Enhancing the Curation of Botanical Data Using Text Analysis Tools

Citation for published version:
Llewellyn, C, Grover, C, Oberlander, J & Haston, E 2012, Enhancing the Curation of Botanical Data Using Text Analysis Tools. in P Zaphiris, G Buchanan, E Rasmussen & F Loizides (eds), *Theory and Practice of Digital Libraries: Second International Conference, TPDL 2012, Paphos, Cyprus, September 23-27, 2012. Proceedings.* Lecture Notes in Computer Science, vol. 7489, Springer Berlin Heidelberg, pp. 480-485.
https://doi.org/10.1007/978-3-642-33290-6_55

Digital Object Identifier (DOI):
10.1007/978-3-642-33290-6_55

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Theory and Practice of Digital Libraries

Llewellyn, C., Grover, C., Oberlander, J., & Haston, E. (2012). Enhancing the Curation of Botanical Data Using Text Analysis Tools. In Zaphiris, P., Buchanan, G., Rasmussen, E., & Loizides, F. (Eds.), Theory and Practice of Digital Libraries. (pp. 480-485). (Lecture Notes in Computer Science). Springer Berlin / Heidelberg. doi: 10.1007/978-3-642-33290-6_55

Enhancing the Curation of Botanical Data Using Text Analysis Tools

Clare Llewellyn[1], Clare Grover[1], Jon Oberlander[1], and Elspeth Haston[2]

[1] University of Edinburgh, Edinburgh, United Kingdom
[2] Royal Botanic Gardens Edinburgh, Edinburgh, United Kingdom
C.A.Llewellyn@sms.ed.ac.uk,
{Grover, Jon}@inf.ed.ac.uk,
E.Haston@rbge.ac.uk

Abstract. Automatic text analysis tools have significant potential to improve the productivity of those who organise large collections of data. However, to be effective, they have to be both technically efficient and provide a productive interaction with the user. Geographic referencing of historical botanical data is difficult, time consuming and relies heavily on the expertise of the curators. Botanical specimens that have poor quality labelling are often disregarded and the information is lost. This work highlights how the use of automated analysis methods can be used to assist in the curation of a botanical specimen library. A prototype tool has been built which allows users to interact with automatically generated geographical locations in order to identify where a botanical plant specimen was collected. An evaluation study provides a comparison of this tool with the traditional methods currently used.

Keywords: text analysis, text mining, geographical location, assisted curation, botany

1 Introduction

The aim of this work is to improve the interaction between users and automatic text analysis tools within a practical context. A tool has been created that allows users to curate botanical data by adding geographical locations. This is achieved via the user interacting with automatically generated locations extracted from information about botanical specimens. The user can correct or make additions to this generated output in order to specify the exact location where the botanical sample was collected.

To pursue the aim of creating a productive usable interface for textual analysis tools, a specific interface for such a tool has been created. The tool and interface are both applicable to many uses, but in order to evaluate it in detail, the focus is botanical science specimen data. The user group that this tool is intended for is staff the Royal Botanic Gardens, Edinburgh (RBGE). The data curation experts at the RBGE expressed a need for the integration of a tool that extracts geographical locations from plant specimen data records into their current work flow. The demand for such a tool gives the opportunity to engage the likely users of such a tool in evaluating the interface, therefore producing valid usability results. The specific objectives of this work were to identify where text analysis tools can be used in the botanical curation workflow, to identify appropriate geographical text analysis tools, to design and implement a prototype tool and to evaluate if this tool improves the ease, speed or accuracy of botanical curation.

2 Related Work

The Royal Botanic Garden of Edinburgh an internationally renowned centre of excellence for plant biodiversity research. The herbaria houses nearly three million specimens representing half to two thirds of the world's flora. It holds specimens collected from across the world and they continue to receive approximately ten thousand new specimens each year [14].

Plant specimens are labelled with data that is relevant from their collection, a description of the plant, the name of the collector, the date it was collected, the habitat conditions it was found in and the geographic location it was collected from. Geographic referencing in this domain means converting the textual descriptions of where a plant was collected into machine readable geographic locations generally using a map based coordinate system. This is either done at the time when the plant is collected by GPS systems or retrofitted from textual descriptions[9]. Historically, locations on plant specimens have been vague. Identifying and correcting plant specimens records that contain errors is time consuming and expensive for curators. Geographic referencing has been described as 'the most significant bottle neck in the digitisation process' [9], therefore improvements to the speed and the accuracy of this process would be valuable.

