

# Applying Semantic Web Technologies to Support Learners in Plant Identification and Taxonomy

<http://dx.doi.org/10.3991/ijet.v7i3.2089>

K. Jordan

University of Cambridge, United Kingdom

**Abstract**—This paper describes a novel application of semantic web technologies to support undergraduate students' learning in Plant Sciences. The pedagogical context focuses upon a field trip, seeking to enhance students' familiarity with plant species they will encounter in the field and making links to over-arching concepts in the wider taught course. Semantic web technologies were selected as a potential match to support this learning context because of the well-defined plant taxonomy underpinning plant classification. The paper will conclude with reflections on the affordances, challenges and issues surfaced by this approach and its alignment with pedagogical theories.

**Index Terms**—Education, Educational Technology, Semantic Web, Taxonomy.

## I. INTRODUCTION

This project sought to create an online tool using semantic technologies to assist students in learning how to identify plants. The tool will support students in understanding the diversity of plant species they encounter on the course's field trip, making explicit plant classification and relating them to the lecture course.

A key part of the teaching within Plant Sciences for second year undergraduate students is a week-long field course in the Mediterranean. This gives the students the opportunity to examine plants, applying the theory that they have been taught during lectures. To do this, it is important to be able to identify the plants they encounter. On the course this is addressed by a walk and lecture en route by an expert in plant systematics.

Plant identification is not a skill that students usually have at this stage in their career [1] so providing extra support and guidance is key. Teaching students to recognise key characteristics enables them to identify plants on the generic level. Current practice addresses this by providing students with a printed field guide, divided into different ways of classifying plants and the salient characteristics to look out for. While this is a high-quality resource created by an expert in the field, this project sought to explore whether the content could be made more explicit and engaging for students in a different, more interactive medium.

An online tool would be beneficial to support learning outside of the field trip. The field trip and its relationship to the rest of the course may be considered a type of 'experiential learning' [2]; a supporting tool could afford extra partial cycles of learning before or after the trip. The main role for the tool would not be in the field trip itself, as the students have access to the experts during this time, and limited availability of mobile devices might mean that not all students could use it. The key advantage of an online tool is that the time with the experts is fleeting; this

would model the experts' knowledge and make it available to students during the rest of the academic year.

Multiple taxonomies and ontologies make it a good match for semantic technologies. Semantic technologies use aspects or affordances of the concept of the semantic web [3], whereby all information is available in standardised machine-readable formats. Essentially, this means the data about different plant species and their defining characters could be translated to a data-driven webpage known as an Exhibit [4]. This would support enhanced search and faceted browsing [5], allowing many different routes through the dataset and making the relationships more explicit. This could be regarded as a type of scaffolding; the many different plant species can be overwhelming to students, but by structuring the data to reflect the experts' strategies of making sense of the diversity of plants, the way that the tool is constructed may provide expert guidance through the data. Having these expert 'inroads' into the material could be argued to allow the tool to be a 'more knowledgeable other' and extend into the students 'zone of proximal development' [6].

## II. DESIGN & DEVELOPMENT

The starting point for the tool was a field guide to the plant species encountered on the course, available through the institutional virtual learning environment (VLE) course site as a PDF document. Liaison with course leaders highlighted the key concepts that the field course relates to in their wider teaching, flagging up the importance of the field course in terms of structuring revision and applying what has been taught in lectures.

The PDF document presents information about different ways of identifying plants to students. The information is clearly presented and authored by an expert in the field. It is divided into two main parts: grouping plants by families, or grouping plants by growth forms. Within each, the key features to identify particular species are then described. However, as it is a PDF file, the data is static, presented as text and images listed within three types of groupings. The information in each section can only be examined independently of the others; it is not possible to explore relationships between the different main groupings of families or growth forms, for example (see Figure 1 for a diagrammatic representation of the information as presented in the PDF).

In contrast, by taking the data from the PDF and translating it into a spreadsheet, where each species is an entity and has attributes reflecting the different family or growth form groups, the collection of species can then be explored and filtered in a more flexible, dynamic way (see Figure 2 for a diagrammatic representation of the information in this case; compare to Figure 1).

