

BIODIVERSITY INFORMATION SYSTEMS USING GEOREFERENCING AND PHYSICAL IDENTIFICATION

Tiago Lira

Faculdade de Engenharia da Universidade do Porto, Porto, Portugal

Cristina Ribeiro, João Correia Lopes

DEI - Faculdade de Engenharia da Universidade do Porto / INESC Porto, Porto, Portugal

Keywords: Biodiversity informatics, Web applications, Georeferencing, Physical object identification, Data standards.

Abstract: Biodiversity information can be found in diverse information systems, ranging from large technical repositories for research purposes to dedicated databases supporting the web presence of natural sites. To deal with individual botanical objects, it is important to adopt the established scientific taxonomy. Providing a satisfying experience with current interface devices is also mandatory. Two problems have been addressed in this work: the unique identification of the objects and the representation of their physical location. The former has been tackled with QR code, a two-dimensional barcode standard, and the latter with the use of a PostgreSQL/PostGIS spatial database.

The paper describes the development and test of a Web information system integrating physical identification and geo-referencing. The application in biodiversity information management shows that low-cost, state-of-the-art technologies can provide appropriate solutions even for small- and medium-sized natural sites, incorporating existing scientific information on species and contributing for a growing record of their occurrences. The system was tested on a university garden and can be used by site maintainers, visitors and researchers. The dataset collected in the site can be exported according to the Darwin Core biodiversity standard for research purposes.

1 INTRODUCTION

The year of 2010 was declared International Year of Biodiversity by the United Nations, both as a celebration and as an invitation to safeguard the variety of life on earth¹.

Biodiversity informatics is gathering researchers from biology and from several information and computer science domains to handle the collection, analysis, search and application of information regarding biological diversity. The massive amounts of data in this field are being organized by international organizations (ITIS, 2010; Wikispecies, 2010; IPNI, 2010; Smithsonian Institution, 2010). In the near future, it is reasonable to expect that application developers will have access to high-quality information on species that they can explore at a convenient level of detail.

There are ongoing international initiatives promo-

ting the organization of large digital botanical resources. The Global Biodiversity Information Facility (GBIF) is an example of an organization aimed at providing an informatics infrastructure to support biodiversity applications (GBIF, 2006). The concept of an individual, as an instance of a species, is present in the Darwin Core biodiversity standard proposed by GBIF. Gathering information on individuals is very important when dealing with applications where the representation of a natural site requires the identification of specimens.

Many existing biodiversity repositories tend to be narrow in scope and have very specialized information on some categories. They may cover just one genus, the flora from a region, or a set of endangered species. They are therefore not suitable for services aimed at lay users, where it is necessary to have non-specialized descriptions of a broad range of species.

In this work we focus on the management of botanical information in natural sites. The people in

¹Introduction to the International Year of Biodiversity, <http://www.cbd.int/2010/about/>.

charge of a site want to represent the data about their specimens, to manage their evolution and to present them to the public in appropriate ways. We have assumed also as requirements the use of rigorous botanical data and the association of geo-referencing information to each object. Currently, it is hard for small and medium-sized natural sites to get botanical information covering their specimens and aimed at lay users, and it is also not common for a site to contribute the data it gathers to some larger repository. As a consequence, these sites release little information about their collections to visitors, as promotional resources are expensive to maintain, and they rarely share their data with other organizations.

The proposed system provides the tools for promotion and scientific contribution, therefore increasing the public interest in them. The concept of information systems for a natural site can be redefined with the application of current technologies. We propose the integration of geo-referencing and physical object identification features to improve the usefulness of these systems. This creates a new approach, both in the way they manage their specimen collection and in the way they interact with site maintainers, visitors and the scientific community. Our system is supported by a spatial information platform and geographical data visualization applications. Its contribution for Biodiversity Informatics research is on two main aspects: the integration of tools that contribute to a more accurate representation of specimens and the ability to share standardized information. The system is demonstrated using the capabilities of modern mobile devices to bring relevant content to the site users.

