche use. One of the drawbacks is their extreme overhead—by accounting for all possible use cases, the system can swell to a size unmanageable by a single developer. (For rough comparison: a basic Wordpress install is in the neighborhood of 125 MB, while the entire GWAAP system developed for this project is around 600 *kB*.) The size of the system on disk is hardly a concern, but the number of files to go bug-hunting through is very much a concern for a developer attempting to add or especially change existing functionality. The size and sprawling focus of a typical CMS is the biggest drawback to them for a project such as this. Although it may be very good at what it does, it is highly likely to be awful at things it doesn’t do, if it can be coerced to do them at all.

The middle ground is the use of a web framework, like Django, Ruby on Rails, or the Microsoft ASP.NET stack. A web framework takes a more balanced (though not necessarily better) approach to built-in functionality. Frameworks will typically solve many of the foundational problems in web application development (like user authentication) but do little more than that. Ruby on Rails is a popular framework because of its “convention over configuration” approach, in which the development model is generally taking the base use case of the Ruby on Rails system and defining any ways it should behave *differently* from its standard configuration. The ASP.NET framework is also popular, as are frameworks like CakePHP and Catalyst (a Perl-based framework). Django is a Python-based web application framework. Like any other approach, it has benefits and drawbacks. It is typically clear syntactically and an experienced Django developer can prototype rapidly with it. It includes a very mature testing suite that builds on top of Python’s already-robust unit testing framework. It does a fine job of providing basic functionality in its custom settings, is well integrated with Eclipse, and provides good command-line tools for interacting with an application under development. Best of all is its ability to abstract (but not hide) details of database implementation. Database tables are defined in Python using inheritance from some base classes provided by Django, and an external tool is used to generate and then run appropriate SQL calls to the database. It is quite possible to develop an entire database-backed application without ever making a manual SQL query.

Django has at least two major drawbacks, however. First, the learning curve for Django is significant. It is a very large, very powerful, well-established system under active development. It is not as simple as, say, learning a new library for Python. Rather, it is more closely comparable to learning an entirely new language based on Python syntax. Second, Django is very rigid in its approach to certain aspects of web application architecture. While its Model-View-Controller separation is very clear and easy to understand, there are fairly extreme levels of complexity within each layer and Django is very strict about the way it wants things done, and the order in which it wants them done in. This is especially problematic when designing a system using software engineering standpoint, which is further documented in Chapter 3.

Ultimately, a system built in Django is able to satisfy all non-functional requirements of the proposed system. It can be backed by a RDBMS, it can be written to interoperate with a legacy system, it puts a high level of control in the hands of the CSSE department, and it can offer thoroughly-tested security.

An additional non-functional requirement unique to this project was to apply software engineering principles to the construction of a web application. To help achieve this goal, construction of the system followed PCSE, a personal software process under development at Auburn University. PCSE is a full-lifecycle process that addressed every section of the system, although some practical modifications were made during the course of construction. The PCSE process, its application to this system, and modifications made to it are discussed in Chapters 3 and 5.

# Chapter III

## Analysis

The first stage of this project was describing the needs that a new system would have to address. PCSE refers to this part of the lifecycle as “Analysis”, and for this project it took the form of a set of use cases derived by Dr. David Umphress (see Appendix A). The use-case document included a diagram of actors and how they might interact with various components of the system and each other, as well as 14 use cases describing specific functionality desired in the final product. Each use case included information like objective, entry and exit criteria, basic and alternate paths, and notes or diagrams expanding or explaining the desired functionality. The use cases document was purposely non-specific regarding implementation details, describing a black-box approach of inputs and outputs that a user would expect from the system.

This document was supplemented by occasional stakeholder meetings to refine the ideas presented in it, including a lengthy initial meeting to fully describe the problems facing the existing GWAAP system and how the use cases were to address those problems. Taken as a whole, this information represented all necessary requirements to complete PCSE’s information-gathering Analysis phase. Because the use cases matched existing pain points, and because they specified some parts of the system’s complete functionality, the Analysis phase more closely matched a real-world scenario in which a solution to a known problem has to be constructed to match loose stakeholder requirements. This is in contrast to a typical (undergrad) academic scenario where a toy system is created to exacting specifications, or a system is created to address a theoretical problem. The flexibility of PCSE was a benefit here, as the process could accommodate this type of Analysis phase as opposed to a system where actual components or function points had already been specified.

## Architecture

The next phase of construction was a process of mapping the system requirements to an overall architecture as well as individual components.

From a design pattern standpoint, Django effectively mandates a Model-View-Controller architecture. Because this pattern is exceptionally well-suited for web applications, it makes sense to base the architecture of the system around it. Django deviates from the standard terminology for this pattern, however. Django Models are equivalent to the M of MVC, although Django does allow models to define their own methods for convenience or clarity. Django introduces confusing use of the term “View”, however. In Django, a View is a Python method that contains the logic for processing requests and rendering HTTP responses to the client. This is analogous to the Controller in standard MVC parlance. The display layer, normally known as the “View” of MVC, is referred to as the Template layer by Django. In short, Model-View-Controller architecture is better described as “Model-Template-View” for the sake of Django. The concepts in each layer map more or less as expected, so it is primarily just a difference in terminology. Because of the well-defined nature of MVC architecture, a good deal of the process was mapping use cases to their necessary components among the three layers. Models were derived from logical needs in the use cases. For instance, the use case specifying the need for commenting on applications led to the Comment model containing a foreign key for the application it points to, an additional foreign key for the user who made the comment, and a character field for the contents of the comment. The foreign key to application is required (since it doesn’t make sense to leave comments in the database for applicants who don’t exist or have been deleted), but the foreign key to the faculty member was made optional from a database perspective to account for admin-level comments not associated with a faculty member, as well as to prevent deletion of users from breaking the comments on an application. During this phase, all known additional models were derived in this way. The need for a few extra models was determined during construction and is outlined below.

