Automated Binary based Tree Species Identification

Ugege Peter E.
Computer Unit,
Dept of Planning, Res., Statistics & Biometrics
Forestry Research Institute of Nigeria
Ibadan, Nigeria

Ugbogu Omokafe A.
Taxonomy Section,
Dept of Forest Conservation & Protection
Forestry Research Institute of Nigeria
Ibadan, Nigeria

ABSTRACT

Although automated species identification for many reasons is not yet widely employed, efforts towards the development of automated species identification systems within the last decade is extremely encouraging; that such an approach has the potential to make valuable contribution towards reducing the burden of routine identification. In this work, we developed a system that uses binary numbers generated from the morphological characters of trees to uniquely identify all Nigerian tree species.

General Terms

Computer Assited taxonomy, Automated species identification

Keywords

Automated; tree species; identification; morphological; binary

1. INTRODUCTION

There are many factors influencing the taxonomic impediment to the study of biodiversity. A major one being that the demand for routine identification in biodiversity studies extends beyond the available resources. In many spheres the volumes of plant or animal specimens that can usefully be obtained, particularly using modern sampling methods, vastly outstrip any capacity to identify this material [1]. This has limited the progress in some aspect of biodiversity research [2].

These demands are likely to steadily increase as the proportion of previously un-described species in local, national or regional floras and fauna declines and as requirement or desirability of biodiversity inventories and other such survey grows [1]. This has led to several solutions being proffered to reduce the burden of routine identification. One of the proffered solutions is automating the identification process in some way [3]. This is generically referred to as Computer Assisted Taxonomy (CAT) [4]. However, the development and application of an automated approach to taxonomic identification has remained a minority interest till date. Among reasons for this are the notions that it is too difficult, too threatening, too different or too costly [1]. It is most encouraging to know that despite these limitations, efforts towards the development of automated species identification systems have been progressive.

In this work, we further buttress the present minority notion that the automation of species identification process is possible and achievable. We have developed a system that uses binary codes generated based on the morphological characters of trees to uniquely identify them.

Though this is not the first time an attempt is made to automate species identification using their morphological characters, our approach is far simpler and less expensive to implement. For instance while previous approaches are centered round the need for a computerized pattern recognition system [5 – 10], ours does not require such. We were able to easily prove the effectiveness of the system by restricting our study to the over one thousand Nigerian Trees species. All a user need is a functional computer system, a ruler and personal ability to supply answers to the questions asked by the system and the tree identification process is completed.

2. MOTIVATION

The task of identifying plant species with sole reliance on dead specimens that have been pressed, dried and/or pricked is becoming more and more tasking by the day and the pressure constantly being mounted on the fragile specimens have continued to increase at times resulting in some of these age long conserved specimen being destroyed. Also in-situ identification in most cases is not that easy; meaning materials for identification may have to be transported to the herbarium before identification can be done. And then there is the fact that a particular identification process may take longer than expected since the task of finding a match among the thousands of specimens in the herbarium cannot be described to be an easy one. Automating the identification process will in no small measure contribute positively to solving these problems.

3. METHOD

Training sets were first generated by carefully considering and selecting the sixteen most distinctive characters of tree species that can most easily be used to uniquely identify them. This formed the bases for generating binary codes on which the identification is done. Each species had a unique code attached to it which is determined by the presence or absence of the selected sixteen morphological characters. The presence of a character is assigned a "1" and the absence is assigned a "0" leading to a sixteen bit binary code being generated. In cases in which two or more species have the same binary code generated for their identification, subsidiary binary codes are generated using other morphological characters not previously considered for selection in the first instance to uniquely identify them. The length of the subsidiary codes ranges from 3 to 4 bits as the case may be. Works by [11] contributed immensely to our being able to draw up the training sets.

3.1 Morphological characters used

The presence or absence of the following morphological characters determines the binary code generated for each species.

1. Spine
2. Latex
3. Other exudates

4. Simple leaves
5. 1-pinnate leaves
6. Opposite leaves or leaflet
7. Leaves or leaflets widest above middle
8. Leaves or leaflets widest below middle
9. Blade of leaf that is very narrow
10. Blade of leaf that is very broad
11. leaf or leaflet whose margin is entire
12. leaves or leaflets whose blades are over 30 cm
13. leaves or leaflets whose blades are under 2.5 cm
14. leaves or leaflets whose stalks are over 5 cm
15. leaves or leaflets whose stalks are under 1.75 cm
16. palmately-nerved

An example of morphological characters used to generate subsidiary codes where necessary could be the following:
1. Some leaves with more than 15 lateral nerves on each side of the midrib
2. Most leaves with not more than 6 lateral nerves on each side of the midrib
3. Stipules present

4. THE IDENTIFICATION PROCESS

A user supplies answers in Yes/No format to questions which the system have been designed to ask based on the selected morphological characters on which the training sets have been created Fig. 1 (a). The system then generates a binary code by converting every "Yes" answer to a 1 and every "No" answer to a 0. The length of this code is usually 16 representing the 16 morphological characters being used Fig. 1 (b). The system searches its inference-rules-base for a satisfactory rule using the binary code generated, if found, the species linked to that code is returned as the name of the species to be identified Fig. 1 (c). The system does not stop searching immediately a match rule is found. Rather all rules are considered before returning a result. This way it is able to detect if there are more than one match rule for a particular generated binary code in which case the system engages the user in another round of questioning based on other sets of morphological characters to generate subsidiary binary codes that will be used uniquely assign the rules.

