

Date: Tue, 27 Apr 1999 13:05:56 -0600
From: Meredith Brown <racer@lanl.gov>
Subject: Blue Alert: Word Processor Conversion Issues

Project Hanford Lessons Learned

Title: Characters Convert Inaccurately Between Word Processors

Date: April 20, 1999 Identifier: 1999-RL-HNF-0018

Lessons Learned Statement: Documents converted from WordPerfect® into Microsoft Word® may contain errors in character mapping that could cause serious problems and potentially unsafe conditions, especially with operating procedures and other safety-related documents.

Discussion of Activities: Documents converted between word processing software packages may contain incorrect characters caused by inaccurate mapping between character sets. Some fractions may not convert accurately. This could create safety issues if operating procedures were released with erroneous information. For example, the fraction 1/4 in WordPerfect® converts to a value of three (3) in Word. An operating procedure that states, "Open valve A-45, Holding Tank Fill Valve, 1/4 turn." would read "Open valve A-45, Holding Tank Fill Valve, 3 turn." after conversion.

Analysis: There are basically two problems. One deals with font substitution and the other deals with the way Word® converts symbols into characters. The problem related to font substitution is an issue with any Windows application. As technology improves, so have the character definition standards. Word® uses a relatively new character definition standard called Unicode that assigns a unique 16-bit number to each symbol or character. WordPerfect® 5.1 uses a character set called OEM (original equipment manufacturer). Older Windows® applications use a standard character set called ANSI (American National Standards Institute). The 3 standards are not the same for all characters, especially for symbols and characters beyond the normal letters and characters on the keyboard. A word processor opening a file with a character set that it does not use may not map each character to the correct symbol in its own set. Users opening a document on a system that does not have the same font set as that used in the original document may see unexpected results on their screen such as boxes or other strange characters, due to font substitution.

The other problem arises from the way Word® treats some symbols as codes and others as characters. The character symbols are more susceptible to changing when the document base font is changed. The character may change to an empty box, or a character from a different font set. To prevent this, the character symbol should be changed to a code symbol.

Any document with any character set created using any version of WordPerfect® may be susceptible to this problem. Other word processors may also cause these errors.

Recommended Actions: All converted documents must be carefully proofread to ensure the documents are accurate and complete before they are released for use. LMSI has developed a MACRO

for Word® that will replace many character symbols with the appropriate code symbols to minimize the impact of the conversion problem. DO NOT RELY SOLELY ON THIS MACRO TO CONVERT DOCUMENTS ACCURATELY. Some character sets do not convert cleanly even with this MACRO. ALWAYS proofread the document carefully after conversion.

Priority Descriptor: BLUE/Information

Functional Categories (DOE): Conducts of Operations, Information Technology

Originator: Fluor Daniel Hanford, Inc.

Contact: Project Hanford Lessons Learned Coordinator; (509) 373-7664; FAX 376-5243;

e-mail: PHMC_Lessons_Learned@rl.gov

Authorized Derivative Classifier: Roger Terry (509) 372-3343

Reviewing Official: John Bickford

Keyword(s): file conversion, incorrect mapping, WP Character Sets

References: Hanford User's Help, for Word®

WordPerfect® is a trademark of Corel Corporation.

Word® and Windows® are trademarks of Microsoft Corporation.