

# Types of identification keys

Gregor Hagedorn, Gerhard Rambold, Stefano Martellos

**Abstract** — A number of terms related to identification tools are introduced and the advantages of selected types of identification keys are compared.

**Index Terms** — Identification tools, single-access key, dichotomous key, polytomous key, lead, couplet, free-access key, multi-access key, matrix key, multi-entry key.

---

◆

## 1 INTRODUCTION

The generalization of individuals (things, events, etc.) into classes is essential to transfer knowledge across individual incidents. When learning a language, we learn the defining features of classes like “table”, “chair”, “shrub”, etc. Similarly, biology defines formal classes for living things (called “taxa”) together with class names (“taxon names”) and defining descriptions. The assignment of an unknown object to a taxon is called “identification” or “determination”. To non-biologists this may be confusing, the term “identification” being more commonly associated with the naming of individuals (as in “ID card” or “record identifier”).

The number of taxa in biology is very large. For example, currently about 900 000 insect taxa alone are recognized. Compared to the average vocabulary of an educated English native speaker of roughly 25 000 words, it is clear that teaching the vast “taxon vocabulary” to biology students was always problematic. Although comparing a collected specimen sequentially with published descriptions or representative specimens is an essential identification method, any “linear search” method comparing one specimen after another soon becomes impractical.

Biologists have therefore developed various forms of “identification keys” to “unlock” the knowledge that would otherwise remain inaccessible. These are essentially “divide and conquer” search algorithms that reduce the result set recursively until the remainder is small enough to be solved by direct comparison. The fastest algorithms are those that provide a division into equally sized partitions (leading to search algorithms that scale logarithmically with

---

*G. Hagedorn is with the Julius Kühn-Institute, Federal Research Centre for Cultivated Plants, Institute for Epidemiology and Pathogen Diagnostics, Königin-Luise-Straße 19, D-14195 Berlin, E-mail: gregor.hagedorn@jki.bund.de. – G. Rambold is with the Department of Mycology, University of Bayreuth, D-95440 Bayreuth – S. Martellos is with the Dept. of Life Sciences, Univ. of Trieste, I-34127, Trieste, Italy. E-mail: martelst@units.it.*

the number of taxa). Biological keys don't always provide this, because other factors (character observation reliability, convenience, cost, etc) conflict with the desire to provide fastest progress. The authors of biological identification keys, however, typically realize that evenly splitting choices are desirable.

## 2 SINGLE-ACCESS, MULTI- OR FREE-ACCESS, AND MULTI-ENTRY KEYS

The most traditional biological identification keys are the easily printable dichotomous (every choice has two alternatives) or polytomous (two or more options at each choice) forms. The structure of these keys typically consists of a series of alternative statements, called "leads". All leads that need to be evaluated for a single decision form a "couplet".

Clearly, dichotomous keys are a special case of polytomous keys. For simple choices involving a single characteristic (a "character" or "feature"), such as "wings: 1. present 2. absent" or "flower colour: 1. red 2. blue 3. pink 4. yellow", the number of options matters little. However, to achieve reliable identification in the face of natural variability and continuous variation, it may be desirable to use complex Boolean statements involving multiple characters. For example, "leaves hairy and flowers red" versus "leaves glabrous or flowers not red" may be used in a case where the alternative may include glabrous plants with pinkish-red flowers, or plants with mixed hairiness but other flower colours. Although it is theoretically possible to construct a polytomous key with Boolean lead statements, practice has shown that the result is often akin to a logical riddle. Many editors therefore recommend or require the use of dichotomous keys.

However, a key may be a mixture of simple polytomous and complex dichotomous choices. The generalizing term "single-access key" [1] is therefore used in the present paper to include both dichotomous and polytomous keys. The equivalent term in computer science is "decision tree".

