TAXOKEY: a universal taxonomic key program using DOS text files and graphics

John A.Byers

Abstract

A program (TAXOKEY.EXE), requiring only a few keystrokes to use, is described as a universal taxonomic key in the identification of plant and animal species by IBM-compatible personal computers. A DOS text file serves as the database for a key that can be dichotomous or with multiple choices. TAXOKEY can optionally display colour or monochrome screen pictures to illustrate the keys. Additional programs are described that are used to make the keys and check for errors, as well as convert .PCX graphic images for use with TAXOKEY. A second text file can be searched by TAXOKEY for information on particular species descriptions, notes and references.

Introduction

Biological observations and experiments are only valid when the organisms in question are accurately identified. In environmental and ecosystem studies, taxonomic keys may be the most practical method of classifying large numbers of organisms into taxonomic groups. The specimens can then be optionally sent to an authority for species identification. More specialized keys permit one to identify an organism to species by progressing from general family or genus characters to species-specific traits. Compared to keys found in books or articles, computer software keys are easier and faster to use as well as update. My first objective was to construct a computer program that is flexible enough to accommodate many different keys with text and illustrations for many taxonomic groups of organisms. This could be accomplished by having the software use DOS text files containing not only the key's subject content but also a few simple codes controlling the key's organization. Names of graphic images (.PCX standard) would be included in the text files to illustrate the keys. Another objective was to make the software intuitive with a minimum of keyboard control features in order to shorten program learning time and simplify use. Such a computerized universal taxonomic key is described here that is of

Department of Plant Protection, Swedish University of Agricultural Sciences, S-230 53 Alnarp, Sweden. E-mail: john.byers@vsv.slu.se

general use to biological researchers, systematists, and educators.

Implementation

Making a text file of a taxonomic key

A DOS text file given any name (e.g. BBEETLE.KEY) serves as the database for a particular key. The author or revisor of the taxonomic key edits the text in either a word processor (saved as DOS text), the DOS text editor EDIT.EXE (included in all versions of DOS 5.0 and above) or TAXO-ED.EXE (a specialized editor described here). The universal key program can only work with a text file that has a special format (an example is shown in Fig. 1).

A 'page' of the key always begins with # and can be followed optionally by one to 10 file names of graphic images for illustrating that page of the key (Fig. 1). The bit-mapped color or monochrome images can be EGA (640 × 350 pixels) or VGA (640 × 480 pixels) contained in two EGA or three VGA files on the hard drive. The EGA graphic files can be any name but must have the extension *.A9* in the key's text file and the corresponding files on disk must have the extensions *.A9* and *.B9* (explained in more detail later). When a VGA file is to be used by a page in the key, the name in the text file must have the extension *.A12*, while the three graphic files have the same parent name but end in *.A12*, *.B12* and *.C12*. While up to 10 pictures can be scrolled for any particular page of the key (always beginning with #), only the first picture has text of choices overlaid from the key's text file. The first page of the key (#*1*) does not have a second # as do all subsequent pages which refer back to a previous page, however, this line must end with a space and asterisk (Fig. 1).

The next line after the beginning of any page has the location on the screen where the text of the first choice should be placed (Fig. 1). These lines begin with the ^ (exponent) symbol followed by the *y*-axis (line position) that starts at the top of the screen, followed by a space, and then the *x*-axis position where the text will start. This is illustrated in Figure 2, which is the TAXOKEY program's screen output controlled by the first 15 lines of Figure 1. The *y*-axis positions are determined by the type of graphic image so that for EGA there are 25

```
#1 F5.A12 F6-10.A12 F7.A12 F1.A12 F2-3.A12 F4.A12 *
^1 2
Revised Oct. 1, 1995
^3 2
abdominal stemites progressively rising posteriorly towards apices
^4 2
of elytra (fig. 5); elytra flat, posterior elytra not sloping ventrally;
^5 2
tibiae produced apically into long, curved tooth (fig. 6)
^6 2
Subfamily SCOLYTINAE *A*
^8 2
abdominal stemites not rising posterior (fig. 7); posterior elytra sloping
^9 2
ventrally; tibia with small lateral spines *B*
#B F9.A12 F7.A12 F6-10.A12 backto #1
^1 2
head completely concealed from above; pronotum granulate anteriorly and
^2 2
punctate posteriorly (fig. 9); third tarsal segment not bilobed (fig. 10)
^3 2
Subfamily IPINAE *C*
^5 2
head partly visible from above; pronotum uniformly sculptured; third
^6 2
segment of tarsi bilobed (fig. 8): Subfamily HYLESININAE *D*
#D F11-12.A12 backto #B
^1 2
anterior elytral margin granulate, ridged (fig. 11) *2B*
^3 2
anterior elytral margin smooth, not ridged (fig. 12) *17B*
```

