
CTAN packages get keywords

Jim Hefferon

As part of a larger effort, I am adding to the web pages on the TUG CTAN node the ability to search by keyword and by hierarchical characterization.

1 Audience

The great majority of people who come to CTAN use \TeX and \LaTeX only as an adjunct to their main work. This kind of visitor comes to us with a specific need, looking to swoop into our holdings, grab just what helps, and then go back to what they were doing.

The good news for this person is that we offer a tremendous number of things — \TeX and CTAN have been around for a long time so probably their problem has been solved and probably we hold that solution. But the bad news is that we offer a tremendous number of things, so leading visitors to exactly the materials that will help them is a challenge.

The current text search on <http://tug.ctan.org/search.html> is a good resource, as is the file name search on that page. But online users are also accustomed to two other query forms that we do not now offer, both of which suit swoopers. The first is a list of keywords, while the second is to have choices in a tree, as with the *Open Directory* (<http://www.dmoz.org>).

I have put up experimental versions of these two. Please have a look, but note that the web addresses are temporary. At the present, the keyword page is at <http://az.ctan.org/keyword> and <http://az.ctan.org/characterization> is the characterizations page. If this system becomes the default then those addresses will change to using tug.ctan.org.

For instance, from the keyword page a user can choose “tables” and get a list of links to packages that work with tables. It now finds 76. The user can narrow the results by selecting up to four keywords at a time and getting the intersection of the sets of associated packages.

Characterizations also work in the natural way. One example: a user picks the **primary** tree and then picks **Document types > Books > Publisher styles**, yielding links to our packages that provide publishers’ book classes.

The top characterization page links to two trees besides **primary**. Many packages do more than one thing so in addition to having the main purpose of the package entered on the **primary** tree, other functionality is on the **secondary** tree (a package can be on at most one branch of the **primary** tree but can be on many **secondary** branches). The other

tree is based on Jürgen Fenn’s topical index of the *Catalogue*.

Figure 1 shows a screenshot of a package information page including keyword and characterization information, along with the standard *Catalogue* data.

2 Architecture

The keywords and characterization features are one aspect of a larger overhaul of the tug.ctan.org web pages. This project is named AZ, after my ham radio call sign KE1AZ, and also for the amusement of calling it AZ \TeX .

Behind the web pages is a database, where most of the data is provided by the *Catalogue*. Keyword and characterization data is in addition to that from the *Catalogue*.

I made up the list of keywords based on questions to `comp.text.tex` and `texhax`, and also based on the content of the packages that we hold — if I ran across a number of packages to do something then I guessed that there was a demand for the functionality.

I also made up the **primary** and **secondary** trees, which have the same entries. The **Fenn** tree is scraped from Jürgen’s HTML page.

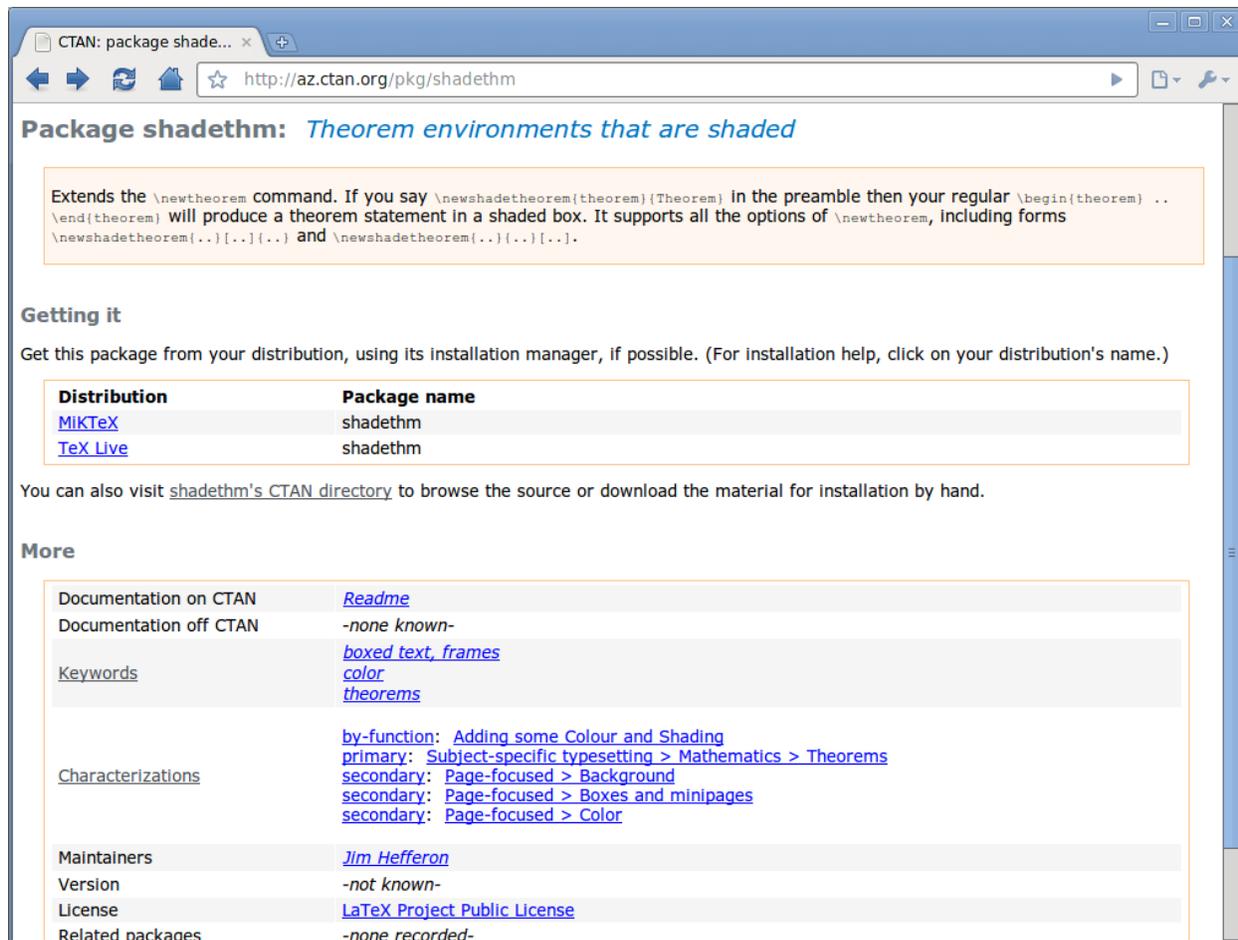
The selection of list entries is influenced by the choice of audience. For instance, a person interested in \TeX development will find that all of the packages aimed at developers have gone into just one keyword and only a few categories. Most of the keywords and categories are for swoopers.

