

ay no other column. The  
by new South Boro  
to the King searched for  
line.

# NEWS

imes). If they of jingo-Telegraph Palmer noted in other wi  
he Daily Star, "pabling or long one: New York tabloid words will  
nes equally well "papee Sunday with Rudolf sub-editors.  
-of Mar

## A database for English poetry

The academic and professional publishers Chadwyck-Healey (headquarters in Cambridge, England) expect to begin publishing in the spring of 1992 their *English Poetry Full-Text Database* on CD-ROM (Compact Disc: Read Only Memory).

When complete, the database will hold the texts of over 4,500 volumes of poetry, the work of 1,350 poets from AD 600 to 1900. In this database, more than a gigabyte (that is, a billion bytes or characters) of verse will be searchable in a variety of ways (see panel). Sir Charles Chadwyck-Healey, founder of the company, describes it as 'The largest and most accessible full-text database yet published in the humanities' and claims that 'it sets a new standard for the creation of textual databases.'

The basic purpose of *English Poetry* is to provide reliable texts in machine-readable form including for each poem any accompanying text by the poet (such as footnotes, endnotes, sidenotes, dedications, prologues, and epilogues). The preview booklet (mid-1991) notes: 'A descriptive markup scheme is used in which textual elements are identified by their function rather than their appearance. For example the titles of poems are marked explicitly as titles rather than as being in a larger typeface or italicized. However, for users who wish to refer back to the printed sources of poems, page breaks, page numbers and other essential information specific to editions [are] included in the database.'

In addition to the generally recognised body of English poetry, the database includes: children's verse; translations of non-English verse (complete before 1800 and selectively thereafter); poetry written during the period covered but not published for whatever reason before 1900; hymns published before 1800; and complete verse dramas by authors primarily known as poets.

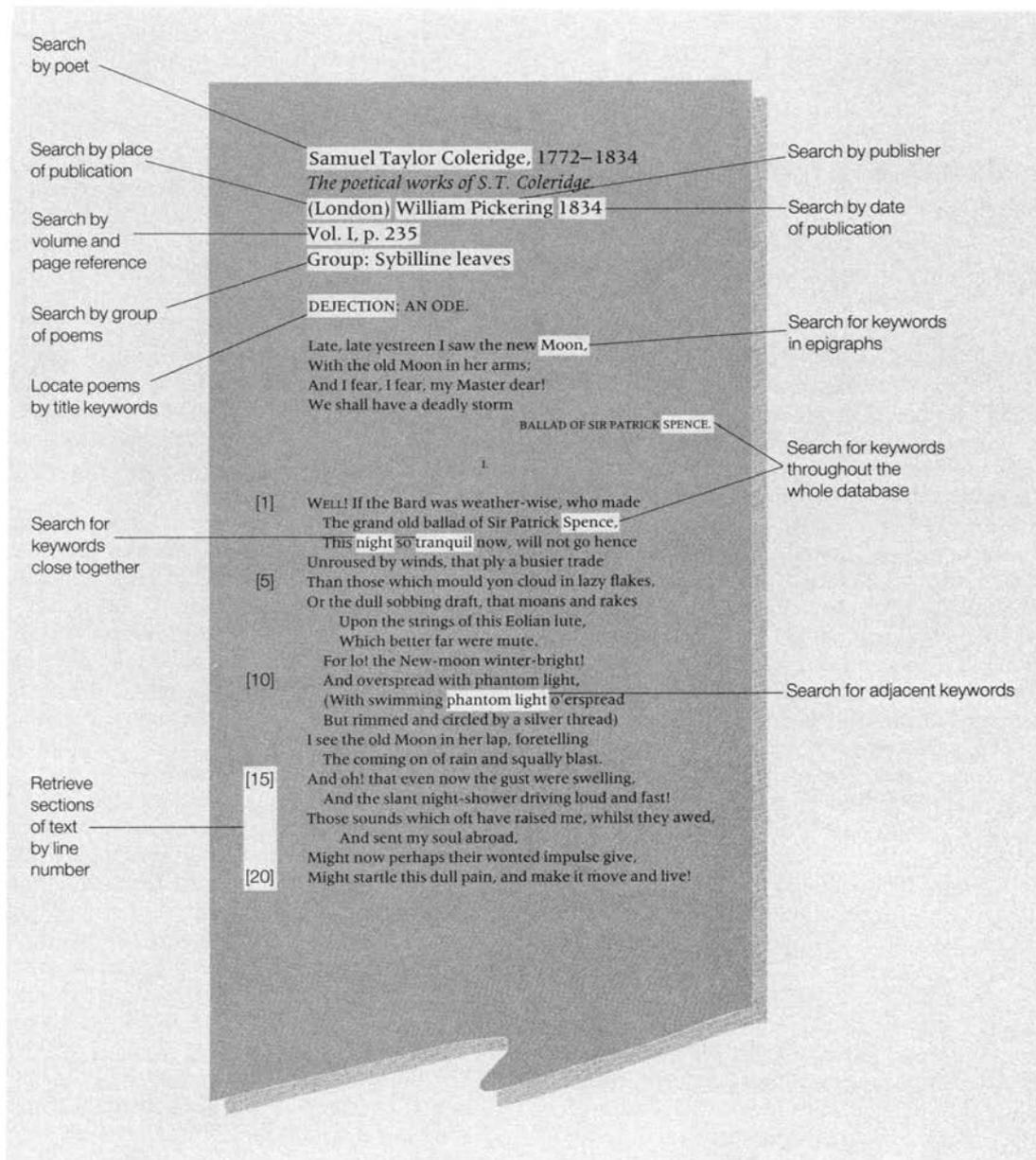
The database covers the works of those writers listed as poets in *The New Cambridge Bibliography of English Literature* (NCBEL, Cambridge University Press, 1969–1972), a