The process of developing a semantic tool to support this comprised two phases: 1. Restructure the data from the PDF into a spreadsheet, categorizing each plant species according to the experts' information in the field guide, including plant systematics, physiology, ecology, and making links to related lectures; 2. Develop the user interface, using Exhibit to display the data.

A. Phase 1: Data modelling

The information from the handout is re-structured into a spreadsheet so that links between data in different parts of the handout can be made explicit. The plant species involved become records, with their synonyms, classification data and key physical characteristics all becoming properties of each record. This can then be converted to an Exhibit-ready format using the 'Babel' web service [4].

B. Phase 2: Visualising the data & developing online tool

The newly-restructured data can then be visualised on a webpage using Exhibit [4]. By configuring the page to the relevant data fields, faceted browsing [5] (see Figure 3) can be supported to allow the dataset to be interrogated using any characteristic as a starting point. The other fields will dynamically change so trends and links between families and physical characters are more explicit. When a species of interest is selected, further details are presented, including active hyperlinks to related lecture-based material in the course VLE.

In order to assist students in their navigation through the data, recommended questions to think about when using the tool were included in the introductory text. These were based on suggestions from the course leader. Once developed, the new tool was embedded within the course VLE and has been accessible for students throughout the academic year.

III. EVALUATION

The concept of 'Illuminative evaluation' was used to frame evaluation of the tool [7]. Illuminative evaluation recognizes that interventions are part of a wider 'learning milieu' and consequences may be unanticipated, so it is best to collect a variety of different types of data and adopt an exploratory approach to research questions [8]. The concept and its principles are particularly relevant when thinking about new technologies and emergent pedagogy.

The evaluation process has included (i) collection and analysis of server log data about levels and periods of use, (ii) survey students to gauge how useful they perceived the tool to be and why, (iii) an interview with the course leader, and (iv) feedback from other academics involved in the course. Feedback on the project was also provided by tutors on a technology-enabled academic practice course, which the tool contributed toward. Over the course of an academic year, site log data (Google Analytics; [9]) revealed two spikes in use of the tool; first, in the week running up to the field course, and second, during the revision period at the end of the year. Unfortunately the response rate to the student survey was low (n=4, which represents ~10% of the course cohort). However, together the sources of evaluation data did provide an insight into the tool and issues surrounding its use; the emergent themes are presented in the next section.

Following evaluation, refinements were made to the tool in light of the results, and the finalized tool was made available for the following years' course.

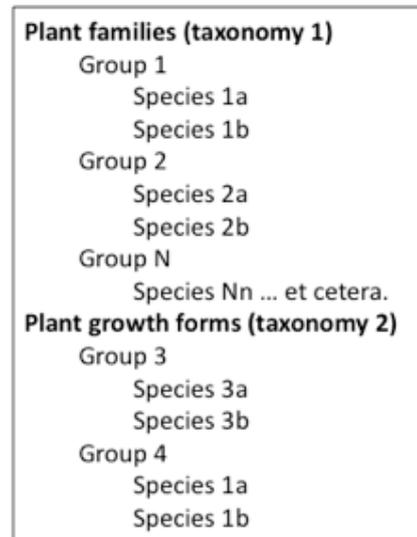


Figure 1. Simplified representation of the way the data is presented in current practice (PDF file). One particular species maps on to both taxonomies but in this format each is treated as a separate list – links and trends between the two are not made explicit.

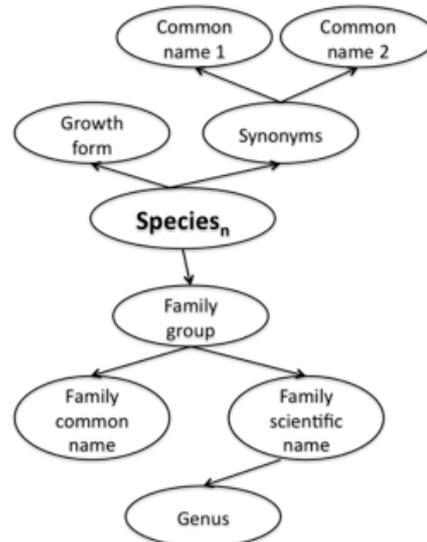


Figure 2. Representation of how the information has been restructured via a spreadsheet ready for use with a semantic web-based tool.