Currently, geographic referencing is conducted using tools such as gazetteers and maps to find the coordinates (latitude and longitude) of the place names that have been identified in the plant specimen records by the curators. Generally several of these tools must be used in coordination in order to obtain an accurate location. This process relies heavily on the expertise of the curator both in using these tools and on the vast knowledge of likely plant locations that they possess. Geographic referencing takes a long time, requires expertise and access to resources. On line tools have been created to assist in this process but they do not fit well into the current work flow. Once the data has been geographically located it allows a botanical scientist to study environmental changes, particularly those which study human impact and climate change.

Locations can be automatically generated through content analysis, natural language processing and text mining. Widespread use of text analysis has not yet been achieved. For example in geographic referencing it is believed that the main barrier to uptake is that accuracy levels usually fall short of the expectations and needs of the user. It is proposed that this problem is rectifiable through the provision of interface extensions to existing text analysis tools to allow the user to correct and enhance automatically created output, thereby combining the efficiency of automatic processing with the accuracy of manual annotation [1].

Metrics for text analysis evaluation currently focus on comparisons with other text mining systems rather than evaluating the usefulness of the tool within a domain [1]. Within the biomedical domain it has been shown that although there is an indication that text mining is useful in curation, there is a question on the actual effectiveness as tools often have low accuracy scores due to the complexity of natural language [1,10,13]. Manual curation is more accurate than automated curation but this can take a much longer [13]. A study by Alex et al. in 2008 [1] was conducted to evaluate the usability of a curation tool that assisted in hypothesising protein-protein interactions through an analysis of textual information. They found that the speed of curation can be increased by a third by assistance of text mining tools.

3 Prototype Tool

3.1 Data

Plant specimens have labels describing the collection details of that specimen. The text from these labels are stored in a database, for an example record see table 1. The conversion of the label to a record is a manual process performed by the curators. This study focused on records from the United Kingdom and Ireland from 1747 to 2010. The total number of records processed was 43,060.

Table 1. Example Record

Field Name	Field Information
Specimen Number	279507
Barcode	E00000425
Current Name	Riccardia chamedryfolia (With.) Grolle
Family	ANEURACEAE
Collection Year	2008
Collector and Number	Chamberlain, David F.; Kungu, Elizabeth M.
Collector	Chamberlain, David F.; Kungu, Elizabeth M.
Collection Date	6 Feb 2008
Country Where Collected	United Kingdom
Locality	Above Fife Coastal Path, near Baddo Rock, NNW of Boarhills
National Grid Reference	NO 561 152
Habitat	North-facing sandstone cliff
Region	1A
Sub Country 3	(VC 85) Fife & Kinro
Sub Country 1	Scotland

Several fields in the record contain geographic information that can be used to determine location - either a recognised location system such as latitude and longitude coordinates, a National Grid Reference, or textual description of the location. Of the records used 881, contain a latitude value, 879 contain a longitude value and 10,537 contain a National Grid Reference, 32,523 had no location code. The National Grid Reference can be converted to latitude and longitude values, although this can lead to inaccuracies.

For the text mining there are several fields in the database record that could be used, those describing locality (specific village), sub country (several fields describing towns, cities, counties and various other levels of location) and habitat (a very precise local location). It was found that, in total, 63.82% of records had some degree of geographical information, and could be geolocated. Other records did not contain any geographic information and it would not be possible to provide locations for these.

3.2 Text Analysis Tools

The data from the database was processed using the Edinburgh Informatics information extraction tools which include LT-TTT2 and the Edinburgh Geoparser [6,7,11,12]. These are well established tools that process text and XML to identify place names and provide geographic coordinates for the locations.

