"Look What Thy Memory Cannot Contain":
The Shakespeare Electronic Text Archive

By Kenneth B. Steele

     Twenty-five years ago, when T.H. Howard-Hill performed his
first test concording of Shakespeare's _Measure for Measure_ at the
Oxford University Computing Laboratory, literary computing required
custom software, mainframe facilities, and thousands of keypunched
cards.[1]  Even as late as 1973, Howard-Hill quite rightly
cautioned that it was still "too difficult, time-consuming, and
expensive for all but the most determined scholar to work with the
computer."[2]  As most academics know, however, in the 1980s the
tables have turned: to "all but the most determined scholar," word
processing alone has made the computer indispensable.  For the more
computer-literate humanist, dissertation abstracts and the _MLA
Bibliography_ are available on-line, the _Oxford English
Dictionary_ has been published on CD-ROM, and the Oxford and
Riverside editions of Shakespeare are available commercially in
electronic form.  Text retrieval software has become increasingly
"user-friendly": anyone who can manage WordPerfect or Nota Bene can
master WordCruncher or TACT in minutes.[3]  Literary computing need
no longer reduce volumes of poetry to fanfold pages of numerals and
z-scores; current computer software moves us, not further away from
the text, but closer to it than ever before, with a level of
precision and exhaustiveness seldom practical without prolonged
research.

     Interactive text retrieval is above all else _interactive_:
many of its primary benefits are immediately evident on-screen, but
difficult to convey in print.  It is remarkably convenient, often a
mere keystroke away from your wordprocessor, able to perform
complex operations instantly, and to insert the resulting
quotations, with citations, directly into your work-in-progress.
The computer becomes an invaluable tool to accelerate and refine
traditional approaches, such as close reading, imagery analysis, or
source study.  It renders printed concordances virtually obsolete,
and removes much of the drudgery from more technical textual or
orthographic research, increasing rather than diminishing scholarly
creativity.  The computer user is not limited to a mere index of
references for each keyword, but instead has instantaneous access
to the complete context of every occurrence of every word, every
partial word (e.g. all words ending in "-ick," all words with the
prefix "pro-," all words containing hyphens, etc.), every complete
or partial phrase, every punctuation mark, and every textual code.
Most software can generate charts or graphs of distributions by
play, scene, character, compositor, genre, or even chronological
period, as an aid to detecting overall patterns and relating them
to the specific occurrences.  The software can search an electronic
text for co-occurrences of words or phrases, such as x with y, x
without y, x and y without z, or x or y with z, and so on, all
within a user-specified range (any number of characters, lines,
scenes, etc).  Furthermore, lists of thematically-related words can
be created, such as Petrarchanisms or metatheatrical references,
and can then be examined or manipulated with immediate results,
enabling a novice to chart Shakespeare's imagery in a fraction of
the time it took Caroline Spurgeon.[4]  Obviously, it would be

virtually impossible for printed concordances to document the infinite number of possible co-occurrences or thematic distributions.

Electronic text retrieval can also revolutionize one's conception of the text.  Re-reading the works of Shakespeare via WordCruncher or TACT is a new stimulus to the critical imagination: a diachronic corpus suddenly becomes synchronic, multiple plays interpenetrate on a single word, and one is reading vertically, by cross-section across various plays, rather than horizontally, from beginning to end of each.  This is, obviously, no way to be _introduced_ to Shakespeare, and is certainly not to be taken as an invitation to "drown [our] book," like Prospero, or deflect attention from theatrical performance, where the plays ultimately come alive.  Text retrieval software does, however, turn a linear text into an imaginative "doodle pad," a testing-ground on which casual browsing and random experiment can lead to new and suggestive ideas.  Just as computerized spreadsheets have revolutionized accounting, making possible a rapid and flexible "what if" approach to financial variables, text retrieval software supplies immediate answers to critical experiment: literary scholars can now ask questions which were previously unthinkable, or unthought-of, accomplishing in moments analyses which might have been lifetime occupations in previous decades.

