Searching for the Ideal MAN Tool

Barbara L. Herron, National Energy Research Supercomputer Center, Lawrence Livermore National Laboratory, Livermore, California

ABSTRACT: *The quantity of online documentation and viewing tools is overwhelming, with the World Wide Web, vendor-supported and local-site documentation and tools, etc. Maintaining the information and tools is equally overwhelming. However, our statistics show that MAN usage far exceeds usage of other online documentation tools. But as we all know, MAN has its own problems, and at the forefront are MAN's many inconsistencies.*

1   Introduction

The National Energy Research Supercomputer Center (NERSC) provides high-performance production computing and networking services to the Energy Research community. The Center meets the needs of its customers by providing super-computer resources and large-scale hierarchical file storage, accessible to the remote customer via a high-speed network. NERSC currently offers our customers use of the computers listed in Table 1.

Table 1. Machines at NERSC as of September 1995

Machine	CPUs	Operating System
CRAY Y-MP C90	16	UNICOS 8
CRAY-2	8	UNICOS 7
CRAY-2	4	UNICOS 7
CRAY T3D[a]	256	UNICOS 8
SGI R8000 Power Challenge	12	IRIX 6.1
SUN SparcCenter 2000	6	Solaris 3.4
Hewlett-Packard 9000/7551	1	HP-UX 9

[a.] The CRAY T3D is located at NERSC, but is financed by the Department of Energy High Performance Parallel Processing Project's Industrial Computing Initiative and by internal funds from within Lawrence Livermore National Laboratory.

We are in the process of procuring a new massively parallel computer. While our procurement is in progress, our customers may obtain limited access to the CRAY T3D.

2   Review of Documentation Tools

Several documentation tools are available for our customers. Some tools are available on the NERSC CRAYs, while some are available from other platforms. This can cause confusion because different information can be obtained from different tools, which are utilizing different information databases (with the exception of the MAN and XMAN tools, which use the same database). Customers will frequently assume the information they are looking for is not available if it is not found using just one of the several documentation tools. Multiple databases of documentation can also cause maintenance problems.

Tools available on machines other than the CRAYs are advantageous because their use does not consume valuable CRAY cycles and disk space. However, requiring users to have an account on another machine used for viewing documentation about use of the CRAY can be confusing and inconvenient. Documentation located on machines other than where the customer is working is likely to go unread because its availability is not known.

The following two sections describe tools available online on the NERSC CRAYs, and those available on other platforms.

2.1   *Online tools available from the NERSC CRAYs*

2.1.1   *DOCUMENT*

At one time, DOCUMENT, a locally developed tool, was NERSC's primary documentation tool. It was used for viewing large locally developed online manuals. DOCUMENT is similar to DOCVIEW, but is not as sophisticated. Entire manuals were printed (to paper or disk) or keyword searches were used to find passages to be viewed or printed. Long lists of keywords were displayed for the user to select. It was often difficult to identify the appropriate keyword for the information being sought. To

view a manual passage, DOCUMENT required the customer to type in the entire keyword, which was sometimes quite a long string. Manuals were frequently copied to a disk file that was searched for a pattern or string (with a favorite search tool, e.g., GREP or an editor) when finding the appropriate keyword was too difficult. Manuals available with DOCUMENT were pure text; inclusion of graphics was not possible. DOCUMENT was removed from all NERSC machines on October 1, 1995, due to lack of use.

2.1.2 DOCVIEW

The DOCVIEW documentation tool provides a method for viewing large online manuals when the printed manual is not readily available. Users search for a particular passage via long lists of keywords. It can be difficult to identify the appropriate keyword for the information being sought. An entire document can be printed to disk, which can be searched online or printed for hardcopy. DOCVIEW is available using a curses or X window display. The X window display contains some inconsistencies that contrast with the behavior of most X window tools (for example, scrolling), and thus several DOCVIEW users prefer the curses mode.

2.1.3 MAN

MAN is the standard Unix (and POSIX) tool we all know and either love or hate. MAN provides good summary information for those already familiar with a command. In some cases, well-written manual pages provide good overall documentation. However, when a particular manual becomes too long, it becomes a cumbersome method to use for reading documentation. In addition there are many such manual pages that are very poorly written.

2.1.4 XMAN

XMAN is an X window manual page display program available in the X11 package. XMAN options are significantly different from other MAN tools. XMAN provides searching, displays pathnames, and provides access to the same set of manual pages used by MAN.