The Geoparser is made up of two main components; the Geotagger which provides place name recognition (identifies text strings as places) and the Georesolver which provides geographic referencing (looks up the names in a geographic gazetteer) [6]. The named entity recognition tool identifies word sequences as place-name entities and marks them up as XML elements. After initial tokenisation and part-of-speech tagging, it uses a rule-based method that takes into account information about part-of-speech, capitalisation, local context and lexicon look-up. The place-name entities recognised by this method are converted to gazetteer queries which are submitted to one of the Unlock or GeoNames gazetteer services. Unlock is sourced from the Ordnance Survey and covers Great Britain while GeoNames has worldwide coverage. As this project is restricted to the UK, Unlock was the service used. The gazetteer service typically returns several entries per place-name and the Georesolver component of the Edinburgh Geoparser ranks these in order of likeliness [6,7,12].

For each record the fields that contained textual location information were converted to an XML file which was processed by the Edinburgh Geoparser provide suggested locations. The complete XML file was used to rank the locations produced so if several terms are found which provide locations that are geographically close together this will improve the ranking of these locations. To assist users in finding an accurate location the fields which provided the more granular locations were presented first, for example a locality is presented before a sub-country.

A goal of the work is to provide a tool that allows users to correct and enhance automatically created output. Therefore when investigating whether automatic text analysis tools can be effective within a practical context creating a productive interaction with the user was essential. An extensive requirements analysis was conducted to ensure a strong user centred approach in the design of the system and integration of the geographic location tools within the botanical workflow.

3.3 Botanical Workflow

Users were involved extensively throughout the creation of these tools, from the design of initial mock ups to the tool itself, and features were adapted to meet the needs of the user.

The requirements analysis started with fairly informal discussions with the curators and researchers. These discussions centered around the current curation processes and the methods for storing and accessing data. Specifically, curators and researchers were asked about the systems that they used and the identification possible insertion points for the geographic location tools was made. Subsequent interviews were more structured in order to determine the precise role of the tool and to ensure that different curators and curation processes could be supported. These interviews looked in more detail at the systems used; one-on-one interviews with several curators were conducted and one participant provided a cognitive walk through of the current process. Workflow information was taken from the documentation of the current process.

3.4 The Tool

The data curation experts requested an automatic tool that extract geographical locations from plant specimen data records and could be integrated into their current work flow. From the analysis of their work processes it was found that the users wished to interact with the data in several ways. Therefore, a tool was created that allowed the users to:

1. Search or browse for a set of records or a specific record
2. Sort the records on several different fields (especially the sub country fields)
3. Visualise the suggested geographic locations both in a map form and a list form
4. Use maps from multiple sources to define a precise location
5. Refer back to the original record
6. Visualise the specific itinerary of a collector (other locations the collector has collected from in a similar date range)

The system included a back-end mySQL database (which holds multiple tables) - mySQL was chosen as this is the database technology used at the RBGE and mirroring this would provide easier interaction between the systems. The data was initially processed through text mining, database matching of similar fields and a National Grid Reference conversion to latitude and longitude. The information produced from this processing was stored in the database.

The system is web based and the users interact with it through a webserver to query the database. The interface provides two views of the result, as a list of locations and as points on a map. The maps used are accessed through APIs - Google Maps and the National Library of Scotland's Ordinance Survey Maps. The results for each query are built 'on the fly' using XSLT to build the specific HTML and JavaScript for each record. The results interface contains:

1. Details from the original record
2. Various layers of maps
3. The locations (from text mining and National Grid conversions) are presented as a list of location names and longitude latitude coordinates and also as pins on a map. Clicking the location name will focus the map on that point
4. A pin which the users may place on the map to select a location for the specimen. They are also able to manually enter a value.
5. A facility to add a comment into the system describing why the user has picked a specific location

During the design phase it was noted that attempts were being made to present a lot of information in a small space especially on the Map page. Therefore, there is extensive use of JavaScript so items can be hidden and revealed when desired.

4 Evaluation

In order to test the hypothesis, whether a textual analysis tool can be used to improve, increase the speed or accuracy of the workflow of curators who are archiving plant specimen data an evaluation study was conducted. The current manual curation process was compared against a tool with textual analysis support. The evaluation was conducted in a manner adapted from a digital library evaluation framework [5].

4.1 Approach

Human computer interaction (HCI) within digital libraries has been studied extensively for the past ten years [8]. The tool was evaluated to ensure that it observed the basic HCI principles of a digital library such as obtaining correct results to a query quickly (precision and recall)[4]. It is important for the user to receive a manageable number of results so that they can see what the general content will be. While performance indicators such as number of key strokes to reach a certain result are important, it does not reveal the true usability of the system [4]. Furh et al (2007)[5] provide an extensive framework for the evaluation of digital libraries. They offer a set of guidelines for digital library evaluation which are adaptable and can be used for this task. They suggest focusing on usability, usefulness (or relevance) and performance, therefore these were the areas evaluated in this work.

