Standardisation of Word Processing for Publication: The Application of ASPIC in Common Word Processing Packages

J.D. Wilcock and R.D. Spicer
Research Centre for Computer Archaeology,
North Staffordshire Polytechnic, Stafford

1. Introduction to ASPIC

ASPIC is a set of standard codes prepared by the Composition Group, a working party set up by the Development and Technology Committee of the British Printing Industries Federation (British Printing Industries Federation, 1984). It consists of a number of non-technical codes or flags which can be keyed as ASCII text on any input device or computer, and which indicate to photocomposition equipment any changes in typeface, layout, etc. required by the author at a specific point, and which simplify subsequent setting. ASPIC can be used before decisions about typographic style or printing method have been finalised. The problem with the increasingly prolific use of word processors for text input is that all the different word processing packages use embedded control codes which are in general non-printable characters, and which are certainly non-standard. ASPIC is designed to satisfy the urgent need for a printing industry standard, and does so by providing a concise range of codes for the most popular types of formatting for typesetting. The Printing Industry Research Association is preparing the groundwork for a substantial standardisation effort, with implications for publishing via electronic and other new media as well as via conventional typesetting and printing. PIRA intends that any new schemes they develop will not conflict with ASPIC, and BPIF is supporting PIRA in this work.

2. Application of ASPIC to WORDWISE (TM)

2.1 Introduction to WORDWISE (TM)

WORDWISE (TM) is a word processor written by Computer Concepts, specially designed to make the best use of the facilities offered by the BBC Model B microcomputer. As such it is widely used, and the facility to produce ASPIC-compatible files with WORDWISE (TM) is necessary. The remarks in this paper refer to WORDWISE (TM) Version 1.2.

A package has been developed which translates a standard WORDWISE (TM) file into an ASPIC-compatible file. This opens the way for authors' texts to be transferred to the printers via floppy disks prepared on a commonly-available microcomputer.

2.2 WORDWISE (TM) to ASPIC

There are three categories of codes which occur in WORDWISE (TM) and ASPIC:

i) Codes used in WORDWISE (TM) with no ASPIC equivalent;
ii) Codes used in ASPIC with no WORDWISE (TM) equivalent;
iii) Codes which may be translated between WORDWISE (TM) and ASPIC.

As a matter of fact, there are relatively few codes in the third category which can be directly translated, so this simplifies the construction of the translator package. Codes in the first category with no ASPIC equivalent are stripped by the package from the WORDWISE (TM) file. Codes in the second category, since they are in ASCII format, may be inserted by direct editing of the file using WORDWISE (TM).

2.2.1 Codes used in WORDWISE (TM) with no ASPIC equivalent

The following WORDWISE (TM) embedded commands are not directly available in ASPIC, and are stripped out by the package:

LM	Left Margin
LL	Line Length
PL	Page Length
TS	Top Space
DH	Defines Heading
HP	Heading Position
BS	Bottom Space
DF	Define Footing
FP	Footing Position
LS	Line Spacing
SS	Single Spacing
CO	Continuous Output
EP	Enable Paging
BP	Begin Page
PN	Page Number
DP	Defines Pound sign
PC	Pad Character
EM	Enable Messages
DM	Disable Messages
PP	Print Page number
GF	Get File

The following commands, while being stripped by the present package, are reserved for future enhancements, to provide formatted code compatible with ASPIC on an output file:

OC	Output Control code (to a specialist printing device)
OS	Operating System command (to be used, for example, to download software fonts to a specialist printing device)
SP	Output blank lines (to reserve space for diagrams)
CP	Conditional page (to be used in conjunction with SP to provide reserved space for diagrams, using the ASPIC commands [qp] quarter page, [hp] half page, and [wp] whole page)
DT	Define TAB stop positions (for tabulating columns)

TAB Tabulation character (to be used in conjunction with DT to tabulate columns).