However, the details of Django implementation caused certain architectural issues to arise after the fact. During this phase, one of the most difficult problems of applying standard software engineering process to a Django application reared its head. As part of Architecture, the inheritance hierarchy for users was developed. Django offers a User base class outfitted with authentication functionality. The GWAAP system, however, has two major classes of user – faculty members using the system to evaluate applicants, and the applicants themselves. The boilerplate authentication needs for both (permissions, sessions, etc.) are similar, but each class of user has access to an entirely different set of use cases and expect a different path through the system. From a purely architectural standpoint, a natural choice would be to have children that inherit from Django’s User base class. One, representing a GWAAP User (i.e. faculty member, administrative staff, GPO, etc.) needs high levels of access, while the other, representing each Applicant, needs low-level access tailored to their personal application. This fell out into two components to be built in the first iteration.

The problem in this situation stemmed from the fact that, as mentioned before, Django has very strong notions of how things “should” be done. Due to a lack of implementation experience with Django, it was unclear during the Architecture phase that subclassing Django User presents a number of problems. (Due to namespace conflicts in the prose description of these classes, the base Django user class will be referred to as DjangoUser, and the GWAAP user class will be referred to as GwaapUser.) First, all classes that inherit from DjangoUser are accessible by querying the DjangoUser object manager. Although this can be accounted for in custom views and test cases, the Django framework itself does not differentiate between GwaapUser and Applicant when pulling them from the database. The primary issue here is that the authentication/session middleware of Django passes Views an object containing the currently logged-in user as a field. No matter what the actual class of that user is, Django can only recognize it as a DjangoUser. This forces awkward casting within each view to gain access to specific attributes or convenience methods of the models.

This is the second major problem of subclassing from DjangoUser: because all users are the same at the framework level, it is possible to give certain classes permissions they would not otherwise be entitled to. Since these permissions are the *only* wayof preventing access to unauthorized user types at the framework level, abnormal user types with non-default permissions can progress to Views they would normally be rejected from. Within Views that attempt to cast users to classes they don’t actually inherit from, database errors will be thrown. This is not in and of itself a security hole, because non-default permissions can only be set by system administrators (who have the power to cause other, more serious security vulnerabilities). Nonetheless, it is troublesome because it is conceptually confusing and because it can result in bizzare HTTP 404 and 500 errors. Notably, superusers automatically get all permissions by default meaning to prevent these unexpected behaviors they need to deactivate their own “is\_gwaap\_user” and “is\_gwaap\_applicant” permissions.

Permissions were ultimately used as the solution to this architectural problem—within Views that need to access specific fields or methods of users, requests are first filtered by permission (redirecting unauthenticated or unauthorized users to the appropriate login page), then pulling the appropriate class of user from the database with manual calls to their object managers. For normal operations, this means that the system *can* be designed to match the architectural expectations of child classes off of DjangoUser. One of the benefits of this is that Applicant models can have database fields or related rows in other tables that are exclusive to them. GwaapUser models do not need entries in the table of Application models or Profile models, and this way those tables are not filled with rows upon rows of null values.

Attaching extra information to Applicant models is accomplished through a hybrid of two Django-specific techniques: UserProfiles and signals. UserProfiles are the canonical Django way of adding extra information to DjangoUser objects. In recent versions of Django, support for true subclassing and explicit addition of extra fields is improving, but documentation on these features is less robust and at times overshadowed by old but popular workarounds for previous versions of Django. Signals are messages sent by the framework upon certain events in the lifecycle of an HTTP request. Methods can register themselves to receive signals that interest them, take whatever action they need to respond to the message, then return control to the Django dispatcher.

In this system, UserProfiles are used to add additional information to Applicant models only. The Django framework provides a get\_profile() method to all DjangoUser objects which returns an instance of the custom profile class defined in the settings file and hooked to a custom model. Arbitrary extra information for User models can be attached to their UserProfile and retrieved by Views during the processing of HTTP requests.

Signals are used at creation time for these models. Django offers (among other types) a “post\_save” signal, in which a specified method receives notification any time a model is saved to the database. Methods that wish to receive this signal will get at least three parameters: a class name expressing what kind of object is being saved, a Python object representing the instance of the model object, and a boolean value expressing whether this object already exists in the database or is being added for the first time. This signal framework is used to handle appropriate instantiation for Applicant objects. One particular method listens for any time a new Applicant object is being added to the database. After being added (and thus after it has been issued a primary key) the method instantiates a new user profile for the applicant and saves it to the database as well (thus ensuring that *it* now has a primary key). An Application and GwaapProfile object are then created as well with foreign keys pointing to the user profile. This process (which is admittedly one of the more convoluted Django idioms) makes possible at least two things. First, it means that GwaapUser objects do not have Applications associated with them, which would not only take up space in the database but also makes no sense. Second, it guarantees that if an Applicant exists in the database, it also has a matching Application object to hold information associated with it. This is very useful because it means that Views using Applicant models do not have to manually check each time for whether the Applicant actually has an Application – they can (correctly) assume that Applicants always have Applications and make calls on it with that expectation.

Two additional questions arose during Architecture. The first was related to a single applicant having multiple applications. The Applicant model logically matches the human who is applying – the model contains information on username and password along with email address, none of which *must* change even if the person applies a second or third time. (This is in contrast to the existing GWAAP system, where applicants must begin the process completely from scratch including a new account.) One option would be to allow applicants to attach an arbitrary number of accounts via foreign key relationships. This would handle the problem of applicants who make more than one application. It would also allow old information to be neatly reused in a new application. For example, an applicant wishing to reuse an old Reference could simply change the foreign key in the existing Reference to point at the new Application object. The problem with this setup is not an engineering one, but a user interface one. For most users (who will only apply once to the department), it adds an additional layer of difficulty in understanding how the parts relate to one another for no benefit. Applicants must set up their applications separately from their account, which for most will be a meaningless or even confusing differentiation. By setting up a one-to-one relationship between Applicant and Application, there does not have to be a differentiation for the most common case. In the rare cases where an applicant does want to reuse their account information with a new Application, this can be achieved manually by an administrator. For this system, it was determined that this one-to-one relationship was the most sensible approach.

