DEVELOPING DIGITAL FIELD GUIDES FOR PLANTS: A STUDY FROM THE PERSPECTIVE OF USERS

A Thesis

Presented to

the Faculty of California Polytechnic State University

San Luis Obispo

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Computer Science

by

Emily R Schwarz

June 2011
### COMMITTEE MEMBERSHIP

| TITLE: | Developing Digital Field Guides for Plants: A Study from the Perspective of Users |
| AUTHOR: | Emily R Schwarz |
| DATE SUBMITTED: | June 2011 |
| COMMITTEE CHAIR: | Franz Kurfess, Ph.D. |
| COMMITTEE MEMBER: | Clark Turner, Ph.D. |
| COMMITTEE MEMBER: | David Janzen, Ph.D. |
Abstract

Developing Digital Field Guides for Plants: A Study from the Perspective of Users

Emily R Schwarz

A field guide is a tool to identify an object of natural history. Field guides cover a wide range of topics from plants to fungi, birds to mammals, and shells to minerals. Traditionally, field guides are books, usually small enough to be carried outdoors. They enjoy wide popularity in modern life; almost every American home and library owns at least one field guide, and the same is also true for other areas of the world.

At this time, companies, non-profits, and universities are developing computer technologies to replace printed field guides for identifying plants. This thesis examines the state of the art in field guides for plants. First, a framework is established for evaluating both printed and digital field guides. Second, four print and three digital field guides are evaluated against the criteria. Third, a novel digital field guide is presented and evaluated.
Acknowledgements

I would like to thank those whose professional and personal dedication made this work possible.

First of all, I would like to thank Dr. Franz Kurfess, my advisor, for his limitless guidance and patience. Without his advice and support I would have never had the courage to take this unconventional topic and run with it. I would like to thank Dr. Clark Turner for making me a better writer and a better thinker. His classes were my most memorable experiences at Cal Poly. I would like to thank Dr. David Janzen for his forward thinking in establishing classes on the Android platform. Not only did my own work benefit greatly from taking his class, I know countless other students have benefited as well. I would like to thank Dr. Clinton Staley for making me a better programmer. I would like to thank Dr. David Keil for the great experience I had in his Field Botany class. It was that class that most inspired the technology herein.

I would also like to thank my friends and family who supported me during this long and difficult process. I would like especially thank my boyfriend, Ricky Martinez, for putting up with me for the past few months. His cooking and humor made the process of writing this thesis possible. I would like to thank my sister, Virginia Schwarz, not only for her extensive editing, but also for teaching me where to put my commas! I would like to thank my parents, my mother for bestowing upon me my love of plants, and my father for bestowing upon me my love of learning. Finally, I would like to thank my grandmother, Hazel Boggs, without her support and encouragement I might have never gone to college at all.
Contents

List of Tables ix
List of Figures x

1 Introduction 1
  1.1 Scope of this Work 1
  1.2 Overview of Field Guides 2
  1.3 Users 5

2 Evaluation Framework 7
  2.1 Ease of Reaching Identification 8
  2.2 Accuracy 9
  2.3 Corroborating Information 10
  2.4 Identification Method 11
    2.4.1 Dichotomous Keys 12
    2.4.2 Single Keys 12
    2.4.3 Multi-access Keys 13
    2.4.4 Automated 14
    2.4.5 Possible Values 17
  2.5 Area of Focus 18
    2.5.1 Possible Values 19
  2.6 Geographic Area 19
    2.6.1 Possible Values 20
  2.7 Number of Species 20
    2.7.1 Possible Values 21
2.8 Percent of Total Species .................................................. 21
  2.8.1 Possible Values ...................................................... 21
2.9 Portability ................................................................. 22
  2.9.1 Possible Values ...................................................... 23
2.10 Ease to Acquire ........................................................... 23
  2.10.1 Possible Values ...................................................... 24
2.11 Extensibility ............................................................... 24
  2.11.1 Possible Values ...................................................... 25

3 Evaluation of Existing Systems ............................................. 26
  3.1 Jepson Manual ............................................................ 26
    3.1.1 Evaluation .......................................................... 28
  3.2 San Diego County Native Plants ....................................... 28
    3.2.1 Evaluation .......................................................... 29
  3.3 Wildflowers of San Luis Obispo, California ......................... 30
    3.3.1 Evaluation .......................................................... 30
  3.4 National Audubon Society Field Guide to North American Trees:
    Western Region .......................................................... 31
    3.4.1 Evaluation .......................................................... 32
  3.5 Discover Life Website: Wildflowers Guide ......................... 33
    3.5.1 Evaluation .......................................................... 34
  3.6 Audubon Wildflowers: A Field Guide to North American Wild-
    flowers iPhone Application ............................................ 35
    3.6.1 Evaluation .......................................................... 35
  3.7 Leafsnap iPhone Application .......................................... 36
    3.7.1 Evaluation .......................................................... 37
  3.8 Summary ................................................................. 37

4 Implementation ............................................................... 39
  4.1 Motivation ............................................................... 39
  4.2 Architecture ............................................................ 41
    4.2.1 Plant Identification Wiki ........................................ 42
    4.2.2 Wiki-to-DB ......................................................... 49
4.2.3 Field Guide Service ........................................... 51
4.2.4 SLO Digital Field Guide ..................................... 52
4.3 Evaluation ......................................................... 57
4.3.1 Evaluation ..................................................... 57

5 Usability Analysis .................................................. 59
5.1 Preliminary Evaluation .......................................... 59
5.2 Usability Study Set-up ........................................... 61
  5.2.1 Plant Collection ............................................. 61
  5.2.2 Usability Study ............................................... 62
5.3 Potential Sources of Bias ....................................... 63
5.4 Results .......................................................... 64
  5.4.1 Identification Trial ......................................... 64
  5.4.2 Usability Survey ............................................. 67

6 Conclusion .......................................................... 69
  6.1 Future Work .................................................... 69

Bibliography ........................................................ 71
A Glossary .......................................................... 75
B Usability Survey ................................................... 76
List of Tables

3.1 Summary of Evaluations ........................................ 38
5.1 Identification Trial Data ..................................... 65
5.2 Identification Trial Data Times ................................. 66
5.3 Identification Trial Data Number Correct ..................... 66
5.4 Identification Survey Data .................................... 68
List of Figures

4.1 System Component Dependencies on Open Source Technologies . 40
4.2 Propagation of identification information through the system . . 42
4.3 Plant Identification Wiki class diagram . . . . . . . . . . . . . . . 48
4.4 Calculating the URL of an Image in the wiki . . . . . . . . . . . . 51
4.5 Overview of SLO Digital Field Guide screens . . . . . . . . . . . 53
4.6 Screens used for Intelligent Search . . . . . . . . . . . . . . . . . 55
Chapter 1

Introduction

Traditionally defined, a field guide is a book with the primary purpose of assisting a reader in identifying of an object of natural history. Field guides play an important role in educating the public about the natural world that surrounds them. They are also used by researchers in a variety of disciplines [33]. However, printed field guides are necessarily limited in the amount of information they can contain as well as the way that information can be structured. In light of recent advances in portable computing, academic institutions and environmental groups [1, 17, 3] are working to translate field guides into a digital form.

1.1 Scope of this Work

This thesis will examine field guides for the identification of vascular plants (Tracheophytes), and the term plant will henceforth be used to refer only to the vascular plants. Non-vascular plants such as mosses and liverworts will not be included, neither will fungi. Since most field guides draw their boundaries along the line between vascular and non-vascular plants, this division is convenient and
appropriate [11]. References to field guides with topics other than plants may be included where these discussions are relevant to plant guides. Despite this narrowed focus, the analysis presented in this thesis may be relevant to a variety of field guide topics, since field guides share many of the same uses regardless of topic [33].

This thesis will present a detailed analysis of the current state of field guides relating to plants and the directions of new technology being applied to this field. It will present a detailed criteria for judging existing field guides. It will use the criteria to evaluate seven existing field guides. It will also present and evaluate a new field guide, SLO Digital Field Guide, and its supporting technologies. I will present a usability study comparing SLO Digital Field Guide to other guides. Finally, it will conclude with a summary and discussion of future work.

1.2 Overview of Field Guides

The first popular field guide was published in the late nineteenth century [26], and they have since become standard in almost every American home [32]. Field guides are an important resource for novices and experts alike for identification of unknown specimens as well as a reference for taxonomic information, description of species characteristics, and distribution information [33]. According to the International Field Guides website [31], 29% of field guides with a single subject identify plants of one kind or another. Wildflower field guides, in particular, are the second most popular type of field guide, trailing only birds.

Field guides have essentially four purposes: inventory, identification, description, and classification [18]. **Inventory** is knowing taxa are present in a given area. This is important information for users who wish to study particular taxa
as restricted to certain areas. Understanding what occurs where is a necessary precursor for more advanced research [34]. **Identification** is determining the names of the taxa we find or study. **Description** is presenting us with knowledge of taxa. **Classification** is understanding how the taxa are related to each other and those outside the region. While classification can be viewed as a subset of description, it is more helpful to see them as presenting two different types of information. Description is the information relating to the properties of the described entity, its characters. Classification presents information about taxonomic relationships. Classification is important because it gives users a way to relate a new organism into an existing framework. This helps identification by enabling users to only remember how an organism differs from other closely related organisms, rather than memorizing all its characters. Classification information can also be used to infer missing information such as if the plant is likely to be poisonous, its possible pollinators, or economic value.

Identification of specimens in the wild is key to collecting distribution information that is essential for conservation [18]. This is true not only in the tropics [28] but in the rest of the world as well. Distribution information for many plants is incomplete and ever changing, even in areas that are assumed by outsiders to be thoroughly explored. For instance, many of the endangered plants listed in the California Native Plant Society’s Inventory of Rare and Endangered Plants would benefit from better distribution data, even though many of the rare plants it contains reside near some of the largest population centers in North America [38]. Better identification methods could make the collection of this data simpler and more reliable.

Despite the importance of species identification to conservation efforts, the identification of an unknown specimen is currently time consuming and difficult
It also requires specialized knowledge and skills that only highly trained experts possess. In the developing world, this difficulty in identification can hinder progress to discover new species before they become extinct. In the developed world, better identification techniques could improve the cost effectiveness of field inventories and on site data collection, both of which are important tools in effective land management and species conservation [36]. Additionally, quality identification methods can increase public enthusiasm and support for conservation and enjoyment of our natural world [15].

In recognition of the importance of identification in the taxonomic process, governments are examining ways to encourage the use of new technologies in plant taxonomy [5, 1]. These investments have made new types of digital field guides available, such as the LeapSnap smartphone app, supported by a National Science Foundation grant [8].