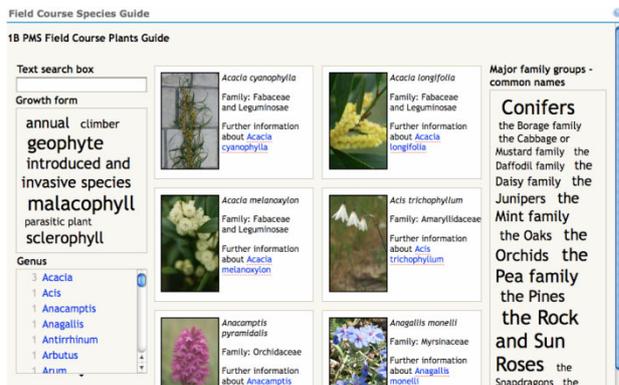


Figure 3. Plant identification field guide as a semantic tool. Note search facets along the left- and right-hand side.

## IV. DISCUSSION

A number of themes emerged from the evaluation data. These included: technology as a stimulus for curriculum design, the addition of a glossary to interpret specialist terms, and sustainability of technology-enhanced learning tool developments.

## A. Curriculum design

Prior to releasing the initial version of the tool to students, the course leader was shown the tool and interviewed. In addition to discussing refinements to the tool, it elicited an unanticipated response. Seeing the links to the lecture material mapped on to the species' guide, he was inspired to make the links stronger and to encourage lecturers' on the course to use examples from the species guide to illustrate the concepts in their lectures wherever possible. This was also highlighted as desirable in the student questionnaire, ranked as the top priority for development. This can be conceptualized as initiating or surfacing a desire for further *constructive alignment* [10] throughout the course. This is an interesting result, as it was not one of the aims of the tool from the outset, but highlights the way that new technologies may have unanticipated benefits when applied to educational settings.

## B. Glossary

This was ranked as the 2nd highest priority for enhancing the tool in the student survey. The structure of the undergraduate course means that the field course is often the first time students encounter plant taxonomy. The field guide contains a large amount of specialist language which as a result may be unfamiliar to students. The new version of the tool includes an extension to the Exhibit containing definitions from specialist terms within the text (see Figure 4).

This point is not unique to this domain but rather may be valuable to consider when applying semantic technologies to other educational settings. Semantic web technologies readily utilize domain taxonomies and ontologies, although it is worth remembering that these represent an experts' way of seeing the domain. While this can be considered an advantage, some thought needs to be given to bridging the gap between novice and expert. A pedagogical tool therefore needs to not only use or represent the expert ontology, but include mechanisms to support learners in developing their own understanding from a novice position.

## C. Sustainability

The issue of how staff will be involved in the ongoing development of the resource was highlighted in the academic developers' feedback. Educational technology interventions run a risk of never being successful in the longer term if steps have not been taken to hand over maintenance to course administrators. Two ways were identified to enhance sustainability:

1. Modification of the data source, from static spreadsheet plus manual upload and conversion, to Google spreadsheet. In the initial tool, the data was structured within an Excel spreadsheet, and converted to the format required for Exhibit (JSON) using the Babel web service [4]. Using this method, the data must be re-converted using Babel whenever changes are made to the spreadsheet. Alternatively, the data can be hosted in a Google spreadsheet, which is automatically converted when the