A second additional concern was the way Reference models would be handled. One of the major use cases of the system specifies that external users are able to interact with the system and register their comments and rankings as a reference for a given applicant. References are given an online form that is meant to take the place of a paper recommendation. Not only does this eliminate a large amount of additional paperwork for all parties, but it standardizes to some degree the applicant qualities addressed by references. By offering a defined ranking system and categories for references to fill out, faculty evaluators can not only compare apples to apples but also have data on qualities that are important for acceptance decisions. One of the biggest questions in this system was how references would authenticate themselves to the system. One solution would be to have them create accounts with the system just like applicants. A benefit of this approach is that Views handling references could be written in the same ways as those handling applicants and faculty users; i.e. checking for appropriate permissions and having built-in access to the User representation of the person accessing the system. A second benefit is the additional level of security and confidence for both references and evaluators. By ensuring that all persons filling out recommendations are authenticated to the system, involved parties have an additional measure of confidence that references are who they say they are, and that contact information can be retrieved if it is suspected that they aren’t.

Here again the problem lies in the user experience with the recommendation process. Asking users who will probably never again interact with the system to create an account with it is non-trivial overhead both for the database and for the people serving as references. Since all applicants are required to have three references, that category of users would quickly swell to much larger than faculty users and applicants combined, yet most of them would never need or use their account with the system again. Second, the whole online reference process is intended to both lessen the burden of recommendation on those third parties as well as increase the value of the recommendations to faculty members evaluating applicants. Adding an account creation and login overhead to the process runs the risk of irritating those users which might jeopardize the validity of their data. It was determined that for this system, references should not need to log in or create accounts with the system to use it. The implementation details of that were purposely deferred to a later iteration.

## Iterative Design and Construction

Having completed Analysis and Architecture, PCSE becomes an iterative process in which the Planning, Construction, and Refactoring steps are completed in cycles, each cycle moving closer to a complete system. PCSE does not mandate the order in which components are built. Therefore, there can be a number of approaches to dividing the implementation into iterations. One approach is to build a minimum viable product in the first iteration, with each subsequent iteration adding additional functionality, polish, or other value. Another approach is to take one or more functions of the system at a time, implement them to production quality, then repeat. Yet another approach is to build components in order of their projected difficulty from hardest to easiest or vice-versa. For this project, it was determined that the implementation of Model components would be the focus of the first several iterations, followed by View components, and finally by Template components, with one or more iterations for polish near the end after the basic functionality of the system was complete. There were several motivations for this. First, the model components were foreseen to be the easiest to implement. Although this turned out to be incorrect, the thought was that practice on the easier components would be desirable as a method of becoming familiar with the complex Django framework. Second, the nature of the framework is such that the Models must be in place for the Views to have anything meaningful to interact with. Third, Models and Views form the core of the system, at least from a business logic standpoint. No viable product can exist apart from them so it seemed backwards to focus on writing a Template layer for a product that did not exist. Finally, as the Template layer represents mostly just the user interface/experience, and this project focused on software engineering principles, a functional system and test suite was deemed of more value than a slick interface, at least early on.

### Iterations 1 through 4

Iteration 1 tackled the core authentication issues of GwaapUser and Applicant, how to subclass them from DjangoUser, and how to attach Applications to Applicants. In the course of implementing these classes, the problems described above of mapping architecture to implementation arose. Due to external circumstances, a long break was taken from implementation, and upon return to the system Iteration 1 was arbitrarily ended and Iteration 2 began with mostly the same goals. The Django User Profile system was discovered in the course of Iteration 2, which led in turn to the discovery and implementation of signals (specifically the post\_save signal) and the architectural revisions to account for the idiosyncrasies causes by subclassing DjangoUser. Test-Driven Development was *invaluable* at this stage. Because Model objects were the first to be constructed, there was no other method of validating their correctness besides test cases generated as part of TDD.

Unfortunately, the practice of TDD for this system (especially for early iterations) bore little resemblance to normal experience with TDD. The process of writing a failing test case was fairly straightforward after becoming familiar with the Django-specific abilities like new asserts, but the time between red-light test cases and green-light test cases was on the order of hours instead of minutes. As repeatedly mentioned, Django has very specific ways it needs things done, and very specific orders it wants them done in. While the Django community site offers a well-written tutorial and extensive documentation, it is still a *very large* framework. It was extremely difficult to take the “magic” methods and procedures from the tutorials and generalize them to application for this project. Failing that, it was necessary to learn how things worked before they could be made to work. This involved assimilating tremendous amounts of information to accomplish tasks that looked simple in tutorials.

The power of the Django Model layer, and its abstracting of the database, actually made this process more difficult. With the benefit of hindsight, it is clear that Django Models are very direct analogues to pure SQL—they can be best thought of as Python syntax to make SQL calls. Unfortunately, that very important fact was not clear from just the tutorials and high-level documentation. This resulted in attempts to mesh knowledge of what should work in Python with reality of what wasn’t working in Django. Additionally, Django’s highly-specific ideas of what order things must be done were troublesome here as well. Notably, definitions of primary and foreign keys must be done not only in logical order for the sake of the database, but also at specific moments in the Python object lifecycle, which is compounded by the fact that relevant details of that lifecycle are “conveniently” abstracted by the Django object managers. Once a conceptual design is formed about exactly what the framework is doing, it is much easier to deal with. Forming that mental model for the first time is a major hurdle, however.