Fig. 1. Example of a DOS text file (first 31 lines of BBEETLE.KEY) of a taxonomic key for European bark beetles (from Grüne, 1979). The first line begins with #1 and optional names of one or more bit-mapped image files (e.g., F5.A12, shown in Figure 2) and ends with an asterisk. The second choice of 'page one' (lines 12–15) ends with *B* that is the destination page beginning with #B (line 16). The text file should end with two hard returns (by pressing [Enter] twice; see text for details on the format of the text file).

possible lines and for VGA there are 30 lines. The x-axis positions range from 1 to 80, left to right across the screen. The next line contains the text of the first choice (less than 80 characters, depending on the x-position). Several lines can make up a choice by using several ^y x position lines each followed by a line of text (Fig. 1). However, the last line of a series of lines comprising a choice must end in a

destination letter or number such as *A* that is enclosed in asterisks (Fig. 1).

All choices, either dichotomous or up to 30 lines (VGA screen), for a page of a key must be grouped together and their ordering from top to bottom corresponds to the desired cursor scrolling sequence (Fig. 1). The destination indicated on another page of choices (e.g., #A as indicated by *A* above, Fig. 1) can be anywhere in the text file and does not have to be in any relation to other pages. All subsequent pages after the first, for example a second beginning with #A, can have up to ten graphic pictures, each separated by a space, but the last characters in the line must refer back to a previous page, in this case #1 (Fig. 1). The words *back to* can immediately precede the second # to aid one's understanding. When a number of choices leads one to a species determination, this choice should end with the *Genus species (Author)* * (ending asterisk, see Figs 1 and 3). The *Author* and parentheses are optional.

Checking a text file of a taxonomic key for errors

The specialized text editor, TAXO-ED.EXE, can be used to rapidly check for proper positioning of text on any bit-mapped pictures. However, this program does not check for appropriate syntax nor for correct addressing between choices and pages of the key. This can be done by running the program TAXO-CK.EXE, which takes the key's text file and makes another file that contains the same lines but also inserts any error messages immediately after the lines with errors. This new text file is named TAXO-CK.ERR (overwriting any previous file with this name). A word processor or TAXO-ED.EXE is used to correct the errors
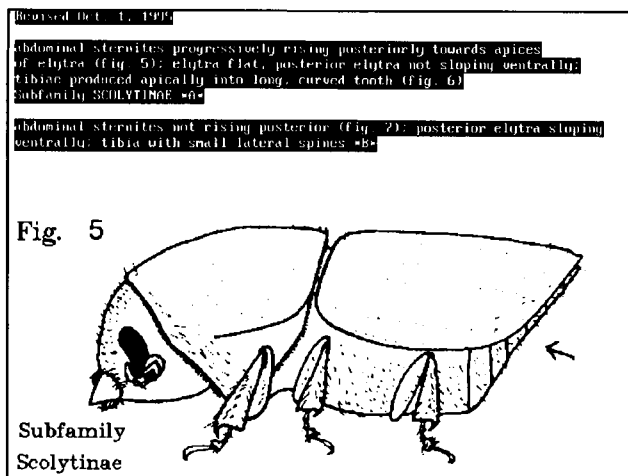

Fig. 2. Computer screen image composed of a scanned picture (Fig. 5 from Grüne, 1979) and the appropriate text from the key's text file (Fig. 1). The scanned picture was arranged by PC Paintbrush and saved as a .PCX image and then converted by PCX.EXE into three *.?12 binary image files loaded by TAXOKEY.EXE.
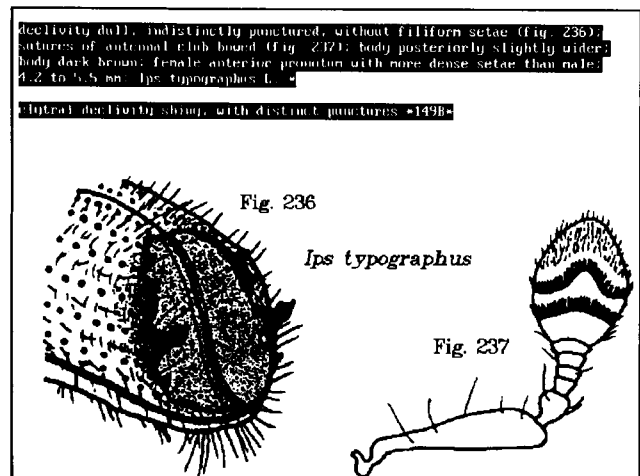