Electronic _editions_ of Shakespeare are now commercially available for use with such software, but recent scholarship, particularly that by Steven Urkowitz and Gary Taylor,[5] has brought about a new textual awareness throughout the scholarly community.  The intriguing distinctions between Q1 and F1 _King Lear_ reveal consistent and deliberate differences in characterization, plot, and poetry, but this play is not a special case: seventeen of Shakespeare's thirty-eight plays survive in two variant forms, and two more, _Romeo and Juliet_ and _Hamlet_, survive in three.  Scholars cannot afford to discard texts as "Bad Quartos" simply because they are awkward to collate, nor should inconvenient variants be dismissed as "indifferent"; as E.A.J. Honigmann remarks, the word "can as aptly describe the beholder as the thing observed."[6]  Although most scholars remain reluctant to "unedit" Shakespeare entirely,[7] the challenging indeterminacy of the original Quarto and Folio texts certainly rewards investigation.

In early 1988, the Centre for Computing in the Humanities at the University of Toronto initiated the Shakespeare Text Archive project, to produce an accessible and convenient textbase of the original Quarto and Folio texts for use on an IBM microcomputer. The Oxford University Computing Service Text Archive generously supplied the Howard-Hill texts of the First Folio and many important Quartos, and CCH computerized a number of so-called "bad" Quartos which had been previously ignored.  The encoding in all the files was then standardized for use with WordCruncher software. The Archive is still incomplete, and the process of verification continues, but currently 55 early texts of Shakespeare's 38 plays occupy a mere 10.2 megabytes of hard disk space on an IBM AT (18.8 mb with the WordCruncher indexes).

Unlike commercially-available electronic _editions_ of Shakespeare, these texts endeavour to be exact "electronic facsimiles" of the original Quarto and Folio texts.  Current

technology cannot yet replace photographic facsimiles, such as the
Norton facsimile,[8] but unlike more traditional reproductions the
text archive can be used in conjunction with any concording, text
retrieval, or collation software.  These texts are quickly proving
themselves a significant new tool for Shakespearean textual
scholarship, offering instantaneous access to all of the
authoritative texts in minute detail with unfailing precision.  The
electronic texts encode such typographical details as pagination,
signatures, italicization, justification, and turned-over lines,
and meticulously retain typographical errors.  For convenience of
reference, codes have been added to identify act, scene, and line
divisions, stage directions, speech prefixes, and title pages.
Hypothetical differentiations of compositor stints and homographs
are encoded but can be disregarded.

    The text archive contains an embarrassment of textual riches,
and for less technical research an electronic edition may prove
both simpler and more productive.  For literary rather than textual
purposes, the user must account for a bewildering array of
duplicate texts, variant spellings, typographical eccentricities,
textual cruxes, and apparent compositorial "errors" familiar to
anyone who has tried to read Shakespeare in facsimile.  The user
cannot merely search for all occurrences of the word "aye," but
must also consider "ay," "aie," "ay," and "I" (and less obvious
misprints).  Ideally, the Quarto and Folio texts will be
electronically linked to an edition, permitting searches of either
textbase in correlation with the other (a process which will
ultimately require hypertext technology).

    The potential of text-retrieval technology can perhaps be best
illustrated by a number of brief examples.  These sample queries
and results are certainly not the most intricate or intriguing
possible, but will hopefully be suggestive.  First, some more
general literary applications, for which an electronic edition
might serve equally well.  If we were developing a theory about
"The Murder of Gonzago" in _Hamlet_, for example, text retrieval
would allow immediate investigation of the 603 references to the
ear in Shakespeare.  Twenty-five of these references seem to carry
distinctly sexual undertones (specific evidence which seems more
solid than vague Freudian generalizations).  In several cases,
moreover, pouring poison or infection in the ear is closely
associated with some form of male sexual jealousy (see Fig. 1): the
ghost of Hamlet Senior reports the "iuyce of cursed Hebona in a
viall" which Claudius poured "in the porches of my eares" (_Hamlet_
Q2 1.5.66-7); which is of course echoed in the dumbshow stage
direction, in which we are told that the murderer "pours poyson in
the sleepers eares" (_Hamlet_ Q2 3.2.120sd).  In later plays, Iago
vows to "poure this pestilence into [Othello's] eare" (_Othello_ Q1
2.3.331), and Pisanio exclaims of his master Posthumous, "what a
strange infection / Is falne into thy eare?" (_Cymbeline_ F1 3.2.3-
4).  This recurrent association in Shakespeare's work, initiated,
we may note, in _Hamlet_, may help to explain the symbolism of
_Hamlet_ as it formed in Shakespeare's mind, and as it reverberated
throughout his career.