This lengthy process of trial and error lasted throughout iterations 1, 2, and 3, as well as most of 4. One of the side effects of such warped TDD is that time technically spent in construction and labeled as such was really sandbox time. Although the work was done in production and test code, a great deal of it was figuring out how to use the framework effectively. This skewed construction times too high at first. A task recorded as taking one hour might have only taken an experienced Django developer ten minutes or less. An additional side effect was the practical reality of doing code experimentation within the production code. Ideally, sandbox work would be done in a completely separate file. The problem is recreated in the sandbox, a solution is developed, and context switches back to production code and the sandbox is deleted. In what quickly became a large system, however, it because increasingly difficult to recreate production problems from scratch after every context switch to the sandbox. In response to this, a more formal approach to the sandbox was developed, which will be described later in the report.

The end of Iteration 2 saw the creation of the first basic Views for the system, namely the login and logout views. By differentiating users at a framework level via Django Permissions, custom Views could make use of decorators provided by Django that allow filtering via permissions. Valid users are allowed to continue in the View, while invalid (unauthenticated or unauthorized) users are redirected to a login page. Basic use of these decorators Iteration 2 was followed by more advanced use in Iteration 3, which fleshed out additional Views including an early Reference model. At this point in construction, Views dealt with hand-made HttpResponse objects containing simple strings to validate that the views were functional.

Iteration 3 also saw the setup of the built-in Django admin system. This collection of modules, which is an optional part of the framework (referred to as one of many packages of “middleware”), offers a web interface to the database, allowing a user to interact with Models through a GUI instead of solely on the command line. The Django admin system is flexible and powerful but also requires a non-trivial amount of setup to make it appropriate for systems with complex behavior (like this one), so it did not entirely supplant command-line tools. Making it usable (and useful) required an additional module in the GWAAP system as well as overriding a few key magic methods, notably the \_\_unicode\_\_() method for several of the Models that needed to be accessed through the admin interface.

Iteration 4 was focused on the Comment and Vote models as well as the views to interact with them, plus two additional views for GwaapUsers to get a list of all applicants in the system as well as details on individual applicants.

At this point in construction, the system had most of its major functionality available from the command line or admin interface. Most Models were defined and working properly, and basic raw-text Views were available for several critical system functions. The functionality of the system was well-tested (around 100 test cases at that point) and separate paths for GwaapUsers and Applicants existed. Following iterations would add the Template layer and focus on several aspects of the user experience.

### Iterations 5 and following

Iteration 5 was the first iteration to include the development of components for the Template layer. Using previous experience as a guide, effort estimation was handled differently from the previous iterations. 15 components were identified for construction in iteration, which seems like an overly large number except for the fact that the Django template layer includes an inheritance API. Base templates can be defined once and child templates can simply inherit from one or more base templates and only override their specific area of functionality. This was accounted for in iteration planning by putting a large percentage of the expected effort on the first (base) template, then assuming that child templates would take less time. This approach yielded fairly accurate effort estimation, especially considering that this was the first iteration for an entirely new type of component.

The architecture for the template layer makes use of three levels of inheritance hierarchy. The first level consists solely of the base template. All other templates are expected to include this template in their hierarchy, so it was a natural place to put things common to every page in the web application. This includes the expected HTML skeleton common to any page, as well as the linking of stylesheets and JavaScript. To speed development, a collection of CSS and JS boilerplate called Bootstrap was used. Bootstrap defines several CSS and JS classes that elements of the DOM can be marked up with to enhance their presentation (and in some cases their functionality). It is a popular and actively-maintained package with good compatibility among its own classes and between different browsers. Because the development of a new stylesheet is not only time-consuming and error-prone but also outside of the direct scope of the project, it was determined that the use of Bootstrap was preferable to an entirely hand-made stylesheet. The project does include a set of custom styles to extend and modify some aspects of Bootstrap, but these are just that – extensions, not a standalone style.

The Bootstrap stylesheet and JavaScript files are linked to by the base template, giving all child templates direct access to their markup. The base template also outlines the basic structure of the HTML, including <head>, <body>, message, and comment blocks (the purpose of the latter two is described later). By handling this all on the base template, most of the non-page-specific markup only had to be written and debugged once. After this component was declared stable, all child templates could assume that their inherited markup was stable, leaving those components to focus only on their own unique functionality.

One of the major ways this technique proved valuable was in the use of the Django message middleware. Views are able to call methods of an automatically-provided Message object to add short messages intended to be displayed to the user in response to important actions or conditions. Messages passed in this way are added to a queue of messages available to the template layer the next time a template is rendered. The message queue can then be iterated over (which also clears it) and the messages can be output to the final HTTP response. To make this process as painless as possible while coding, some additional message-handling code was added to the base template. This block of code is run every time the base template is rendered (which is every time that *any* template is rendered in the system). Inside, the template iterates over the message queue, detects what type of message has been passed (“success”, “info”, or “error”), creates an HTML <div> element with appropriate coloring, and renders each such div in a standard format and location on the screen along with a small “X” to dismiss the message. This technique, called “toast” windows in many systems, is a convenient way to keep the user apprised of their progress through the system and the results of interacting with various parts of it. More importantly, it provides a way to give user feedback without having to design extra, explicit views for simple messages.

For example, consider the case of an applicant uploading a document like their letter of intent. There are at least two exit states: success, wherein the document was successfully uploaded and now resides on the server, and failure, wherein *something* went wrong and the document has not been uploaded. Furthermore, there are several ways to exit with failure. The user could have attempted to upload a document in an unsupported file format, could have submitted a form with no document attached at all, or could have attempted to upload a document to an application flagged by an administrator. In any case (success or failure), the user needs some way of knowing what happened. Without feedback, they simply get redirected to some page with no certainly of whether their action had the intended effect or not. One way many systems handle this is a simple success or failure page with a message and a link to the next reasonable page (the account page on success, the upload page on failure, etc.). This is reasonably effective, but has at least two drawbacks: it involves an extra click for the user, and in the specific case of this framework, requires extra View and Template code to be written for each and every failure or success case throughout the entire web application. This would mean an explosion in test cases and View methods even if a single template was developed using variables to populate its contents. The toast window system was a natural answer to this, and tying the toast messages to the base template means they are always available to any view at any time.

