

3, are separated. Meanwhile, compound III, which does not belong to that class, was placed in the cluster with compound II, a more expected error of type M.

A traditional form of categorizing this set of compounds was performed as follows. The 517 cluster leaders of clusters bigger than size 9 were perused to identify chemically central features with large representation. These 25 or so features were then used for substructure search on the whole file of 12 900 compounds.

The 25 classes obtained in this way ranged in size from 100 to 400 compounds and together comprised about 40% of the 12 900 compounds. By a comparison to Table V, we see that these classes are broken up a lot. Even at the 20% threshold where the clusters are not so pure, it is clear that those sizes cannot be accommodated. So, the machine clustering can be used as a first step in a more traditional classification procedure. It may be that substantially more extensive categorization would be difficult due to diversity.

From this study we shed some light on the difference between mechanical and traditional clustering. Again we see

it is difficult to classify a set of diverse compounds.

REFERENCES AND NOTES

- (1) Hodes, L. Clustering a Large Number of Compounds. 1. Establishing the Method on an Initial Sample. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 66-71.
- (2) Boyd, M. R. National Cancer Institute Drug Discovery and Development. In *Accomplishments in Oncology*. Vol. 1, No 1. *Cancer Therapy: Where Do We Go From Here?* Frei, E. J., Freireich, E. J., Eds.; J. B. Lippincott: Philadelphia, 1986; pp 68-76.
- (3) Whaley, R.; Hodes, L. Clustering a Large Number of Compounds. 2. Using the Connection Machine. *J. Chem. Inf. Comput. Sci.* **1991**, *31* (preceding paper in this issue).
- (4) Willett, P.; Winterman, V.; Bawden, D. Implementation of Nonhierarchical Cluster Analysis Methods in Chemical Information Systems: Selection of Compounds for Biological Testing and Clustering of Substructure Search Output. *J. Chem. Inf. Comput. Sci.* **1986**, *26*, 109-118.
- (5) Lawson, R. G.; Jurs, P. C. Cluster Analysis of Acrylates to Guide Sampling for Toxicity Testing. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 137-144.
- (6) Hodes, L. Computer Aided Selection for Large Scale Screening. In *Comprehensive Medicinal Chemistry*. Vol. 1. *General Principles*; Hansch, Sammes, Taylor, Eds.; Pergamon Press: New York, 1990; pp 279-284.

COMPUTER SOFTWARE REVIEWS

Winword

CLEMENS JOCHUM

Beilstein Institute, 40-42 Varrentrappstrasse, D-6000 Frankfurt, Germany

Received January 30, 1991

I have to confess that I have always been a Wordstar¹ addict. When I started working with personal computers in the late seventies under the CP/M operation system, Wordstar was the only decent word-processing program around. When I switched to IBM-compatible DOS-based PC's I was glad that Wordstar was also ported (although it worked slower than on my 4-MHz Z-80 machine!) and I did not have to learn a new word-processing program.

In the following years the competition for Wordstar became continuously stiffer. Within a short time other word-processing programs such as Wordperfect,² Microsoft Word (for DOS),³ or Chemtext⁴ had surpassed Wordstar in features and speed. They did not, however, offer so many more features that I considered it worthwhile switching. (Apparently Wordstar Corp. was convinced that they had only customers like myself, i.e., who are too lazy to learn a new word-processing program, so the company never did very much to improve Wordstar.)

During the last several years my colleagues and I tried out various other word-processing programs, but, as mentioned above, we could not convince ourselves to switch. Having a large local area network (LAN) with several servers, a switchover would mean that all other word-processing users at the Beilstein Institute would have to standardize on a new program, all Wordstar files would have to be converted, etc.

My initial experience with Word for Windows (Winword) version 1.0⁵ was mixed. On one hand I liked all the features that make Winword not only a very good word-processing program, and it also contains many features usually found only in Desktop publishing (DTP) programs. On the other hand Windows 3.0 was not yet available and running it under

Window-286 or Windows-386 (version 2.11) was cumbersome since one was constantly running out of memory. The whole situation changed with the release of Windows 3.0. This graphics user interface (GUI) and Winword (now version 1.1) are such a powerful combination (provided that you have the right hardware) that I finally said goodbye to Wordstar.