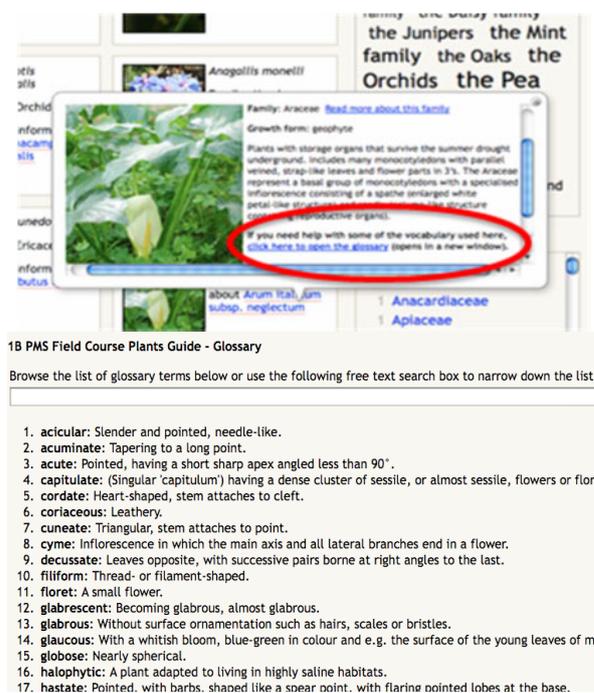


Figure 4. Screenshots showing the glossary extension to the tool (bottom), and how it is linked to from records in the species guide (top).

Exhibit is loaded, so any changes made to the spreadsheet appear in the Exhibit automatically. Using a Google spreadsheet rather than a JSON file has some associated risks: it can increase the time taken for the Exhibit to load, and if Google is down, the data won't load. However, these risks are low and this is a trade-off toward making the data update process much shorter.

2. Links to VLE-based lecture notes and resources. Each year, the courses' site within the VLE is duplicated, creating a new copy for the incoming cohort. The relative file paths within the site remain the same, although numbering of lectures may differ slightly from year to year. To ease the transition from one year to another, the links in the tool (which initially linked straight to PDFs of each lecture) can be replaced with links to the folders for each module, which remain more consistent in naming from year to year [11]. The links can then be updated each year by search-and-replace within the spreadsheet to replace the current site ID with the new one.

## V. CONCLUSIONS

Although progress has been made in some specific areas, such as case-based learning [12, 13] or university administration [14], the ways that semantic technologies can enhance education are not well understood or theorized yet [15]. In developing this semantic tool, it has been challenging to think about the learning theory framing it.

From discussions with the course leader, it is clear that the primary goal of the field course is to bring together the abstract concepts taught in the lecture course, for students to be able to apply what they have learnt first-hand. Plant identification skills are a secondary objective. This is not to say that the tool is not useful; rather, that more emphasis needs to be put on making explicit the links to the lecture course. This resonates with learning theory around scaffolding, guiding students through the material mirroring the way, which the teaching staff do [6, 16].

As it is an online tool, it is natural to consider the tool in terms of connectivism [17], a learning theory related to the notion that the Internet is a form of transactive memory [18]. In the context of this tool, it is the links internally within the content that are important, and in terms of learning about plant taxonomy, mastering and understanding these internal relationships within the dataset are crucial. Constructivism could be a better theoretical lens for viewing this, particularly in terms of the work of Jerome Bruner and the role of structure and categorization in learning [19, 20]. The difference between connectivism and constructivism here is the purpose of learning about links between the data; the tool is not a replacement for or equivalent to knowing the data, but to help instruct on how to think like an expert about plant taxonomy.

While this project began by focusing on a very specific learning context – that is, the identification of plants and application of knowledge within a Plant Sciences undergraduate degree course – its development and evaluation has surfaced issues which maybe helpful for others considering deploying semantic web technologies for educational purposes. As an emergent technology, use of semantic web applications are not yet established in Higher Education. It is not possible to anticipate the ways that emergent technologies such as this will impact upon academic practice; for example here, by making the links between species and the taught course more explicit and presented in a novel way, staff were prompted to reconsider curriculum design and constructive alignment.

The potential to utilize domain ontologies is regarded as an affordance of semantic technologies for education. However, it is also important to remember that in a pedagogical context, these represent an expert view and learners will not be as familiar with the terms underpinning the taxonomy. It also highlights the need for educational semantic technologies to support the transition from a novice to expert way of knowing the subject.

## ACKNOWLEDGMENT

K.J. thanks staff and students in the Department of Plant Sciences at the University of Cambridge for their input into developing this tool. Also to staff in the Learning Development Center at City University, London for their guidance as part of their programme in technology-enabled academic practice.