research tool that describes itself as 'a blueprint for research'. The preview booklet notes that the database 'also includes the few writers of poetry not cross-referenced by NCBEL, for example Emily Brontë and Aphra Behn'. In addition, following NCBEL, the database contains the work in English of Welsh, Scottish and Irish poets. Poets who were active before 1900 are included but poets principally active in the twentieth century are excluded. Although no reasons are given for a cut-off date of 1900, it seems likely that copyright was a major influencing factor.

The company are aware of some of the shortcomings of NCBEL as a reliable guide to the full extent of English poetry, acknowledging that some categories of poets, most notably women poets, are underrepresented. It is intended that such omissions be rectified at a later stage by issuing supplements to the main database 'in response to the suggestions of scholars'. The publishers maintain that a major advantage of a database of this kind is the ability to merge new texts seamlessly into the existing corpus, provided that the new texts are prepared following the same coding rules. Because the coding is based on international standards and is detailed in an accompanying handbook, it is suggested that individuals will even be able to supplement the database themselves should they wish particular editions of particular poets' works to be included.

The database can be searched by poet, by titles and headings, by bibliographical details (place and date of publication, publisher, etc.) or by any particular word or combination of words. The software will also allow combinations of searches, will give counts of word-frequency and, by displaying any selected keyword in context, function as a concordance of English Poetry. It will be possible to download and print-out texts, enabling the creation of customized anthologies for courses and other purposes, and to add comments and questions, etc., as needed. For those wanting to carry out stylistic, syntactic or other kinds of analysis the database will provide a rich source of raw data.

The editorial guidelines for the databases



have been formulated with the help of an Editorial Board consisting of the following scholars: Professor John Barnard of the University of Leeds; Professor Derek Brewer of the University of Cambridge; Lou Burnard of the Oxford Text Archive, the University of Oxford; Dr Howard Erskine-Hill of the University of Cambridge; Dr Daniel Karlin of University College London; and Michael Sperberg-McQueen of the Computer Centre, the University of Illinois at Chicago.

Alison Moss, the company's Editorial

Director, commented: "Most editions selected for English Poetry will have to be contemporary with their authors since much English poetry was not reprinted after its author's death. To provide a degree of editorial consistency across the database a decision has been made to favour contemporary editions for all poets, including those for whom alternatives are available. The editors also favour collected editions to ensure uniformity in the presentation of a particular poet's work. However, we

*continued on page 63*