Let me begin with the negative points: You need fairly powerful hardware to run Winword efficiently. You should have at least a 12-MHz 80286-based computer (it is more fun to use Winword on a 20-MHz 80386 machine) and at least 2 MBytes of RAM. 4-8 MBytes of (extended—not expanded!) memory make all Windows-based programs much more efficient since several MBytes can be assigned to Windows disk-caching device driver Smartdrive. This speeds up Winword considerably. This review was written with Winword on a Compaq SLT-286 with 3.5 MBytes of RAM whereby 1 MByte was assigned to Smartdrive. The speed is absolutely sufficient for my two-finger-typing.

Another negative aspect of the program is that the very powerful Basic-like Macro language is not described in the manual. It is only described in the technical manual which has to be ordered separately. If you are an experienced Basic programmer you can figure out most of the language by simply using the sample macros. The macros also allow you to completely customize the user interface. You remove any command from any pull down menu or add other ones which you have written yourself with the Word-Basic macro language. This way you can customize one version for your secretary, one for scientific writing, one for programming, etc. However, I would not recommend changing the pull down

menus too much since they would then no longer correspond to the manual or the tutorial.

What makes Winword so different from other word-processing programs? First it is completely WYSIWYG (what you see is what you get). You can combine graphics with your text. Winword comes with import filters for most graphics formats such as Tiff, PCX, HPGL, and so on. Unfortunately there is not yet a filter for EPS (encapsulated postscript) format. If you have graphics which are not supported by Winword, a program called HiJaak⁶ helps to convert practically any graphics format into almost any other one.

If you decide to use Winword you have probably used another word-processing program before. It is therefore very important that you are able to import your former word-processing files into Winword. Microsoft's new word processor comes with filters for most word-processing programs such as Wordperfect, MS Word for DOS, Multimate, and so forth. It also comes with a filter for Wordstar. But only for the older versions 3.3 and 4.0 and not for the new version 5.0 whose format differs significantly from 4.0. To convert my Wordstar 5.0 files I first had to use Wordstar to export its 5.0 files into 4.0 format before I could import them into Winword.

A more preferable way to convert your files is to buy a program called Software Bridge⁷ which includes virtually any word-processor format you can imagine. Unfortunately, Software Bridge does not yet support the Winword .DOC format directly. So, you can either convert your format to Microsoft's ASCII transfer format RTF or to MS-Word 5.0. Strangely enough a Software Bridge converted RTF-file was imported without any problems by Winword. Pagemaker⁸ (see review⁹), however, did not accept it. On the other hand, if I exported a file from Winword in RTF-format I had no problem to import it into Pagemaker. Are there different versions of RTF??

For a chemist it is most important that text and graphics be combined easily. With Winword this can be accomplished in several ways. If you use a DOS-based chemical structure drawing program you have to store the structure on disk in one of the above-mentioned graphics formats. If you are using a Windows-based chemical drawing program, importing structures is even easier. For example, one creates a chemical structure using ChemWindow.¹⁰ The structure can then be transferred to the clipboard by marking with the left mouse button and transferred into the Windows Clipboard using the Edit Cut or Edit Copy command. Next you switch to Winword and move the cursor to the place where you would like to have the structure. Use the Edit Paste command to insert the structure at this place. If your screen is large enough and has a sufficient resolution you can run Winword and ChemWindow side by side. Although this is not a ChemWindow review I can strongly recommend this program because it is extremely simple to use and one can draw very fine looking structures.

Winword has very few drawbacks when compared with a standard DTP program such as Pagemaker. There are no drawing tools in Winword and you cannot let text flow around graphics in as many ways as with good DTP programs. You also have less possibilities in drawing frames around text or titles. When I produce overhead transparencies (usually in the last minute before a talk) I still prefer Pagemaker over Winword. Overhead transparencies usually contain comparatively little text so I do not miss the Spelling Checker too much. On the other hand the drawing and framing tools come in very handy particularly when you produce an overhead transparency.