2.2.2 Codes used in ASPIC with no WORDWISE (TM) equivalent

Many of the ASPIC codes have no direct equivalents in WORDWISE (TM), and so a facility in the translator package is not required for them. These codes may be inserted as direct ASCII text using the standard WORDWISE (TM) facilities. They include all levels of heading commands and their corresponding end heading commands, all levels of text commands, font changes, right justification (ragged left) [jr], line feed without return [[, the end-of-paragraph line centre command [lc] and end-of-paragraph line right command [lr], adjusting points, em and en space, em dash [md] and en dash [nd], superiors and inferiors (although these might be available on a specialist printing device using print roller feeds), caption start [cs] and end [cx], gaps of fixed [g#] and unknown depth [g?], insert leader dots [id], insert baseline rule [ir], insert space [is], insert vertical space [iv], specialised character [p#] and unknown specialised character [?#].

In Supplementary ASPIC none of the features is compatible with WORDWISE (TM). These include accents, Greek letters, most of the special characters, tabular commands, and most of the mathematical symbols. Some of the special and mathematical symbols are, however, implementable in WORDWISE (TM) (see below).

It is a pity, therefore, that many of the sophisticated features of Supplementary ASPIC will not be available on the majority of word processing packages, and WORDWISE (TM) is no exception to this.

2.2.3 Codes which may be translated between WORDWISE (TM) and ASPIC

The following is a detailed description of those WORDWISE (TM) embedded codes and characters which are directly translated in the present package to ASPIC codes in ASCII:

IN# Indent. The # variable is used for a calculation to the nearest suitable ASPIC indent code [i1], [i2], etc.
IN0 Indent zero. Translated to [ix]
CI Cancel Indents. Translated to [ix]
CE Centre. Translated to [jc]
JO Justification on. Translated to [jj]
NJ No Justification. Translated to [j1]
TI Temporary indent. This is assumed by the package to be the start of a new paragraph with indent, and is translated as return followed by []
TI0 Temporary indent zero. This is assumed by the package to be the start of a new paragraph with no indent, and is translated as return followed by]]

The remainder of the translation is concerned with specific characters. Quotation marks ' and " are translated direct. Note that the feet and inch symbols are indistinguishable from these in WORDWISE (TM), so are not translated as [ft] and [in] - if desired these codes must be keyed in. Closing quotes are a special case in ASPIC: the group of two characters '. is translated as three characters ''. and ". is translated as three characters "".

The following characters are also translated:

*	Asterisk [as]
@	At sign [at]
(c)	Copyright (three characters) [co]
#	Hash [ha]
+	Plus sign [+]
-	Minus sign [-]
_x_	Multiplication sign (3 characters space,x,space) [*]
<	Less than [lt]
>	More than [mt]

The package is written in BBC BASIC, and is therefore convenient for use on BBC microcomputers which have WORDWISE (TM) sideways ROM chips. A test run of the package is illustrated by the printouts in Figure 1.

3. Application of ASPIC to WordStar (TM)

3.1 Introduction to WordStar (TM)

WordStar (TM) differs greatly from WORDWISE (TM) in that formatting commands are not generally embedded into the text; the most commonly used formats are obtained directly from manipulation of the on-screen ruler and by various toggles which can be switched by the operator at any stage of the editing. Thus the program actually alters the text file dynamically, adding spaces and new lines automatically while editing takes place. Because the commands cannot be stored with the text, a re-edit of the file with different settings would remove the formatting performed previously.

Thus the sort of replacement or translation described above for WORDWISE (TM) is limited only to a few WordStar (TM) features: the printing commands (prefixed with CTRL-P), and some of the dot commands (which use a dot as the first character of a line to distinguish them).

WordStar (TM) does perform some rather perverse modifications to text which, to the uninitiated, can be incomprehensible; and it sometimes causes havoc when files are transferred to another system. In the document mode, word-wrapping sets the top bit of the last character of each word, and any dynamic spaces which it inserts also have their top bit set. The so-called 'soft' newlines are generated with a 8DH,0AH string rather than the conventional 0DH,0AH of the 'hard' RETURN. Thus WordStar (TM) can

*@(c)#+- x <>'"'.". This is a test message for translation by V.READCD.