In short, the base template gives a consistent look to the application, as well as consistent availability of CSS classes and template blocks for structuring page content. However, a second level of the template hierarchy was devised to differentiate between the faculty portion of the application and the application portion. Even with a consistent UI, not all parts need exactly the same functionality. For instance, clicking on the “Home” button in the navigation bar should take faculty users to a different page from applicants. Similarly, the contents of the navbar should be different for the two classes of user. This behavior could have been implemented on the lowest level of templates, but it would have required a great deal of code duplication which would be more susceptible to mistakes as well as more painful to change if site-wide changes were required.

With all of the above functionality in place, the third and lowest level of the template hierarchy is occupied by the child templates defining page-specific UI elements. This includes everything else from form fields to the display of applicant details. At this level, the Django Template layer is powerful, but no so powerful that application logic can be put in it. This is one of its strengths. The main control structures in the Template layer are an if/else block and a for block, along with a brace notation that tells the template parser where to look for and insert variables. Variables are passed to the template layer as a Python dictionary. Whenever the template parser encounters a variable, it checks to see if the dictionary contains that string as a key, and populates the variable block with those contents if so.

The parser is capable of following limited dot-notation relationships as well. This is used in the templates for this system when an object (usually representing some Model passed from the view) is placed directly in the dictionary, as is typically the case with collections that need to be iterated over by the template. For example, in the “display all applicants” View that faculty evaluators have access to, a collection of all applicants currently active in the system is placed in the dictionary as a QuerySet (which is essentially a Python list wrapped in some extra Django functionality). The template iterates over each entry in the QuerySet to get one applicant at a time, then uses dot-notation lookups to populate specific holes in the template for each row in the table.

The other notable bit of advanced template usage for this project is the idea of template includes. These are somewhat like template inheritance but from a different perspective and usually with a smaller scope. When include blocks are encountered during template parsing, the parser loads them from the filesystem and parses them in the same context as the calling template, optionally with a reduced or renamed set of variables. Whereas inheritance is used when the framing aspects of a page need to be the same, includes are used when small parts need to be consistent within pages or sections that are generally different from one another. In this project, includes were used to display the different fields of recommendations in a consistent manner. Recommendations for applicants include five different fields asking the recommender to rank the applicant. Although the names for each field are different, the ranking levels are the same for all fields, and the display for them should be consistent. To achieve this, the five fields were placed in a table in the page template, with holes in the template for the contents of the fields. Populating each hole required a 6-level if-statement to account for the various possible responses that could be pulled from the database. Because that logic was consistent among all five of the fields (and verbose), it was extracted and placed into an include. For each row in the HTML table, the relevant variable was passed to the included code which in turn filled the hole with the appropriate response. The benefits of this were that the logic only had to be defined once (and therefore could be changed in a single place if necessary), and that the page template was not cluttered by five large, identical if-else blocks.

The most significant component pair of Iteration 5 was the View and Template for the “View Application” use case. This use case represents applicants’ ability to check on the completion of the various parts of the application document and to take action to complete it as necessary. The template for this page consists of an HTML table with 7 rows in addition to its header. Four of the rows are direct representations of the status of their database fields. “Complete” fields yield a green cell with a check mark, “Not Applicable” fields yield a grey background with the text “N/A”, and all other database contents yield a red cell with an “X” icon. If the database field is set to “Contact Administrator”, this page will also yield an error message with that text. The bottom two rows of the HTML table represent the status of the applicant’s resume and letter of intent uploads. These cells are considered complete if a valid document has been uploaded, and incomplete if not.

The other row represents the status of Recommendations. Because an applicant must have three references fill out these recommendations, the conditions for determining this status are more complex. If the applicant has not yet issued three recommendations, the status cell is filled with the incomplete contents. In addition, a third “Actions” column will have one button that yields details on why the cell is incomplete, and a second button that allows applicants to specify another reference. Once three references have been specified, the status column changes to a special “pending” cell with an orange background and the icon of a clock. The details button yields information to the effect that while the applicant can take no further actions, the cell cannot be marked complete until all references have filled out their recommendations. The applicant cannot specify additional references at this point, although in the event that a reference refuses to fill out a recommendation, an administrator can delete one of the incomplete references thus allowing the applicant to specify another. Once all three references have completed their form, the cell changes to green and the row is marked complete.

This page in particular necessitated more complex interaction between the View and Template layer than for most other pages. Specifically, dynamically determining and populating not only the colored status cells but the action buttons required a significant amount of HTML markup in the Python code for the Views, which then had to be placed correctly in the final HTTP response so that the DOM could render the contents properly. Two techniques were developed to help cope with these problems. First, the *highly* error prone generation of dynamic HTML elements was placed in a factory method accessible by the View. The factory method takes the custom strings needed to generate and differentiate certain elements, then plugs it in to a long string amounting to a complete HTML element. This allowed for a single place to assemble and test the HTML code (made more problematic by having to constantly open and close three different Python quote types) and then allowing one well-tested factory to generate all subsequent HTML elements of that type. Furthermore, the return value of the factory method could be stored as a Python variable, making it much easier to deal with in code than a large, multi-line string literal. The second technique was developed to handle the arbitrary JavaScript that was necessary for some of the components to operate properly. Because the JS on the page needs to be rendered in an area reserved for the base template, and because the exact content of that JS could not be abstracted for all pages, a special block was created on the base template. The block, titled “extra\_onready\_js” could be filled by any view that needs a place to put its own specific JS code. In the base template, this block was then placed inside a <script> tag, and specifically inside a pre-defined method guaranteed to be run on each page load. If this block is empty (as it is for most pages), it does not slow down the page because it is effectively a blank line inside a minimal function. On the other hand, pages that need to define their own JavaScript can do so straight from Python. This technique was used on the View Application page to enable the mouseover behavior for the Detail buttons, among other places.