    More complex research is made possible by creating lists of
related words, such as animal references, Petrarchan images, or
Italian proper names.  Obviously the approach is somewhat
mechanical and limited to explicit references rather than more
subtle allusions (unless the investigator is already aware of these
subtleties), but it does provide a rapid survey to help direct

one's future efforts.  For example, a preliminary study of
theatrical metaphors (or metatheatricality) across the corpus could
begin with a list of theatrical terms (see Fig. 2).  My word list
matches 3384 occurrences, [9] largely in the tragedies and good
Quartos, and concentrated in the Elizabethan plays.  The histories
and the First Folio seem comparatively lean in metatheatrical
references.  _Hamlet_ and _A Midsummer Night's Dream_,
understandably, have by far the greatest concentration of
theatrical language, owing to their internal performances, and
their overall thematic concerns.  The software will supply the
specific quotations which cannot be ignored in formulating a
theory, regardless of their location in the canon.

     Any electronic edition would permit the preceding
investigations, and would simplify matters by modernizing
orthography and emending textual difficulties, but would obscure
the distinctions between the variant texts.  For textually-
sensitive research into matters of bibliography or revision, the
Text Archive becomes indispensable.  Modern editions generally
insert stage directions as they are deemed necessary; the Quarto
and Folio texts contain only those 6276 lines of stage directions
which could possibly be authorial.  It is immediately striking that
no comedy has more than about one hundred lines of directions,
while history plays and tragedies especially almost never have
fewer than one hundred.  Stage directions are heaviest in _Antony
and Cleopatra_ F1 (202), _Coriolanus_ F1 (176), _The Two Noble
Kinsmen_ Q1 (194), and _Richard III_ F1 (191).

     Permissive stage directions, once considered evidence of
authorial manuscript or "foul papers," can be quickly isolated and
analyzed.  Shakespeare's infamous negative capability found
frequent expression in stage directions which include words such as
"others" or "other lords": 96 such stage directions are present in
the early texts.  Surprisingly, however, these seem to be
concentrated in the so-called "bad" Quartos and the Folio, not the
"good" Quartos usually called foul paper texts.  The overall
concentration of "other" characters seems to follow the
concentration of stage directions themselves, in the tragedies and
histories -- where processions, spectacle, and crowd scenes are
prevalent.  Permissive stage directions also include numerical
ambiguities, such as "one or two" (7 occurrences), "two or three"
(22), "three or four" (16), "four or five" (3), "five or six" (4),
and even "seven or eight" (1).  These stage directions are most
prominent in _2 Henry VI_ F1, _Romeo and Juliet_ Q2/F1 (Q1 has
none), and _Coriolanus_ F1 (with 4 each), _2 Henry IV_ Q1 (F1 has
none), _Hamlet_ Q2/F1 (Q1 has none), _All's Well that Ends Well_
F1, _Antony and Cleopatra_ F1, and _Pericles_ Q1 (with 3 each).
This form of permissiveness does indeed seem concentrated in
authorial copy, and takes even more interesting forms:
"Bullingbrooke or Southwell reades" (_2 Henry 6_ (F1) 1.4:21);
"Enter Menenius to the Watch or Guard" (Coriolanus (F1) 5.2:0); and
"...Branches of Bayes or Palme in their hands" (Henry 8 (F1)
4.2:90).  _Coriolanus_ F1 has more such stage directions, including
"or" or "other", than any other play text (12), followed by _Antony
and Cleopatra_ F1 (9), _2 Henry VI_ F1 (8), and with six each,
_Richard III_ F1 (Q1 has 5), _Romeo and Juliet_ Q2/F1 (Q1 has 2),
_Hamlet_ Q2 (F1 has 5, Q1 has 2), and _Timon of Athens_ F1.