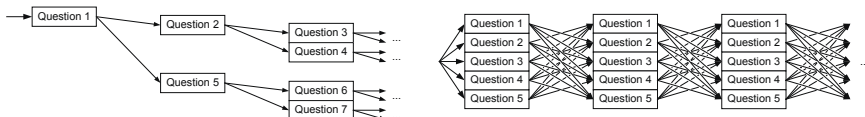


Fig. 1 – User interaction steps in a single-access key (left, the sequence of steps follow the data structure) and a free-access key (right, the sequence is determined by the user). From [2].

An alternative to a single-access key is the *free-access key* (also known as multi-access key, matrix key, or, incorrectly, "synoptic key"<sup>1</sup>). Whereas in a single-access key a fixed sequence of choices (decisions) is defined by the author (providing a single path to each result), in a free-access key the sequence of choices is up to the user. In every step, the user can select from the list of characters offered, and choose a matching state or value. Thus a free-access key is the set of all possible single-access keys that arise by permutating the order of characters (Fig. 1). Although printable free-access keys exist, they

<sup>1</sup> The term „synoptic key“ has traditionally been used for single-access keys that reflect the taxonomic hierarchy. Its use for multi-access keys (especially printable ones) should be avoided.

are most suitable for computer-aided identification tools, and have a long development history [3]. Examples are DELTA-IntKey [4], Lucid [5], NaviKey [6], Xper<sup>2</sup> [7]. The Flash-based IBIS-ID [8] was newly developed in *KeyToNature*.

In a free-access key, the choice of characters is repeated at every step. A related form, the multi-entry key, allows free choice of characters in a first step (a “multi-character-query-form”), followed either by a field-guide-like listing of remaining taxa, or by a dynamically generated (filtered) single-access key (as in the FRIDA/Dryades keys [9]).

	<b>Single-Access</b>	<b>Free-Access</b>	<b>Multi-Entry</b>
<b>Information reduction</b>	High	None (complete information is optimal)	Variable (none if all characters are available in the initial step)
<b>Average identification speed</b>	Depends on the creators of the key	Depends on user’s background knowledge; may exceed average	Variable between single and multi-access
<b>Complex statements (and, or, etc.)</b>	Yes (not recommended for polytomous keys)	No	“No” in entry-form, “yes” in following single-access key
<b>Question-answer style</b>	Possible for simple statements	(Implicit in character state or value choice)	(Implicit in character state or value choice)
<b>Difficulty of choosing next decision</b>	None	Often high for beginners	Variable; depends on completeness of initial entry form
<b>Skipping unanswerable choices</b>	Difficult; all alternative paths must be followed to the end	Easy	Easy in entry-form, difficult in an optional single-access part
<b>Resources required for construction</b>	Low for first draft. Good keys require high expertise	High investment until first version can be tested	High for data matrix, but the size of the matrix is variable

Tab. 1 – Comparison of some identification methods. The comparison is aimed at manually created single-access keys (those generated from a data matrix are not considered here).

An evenly splitting single-access key requires fewer decisions from beginners than a multi-access key. The latter may in fact generate faster progress in terms of “steps”, but requires additional decisions as to which character to use next. Even if the character list is ordered by character suitability and fastest progress, beginners will be tempted to use execute their free character choice. This is problematic if many characters are not yet understood. In contrast, the fixed

identification path of a single-access key also fixes which terms and concepts must be learned first. A disadvantage of single-access keys is that identification may be impossible if a choice cannot be decided at all. This may occur because a character cannot be observed (e. g., a developmental stage is not present in the specimen), or because the options are not communicated clearly enough. The resulting frustration can be high, especially for beginners.

Both free-access and multi-entry keys truly excel in their performance when used by experts. For these, character selection is intuitive and fast. By choosing characters, for which a rare state is present in the specimen, identification progress can then be on an order of magnitude faster than using a single-access key. This is already possible with moderate experience, since states that were never observed by a user before are, by definition, rare. Tab. 1 gives an overview of some differentiating features.