Academic institutions that do botanical research also see the benefit of supporting new technology both from a scientific and educational standpoint. Herbaria and museums have overlapping interests with academic institutions. Additionally, herbaria see added value in systems that provide a framework for the digitalization of their often huge collections [1]. These efforts complement and encourage the digitalization of field guides.

Field guides are also poised to benefit from the increased use of amateurs collecting scientific data. Called “Citizen Science” this practice teams scientists looking to collect data with volunteers who can help collect data [7]. While this practice has been around for at least a century in the birding community, it is gaining popularity in all sorts of ecological studies. Public monitoring of plant populations could reduce the costs of state or federally funded plant inventories, increase public awareness of invasive species [39], and provide a much
larger dataset than conventional methods [7]. Quality field guides are especially important for educating and training volunteers [34].

1.3 Users

The consumers of field guides are a diverse group. The *Jepson Manual*, for instance, lists possible users as “beginning students, self-taught amateurs, environmental consultants, native-plant gardeners, employees of diverse government agencies, ecologists, and academic systematics [11].” While far from complete, this gives a rough picture of the diverging motivations and interests of potential users of field guides. Regardless of their differences, all users can be roughly divided into two groups, experts and novices.

In this thesis the term **Expert** will describe users with the following characteristics. They are skilled in botanical terminology and traditional identification methods [17, 1]. They have a high degree of visual expertise [17]. Visual experts can see an organism as a whole and have a deep understanding of its structure. They are able to provide systems with information based on their expertise. For example, while they may not know the species or genus of a specimen, they will likely know the family. Systems can then use the experts’ knowledge to narrow viable identifications [1].

**Novices** on the other hand lack knowledge of botanical terminology and are unfamiliar with traditional identification methods, such as the dichotomous keys. Because novices do not usually work within plant taxonomy, they have little incentive to learn the difficult and arcane botanical terminology [17]. They have some visual knowledge, but it is incomplete [1]. Novices might be botanical hobbyists, or they could be researchers in fields related to botany that do not
require the training in botanical terminology or practice. It should also be noted that experts in one particular area of taxonomy may have the characteristics of a novice if they are working outside their area of expertise. In some areas of taxonomy, novices may out number experts [34, 26]. This makes reaching novices a potential improvement in identification technology because the best identification systems are currently reserved for experts.
Chapter 2

Evaluation Framework

This chapter presents a novel framework for comparing existing field guides of all formats. After the description, each characteristic has an explanation of why it was or was not evaluated in the following chapter. The characteristics Ease of Reaching Identification, Accuracy, and Corroborating Information are not evaluated in the evaluations but they are revisited in the Usability Analysis chapter.

The rest of the characteristics are evaluated in the next chapter and include a definition of different ratings field guides can receive. For some characteristics, this rating is a number or percentage. All of these have been arranged in a way that larger numbers represent a fuller implementation of the characteristic. For other characteristics, the possible ratings are given as a list of distinct sets of characteristics identified by labels, called states. For these characteristics, the states are ordered such that fuller or more preferable implementations appear later in the list.
2.1 Ease of Reaching Identification

The most important characteristic of a field guide is the ease in which users reach a correct identification. This can be measured by time [1] or by number of steps [17] to complete the identification. Users of field guides will likely identify many specimens over the course of a single day in the field, so any reduction in time spent per identification will be compounded many times. However, neither time nor number of steps is adequate by itself to fully represent the toll on users. An identification process that takes a longer time could be preferable if, for example, the system requires less user input but requires more time to process a result. Likewise, a process that takes more steps may be preferable to one that takes less if the one with fewer steps requires more arduous reasoning on the part of the users than the one with more steps. These trade-offs are best studied more extensively in the context of user interface research.

Visual identification is a powerful method of identification, and according to research by Kirchoff et al. [17], it is the easiest and most natural form of identifications. Because even novices have some visual expertise, keys that use drawing and photographs to aid in identification will appeal to the largest audience. While it is true that photographs are the cornerstone of most field guides, it is also true that some additional structure is necessary to reduce the time users spend browsing photographs for a look-alike. It is difficult to quantify at what point adding more photographs adds to the ease of identification and at what point it begins to reduce it.

In most field guides, the ease of identification is inversely correlated with the total number of species in the system. In paper field guides, fewer species means fewer photos to look through. If a guide is appropriately targeted to a local
area, it can greatly reduce the number of species it needs to include, and thus increase the ease which users reach identifications. Field guides for larger areas will have difficulty reducing their species count as doing so means the users are more likely to find a specimen that cannot be identified. For guides that use keys that present users with questions to reach an identification, fewer species means fewer questions need to be asked.

Because of the difficulty of differentiating solid analysis of this character from preferences of the evaluator, this characteristic will not be rated explicitly within the evaluations.

### 2.2 Accuracy

Accuracy is a measurement of the average number of correct versus incorrect identifications users make while using a particular field guide. This can be determined through user trials [1], or if the system already has a formal mathematical model based on accuracy, it can be simply extracted from the system [17].

An important landmark will come when digital field guides can demonstrate that users make more accurate identifications with digital guides than with the best traditional paper guides such as the *Jepson Manual*. This is a difficult measure to demonstrate though as systems with fewer plants have an innate advantage. Additionally, varying levels of user sophistication may make such comparisons difficult since different levels of user experience could drastically change users’ abilities over different systems. User experience with a single system may make a difference as well. As users become experts in more sophisticated systems their accuracy will likely grow over time. Another source of variability in accuracy measurements is likely the plants that are used in tests. Certain plants
may favor certain systems of identification.

Accuracy is difficult to determine in a single study on a single group of users. A complete evaluation of accuracy must include the accuracy of a system over all its possible users, at all skill levels, familiarity levels, and use cases. Additionally, there is no standard set of plants to measure accuracy with. Because of the variability in measuring accuracy, and its difficulty in being inferred directly, it will not be evaluated within the framework. However, the Usability Analysis presents a limited measure of accuracy within the context of the usability study.

2.3 Corroborating Information

After making a preliminary identification, users are often encouraged to review a number of details about the species to confirm it is really the one that best matches their specimen. This information could be a list of defining characteristics of the species, descriptions of its habitat, locations where it is found, or other facts about the species. This feature appears prominently in many guides and is often displayed as a selling point.

Corroborating information is also where the other purposes of field guides, Inventory, Description, and Classification are presented. Field guides are important repositories of knowledge for researchers and the Inventory and Description data are often cited in scientific works [33].

Corroborating information is clearly important to the identification process and the value of the guide as a whole. However, it is unclear exactly what information is most helpful to different groups of users and how much of the information should be displayed at a given time. After an identification, displaying too much
information about a species might make it hard for users to find the information most relevant to the identification. Because of the difficulty of determining what corroborating information is most useful, this characteristic is not evaluated within the framework.

2.4 Identification Method

Field guides differ most prominently in the way in which users make their identification. Identification is defined as assisting users in obtaining the identity of a given unknown plant specimen to at least the species level and, if available, the subspecies or variety. This section will define both traditional and new methods for identification. Dichotomous keys and single keys are the traditional methods of identification for published field guides. Introduction of computer-based technologies has made two new types of identification a possibility: multi-access keys and automated identification.

Even for computer-aided identification, users will be expected to interact with the system. Because a human user will be interacting with the system, it is not necessary for the system to make the full identification itself; it could present users with suggestions rather than a single determination [1]. This is appropriate since in many systems, users will have much more information than the system as to the nature of a specimen. This is also reasonable as both experts and novices have some degree of visual knowledge. If they are presented with visual information, they are likely to be able to choose the correct identification out of a set of possible species.
2.4.1 Dichotomous Keys

A dichotomous key presents users with a series of mutually exclusive pairs of statements, essentially a binary tree. Users determine the correct statement and proceeds to the next statement along the correct branch. Rarely both statements may be true; in this case the key will have two leaves that arrive at the same taxon [11]. Dichotomous keys are usually aimed at experts, and their proper use is regularly taught in botany classes at Cal Poly.

Users of dichotomous keys are completely at the mercy of a highly linear process. In most cases if they come to a statement pair for which they are unsure, they reduce their probability of finding the right answer by 50%. Rarely though, when the answer to a particular question is particularly ambiguous the key author will include a “trap door” on the wrong side of the tree so to correctly identify a specimen given that wrong answer. This, however, can drastically increase the size of the key so that it is used only in rare cases.

2.4.2 Single Keys

The most popular method of identification in paper field guides is the single key. Users of this type of key examine a single obvious portion of the plant for a small number of easy-to-recognize characters, such as flower color and flower shape [35]. The information within the guide is indexed by these features. This kind of identification is problematic because it is difficult to include large numbers of species. If users find a yellow, radially symmetrical flower, then they must scan through all of the yellow, radially symmetrical flowers to find the one they are looking for.
Single character field guides are often filled with photographs or drawings to make searching easier. However, they can not present a very large number of species because of the space needed for images. Moreover, as the number of images increases, so does the time it takes users to find a particular image. Traditional guides make a trade off in covering large geographical areas at the expense of including less frequent species [35] or in covering small geographical areas more thoroughly [16]. This makes single character field guides a tool for those users interested in either identifying only the most common species or those willing to make approximate identifications to the nearest “similar” species.

This approximate identification may not be adequate for scientific or governmental processes because it is not uncommon for an endangered species to share habitat with a nearly identical common species. One such case would be Layia jonseii protected under the Endangered Species Act and Layia platyglossa its common cousin, both of which are found growing together in serpentine outcroppings [38, 11].

2.4.3 Multi-access Keys

Multi-access keys can can mitigate some of the difficulties experienced with single keys by allowing users to select which characteristic they would like to answer questions about [9]. Some of the first efforts to improve electronic identification were along this line [2]. Initial efforts used expert systems to provide the users a more flexible way of answering questions. Multi-access keys do not enforce a specific ordering of the questions they present, so unlike dichotomous keys, it’s much harder to get “stuck” and be unable to complete an identification.

Multi-access keys may also allow users to answer questions incorrectly and still
reach a correct identification. This sort of tolerance, however, is not necessarily present in all multi-access keys.

Multi-access keys retain some of the problems of the previous keys, for instance, they grow larger and harder to use the more plants are added to the system.

One potential improvement computer technologies could make to this type of guide is the ability to reduce the number of questions that need to be asked by taking into account previous answers. Discover life accomplishes this by filtering out those questions which all the remaining candidates fall under a single answer [27]. This approach is simple but effective. However, there seems to be room for improvement by adding additional methods to select which questions would have the most impact on the remaining search.