2.2 Documentation tools available on other platforms

2.2.1 CrayDoc (CDOC)

CrayDoc, Cray Research's workstation-based documentation viewer, has not received much use by either NERSC staff or customers. Because CrayDoc is only available on a workstation, most of our customers are not yet aware of it, in spite of our advertising its availability. Most of the large manuals available from DOCVIEW are also available using CrayDoc.

2.2.2 DynaText

Our locally developed large online manuals will soon be available using Electronic Book Technology's (EBT) DynaText. This is the same tool used to display Cray Research manuals via CrayDoc. Using EBT's software, we will be implementing manuals viewable from one of our workstation machines using DynaText, and these same manuals may also be viewed using the World Wide Web (WWW).

2.2.3 The World Wide Web and the browser of your choice

The WWW can be a treasure chest of information. Browsing the Web can also be time intensive and can lead to incorrect, outdated, or offensive information. Useful information, once found, frequently disappears or is moved without a fowarding address. However, if well managed and properly maintained, a WWW server can provide useful information that is viewed by many people around the world.

NERSC converted its local manuals to the HyperText Markup Language (HTML) used throughout the WWW and made them available on the WWW, while we are transitioning to EBT's software. We have restricted access to many of the manuals because of vendor-proprietary information or for security purposes. We are also working on putting the MAN manual pages from each NERSC machine on our WWW server, also with restricted access. Currently, our methods of implementation on the WWW have led to maintaining multiple databases in some cases.

3   Why Search for an Ideal MAN Tool?

Statistics gathered at NERSC show that usage of MAN far exceeds usage of any other documentation tools available on the CRAYs. Table 2 shows access statistics for four online documentation tools from our most heavily used CRAY. As you can see, DOCUMENT usage does not warrant any further effort toward this now outdated tool. And DOCVIEW usage is even lower than DOCUMENT.

Table 2. Statistics of online documentation tool usage on NERSC's CRAY Y-MP C90 over a 391 day period.

CRAY Y-MP C90 Online Information Tools	Unique users per day		Invocations per day		Hours of use per day	
	Average	Percent[a]	Average	Percent[a]	Average	Percent[a]
MAN	32.1	89%	92.4	92%	28.4	96%
DOCUMENT	2.88	8%	6.26	6%	0.61	2%
DOCVIEW	1.17	3%	2.08	2%	0.57	2%
XMAN[b]	0.03	-	0.05	-	35.8	-

a. Indicates the percentage of use for these three CRAY C90 online documentation tools.
b. Usage statistics for XMAN are for a different time period, composed of 183 days. The percentage is not meaningful and therefore not provided.

It is easy to conclude that MAN is the tool customers know how to use and that it is the tool of choice when information is needed. I suspect this is because MAN is available on all Unix platforms, it is conveniently available on the machine on which users are working, and the information is in a familiar format.

4   MAN's Problems

Although MAN is a standard Unix tool (required by the POSIX.2 standard), the options supported among MAN tools on different platforms vary widely, to the point where it is inconsistent across platforms. In addition, MAN needs to be updated to the file formats widely used today.

I compared MAN tools across a number of platforms by executing MAN and studying the MAN manual pages. I have only studied the manual pages in cases where I have not been able to execute MAN (e.g., BSD, Convex, Paragon, and SP2). Table 3 shows the available options, plus some other features, which demonstrate the similarities and differences between MAN tools.

4.1   The POSIX standard

The POSIX.2 standard says very little about MAN. References to MAN in the standard say only:

1. MAN shall exist.

2. MAN shall display the command line syntax for POSIX tools.

3. MAN shall have a **–k** option to search and display a brief purpose for POSIX tools.

4. The PAGER environment variable shall have an effect on the execution of MAN.

5. If the PAGER environment variable is not set, the output shall be filtered through either MORE or another paging utility.

One other interesting item from the POSIX.2 standard within its requirements for MAN is the following quote from section E.5.16:

> *"It is recognized that the MAN utility is only of minimal usefulness as specified."*

To my knowledge, changes to this area of the POSIX standard are not being considered. Further requirements for MAN were reportedly not made because of vendor pressure. However, a tool by the same name that is available on a wide variety of platforms, and one as important as MAN, which displays usage information for most subjects involved in the use of a computer, needs more standardization across platforms. In my survey of several platforms I discovered several options that are almost uniformly provided (for example, use of the MANPATH environment variable, support of NROFF and TROFF formats, and in many cases an option to force display of all of the multiple versions of a manual page with the same name).

4.2   MANPATH

All of the MAN tools I surveyed use the MANPATH environment variable. The Convex tool goes even further by using the PATH variable if MANPATH is not set (which seems a very good use of PATH). However, the POSIX.2 standard does not include a specification and requirement for MANPATH "... *because no attempt is made to specify naming conventions for manual page files, nor even to mandate that they are files at all.*" However, I've yet to see or use a MAN tool that doesn't use files for its manual pages located in `man` subdirectories.