I also tried, in selecting the list entries, to avoid entries with too many associated packages, or too few. This set of keywords and characterizations is actually a second draft. I did a dry run with an earlier set last year and made adjustments to get to the current list. (Some of the entries seem to me to be time-sensitive, but not too many. I hope these sets last for a while.)

3 Accomplishment

CTAN holds about 4000 packages. I entered the keyword and categorization data using a web form, and while I sometimes needed to examine a package closely, more often I used only the *Catalogue* description. Still, even though a typical package was done quickly, the sheer number of packages means that the entire job took a long time. (I entered these in the evening while watching the Celtics play basketball and it took me all season, so the job size is several months of a couple of hours, three nights a week.)

I hope the list can be maintained in part by package authors entering or editing the data via a



CTAN: package shade... x

http://az.ctan.org/pkg/shadethm

Package shadethm: *Theorem environments that are shaded*

Extends the `\newtheorem` command. If you say `\newshadetheorem{theorem}{Theorem}` in the preamble then your regular `\begin{theorem} .. \end{theorem}` will produce a theorem statement in a shaded box. It supports all the options of `\newtheorem`, including forms `\newshadetheorem{...}{...}{...}` and `\newshadetheorem{...}{...}{...}`.

Getting it

Get this package from your distribution, using its installation manager, if possible. (For installation help, click on your distribution's name.)

Distribution	Package name
MiKTeX	shadethm
TeX Live	shadethm

You can also visit [shadethm's CTAN directory](#) to browse the source or download the material for installation by hand.

More

Documentation on CTAN	Readme
Documentation off CTAN	-none known-
Keywords	boxed text , frames color theorems
Characterizations	by-function: Adding some Colour and Shading primary: Subject-specific typesetting > Mathematics > Theorems secondary: Page-focused > Background secondary: Page-focused > Boxes and minipages secondary: Page-focused > Color
Maintainers	Jim Hefferon
Version	-not known-
License	LaTeX Project Public License
Related packages	-none recorded-

Figure 1: Package description including keywords and characterizations, as displayed in a browser.

web form at the time that they upload; this is another feature of the AZ system.

After looking at my keywords and characterizations you may think of another or better list. For instance, you may think that a characterization into “works with L^AT_EX” or “not” or “not applicable” would help. I’m open to suggestions but above I’ve described the amount of scutwork required so that you will know that if you want to suggest entering and maintaining it then you would be committing to a significant effort.

4 Acknowledgements

I would like to thank my CTAN colleagues, Rainer Schöpf and Robin Fairbairns; in particular, Robin has done many years of yeoman’s work on the *Catalogue*. Karl Berry was also invaluable here, as in so many places.

- ◇ Jim Hefferon
Saint Michael’s College
Colchester, Vermont USA
ftpmaint (at) tug dot ctan dot org