This work explores the capabilities of information systems for efficient management of objects on large spaces, creating a collection of geo-referenced and physically identified elements, making them searchable and making their information immediately accessible. While this approach can be useful in many fields, the significant dispersion of specimens in botanical collections makes them a challenging application domain.

1.1 Contributions

The main contributions of this work are:

- the evaluation of physical identification methods in botanical collections;
- the integration of spatial information in the database supporting a natural site;
- the development of an information system for small and medium-sized natural sites that offers innovative features for object identification and geo-referencing;
- the use of a data model for natural sites accounting for both the botanical information and the site-specific management information;
- the use of biodiversity information standards for exporting data from a natural site;
- the instantiation of the system in a real-world botanical site in order to evaluate its functionality.

2 EXISTING SYSTEMS

Biodiversity information platforms are being created to support the storage and access to large biodiversity datasets. The systems that manage information on biological specimens range from large scientific repositories, used in research, to the databases that support the management of natural sites, to informal collections of information contributed by amateurs.

Large scientific repositories are supported by powerful databases and are now focusing on the sharing of information between organizations. A problematic barrier is found in taxonomic differences and data model inconsistencies (Chavan et al., 2005); for this reason many repositories are contributing to the definition of standards to share and transfer their information. The use of spatial data is also becoming common, as the information on the location and distribution of specimens is valuable for research in this field. As for the services presented to users, the research-oriented systems focus more on search functions to retrieve technical information from their databases; there usually isn't much concern about the details of user interaction (Neale et al., 2007).

The GBIF² network is a good example of a large-scale repository on biodiversity. This international initiative displays information from hundreds of organizations around the world. It includes data from both natural history collections and from observational records. Combining information from diverse sources, it uses the Darwin Core standard as a preferred format for data exchange.

Repositories are now giving attention to spatial information, registering the geographic location of the specimens and allowing its visualization in graphic format. As an example, the Ocean Biogeographic Information System (OBIS, 2010) includes in their search results a distribution map for the specimens (see Figure 1).

²GBIF: Infrastructure, <http://www.gbif.org/informatics/infrastructure/>.

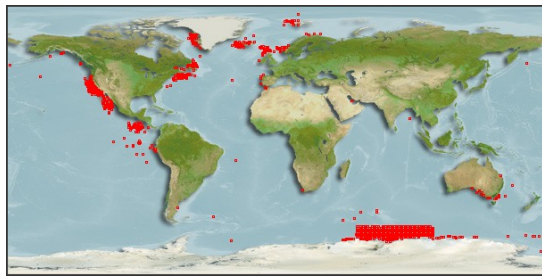


Figure 1: Distribution map generated by the OBIS system.

The systems that support the information for natural sites, on the other hand, are usually less complex in their services, and tend to give simple descriptions of the site with the intention of inviting visitors. Still, some of these systems present some spatial information, in the form of dynamic maps of the space, giving visitors a preview of the site. Graphical presentation of the information is more important here, and no great detail on the collection's specimens is required. An example with a remarkable graphical interface for a virtual visit is the Monticello Explorer website (Thomas Jefferson Foundation, 2010) shown in Figure 2; the user can navigate the interactive map and see an interesting description of the highlights in each area.



Figure 2: Monticello Explorer's interactive map.

3 AN INFORMATION SYSTEM FOR NATURAL SITES

The developed information system offers innovative features for managing and publishing information about a botanical site. The main characteristic of this system is the implementation of a data model capable of combining the internal information of the site, the visitor-oriented information, and the scientific information in a coherent manner. The approach is based on current technologies, namely geo-referencing and physical identification.

The Darwin Core biodiversity standard (Darwin Core Task Group, 2009) is used for exporting

datasets, contributing them to the scientific community and making it easy to share information with other organizations.

The proposed information system is designed for small and medium-sized natural sites, allowing them to manage their information in an effective way. It includes a spatial database and a web application which presents information on botanical specimens.