The evaluation was conducted with ten participants all of whom work at the RBGE. Eight of these participants were specimen digitisers and two were botanists. Each participant had experience in geographically locating specimens and all had used the computer and the internet on at least a weekly basis.

Each participant was asked to geographically locate the collection sites of eight specimens. Four specimens by using the tool created in this project and four by using any other method they desired (henceforth referred to as the 'traditional method'). They were provided with a computer with internet access, several maps of the UK, a UK atlas in book form and a UK gazetteer also in book form. For each participant several tabs were opened in an internet browser giving access to Google maps, Ordnance Survey Maps, The British Place Name Gazetteer and Street Map. The participants were instructed that they could use any resource available on the web. For the tool, the users were provided with the bar code of the specimen that they could type in. For the traditional method, the users were provided with print outs from the database of the information recorded for each specimen.

The participants were paired. One from each pair geographically located four samples using the tool and the other using the traditional method for those same samples. The order was reversed for the pair so that one used the tool first then the traditional method and vice versa, so times would be comparable. It was decided as there were two botanists and eight digitisers, the botanists would be paired together.

4.2 Tasks

Each participant was asked to perform eight tasks. The evaluator made notes throughout the task to determine if the task was completed successfully and to count the number of missteps made in the task. Success was defined as when the user thinks they have successfully located the collection point, they were asked to state aloud when this had occurred so the time taken could be recorded.

The data used for evaluation was data with known locations (a random sample from the 881 RBGE records that contained a latitude and longitude values). This was then used to provide an accuracy for each task. During the process it was noticed that the data in some of the records contained errors. For example, when mapped, the longitude and latitude values located the specimens in the sea. The specimens with obvious mistakes were removed from the evaluation, but it is worth noting that in the records remaining there is no guarantee that the data is completely accurate.

A post task interview was used to provide qualitative information on the participants opinion of the system.

4.3 Results

Usability is measured by looking at the effectiveness, adaptability, enjoyability and learnability of the tool.Effectiveness was determined by whether the system could provide information effectively - it was measured by how many tasks could be completed [4,5]. The results are presented in table 2, they suggest that the tool performs slightly better than the traditional method.

Whether the system is adaptable was measured by how the users rated the different features and whether they could adapt the experience to their own preferences. The type of question used to determine this included; Could you determine the origin of the locations? Could you easily find the features? Did the features work the way you expected them to? Did you find the map easy to use? Comments on these features suggested that they were generally well liked. Participants commented that the different maps were used in different ways and this was not liked, they thought the map use would get better with time. They liked the different colours of the pins, they thought that this made it easy to use. Participants also had a preference dependent on what maps they had used before. There were two participants who would have preferred a larger map. The Ordnance Survey map, although slightly harder to use, was well liked and primarily this was the map feature used to identify the location.

Enjoyability and learnability were measured through satisfaction scores of ease of use, visual appearance, contents, structure, error corrections and usefulness of help information. Questions such as; Was it easy to find the locations? When you made mistakes could you correct them? Did you enjoy using the tool? Is the tool nice to look at? Is the tool clear and easy to navigate? The tool scored highly in this category, the users liked the tool and found it easy to use. Participants found the tool easier to use than the traditional method.

Table 2. Number of Locations Identified

Participant Number	Traditional Method	Tool
1	4	4
2	4	4
3	4	3
4	4	4
5	4	4
6	4	4
7	3	4
8	4	4
9	3	4
10	0	2
Total	34	37

Usefulness is evaluated through the relevance of provided content. Does the content assist with the task defined in a satisfactory way [4,5]? The type of question asked were; Did the site help you geographically locate the specimen? Was the map feature useful? Were the text mining suggestions useful? Was the information presented in a way that was useful? Also investigated was whether the content provided lead to participants accurately locating the samples.

The text mining suggestions were not considered completely accurate, as many false positives were returned in order to include as many true positives as possible, but the text mining suggestions were still considered helpful. The users were willing to tolerate a degree of inaccuracy in the suggestions. The maps were considered very useful.

