

B. R. Schneider, Jr  
English Department  
Lawrence University  
Appleton, WI 54912  
90-2-3000-635

es. Inc.

FIRS  
CH



0.75

PB METER  
7152259  
U.S. POSTAGE

~~Professor John C. Greene  
English Department  
University of Southwestern Louisiana  
Lafayette, LA 70504-4691~~

FIRST  
CLASS

**Do not bend or fold.**

SITAR

CH

LSTB Sample  
Test.npp Sitar.exe  
Sitar.DOC Sitar.fil

Dysan®

DOUBLE SIDED • DOUBLE DENSITY  
48 TPI  
REORDER #802060

 Dysan®  
CORPORATION

3 April 1990

To view sample "type test.npk / more" at DOS prompt  
SITAR

SITAR is a System for Interactive Text Analysis and Retrieval that has some advantages over commercial packages and some disadvantages. Unlike most commercial packages, it is totally independent of the structure of the file it is searching. It is not, for instance, restricted to retrieval of items by line, unless you want it to be. It is sensitive to ordinary features of prose, such as punctuation and paragraphing, as well as to any special features of the text in hand, such as the protocols of a bibliography entry. It enables the user to control the context in which he wishes a match to appear, by choosing where he wants the context to begin and end. It doesn't require any pre-processing of the text file, but you may find that some pre-processing will expedite the work in hand. It has a global search and replace capacity not confined to replacing a letter, word or phrase, but everything within declared boundaries. It will, for instance, replace any sentence containing a given key word or any sentence containing a number. Or, if you wish, it will enable you to edit every string that your search pattern finds.

It can also replace all but certain characters masked in the replacement. It uses six different kinds of "wild cards," characters that will match on a given set of characters, like any number. For instance, if you want to find all the zip codes in an address list, you can specify five number codes in a row.

Its main disadvantage is that it searches sequentially, checking every character from beginning to end until it finds a match. It doesn't work from an index to the text, by which commercial packages go directly to a match instead of searching up to it. Therefore, I suspect it is considerably slower (though not slower than WordPerfect SEARCH); but its loss in speed is a gain in freedom and power. By means of REPLACE you can number lines, paragraphs, or pages, and then write patterns that will include the numbers.

SITAR retrieves from or replaces in a text file four basic kinds of patterns:

```
key
key-
·key
key-[string]
[string]-key-[string]
```

In this representation, "..." stands for the ellipsis, normally written "...", which indicates that text intervening between the other elements is being ignored. This dot is invoked by the ESCape key, and by means of it we "escape" the text between the boundary strings and the key. The word "key" designates the character string for which SITAR searches first. If SITAR finds a match to the key, it then supplies the "hit" (matched key) in a context determined by the "[string]" delimiting the left and right boundaries of the context. Brackets in the representation above indicate that these left and right strings are optional. If no left or right delimiters are stated, the context on either side is theoretically all the text there is on either side of the key. That much context is just what we have to begin with, so SITAR

## How it Works

To start the program, type "sitar" at the DOS prompt. When the program comes up it displays the following masthead across the top of the screen:

```
SITAR: Copyright (c) 1987, 1988, by B R Schneider, Jr, Lawrence University.  
Block 45000 Limit 1000 Set xyz Mode n Infile C:\xyzxyzx.lmn Outdir  
\abc\bca\
```

To change a setting type 1st letter of parameter:

You are now in a position to change any item in the list of settings under the copyright statement:

**BLOCK:** The block size, given in bytes, is the number of bytes of text that SITAR takes from the disk file each times it goes to get some more for searching. The absolute limit is 65528 bytes, the longest string allowed under DOS. Since files may grow in size in REPLACE mode, 45000 is relatively safe block size. SITAR output is in byte files but since byte mode does not respond to any codes in a file (it merely records them), it can take any file as input. E.g., if SITAR takes in a text file, the CR/LFs will appear as codes on your screen. If at settings prompt, you type "b", the program responds, "**New block size?**" waits for you to enter a new figure. Do not use any commas or decimals.

**LIMIT:** This variable designates the amount of text that SITAR will display if a pattern match exceeds this number. If, after a hit on the key, the program finds that either the left or right delimiter is farther away than the limit you have designated, the program chops that side of the string hit to fit the limit. If at the settings prompt you type "l", the program responds "**New limit?**" The full screen of a typical monitor screen contains 2000 bytes (25 lines x 80 characters).

Since this limit applies to both left and right extensions from the key, if you set "l" to 1000, each hit will be limited to a screenfull. When a left or right context exceeds the specified limit, SITAR chops off the context at the specified limit and prints a special double bar in the output to indicate that it has done so.

**SET:** By means of this parameter you can make up your own wild cards, using any of the wild cards already defined plus any list of matching characters you wish to add. When you use the set character "f" (created by CTRL/S) in a key or delimiter string SITAR will match on any member of the set in that position in the string. For example, suppose you wish to capture all entries of an address list that have a ZIP beginning with 54 and have a 7, 8, or 9 in the 3rd digit. At the settings prompt, press "s" and answer the "**New set?**" question with "789". Then write the key of the pattern request thus: "54fnn". ("n", CTRL/N, stands for any number.)