3.1 Specimen and Species Information

A major goal in this work is to supply useful descriptive information to visitors and site workers. They can view information pages, used to display the data related to the specimens. These pages can be accessed from the various services of the system. An individual page focuses on a specific specimen or set of specimens, and includes its description, photographs, general information about the corresponding species and its taxonomic classification. Figure 3 gives an example of one of these pages, where species information uses the scientific descriptors in Latin.

Specimen:

Identification: 1
 Description: Front garden oak
 Notes:
 Last modified: 2010-05-13

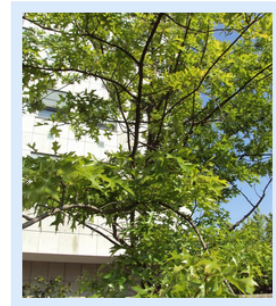
Species information:

Scientific name: *Quercus coccinea*
 Common name: Scarlet oak

Form: Tree
 Life cycle: Deciduous
 Reproduction: Sexual

Roots: Primary
 Leaf type: Simple
 Leaf shape: Ovate
 Leaf margin: Lobate
 Inflorescence: Catkin
 Flowering months: 4 to 5
 Fruit type: Nut

Origin: North America



Classification

Quercus coccinea
 Kingdom: Plantae
 Division: Magnoliophyta
 Class: Magnoliopsida
 Order: Fagales
 Family: Fagaceae
 Genus: Quercus

Figure 3: Information page for an oak tree.

3.2 Spatial Services

This work explored the geo-referencing of the site specimens, allowing the user to find the physical position of any specimen. This information can be easily inserted and corrected by the site staff, and some features were developed based on this information.

The main spatial feature is the virtual tour map, where the user can view a dynamic map of the site,

with specimens represented by markers and specimen sets by polygons. Clicking in these markers the user will have basic information in popup windows or go to the full information page of the specimen. This feature also includes the suggestion of tour paths (see Figure 4), formed by sequences of interesting points chosen by the site staff.

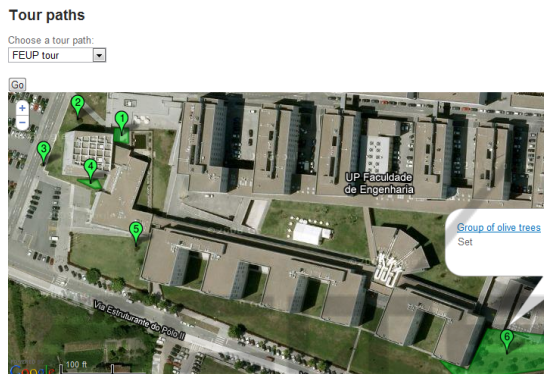


Figure 4: A suggested tour path.

A more demanding use of the spatial database capabilities is the geographic search feature. Here the user selects a polygon in the map, which results in a spatial query that returns the elements contained or intersected by that polygon (see Figure 5).



Figure 5: Geographic search.

These features are the more obvious examples of the potential of spatial data for natural sites. Other services might be developed, depending on the scope and needs of an individual site.

3.3 Physical Identification and Mobile Services

Physical identification allows a user to see a specimen and access its information (and even update it)

on site. The QR code (Denso Wave, 2010), a two-dimensional barcode standard, was chosen for this purpose because it can be read by most mobile devices with standard software. Any user with a common cell phone equipped with a camera and basic Internet access can use this service.

For visitors, the local tour assistant feature is supported by this technology. The user identifies the specimen with his mobile phone (see Figure 6), and the device connects to the corresponding information page in the Web application. This supplies casual visitors and researchers with relevant information during their visit to the site, instead of having to search for it later. For the site staff, there are other uses for this technology. They can edit the specimen's information when they are working, register operations they did or upload photos of the specimen, all from their mobile phones and using the identification codes. This ability to easily identify a specimen and access its related contents creates a richer interaction with the system, making it more useful. The code tags are generated automatically, using an external API, ready to be printed (Google, 2010b).



Figure 6: Identification using a QR code.