5. THE SYSTEM OVERVIEW

The system components include a GUI, a search tool, a binary code generator, an inference-rules-base, a database and a query tool Figure 2. The GUI is the means through which a user interacts with the system. Identification, getting information on already known species and updating of the system information base are done here. The update of the system information base is however restricted to the system administrator only. For the identification of a tree species, the binary code generator uses the answers supplied by a user to generate a binary code which is passed to the search tool. The search tool uses this code for identifying the species based on combinations of rules in the inference-rules-base specified by the code. If the identification is successful, the result is passed straight to the user while information about the identified species could be extracted from the database by the query tool at the request of the user. To get information on already known species, the query tool is responsible for extracting the needed information from the database and sending it to the user through the GUI.

Spine	Yes
Latex	No
Other exudates	Yes
Simple leaves	No
1-pinnate leaves	No
Opposite leaves or leaflet	Yes
Leaves or leaflets widest above middle	No
Leaves or leaflets widest below middle	No
Blade of leaf that is very narrow	No
Blade of leaf that is very broad	Yes
leaf or leaflet whose margin is entire	Yes
leaves or leaflets whose blades are over 30 cm	No
leaves or leaflets whose blades are under 2.5 cm	Yes
leaves or leaflets whose stalks are over 5 cm	No
leaves or leaflets whose stalks are under 1.75 cm	Yes
palmately-nerved	Yes

(a)

1	0	1	0	0	1	0	0	0	1	1	0	1	0	1	1

(b)

Rinoria oblongifolia	0	0	0	1	0	0	1	0	0	0	0	1	0	1	0	0
Rinorea longisepala	0	0	0	1	0	0	1	0	0	0	0	1	0	0	1	0
Dichrostachys cinerea	1	0	0	0	0	1	0	0	0	0	1	0	1	0	1	0
Strychnos innocua	1	0	0	1	0	1	0	0	0	1	1	0	0	0	1	1
Acacia gourmaensis	**1**	**0**	**1**	**0**	**0**	**1**	**0**	**0**	**0**	**1**	**1**	**0**	**1**	**0**	**1**	**1**
Mayetenus senegalensis	1	0	0	1	0	0	1	0	0	1	1	0	0	0	1	0
Bridelia ferruginea	1	0	0	1	0	0	0	0	0	1	1	0	0	0	1	0
Bombax costatum	1	0	0	0	0	1	1	0	0	0	0	1	0	0	1	0

(c).

Figure 1: The automated tree species identification process. (a) depicts answers supplied by a user in Yes/No format. (b) binary code generated based on answers in (a). (c) depicts a search for a match by the system using the binary code in (b), finding the match and eventually identifying the species to be *Acacia gourmaensis*.

DISCUSSION

The system framework with its components and interfaces has been designed taking cognizance of the fact that targeted users are not at the moment very vast in the use computers. As such users are presented with very simple and easy to use interfaces such that even a beginner to the use of computers can easily use it. Beside the answers supplied by users are in simple "yes" and "no" format. Only in some other cases depending on the action they want to perform with the system are they asked for the scientific name of the species of which they are already conversant with. The system is also designed to allow for the identification of newly discovered species simply by updating its knowledge base.

The system which is currently being test run at Forest Herbarium Ibadan (FHI) in the Taxonomy Session of Forestry Research Institute of Nigeria can presently identify over 70% of the entire Nigeria tree species based on the information currently in its information base. However its framework has been designed to enable it achieve 100% identification of Nigerian tree species once its information base is adequately updated. It incorporates the search of information on already known species and has the ability to display needed information on identified species inclusive of pictures.

One major advantage of the system is the fact that it is helping to preserve the fragile plant specimens in the herbarium as they are no longer often referenced for information and identification.