**MODE:** SITAR operates in three basic modes with several options in each. At the settings prompt, when you press "m" the following question pops up:

```
New mode? (n[ormal], r[esearch], x, s [s or p])
```

You now have the following options:

"n" produces normal or essential SITAR operation, in which the program searches for matches on the pattern statement and simply displays them on the screen. If you choose this option, SITAR answers "Pattern?" After you answer with a pattern statement like those described above, the program supplies a screenfull of hits and then asks "More?" if you answer "y", it gives you the next screenfull. If you answer "n" it stops displaying hits but continues to save them. If you answer by hitting the RETURN (ENTER) key the program cancels the search and returns you to the pattern statement.

"nx": same as above, but unpacks dates in London Stage files and displays them in full at the beginning of each section in the pattern match.

"r" searches the file just as in normal mode but enables editing of the hits in three ways. When you choose this option the program asks, "Pattern to be replaced?" You reply with a pattern statement as above. It then asks, "String to replace it with?", and here's where the options occur. If you simply want to replace "Henry" with "Charles" every time "Henry" occurs then you write "Henry" in answer to the first question and "Charles" in answer to the second. Unless you type "n" at the "More?" prompt, the replace routine will display all the revised output a screenfull at a time.

To replace only part of a string, mask the part that should remain as is with CTRL/X, which will appear as a solid block in the pattern. Warning: if you replace a pattern with a longer string and the masking character follows the expanded part, a mask won't protect the corresponding part of the search pattern. (Of course, but I found this out the hard way.)

If you wish to edit the hit when you get to it, respond "edxx" to the second question. The program will display the text on the screen and enable you to move the cursor, insert, and delete. Dire will result if the pattern match is longer than one screenfull. If that is so, cancel (CTRL/BREAK), change settings, change delimiters.

"pkxx" at the "String to replace it with?" prompt invokes a special London Stage routine that changes all partial dates to full dates, and packs them into two bytes at the head of every section. Use the pattern statement "\*.*f*" in which "*f*" stands for the set "Ü." (In most sections the date comes between the beginning of the section and the first letter of the title, the first Upper-case letter. But there are some sections without any caps at all--"*d*." and the date should be plugged in between the section code and the period.)

WARNING: patterns with ellipses may cause SITAR to search back beyond the previous hit to find a matching delimiter, thus damaging the file. In this case the program chops off the excessive portion and prints a double bar in the text at that point to let you know it has done so. Since this double bar indicates that you haven't completely replaced the pattern, don't save the output. Or as soon as you see the double bar, cancel (CTRL/BREAK), revise your pattern and try again. It may be that you simply cannot do what you intend to do.

"x" invokes The London Stage "eXplode" routine which makes sort records for every performance of some thing by someone, and every name or title marked in parenthetical text. It comes up with a "Pattern?" request, as in normal mode, and you must state "\*\*\*", which brings in one section at a time for processing.

"xs" Shows each exploded section, one at a time. Use RETURN to pass from one to the next.

"xp" sends exploded output to the Printer.

"s" Selects only those exploded records that contain a given key. The pattern used in this option is "\*KEY\*".

"ss" Shows each it on the screen.

"sp" prints each hit.

**INPUT:** If you wish to change the input directory and file, press "i", which produces the question, "New input directory & file?" and prints the current drive/directory. You may edit this information as you wish.

**OUTPUT:** When you press "o", the query "New output directory?" appears. At this point you can change only the directory. You give the file name after the program has finished and you know what you've got.

When you have made all the settings changes you wish to, press RETURN and SITAR will go on to the next step. These settings are stored in SITAR.FIL when you exit the program. SITAR.FIL must always be in the same directory as SITAR.

When SITAR completes the requested job, it reports the number of hits and asks "Save output? (y or n)" RETURN or "n" returns you to the pattern statement without saving the output. "y" results in "Output file name?" If the file you choose already exists, you are warned and you can either write over the existing file or name a new one for the output.

WHEN YOU RESPOND TO ANY SITAR REQUEST BY PRESSING return, THE PROGRAM REVERTS TO ITS PREVIOUS QUESTION. TO EXIT SITAR, PRESS return AT THE PATTERN PROMPT. To return to the settings from the pattern prompt press the F1 key.

CTR<sub>b</sub>/Break stops pgm while parsing

At prompt F1 stops PGM

### Key to London Stage Coding

\*p begins entry for a performance at a theatre on a day. Date and theatre follow this heading before the contents of performance begin.

\*a signals afterpiece, usually short, farcical, and/or musical

\*c signals a comment on the day's performance

\*m signals musical performance entry

\*d dancing

\*e entertainment, like leaping over 11 barrels

[xyzzyzx] material bracketed in original source, The London Stage,  
1660-1800

[xyzzyzx%] material bracketed by London Stage Project to enable extraction of casts.