To offer these features, the Web application must be suited for mobile devices. There are several options to solve this problem (Ekris and Vugt, 2010; Mehta, 2008; Harris, 2008). In this case, as all the services are intended to work in the mobile version as well, a separate stylesheet was created to format the pages according to the specification of mobile device screens. The system detects the kind of device being used, either by the browser properties or the screen size and selects the correct stylesheet. This directs the user to the services in his mobile browser (see Figure 7). The support of the different features depends on the specifications of each device; while the basic features like information pages can be used even in low-end mobile phones, the more resource-consuming features like the dynamic map will re-

quire more powerful devices to work correctly (current smartphones should be able to use all the features in the mobile version).



Figure 7: Mobile version of the Web application.

Mobile technologies are constantly evolving and its use will predictably continue to grow in the near future (Kendrick, 2010). The combination of mobile technologies with QR codes makes information in large open spaces easily available to visitors.

3.4 Scientific Information

While the above features are sufficient for casual visitors, researchers have a need for more formal and organized information. The support for the Darwin Core biodiversity standard allows the site to export its datasets in a format that can be read by other organizations, thus making it possible to contribute to the ongoing efforts on botanical repositories. Researchers can download the site's data in the specified format, either for research or for integration in larger repositories.

When the user requests the data for a selected specimen, the system provides an XML file with the relevant terms of the Darwin Core description³. This includes specimen identification, details on its geographic location, the scientific name of the species and its taxonomic classification (see Figure 8).

3.5 Search and Additional Features

To fulfill the needs of the system, some additional features were implemented, including various search options. Apart from the above mentioned geographic search, the user can search by names (specimen description, common and scientific names of the

³Darwin Core XML Guide, <http://rs.tdwg.org/dwc/terms/guides/xml/index.htm>.

```
<dcterms:Location>
<dwc:decimalLatitude>41.1782966451506</dwc:decimalLatitude>
<dwc:decimalLongitude>-8.59814286231995</dwc:decimalLongitude>
</dcterms:Location>
<dwc:Taxon>
<dwc:scientificName>Quercus rubra</dwc:scientificName>
<dwc:kingdom>Plantae</dwc:kingdom>
<dwc:phylum>Magnoliophyta</dwc:phylum>
<dwc:class>Magnoliopsida</dwc:class>
<dwc:order>Fagales</dwc:order>
<dwc:family>Fagaceae</dwc:family>
<dwc:genus>Quercus</dwc:genus>
</dwc:Taxon>
```

Figure 8: Excerpt from the Darwin Core description of a specimen.

species), and attributes (the user selects one or more characteristics from the available options, and the system retrieves the specimens that verify them). Basic user-related features were also implemented, including authentication, access control, and user management, to achieve a complete application that can be put to real use.

4 ARCHITECTURE

The system is implemented as a web application developed in the *PHP* language and using the *Zend Framework*. *Zend* is a flexible framework that features the Model-View-Controller design pattern and an extensive library of classes. Its use leads to the creation of a solution with good portability, meaning that it can be easily adapted to any natural site, requiring only a suitable web server and little configuration. From the available choices, the *Zend Framework* was selected for its flexibility of use and good documentation (Lapsha, 2008).

The data model for natural sites comes from the work of João Silva (Silva, 2009), and was refined to better suit the current needs. As shown on Figure 9, it can be divided into 4 main parts: the *Park* package includes classes related to the site and its users, the *Identification* package focuses in the data about specimens, the *Taxonomy* package deals with classification and species information, and the *Geography* package holds the spatial data.

The spatial database is based on this model. The *PostgreSQL/PostGIS* technology is used for supporting the geographical information. *PostgreSQL* is an open-source database system renowned for its reliability, stability and scalability. *PostGIS* (Obe and Hsu, 2010; *PostGIS*, 2010), its spatial extension, is a solution for storing and managing geographic data certified by the Open Geospatial Consortium (OGC).