6. CONCLUSION

In this work, we have tried to further prove that automated species identification is possible. Although building a single system that would replace the routine identification of all biological species is unthinkable at the moment, a lot could be achieved in this direction by trying to automate this process for sub-domains and sub-sub-domains, and so on down the line within the entire biological domain. We focused solely on Nigerian trees species. The routine process for identifying tree species was described to be tedious and requiring much efforts and experience [12]. Our system apart from buttressing the call for the automation of species identification process has also simplified this identification process. However let us conclude by saying that the identification of individuals of described species will continue to depend heavily on routine identification based on dead specimens that have variously been pressed, dried and/or pricked as they are the materials available and suitable to work with by systematists/taxonomists who are currently saddled with the responsibility of species identification. This would only change if we continue with our efforts towards research in the automation of the existing routine identification process.
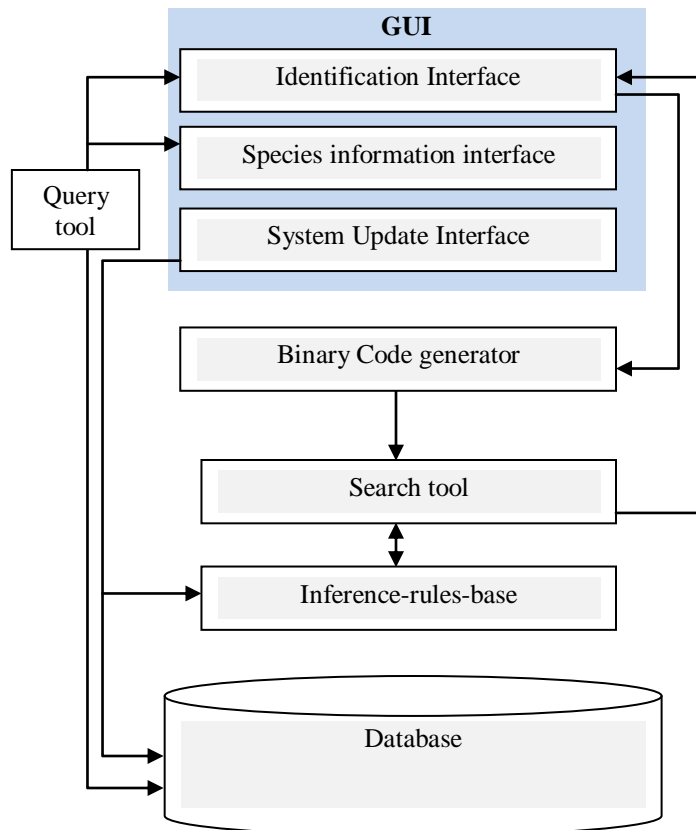
7. ACKNOWLEDGEMENT

Figure 2: The System Architecture

8. REFERENCES

[1] Gaston K. J. and O'Neill M. A. 2004 Automated species identification: why not?, Phil. Trans. Roy. Soc. London, vol. B359 (March 2004) pp. 655–667

[2] France I., Duller A. W. G., Duller G. A. T. , and Lamb H. F. 2000 A new approach to automated pollen analysis *Quatern. Sci. Rev.* vol. 19, 537–546

[3] Gauld I. D., O'Neill M. A. and Gaston K. J. 2000 Driving Miss Daisy: the performance of an automated insect identification system," in Hymenoptera:evolution, biodiversity and biological control, A. D. Austin and M. Dowton, Eds. Collingwood, VIC: CSIRO, pp. 303–312.

[4] Chesmore D. 2000 Methodologies for automating the identification of species, Proc. Inaugral Meeting of the BioNET-International. Group for Computer-Aided Taxonomy (BIGGAT), D. Chesmore, L. Yorke, P. Bridge and S. Gallagher, Egham: BioNET-International Technical Secretariat, (2000), pp. 3–12 .

[5] Banarse D.S., France I. and Duller A. W. G. 2000 Analysis and application of a self-organising image recognition neural network, Adv. Engng Software, vol. 31, pp 937–944.

[6]   Wu J. and Zhou Z. H. 2002 Face recognition with one training per person. Pattern Recognistion Lett., vol. 23, pp 1711 – 1719,

[7]   Gou B., Lam K. M., Lin K. H., and Siu W. C., 2003 Human face recognition based on spartially weighted Hausdorff distance, Pattern Recognistion Lett., vol. 24, pp 499 –507

[8]   He Y., Tian J., Luo X., and Zhang T., 2003 Image enhancement and minutiae matching in fingerprint verification. Pattern Recognistion Lett., vol. 24, pp 1349 – 1360

[9]   Lu G., Zhang D. and Wang K., 2003 Palm print recognition using eigenpalms features. Pattern Recognistion Lett., vol. 24, pp 1463–1467.

[10]   Tsalakanidou F., Tzovaras D., and Strintzis M. G. 2003 Use of depth and colour eigenfaces for face recognition. Pattern Recognistion Lett., vol. 24, pp 1427–1435

[11]   Keay R. W. J., Onochie C. F. A., and Stanfield D. P.1964 Nigerian Trees Vol. II. Nigerian National Press Ltd., Apapa.

[12]   Ugege P. E. Ugbogu O. A. Jayeola, O. A. Halidu S. K, and Oyeleke O. O. 2007 Computer Aided Tree Species identification System. Journal of Forestry Research and Management, vol 4., pp 18-24