Fig. 3. Computer screen image of scanned pictures (Figs 236 and 237 from Grüne, 1979) and text from the key's text file, BBEETLE.KEY (from Grüne, 1979). Pressing [Enter] while the cursor is on a *Genus species* * causes the program to search a text file of species descriptions, notes and references (e.g. BBEETLE.REF).

and remove the error messages. This file should then be saved as the name of the original file of the key. Conditions that cause an error are whether: (i) a choice on a page specifies a non-existing destination page or one that is not bracketed with asterisks (e.g., *A1* is correct); (ii) multiple page destinations occur when only one is allowed (the offending line numbers are given); (iii) the $^\wedge y$ x positions for choices are outside the appropriate ranges; (iv) a page refers back to a non-existing page; (v) a page refers back to a previous page number of which there are multiple copies (the line numbers involved are given); (vi) a later page refers back to a page that does not have the later page as its destination; and (vii) the graphic file endings are incorrect (they must end in *.A9* or *.A12*).

Additional components of a universal key

The main program (TAXOKEY.EXE) begins by opening a text file, TAXOKEY.DAT, that contains specific data. This file is made with a word processor and contains five lines that can be of any width in characters. Line (1) specifies the name of the text file for the key while line (2) has the name of a text file containing the species descriptions, notes and references. This file of species descriptions also is made with a text editor (or TAXO-ED.EXE) but the width of the text lines should be no more than 80 characters (width of the screen). This file is converted each time the main program is run into a random access file so that scrolling is possible. Line (3) of TAXOKEY.DAT has the title of the key which can be divided by commas so that up to three lines are shown on the main program's opening title screen. Line (4) is for the author's name and address which also is divided by commas that can designate up to six lines on the opening title screen. Line (5) consists of two numbers separated by a space that specify the text and border colors, respectively (text colours are standard QBASIC colours from 9 to 15 and border colours from 2 to 15).

Drawing, scanning and converting graphic images

Illustrations for a key can be drawn by computer using Windows Paintbrush, the standard drawing program included in the Windows accessories. Drawings can be saved in monochrome or 16 colours in standard .PCX format (ZSoft Corp., Marietta, GA, USA) for later conversion to image files compatible with TAXO-KEY.EXE and QuickBASIC by the program PCX.EXE (explained below). Images can be scanned with hand- or flatbed scanners directly into .PCX image format or as .TIF files that are then converted by the Shareware program Graphic Workshop (Alchemy Mindworks Inc., PO box 500, Beeton, Ontario L0G 1A0, Canada). PC paintbrush and a Marstek M-800 Plus hand-scanner were used to scan drawings from published keys into .PCX computer images (Figs 2 and 3). At 200 dpi (dots per inch) a scanned image of 8.128×6.096 cm fits exactly in the 640×480 pixels of the VGA screen. Paintbrush can then be used to combine, move, or erase graphics and add text.

The saved images in .PCX format then are converted to BASIC-compatible EGA or VGA files with appropriate extensions (*.A9* or *.A12* etc.) by the program PCX.EXE. These image files are each 51 211 bytes or 153 633 bytes total per image (VGA). Thus, several megabytes or more of disk space may be required for uncompressed images of a key. In order to transport numerous graphic files between computers and users, I made the program ZIP-WILD.EXE which works with the popular compression software PKZIP.EXE (PKWARE Inc., Brown Deer, WI, USA). ZIP-WILD.EXE with the aid of PKZIP.EXE takes all graphic files of wildcards *.?12 or *.?9 and compresses them to standard *.ZIP files. With the companion shareware program, PKUNZIP.EXE, the *.ZIP files can be uncompressed on the same or a second computer. The 51 211 byte graphic files typically compress to only 2–5% of their former size, for example, 519 bark beetle image files with 26.6 Mbytes were compressed to only 1.3 Mbytes and fit on one 1.44 Mbyte disk.