From an author's perspective, matrix-based keys require a high initial investment to research and fill a large character  $\times$  taxon matrix. In contrast, single-access keys require less formal investment. Due to the inherent information reduction (most characters apply only to a relatively small subset of taxa), a reviewable key is faster to produce and proof-reading is less time-consuming than the creation of an equivalent data matrix containing all characters for the same group of taxa. However, a successful single-access key depends strongly on the expertise of the author to choose characters that are convenient, cost-effective, reliable across all taxa in the subtree, and available throughout a large period of the developmental cycle of the organism. Single-access keys may therefore require several cycles of testing until initially overlooked problems have been fixed; their production can be akin to the "debugging" of software code.

Furthermore, the creation of matrix-based keys generally requires learning a special-purpose application like DELTA or Lucid, whereas single-access keys may be created in a text-processing application. Therefore, although newly created single-access keys may occasionally be problematic to use, they offer considerable benefits to both producers and consumers.

Single-access keys, until recently, have been developed only rarely as computer-aided, interactive tools. Noteworthy developments in this direction are the commercial Lucid Phoenix application [10], the FRIDA/Dryades software [9], [11], the *KeyToNature* Open Key Editor" [12], and the open source WikiKeys and jKey [13] application on biowikifarm [14].

### **3 STRUCTURAL VARIANTS OF SINGLE-ACCESS KEYS**

Two additional structural variants of single access keys are relevant when building information models [2]: 1. Couplets may consist of a question, with the leads providing contrasting answers. This question-answer-style is often appealing to beginners. However, complex statements (involving more than one character and Boolean expressions) are not possible. Whereas a mixture of simple polytomous and complex dichotomous couplets in a single key is quite intelligible, a mixture of simple question-answer-style with question-less complex dichotomous couplets is not. 2. The desire for fast identification progress using

convenient characters often conflicts with character variability in a subset of organisms. As long as the character is reliable for the majority of taxa, a frequent solution to this problem is to key out taxa with variable character expression multiple times. This may affect only the terminal taxa, or entire branches of the keys. Whereas the first case will often simply be handled by true duplication, multiple references to entire branches of a decision tree turn a “tree” structure into a directed (and generally acyclic) graph (DAG) and requires careful attention when modelling information models or software. In biology a DAG is sometimes called a “reticulated” identification key.

**Linked Key Style (also called “parallel”, “juxtaposition” or “bracketed” style):**

- 1. Ovule solitary, basal .....2
- Ovules numerous, axile or free-central .....3
- 2. Perianth green, membranous or absent; filaments free .....**Chenopodiaceae**
- Perianth translucent and papery; filaments often united below .....**Amaranthaceae**
- 3. Placentation axile; leaves alternate .....**Saxifragaceae**
- Placentation basal or free-central; leaves usually opposite ..... 4

**Nested Key Style (also called “yoked” or “indented” style):**

- 1. Ovule solitary, basal
  - 2. Perianth green, membranous or absent; filaments free .....**Chenopodiaceae**
  - 2. Perianth translucent and papery; filaments often united below .....**Amaranthaceae**
- 1. Ovules numerous, axile or free-central
  - 3. Placentation axile; leaves alternate .....**Saxifragaceae**
  - 3. Placentation basal or free-central; leaves usually opposite

Fig. 2 – Examples of the linked and nested styles of branching keys in lead style; see [2] for derivation.

## 4 PRESENTATIONAL VARIANTS OF SINGLE-ACCESS KEYS

The dominant presentation styles of single-access keys are shown in Fig. 2. In “linked” keys the connection between couplets is achieved by a linking reference (at the right side) to a couplet ID (left). In nested keys direct nesting of couplets replaces the explicit linking. Nested keys are more commonly known as “indented”, but unfortunately this refers to an accidental (albeit frequent) rather than essential quality [3]. Nested keys may be printed without indentation to preserve space (relying solely on corresponding lead symbols) and linked keys may be indented to enhance the visibility of the couplet structure. Further presentational (“solid keys”, “graphical style”) and semantic (“artificial” or “diagnostic” versus “natural”, “synoptic”, or “phylogenetic”) variants exist; see [2].

## 5 SUMMARY

The order of couplets (choices) in an identification tool may be defined by the creator (single-access key), or may be freely selectable by the user (free-access key). A multi-entry key is an intermediate form that may combine advantages of both forms if only a small character subset is included in the multi-entry phase.