2.4.4 Automated

There are two possible technologies that would revolutionize the task of identification if their full potentials were realized: photographic identification and DNA barcoding. Both of these technologies could potentially identify organisms completely and require minimal human interaction. Currently both of these technologies are in their infancy.

Photographic Identification

Photographic identification is the use of technology to identify plants from photographs. The only popular use of this technology thus far is the LeafSnap iPhone application [8]. This technology is likely to expand, however, because it uses photo classification techniques that are applicable to a wide variety of image
recognition applications [1].

Currently the leaf photos used by LeafSnap must be taken from a particular angle, essentially posed rather than attached to the plant. An improvement, possibly through 3-dimensional visualization would allow the system to identify the leaves anywhere in a photograph such as in a picture of the complete plant. Another improvement could be the recognition of characteristics other than leaves, such as flowers.

While this technology is promising, it remains to be seen whether it can be generalized to plants other than trees. Many other plants such as grasses have leaves which carry very little identifying information. There is also the problem of posing the sample. Unless photo identification is possible without touching or picking the specimen, certain environmental interests might oppose their use. Since the current implementation is with trees, a few picked leaves is a minor inconvenience to the tree. For other species such as wildflowers, plucking a leaf might cause real damage to the plant or habitat, especially if the current technology becomes widely used. There is also the problem that not every plant is good to touch, identifying stinging nettles or poison oak with this technology might cause trouble for users.

Photographic identification could also be used to enhance other identification methods. It could, for example, be incorporated into a multi-access key to distinguish particular characteristics of the plant and automatically answer questions within the key.
DNA Barcoding

DNA barcoding is the process of identifying an organism based upon a small segment of DNA from a particular location within the organism’s full DNA profile [4]. The location is determined by an area that is both easy to process and gives the best variation to allow for species level identification. Unlike the potentially inaccurate methods for photo identification, DNA barcoding has been shown to be highly reliable. Moreover, it has several advantages over methods that rely on morphology, including complete photographic identification as discussed above. These include not needing an intact or complete sample and the ability to differentiate species with few morphological differences [25].

The vision of DNA barcoding is presented in Dr. Daniel Janzen’s paper “Now is the Time” [15], a world in which everyone has access to a “gadget” that can instantly identify any organism from a small amount of genetic material. In that world the users of such a gadget could be novices, hobbyists, everyday citizens, as well as experts. This would free the experts to pursue analysis while the novices could make useful contributions by collecting the data all around them. The author earnestly argues that, “better keys, more keys, more images on the Web, more Web sites, species pages, more descriptions, more phylogenies, more specimens, more anything,” is not truly sufficient to connect everyday people to the biodiversity all around them. Such predictions obviously have implications for the developers of keys, images, and websites.

Progress in creating a DNA barcode for plants is lagging behind that of animals due to the difficulty of finding a DNA segment that meets all the criteria [13]. However, recent studies are working to build a consensus in this area, and it is likely that the DNA site for plant barcodes will be agreed on within the next
year. Current methods for barcoding are making improvements in their speed and cost effectiveness. According to the Consortium for the Barcode of Life, the laboratory cost of processing a sample is a few dollars; however, such a laboratory requires a much larger initial investment for equipment [6]. Generating a barcode can be done in as little as two hours [14]. This is a long way from the few minutes that researchers envision in the future.

Cameron et al. [6] points out that there is no reason to believe that barcoding will ever be as quick, portable, and cheap as proponents such as Dr. Daniel Janzen [15] suggest. They also note that barcoding requires a tissue sample which makes it inappropriate for identifying certain kinds of specimens such as wild living animals. Depending on the size of necessary tissue sample, it might be destructive to the plants sampled placing it at odds with conservation especially if it is available to everyone as presented in Janzen’s vision. Affordability is also an issue for its use by the general public. Researchers and government inspectors may be willing to pay a certain price for such a gadget, but without signification public interest in the product it may never reach general consumers. Cameron et al [6] goes on to argue that such an interest just isn’t there, and the public would be better served by “well illustrated visual field guides.”

2.4.5 Possible Values

Field guides can be assigned to one of these possible states based upon their identification method.

- **None**  The guide does not help users identify plants in any structured way.

- **Dichotomous Key**  Users must follow a dichotomous key to reach an identification.
• **Single Key**  The guide uses a single key that may contain one or more characteristics for identification.

• **Multi-access Key**  The guide provides multiple characteristics that users can select or skip for identification.

• **Automated**  The guide uses an automated process for identification that does not require user intervention other than providing initial input and viewing a list of results.

### 2.5 Area of Focus

Field guides most often identify only a subset of the plants located within their Geographic Area. The most common sets are “Wildflowers” and “Trees.” While these may seem like simple distinctions, both subsets are open to interpretation. Since “wildflower” has no botanical meaning, it can be difficult to define this group precisely. According to [35], “Wildflowers are wild plants with flowers that may be found almost anywhere, from cracks in city sidewalks to ... mountain meadows.” This definition describes any wild-growing member of the division Angiospermae of the kingdom Plantae, including many plants that are not generally thought of as wildflowers, such as members of the family Poaceae (grasses) and trees from numerous families. A dictionary definition is roughly the same, ”any flowering plant that grows in the woods, fields, etc [37].” A more practical definition would be, “Any wild plant with large, showy, or auspicious flowers.” In practice you will not find plants with small hidden flowers, such as oaks, in wildflower field guides. This definition also brings to light the difficulty in determining if a given plant is a wildflower and thus could be identified in a
wildflower field guide. At least the botanical term Tree has a more specific definition, “a woody plant with an erect perennial trunk at least 3” in diameter at breast height (4.5’), a definitely formed crown of foliage, and a height of at least 13’ [22].” Trees may grow as shrubs (smaller than 13’ with multiple branching trunks) under severe cold or dry climates, adding variability to this definition as well. Users who wish to identify plants that are not trees or wildflowers may have difficulty finding resources although guides that cover a larger range of plants do exist [31, 21].

Covering a larger number of plant subsets may seem advantageous to users, but this is only true if the additional subsets do not hinder the overall usability of the field guide. If adding additional subsets makes all the plants more difficult to identify (or increases cost or portability), it may be better to have several guides that are available separately rather than one combined guide.

2.5.1 Possible Values

- **Single [Area]** The guide only contains information on a single type of plants, with the type (as reported in the guide) included in brackets.

- **Multiple [Areas]** The guide contains information on a multiple types of plants, with the types (as reported in the guide) included in brackets.

- **All** The guide is not specific to any type or types of plants.

2.6 Geographic Area

Every field guide covers a specific geographical area. This area can range from the size of a county [21] to the size of a continent [3]. Field guides which cover
large areas have the advantage of being useful in more places, from a walk in one’s own city to a camping trip in a different state. However, as with including multiple Areas of Focus, increasing the area covered can make the field guide harder to use, especially if there are space limitations and some plants must be left out. Diversity of area also plays a role here as well; areas that are more diverse will require more or larger field guides to accommodate a large selection of the species. For example, California is comparatively diverse with 4844 species within 159,000 square miles. In contrast Texas has 4498 species within 268,270 square miles, a much larger area with only 93% the species of California [12]. Thus Geographic Area must be understood as a combination of square mileage and the location it represents.

2.6.1 Possible Values

The number of square miles covered by the guide along with a place name of the area.

2.7 Number of Species

The Number of Species is the total species and subspecies included within the guide. Most guides will differentiate between subspecies of a single species if subspecies are available. For this factor as well as Percent of Species, this thesis adopts a more general definition of the term species: either a species proper or an intraspecific taxa (subspecies or variations being most common). Since different areas contain different total numbers of species, field guides can only be directly compared by number of species if they cover the same geographic area (both size
and location) as described above.

### 2.7.1 Possible Values

The number of species and intraspecific taxa represented within the guide.

### 2.8 Percent of Total Species

Where possible, we have found sources for the total number of species present in the geographic area of a guide and divided the Number of Species present in the guide by that total number. Because it is difficult to determine the total actual numbers of species within each Area of Focus, all totals are based on the total of all plants in a given area. Because there may be disagreement on the status of intraspecific taxa, both the source of the total and the field guide itself may contain varying numbers of these taxa. This percentage therefore cannot be exact because of these disagreements. It should, therefore, be taken as a rough measurement rather than a hard fact.

#### 2.8.1 Possible Values

The percent of the total plant species and intraspecific taxa in a Geographic Area represented within the field guide, or blank, if no such totals could be found.
2.9 Portability

Portability is the ability of users to take the field guide with them on hikes or outings, where there may be no connection to Internet and users are likely to be carrying the guide along with their other supplies and equipment. Therefore, portability has two considerations.

First, usefulness within the field. This can be summed up in the question: is the guide less useful in the field than it would be in another, more controlled, setting? Needing a constant Internet connection could be a liability for digital guides if they weren’t designed specifically to be used in the wilderness, however, there are other ways a field guide can be impacted in the field. For example, the guide could require hardware that is too delicate for use in the field. Here we can assume that users are able to protect the guide from water or rain, but not from the more subtle elements of temperature, direct sunlight, or dust. There is actually a wide range of suitability for use in the field: some paper guides may be more durable, or some smart phones harder to read in the sun than others. However, the study of these guides and their various responses to the elements are outside of the scope of this paper, and only significant problems which would make a noticeable change in the guides functioning will be considered.

The second consideration is the weight and dimensions of the guide. A heavy, awkward guide will be difficult to carry into the field. Smaller and lighter is better. This factor has the potential to be very important in adoption of the guide since space and weight are important factors in most wilderness excursions. Since in all our sample guides size indicated weight (larger guides were all heavier) only size was taken into account in the rating.
2.9.1 Possible Values

- **Low**  The guide is too large to be easily carried (volume greater than 96 cubic inches) or is significantly impaired by use outdoors.

- **Medium**  The guide is small enough to be easily carried (volume less than 96 cubic inches) or may have some small impairments to non-crucial features when used outdoors.

- **High**  The guide is pocket sized (volume less than 12 cubic inches) and has no change of functionality when used outdoors.

2.10 Ease to Acquire

Not all field guides are equally easy to acquire. A field guide may be out of print or run on hardware that is not widely used. There are two components to Ease to Acquire: first, the cost and availability of the guide itself, and secondly, the cost and availability of the technology that allows the guide to run. The real cost of the technology relating to a field guide can be thought of as a measure of its existing ownership versus its price. For example, if many people already own the technology that is used then the cost of the technology within the field guide cost calculation is low, since only a smaller percent of users would need to buy the technology explicitly for the guide. Conversely, if the technology is not widely owned, its price becomes more important because more users will have to acquire that technology specifically to use the guide. An example of widely available technology with near universal adoption would be a personal computer connected to the Internet. Widely available technology with a strong adoption rate includes smartphones and tablet computers.
2.10.1 Possible Values

- **Low**  The guide is either very costly (over $100), out of print, or runs on expensive hardware (over $100) that is not widely owned.