Iteration 6 added some additional polish to the system while tackling two new design problems. The first of these was the implementation of the complete recommendation/reference system. As mentioned, it was determined that this system should not require users to make account with the system. Nonetheless, it is necessary that users still be authenticated (in the non-Django sense) and that the system offers an effective user experience. For instance, references need to see the name of the person they are recommending on the page to prevent confusion especially in the event that they are asked to serve as a reference for more than one applicant through the GWAAP system. At the same time, there needs to be a high degree of certainty that users cannot somehow find their own recommendations in the system and fill them out. It is also desirable that users cannot find out who else is applying and who their references are by searching through URIs in a systematic manner. Exposing URIs comprised of the primary keys of applicants and the primary key of the Reference objects in the database are acceptable if there is some sort of password protection or other validation between users and the display of the recommendation form, but not acceptable if the reference form is immediately available upon visiting the correct link.

The solution used in this system is to create a unique identifier string for each Reference object in the database at the time it is first saved. Next, a base URI is exposed by the Django dispatcher to catch attempts to access Reference objects. A regular expression extracts the unique identifier string, pulls the matching Reference object from the database, then populates a form template and displays it to the recommender. This method achieves all goals necessary for the system. First, it allows recommenders to interact securely with the system without creating an account with it. Second, if the key space of the unique identifier string is large enough, it prevents unauthorized users from accessing Reference objects that they shouldn’t.

Using a hashing method to generate the unique identifiers is one possibility, but the chance exists (however small) that two different Reference objects may hash to the same value, thus creating a conflict when the system tries to pull exactly one matching object from the database and receives two. To prevent such collisions, new additions to the table of Reference objects would need to check to make sure that no other Reference objects existing in the database already share that unique identifier. This means the process of inserting a new Reference object will have to include a linear search using this method. Since that is the case, a simpler method of generating a random string of ten upper- and lower-case alphanumeric characters is used, checking to make sure that string does not already exist in the database, and then using it as the unique identifier. This is still a linear search but has the benefit of increasing the key space for the same number of characters while not having to depend on a hashing algorithm.

Finally, this unique identifier is sent in the email to references already included in a link to access the reference forms. This means that potential references have only to click a single link in their email and get direct, secure access to the recommendation form. Because the key space for this string is 10^62 (ten characters randomly selected from all numbers and upper and lowercase letters), the chance of brute force access to even a single random reference in the database is vanishingly small; the chance of brute forcing access to a *specific* reference, smaller still.

The second, albeit smaller, design concern in this iteration was the final page of the applicant area of the system – the “Home” page, or Applicant Status page. This primary feature of this page is a graphic indicating how far along the applicant is in the application process. The applicant will always be at one of six linear milestones, which are held in the database as small integers. The human-readable milestones these integers represent are declared as a tuple of values in Python, each value being itself a tuple pairing the integer value with a human-readable explanation of the milestone. Using this method, the Django admin interface is able to populate itself with the string values for convenience, while the database only has to hold the numeric representations. This fact is used on the Applicant Home page to determine the size of the progress bar graphic that represents which milestone the applicant is at. Instead of pulling the string from the tuple, the progress bar pulls the integer value, multiples it by a percentage, and then uses the resulting value directly in the CSS definition of the progress bar width. Additionally, the template includes six holes to indicate whether the corresponding help text should be bolded or not. Since the template parser renders nonexistent variables as blanks, the variable named for the integer value in the database includes the markup to make the text bold, while the other variables render nothing and therefore do not have their corresponding text bolded.

## Sandboxing complex system functionality

Although it is not part of the linear flow of any iteration, PCSE describes an additional implementation block called the “sandbox.” In this block of time, context is switched a blank, unconnected module in order to test a new or poorly understood technique without compromising the integrity of production code. Sandbox time can occur at any point in any iteration, although it usually occurs during construction to address problems encountered while coding. When the source of confusion is a syntax error or language issue, the blank-slate approach of the sandbox is highly desirable. However, there were many times during the implementation of this system where problems arose precisely from the implementation of the system itself, or from the interaction of various parts of the Django framework. In these cases, a typical sandbox approach was infeasible after about halfway through Iteration 2—basically, once the system had grown too large to accommodate reconstructing relevant parts for the sake of the sandbox.

To address this concern, a source control solution was devised. Whenever there was a need to break the code out into a sandbox, the system was briefly refactored into a stable state. This state was committed to a Git repository. Next, a completely separate instance of Eclipse was opened (to prevent namespace conflicts that arose from having two Eclipse projects with the same name in a workspace). Finally, the stable version of production code was imported into the sandbox instance of Eclipse, and sandboxing could be performed on the full system without violating production code. Lessons learned could be immediately applied back to the production code and another commit could be made, then the sandbox could be entirely wiped out without loss.

# Chapter IV

## Validation

To determine that the system solved the problem as outlined, two major methods were used.

### Deriving system components from use cases

First, the system was compared at every stage to the use cases. Each proposed component was derived (either directly or indirectly) by linguistic analysis on the use cases. Given the line from the “Submit Comment” use case, “The faculty member submits a comment,” for instance, three components could be derived—Faculty Member and Comment as Models, and Submit Comment as a View/Template. Other components were derived from non-functional requirements. For example, the base template component was derived from the non-functional requirement that the interface be unified among the several pages of the application.

