Debtags is ready

How to add Debtags to your everyday Debian work

Enrico Zini
enrico@debian.org

21 June 2007

What is Debtags

- A vocabulary of categories (*tags* for short).
- A database associating packages to *multiple* tags.
- *Cleverness*. Lots of cleverness.

Faceted classification
Clever tag vocabulary

- Packages are classified from different points of view.
- A point of view is called a *facet* of a package.
- Every facet has a set of tags.
- Every tag answers the question: "what does the package look like from this point of view?"

Enrico Zini enrico@debian.org   Debtags is ready

Example vocabulary 1/2

```
admin::accounting - Accounting
admin::automation - Automation and scheduling
devel::bugtracker - Bug Tracking
devel::buildtools - Build Tool
field::arts - Arts
field::astronomy - Astronomy
game::adventure - Adventure
game::arcade - Action and Arcade
interface::commandline - Command Line
interface::text-mode - Text-based Interactive
interface::web - World Wide Web
interface::x11 - X Window System
role::devel-lib - Development Library
role::documentation - Documentation
role::program - Program
role::shared-lib - Shared Library
use::browsing - Browsing
use::chatting - Chatting
use::checking - Checking
use::compressing - Compressing
...
```

Example vocabulary 2/2

```
Tag: role::program
Description: Program
 Executable computer program.

Tag: scope::utility
Description: Utility
 A narrow-scoped program for particular use case or few
 use cases.
 It only does something 10-20% of users in the field will need.
 Often has functionality missing from related applications.

Tag: protocol::ethernet
Description: Ethernet
 Ethernet is the most popular networking technology
 for creating local area networks (LANs).
 .
 The computers in an Ethernet network communicate over
 twisted-pair or fibre cables and are identified by their
 MAC address. Several different types of Ethernet exist,
 distinguishable by the maximum connection speed.
 The most widespread types today are 100MBit/s
 (100BASE-*) or 1GBit/s (1000BASE-*).
 .
 Link: http://en.wikipedia.org/wiki/Ethernet
```

Example tag database

```
oaf: devel::library, devel::rpc, network::client, network::server, role::program,
  suite::gnome, works-with-format::xml
obconf: interface::x11, role::program, scope::utility, special::not-yet-tagged,
  special::not-yet-tagged::o, uitoolkit::gtk, use::configuring, x11::application
obexftp: admin::file-distribution, network::client, role::program,
  special::not-yet-tagged, special::not-yet-tagged::o
ocaml: devel::compiler, devel::lang:ocaml, implemented-in::ocaml,
  role::metapackage, works-with::software:source
ocaml-base: devel::lang:ocaml, devel::runtime, implemented-in::ocaml,
  interface::x11, role::shared-lib, uitoolkit::tk, x11::application
ocaml-base-nox: devel::lang:ocaml, devel::runtime, implemented-in::ocaml,
  interface::text-mode, role::shared-lib, uitoolkit::ncurses
ocaml-book-en: devel::doc, devel::lang:ocaml, implemented-in::ocaml,
  role::data, role::documentation
```

The official debtags data source

Tag workflow:

1. Tags are entered at
   `http://debtags.alioth.debian.org/todo.html`,
   or through an e-mail submission bot.

2. Unreviewed tags can be downloaded from
   `http://debtags.alioth.debian.org/tags/`
   `tags-current.gz`

3. Tag submissions are manually reviewed and committed to
   `svn://svn.debian.org/debtags/tagdb/tags`

4. The reviewed tag database is converted into an override
   file and uploaded to the archive as an automatically
   processed byhand upload.

The official debtags data source

Vocabulary workflow:

1. The vocabulary is maintained at
   svn://svn.debian.org/debtags/vocabulary.
2. It can be downloaded from http://debtags.alioth.
   debian.org/tags/vocabulary.gz.
3. It is also included in the debtags package.

More data sources

There are things that people want but do not fit in the official tag source:

- Subjective tags, such as ratings.
- Tags specific to one use target, that CDDs or big local installations may want to provide for their users.

More data sources

It can already be done:

1. Tag sources are listed in /etc/debtags/sources.list
2. debtags update will acquire tags from the various sources and merge them together
3. The merged, validated and indexed tag database is stored in /var/lib/debtags/package-tags
4. The merged, validated and indexed tag vocabular is stored in /var/lib/debtags/vocabulary

More data sources

Example `/etc/debtags/sources.list`:

```
# Reviewed tag data from the APT Packages file
tags apt://

# Unreviewed tags from Alioth
#tags http://debtags.alioth.debian.org/tags/

# Package rating from http://www.iterating.org
#tags http://www.iterating.org/tags

# Example of locally maintained custom tag data
#tags file:/etc/debtags/mytags
```

The web-based tagging interface

Lots of clever goodies:

- Tips.
- Editable diff.
- Context-sensitive tips.
- *Really clever* tag search.
- Suggestions inferred from package descriptions.
- TODO-list view with more popular packages first.
- Search-as-you-type view of the entire vocabulary.

The web-based tagging interface

An experiment in getting the maximum quality out of the maximum freedom:

1. Tags are freely added.
2. Suggestions from expert taggers are encoded in tips and context-sensitive tips.
3. Structure is automatically inferred and fed back to the taggers through autogenerated context-sensitive tips.

Nice things to do with tag data

Some things that have already been done:

- Providing info that would be better outside of the package description (e.g. "contact list written in Python with GTK.").
- Cleaning up search results (e.g. hiding shlibs and -data packages).
- Simple filtering (GUI packages, games, commandline programs, C++ development...).
- Domain-specific package views (e.g. a game browser).
- Smarter keyword searches (you can look for "image editor" and actually *find* gimp).
- Finding similar packages.

Software

For users:

- apt-cache
  - Shows tags.
- aptitude
  - Shows tags.
  - Has an experimental tag browser.
- ept-cache
  - Tag filters.
  - Tag-enhanced keyword searches.
- debtags
  - Maintenance of the local tag database.
  - `debtags smartsearch`
- `http://debtags.alioth.debian.org`
  - Smart search at `http://debtags.alioth.debian.org/ssearch.html`

Software

For programmers:

- libept-dev
  - easy to use C++ library (hint: aptitude, synaptic and adept are written in C++).
  - does not mandate a way of using APT, so it can be used without changing the program infrastructure of existing software (such as aptitude, synaptic and adept).
  - designed to be easily bindable to other programming languages.
- python-debian
  - does not use the indexed database, and loads all the tag data into memory.
  - excellent for prototyping Debtags-based ideas thanks to Python having fast native set types.

Future

Future directions discussed during DebConf:

- Automated maintenance of tags not yet edited by humans.
- A new facet of level-of-support tags.

Conclusion

- 70% of the packages are already tagged.
- The dataset is already useful.
- It used to be an experiment.
- Now it's ready.
- Use it.