

if enough candidates are found (Step 530). If enough candidates are found, the process outputs just the distance one corrections and terminates (Step 540). Otherwise, the process 500 continues with soft distance two variant testing (Step 600), as discussed further below in conjunction with FIG. 6, and only after this testing output the results and terminate (Step 540).

[0040] FIG. 6 describes the process 600 of testing variants of the candidate word against hash tables derived from the dictionary for "soft" distance two misspellings. The algorithm 600 starts in block 610 and obtains the candidate word. Following that, the unchanged candidate word is tested against the deletion and deletion-transposition hash tables in Step 620. Next all single step transpositions and single step deletions are generated (Step 630), and then tested against the single step transposition and single step deletion hash tables (Step 640). Finally, the accumulated matches are output in Step 650.

[0041] System and Article of Manufacture Details

[0042] As is known in the art, the methods and apparatus discussed herein may be distributed as an article of manufacture that itself comprises a computer readable medium having computer readable code means embodied thereon. The computer readable program code means is operable, in conjunction with a computer system, to carry out all or some of the steps to perform the methods or create the apparatuses discussed herein. The computer readable medium may be a recordable medium (e.g., floppy disks, hard drives, compact disks, or memory cards) or may be a transmission medium (e.g., a network comprising fiber-optics, the world-wide web, cables, or a wireless channel using time-division multiple access, code-division multiple access, or other radio-frequency channel). Any medium known or developed that can store information suitable for use with a computer system may be used. The computer-readable code means is any mechanism for allowing a computer to read instructions and data, such as magnetic variations on a magnetic media or height variations on the surface of a compact disk.

[0043] The computer systems and servers described herein each contain a memory that will configure associated processors to implement the methods, steps, and functions disclosed herein. The memories could be distributed or local and the processors could be distributed or singular. The memories could be implemented as an electrical, magnetic or optical memory, or any combination of these or other types of storage devices. Moreover, the term "memory" should be construed broadly enough to encompass any information able to be read from or written to an address in the addressable space accessed by an associated processor. With this definition, information on a network is still within a memory because the associated processor can retrieve the information from the network.

[0044] It is to be understood that the embodiments and variations shown and described herein are merely illustrative of the principles of this invention and that various modifications may be implemented by those skilled in the art without departing from the scope and spirit of the invention.

What is claimed is:

1. A method for correcting spelling of at least one candidate word, said method comprising:

obtaining at least one variant dictionary hash table based on variants of a set of known correctly spelled words, wherein said variants are obtained by applying one or

more of a deletion, insertion, replacement, and transposition operation on said correctly spelled words; obtaining from the at least one candidate word one or more lookup variants using one or more of said deletion, insertion, replacement, and transposition operations;

evaluating one or more of said at least one candidate word and said lookup variants against said at least one variant dictionary hash table; and

indicating a candidate correction if there is at least one match in the at least one variant dictionary hash table.

2. The method of claim 1, further comprising the step of obtaining a dictionary hash table having entries in a dictionary of known correctly spelled words and wherein said dictionary hash table and said at least one variant dictionary hash table are based on said dictionary and are comprised of at least one distance one variation for each of said entries, wherein said distance one variation comprises one or more of a deletion, insertion, replacement, and transposition operation performed on said entries; and

wherein the step of evaluating one or more of said candidate word and said lookup variants against said at least one variant dictionary hash table further comprises the step of evaluating one or more distance one variants against said at least one variant dictionary hash table.

3. The method of claim 2, wherein said distance one variation comprises a replacement operation to generate a replacement hash table having entries of single character wild card replacements of said entries in said dictionary and said method further comprises the steps of generating single character replacements and insertions of said candidate word and comparing said single character replacements and insertions against said replacement hash table.

4. The method of claim 3, wherein the replacement hash table is obtained by:

generating variants of each word in the dictionary, each variant is comprised of replacing any one character in the word with a wild card character and leaving other characters unchanged, thereby generating W variants for each word of length W; and

for each generated variant of a word in the dictionary, storing a key-value pair in a hash table, wherein a key is a generated variant having a value that is the word itself.

5. The method of claim 2, further comprising the steps of generating one or more distance one variants of said at least one candidate word and testing said distance one variants against one or more of said dictionary hash table and said at least one variant dictionary hash table, and accumulating matches.

6. The method of claim 5, wherein said one or more distance one variants comprises adjacent character transpositions of said at least one candidate word obtained by generating all variants of the candidate word wherein any one pair of adjacent characters are interchanged, and the remaining characters are left unchanged.

7. The method of claim 5, wherein said one or more distance one variants comprises single character deletions of said at least one candidate word obtained by generating all variants of the candidate word where any single character is deleted and other characters are unchanged.

8. The method of claim 5, wherein said one or more distance one variants comprises single character wild card