In addition to deriving components from the use cases, a second pass through the use cases was made to determine that every action described in the use cases could be accomplished through at least one portion of the system. This was a formalization of deciding what use cases were available through the built-in Django admin functionality, and which were only possible through custom Views. For example, the use cases specify that both applicants and administrators must be able to update applicant profile information. Because administrators have access to the Django admin interface, this use case did not require a custom solution. However, because applicants do *not* have access to any portion of the admin interface, custom View and Template components were derived to satisfy this requirement.

One use case described visually the relationships between actors and their access to the system. To use the example above, both Administrator and Applicant types have access to the “Update Applicant Profile” use case and this is shown by a line connecting their respective stick figures to a circle containing the name of the use case. Validating that these requirements were met was a matter of ensuring that each user type could be created from some mix of permissions granted through the admin interface or by signals in the system at object creation time. For example, Faculty Members can be generated by creating a base User object and adding Comment permission, while Committee Members are generated by doing the same plus adding a Vote permission. Administrator-level users can be created by adding DjangoUser objects to the system then having a superuser give them some additional system-specific permissions.

The final method of validating the system was by comparing expected outputs expressed in the use cases to the actual outputs of the system. In some cases, this was as simple as validating that the text displayed in the use cases matched text displayed on screen. For use cases including sample graphics though, this was a less-rigid process of evaluating the visual output of the system and ensuring that it matched (at least on a conceptual level) the sample graphics from the use cases.

### Regular stakeholder meetings

The second method of system validation was by regular meetings with primary stakeholder for the system. This allowed for quick feedback loops early in the design, as well as requests for design changes as necessary. As the implementation of the system matured, the tight feedback loop and ability to combine or modify use cases as appropriate became increasingly more valuable. After approximately Iteration 4, meetings also included system demonstrations to ensure that the system was progressing according to stakeholder expectations (both functional and non-functional).

## Verification

By virtue of constructing using Test-Driven Design, a full suite of test cases was developed for the system. Approximately 180 test cases were devised to test all three layers as thoroughly as possible.

Model testing was fairly straightforward. Because Django abstracts the database tables and rows into Python objects, these objects were directly available for test case asserts. The intermediary of the framework meant that standard Python idioms could be applied to data that was ultimately backed by a database. For example, test cases had access to Python built-in methods like len() for checking the number of rows returned for particular queries, etc. As the Model layer is primarily a data definition layer in Django, most of the model-related test cases checked for the existence of appropriate fields, correct foreign key relationships, correct data storage, and similar issues. It is impossible to overstate the value of the Django framework in converting database contents to testable Python objects.

View testing was slightly more complex, since ultimately the View layer takes HttpRequest objects as inputs and returns HttpResponse objects as outputs – essentially, glorified string values. Again, the Django framework provided several facilities to ease this process, though. First, HTTP header fields and status codes are accessible through dot-notation on the relevant Django objects. This eliminates a great deal of string parsing to get at key information. For instance, some test cases were designed to check for the existence of expected status codes (i.e. a test case purposely calling for a View that doesn’t exist should get a 404 status code, etc.). Django offers these status codes as a status\_code attribute off of the returned HttpResponse object. Even more important is Django’s provision of a Client object to interact with views as if testing against a (very abstracted) web browser. Using the Client object, test cases are able to make requests of the system as if they are querying a server. This is useful for ensuring that proper URIs exist for the relevant Views and that the Django dispatcher is serving them as expected. Additionally, it allows for both GET and POST requests on Views, the latter of which taking a Python dictionary as variables for the POST request (for instance, form contents for testing that the Views handle form data correctly). The Client object also allows for simulated login, which was invaluable for testing that the correct types of users could access the correct areas, or more properly that the wrong types of users could not. Finally, the Django testing suite offers a few extra asserts (compared to the standard PyUnit asserts) including the assertContains() method. This is a relatively simple method that searches through the text content of an HttpResponse object for a specified string. Though it is conceptually simple, it is powerful in practice, especially in conjunction with the testing method devised for the template layer.

## Unit testing for the template layer

One of the methods used to validate the template layer was the comments block. In the base template, an additional block called “comments” was included inside an HTML comment on the template. Because the Django parser ignores HTML comments (Django templates have their own notation for comments), the parser can populate an HTML comment just like any other block in the template. In this way, the actual display of the page to the user can be decoupled from the need for certain test strings to appear on the page. For instance, one test case tested for the presence of the string “View Applicants” on the ViewApplicants template. This way the test case could validate that the correct template is loading. Then, that string was inserted inside the comments block on the appropriate page to ensure that it always appears in the template, even if the display of the page is changed such that “View Applicants” no longer appears as such to the user. Furthermore, through the use of Django template variables within the comments block, arbitrary comments can be added at runtime by the View layer itself. In another template, an “extra\_comments” tag was added to look for a string by that key in the dictionary. The View can then add comments as necessary to verify that the template is correctly pulling data from models, etc. For example, a test case was devised to test that the template layer correctly populated an HTML table with the proper number of rows pulled from the database. In the view, a simple counter was incremented each time another model was placed into a row in the table. Then, the counter was inserted into the extra\_comments block along with an identifying string. Finally, the test case can check for the presence of the correct counter value after the identifying string, thus testing that the table is populated with the correct data.

Obviously, the technique is unsuitable for sensitive data that should never appear in the template layer at all, because it is still visible to users who view the source of the page. Still, it works well for digesting information that is already available on the page, but that would be extremely difficult to extract by simply analyzing the text content.

