

METHOD FOR CAPITALIZATION CHECKING DURING SPELLING VERIFICATION

D. A. Hamilton, A. M. Herzik and R. C. Nielsen

An approach taken by some existing spelling checking systems is to have all alphabetic characters in lower case form when storing words in the dictionary or supplemental dictionaries. When a word within text is to be compared against the dictionaries, all alphabetic characters are then represented in a lower case form. This approach does not allow for recognition of capitalization errors. Following is a list of examples of capitalization errors which are not detected by existing systems.

robert
Mcknight
iBM
aNimal

Disclosed is a solution to the above problem. A dictionary is provided as part of the spelling checking system wherein the words within the dictionary have the alphabetic characters in lower case except for characters which must be in upper case in order for the word to be correct (e.g., in "Robert", the "R" must be in upper case).

The user is provided the capability to store words "as is" (upper and lower case characters are stored as they appear) in supplemental dictionaries. These words would be character strings unique to user application and not found in the dictionary supplied to the user.

In operation, words are compared to the dictionaries to determine if the words are valid, using the following rules:

1) If the alphabetic characters in the words are (a) all lower case, (b) a mixture of upper and lower case other than initial upper case alphabetic with all other characters lower case, or (c) if the word contains non-alphabetic characters, THEN the dictionary is examined for an exact match against the word. If it is not found, verification fails.

2) If all characters are upper case, and a dictionary entry exists for that word with (a) all characters lower case, (b) only the first character upper case, or (c) all characters upper case, THEN the word verifies as valid.

3) If only the first character is upper case, and a dictionary entry exists for that word with (a) all characters lower case or (b) only the first character upper case, THEN the word verifies as valid.