International Journal of Experimental Research and Review (IJERR)

ISSN: 2455-4855 (Online) ***Original Article***

Received: 20th October, 2019; **Accepted:** 22th November, 2019; **Published:** 30th December, 2019

DOI: https://doi.org/10.52756/ijerr.2019.v20.004

An overview on Taxonomic Keys and Automated Species Identification (ASI)

Mitu De[1] and Santi Ranjan Dey[2*]

[1]Department of Botany, Gurudas College, Kolkata 700054, West Bengal, India; [2]Department of Zoology, Rammohan College, Kolkata-700009, West Bengal, India

[*]**Corresponding author:** srdey1@rediffmail.com

Abstract

A core goal of taxonomy and systematics, entomology, field botany, horticulture, zoology and many agriculture courses involves learning to identify plants and animals. But current syllabi and time constraint allow students to see only a limited amount of taxonomic variability. Usually only experts such as taxonomists and trained technicians can identify taxa accurately because it requires special skills acquired through extensive experience. Taxonomic keys are essential tools for species identification, used by students and professionals. The development of computer-based, multi-media keys provides one means of addressing this critical identification and diagnostic function. Automated species identification (ASI) is a method of making the expertise of taxonomists available to ecologists, parataxonomists and others via digital technology and artificial intelligence. Today, most automated identification systems rely on images depicting the species for the identification. Although computer programs will not replace classical plant identification methods, they have the potential to make these methods more effective.

Keywords: ASI, computer based tools, identification, taxonomic keys.

Introduction

Species identification is not exclusively the job of taxonomists and ecologists. It is required or useful for large parts of society, from professionals (such as landscape architects, foresters, farmers, conservationists, and biologists) to the general public (like eco-tourists, hikers, and nature lovers). The identification of biological specimens remains central to the study of ecology and our understanding of the complex roles of biodiversity in the functioning of the earth's biosphere. Without the ability to identify, and differentiate between, the taxonomic units of the biosphere it will be difficult to understand the ecosystem services on which the entire mankind depends.

Species identification is traditionally based on taxonomic studies which were primarily based on morphological studies (Martineau et al., 2017). Usually only experts such as taxonomists and trained technicians can identify taxa accurately, because it requires special skills acquired through extensive experience. But the identification of flora and fauna by conventional means is difficult, time consuming, and (due to the use of specific botanical terms) frustrating for novices. The world-wide decline in taxonomic expertise and services has meant that many identification services have either become unavailable or prohibitively expensive, especially in developing countries. The declining and partly nonexistent taxonomic knowledge within the general public has been termed "taxonomic crisis" by B. Dayrat in 2005 in a publication in the 'The Biological Journal of the Linnean Society' which is a direct descendant of one of the oldest biological journal in the world, the Transactions of the Linnean Society.

The so-called 'taxonomic impediment' recognises a bottleneck in the development of our understanding of the biosphere resulting from a shortage of skills in taxonomy and systematics (Reaka-Kudla et al., 1996). In some cases a trained technician could make routine identifications using morphological "keys" (step-by-step instructions of what to look for), but in most cases an experienced professional taxonomist is needed (Martineau et al., 2017).

Taxonomic keys are essential tools for species identification, used by students and professionals. Griffing (2011) showed that the development of traditional paper-based keys started at least 320 years ago when, in 1689, Richard Waller made a presentation to the Royal Society entitled "Tables of the English Herbs reduced to such an order, as to find the name of them by their external figures and shapes'. The development of computer-based, multi-media keys provides one means of addressing this critical identification and diagnostic function. Use of modern computer programs may simplify and augment the process of learning taxon identification.

Taxonomic Key

One of the least appreciated duty of taxonomists is the construction of tools that allow others to identify the members of a flora or fauna. A key is a device, which when properly constructed and used, enables a user to identify an organism. Keys are devices consisting of a series of contrasting or contradictory statements or propositions requiring the identifier to make comparisons and decisions based on statements in the key as related to the material to be identified.

Taxonomic key is a device for quickly and easily identifying to which species an unknown plant belongs. The key consists of a series of choices, based on observed features of the plant specimen. It provides a choice between two contradictory statements resulting in the acceptance of one and the rejection of the other. A single pair of contradictory statements is called a couplet and each statement of a couplet is termed a lead. By making the correct choice at each level of the key, one can eventually arrive at the name of the unknown plant.