     Verbal evidence such as spelling or hyphenation can also be
examined electronically.  The fifteen occurrences of the

obsolescent plural "eyen" (in its various spellings) reveals that in all cases but one (the questionable text of Q1 _Pericles_), the word is used primarily for purposes of rhyme (and eight times to rhyme with "mine").  Furthermore, it seems clear from Shakespeare's use of the word in _A Midsummer Night's Dream_ and _As You Like It_ that he considered it preposterous, appropriate in circumstances such as Bottom's performance as Pyramus, or Phebe's love poem to Rosalind.

The 8561 hyphenated words in the Quarto and Folio texts can immediately by summoned up by the retrieval software, and it can quickly be determined that the concentration is heavier in the Folio (probably because of narrower column widths).  _The Merry Wives of Windsor_ F1, in particular, has 532 hyphenations, exactly ten times the number of Q1, and they are concentrated in scenes 1.3-4.2 and 5.5.  Strangely, most of the hyphenations seem not to be required by lineation, but by compositorial spelling-habits (although this is a very preliminary observation).

Punctuation study is quick and painless on the archive texts. The 138,198 commas can be viewed immediately (although you would have to page through more than 23,000 WordCruncher screens to see them all), as can the 104,928 periods, 26,974 colons, 5820 semi-colons, 15,785 question marks, and 1162 exclamation marks. Distributions can be examined to identify compositorial habits, or more generally to observe that the Folio seems light in periods, but heavy in semi-colons and colons.  Such study has already been done manually, of course, but never has so much raw material been so accessible: rather than hunting through decades of printed criticism to discover what Charlton Hinman concluded, one can instantly examine the texts themselves in any way desired.

The Shakespeare Text Archive project is currently dependent upon the labour of a single volunteer, and as a result it is still some distance from completion.  A number of texts have yet to be obtained from Oxford, and a number of texts will have to be entered manually.  Once the Archive contains electronic texts of every significant early edition of the poems and plays, and has been carefully proofread yet again, it will be made available to the scholarly community as a whole through the Oxford University Computing Services Text Archive for a nominal fee. Ultimately, it will also incorporate non-copyright texts of Shakespearean source materials, and either an encoding of important emendations or a parallel edited text.

Many of the Text Archive's ultimate aims will be realized only when a more advanced software engine is developed: currently WordCruncher is unable to index non-sequential hierarchies (i.e., it can index Play/Act/Scene/Line but not speaker or compositor stints, which are scattered throughout the texts), and TACT, which would solve this difficulty, is unable to manage the size of the complete textbase.[10]  Ideally, software will be developed to perform on-screen collation of the variant texts, automated comparisons of multiple inquiries, and the analysis of repetition itself, as an abstraction: for example, the distribution of rhetorical repetitions such as anaphora, epistrophe, epanalepsis, anadiplosis, and perhaps even alliteration and rhyme.  The text files, in standard ASCII format, have weathered twenty-five years of technological revolution, which have seen ten times the computing power of Howard-Hill's first mainframe appear on the

average academic desktop, and in all likelihood their usefulness will continue despite the unimagined technological developments to come.

The "electronic facsimiles" of the Quarto and Folio texts are a powerful and flexible approach to the textual indeterminacy of Shakespeare's works.  Random Cloud has insisted that critics must look "to process in creation rather than to hypostatized artefacts," and "away from the editor's ideal single version -- the so-called 'definitive text' -- to the author's actual multiple versions: an infinitive text,"[11] which he elsewhere defines as "a polymorphous set of all versions."[12]  The Shakespeare Text Archive, which may ultimately become an electronic variorum, is perhaps a first step toward realizing the goal of a fluid "infinitive text" in electronic form.

 |l65       Vpon my secure houre, thy Vncle stole
 |l66       With iuyce of cursed Hebona in a viall,
 |l67       And in the porches of my eares did poure
 |l68       The leaprous distilment, whose effect
 |l69       Holds such an enmitie with blood of man,
 |l70       That swift as quicksiluer it courses through

