

PUBLIST - A bibliographic database utility

Ronald F. Peierls

Department of Applied Science, Brookhaven National Laboratory

Introduction

A few years ago, the Department of Applied Science perceived a need to automate activities related to publications by using some computer based system. Among the objectives were that:

- it should be easy for a secretary or someone without extensive computer skills to use the system
- it should run on PCs (at that time DOS based), Macintosh, and Unix systems, so that different groups or individual investigators could use it on their platform of choice
- it should be flexible enough to track evolving views of what information was needed
- it should be able to generate output in different formats for different purposes
- the information should be able to be selected from and sorted by a wide variety of keys
- individual items should be able to be updated with new information or deleted

Since no off-the-shelf package existed at that time, something would have to be written. One possibility would be to use a commercial database package, but that would restrict portability to machines having the package installed, and a user interface would have to be developed. Also, since one publication can have multiple authors, a simple flat database file would be inadequate.

A number of public domain packages for handling bibliographic data exist; *bib*, *refer*, *tib*, *BibTeX*, etc. Although they are generally specific to some particular operating system or formatting language, they are quite mature and have addressed most of the difficult issues of classifying publications. It seemed wise to take as much advantage as possible of their architecture.

The language chosen for implementation was AWK, because of its suitability for manipulating text files, and the existence of a public domain version (*gawk*) available on all the requisite platforms. If the project were to be started today, PERL would probably be a more widely used choice.

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, make any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

Overview of current package

Publication information is stored in an ASCII text file. A blank line followed by a line beginning “%n” starts a new publication. Each line beginning with a “%” starts a new field; the field type is determined by the character immediately following the “%”. The only reserved field types are “%n” identifying a new publication, and “%t” designating the publication type.

Each publication type is defined by a table, listing its allowed field designators, the name associated with each, the order in which they are specified, whether they are required or optional, and their default value if they are required. These tables drive the scripts which control the addition of new entries. In its present character based version, the user is prompted for the content of the relevant fields.

Another set of tables is used to determine the format in which the information is printed out. For example one such style table would be used for an annual report, another for a bibliography for a published article or book, etc. Each table specifies for each publication type, which fields are printed out, in which order, and allows the specification for each field of a character string to be printed before the field and one to be printed after it. This allows punctuation to be inserted.

The present implementation is character based, using a series of shell scripts (BAT files on the PC) to provide a prompt driven interface allowing the user to create a bibliographic file, add publications to it, edit fields in existing publications, merge two such files, extract a subset based on any field criteria, and sort a given file.

Plans for Web interface

Two enhancements which have been desirable from the beginning are the use of a GUI, rather than a prompt based interface, and the use of the output style tables to specify font information as well as punctuation. The difficulty with both of these was portability across different platforms (MS Windows, X-Windows, Mac, etc) and word processors.

The near universal acceptance of the Web browser as a user interface, and HTML as a formatting standard, makes a portable implementation feasible.

The underlying file structure and table specification will be preserved: they have proved flexible and effective, and provide upwards compatibility. The output style tables can be used to specify HTML codes to format fields as needed. The cumbersome incremental prompts now used for data entry or editing will be replaced by a series of forms, with pull down lists to select fields or other attributes. New capabilities may be added which would have been too cumbersome to implement or to use in the character based environment.

The existing system uses a combination of shell (or BAT) scripts and AWK scripts. By converting to PERL, now the most widely used scripting language for Web tools, only one type of script would be needed. As a first step the existing AWK scripts can be converted by the *a2p* utility.

This work was authored under contract number DE-AC02-76CH00016 with the U.S. Department of Energy.

Attachments:

Publication type tables

Sample data file

Output format style table

Formatted output of sample

Instructions for PC version

PUBLICATION TYPE SPECIFICATION

J [J]ournal Article				
key	Prompt	Next Field	Required	Default
A	Author	T	1	
T	Title	J	1	
J	Journal	V	0	
V	Volume	N	0	
N	Number	P	0	
P	Pages	D	0	
D	Date	b	0	
b	BNL_No	d	1	#####
d	Division	m	1	***
m	Month_Entered	Q	1	—
Q	End	X	1	Yes
S	Series	N	0	
a	Annual_Report_Year	Q	0	
f	Funding_Source	Q	0	
c	B&R_Category	Q	0	
K	Search_key	Q	0	
g	Group	Q	0	

P [P]roceedings of Conference				
key	Prompt	Next Field	Required	Default
A	Author	T	1	
T	Title	C	1	
P	Pages	D	0	
D	Date	b	0	
b	BNL_No	d	1	#####
d	Division	m	1	***
m	Month_Entered	Q	1	—
Q	End	X	1	Yes
L	City_of_Publication	D	0	
C	Conference_Title	L	0	
E	Editor	P	0	
I	Publisher	L	0	
S	Series	N	0	
N	Number	Q	0	
a	Annual_Report_Year	Q	0	
f	Funding_Source	Q	0	
c	B&R_Category	Q	0	
K	Search_key	Q	0	
g	Group	Q	0	