Types of Taxonomic Keys

i. Single-access keys/ dichotomous key

Dichotomous keys are a type of 'single-access' key because there is a single starting point at the root of the key. Keys in which the choices allow only two (mutually exclusive) alternative couplets are known as dichotomous keys. The traditional paper-based dichotomous key remains a fundamental tool in the armoury of many biological recorders. Identification of a specimen is achieved by starting at the root of the key and choosing between pairs of descriptions, known as couplets, which normally consist of contrasting states of one or more morphological characters. Each description in a couplet is also known as a 'lead'. Depending on which lead in a couplet most closely matches the specimen, the user is directed to another couplet in the key or to the identity of the specimen. An useful way of conceptualising such a key is as a decision tree and dichotomous keys are occasionally presented as such (Tilling, 1984).

In constructing a dichotomous key, contrasting characters are chosen that divide the full set of possible species into smaller and smaller groups i.e. the statements typically begin with broad characteristics and become narrower as more choices are required. Each time a choice is made, a number of species are eliminated from consideration and the range of possible species to which the unknown specimen may belong is narrowed. Eventually, after sufficient choices have been made, their range reduces to a single species and the identity of the unknown plant is revealed.

There are two major ways of presenting single-access keys on a printed page: 'parallel' and 'yoked'. In parallel keys, the leads of a single couplet are presented one after the other. An advantage is that the contrasting leads can be easily compared, but a disadvantage is that a lead will often need to direct the user to a couplet some distance away in the key. Parallel keys are sometimes referred to as 'bracketed' keys.

In the yoked style of presentation, leads of couplets are split so that leads of successive couplets which result in an identification can be grouped together. An advantage of this style is that for some identification (preferably those most frequently required) all the positive leads which result in that identification are presented together, one after the other.

ii. Multi-access keys/ Poly Clave Keys:

'Multi-access' keys operate in a very different way. In a multi-access key, the user can consider the morphological characters in the key in any order, matching them against the character states from their specimen. This type of key, which is relatively a new alternative to dichotomous keys and becoming increasingly popular, especially because of the ease of computerizing them, is termed multiple access or poly clave or synoptic key. The advantage of these keys is that they allow the user to enter the key at any point.

This key is based on the identification of organisms by a process of elimination. In a written poly clave key there is a series of characters and character states. Each state is followed by a number or code for the species that possess that feature.

The user needs to select any character and then copy down the list of species that possess the feature. Then the user has to select another character and eliminate any species that is not common to both lists. This process has to be continued until the specimen is identified.

Automated species identification (ASI)

Automated Species Identification (ASI) involves the use of a computer to aid species identification (Ayob and Kadir, 2016). Inputs may range from image sensors such as camera or sounds. Automated species identification is an area of rapid technical development which includes assigning specimens to known taxa based on photographic images or audio recordings – techniques of great potential value to biological recorders. Automated species identification is a method of making the expertise of taxonomists available to ecologists, para-taxonomists and others via digital technology and artificial intelligence. Today, most automated identification systems rely on images depicting the species for the identification.

Non-automated techniques for identifying organisms based on photographs and audio recordings are well-established. The iSpot project (iSpot, 2013) is a widely-used web resource where digital photographs of organisms can be posted and identified by a very broad community of users, including many experts.

Computer-based biological identification

The roots of computer based biological identification can be traced back to the 1960s and tools of types we would recognise today began to appear in the 1970s (Pankhurst, 1991; Edwards and Morse, 1995). In recent years, computer science research, especially image processing and pattern recognition techniques, have been introduced into plant taxonomy to eventually make up for the deficiency in people's identification abilities. The rapid development and adoption of information technology over the last few decades has affected almost every aspect of human endeavour, including taxonomy. Automating the identification of insects was seen by P.J. D. Weeks and co workers affiliated with Natural History Museum, London as a new solution to an old problem (Weeks et al.,1997a).

DAISY (the Digital Automated Identification System) is a prototype novel automated identification system. In its pilot phase, the DAISY algorithms were developed to discriminate five species of parasitic wasp, based on differences in their wing structure. DAISY, the image- based system designed for the non-expert user. Using images of wing venation and pattern, it was first tested with parasitic wasps, then other insect groups, including biting midges (Weeks et al., 1999b). Applications of computer based biological identification range from identification of fungal pest *Tilletia indica* (Chesmore et al., 2003), and different insect groups viz., ladybirds (Ayob and Chesmore, 2012), honeybees (Daly et al., 1982), solitary bees (Roth et al., 1999) and moth (Mayo and Watson, 2006). In plant science research, one of the pioneering study by Clark of the University of Surrey aims for the identification of mature specimens taken from the crown of the tree (Clark, 2004, 2007).

Many modern keys, both written and computer-based, include multiple illustrations and/or high-quality photographs. These images are used in conjunction with written descriptions to reduce ambiguity in identification process. Early during short guides and identification keys contained few, if any, images. Scharf (2009) reviewed the use of images in these guides and concluded that this lack was not because early authors did not recognize the importance of image use, but was rather due to publishing constraints (Scharf, 2009).