    (Hamlet (Q2) 1.5:67)


─────────────────────────────────────────────────────────


           him: anon come in an}
      *{#other man, takes off his crowne, kisses it, pours
          poyson in the sleepers eares},
      *{and leaues him: the Queene returnes, finds the King
          dead, makes passionate}
      *{action, the poysner with some three or foure come in

    (Hamlet (Q2) 3.2:120sd)


─────────────────────────────────────────────────────────

 |l330      And she for him, pleades strongly to the Moore:
 |l331      I'le poure this pestilence into his eare,
 |l332      That she  repeales him for her bodyes lust;
 |l333      And by how much she striues to doe him good,
 |l334      She shall vndoe her credit with the Moore,

    (Othello (Q1) 2.3:331)


─────────────────────────────────────────────────────────

 |l3       Oh Master, what a strange infection

 |l4       Is falne into thy eare? What false Italian,
 |l5       (As poysonous tongu'd, as handed) hath preuail'd
 |l6       On thy too ready hearing? Disloyall? No.

    (Cymbeline (F1) 3.2:3)

```
_____

Fig. 1:WordCruncher report showing occurrences of ears in
association with poison imagery.

Report for:  theater,thea-tre,stag'd,perfourme,costume,revells,
             apeer's,plat-formes,platforme,gloabe,page-ant,part,
             parte,partes,parts,applau'd,applaud,applaude,
             applauded,applauding,applause,applauses,audience,
             auditor,auditorie,auditors

Total References in List: 3481


                  Frequency   -- Percentages --
Range Names        Count  Actual Expect Difference
-----------------------------------------------------

First Folio         2098    60%    66%      -6%
Good Quartos        1014    29%    26%       3%
Bad Quartos          266     8%     7%       1%
Minor Poems          103     3%     2%       1%
Comedies             942    27%    27%       0%
Histories            702    20%    27%      -7%
Tragedies           1221    35%    31%       4%
Romances             318     9%     9%       0%
Authorial           2326    67%    70%      -3%
Elizabethan         2502    72%    70%       2%
Jacobean             935    27%    30%      -3%
Prefatory             44     1%     0%       1%


                  Frequency   -- Percentages --
Play               Count  Actual Expect Difference
-----------------------------------------------------

Folio                 44     1%     0%       1%
1 Henry 6 (F1)        30     1%     2%      -1%
2 Henry 6 (F1)        37     1%     2%      -1%
3 Henry 6 (F1)        35     1%     2%      -1%
Richard 3 (Q1)        42     1%     2%      -1%
Richard 3 (F1)        40     1%     2%      -1%
Venus&Adonis (Mod)    20     1%     1%       0%
Comedy of Errors(F1)  19     1%     1%       0%
Sonnets               83     2%     1%       1%
Titus Andronicus(Q1)  37     1%     2%      -1%
Titus Andronicus(F1)  31     1%     2%      -1%
Taming the Shrew (F1  46     1%     2%      -1%
Two Gentlemen (F1)    27     1%     1%       0%
Love's Labours (Q1)   69     2%     2%       0%
Love's Labours (F1)   70     2%     2%       0%
King John (F1)        55     2%     2%       0%
Richard 2 (Q1)        37     1%     2%      -1%
Richard 2 (F1)        37     1%     2%      -1%
Romeo & Juliet (Q1)   41     1%     1%       0%
Romeo & Juliet (Q2)   69     2%     2%       0%
Romeo & Juliet (F1)   65     2%     2%       0%
Midsummer (Q1)       109     3%     1%       2%
Midsummer (F1)       107     3%     1%       2%
Merchant of Ven (Q1)  79     2%     2%       0%
```

```
Merchant of Ven (F1)      72      2%      2%       0%
1 Henry 4 (Q1)            51      1%      2%      -1%
1 Henry 4 (F1)            51      1%      2%      -1%
Merry Wives (Q1)          14      0%      1%      -1%
Merry Wives (F1)          30      1%      2%      -1%
2 Henry 4 (Q1)            66      2%      2%       0%
2 Henry 4 (F1)            74      2%      2%       0%
Much Ado (Q1)             48      1%      2%      -1%
Much Ado (F1)             48      1%      2%      -1%
Henry 5 (F1)              77      2%      2%       0%
Julius Caesar (F1)        52      1%      2%      -1%
As You Like It (F1)       65      2%      2%       0%
Hamlet (Q1)              107      3%      1%       2%
Hamlet (Q2)              194      6%      2%       4%
Hamlet (F1)              168      5%      2%       3%
Twelfth Night (F1)        44      1%      2%      -1%
Troilus & Cress (Q1)      76      2%      2%       0%
Troilus & Cress (F1)      80      2%      2%       0%
All's Well (F1)           36      1%      2%      -1%
Measure (F1)              54      2%      2%       0%
Othello (Q1)              59      2%      2%       0%
Othello (F1)              63      2%      2%       0%
King Lear (Q1)            47      1%      2%      -1%
King Lear (F1)            43      1%      2%      -1%
Macbeth (F1)              41      1%      1%       0%
Antony (F1)               93      3%      2%       1%
Coriolanus (F1)           73      2%      2%       0%
Timon (F)                 38      1%      2%      -1%
Pericles(Q1)              62      2%      1%       1%
Cymbeline (F1)            63      2%      2%       0%
Winter's Tale (F1)        79      2%      2%       0%
Tempest (F1)              41      1%      1%       0%
Henry 8 (F1)              70      2%      2%       0%
2 Noble Kinsmen (Q1)      73      2%      2%       0%
```