## Note on system functionality tested externally

Four aspects of the system were assumed to be tested elsewhere. First is the Django framework itself. In fact, Django comes with about 300 test cases included that can be run as a standalone test suite or as part of an application’s individual test suite. All test cases pass on the development system. Second is the CSS and JavaScript components that make up Bootstrap. Because Bootstrap is actively in use and regularly updated, this component was assumed to work properly given correct HTML markup. Third is the web browser on which the application will be run. Because the system takes advantage of no browser-specific functionality (limiting its markup to a subset of cross-browser compatible HTML), it is assumed that the system will work in all major web browsers. (Note that testing *was* done to ensure that HTML tags were syntactically correct, that anchors pointed to the correct links, etc.). The fourth and final externally tested component is the database. It is entirely beyond the scope of this project to validate that the RDBMS backend is working properly, but it is also very safe to assume that it is.

# Chapter V

## Lessons learned in Django

The process of learning Django while developing this system made it somewhat unique compared to implementing it in a pure subset of Python. Thanks to the exceptional documentation and community of the framework, most of the system exhibits best-practices for major functionality. However, the User authentication system could be improved. There is no single standard way in Django 1.3 (the version used for this project) to handle additional information tied to users, especially subclasses of User. The most idiomatic way according to the Django community site is the UserProfile functionality, but this is designed specifically for applications that directly instantiate DjangoUser to represent users on the site. Because of the creation of the two child classes, several other workaround had to be devised to handle the idiosyncrasies that arose. No single workaround for these problems exists or even seems to be favored (for Django 1.3 at least). There was a fair amount of trial and error involved, along with implementing the system around existing design decisions for the User classes. If there is one area of the system that could be improved by starting over, this would be it.

One of the biggest early obstacles was interacting with the database. Django does an exceptional job of handling database interactions for the developer – so exceptional, in fact, that it is easy to make poor design decisions by applying object-oriented reasoning to what is in reality a database schema. The takeaway from this is to enjoy the database convenience that the framework offers, but to always remember that it is in fact a database just under the surface, and to think in terms of SQL, not object orientation, when defining models.

Testing FileFields for models was very challenging to handle in a non-trivial way. For this project, files are tested by actually creating dummy files and passing them as POST data to Views during testing, then asserting that the presence of the file exists in the correct place after the View runs. The problem with this method is that it actually creates files on the filesystem. In practice, this meant that once those test cases were determined to run correctly, they were commented out until the file system could be cleaned up. Every time they are run they require manual cleanup afterwards. Some sort of file mockup package would have been helpful for this testing. At least one such package appears to exist for Python, although time did not permit its use in this project.

Finally, it was interesting (although not necessarily Django-specific) to have to interact with a legacy system during this project. For instance, Applicant Profiles have a field called “ENTER\_QTR”, referring to the “quarter” they begin their studies. Auburn has not been on a quarter system for some time now, but the name persists in the old GWAAP database. To ensure that the new system is interoperable, the field in the new system shares the same (outdated) name. It was interesting to have to make a known “bad” decision in the new system in order to coexist with the roots in the new system.

## Applying PCSE to web development

One process problem arising from the use of the Django framework is the issue of how to track time spent in the sandbox versus time tracked as construction. Typically, any time spent writing test code or production code would be counted as part of the construction phase. However, a significant amount of time in this project was spent trying to perform simple tasks in production code using Django idioms. As described earlier, this could cause extremely long periods between failing and passing test cases.

Since construction *technically* consisted of writing production and test code during that time, the choice was made during iteration 1 to document this time as construction time. Within the scope of this project, that decision worked for estimating effort in subsequent iterations. However, a problem would arise if data from this project were used to estimate effort in a second Django project. Because PCSE construction data tracking was not granular enough to distinguish “learning-Django-construction” from “normal-TDD-construction”, using this data again would produce skewed effort estimation since the learning curve for Django would be lessened. This would likely hold true for any project heavily based on a framework that was being used for the first time (or even a project in a new language). For this reason, if the project were started again today, time spent learning Django (even time spent technically writing production code or tests) would be recorded as sandbox time. That would enable accurate time estimation for subsequent iterations by combining sandbox and construction time. In a subsequent project time effort estimation could be based only on construction time.

Testing of the template layer could also be expanded for the development of other web applications. Although the content of the template layer can be validated through PyUnit using the comment block technique described above, there are also in-browser testing frameworks designed to pick up where testing in this system leaves off. Specifically, these are used for testing actual DOM elements and even CSS markup. For a higher confidence level in the user experience with the templates, one of these tools could be used.

## Future work

At least three expansions to the current system could be made to improve functionality. First, the current file server is intended for debugging only and should be replaced with a server (or at least a View) with permissions tailored more properly to its use. In its current configuration, files uploaded to the system (including all resumes and letter of intent) are publicly available if their correct URI is retrieved. Second, some additional verbosity in detail buttons and error messages would be welcome. Although the system does its best to cover the most common errors, it is still possible to give it such malformed input that it will throw an “ugly” error, like an HTTP 404 or 500 response. Most of these errors could be eliminated through the power of the Django dispatcher if desired. Third, an automated method for the creation of applicants through the GWAAP dump files is desirable. Ideally, the system could expose a URI that accepts POST data consisting of a CSV file from the Graduate School and automatically creates new Applicant objects in the system and emails their new owners with setup details. It would also be easy to create a basic command-line tool for this in Python.

Finally, a major expansion of the system would be to modularize or abstract its CSSE-specific details so that it could be used by other departments. The Django framework already makes use of a settings file where deployment-specific details of the application can be stored. With some major refactoring, the system could be rewritten to take advantage of similar functionality and be expanded such that migrating it to a new department would require only a custom settings file. This would mean the main aspects of the system could be held on a single authoritative server, enabling all departments using the system to automatically benefit from updates to the system while still maintaining their department-specific behavior.

# Appendix A

[original use cases]

# Appendix B

[test case report]

# Appendix C

[Screenshots comparing old GWAAP to new GWAAP]

# Appendix D

[PCSE process documents (scans and digital copies)]