## REFERENCES

- [1] G.E. Burrows, "Testing times for Plant Family Recognition," *Bioscience Education E-Journal*, 16-3c, 2010.
- [2] D.A. Kolb, *Experiential Learning*, Englewood Cliffs, NJ.: Prentice Hall, 1984.
- [3] T. Berners-Lee, J. Hendler and O. Lassila, "The Semantic Web," *Scientific American*, 284(5), pp. 34–43, 2001. <http://dx.doi.org/10.1038/scientificamerican0501-34>
- [4] D.F. Huynh, D.R. Karger and R.C. Miller, "Exhibit: Lightweight structured data publishing," *Proceedings of the 16<sup>th</sup> international conference on World Wide Web*, 2007.
- [5] M.A. Hearst, "Design recommendations for hierarchical faceted search interfaces". In A.Z. Broder and Y.S. Maarek, editors, *Proc. SIGIR 2006 Workshop on Faceted Search*, pg.26-30, 2006.
- [6] L.S. Vygotsky, *Mind in society: The development of higher psychological processes*. Cambridge, MA.: Harvard University Press, 1978.
- [7] M. Parlett and D. Hamilton, "Evaluation as illumination: A new approach to the study of innovatory programs occasional paper. *Centre for Research in the Educational Sciences, Edinburgh University*, 1972.
- [8] J. Bennett, *Evaluation methods in research*. London: Continuum, 2003.
- [9] V. Rolfe, How to monitor the use and reuse of open educational resources using Google Analytics. *Proc. EDULEARN10*, pg.2320, 2010.
- [10] J. Biggs and C. Tang, *Teaching for quality learning at University*. 3<sup>rd</sup> Ed. Maidenhead, UK: McGraw-Hill and Open University Press, 2007.
- [11] M. Soanes and J. Toner, "Scaffolding unstable and complex structures: the semantic web and complex programme and content design," *Learning at City Conference, City University, London*, 23<sup>rd</sup> June 2011.
- [12] M. Tscholl, F. Tracy and P. Carmichael, "Case methods, pedagogical innovation and semantic technologies". *1<sup>st</sup> International workshop on semantic web applications for learning and teaching support in Higher Education (SemHE'09), ECTEL'09, Nice, France*, 30<sup>th</sup> September 2009.
- [13] A. Martinez-Garcia, F. Tracy, M. Tscholl, S. Morris, and P. Carmichael, "Case based learning, pedagogical innovation and semantic web technologies," *IEEE Transactions on Learning Technologies*, 4(3), in press.
- [14] T. Tiropanis, H. Davis, D. Millard, M. Weal, S. White and G. Wills, *Semantic technologies in learning and teaching (SemTech) report*, Joint Information Systems Council (JISC).
- [15] S. Schaffert, T. Berger, W. Hilzensauer, and S. Schaffert, "Underlying concepts and theories of learning with the semantic web", *Proc. Technology Support for Self Organised Learners 2008 (TSSOL08)*, pg.67-83, 2008.
- [16] S. Lajoie, "Extending the scaffolding metaphor," *Instructional Science* 33(5-6), pg.541-57, 2005. <http://dx.doi.org/10.1007/s11251-005-1279-2>
- [17] G. Siemens, "Connectivism: A learning theory for the digital age," 2004. Online at: <http://www.elearnspace.org/Articles/connectivism.htm> (Accessed 25th July 2011).
- [18] J. Bohannon, "Searching for the Google effect on people's memory," *Science*, 333(6040), pg.277, 2011. <http://dx.doi.org/10.1126/science.333.6040.277>
- [19] M.K. Smith, "Jerome S. Bruner and the process of education," *The encyclopedia of informal education*, 2002. Online at: <http://www.infed.org/thinkers/bruner.htm> (Accessed 18th August 2011).
- [20] J. Bruner, *The Process of Education*. Cambridge, Massachusetts: Harvard University Press, 1960.

## AUTHOR

**K. Jordan** is with the Faculty of Education, University of Cambridge, Cambridge, CB2 8PQ UK (e-mail: klj33@cam.ac.uk).

Submitted, April, 9, 2012. Published as resubmitted by the authors on August, 7, 2012.