- **Medium**  The guide is medium cost (between $30 and $100), and either uses no technology or uses technology that is widely available with a strong adoption rate. If the guide is printed, it must not be out of print.

- **High**  The guide is low cost (less than $30) and either uses no technology or uses technology that is widely available with near universal adoption. If the guide is printed, it must be widely available and not out of print.

2.11 Extensibility

Extensibility is the ability of the field guide to contain new knowledge. This can come in many forms. The most simple is the idea of a developer releasing a new version of a field guide application. If the new version of a field guide is free for users of an older version, then old guide has been extended (as is the case with software updates). This sort of extensibility is desirable, as stated in the introduction, as scientific names and distributions often change over time. A higher level of extensibility allows users to contribute new knowledge themselves, usually in the form of adding plants or identification information. Because extensibility refers to the system as a whole, items that users create but that do not contribute to the overall content of the system do not meet the definition of extensibility. An example of a user created item that does not meet this definition would be a note on a plant or a specific species sighting that is not shared with others. This level of extensibility may or may not be useful for users. Some
users may prefer an authoritative guide created by experts rather than one that includes user contributions. Field guides that are less compete, as measured by Number of Species and Percent of Species, will likely benefit the most from user contributions, in order to bolster both of the other metrics. A final level of extensibility would be the ability for users change the field guide itself, rather than just its contents. An example of this level of extensibility would be a digital field guide with open source.

2.11.1 Possible Values

- **None**  The guide can not be extended once released or bought.
- **Developer**  The developers of the guide may extend the guide.
- **User**  Users can enter new content into the guide.
- **Meta**  Users can change the structure and functioning of the guide, including what information it can contain.
Chapter 3

Evaluation of Existing Systems

This section will evaluate seven field guides. There guides were selected because they show a range of the field guides available. Every identification technology described above is represented. The number of field guides available [32] makes finding a representative sample a difficult task. There was the additional concern of evaluations, it is far more difficult to evaluate a guide that is not from a local area since it can not be used in general practice.

3.1 Jepson Manual

The Jepson Manual was written in 1993 and is the definitive guide to the plants of California [11]. It can hardly be called a “field” guide since it is large and impractical to take on long hikes. It also requires some amount of botanical training and knowledge to use. Despite these shortcomings, it is evaluated here because of its authoritative stance and usefulness to botanists. If a user wants to unequivocally identify any California plant, the Jepson Manual is best suited to this task.
The *Jepson Manual* is difficult and time consuming to use. It requires at most 28 decisions to key any plant in California to its proper family but on average requires about 9 or 10. Because of space considerations abbreviations are common as well as technical terminology. This terminology also helps reduce ambiguity. For example, it is more efficient and more exact to refer to a “bract” than a “small, leaf- or scale-like structure associated with an inflorescence or cone.” While a few such terms might be picked up by a novice, the thirteen page glossary and two pages of abbreviations assures most descriptions in the manual can only be understood by an expert.

This complexity can also have a negative impact on the accuracy displayed by users, especially novices. Since correct identification relies on written keys, the users ability to answer the questions asked with complete precision is necessary. Questions often ask minute features that users must have a microscope or field lens to see, or features that could be absent on a specimen, such as the fruit if the specimen is not currently fruiting. A single wrong answer precludes finding the correct identification, and even if a possible identification is found, the description may be difficult to understand and the picture may be missing or incomplete.

Space was a major concern in the inclusion of corroborating information. The need to include every species means that there are no photos, rather, line drawings which feature only a selection of the included plants and of those many include only a few characters (a seed, a leaf, a fruit). However, for experts a precise botanical description is given and this might be just the kind of information desired.
3.1.1 Evaluation

- **Identification Method:** Dichotomous Key

- **Area of Focus:** All  All plants within California are covered.

- **Geographic Area:** 159,000 square miles, California  Covers the entirety of California.

- **Number of Species:** 5867

- **Percent of Total Species:** 100%  Since the Jepson Manual is the definitive source for California plants, it necessarily contains them all.

- **Portability:** Low  Dimensions at 11”x8.5”x2.25”, total volume over 210 cubic inches. No restrictions to use outdoors.

- **Ease to Acquire:** Medium  Is a printed guide and requires no special hardware. Retails for $100.

- **Extensibility:** None  As a printed guide it cannot be extended.

3.2 San Diego County Native Plants

*San Diego County Native Plants* was published in 2006 [21]. It contains all varieties of plants found within San Diego county. The plants are divided into two groupings: trees and shrubs and then herbs. The groups are broken into broad categories of taxonomic classification, gymnosperms, dicots, and monocots. From there families are listed alphabetically with species alphabetical as well. Experts might be able to navigate this network, assuming they can get down to family and then browse, but amateurs will have little luck here. Interestingly this system
allows this guide to be free of the need for a single identifying character to unite
the content which allows it to include an amazing variety of species beyond the
typical wildflowers and trees.

Because of its lack of a real identification method, its Ease of Identification
will necessarily be low. Accuracy is debatable; the photos are good and its in-
clusiveness assures that there are plenty other plants to compare each possible
identification with. However, since it is not organized in a way that necessarily
groups similar looking plants together, an identification could be made without
noticing a similar plant in a different section. Corroborating information is lim-
ited to photos and a brief description.

3.2.1 Evaluation

• Identification Method: None

• Area of Focus: All All plants within San Diego County.

• Geographic Area: 4,525 square miles, San Diego County

• Number of Species: 1088 Lists “over 1000”, 1088 was obtained by count-
ing.

• Percent of Total Species: 52% By its own estimation there are roughly
2100 species in the area, this may be large compared to the Jepson Manual,
but it also points out the San Diego County has more wild plant species
than any other area within the continental United States.

• Portability: Medium Dimensions at 8.5”x5.75”x.75”, total volume roughly
36.7 cubic inches. No restrictions to use outdoors.
• **Ease to Acquire:** Medium Is a printed guide and requires no special hardware. The 2006 (2nd) edition is out of print, but the 3rd edition retails for $34.95.

• **Extensibility:** None As a printed guide it cannot be extended.

### 3.3 Wildflowers of San Luis Obispo, California

*Wildflowers of San Luis Obispo, California* was published in 2010 as a joint effort by the City of San Luis Obispo and the California Native Plant Society. It is clearly a book for novices, it has large pictures, relatively little botanical information, and is organized by color.

Ease of identification is likely high because it targets a very specific area. It has few plants and thus can be generous in the space it gives to each, including a large font size for the description. Accuracy is also likely to be high because of its limited scope. Corroborating information is limited to good pictures but it also contains the months when the flower blooms, which in this limited scope is likely adequate.

#### 3.3.1 Evaluation

• **Identification Method:** Single Key

• **Area of Focus:** Wildflowers

• **Geographic Area:** 147 square miles, San Luis Obispo City and surrounding 5-miles on all sides Assuming the 10.8 sq mi city is a circle and increasing its radius by five miles yielded this result.
• **Number of Species:** 268 Lists “over 190”, this discrepancy could be because the 190 does not include subspecies. It has a fixed number of four plants per page. 268 was obtained by calculating the number of pages with plants and multiplying by the number of plants per page.

• **Percent of Total Species:** More than 16%
Calflora (http://www.calflora.org/) lists 3001 species in San Luis Opispo County, but it also overstates the values for the previous guides (Jepson and San Diego County) by a factor of 1.8. This might be due to the inclusion of non-reproducing populations of non-natives or due to synonyms in Calflora’s lists. Assuming the average over-estimation of a factor of 1.8, the total species should be roughly 1667. Since this guide only represents area near the city, rather than the entire county, the the county number is an overestimate.

• **Portability:** High Dimensions at 8.5”x5.5”x.25”, total volume roughly 11.7 cubic inches. No restrictions to use outdoors.

• **Ease to Acquire:** High Is a printed guide and requires no special hardware, retails for $12.95.

• **Extensibility:** None As a printed guide it cannot be extended.

### 3.4 National Audubon Society Field Guide to North American Trees: Western Region

Published 1980 as part of the National Audubon Society’s line of field guides, *Field Guide to North American Trees: Western Region* is a typical published
guide. It identifies trees by leaf shape, fruit shape and flower color. Despite having 540 photos for identification, it contains only 309 species, as many species have multiple photos in different areas of the key.

Like *Wildflowers of San Luis Obispo, California*, this guide is very restricted in its topic. When identifying trees it’s likely to be fairly usable, although the plainness of the subject matter might make identification harder. Its accuracy might also be affected by the difficulty of the subject. The guide presents excellent corroborating information. Each tree has a detailed description, a small drawing, and information regarding the plants’ range. In fact, each tree gets on average a page of space full of detailed information.

### 3.4.1 Evaluation

- **Identification Method: Single Key** Although it has three keys, leaf shape, fruit shape, and flower color, the keys are separate and do not form a multi-access key.

- **Area of Focus: Trees**

- **Geographic Area:** 4,000,000 sq miles, Western North America
  
  Roughly half the size of North America.

- **Number of Species:** 309

- **Percent of Total Species:** less than 5% Because of its large, non-standard area, it is difficult to get a reliable figure on the total number of species present. Since it includes California, this number must be less than 5%. 
• **Portability:** Medium Dimensions at 7.75”x4”x1”, total volume roughly 31 cubic inches. No restrictions to use outdoors.

• **Ease to Acquire:** High Is a printed guide and requires no special hardware, retails for $19.95.

• **Extensibility:** None As a printed guide it cannot be extended.

### 3.5 Discover Life Website: Wildflowers Guide

The online identification website Discover Life [27] focuses solely on identification but presents relatively little information about the species identified. Rather, they present an amalgamation of varying online resources. This can be a problem because the pages contain some corroborating information they can be difficult to navigate. Separating this information from the identification makes the process less efficient because reviewing corroborating information is an important part of the identification task.

Key writers using Discover Life have many of the same worries as writers for traditional keys in that the key writer must include as much flexibility as needed to anticipate wrong answers by users. This is a large burden to put on the key writer, and it contains the assumption that the key writer is an expert.