Many users have limited hard drive space and probably will not want to use upwards of 25 Mbytes for the uncompressed image files, although these can be viewed almost instantly. Thus, TAXOKEY looks for the file PKUNZIP.EXE in the current directory and if present then TAXOKEY assumes that all image files are compressed in *.ZIP files. In this case, TAXOKEY uses PKUNZIP.EXE to uncompress the particular image files specified in the text file for the key and then loads these uncompressed images as usual. These uncompressed image files are deleted as soon as they are displayed to avoid filling the disk. The user must decide whether it is more important to have maximum speed for viewing pictures (no PKUNZIP.EXE in the current directory and many megabytes of *.?12 image files) or slower viewing but requiring only 1 or 2 Mbytes of disk space. The uncompressed images would be best for use with a writable optical disk.

Running the program TAXOKEY.EXE

The program is easy to use: type *taxokey* and the starting menu shows the name and address of the authors until a key is pressed. The first page may or may not have graphic pictures and from one (usually two) to several choices (Fig. 2) that can be selected by the [up/down] cursor keys and then pressing [Enter]. Pressing the [Tab] key will show additional graphic pictures if available. Forward progress through the key is simply a matter of selecting choices as

above. The [Backspace] key can be pressed at any time to go back one page or one picture. When the progression ends with a Genus species * (Fig. 3), pressing [Enter] causes a search for the *Genus species* or *G. species* in a text file containing species descriptions, notes and references. The search can be continued by successive presses of [Enter]. Line by line scrolling is possible at any time, and pressing [Esc] returns one to the main key. The user can immediately return to the first page of the key by pressing [Home]. Pressing [F1] shows the brief help, [F5] a detailed manual (in TAXOKEY.TXT), and [Esc] exits the program.

Discussion

The ease of use of TAXOKEY is in contrast to several more comprehensive and complicated program systems. The most widely known is called 'DELTA' (DEscription Language for Taxonomy) (Dallwitz, 1980a, b). This system has several programs (CONFOR, PAUP, KEY, INTKEY, DELFOR, and others) which because of the numerous facilities require many hours to learn (Watson *et al.*, 1986; Askevold and O'Brien, 1994). The complexity of the system is demonstrated by the detailed and syntactical descriptions found in Askevold and O'Brien (1994).

Several computerized keys that are specialized for various groups of organisms have been reported (Zo, 1982; Doering *et al.*, 1988; Terranova and Roach, 1989; Immelman, 1991). Hypercard database systems are another approach to organization of taxonomic data and have been used in teaching (Skov, 1989; Theodoropoulos and Loumos, 1994). Expert systems include those in which data are entered about a specific group of organisms and the program is instructed to generate a series of questions that when answered by the user allows the program to suggest a species name for a particular specimen (Wilson and Partridge, 1986; Woolley and Stone, 1987; Allkin *et al.*, 1992; Estep *et al.*, 1993; Dykes *et al.*, 1994). The expert systems, sometimes called artificial intelligence systems, can be highly specific with a fixed database or quite flexible in database and program structure.

The 'PANKEY' system (Pankhurst, 1986), can work alone or with DELTA as a universal key and currently comprises 12 programs including data capture, DELTA editor, diagnostic description, character, clustering and cladistic analyses, key construction, and an interactive expert system (R.J.Pankhurst, Taxonomic Systems, 23, Royal Crescent, Edinburgh EH3 GQA, UK). These programs are from £95 to £145 each (£1095 total). This suite of programs is again for experienced users and dedicated taxonomists due to the considerable cost and

necessarily slow-learning curves, but they offer the most comprehensive facilities.

Use of taxonomic keys in books can be tedious and time-consuming because of the need to skip forward and backward through numbered parts of different pages of the key and illustrations—with the possibly of losing one's place. TAXOKEY allows the user to easily navigate through a key by keeping track of up to 150 levels of branching in a dichotomous key. A particular path of choices entered by the user that describes the progression from the first page to the current place within the key is 'remembered' by the program in a memory array (e.g., 121222112221). Using [Enter] or [Backspace] moves one forward or backward through the array allowing the program to display appropriate pages as further specified in the key's text file. The values in the array are not altered unless the user selects a choice different from that in the current path. This means that in theory up to 2.85×10^{45} choices are possible if all 150 levels (n) are utilized fully, according to:

$$\sum_1^n 2^n.$$

The maximum number of direct paths to species identifications is equal to 2^n which is also the maximum number of species that can be considered. However, no keys are expected to use all 150 possible levels fully, for example, the identification of *Ips typographus* in the bark beetle key (Fig. 3) requires 12 levels of choices while *Scolytus kirschi* needs only three levels.