A [A]bstract or presentation at conference				
key	Prompt	Next Field	Required	Default
A	Author	T	1	
T	Title	C	1	
D	Date	b	0	
b	BNL_No	d	1	#####
d	Division	m	1	***
m	Month_Entered	Q	1	—
Q	End	X	1	Yes
.L	Conference_Location	D	0	
C	Conference_Title	L	0	
a	Annual_Report_Year	Q	0	
f	Funding_Source	Q	0	
c	B&R_Category	Q	0	
K	Search_key	Q	0	
g	Group	Q	0	

T [T]echnical Report				
key	Prompt	Next Field	Required	Default
A	Author	T	1	
T	Title	R	1	
D	Date	b	0	
E	Editor	D		
b	BNL_No	d	1	#####
d	Division	m	1	***
m	Month_Entered	Q	1	—
Q	End	X	1	Yes
R	Report#	E	0	
S	Series	N	0	
N	Number	Q	0	
a	Annual_Report_Year	Q	0	
f	Funding_Source	Q	0	
c	B&R_Category	Q	0	
K	Search_key	Q	0	
g	Group	Q	0	

I [I]nformal Report				
key	Prompt	Next Field	Required	Default
A	Author	T	1	
T	Title	R	1	
D	Date	b	0	
b	BNL_No	d	1	#####
d	Division	m	1	***
m	Month_Entered	Q	1	—
Q	End	X	1	Yes
.R	Report#	D	0	
S	Series	N	0	
N	Number	Q	0	
a	Annual_Report_Year	Q	0	
f	Funding_Source	Q	0	
c	B&R_Category	Q	0	
K	Search_key	Q	0	
g	Group	Q	0	

B [B]ook				
key	Prompt	Next Field	Required	Default
A	Author	B	1	
D	Date	b	0	
b	BNL_No	d	1	#####
d	Division	m	1	***
m	Month_Entered	Q	1	—
Q	End	X	1	Yes
B	Book_Title	I	0	
I	Publisher	L	0	
L	City_of_Publication	D	0	
E	Editor	P	0	
S	Series	N	0	
N	Number	Q	0	
a	Annual_Report_Year	Q	0	
f	Funding_Source	Q	0	
c	B&R_Category	Q	0	
K	Search_key	Q	0	
g	Group	Q	0	

C [C]hapter in Book				
key	Prompt	Next Field	Required	Default
A	Author	T	1	
T	Title	B	0	
P	Pages	I	0	
D	Date	b	0	
b	BNL.No	d	1	#####
d	Division	m	1	***
m	Month_Entered	Q	1	—
.Q	End	X	1	Yes
B	Book_Title	E	1	
I	Publisher	L	0	
L	City_of_Publication	D	0	
E	Editor	P	0	
S	Series	N	0	
N	Number	Q	0	
a	Annual_Report_Year	Q	0	
f	Funding_Source	Q	0	
c	B&R_Category	Q	0	
K	Search_key	Q	0	
g	Group	Q	0	

%n ADM006
%t J
%A Laszlo, Pierre
%A Petrakis, Leonidas
%T On risk
%J Chemistry in Britain
%b 48135
%d ADM
%m 92-11
%f 05004
%c KP-05

%n APS071
%t A
%A Johnson, Brant M.
%T Atomic physics research
%C 1992 DOE/BES Atomic Physics Program Workshop
%L Cornell University, Ithaca, NY
%D October 15-16, 1992
%b #####
%d APS
%m 92-07
%f 05458
%c KC-03-01-03

%n APS072
%t A
%A Rivers, M. L.
%A Sutton, S. R.
%A Jones, K. W.
%T Introduction to synchrotron radiation, synchrotron sources and techniques.
%C 1992 ASA-CSSA-SSSA Annual Meeting
%L Minneapolis, MN
%D November 1-6, 1992
%b 47778
%d APS
%m 92-07
%f 05455
%c KC-03-02-02

%n ASD089
%t I
%A Xi, Xiaolin

%T Energy Development and CO2 Emissions in China
%D June 1993
%b #####
%d ASD
%m 93-06
%f 05166
%c PE-04-01

%n ASD056
%t J.
%A Morris, S. G.
%T Trends in quantitative cancer risk assessment.
%J Environ. Health Perspec.
%V 90
%P 297-98
%D 1991
%b #####
%d ASD
%m 92-07
%f 05139
%c HA-01-15

%n ASD058
%t J
%A Marr, Robert B.
%A Pasciak, Joseph E.
%A Peierls, Ronald F.
%T Parallel computation with remote procedure requests using
structured messages.
%J Concurrency Practice and Experience
%b #####
%d ASD
%m 92-08
%f 05290

%n BPS049
%t A
%A Long, S. P.
%A Nie, G-Y.
%A Baker, N. R.
%A Drake, B. G.
%A Farage, P. K.
%A Hendrey, G. R.
%A Lewin, K. H.

%T The implications of concurrent increases in temperature, CO₂
and tropospheric O₃ for terrestrial C₃ photosynthesis.