Although it includes many local flowers, testing made it clear that very few flowers within the Wildflowers guide of Discover Life can actually be identified. This is likely because many plants are included in the guide but have no identifying information which they can be eliminated. Therefore the search results are constantly clogged by thousands of potentials, diluting the real matches. Ease of identification is negatively affected because users will have to answer many
questions and even then will be unlikely to find a match. This will also have a negative impact on accuracy. It is hard to determine the quality of corroborating information. On the positive side, it displays a lot of information from various websites. On the negative side, it displays all of the pages (including duplicate information) at once, one after another, on a huge single page. It is difficult to navigate and unpleasing ascetically.

3.5.1 Evaluation

- Identification Method: Multi-access Key
- Area of Focus: Wildflowers
- Geographic Area: 9,540,000 sq miles, North America
- Number of Species: 3749
- Percent of Total Species: less than 19% The Flora of North America North of Mexico [10] gives the total number of species to be about 20,100. The number of plants in North America including Mexico must be greater than this.
- Portability: Low As a website, its use depends on access to the Internet.
- Ease to Acquire: High The website is free to use and can be accessed via a personal computer and Internet connection.
- Extensibility: User
3.6 Audubon Wildflowers: A Field Guide to North American Wildflowers iPhone Application

Audubon Wildflowers is an extension of the popular printed series of guides to which *National Audubon Society Field Guide to North American Trees: Western Region* belongs.

Since it includes a multi-access key, it is likely to have a high ease of identification. There are many search functions as well as browsing functions with complement the key. Accuracy is also likely to be high for the same reasons. There are a lot of plants in this guide, however, so that might hurt accuracy. There is a question in the multi-access key to show only plants from a certain location, but all the location areas are large chunks of the United States. A more narrow location selection would help balance against the high numbers of plants in the guide. Corroborating information is displayed in an organized way, and content-wise is similar to a printed guide with range maps, pictures and descriptions.

3.6.1 Evaluation

- Identification Method: Multi-access Key
- Area of Focus: Wildflowers
- Geographic Area: 9,540,000 sq miles, North America
- Number of Species: 1632
- **Percent of Total Species: 8%** The *Flora of North America North of Mexico* [10] gives the total number of species to be about 20,100. The number of plants in North America including Mexico must be greater than this.

- **Portability: High** Using the iPod touch as a typical representative of the technology used, the dimensions are 4.5”x2.5”x.25” for a total volume of 2.8 cubic inches.

- **Ease to Acquire: Medium** Uses smartphone technology, retains for $9.99.

- **Extensibility: Developer**

### 3.7 Leafsnap iPhone Application

The LeafSnap iPhone application use technology to identify tree specimens through the shape of their leaves. Users take a picture of the leaves, and the system uses shape recognizing algorithms to match the leaf to the leaves in the system database. The system then displays the top matches for users to choose from.

Ease of identification is going to be high. User will likely enjoy technology that allows them to identify a plant without the typical amount of user involvement. Accuracy is unknown. The leaf must be posed and correctly orientated to make an identification. It is unknown how these factors will affect identifications make by users in the field. Corroborating information includes specially prepared photographs that show the plants in a variety of positions meant to aid in identification.
3.7.1 Evaluation

- **Identification Method:** Automated
- **Area of Focus:** Trees
- **Geographic Area:** 373 sq miles, New York City and Washington D.C.
- **Number of Species:** 184
- **Percent of Total Species:** Unknown This is a non-standard area where invasive species and cultivated plants are the most common.
- **Portability:** High Using the iPhone as a typical representative of the technology used, the dimensions are 4.5”x2.4”x.48” for a total volume of 5.6 cubic inches.
- **Ease to Acquire:** Medium Uses smartphone technology, is free to download.
- **Extensibility:** Developer

3.8 Summary

Table 3.1 presents a summary of the evaluations of the field guides.
<table>
<thead>
<tr>
<th>Guide</th>
<th>Identification Method</th>
<th>Area of Focus</th>
<th>Geographic Area</th>
<th>Number of Species</th>
<th>Percent of Total Species</th>
<th>Portability</th>
<th>Ease to Acquire</th>
<th>Extensibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jepson Manual</td>
<td>Dichotomous Key</td>
<td>All</td>
<td>159,000 sq. mi. California</td>
<td>5867</td>
<td>100%</td>
<td>Low</td>
<td>Medium</td>
<td>None</td>
</tr>
<tr>
<td>San Diego County Native Plants</td>
<td>None</td>
<td>All</td>
<td>4,525 sq. mi. San Diego County</td>
<td>1088</td>
<td>52%</td>
<td>Medium</td>
<td>Medium</td>
<td>None</td>
</tr>
<tr>
<td>Wildflowers of San Luis Obispo, California</td>
<td>Single Key</td>
<td>Wildflowers</td>
<td>147 sq. mi. San Luis Obispo City and surrounding areas</td>
<td>268</td>
<td>More than 16%</td>
<td>High</td>
<td>High</td>
<td>None</td>
</tr>
<tr>
<td>Nation Audubon Society Field Guide to North American Trees: Western Region</td>
<td>Single Key</td>
<td>Trees</td>
<td>4,000,000 sq. mi. Western North America</td>
<td>309</td>
<td>Less than 5%</td>
<td>Medium</td>
<td>High</td>
<td>None</td>
</tr>
<tr>
<td>Discover Life Website: Wildflowers Guide</td>
<td>Multi-access Key</td>
<td>Wildflowers</td>
<td>9,540,000 sq. mi. North America</td>
<td>3749</td>
<td>Less than 19%</td>
<td>Low</td>
<td>High</td>
<td>User</td>
</tr>
<tr>
<td>Audubon Wildflowers iPhone Application</td>
<td>Multi-access Key</td>
<td>Wildflowers</td>
<td>9,540,000 sq. mi. North America</td>
<td>1632</td>
<td>8%</td>
<td>High</td>
<td>Medium</td>
<td>Developer</td>
</tr>
<tr>
<td>LeafSnap iPhone Application</td>
<td>Automated</td>
<td>Trees</td>
<td>373 sq. mi. New York City and Washington D.C.</td>
<td>184</td>
<td>Unknown</td>
<td>High</td>
<td>Medium</td>
<td>Developer</td>
</tr>
</tbody>
</table>

Table 3.1: Summary of evaluations of existing systems.
Chapter 4

Implementation

This chapter presents a novel field guide, SLO Digital Field Guide, and its supporting technologies. SLO Digital Field Guide and its supporting technologies are herein referred to as the “system.” This system has no overall formal name as it is, in essence, a collection of technologies designed to be as loosely-coupled as possible. Each component within the system could be extended, modified, or replaced with only interface changes needed in the directly related components. All components have been open sourced and made available to the public.

4.1 Motivation

As seen in the previous chapter, the field guides surveyed, even digital ones, do not have extensibility as a core feature. They are designed and written by experts. Even though Discover Life allows users to contribute guides, the process is tightly controlled and mediated [27]. Wikipedia has some information on plants, but species level information is often lacking and since the data is unstructured it is difficult to use for identification. Wikipedia is better suited to description
and classification than to identification and inventory. There is no central definitive free source online to identify plants. The lack of a central free source for identification leaves the market to many non-free guides, although most of the information contained in these guides is available in a free form somewhere else. If this free information could be gathered and combined in a useful way, it would make identification free and accessible to everyone.

The open source movement has changed the way applications are developed and shifted the traditional relationship between developers and users [30]. Although open source projects do not necessarily have to center around user contribution and involvement, this is the most successful way for them to operate. This model could be of great use in designing a field guide. An open source model
could mean that the features of the guide could grow and be directed by a community effort. Features for the identification of less eye-catching plant types like scraggly bushes and weedy grasses could be developed since they would no longer be subject to the cost analysis that keeps them out of most for-profit guides. An open source field guide represents the highest level of field guide extendability: users will not longer be limited to adding content. They will have the potential to direct every aspect of the guide. Best of all, an open source license can guarantee that the guide is free, and remains free.

The system, as presented herein, is a complete implementation, although a limited one at best. The purpose of this system is to serve as motivation and a foundation for future development towards a popular, widely used, open source, free field guide. It has been built to run on open source technologies, and all of its own source as been released as open source. Figure 4.1 demonstrates the relationship of each system component to the open source technologies it relies on.

### 4.2 Architecture

The system is divided into four components. The first is the **Plant Identification Wiki** which is a Semantic MediaWiki installation. This component allows users to enter information about plant species. The next component is **Wiki-to-DB**, a small Java program that queries the Plant Identification Wiki and inserts the species information into a MySQL database. The third component is the **Field Guide Service**, a PHP application that functions as a webservice. The final component is **SLO Digital Field Guide**, an Android app that takes data from the Field Guide Webservice and displays it to users. Figure 4.2 shows
the path the species data takes through the system components.

4.2.1 Plant Identification Wiki

Plant Identification Wiki is a Semantic MediaWiki installation. Semantic MediaWiki is an extension to the popular MediaWiki wiki technology [19]. This component’s task is to allow users to enter and view identification information. It serves as the source of data for the SLO Digital Field Guide app. However, it is designed in such a way that other applications could be built using the data it contains.

All content in the wiki is licensed under the Creative Commons Attribution-ShareAlike license, including photos and text. This license allows commercial uses of the Plant Identification Wiki materials. This means that commercial field guide creators can use this data in their works. There are two justifications for allowing commercial use. First, if commercial use does occur, it is likely to “give back” to the wiki in the form of adding or editing content. This could result in making the wiki more complete and correct, making it more useful for all
users. Second, commercial use may increase the exposure of the wiki, drawing more users. The drawback of this licensing is that users who want to use the site must agree to release their content under these licenses. This could hamper contributions, especially of photos. There are plenty of great plant photos online with restrictive copyrights. The Plant Identification Wiki does not need to be another source of these. Rather, it would be more useful for it to contain plant images that can be used for any purpose. Additionally, Wikipedia has the same standards for its licensing, commercial use is allowed on Wikipedia content and images. This means that applications that use Wikipedia will also be encouraged to use Plant Identification Wiki, since material from both of these wikis will be compatible at the licensing level.

Technologies

Semantic MediaWiki extension to MediaWiki is the central component of the Plant Identification Wiki. Using the Semantic MediaWiki allows this component to collect structured data and also take advantage of other wiki functions, such as revision and user access controls. MediaWiki was a strong choice for wiki technology, it powers large wikis such as Wikipedia and has a large user base who are familiar with its syntax. MediaWiki is also open source and under constant development.