Accuracy was compared between the tool and the tradition method. Initially it was found that there was no significant difference between the accuracy of the two systems. However, there is a significant positive correlation between the tasks showing that if a task was difficult to perform with the tool it also was with the traditional method. Observations suggest that this may be because if a user knows an area they will find it easy to locate a specimen in that area. Further work would be needed to establish if there is a link between knowledge of an area and accuracy in geo-locating in that area.

In order to look more closely at the accuracy achieved using the tool a further experiment was conducted. All participants in the test were asked to geo-locate all of the samples used in the initial evaluation using the tool. These locations were then clustered and the location which was the furthest from the others was left out, as was any location more than 25km away from the average point. Using an average location point from those left a significant increase in accuracy was found when using the tool. The results show that the tool in this case is, statistically significantly, more accurate than the traditional methods ($p= 0.012$, $t=-2.742$, $df= 23$). Thus the tasks may be difficult for specific individuals but when an average is taken over the whole group the result will be accurate. A number of participants did not complete all the tasks using the traditional method so it is unknown if an increase in accuracy would occur here as well. The resources needed for the traditional method are not widely available, but as the tool is available on the web and it is easy to use, people may be more likely to use it. This could provide an interesting future direction for this work, developing a crowd sourced facility.

Performance and efficiency of the tool was evaluated by assessing the efficient retrieval of information. It was measured by how much time it took to correctly complete tasks. [4,5]. Each task was timed for each participant. In addition participants were asked to rate performance for each task.

The performance, which was judged by speed of task completion, was better on average than with the tool (see table 3). A paired sample t-test shows that the difference is not significant ($p=0.539$, $t=-0.639$ $df=9$). However, there is a positive correlation; those who did well in the traditional tasks did well in the tool tasks but this is not significant (correlation = 0.583, $p = 0.077$). This was particularly true of the two botanists in the study they adopted a different approach to the other members of the group because of their experience in plant and habitat identification. They were very fast both with and without the tool but noticeably faster without the tool. It could therefore be proposed that the tool is of more use to a less expert user. A further study with more participants and tasks may be able to determine if there is a statistically significant difference.

Table 3. Average Speed for Tasks (in seconds)

	Minimum Time	Maximum Time	Mean Time	Standard Deviation
Tool	139.25	263.75	190.65	42.47748
Traditional	87.75	300.00	202.93	74.7781

A further experiment was conducted to investigate if the number of text mining locations offered had an effect on the time taken to complete the task to see if there was an optimal number of locations. Initially the total number of locations provided to the user was considered, every single latitude and longitude pair for every place name. Analysis indicates that it is possible that there is a positive effect of either offering very few or very many suggestions (below 2 and above 5). It could be postulated that when the number is small the user trusts those suggestions and when the number is high the good suggestions outweigh the bad. If the user is offered 3 or 4 suggestions they may become confused. Further investigation with a larger sample size would need to be conducted to find out if this is the case.

The total number of locations offered was contrasted with the number of unique locations. The tool often suggests a number of individual latitude and longitude locations for a single place name (as many location names are reused). A unique location is classified as a single place name (no matter how many suggestions are offered for that name). It was found that 2 unique locations may be beneficial,possibly because they are used to provide confirmation, especially if the 2 locations are close to each other. With 1 location the task may take longer, as the user may need to consult other features. With more than 3 locations the task may take longer, as the locations may be contradictory. Further study would need to be undertaken in this area to ascertain if this is the case.

5 Conclusions

The specific objectives of this work were to identify where text analysis tools can be used in the botanical curation workflow, to design and implement a prototype tool and to evaluate if this tool improves the ease, speed or accuracy of botanical curation.

A tool was created that allowed users to interact with automatically generated geographical locations in order to correct or make additions to the output. As requested, by the data curation experts the tool has been integrated into the current workflow.

An evaluation was conducted through the comparison between traditional curation of geographical locations and assisted curation. In this evaluation, using the tool, has been shown that average speeds are quicker, but not significantly so. Further evaluation studies may be able to statistically prove this in the future with a larger study.

The tool scored highly for both usability and usefulness. The participants in the evaluation liked the tool and found it easy to use - they preferred it to the traditional method of using multiple data sources.