The real size limitation for a key is how long it takes a computer to do a linear search of the key's text file. The bark beetle key, BBEETLE.KEY, is 83 Kbytes with 161 pages (average of 15 text lines per page) and takes a 486-computer about 0.12 s to find a designated page. Full utilization of 10 choice levels would allow coverage of 1024 species with 2^n-1 pages, and by comparison would take about 0.76 s to look for a certain page. Similarly, a 20-level dichotomous key could have at most 1 million species but would require a hugh text file of 539 Mbytes and take 13 min on average to find a specified page.

Compared to published keys, computerized keys are easier to restructure and modify after taxonomic revisions. Thus, several specialists could work together over the Internet and update or improve a software-based key rapidly and with less effort. Computerized keys are easy to distribute by disk, e-mail and the Internet. The use of illustrations in these keys makes it possible even for non-specialists to successfully identify organisms. Taxonomists would have incentives to distribute their work in computer keys since they could advertise their name and institution.

The program system can be obtained by sending one

formatted disk (1.2 M or 1.44 M) to me, while an additional disk is required for the compressed VGA image files of each example key: European bark beetles (Grüne, 1979) or Grasshoppers of Colorado (Capinera and Sechrist, 1982). Donations of $5 for shipping and handling would be appreciated. The software can also be downloaded from the internet at: http://alyssum.stud. slu.se:8001/~johnb/software.html.

Acknowledgements

References

Allkin,R., Moreno,N.P., Gama-Campillo,L. and Mejia,T. (1992) Multiple uses for computer-stored taxonomic descriptions: Keys for Veracruz. *Taxon*, **41**, 413–435.

Askevold,I.S., and O'Brien,C.W. (1994) DELTA, an invaluable computer program for generation of taxonomic monographs. *Ann. Entomol. Soc. Am.*, **87**, 1–16.

Capinera,J.L. and Sechrist,T.S. (1982) Grasshoppers (Acrididae) of Colorado: Identification, biology and management. *Colorado State University Experiment Station, Bulletin,* **no. 584S**, 1–161.

Dallwitz,M.J. (1980a) A general system for coding taxonomic descriptions. *Taxon*, **29**, 41–46.

Dallwitz,M.J. (1980b) User's guide to the DELTA system. Report no. 13. CSIRO Division of Entomology, PO Box 1700, Canberra City, ACT 2601, Australia.

Doering,B., Ehrhardt,S., Leuchke,F.K. and Schillinger,U. (1988) Computer-assisted identification of lactic acid bacteria from meats. *Syst. Appl. Microbiol.*, **11**, 67–74.

Dykes,G.A., Britz,T.J. and Von Holy,A. (1994) Numerical taxonomy and identification of lactic acid bacteria from spoiled, vacuum-packaged vienna sausages. *J. Appl. Bacteriol.*, **76**, 246–252.

Estep,K., Slugs,R. and Syvertsen,E.E. (1993) 'Linnaeus' and beyond: Workshop report on multimedia tools for the identification and database storage of biodiversity. *Hydrobiologia*, **269**, 519–525.

Grüne,S. (1979) *Brief Illustrated Key to European Bark Beetles.* Verlag M&H Schaper, Hannover.

Immelman,K.L. (1991) Synopsis of the genera *Nesaea* and *Ammannia* (Lythraceae) in Southern Africa. *Bothalia*, **21**, 35–50.

Pankhurst,R.J. (1986) A package of computer programs for handling taxonomic databases. *Comput. Appl. Biosci.*, **2**, 33–39.

Skov,F. (1989) Hypertaxonomy: A new computer tool for revisional work. *Taxon*, **38**, 582–590.

Theodoropoulos,G. and Loumos,V. (1994) Parasitology tutoring system: A hypermedia computer-based application. *Comput. Meth. Prog. Biomed.*, **42**, 113–118.

Terranova,A.C. and Roach,S.H. (1989) A computer-aided key for distinguishing South Carolina species of the genus *Phidippus* (Aranea: Salticidae). *J. Agric. Entomol.*, **6**, 23–30.

Watson,L., Aiken,S.G., Dallwitz,M.J., Lefkovitch,L.P. and Dubé,M. (1986) Canadian grass genera: Keys and descriptions in English and French from an automated data bank. *Can. J. Bot.*, **64**, 53–70.

Wilson,J.B. and Partridge,T.R. (1986) Interactive plant identification. *Taxon*, **35**, 1–12.

Woolley,J.B. Stone,N.D. (1987) Application of artificial intelligence to systematics: SYSTEX, a prototype expert system for species identification. *Syst. Zool.*, **36**, 248–267.

Zo,Z. (1982) The sequential taxonomic key: An application to some copepod genera. *Hydrobiologia*, **96**, 9–14.