Geographical information visualization, using a JavaScript-based dynamic map, uses the *Google Maps* service (Google, 2010a). Its combination with

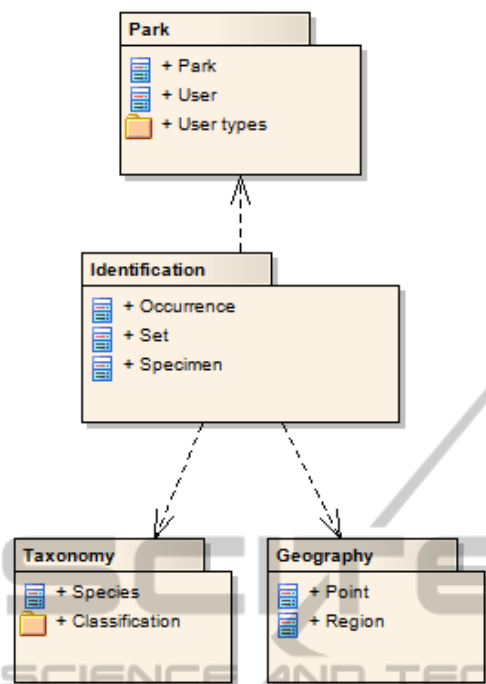


Figure 9: Package diagram of the data model structure.

the spatial database provides geo-referencing, making it possible to insert and visualize the specimen's physical location in an intuitive way, and allowing for other operations like the spatial search feature. The required spatial calculations are accomplished by PostGIS functions (BostonGIS, 2010), making use of the capabilities of the database; the results are returned as text coordinates to be used for visualization in the dynamic map.

As an identification method, the system uses the previously mentioned QR code. There are various alternatives for specimen identification. RFID tags (Smith and Konsynski, 2003; Want et al., 1999) can be an effective solution, but require costly equipment and would be usable only by the park maintainers. The barcode system, on the other hand, is inexpensive and readable on most mobile phones⁴, extending the service to the visitors of the natural site. The QR code standard was compared to other barcode types and chosen for its efficiency and for being supported in the majority of mobile devices (Kato and Tan, 2005; Susono and Shimomura, 2006).

⁴QR-Code Readers, <http://www.mobile-barcodes.com/qr-code-software/>.

5 DATA COLLECTION

A practical aspect which has required some concern was how to populate the database with accurate species information. Users should not be forced to create information such as species taxonomies, which are well established in the scientific community, or even generic descriptors such as specimen size or shape, which are typically already described elsewhere. We have collected general data on taxonomy and species descriptions that can be applied to any natural site. A case study was used to test the system on a real garden.

5.1 Species and Taxonomic Information

After contacting various organizations, the *Illinois Plant Information Network* (Iverson et al., 1999) allowed the use of their data, which included common and scientific names and various descriptive botanical information on plant species. The data was distributed on thousands of text files with organized formatting but somewhat variable content. A small Java application was developed to extract the information and prepare it for insertion in the database. Data was obtained for more than 3000 natural species.

The taxonomic information for the *Plantae* kingdom was imported from a previous work (Silva, 2009). With this data, the system shows the taxonomic classification for each species in the site.

All this information is associated with the plant species, and can be used for any natural site, independently from its specimens. While not all the species have a description, many are covered by the collected data, and others can be completed manually.

5.2 Case Study

The system was applied to one of the university gardens, which is in the process of expanding its specimen collection. The garden has several species and is large enough to take advantage of the developed services.

This deployment of the system allowed us to test its features in a real situation. For that purpose data was collected on various specimens of the garden, including descriptions and photographs, which were associated to the corresponding species. Some of the species were not in the system yet and were inserted with data from various sources (Fernandes and Carvalho, 2003). The geographical coordinates were collected for each specimen, using a *GPS* device. As the device presented significant precision errors, part of

the data was corrected using the application's dynamic map. The collected spatial data was tested in the virtual tour service, visiting the specimens for the frontal area of the garden (as shown on Figure 10). Identification codes were also printed for some of the specimens allowing us to test the local tour assistant and related mobile Web contents.