This is a test of indent paragraph with a quite long sentence.

Paragraph without indent.

[i1] [i2]
[ix] [ix]
[jc]
[jj]
[jl]

[as] [at] [co] [ha] [+] [-] [*] [it] [mt] '"''."". This is a test message for translation by V.READCD.
[]
This is a test of indent paragraph with a quite long sentence.
]]
Paragraph without indent.

Figure 1

identify those spaces and newlines for which it is responsible, and can distinguish them from those entered by the user. It is this bit-setting which sometimes makes a WordStar (TM) file appear on the screen with Greek letters, reverse-video or other illegal characters when it is listed as a stand-alone file.

3.2 WordStar (TM) to ASPIC

ASPIC is capable of stripping these bit-setting anomalies out of a file, and also of removing control codes, but this means that the ASPIC codes would have to be inserted at the printer's premises, taking control away from the author, and probably incurring increased costs. The program described below will strip codes unique to WordStar (TM), and replace those which translate into ASPIC.

The program permits quite a variety of functions to take place, and it is a useful utility to any WordStar (TM) user, whether or not ASPIC is involved. These functions are invoked by the usual means of 'switches' in the command line, and can take place either all together in one pass or, if required, in separate passes. A report is optionally generated at the end of each pass, giving the line number and the first ten characters of that line for each substitution, so that checks can be made for any ridiculous changes.

3.2.1 Processes in the WordStar (TM) to ASPIC program

The separate processes are:

i) Removal of 'soft' spaces inserted by WordStar (TM)

ii) Stripping of all 'top bits' to leave pure ASCII

iii) Replacement of control codes and 'hard' newlines with corresponding ASPIC codes

iv) Substitution of 'soft' RETURNs with 'hard' RETURNs

v) Replacement of unrecognised control codes with empty ASPIC '[]' to permit manual insertion of ASPIC codes

vi) Complete removal of all control codes (this process can be used either after process iii or after process v)

vii) Removal (or replacement with '[]') of all unrecognised dot command lines

viii) Removal of EOF character from the body of the file; this is to permit recovery of files which have been damaged or badly concatenated. An EOF character is placed at the physical file end

ix) Report of substitutions made, with line numbers and the first few words of each line

x) Line count, word count, number count and hyphen count printed

xi) ASPIC to WordStar (TM) reverse translation

xii) Removal of ASPIC codes for plain text checking.

The above list is not in any particular order, and care would obviously be needed when more than one pass is invoked with different commands to ensure the correct sequence, although a single pass (with multiple commands) would prohibit illegal combinations.

The program is written in C, and therefore is available for most common micros. The code translation tables are placed together in an area of the program to permit alteration by the user with a suitable debugging tool.

Although only a few of the wide-ranging ASPIC commands are supported, the special needs of WordStar (TM) files are supported by the program, which will remove control codes to allow the transfer of files to any other word processing environment. It will also strip control codes from other word processing systems for use by WordStar (TM), and indeed it is ideally suited to turning document files into non-document files.

4. Conclusion

A preliminary package has been written in BBC BASIC for the translation of WORDWISE (TM) standard files to compatible ASPIC code. With use of the standard WORDWISE (TM) editing features to insert desired ASPIC codes which are not implemented in WORDWISE (TM), it is possible to provide machine-readable input to an ASPIC equipment. Several enhancements are planned.

A second program has been written in C for the translation of WordStar (TM) standard files to compatible ASPIC code. It may be used on any microcomputer with both C and WordStar (TM) available.

It is hoped that archaeologists will use the cheaper publication facilities offered by automated printing methods, but to do so using the convenience of word processors will imply some standardisation of codes, and the total removal of embedded control codes from machine-readable files. The ASPIC system offers this standardisation, so a translator must be provided for each word processor. The systems described above are a first attempt to provide this for the packages WORDWISE (TM) and

WordStar (TM).

Reference

British Printing Industries Federation. 1984

The ASPIC Handbook
A guide to Authors' Symbolic Pre-press Interfacing Codes
BPIF Development and Technology Committee
Composition Group. January 1984