The accuracy of the geographic location was compared between the tool and the traditional method. Initially, it was found that there was no significant difference in the accuracy of the two systems. There is a significant positive correlation between the tasks showing that if a task was difficult this was independent of the method used.

When many participants used the tool to identify a location and values are clustered, leaving out the least similar location, a significant increase in accuracy is found. This shows that while the tasks may be difficult for individuals higher accuracy can be gained from using locations from a number of individuals, they will collectively locate the specimen accurately.This suggests that this is an ideal tool for use with crowd sourcing this task. Investigating if this is viable would be an interesting area for future work.

The analysis suggests that there is an ideal number of total text mining suggestions and unique text mining locations that reduces the burden on the user and leads to more efficient geographic location.

Certain participants were better a geo-location than others. Results indicate that other less experienced users performed better using the tool. The higher the participant knowledge was of botany and the area the botanical specimen came from, the quicker and more accurate they were. A useful future direction for the work would be to identify strategies for enabling experts to do this work, for example, asking a local gardening group to locate specimens from their local area.

References

1. Alex, Beatrice, Claire Grover, Barry Haddow, Mijail Kabadjov, Ewan Klein, Michael Matthews, Stuart Roebuck, Richard Tobin, and Xinglong Wang. 2008. Assisted curation: does text mining really help? Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing: 556-567
2. Allen, William H. 1993. The Rise of the Botanical Database. BioScience 43 (5) (May 1): 274-279. doi:10.2307/1312059
3. Auer, Soren, Jens Lehmann, and Sebastian Hellman. 2008. Linked GeoData–Adding A Spatial Dimension to the Web of Data. In Proceedings of 8th International Semantic Web Conference (ISWC 2009). http://www.informatik.uni-leipzig.de/ auer/publication/linkedgeodata.pdf.
4. Blandford, A., and G. Buchanan. 2003. Usability of digital libraries: a source of creative tensions with technical developments. Article. http://eprints.ucl.ac.uk/16648/.

5. Fuhr, Norbert, Giannis Tsakonas, Trond Aalberg, Maristella Agosti, Preben Hansen, Sarantos Kapidakis, Claus-Peter Klas, et al. 2007. Evaluation of digital libraries. International Journal on Digital Libraries 8 (1): 21-38. doi:10.1007/s00799-007-0011-z.
6. Grover, Claire, Sharon Givon, Richard Tobin, and Julian Ball. Named Entity Recognition for Digitised Historical Texts. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.166.68.
7. Grover, Claire, and Richard Tobin. Rule-Based Chunking and Reusability. https://www.inf.ed.ac.uk/publications/online/0801.pdf.
8. Jeng, J. (2005) Information technology and libraries, vol 24; numb 2, pages 46-56, What Is Usability in the Context of the Digital Library and How Can It Be Measured?
9. Johnson, Norman F. 2007. Biodiversity Informatics. Annual Review of Entomology 52 (1): 421-438. doi:10.1146/annurev.ento.52.110405.091259.
10. Rebholz-Schuhmann, Dietrich, Harald Kirsch, and Francisco Couto. 2005. Facts from TextIs Text Mining Ready to Deliver? PLoS Biol 3 (2) (February 15): e65. doi:10.1371/journal.pbio.0030065.
11. Tamames, Javier, and Victor de Lorenzo. 2010. EnvMine: A text-mining system for the automatic extraction of contextual information. BMC Bioinformatics 11 (1): 294. doi:10.1186/1471-2105-11-294.
12. Tobin, Richard, Claire Grover, Kate Byrne, James Reid, and Jo Walsh. 2010. Evaluation of georeferencing. In Proceedings of the 6th Workshop on Geographic Information Retrieval, 7:17:8. GIR 10. New York, NY, USA: ACM. doi:10.1145/1722080.1722089.
13. Winnenburg, Rainer, Thomas Wchter, Conrad Plake, Andreas Doms, and Michael Schroeder. 2008. Facts from text: can text mining help to scale-up high-quality manual curation of gene products with ontologies? Briefings in Bioinformatics 9 (6) (November 1): 466 -478. doi:10.1093/bib/bbn043.
14. Edinburgh Royal Botantic Gardens website accessed 29th March 2012 http://www.rbge.org.uk/