Before we acquired Winword we used Wordstar in combination with Pagemaker for producing manuals. Wordstar was used to type and edit the text. With Pagemaker the

graphics were inserted at the right place. There was always a problem if some major text changes had to be made after all the graphics had been inserted. Although Pagemaker allows you to insert text, it does not contain a spelling checker or a string search and replace command. (According to Pagemaker's producer, Aldus, this is going to change with the upcoming Pagemaker version 4.0 which will be out soon and which will contain all these features.) With Winword we do not need Pagemaker for the "finishing touch". As described above, graphics and text can be mixed without any problems.

Winword also has an extremely convenient feature for producing a table of contents and outlines. You mark all your headings and subheadings of an article or manual with a special indent. For example, headings are marked with a left indent, subheadings with a right indent, sub-subheadings with two right indents and so on. After this has been done, one simply chooses the right place for the table of contents and generates it with one key stroke. The page numbers and titles in the table of content section are automatically updated whenever you update the text or a (sub) heading. Through the Windows dynamic data exchange feature (DDE) you can link other programs directly with Winword. For example, you can insert a Excel spreadsheet into the text and link it through DDE with Excel. Whenever you update the spreadsheet with Excel, it will automatically be updated in your text.

It is almost needless to say that Winword offers all the standard word-processing features: underline (word underline, solid underline, double underline), boldface, italics, superscripts, subscripts, spelling checker, thesaurus and so on. You can either work in true WYSIWYG mode or in a (faster) standard text display ("draft mode"). In draft mode a Windows standard font is used so you cannot recognize font type and size directly on the screen. For my typing I find the WYSIWYG mode fast enough even on a 12-MHz type AT. I especially like to see the different font sizes which I sometimes use for headings or superscripts.

A feature I find very useful and convenient is the glossary. This feature allows you to permanently store certain character combinations or paragraphs (e.g., an address) which you frequently use. You call them up by a letter combination which you define yourself. This is demonstrated on the following example: We frequently use the name of Beilstein's substructure search program S⁴ in publications and manuals. To type "S⁴" you have to type "S" then change the font size from 12 point to 6 point, activate superscript, type "4", deactivate superscript, and change the font size back from 6 point to 12 point. This is a fairly elaborate combination of key strokes and mouse movements. You can make life much easier in defining "S⁴" in the glossary which stores this combination of key strokes and mouse movements under, for example, "s4", i.e., small s, number 4. To insert, simply type "s4" and press function key F3 and "s4" will be converted into "S⁴".

If you don't like your current word-processing program there are still good reasons not to switch to another: It usually costs a lot of time and effort to learn all the commands of a new powerful word processor. I have tried Wordperfect several times since my brother had recommended it so highly, but I simply could not memorize the function key assignments (can you?) because I could not see any logic behind its assignments. I am happy to say that this is not true for Winword.

Changing to Winword is easy if you are already familiar with Microsoft Windows, since it will then take you less than a day to become accustomed to Winword. The excellent tutorial will speed up the learning process considerably.

If you are not familiar with Windows, it may make a lot of sense for you to get used to this powerful GUI. There are several chemical structure editors already available (see above)

under Windows. STN Express is now available under the same user interface. There are various data systems for analytical instruments. Molecular Design Limited is announcing new Windows-based software and soon the Beilstein Institute will also offer its current software and database products under a Windows interface. It will probably take you a day to become familiar with Windows and then the same time as above to learn Winword.

In conclusion Word for Windows is a very powerful, but yet easy to use, word processor. It is particularly well suited for chemists because graphics and text can be combined very easily. The outline feature makes it simple to write scientific reports. A major drawback is that you cannot use it on your old IBM XT. The minimum hardware requirements include a fast AT-type computer, at least 2 MBytes of RAM and a fairly fast graphics card (for working without eye strain at least VGA or higher resolution). Winword should only be used with Windows 3.0 (or higher in the future) since it requires too

much DOS memory if run under Windows-286 or Windows-386.

