



innovative guide software
at paperback guide prices

Request for Collaboration on Diet Filter Database

Greg Sepesi
sepesi@eduneer.com
03 December 2007

1. Background

The adage about the “Man Bites Dog” story being more newsworthy than the “Dog Bites Man” story illustrates how events are more informative when they are less common. This relationship between an event’s information $I(x)$ and its likelihood $p(x)$ is quantified in information theory as

$$I(x) = \log_2 (1/p(x))$$

In the context of field guides, the most unlikely characteristics are the most informative when identifying a species. For example, because wildflowers with only two petals are uncommon, observing that a wildflower has only two petals is very informative during the identification process. Conversely, because white is the most common color of wildflowers, the observation of white petals is not particularly informative.

Each characteristic, such as size or color or petal count, that can be filtered (or keyed¹) by a field guide has its statistical outliers (i.e., those species with an uncommon value for that characteristic). These statistical outliers are the species that are trivial to identify. To maximize the number of species that are trivial to identify, field guides should offer **filter capabilities over as many characteristics as possible**.

This document is a request to North American bird watchers to gather diet information into an on-line spreadsheet for the implementation of diet filters. This and other collaboration requests to help improve field guides are at <http://www.eduneer.com/peck/>

2. Motivation

- **improve the Peck bird information manager** – In the Peck application user survey, the most common request is for improved filtering. As you might expect, the implementation of extensive filtering is time consuming. From a developer’s perspective, is the user’s interest enough to justify the work? And which filters are of most interest to users? A collaborative effort in gathering filter information is somewhat unconventional, but it effectively yields the answer: the filters for which users gather filter information are the same filters in which the user interest justifies the implementation work.
- **improve future field guides** – Because field guide developers have rarely shared filter information, little progress has occurred in electronic field guide filtering capabilities. To promote progress, the results of this collaborative effort will be published under the GNU Free Documentation License, which promotes open use and improvements.

¹ Common in paperback field guides, keys are decision trees concerning a specimen’s characteristics. Each answer narrows the list of candidates for identification. On the other hand, filters are common in electronic field guides. Filters also employ questions about a specimen’s characteristics to narrow the list of candidates, but they are more flexible in the order of questions.

3. Spreadsheet

The on-line spreadsheet for entering diet filter information is at

<http://spreadsheets.google.com/pub?key=ph11Ma-p7o2crxttifUtnHw>

In the spreadsheet, there is a row for each species documented by the Peck application.

The first three columns should not be edited:

- //17 – a photograph index used by the Peck application
- COMMON NAME – the common name of the species
- FILE NAME – the file name for the photograph

The columns to be edited by collaborators are

- COLLABORATOR – Use your full name, your initials, or an alias, but be consistent.
- DATE – Use the YYYY-MM-DD format, which has the convenient property of alphabetical order being the same as chronological order. For example July 24, 2007 would be written as 2007-07-24.
- PRIMARY DIET – In any order, list the diet characters that best represent the primary diet of the species. The diet characters are defined in Table 1. For example, the American Robin’s primary diet is fruit. Therefore, its PRIMARY DIET field is ‘u’.
- SECONDARY DIET – In any order, list the diet characters that best represent the secondary diet of the species. The diet characters are defined in Table 1. For example, the American Robin’s secondary diet consists of insects and worms. Therefore, its SECONDARY DIET field is ‘iw’.

DIET	DIET CHARACTER
<u>a</u> mphibians	a
<u>a</u> quatic invertebrates	q
<u>b</u> irds	b
<u>c</u> arrion	c
<u>e</u> ggs	e
<u>f</u> ish	f
<u>f</u> lowers	l
<u>f</u> ruit	u
<u>g</u> reen plant matter	g
<u>i</u> nsects	i
<u>m</u> ammals	m
<u>n</u> ectar	n
<u>n</u> uts	t
<u>r</u> eptiles	r
<u>s</u> eeds	d
<u>s</u> nails	s
<u>w</u> orms	w

Table 1. Diet Characters

4. Process

- Join the Google newsgroup at <http://groups.google.com/group/peckdb>. To discourage spam within the group, you will be asked a simple question when joining.²
- Post a message to the group to specify which spreadsheet you are interested in editing. Also mention the collaborator name you plan to use.
- Wait for an email confirmation that you have been added to the collaborator list.
- Add your contribution(s) to the on-line spreadsheet.
- After the spreadsheet becomes 85% complete, I'll implement the filter in the Peck application.

5. Editing Tips

- If you prefer the species in AOU order, sort the on-line spreadsheet by its first column.
- Avoid overwriting the work of others. If you happen to be editing the spreadsheet at the same time as another collaborator, you'll see the fields automatically updating. To avoid overwriting each other's work, you could put your initials on ten or so rows before you fill them in.

² The question is about why you want to join the group. Spammers are so busy writing, they typically don't take the time to read.