Virtual tour

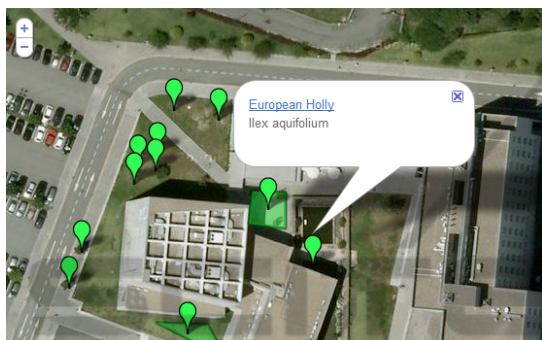


Figure 10: Virtual tour of FEUP's gardens.

6 EVALUATION

The goal of this work was not only the creation of a solution, but also an evaluation of the applicability of the selected technologies in this domain.

Physical identification proved to be an effective service when associated with mobile device contents. The QR codes in particular can be easily generated and printed, and the tests suggest that most common mobile phones can use them. With the evolution of mobile Internet and devices, this service is expected to become even more practical in the future.

The geo-referencing of specimens is a useful feature both for visitors and for site workers, and spatial data can also be important for research purposes. The value of this data increases for larger spaces, where the difficulty of locating a specimen is much more significant.

Exporting data in Darwin Core format is relevant for specialized users and can be valuable for research purposes. The usefulness of the data will depend on its content, the amount of information collected, and the interest of the species in the collection. Nevertheless, the system allows for easy sharing of the data in a standardized format supported by large global organizations.

The success of any system is, of course, dependent on the reaction of its users; therefore, a more detailed evaluation with different user groups would be useful to better assess its practical value. A group of users in

the fields of Biology and Botany would be specially relevant, to test the developed features and identify possible points for improvement. The virtual tour and the local tour assistant features should be tested with visitors.

7 CONCLUSIONS

Biodiversity information systems are evolving in the direction of spatial information. The Darwin Core standard is also becoming more common in the sharing of scientific data. Information systems for natural sites are starting to use new technologies and there are still vast opportunities for improvement.

The work explores two technologies for natural sites. Physical identification of specimens supplies visitors with relevant content in an easy way and improves the efficiency of information management tasks for the staff. Geo-referencing allows the collection and storage of precise information on the localization of the specimens, which can then be visualized by the users. The support of the Darwin Core standard enables the site to share datasets with the scientific community.

This work has resulted in a Web information system focused on improving specimen information management and offering distinct services for visitors, with the support of current technologies. The system is flexible and can be applied to botanical gardens and other natural sites. Hopefully the result will have an impact on biodiversity information systems. These technologies can improve the services of natural sites and their use in this context can be further explored.

The broad scope of the study also creates opportunities for continuation, expanding the work for different uses and related areas. One possibility is to connect the application to a biodiversity repository (some of which already provide APIs for data retrieval), to get relevant information on species or taxonomy whenever necessary. The application might also upload its data periodically to the repository, making it available for research. Another possibility is to expand the system to the animal kingdom, for example in a small zoo, with some modifications on the identification and geo-referencing methods. The spatial capabilities of the database can be further explored as well; global maps may show the origin of the collection's species, or the retrieval place of the specimens. Additionally, the increasingly common GPS capabilities of mobile phones can be used in a location-oriented tour assistant that will point interesting elements near the user's position, or help the staff to

geo-reference a specimen using the coordinates of the device.