%C IX International Congress on Photosynthesis

%L Nagoya, Japan

%D August 30-September 6, 1992

%b #####

%d BPS

%m 92-07

%c KP-05

###

SAMPLE STYLE TABLE

Publication typeJ		
Field Order	print after	print before
A	,	
T		
J	,	
V	,	
N	,	
P	,	
D	.	

Publication typeP		
Field Order	print after	print before
A	,	
T		
C	,	In
E	,	Ed. by
I	,	
L	,	
P	,	
D	.	

Publication typeA		
Field Order	print after	print before
A	,	
T		
C	,	Presented at
L	,	
D	.	

Publication typeT		
Field Order	print after	print before
A	,	
T		
E	,	Ed. by
R	,	
N	,	No.
D	.	

Publication typeI		
Field Order	print after	print before
A	,	
T		
D	.	

Publication typeB		
Field Order	print after	print before
A	,	
B	,	
I	,	
L	,	
D	.	

Publication typeC		
Field Order	print after	print before
A	,	
T		
B	,	in
P	,	
I	,	
L	,	
D	.	

FORMATTED OUTPUT OF SAMPLE FILE

Laszlo, Pierre, Petrakis, Leonidas, On risk Chemistry in Britain,

Johnson, Brant M., Atomic physics research Presented at 1992 DOE/BES
Atomic Physics Program Workshop, Cornell University, Ithaca, NY,
October 15-16, 1992.

Rivers, M. L., Sutton, S. R., Jones, K. W., Introduction to
synchrotron radiation, synchrotron sources and techniques.
Presented at 1992 ASA-CSSA-SSSA Annual Meeting, Minneapolis, MN,
November 1-6, 1992.

Xi, Xiaolin, Energy Development and CO2 Emissions in China June 1993.

Morris, S. G., Trends in quantitative cancer risk assessment. Environ.
Health Perspec., 90, 297-98, 1991.

Marr, Robert B., Pasciak, Joseph E., Peierls, Ronald F., Parallel
computation with remote procedure requests using structured
messages. Concurrency Practice and Experience,

Long, S. P., Nie, G-Y., Baker, N. R., Drake, B. G., Farage, P. K.,
Hendrey, G. R., Lewin, K. H., The implications of concurrent
increases in temperature, CO2 and tropospheric O3 for terrestrial
C3 photosynthesis. Presented at IX International Congress on
Photosynthesis, Nagoya, Japan, August 30-September 6, 1992.

DAS PUBLICATION LIST PROGRAM

PUBLICATION LISTS

Publication lists are text files. Publications are separated by blank lines. Each publication has a number of different fields, each field containing a particular type of information about the publication, such as author, journal, publication date, etc. A line beginning with “%” starts a new field. The next character after the “%” selects the type of field and is followed by a blank. The remainder of the line (together with any following lines up to the beginning of the next field) forms the content of that field, its *value* for that publication. The fields can appear in any order, except that the serial number (the “%n” field) must come first. A publication may have more than one field of the same type, (for example, multiple authors). Where multiple values do not make sense, (for example, BNL numbers) the last occurring value is used and the rest ignored.

PUBLICATION TYPES

One field required for each publication is the “%t”, or *type* field. The publication type (e.g. journal article, book chapter, informal report) may affect the format and order in which a list is printed for different purposes. For each type of publication there is a set of fields which are appropriate. Some may be required, some optional. Required fields may have default values associated with them which will be used if no other value is supplied.

ENTERING PUBLICATIONS

A publication list may be prepared as an ASCII text file using any word processor, being careful to maintain consistency between the publication types and field values, but it is more easily prepared using the interactive program supplied. The program begins by asking for the publication type, supplying a list if queried. It will issue a prompt with the name of a field to enter. Some fields are required, others are optional. The "%n" field is generated automatically. The order in which the prompts occur, and the default values for required fields are determined by a table which can easily be changed. (The currently used table is attached.) In response to the prompt the following entries are possible:

(Enter) If the field is required the default (shown in square brackets at the end of the prompt) will be used, otherwise the field will be skipped.

^K (Control-K) The current publication entry will be killed.

? The list of fields to date will be displayed.

^X (Control-X) The program will be aborted

=x The prompt will change to field x. If x is omitted the last previous field will be used.

any other text The text will be used for the prompted field. If the line ends with "\" then no new field prompt is issued but the next line will be added to the list as a continuation line for the current field.

The "%Q" field does not appear in the output, but just represents a chance to verify the entry .

PC SYSTEM

1. Insert disk in drive[A]
2. type:
A: <Enter>
3. To start a new list type:
create (*list file*)<Enter>
4. To add more publications to a list type:
add (*list file*) (*revised list file*) <Enter>
(If (*revised list file*) is omitted OUT will be used as a default.)
5. To edit a list type:
edit (*list file*) (*revised list file*) <Enter>
(If (*revised list file*) is omitted OUT will be used as a default.)
6. To format the list type:
fprint (*list file*) (*formatted list file*) <Enter>
If (*formatted list file*) is omitted, then the formatted output will be displayed on the screen.
7. To print the a list file or formatted list file type:
print (*file name*) <Enter>