Structural criteria for single-access keys are: a) whether the leads in a couplet are limited to two (dichotomous) or not (polytomous key); b) whether couplets are limited to a single character or combinations of multiple characters, involving Boolean operators such as 'and', 'or', or 'not', are supported; c) whether taxa may be keyed out in multiple places, and whether redirections into entire sections (or "branches") of the key are supported ("reticulated key"); and d) whether leads in couplets are complete statements or split into a question with the couplet and leads providing the answers. Certain presentational forms (nested key versus linked keys) are not structurally relevant.

#### ACKNOWLEDGEMENT

This work was supported by the *KeyToNature* Project, ECP-2006-EDU-410019, in the *eContentplus* Programme. We thank P. L. Nimis and B. Press for review and important suggestions.

#### REFERENCES

- [1] J. Winston, *Describing Species*. Columbia University Press, 1999.
- [2] G. Hagedorn, *Structuring Descriptive Data of Organisms — Requirement Analysis and Information Models*. Ph. D. Thesis, Universität Bayreuth, 2007.
- [3] R. J. Pankhurst, *Practical Taxonomic Computing*, 1991.
- [4] DELTA – DEscription Language for TAxonomy <http://delta-intkey.com/>, 2010-07.
- [5] Lucidcentral.org <http://www.lucidcentral.com>, 2010-07.
- [6] D. Neubacher and G. Rambold, NaviKey – a Java applet and application for accessing descriptive data coded in DELTA format, 2005 (onwards). <http://www.navikey.net>, 2010-07.
- [7] V. Ung, G. Dubus, R. Zaragüeta-Bagils and R. Vignes Lebbe, "Xper<sup>2</sup>: introducing e-taxonomy". *Bioinformatics*, 26 (5): 703-704, 2010; doi: 10.1093/bioinformatics/btp715; see also <http://lis-upmc.snv.jussieu.fr/lis/?q=en/resources/software/xper2>.
- [8] M. Giurgiu, G. Hagedorn and A. Homodi, "IBIS-ID, an Adobe FLEX based identification tool for SDD-encoded multi-access keys". *Proc. of TDWG 2009*, 9-13 Nov. 2009, Montpellier, p. 90, 2009.
- [9] S. Martellos and P. L. Nimis, "KeyToNature: Teaching and Learning Biodiversity. Dryades, the Italian Experience". In: M. Muñoz, I. Jelinek, F. Ferreira (eds.), *Proceedings of the IASK International Conference Teaching and Learning 2008*, pp. 863-868, 2008.
- [10] Lucid Phoenix (<http://www.lucidcentral.org/LinkClick.aspx?link=152>), (2010-07).
- [11] S. Martellos, "Multi-authored interactive identification keys: The FRIDA (FRiendly IDentificAtion) package", *Taxon*, vol. 59 (3), pp. 922-929, 2010.
- [12] S. Martellos, E. v. Spronsen, D. Seijts, N. Torrescasana Aloy, P. Schalk and P. L. Nimis, "User-generated content in the digital identification of organisms: the KeyToNature approach", *Int. J. Information and Operations Management Education*, vol. 3, 3, pp. 272-83, 2010.
- [13] S. Opitz, G. Hagedorn, "The jKey wiki key player and builder". *Proc. of TDWG 2009*, 9-13 Nov. 2009, Montpellier, 2009.
- [14] G. Hagedorn, G. Weber, A. Plank, M. Giurgiu, A. Homodi, C. Veja, G. Schmidt, P. Mihnev, M. Roujinov, D. Triebel, R. A. Morris, B. Zelazny, E. van Spronsen, P. Schalk, C. Kittl, R. Brandner, S. Martellos and P. L. Nimis, "An online authoring and publishing platform for field guides and identification tools". In: P. L. Nimis and R. Vignes Lebbe (eds.), *Tools for Identifying Biodiversity: Progress and Problems*, pp. 13-18, 2010.