According to Pier Luigi Nimis and Martellos. Stefano of the University of Trieste, a public research university in Trieste in the Friuli-Venezia Giulia region in northeast Italy, the new computer based tools have several advantages over traditional paper-printed keys viz. (Nimis and Stefano, 2009).

1) They can be made much easier and also be used by non-experts as they are disentangled from systematics.

2) They give access to a large number of images, hyperlinks and metadata,

3) They run on several media, including PDAs and smartphones, which permit their use directly in the field,

4) They can be modified and personalized by the user.

Image-based methods

Image-based methods are considered a promising approach for species identification (Gaston and O'Neill, 2004). David Evans Walter of the Colorado State University and Shaun Winterton of the California Department of Food and Agriculture believe a revolution in computer diagnostics is now under way that may result in the replacement of traditional keys by matrix-based computer interactive keys that have many paths to a correct identification and make extensive use of hypertext to link to images, glossaries, and other support material (Walter and Winterton, 2007).

In 2007 a study was published by S. G. Wu and co workers on imaging processing with leaf recognition algorithm for plant classification using Probabilistic Neural Network (PNN) to build general purpose automated leaf recognition for plant classification (Wu et al*.,* 2007). An important use of images in field guides is as a supplement to the identification keys. Some of the most usable guides of this type include illustrations in their keys to aid in distinguishing between the character state choices. The use of illustrations makes the characters and character state definitions easier to understand. Some guides that provide this type of illustrated key include field guide to the Orchids of Costa Rica and Panama (Dressler, 1993) and field guide to native Oak species of Eastern North America (Stein et al., 2003).

Discussion

Despite advances in computer technology and the increasing availability of multiple-access taxonomic keys, traditional dichotomous keys remain the most often used taxonomic identification tools. Alpha-taxonomy may have been displaced to the bottom of the systematic alphabet, but the ability to identify species is a prerequisite for most biological sciences (Gotelli, 2004), is critical in pest management and quarantine, and in some cases is now required by legislation covering conservation surveys and international transfer of specimens (Marshall, 2003).

The nuts and bolts of taxonomy, i.e., the description of new species, placement in a classification, and production of identification tools (alpha-taxonomy), is not the purview of journals with high Institute for Scientific Information impact factors. Student's interest and ability to identify plants is declining at a time when there is a great need for these skills. Computer programs may not replace classical plant identification methods but it could help in reviving the interest of students in identification of flora and fauna.

Nomenclatural-taxonomic changes, progress in exploration, the discovery of new species, often render a key outdated within a few years. Computer-aided keys, however, can be updated and corrected in real time. Taxonomists rely on taxonomic keys to help identify known organisms and determine whether they have discovered a new organism entirely. The use of pictorial keys in practice should not be to identify a specimen but to say what a specimen is not. In other words, they are best used as a process of elimination. Eventually, one may end up with a short-list, when more detailed descriptions of the taxa in question should be consulted (Platt, 1984). In recent times progress is also being made on replacing keys entirely by optical matching of specimens to digital databases and DNA sequences. These new tools may go some way toward alleviating the taxonomic impediment to biodiversity studies and other ecological and evolutionary research, especially with better coordination between those who produce keys and those who use them and by integrating interactive keys into larger biological Web sites (Walter and Winterton, 2007).

Conclusion

Usually only experts such as taxonomists and trained technicians can identify taxa accurately, because it requires special skills acquired through extensive experience. Classical taxonomic keys are based on taxonomic terms which only a person with expertise and theoretical knowledge can use. So it becomes difficult for a non taxonomist to identify flora or fauna. The development of computer-based, multi-media keys and automated species identification (ASI) provides the means of addressing this critical identification and diagnostic function. A person not trained in taxonomy may use these image based methods and other automated species identification means for identification. Computer programs may not replace classical plant identification methods entirely, they have the potential to make these methods more effective.

Reference

Ayob, M. Z. and Kadir, M. K. A. (2016). Computer Aided Taxonomy: A Case Study on the Automated Identification of Invasive Ladybirds in the UK. *Asian Journal of Scientific Research.* 9: 266-272.

Ayob, M. Z. and Chesmore, E. D. (2012). Hybrid feature extractor for Harlequin ladybird identification using color images. Proceedings of the IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology, May 9-12, 2012, San Diego, CA., USA., Pp: 214-221.

Chesmore, D., Bernard, T., Inman, A. J. and Bowyer, R. J. (2003). Image analysis for the identification of the quarantine pest, *Tilletia indica*. *OEPP/EPPO Bullet.*. 33: 495-499.