4.3   Multiple manual pages with the same name

Several manual pages for MAN that I read do not specify what action is taken when multiple manual pages of the same name exist. When a tool does not display multiple manual pages, it can be misleading to a user who might believe only one manual entry exists when in fact several exist. Information on how multiple manual pages are handled should be described in all MAN manual pages.

The **–M** *pathname* option is available in many MAN tools and is very convenient when there are multiple manual versions with the same name. On platforms where multiple manual versions are not displayed by the MAN tool, this option (or something like it) are important for locating the proper manual information.

The **–w** option is also available in some MAN tools to list the pathname rather than the contents of manual pages. This option, or some other documented method, is necessary to assist the user to find the appropriate path to use with the **–M** option.

On some platforms, this problem is solved by the use of the **–a** option to force all versions of the requested manual page to be displayed.

I suggest that a better solution would be to list all the multiple versions, then allow the user to choose which of the manuals to view. Those tools supporting the **–w**, **–a**, and **–M** options come close to achieving this feature; but strictly speaking, none of the MAN tools I studied offer this useful feature.

4.4   Manual file formats

Very few of the manuals I read for MAN state anything about what types of file formats may be used. Although some MAN tools require an NROFF format, many will also accept an ASCII text file. Most MAN tools will support output in a TROFF format. A hyptertext language is used with one MAN tool I surveyed. With the wide use of HTML and support and use of SGML looming, these languages are a good choice for a file format which MAN should support.

The Hewlett-Packard MAN tool will recognize and expand compressed files, which can be a truly valuable option. I've never used a computer that had enough disk space (but I have used one computer so limited on disk space that most of the online documentation was removed). So why not support compressed online documentation to free up some valuable disk space?

5   NERSC's Requirements for MAN

In a meeting of interested NERSC staff, the following list of required and desired features for a MAN tool were developed. **Any further input to this list is welcomed.**

5.1   *Required features*

• MAN must meet or exceed the POSIX.2 requirements, which include use of a MANPATH and PAGER environment variables, if they are set.

• A search capability, using at least the **–k** option with a `whatis` file, must be provided.

- MAN, or the paging utility used with MAN, must allow scanning or searching backwards and forwards.

- MAN must recognize and provide a method to display all copies of multiple manual pages with the same name.

- NROFF, TROFF, and ASCII text file formats must be recognized and displayed properly.

5.2 *Desired features*

- If the MANPATH environment variable is not defined, use the PATH variable.

- Allow defining the method (text, curses, X, etc.) used to display the manual pages, using an option that defines the display or automatically determines the method of display (e.g., via the PAGER or TERM environment variables, or using some other method).

- When multiple manual pages of the same name exist, provide a method for choosing one, several, all, or none to be displayed.

- Use multiple `whatis` files (one in every `man` subdirectory if it exists).

- Allow displaying HTML or SGML file formats.

- Recognize and read compressed file formats.

- Allow reading manual pages from another platform or from a foreign file system. For example, using the `-M` option to read some AFS or NFS mounted manual pages.

6   Summary

MAN tools vary widely in the features and options they offer. The existence of a MAN tool is standard, but MAN behavior on different platforms is inconsistent. The POSIX.2 standard does not require many of the favorable behaviors and options used by almost all MAN tools. To my knowledge there are no plans for changing the POSIX.2 standard concerning the MAN tool. MAN is outdated in the file formats it recognizes and in the method of display to the terminal.

NERSC would like a more up-to-date MAN tool to install on the multiple machine platforms we provide, so our customers can rely upon consistency in a distributed computing environment.

7   Acknowledgments

8   Reference

Technical Committee on Operating Systems and Application Environments of the IEEE Computer Society, *IEEE Standard for Information Technology—Portable Operating System Interface (POSIX)—Part 2: Shell and Utilities* (Volumes 1 and 2), IEEE Std 1003.2-1992, Institute of Electrical and Electronics Engineers, Inc. (1992) 558-561, 990.

9   Bibliography

Man [manual pages], Berkeley 6.14, Regents of the University of California (1991).

Man [manual pages], Version 3.0, Tom Christiansen, CONVEX Computer Corporation (1990).

Man [manual pages], SR-2001 8.0, Cray Research, Inc. (September 1994).

Man [manual pages], Hewlett-Packard Company, HP-UX Release 9.0 (August 1992).

Man [manual pages], SunOS 5.4, Sun Microsystems Inc. (February 1994).