The Semantic MediaWiki extension allows users to annotate wiki pages with explicit, machine readable information [19]. It provides three advantages over conventional wikis. First, it provides consistency of content, allowing pages to reuse property values from other pages. This ensures pages that display the same piece of information are linked, so the information can stored in one page, but displayed on many. Second, it provides a way to access knowledge by allow-
ing queries of the information. These queries can even be embedded into pages, facilitating dynamic content generation that can summarize disparate facts and categories appearing throughout the wiki. Third, it provides re-usability of knowledge. Semantic storage and export functionality mean that information within the wiki can be integrated into other applications. For the Plant Identification Wiki this is the most important factor since it enables other applications, like SLO Digital Field Guide, to share in the information it contains. This can also be a motivating factor to encourage users to contribute. If users see their data outside the wiki in applications they use, they may be more motivated to extend the wiki, so they can reap the benefits in the outside applications.

In addition to the Semantic MediaWiki extension, WikiForms was installed into the Plant Identification Wiki to allow users more control over editing wiki pages. WikiForms allows the creation of forms that feed into wiki templates. Wiki templates are structures within pages that allow content to be displayed in a visually appealing way and add consistency to the layout of pages across the wiki. Templates are featured prominently on Wikipedia, where they are responsible for the “infobox” that appears on the right hand side of many pages. Semantic Forms allows categories to be bound to specific templates. Input forms are then bound to the templates. The result is that all pages which are part of a specific category have both a standard layout (the template) and a custom input form to populate the template. The input form is highly customizable and has different inputs for different values types: text boxes, drop down menus, checkboxes and others. Input can also be defined as a file, which links to file upload functionality. Semantic Forms make editing the wiki much easier since users no longer have to know wiki syntax to make edits. It is especially helpful when using Semantic MediaWiki as the extension adds some complexity to the
already intricate wiki syntax.

After its installation, the Plant Identification Wiki was serving pages very slowly. Each request took several seconds and sometimes upwards of ten seconds. This behavior was unacceptable for a user-facing component. APC, a opcode cache for PHP was installed to remedy this problem. APC is a PHP accelerator that stores PHP bytecode to prevent PHP from constantly re-parsing code. APC is especially important in MediaWiki, due to MediaWiki’s complex (many paged) PHP code. Additionally, APC can be used for caching objects queried from the database to prevent connecting to the database as frequently. This object caching using APC was enabled within the Wiki. After installing and configuring APC, page load times are consistently less than a second.

Implementation

This component was not originally conceived as a wiki. The collection of structured data is not the typical use of a wiki. Structured data collection was a necessity as the main critique of current resources is their lack of structured data. Authoritative online sources generally do not save descriptive characteristics in a structured way, rather, they store descriptions in a non-structured paragraph of text. Such sources cannot allow users to search by characteristics, such as flower color, because these characteristics are not stored as separate data items. This is hardly surprising, as the particular color of a species’ flower is of little scientific interest. Flower color, and other structured data items like it, are necessary to create a multi-access key or a system where uses can search by characteristics. The Semantic MediaWiki extension along with Semantic Forms provided the functionality to collect structured data, along with the additional benefits of time-tested content and collaboration management. Overall, utilizing this technology
provided significant advantages and time savings over implementing a full web application from scratch.

Semantic MediaWiki adds the semantics of an ontology to the wiki. It achieves this by adding additional meaning to wiki components so that they represent different aspects of an ontology modeled in the OWL web ontology language (for more information about OWL see [23]). Properties, values inserted into pages, are the basic tool for adding semantic content, and function like OWL properties. Wiki categories become like OWL classes, and pages are like OWL instances. Using templates and forms allow certain properties to be bound to specific categories. This, in effect, creates a class description for that category.

There are some limits to the modeling capabilities of Semantic MediaWiki due its insufficient handling of queries over namespaces. Wiki pages all belong to namespaces, which are prefixed to the front of the page name. MediaWiki installations come with a number of pages in the Special namespace, for example, Special:AllPages, which lists all pages in the wiki. Pages that define categories are in the Category namespace, and pages that define properties are in the Property namespace. Most user generated pages are in the default namespace. The default namespace has no namespace prefix. The problem arises when trying to query within a namespace other than the default namespace. When querying the default namespace, it is possible to select only pages from a certain category. However, when querying a specific namespace, it is not possible to query a specific category, and thus queries will always be run over all pages within a non-default namespace.

This problem presented itself during the process of modeling Flower Color. Flower Color was being modeled as a category, with pages in the category being the various possible flower color: Red, Orange, Yellow, and so on. Flower Color is itself a member of the category Characteristic, which contains various identifi-
cation characteristics for plants: Flower Color, Flower Shape, Number of Petals, and Stature. This set up made it impossible to query for all the identification characteristics for the following reasons: first, they were in the Category namespace, and second, they were members of a category, Characteristic. Since queries cannot define both a namespace and a category, this query was inexpressible to the query engine. This problem was avoided by making Flower Color a part of the default namespace rather than a category. The various flower colors, Red, Orange, etc. were then related to Flower Color by the property Is a value of.

Negating such pitfalls in the design of the wiki’s structure was a non trivial task. A similar maneuver was conducted relating to properties since these are described in the Property namespace. The solution was to create two pages for each property, one in the Property namespace with no category and one in the default namespace with the category Identification Relation.

Figure 4.3 shows a class diagram for the organization of the wiki. Classes represent categories in the wiki. There are five categories in the wiki: Plant, Angiosperm, Characteristic, Identification Relation, and Identification Property. The Plant category is not used directly. Subcategories of Plant are used to assign specific templates and forms. At this time, all plants in the wiki are in the subcategory Angiosperm. The Angiosperm template has fields for all the Identification Properties, plus additional fields for common name, sorting, image, and image attribution.

Characteristics have fields for the associated identification question, how easy the question is to answer, and which Identification Relations make use of the characteristic. The template for the characteristic automatically includes a query in the page which displays a table of those Identification Properties which are values of the characteristic.
Identification Relations are reified properties in the default namespace such as Has Flower Color or Has Number of Petals. Each Identification Relation has a Fuzzy Certainty property which is used to determine the applicability of a value to a characteristic. For instance, Has Flower Color has a 100 Fuzzy Certainty and asserts that the plant indeed has that Flower Color. If users wanted to include a Flower Color that is not present in a plant but could be mistaken (imagine the case of a flower with a dark spot— the spot is black but might be identified as brown in passing), they can instead use Has Flower Color Similar To, which has a lower value for Fuzzy Certainty.

Identification Properties are the values for the Identification Relations, the specific values a Characteristic can take. The different flower colors, for example, are Identification Properties. They have properties which include a human
friendly label, which Characteristic they are related to (using Is a Value of), and an optional icon and description. The human friendly label is usually the page name. Since the page name must be unique, it can get very long and specific. In this case, the label is a shortened version of the page name.

Currently there are 20 plants, all Angiosperms, in the wiki. There are a total of four Identification Properties and four Characteristics. There are 21 Identification Properties.

### 4.2.2 Wiki-to-DB

Wiki-to-DB is a small Java program that queries the Plant Identification Wiki and inserts the results into a MySQL database. It is necessary to store the wiki data in a MySQL database for use by downstream applications (i.e. SLO Digital Field Guide) for two reasons. First, the query functionality of Semantic MediaWiki is currently limited. It is easier, simpler, and more standard to query a MySQL database than to query the wiki. Second, if the data for downstream applications came directly from the wiki, then it would be subject to the irregular whims of wiki editing. A missing property or relation, or worse, a relation that is not semantically valid could cause errors in the downstream applications. The wiki data needs a place were it can be stored in a form where its consistency can be guaranteed. This component cleanses and validates the wiki data before it inserts it into the database.

Additionally, Wiki-to-DB ensures that the data structures for downstream applications does not have to match exactly with the wiki structure. The wiki, as discussed in the last section, often needs special design considerations to properly express certain concepts. By preforming a translation of the data during the insert
process, this component ensures that the database is not hindered by design
considerations for the the wiki. Overall, Wiki-to-DB promotes the independence
of system components and encourages reuse and expansions of the different parts
without disrupting the system as a whole.

This component could, however, be a liability for future wiki improvements
since it embodies a mapping between the data structure of the wiki and the data
structure of the MySQL database. If the wiki was refactored or its structure
changed, this component would be unable to translate the data properly. How-
ever, since it is relatively small (less than 500 lines) it would be easy to edit. It
is also not set up to run automatically. Instead, it can be run with supervision
when it is needed.

Wiki-to-DB uses the Special:Ask page of the Semantic MediaWiki to conduct
its queries. This page can be used by a human user to manually construct queries
and view result tables. It can also be used by a program, by inserting the relevant
fields into the GET parameter in the URL. The Special:Ask page is also able to
export in CSV for easy parsing of the results. For processing the CSV in Java,
Wiki-To-DB uses the OpenCSV library.

An example of the translation Wiki-to-DB preforms is the case of translating
filenames stored for images in the wiki to full URLs for the files. Inside the
wiki, only the file name is stored, and this is the only information needed to
access or display the file. However, the full URL is needed if the image is to be
accessed outside of the wiki. By default, MediaWiki stores images in hierarchy
of directories under the Image directory in the installation. The directories that
contain an image are determined by a substring from the MD5 hash of the image
filename. This distribution of images over many directories makes it easier to
handle spreading the images over my servers in the case of an extremely large
Figure 4.4: Calculating the URL of an image in the wiki.

wiki. The process of calculating the complete URL is as follows. First, the spaces in the filename must first be translated into underscores. Then a MD5 hash is computed on the underscored filename. The first character in the hash is used to calculate the base directory, and then the first and second characters become the name of the second nested directory with holds the file. Finally the computed directories and underscored filename are appended to the known Image directory in the wiki installation. Wiki-to-DB uses the Apache Commons Codec library to conveniently compute the hash. Figure 4.4 shows this process for the file name “Eschscholzia californica.png”.

4.2.3 Field Guide Service

After the field guide data has been stored in the database, it can be accessed through the Field Guide Service. This is a small collection of PHP pages that serve textual exports of the four different entities represented in the wiki. PHP has chosen for this job because of the ease and convenience of set up and implementation. PHP was specifically chosen over Java for this component for performance reasons; there was no reason to set up a Java environment to serve what are, essentially, CSV files.
The separators for the export are the semicolon (;) for records and the pile character (|) for fields within records. These separators were chosen because they were unlikely to appear in the exported text. However, as this component grows, it will likely need to be expanded with either a fully implemented CSV module or a way to export XML as part of a fully capable webservice.

The PHP pages in the Field Guide Service have also been arranged so that additional pages can be added as needed. Right now the page that exports all the plants is <field guide service directory>/plant/all.php. If additional functionality is needed for plants, it can be added to the plant directory. This structure is true of all of the PHP pages.