REFERENCES AND NOTES

- (1) Wordstar 5.5, Wordstar International Corp., 33 San Pablo Ave., San Rafael, CA 94903.
- (2) WordPerfect 5.1, 1555 N. Technology Way, Orem, UT 84057.
- (3) MS Word 5.0, Microsoft Corp., One Microsoft Way, Redmond, WA 98052-6399.
- (4) Chemtext, Molecular Design Ltd., 2132 Farallon Dr., San Leandro, CA 94577.
- (5) Available from Microsoft Corp., One Microsoft Way, Redmond, WA 98052-6399. The price of Winword is \$450.00.
- (6) HiJaak 1.1, Inset Systems Inc., 71 Commerce Dr., Brookfield, CT 06804.
- (7) Software Bridge, Systems Compatibility CORP., 401 N. Wabash, Suite 600, Chicago, IL 60611.
- (8) Pagemaker 3.01, Aldus Corp., First Ave. S., Seattle, WA, 98104.
- (9) Jochum, C. Pagemaker. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, (1) 40-42.
- (10) ChemWindow, SoftShell International Ltd., 2754 Compass Dr., Suite 375, Grand Junction, CO 81506.

The NIST/EPA/MSDC Mass Spectral Database, PC Version 3.0

STEPHEN R. HELLER

Agricultural Research Service, BARC-W, Building 005, 10300 Baltimore Boulevard,
Beltsville, Maryland 20705-2350

Received February 2, 1991

In the fall of 1973 the first version of the Mass Spectral Search System (MSSS) was made publicly available to the scientific community on the General Electric Mark III computer network (1). At that time it consisted of slightly under 9000 nonunique spectra and required a large time-sharing system to store the database and programs. Now, some 18 years later, the database has increased to almost 54 000 unique spectra, and the entire system can be searched almost as quickly and flexibly on an IBM PC, either on a hard disk or on a CD-ROM, as it was searchable on the 1973 million dollar computer system. The June 1990 release, Version 3.0,¹ is the one reviewed here.

The National Institute of Standards and Technology (NIST—formerly NBS, the National Bureau of Standards) Standard Reference Data Program (SRD) has created an excellent and inexpensive source of mass spectral data and search software for the scientific community. In addition, the Aldrich Chemical Co., as part of their database activities, has created an identical CD-ROM version of the database. The NIST system is well designed to serve the researcher who needs access to a large mass spectral database for spectral identification or as an aid in structure determination of unknowns from mass spectral data. It is also a very useful tool for the classroom, both for courses on mass spectrometry or spectral interpretation, as well as any course in organic chemistry.

The details of the database have been described previously²⁻⁶ and consist of 53 994 EI (electron impact) only spectra. Each spectrum has a complete list of masses (m/z values) and intensities, a chemical name, as well as a display option for showing synonyms, molecular formula (and partial formula), molecular weight, CAS Registry Number, source of the spectrum, a Quality Index (QI) assigned to the spectrum,⁷ and, when plotted out, a structure diagram (for over 95% of the compounds in the database), but there is no CAS Registry Number or other identifying number on the spectrum, only the name, molecular formula, and molecular weight. Regular updates of the database, which will include additional spectra

as well as replacement of existing spectra with higher quality data, addition of missing chemical structures, and improvements to the software, along with new software capabilities, have occurred over the past three years since the first release of this system for the IBM PC.⁶ Enhancements included in this latest version are the ability for users to add their own spectra to the system as a separate, searchable, private database, and a graphics display (both on the screen and printed) for 53 994 compounds in the database.

The search software is almost as flexible and extensive as the original online system. What few search capabilities are still missing are being worked on, and it would seem that in a short period of time this PC version will exceed the capabilities of the highly versatile online system. The menus are easy to use, and the help messages are indeed helpful. The system has been well designed, and the system designer and programmer, Dr. Steve Stein of NIST, has listened carefully and well to the many suggestions from the hundreds of users who have purchased the system in the past three years. The 29-page manual (which is the same for the hard disk and CD-ROM versions), with two appendices, is quite readable. As seen in Figure 1, which lists all the system options, one can search the database by ID number (an internal numbering system code found on each spectrum), the CAS Registry Number, the chemical name, molecular formula (partial or complete), molecular weight, abundances of 10 major peaks, and by means of a complete sequential search of the entire database. About the only mass spectral search option currently missing is a search by neutral losses. Only the first 35 characters of a chemical name may be used in a search, but this does not seem to be too severe a limitation—how many users for example, can correctly type longer chemical names? The results of a search or spectrum look-up can be very quickly displayed (m/z values and abundances) or plotted. An example of the peaks and intensities in 1-docosanol, acetate, printed out and plotted are shown in Figures 2 and 3, respectively. The plots shown in Figure 3 took about 20 s to