REFERENCES

- BostonGIS (2010). Part 1: Getting started with PostGIS. Available at http://www.bostongis.com/?content_name=postgis_tut01, last accessed September 2010.
- Chavan, V., Rane, N., Watve, A., and Ruggiero, M. (2005). Resolving taxonomic discrepancies: role of electronic catalogues of known organisms. *Biodiversity Informatics*, 2:70–78.
- Darwin Core Task Group (2009). Darwin Core. Available at <http://rs.tdwg.org/dwc/index.htm>, last accessed July 2010.
- Denso Wave (2010). QRcode.com. Available at <http://www.denso-wave.com/qrcode/index-e.html>, last accessed July 2010.
- Ekris, J. and Vugt, R. (2010). Optimizing a website for mobile devices — modern nomads. Available at <http://modernnomads.info/wiki/index.php?page=Optimizing%20a%20website%20for%20mobile%20devices>, last accessed July 2010.
- Fernandes, M. F. and Carvalho, L. M. (2003). *Portugal Botânico de A a Z — Plantas Portuguesas e Exóticas*. Lidel, 3 edition.
- GBIF (2006). GBIF strategic and operational plans 2007-2011. Available at http://www.gbif.org/strategic_plans.pdf, last accessed September 2010.
- Google (2010a). The Google Maps JavaScript API. Available at <http://code.google.com/apis/maps/documentation/javascript/>, last accessed July 2010.
- Google (2010b). QR-codes — Google chart tools. Available at <http://code.google.com/apis/chart/docs/gallery/qr-codes.html>, last accessed July 2010.
- Harris, S. (2008). Mobilising a website, part 2: Strategies. Available at <http://pointbeing.net/weblog/2008/07/mobilising-a-website-part-2-strategies.html>, last accessed September 2010.
- IPNI (2010). The international plant names index. Available at <http://www.ipni.org/>, last accessed September 2010.
- ITIS (2010). Integrated taxonomic information system. Available at <http://www.itis.gov/>, last accessed September 2010.
- Iverson, L., Ketzner, D., and Karnes, J. (1999). Illinois plant information network. Available at <http://www.fs.fed.us/ne/delaware/ilpin/ilpin.html>, last accessed September 2010.
- Kato, H. and Tan, K. T. (2005). 2D barcodes for mobile phones. In *Proceedings of the 2nd International Conference on Mobile Technology, Applications and Systems (Mobility 2005), Nov. 15-17, 2005, Guangzhou, China*.
- Kendrick, J. (2010). The (R)evolution of mobile. Available at <http://jkontherun.com/2010/04/01/the-revolution-of-mobile/>, last accessed September 2010.
- Lapsha, S. (2008). Zend framework vs. CakePHP, symfony, seagull, WACT, prado, TRAX, eZ and CodeIgniter. Available at <http://sevalapsha.wordpress.com/2007/11/13/>, last accessed July 2010.
- Mehta, N. (2008). Chapter 4 — adapting to user devices. In *Mobile Web Development*, pages 20–28. Packt Publishing.
- Neale, S. H., Pullan, M. R., and Watson, M. F. (2007). Online biodiversity resources — principles for usability. *Biodiversity Informatics*, 4(0).
- Obe, R. O. and Hsu, L. S. (2010). *PostGIS in Action*. Manning Publications Co.
- OBIS (2010). Ocean biogeographic information system—about the data. Available at <http://www.iobis.org/tech/>, last accessed July 2010.
- PostGIS (2010). PostGIS: Documentation. Available at <http://postgis.refrains.net/documentation/>, last accessed July 2010.
- Silva, J. (2009). Sharing botanical information using geospatial databases. Master's thesis, Faculdade de Engenharia da Universidade do Porto.
- Smith, H. and Konsynski, B. (2003). Developments in practice X: Radio frequency identification (RFID) — an internet for physical objects. *Communications of the AIS*, 12(19):301–310.
- Smithsonian Institution (2010). Index nominum genericorum (ING). Available at <http://botany.si.edu/ing/>, last accessed September 2010.
- Susono, H. and Shimomura, T. (2006). Using mobile phones and QR codes for formative class assessment. *Proceeding of Current Developments*, pages 1006–1009.
- Thomas Jefferson Foundation (2010). Monticello explorer. Available at <http://explorer.monticello.org/>, last accessed July 2010.
- Want, R., Fishkin, K. P., Gujar, A., and Harrison, B. L. (1999). Bridging physical and virtual worlds with electronic tags. In *Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit*, pages 370–377.
- Wikispecies (2010). Wikispecies, free species directory. Available at http://species.wikimedia.org/wiki/Main_Page, last accessed September 2010.