------------------------------------------------------------

Fig. 2:WordCruncher distribution report for theatrical metaphors
across the Shakespeare canon.

N O T E S

1.      Howard-Hill describes the process in detail in his article,
"The Oxford Old Spelling Concordances," _Studies in
Bibliography_ 22 (1969): 143-64.  The later history of the
texts is described in an anonymous note, "Shakespeare and
the Computer," _ALLC Bulletin_ 8:1 (1980): 72.

2.      T.H. Howard-Hill, "A Common Shakespeare Text File for
Computer-Aided Research: A Proposal."  _Computer Studies in
the Humanities and Verbal Behaviour_ 4:1 (1973): 54.

3.      Electronic Text Corporation's WordCruncher has been
thoroughly reviewed by John J. Hughes in his article,
"WordCruncher: High Powered Text-Retrieval Software," in
_Bits and Bytes Review_ 1:3 (February 1987): 1-8.  TACT is
a new public domain program distributed by the Centre for

Computing in the Humanities at the University of Toronto.

4.     With apologies to Caroline F.E. Spurgeon, _Shakespeare's
Imagery and What It Tells Us_ (Cambridge: Cambridge
University Press, 1935).

5.     See, in particular, Steven Urkowitz, _Shakespeare's
Revision of King Lear_ (Princeton: Princeton University
Press, 1980), and Gary Taylor and Michael Warren, eds. _The
Division of the Kingdoms: Shakespeare's Two Versions of
King Lear_ (Oxford: Clarendon Press, 1983).

6.     E.A.J. Honigmann, _The Stability of Shakespeare's Text_
(London: Edward Arnold, 1965) p 167.

7.     As proposed by Randall McLeod, "Unediting Shak-speare"
_Sub-Stance_ 33 (1982): 26-55.

8.     Charlton Hinman, ed.  _The Norton Facsimile: The First
Folio of Shakespeare_ (New York: Norton, 1968).

9.     In this case, my choices were the various spellings of act,
appear, comedy, costume, globe, illusion, mask, pageant,
part, perform, platform, play, revel, scaffold, scene,
show, stage, theater, and tragedy.

10.    A "beta-test" version of TACT will shortly be tested with
the Shakespeare Text Archive, and may be available by the
time of printing.

11.    Random Cloud, "The Psychopathology of Everyday Art," 100-
168 in G.R. Hibbard, ed. _The Elizabethan Theatre IX_
(Waterloo: P.D. Meany, 1986). p. 111.

12.    Random Cloud, "Commentary: The Marriage of Good and Bad
Quartos," _Shakespeare Quarterly_ 33 (1982):421-31, p. 422.