Man [manual pages], X Version 11, Release 5, Chris Peterson, Massachusetts Institute of Technology (1993).

"Minutes of the MAN Page Viewers in the NERSC Environment Meeting," Barbara Herron, National Energy Research Supercomputer Center, Livermore, California, private communication (April 6, 1995).

Table 3: Comparison of MAN tool options, part 1.

Platform[a]	Searching	Location	Section	Selection
POSIX.2	`-k` *keyword*			
Berkeley		`MANPATH` or `-M` *path* `-m` *path* (augment the directories searched) `-w` (list pathnames rather than contents)		`-a` (display all manuals that match title for the specified section)
Convex	`-k` *keyword* `-K` *egrep-expression* `-g` *perl-expression* `-f` *keyword* `whatis` database used for searching	`MANPATH` or `-M` *path* or `PATH` `MANALT` (directory to check for alternate sets of man trees) `-w` (list pathname rather than contents) `-l` *pathname* (display specific file)	`[-S]` *sections* `MANSECT` (default ordering for section and subsection sorting)	`-a` (display all manuals that match title) `-h` (ignore the whatis database in searching)
Cray	`-k` *keyword* `-f` *keyword* `whatis` database used for searching	`MANPATH` or `-M` *path* or system default	*section*	Not all manuals that match title are displayed with UNICOS 8 `-i` *keyword* (display summaries from index)
Hewlett-Packard	`-k` *keyword* `-f` *filename* `whatis` database used for searching	`MANPATH` or system default	*section*	Only the first manual that matches title is displayed
Paragon	`-k` *keyword* `-f` *keyword* `whatis` database used for searching	`MANPATH` or `-M` *path* or system default `-w` (list pathname rather than contents)	*section*	
SGI	`-k` *keyword* `-f` *filename* `whatis` database used for searching `-r` (treat names as expressions for searches)	`MANPATH`, `-M` *path*, `-d` *fullpath* or system default `-w` (list pathname rather than contents)	*section*	Uses *, ?, and [...] wildcards to specify *title* Displays all manuals that match the title *title* is case-insensitive
SP2	`-k` *keyword* `-f` *command* `whatis` database used for searching	`MANPATH` or `-M` *path* or system default	*section*	
Sun	`-k` *keyword* `-f` *filename* `windex` database used for searching	`MANPATH` or `-M` *path* or system default `-F` (disable use of windex lookup, force using `MANPATH`)	`-s` *section* `man.cf` file located in each man subdirectory	`-a` (display all manuals that match title) `-l` *title* (list all matches)
XMAN (X11)	`<Ctrl>s` *keyword*	`MANPATH` or system default	displays sections to choose from	pathname of manual page displayed when viewing a manual page

a. Information was obtained from the manual provided for the MAN tool on a selected machine for each of the listed platforms.

Table 3: Comparison of MAN tool options, part 2.

Platform[a]	Output	Formatting	Miscellaneous options
POSIX.2	PAGER or more		
BSD (from GNU)	PAGER or more -c does not pipe through more -h (display only the synopsis)		MACHINE (display manuals only for this architecture)
Convex	PAGER or default selected at installation -D (strips embedded backspace and the preceding characters for output)	-t (format with troff) -T (specify typesetting program) TROFF (specify typesetting program)	-i (displays section and subsection index for title) -v (version) -u (usage) -d (debug)
Cray	PAGER or more -s - does not pipe through more	-t (format with troff) -T macro (specify macro package for formatting) TROFFCMD (specify typesetting program) TCAT (name of utility to print troff output)	-q (displays a quick-reference)
Hewlett-Packard	PAGER or more - does not pipe through more	Recognizes and uncompresses compressed files	
Paragon	PAGER or more - does not pipe through more	-t (format with troff)	
SGI	PAGER or more MANPAGER overrides PAGER -c copies directly to stdout rather than use more, PAGER or MANPAGER	-t (typeset and send to printer) TROFF (specify typesetting program when -t is used) TCAT (name of utility to print troff output) -T macro (specifies macro package for formatting) MANFMTCMD (specify command used to format) -p (prints to stdout commands that would be used to format and display) Formatted entries are in compressed format	
SP2	PAGER or more -s - pipes output through more -s		
Sun	PAGER or more -s	-t (format with troff) TCAT (name of utility to display troff output) TROFF (specify typesetting program) -T macro (specifies macro package for formatting) -r (reformat but do not display)	-d (debug)
XMAN (X11)			Several options and widgets specific only to an X windows tool are provided

a. Information was obtained from the manual provided for the MAN tool on a selected machine for each of the listed platforms.