4.2.4 SLO Digital Field Guide

The Android app, SLO Digital Field Guide, is the final and most complex component of the system. The Android app has been released to the Android App Market where users are can download it for free.

Features

SLO Digital Field Guide has a number of features including:

- Browsing a list of plants.

- Engaging in an “Intelligent Search” where the system asks the user the next question determined to be the most useful.

- Creating a record of a “sighting” where the system records information about a plant that was seen, including the date and GPS location.
Figure 4.5: Relationships between the screens in SLO Digital Field Guide.
• Reviewing and deleting sightings.

• Viewing sightings on a Google map.

• Creating your own Plant record in the application if the plant you are looking for is not included.

The feature list was purposely pruned since the overall goal was to create something that is usable rather than a complete implementation of all the best characteristics of field guides. A view of the various screens and their relationships can be seen in figure 4.5.

**Intelligent Search**

The intelligent search feature’s task is to determine the correct identification by presenting questions to users. It’s task is essentially inferring a decision tree that will lead to the correct identification. The decision tree represents which questions will be asked and what order they will appear. J.R. Quinlan’s "Induction of Decision Trees" presents a foundation for understanding decision trees used for categorization [29]. The decision tree that is needed for the problem of Intelligent Search has several important constraints.

• Users should have to answer the fewest possible so the decision tree should be as compact as possible.

• Users may skip questions.

• Users may give multiple answers to a question and plants may have multiple characteristics of the same type. For example, a flower may be both orange and yellow.
Figure 4.6: Screens used for Intelligent Search.
• The database may change to include more plants, so the questions must be calculated to be relevant to the current contents of the database.

• Users are to be presented with a list of possible answers with weights rather than a single determination.

• Possible values for characteristics are never “hidden.” These values are still displayed even if there is no matching plant within the current set of possible plants. This allows users the ability to change their identification mid-search if they answered an early question wrong.

• Two plants may have all the same characteristics.

In an effort to conserve processing power, SLO Digital Field Guide infers the decision tree entirely based on SQL statements. The Intelligent Search process takes two steps. First, determine and present the next question believed to be the most relevant. Second, when the user answers the question, check if an identification can be made or if there are no more questions to ask. If so, display a list of the top answers for the user to choose. If not, repeat the first step.

The next question is determined by taking all plants with at least one matching answer and finding the characteristic that has the largest total number of values. To determine if an identification has been made, the the plant with largest sum of weights (fuzzy logic values from the wiki) for the answered questions is found. If that plant has a weight more than one hundred greater than the next best plant, then the best plant is returned as the identification. This process is relatively simple and quick to calculate, but may not be as reliable as a more formal approach, such as calculating the information gain [29]. However, because of the constraints listed above, this problem requires significantly more flexibility than the classical methods for calculating decision trees provide.
The Intelligent Search feature also allows users to review all the questions in the system and view and edit their answers. There is also functionality to “skip ahead” to matching plants to view those plants which as currently the best matches given the questions that have been answered thus far. Figure 4.5 shows the screens used for Intelligent Search.

4.3 Evaluation

This section evaluates SLO Digital Field Guide against the framework presented in the second chapter.

Ease of identification should be high due to the restricted subject matter. However, the process of asking questions until the questions run out or a single match is found might be laborious, especially once more questions are added to the system. Accuracy is difficult to judge. The small number of plants might improve accuracy. Pictures, especially in the identification list, are smaller than the typical paper field guide, and might decrease accuracy. Corroborating information is limited. After an identification is found there is a photograph and a list of characteristics, but no additional information.

4.3.1 Evaluation

• Identification Method: Multi-access Key

• Area of Focus: Wildflowers The area of focus could be classified as all, but the system currently contains only wildflowers.

• Geographic Area: 3,304 square miles, San Luis Obispo County
• Number of Species: 20

• Percent of Total Species: 1% Uses the 1667 plants within San Luis Obispo County, as figured in the evaluation of Wildflowers of San Luis Obispo, California.

• Portability: High Using the HTC Evo as a typical smartphone device it needs to run, dimensions at 4.75”x2.25”x.5” volume is 5.3 cubic inches. As long as the use downloads the plants to their phone’s storage device, it has no restrictions on use outdoors.

• Ease to Acquire: Medium Free to download, uses smartphone technology.

• Extensibility: High Open source technologies mean that all aspects of the guide are extensible.
Chapter 5

Usability Analysis

To gain additional insight into the characteristics Ease of Reaching Identification and Accuracy and collect feedback on the design of SLO Digital Field Guide, a usability study was preformed. This study centered around SLO Digital Field Guide and compared it against two other field guides. Before the usability study was performed, each field guide was evaluated as to whether is should be included in the study. The study was conducted by instructing volunteers to identify plants using the selected guides, recording the identification results, and then presenting them with a survey about the effectiveness of the guides they used. After the study was conducted, results were evaluated. This lead to some usability enhancements for SLO Digital Field Guide based on the feedback received.

5.1 Preliminary Evaluation

The purpose of the preliminary evaluation was to determine whether the field guides should be included in the usability study alongside SLO Digital Field
Guide. It was not possible to include all the guides because of the limited size of the study, and also because some of the guides only identify non-local plants and thus could not be evaluated. SLO Digital Field Guide would be the primary target of the evaluation. The only plant category with enough entries in SLO Digital Field Guide to be possible for evaluation was wildflowers, so the other guides had to include wildflowers to be applicable for comparison. Additionally, the other guides were selected if: first, it included enough local plants to be used to identify unknown plants from Poly Canyon (a local hiking trail on Cal Poly owned land), and second, if it was at least usable enough to be used by non-experts to reach an identification in a reasonable amount of time. A reasonable amount of time was defined as less than five minutes since the study was meant to be conducted in ten minutes total.

- **Jepson Manual:** **Reject** Its use of technical jargon precluded its inclusion as it cannot be efficiently used by novices.

- **Wildflowers of San Luis Obispo, California:** **Include** Its large pictures and local selection made it a great candidate. It is also, by the evaluator’s preference, the best book available for local identifications.

- **San Diego County Native Plants:** **Reject** Although it might have been possible to find local plants, its lack of a structured identification method means that it would take users far too long to identify a plant since they would have to flip through the entire book.

- **National Audubon Society Field Guide to North American Trees-Western Region:** **Reject** Eliminated due to its focus on trees rather than wildflowers.
• **Discover Life Website**: *Reject* Because it fails to eliminate most plants through the identification process, would be too laborious or even impossible to identify plants with this website.

• **Audubon Wildflowers- A Field Guide to North American Wildflowers iPhone Application**: *Include* This application works well and is very similar to SLO Digital Field Guide.

• **LeafSnap**: *Reject* LeafSnap does not identify leaves from California and its focus is on trees rather than wildflowers.

### 5.2 Usability Study Set-up

The usability study was designed to compare identification technologies of the selected guides to SLO Digital Field Guide. This section describes the procedures for the usability study.

#### 5.2.1 Plant Collection

The day before the study, four wildflower samples were taken from the upper area of Poly Canyon, past the entrance to the Architectural Village. This collection followed best practices set out in Dr. Keil’s collection guidelines from Botany 313, Taxonomy of Vascular Plants, which prohibit taking samples from lower Poly Canyon. The wildflowers collected were the first four species seen with noticeable flowers and a large enough population to collect without damaging the community. They were: Anthemis cotula (Mayweed), Lotus corniculatus (Bird’s-Foot Trefoil), Mimulus guttatus (Common Monkey Flower), and Stachys bullata (Common Hedge-Nettle). Afterwards, each plant was identified using
Wildflowers of San Luis Obispo, California, and its identity cross-checked online at various photographic repositories.

Each plant was known to be contained within Wildflowers of San Luis Obispo, California. Audubon Wildflowers was searched for each plant, it contained all except Stachys bullata. SLO Digital Field Guide was searched, it did not contain Lotus corniculatus and Mimulus guttatus. These plants were added to the Plant Identification Wiki, then populated down to SLO Digital Field Guide.

Each plant was tagged with a letter, except Stachys bullata which was discarded on the grounds that all the guides should contain all the possible matches.

5.2.2 Usability Study

A total of six volunteers completed the usability study. The process is described in this section. The process took place for each volunteer separately.

The volunteer was presented with one flower and one field guide. The ordering of the flowers and guides were precalculated, so that the maximum number of combinations of different flowers to field guides would be achieved in the fewest number of volunteers.

First, the experimenter gave a brief description of how to use the guide. This included pointing out the basic functions of the guide. For Wildflowers of San Luis Obispo, California the color key was explained. For Audubon Wildflowers, the browse and search features where shown. For SLO Digital Field Guide, the Intelligent search option was shown, along with the associated menu options.

The volunteer was then asked to use the field guide to identify the plant, while the experimenter recorded the time the identification took. After the plant
was identified or the user gave up, the volunteer wrote down the plant’s common name. If the volunteer took longer than five minutes, they were prompted that they may either: continue as long as they like, make a best guess, or give up. Additionally, any questions the volunteer asked the experimenter during the identification were recorded by the experimenter.

This process was repeated one more time, with a different field guide and a different plant with the same volunteer.

After the two identifications, the user was told whether their identifications were correct. They were then asked to complete a survey comparing the two guides they used.

A copy of the survey can be found as Appendix B. Each survey was identical, except the names of the field guides were replaced and re-ordered to match the order of the field guides the volunteer used. Because of this set up, each volunteer only saw two out of the three possible plants and two out of the three possible field guides. However, SLO Digital Field guide was shown once to all volunteers since the purpose of this study was to evaluate its relative fitness to the other guides.

5.3 Potential Sources of Bias

This section’s purpose is to address possible sources of bias within the experimental results of the usability study.

The most natural source of bias is the low number of participants in this study. Although a larger study would have generated more data, the purpose of this study was to have users evaluate SLO Digital Field Guide, not to come up
with verifiable numbers for evaluation time and accuracy. As discussed in the Evaluation Framework chapter, there are multiple inherent biases when measuring ease of identification and accuracy. The numbers presented in the results section should be seen as discussion points, rather than specific proof of any guides superiority.

The next source of bias is the low numbers of plants within SLO Digital Field Guide. This must be kept in mind when asserting the fitness of this program. Since guides with fewer plants perform better in general, all results relating to SLO Digital Field Guide should be viewed in light of the lower bar it had set for it. However, this source of bias was partially controlled by not allowing users to use the Browse Plants feature while identifying, in effect obscuring the small numbers of plants from the user.

5.4 Results

This section discusses the results of the usability study. The first section covers the trial portion, the part of the study where users were timed and asked to perform identifications. The second section covers the feedback portion.