Clark, J. Y. (2007). Plant Identification from Characters and Measurements Using Artificial Neural Networks. In: Automated Taxon Identification in Systematics: Theory, Approaches and Applications, MacLeod, N. (Ed.). CRC Press, London, ISBN: 9781420008074. Pp: 207-224.

Clark, J. Y. (2004). Identification of botanical specimens using artificial neural networks. Proceedings of the IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology, October 7-8, 2004, La Jolla, CA., USA. Pp: 87-94.

Daly, H. V., Hoelmer, K., Norman, P. and Allen, T. (1982). Computer-assisted measurement and identification of honey bees (Hymenoptera: Apidae). Ann. Entomol. Soc. Am. 75: 591-594.

Dayrat, B. (2005). Towards integrative taxonomy. *Biol. J. Linn. Soc.* 85(3): 407–415.

Dressler, R. L. (1993). Field guide to the orchids of Costa Rica and Panama. Ithaca, NY: Cornell University Press. Pp. 260.

Edwards, M. and Morse, D. R. (1995). The potential for computer-aided identification in biodiversity research. *Trends in Ecology & Evolution.* 10 (4): 153–158.

Gaston, K. J. and O'Neill, M. A. (2004). Automated species identification: why not? *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 359(1444): 655–667.

Griffing, L. R. (2011). Who invented the dichotomous key? Richard Waller's watercolors of the herbs of Britain. *American Journal of Botany.* 98 (12): 1911–1923.

Gotelli, N. J. (2004). A taxonomic wish-list for community ecology. *Philos. Trans. R. Soc. London. B*359: 585–957.

iSpot (2013). iSpot - your place to share nature [online]. Available from: http://www.ispot.org.uk/

Marshall, S. A. (2003). The real cost of insect identification. *Newsl. Biol. Surv. Can*. (Terr. Arthropods). 22 (1): 1-38.

Maxime, M., Donatello, C., Romain, R., Ingrid, A. and Damien, M. (2017). A survey on image-based insect classification. *Pattern Recognition (Elsevier).* 65: 273-284.

Mayo, M. and Watson, A. (2006). Automatic species identification of live moths. *Proceedings of the 26th SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence,* December 11-13, 2006, Cambridge, UK. Pp: 195-202.

Nimis, P. and Stefano, M. (2009). Computer-aided Tools for Identifying Organisms and their Importance for Protected Areas. *Eco. Mont.* 1: 55-60.

Pankhurst, R. J. (1991). Practical taxonomic computing. Cambridge University Press. Pp. 202.

Platt, H. M. (1984). Pictorial taxonomic keys: Their construction and use for identification of free living marine nematodes. *Cahiers De Biologie Marine Tome.* XXV : 83-91.

Reaka-Kudla, M. L., Wilson, D. E. and Wilson, E. O. (1996). Biodiversity II: Understanding and Protecting Our Biological Resources (Google eBook). Joseph Henry Press. Pp. 551.

Scharf, S. (2009). Identification keys, the 'natural method,' and the development of plant identification manuals. *Journal of the History of Biology.* 42: 73–117.

Roth, V., Pogoda, A., Steinhage, V. and Schroder, S. (1999). Integrating feature-based and pixel-based classification for the automated identification of solitary bees. *Proceedings of the Annual Convention of the German Society for Pattern Recognition,* September 15-17, 1999, Bonn. Pp. 120-129.

Stein, J. D., Acciavatti, R. E. and Binion, D. (2003). Field guide to native Oak species of eastern North America. [Morgantown, W. Va.]: U. S. Forest Service, Forest Health Technology Enterprise Team. Pp.166.

Tilling, S. (1984). Keys to biological identification: their role and construction. *Journal of Biological Education.* 18 (4): 293–304.

Walter, D. E. and Winterton, S. (2007). Keys and the Crisis in Taxonomy: Extinction or Reinvention? *Annual Review of Entomology*. 52: 193 – 208.

Weeks, P. J. D., O'Neill, M. A., Gaston, K. J. and Gauld, I. D. (1997b). Automating insect identification: exploring the limitations of a prototype system. *Journal of Applied Entomology*. 123(1): 1-8.

Weeks, P. J. D., Gauld, I. D., Gaston, K. J. and O'Neill, M. A. (1997a). Automating the identification of insects: A new solution to an old problem. *Bull. Entomol. Res*. 87: 203-211.

Wu, S. G., Bao, F. S., Xu, E. Y., Wang, Y. X., Chang, Y. F. and Xiang, Q. L. (2007). A leaf recognition algorithm for plant classification using probabilistic neural network. *Proceedings of the IEEE International Symposium on Signal Processing and Information Technology,* December 15-18, 2007, Giza, Egypt. Pp: 11-16.