\$name= name in bracketed text or comments

+title= title in bracketed text or comments

\*S - singing and then one others

D R A F T

AGREEMENT: Southern Illinois University Press and  
Lawrence University

Southern Illinois University Press grants Lawrence University exclusive rights to produce and maintain any computer-readable form of *The London Stage, 1660-1800* (edited by William Van Lennep, Emmett L. Avery, Arthur H. Scouten, George Winchester Stone, Jr., and C. Beecher Hogan, first edition, 5 parts, 11 volumes, copyright Southern Illinois Press, Library of Congress catalogue card 60-6539) to be used for the purpose of mechanical information retrieval as an aid to scholarly research; in return for this privilege Lawrence agrees to produce and maintain a London Stage Information Bank and to make available to Southern Illinois University Press and to any academic researchers any set of information contained therein which is amenable to a mechanical search at a price of no more than Lawrence's cost for each information search.

This exclusive franchise lapses if Lawrence does not produce an operating information bank by July 1, 1973.

In the event that Southern Illinois University Press desires to have produced and maintained an information bank consisting of the Biographical Dictionary of London Performers, 1660-1800, now being compiled by Edward A. Langhans, Philip Highfill and Kalman Burnim and scheduled for publication by Southern Illinois University Press, it will grant Lawrence University the first opportunity to obtain exclusive rights to do so.

- Friday 25*  
DL DANCING. *Four French Peasants* by Weaver, Wade, Shaw, Topham.  
COMMENT. Written by Beaumont and Fletcher.
- LIF CIMBELINE. *Cast not listed.* Also *THE JEALOUS DOCTOR*; or, *The Intriguing Dame*. Fossil - Boval; Plotwell - Thurmond Jr; Underplot - Lun [J. Rich]; Ptisan - Griffin; Mrs Townly - Mrs Schoolding.  
*DANCING.* By du Pre and Mlle Gautier, from the Opera at Paris, being the first time of her appearing upon the English Stage.  
COMMENT. At the Desire of several Persons of Quality. Mainpiece: Written by Shakespear, and since revis'd with several Alterations. Admission: 5s., 3s., 2s. Benefit John and Christopher Rich. Receipts: £61 8s.
- Saturday 26*  
DL THE ORPHAN; or, *The Unhappy Marriage*. Castalio - Wilks; Polydore - Booth; Chamont - Mills; Acasto - Boman; Chaplain - Cibber; Monimia - Mrs Porter; Page - Miss Lindar. Also *THE LOVES OF MARS AND VENUS*. *Cast not listed*, but see 12 Oct.  
COMMENT. At the Desire of several Ladies of Quality. Mainpiece: With an entire Set of New Scenes from the Opera never us'd here before, and proper to the Play.
- LIF THE FATAL MARRIAGE; or, *The Innocent Adultery*. Baldwin - Keene; Biron - Husband; Carlos - Cory; Villeroy - Leigh; Frederick - Bullock Jr; Fernando - Griffin; Jaquelin - Pack; Sampson - Bullock Sr; Isabella - Mrs Rogers; Julia - Mrs Bullock; Victoria - Mrs Thurmond; Nurse - Mrs Kent.  
*DANCING.* As 24 Oct.
- Monday 28*  
DL THE STRATAGEM. Aimwell - Mills; Archer - Wilks; Sullen - Bickerstaff; Gibbet - Cibber; Foigard - Bowen; Scrub - Norris; Boniface - Cross; Mrs Sullen - Mrs Porter; Dorinda - Mrs Mountfort; Cherry - Mrs Bicknell.
- LIF THE EMPEROR OF THE MOON. Scaramouch - Bullock; Harlequin - Spiller. Also *THE COBLER OF PRESTON*. Cobler - Spiller.  
*DANCING.* As 22 Oct.  
COMMENT. Mainpiece: With all the Scenes, Machines, and Decorations proper to the Play.
- Tuesday 29*  
DL THE LANCASHIRE WITCHES. *Cast not listed.* Also *THE WHAT D'YE CALL IT*.  
*DANCING.* By Weaver, Wade, Shaw, Topham, Mrs Bicknell.  
COMMENT. Mainpiece: With all the Songs, Dances, and other Decorations proper to the Play.
- LIF THE CITY POLITICKS. *Cast not listed.* Also *THE JEALOUS DOCTOR*. Punch - Cook; Scaramouch - Thurmond Jr; Harlequin - Moreau; Pierrot - Griffin; Columbine - Mrs Bullock; but see 25 Oct.  
*DANCING.* As 22 Oct.
- Wednesday 30*  
DL THE JOVIAL CREW; or, *The Merry Beggars*. *Cast not listed*.  
*SINGING.* As 25 Oct.  
*DANCING.* By Prince, Birkhead, Wade, Miller, Mrs Santlow, Mrs Bicknell, Mrs Willis. A *Cbacone*, a *Minuet*, and a *Jigg* by Miss Lindar, being the first time of her Dancing on any Stage.
- LIF A WOMAN'S REVENGE; or, *A Match in Newgate*. *Cast not listed.* Also *HOB*; or, *The Country Wake*. *Cast not listed.* Also *THE JEALOUS DOCTOR*. *Cast not listed*, but see 29 Oct.