5.4.1 Identification Trial

Volunteers were matched randomly to plants and guides using a series of identifying numbers. These numbers correlate to different combinations of: first guide, second guide, first flower, second flower. The total number of combinations was 24, with the assumption that SLO Digital Field Guide would always appear once. The identifying numbers (and hence what plants and guides the volunteer
<table>
<thead>
<tr>
<th>ID</th>
<th>Guide</th>
<th>Plant</th>
<th>Time</th>
<th>Result</th>
<th>Guide</th>
<th>Plant</th>
<th>Time</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Wild</td>
<td>A</td>
<td>1:33</td>
<td>Correct</td>
<td>SLO</td>
<td>B</td>
<td>1:40</td>
<td>Correct</td>
</tr>
<tr>
<td>2</td>
<td>SLO</td>
<td>A</td>
<td>1:31</td>
<td>Correct</td>
<td>Aud</td>
<td>B</td>
<td>3:53</td>
<td>Incorrect</td>
</tr>
<tr>
<td>3</td>
<td>SLO</td>
<td>A</td>
<td>5:37</td>
<td>Incorrect</td>
<td>Wild</td>
<td>B</td>
<td>2:49</td>
<td>Correct</td>
</tr>
<tr>
<td>6</td>
<td>SLO</td>
<td>B</td>
<td>1:32</td>
<td>Correct</td>
<td>Aud</td>
<td>C</td>
<td>5:10</td>
<td>Incorrect</td>
</tr>
<tr>
<td>11</td>
<td>SLO</td>
<td>C</td>
<td>2:18</td>
<td>Incorrect</td>
<td>Wild</td>
<td>A</td>
<td>1:01</td>
<td>Correct</td>
</tr>
<tr>
<td>16</td>
<td>Aud</td>
<td>B</td>
<td>4:16</td>
<td>Incorrect</td>
<td>SLO</td>
<td>A</td>
<td>0:57</td>
<td>Correct</td>
</tr>
<tr>
<td>17</td>
<td>Wild</td>
<td>C</td>
<td>1:15</td>
<td>Incorrect</td>
<td>SLO</td>
<td>B</td>
<td>2:25</td>
<td>Correct</td>
</tr>
<tr>
<td>20</td>
<td>Aud</td>
<td>C</td>
<td>8:20</td>
<td>Incorrect</td>
<td>SLO</td>
<td>B</td>
<td>1:03</td>
<td>Correct</td>
</tr>
<tr>
<td>21</td>
<td>Wild</td>
<td>A</td>
<td>3:02</td>
<td>Incorrect</td>
<td>SLO</td>
<td>C</td>
<td>3:05</td>
<td>Correct</td>
</tr>
<tr>
<td>24</td>
<td>Aud</td>
<td>A</td>
<td>3:53</td>
<td>Correct</td>
<td>SLO</td>
<td>C</td>
<td>1:30</td>
<td>Correct</td>
</tr>
</tbody>
</table>


used) were assigned selecting those combinations which would create a roughly even representation of the guides and plants, even with a low number of total participants.

Table 5.1 shows the different combinations for each user, as well as the time the identification took and its result. Incorrect results include the user giving up during the identification process. There are a few things we can assume based upon the data. Users tended to get the first plant wrong and the second right. The average time for the first and second identifications was 3:21 and 2:21 respectively and the percent correct was 40Times vary widely for the same user. Everyone got at least one plant correct, and nearly all got one correct and one wrong. Since each field guide and plant were shown an approximately equal number of times first and second, the ordering of the guides will not be considered in further analysis.
Every guide evaluated every plant, although some plants and guides were evaluated in certain combinations more than others. The data sorted by guide and plant is presented in tables 5.2 and 5.3.

From this data, there a few observations that can be made. This data lends support for the theory that field guides with fewer plants are easier to use. SLO Digital Field Guide and Wildflowers of San Luis Obispo, California had the fewest incorrect identifications and lowest times. Plant C appears to be the most difficult plant to identify. At least in its current state, we can make the
determination that SLO digital field guide has no major functionality problems that prevent users from making identifications.

5.4.2 Usability Survey

In addition to the identification trial, volunteers were asked to complete a survey comparing the two guides they used. The results are tallied in table 5.4. The survey results are consistent with the identification trial. SLO Field Guide was the preferred guide, with Audubon Wildflowers being the least favored.

The feedback on SLO Digital Field Guide from questions asked during the identification trial and feedback on the usability survey contained two major criticisms. First, users requested a button to return back after viewing a plant. This functionality was accomplished through the use of the back button on the Android. This confused users. They requested a back button on the screen. Secondly, some of the terminology used in the questions was not known to some users. They requested pictures and/or descriptions for some of the more difficult terms. Both of these suggestions have been incorporated into the final version of the app.
<table>
<thead>
<tr>
<th>Question</th>
<th>Responses</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Use of field guide before today</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>4</td>
<td>40%</td>
</tr>
<tr>
<td>Once or twice</td>
<td>4</td>
<td>40%</td>
</tr>
<tr>
<td>Occasionally</td>
<td>2</td>
<td>20%</td>
</tr>
<tr>
<td>Frequently</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td><strong>Rating of Wildflowers of San Luis Obispo, California</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Useless/ did not work</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>Of limited use</td>
<td>1</td>
<td>20%</td>
</tr>
<tr>
<td>Somewhat useful</td>
<td>2</td>
<td>40%</td>
</tr>
<tr>
<td>Useful</td>
<td>2</td>
<td>40%</td>
</tr>
<tr>
<td>Very useful</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td><strong>Rating of Audubon Wildflowers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Useless/ did not work</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>Of limited use</td>
<td>2</td>
<td>40%</td>
</tr>
<tr>
<td>Somewhat useful</td>
<td>3</td>
<td>60%</td>
</tr>
<tr>
<td>Useful</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>Very useful</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td><strong>Rating of SLO Digital Field Guide</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Useless/ did not work</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>Of limited use</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>Somewhat useful</td>
<td>1</td>
<td>10%</td>
</tr>
<tr>
<td>Useful</td>
<td>4</td>
<td>40%</td>
</tr>
<tr>
<td>Very useful</td>
<td>5</td>
<td>50%</td>
</tr>
<tr>
<td><strong>Preference of Guide</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Highly prefer Other Guide</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>Prefer Other Guide</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>No preference</td>
<td>1</td>
<td>10%</td>
</tr>
<tr>
<td>Prefer SLO Digital Field Guide</td>
<td>6</td>
<td>60%</td>
</tr>
<tr>
<td>Highly prefer SLO Digital Field Guide</td>
<td>3</td>
<td>30%</td>
</tr>
</tbody>
</table>

Table 5.4: Survey questions and responses.
Chapter 6

Conclusion

This thesis presented three contributions to the state of field guides for plants. First, it presented a novel framework for the comparison of field guides, paper and digital. Second, it reviewed seven existing field guides using the framework. Finally, it presented a novel field guide, SLO Digital Field Guide, and its supporting technologies. Additionally, it presented a usability study of SLO Digital Field Guide and two other guides. This study confirmed that SLO Digital Field Guide is a viable identification solution.

6.1 Future Work

This work could be extended in a variety of ways. First, some of the technologies they were used by SLO Digital Field Guide could be extended. Since these technologies were all open source, there is significant room for user contributions. Second, SLO Digital Field Guide and the Plant Identification Wiki could be extended in various ways. Finally, the standards for evaluating field guides could be extended and clarified by the field guide community.
Semantic MediaWiki is the core technology for the Plant Identification Wiki. This component is in need of features for better querying of semantic data, as well as additional features in other areas. This technology is extremely promising as it combines the established benefits of wikis with the power of semantic reasoning.

Both SLO Digital Field Guide and the Plant Identification Wiki can be extended in many ways. SLO Digital Field Guide could be extended with more features to allow users to share their data. Additionally, new identification methods such as photographic identification could be added to increase usability. The Plant Identification Wiki could be extended in both its catalog of plants as well as the structure of its data. Adding more characteristics could make it more general for identifying a wider variety of plants.

Finally, the field guide community must adopt better standards for measuring and discussing the usability aspects of guides, especially for characteristics such as Ease of Reaching Identification and Corroborating Information. A common set of plants to test field guides with would go a long way to making comparisons between field guides easier. Additionally, as new technologies are applied to field guides, they need to be tested on not only how well they implement the technology but also on how effective they are at increasing usability when compared to existing solutions.
Bibliography


Appendix A

Glossary

This section defines the most commonly used terms within this thesis.

**Characteristic**  A single bit of information about a plant, such as its flower color.

**Dichotomous Key**  A identification method where a user must repeatedly select between two possible sets of characteristics until an identification is be reached. Users may not skip states.

**Field Guide**  A tool with the primary purpose of identifying an object of natural history.

**Identification**  The process by which a user determines the species of an unknown specimen.

**Plant**  A tracheophyte. A member of the kingdom Plantae with vascular tissue.

**Taxon**  A group of one or more organisms.
Appendix B

Usability Survey

The following page contains a duplicate of the surveys given to volunteers in the usability study. Some of the surveys had different field guide names and different orderings to match the field guides they were asked to evaluate.
Field Guide Usability Study Survey

Please circle a multiple choice answer for each question. Comments are optional but appreciated, feel free to use the back of this page for comments.

1) Before today have you ever tried to identify a plant species?
   a) No
   b) Once or twice
   c) Occasionally (three or more times, but less than once a month)
   d) Frequently (more than once a month)

2) The first field guide you used was WILDFLOWERS OF SAN LUIS OBISPO, CALIFORNIA- how would you rate this guides usefulness relating solely to identifying plants?
   a) Useless/ did not work
   b) Of limited use
   c) Somewhat useful
   d) Useful
   e) Very useful

Comments regarding WILDFLOWERS OF SAN LUIS OBISPO, CALIFORNIA:

3) The second field guide you used was SLO DIGITAL FIELD GUIDE- how would you rate this guides usefulness relating solely to identifying plants?
   a) Useless/ did not work
   b) Of limited use
   c) Somewhat useful
   d) Useful
   e) Very useful

Comments regarding SLO DIGITAL FIELD GUIDE:

4) Of the two field guides, which would you prefer to use?
   a) Highly prefer WILDFLOWERS OF SAN LUIS OBISPO, CALIFORNIA
   b) Prefer WILDFLOWERS OF SAN LUIS OBISPO
   c) No preference
   d) Prefer SLO DIGITAL FIELD GUIDE
   e) Highly prefer SLO DIGITAL FIELD GUIDE

Explanation/ comments about preference: