

REPLACED SHEETS	SG
☒ Description : paged ...4... to paged ...15...	
☒ Claim(s) : paged ...41... to paged ...51...	
☐ Drawing(s) : paged to paged	
☒ Abstract : paged ...52... to paged ...53...	
DATE OF RECEIPT	

1

LANGUAGE PROCESSING APPARATUS AND METHOD

BACKGROUND OF THE INVENTION

FIELD OF THE INVENTION

The present invention relates to a language processing apparatus and method. The invention mainly relates to a Chinese word inputting apparatus and a Chinese word processing method of a computer and a word processor. The Chinese word inputting apparatus and the Chinese word processing method enable easy, efficient input by an operator having limited knowledge of the Chinese Pinyin system and of a correct Pinyin alphabetic string.

CONVENTIONAL RELATED ART

A Pinyin inputting system has been used in Chinese word processors, with a character pattern cord or GB cord used as an auxiliary input.

Pinyin represents a pronunciation of Chinese characters, using an alphabet system. In Pinyin inputting systems, the pronunciation of Chinese characters is input into the system in Pinyin alphabetic writing via a keyboard and converted into a desired Chinese character string.

- Conversion of a Pinyin alphabetic string into a corresponding Chinese character string does not always

result in the desired Chinese character string, and a plurality of Chinese character strings corresponding to the same Pinyin alphabetic string may exist. In this situation, the plurality of possible Chinese character strings corresponding to the Pinyin alphabetic string is displayed by the system, and an operator of the system selects a desired character string.

A conventional Pinyin inputting system used in a Chinese word inputting apparatus is described using the following examples:

Fig. 28A illustrates a conventional input example 15. In this example, as there are at least three different Chinese characters corresponding to one Pinyin alphabetic writing input to the system, the operator is required to select a desired Chinese character from the three Chinese characters displayed.

When there are a plurality of Chinese characters corresponding to a Pinyin alphabetic string, tone data representing the accent of a Chinese word is appended to the Pinyin alphabetic string and input to the system. The number of corresponding Chinese characters may be reduced, facilitating the operation of the system to a certain extent. There are five kinds of tone data, each of which is provided with numerical information and is designated by numerals such as, for example, 0 to 4 in

accordance with differences of the accent.

Fig. 28B also illustrates a conventional input example 16. As shown in example 16, the Pinyin inputting system using tone data to indicate an accent in a Pinyin alphabetic string is effective when a corresponding Chinese word is comprised of more than two Chinese characters and has a homonym. The tone data do not have to be appended to the entire Pinyin alphabetic string. It may be appended to part of the Pinyin alphabetic string as shown in example 16.

Fig. 28C also illustrates a conventional input example 17 using a Pinyin inputting system similar to that described above.

This system, however, utilizes only initials of a Pinyin alphabetic string corresponding to the desired Chinese characters.

Therefore, the number of inputting operations can be reduced. This system is especially rendered effective where a desired word is comprised of more than two Chinese characters. In the input example 17, two initials of the Pinyin alphabetic string for the desired word are input. The word corresponding to the two initials is selected from a dictionary of the system.

-Another conventional word processing system is disclosed in Japanese Patent Application No. HEI4-167159.

The disclosed system also converts an entered Pinyin alphabetic string into a Chinese character string. However, when an incomplete or inaccurate Pinyin alphabetic string is entered by an operator having limited knowledge of Pinyin, the disclosed system cannot convert the Pinyin alphabetic string into the correct Chinese character string.

In the prior art, there also exists an English dictionary retrieval tool for an English word whose spelling is uncertain, called "The American Heritage Dictionary". A user of this tool enters, for example, "invent", when he is not sure of the spelling of a desired word beginning with "invent". As shown in Fig. 29, "The American Heritage Dictionary" searches for words beginning with "invent", and displays them on a display unit. The user, if required, selects any of the words displayed, and refers to the dictionary for its meaning.

Fig. 30 is an example where a word including one arbitrary letter between "m" and "t" is retrieved and displayed by the retrieval tool described above. When the user enters "m?t", several possible words illustrated in Fig. 30 are displayed. A word including an arbitrary letter string between "m" and "t" can also be obtained. When the user enters "m*t", more than 500 possible words corresponding to the above-mentioned letter string are

displayed. As described before, the conventional English word retrieval tool displays a large number of possible alphabetical letter strings, among which the user identifies the one he originally desired to select.

However, the above-mentioned retrieval system has the disadvantage that, although it can be applied to a language such as English or French which is represented by a single character form, it may not be applied to a language such as Japanese or Chinese represented by a plurality of character forms. Because, in a language represented by a plurality of character forms, one character form must be converted into another character form.

The term "character form" used herein is defined as "a set of letters or characters used in the writing system of a language".

Hereinafter, kanji or the use of Chinese characters is regarded as one character form, while the use of characters such as Japanese hiragana or katakana, representing the pronunciation of kanji, is regarded as another character form.

Japanese Romaji(Roman letters) or Chinese Pinyin, representing the pronunciation of Chinese characters or Chinese words respectively, is also regarded as one character form.

Basically, the conventional Chinese word inputting apparatus and Chinese word processing systems described above do not operate unless the operator knows the correct Pinyin spelling of a word to be input to the system.

When a nonexistent, or incorrect Pinyin alphabetic string is input, it can not be converted into a correct Chinese character string.

In conventional word processing systems such as, for example, the one for processing English which is represented by a single character form, an operator is able to retrieve a desired word easily, even when the operator is not sure of the correct spelling of the word. However, with languages such as Chinese or Japanese which are represented by a plurality of character forms, operators of word processing systems cannot obtain the correct outputted character form unless they know the correct spelling of the character form to be entered.

SUMMARY OF THE INVENTION

The present invention is made to solve the above-mentioned problems. It is, therefore, an object of the present invention to provide a Chinese word inputting apparatus and a Chinese word processing method which

enable easy and efficient input by an operator having limited knowledge of the Pinyin system, e.g., having limited knowledge of the finals of a Chinese word. The term "finals" used herein is employed to mean "a main vowel with or without a final consonant or a semivowel after an initial consonant or consonant cluster of a Chinese syllable".

Further, it is another object of the present invention to provide a Chinese word inputting apparatus and a Chinese word processing method that are compatible with conventional Pinyin inputting systems, and which enable easy and efficient input for both the conventional Pinyin inputting systems and the one according to the present invention.

It is a further object of the present invention to provide a language processing apparatus and a language processing method which enable easy and efficient input when it is required to convert one character form of an input language word into another.

To achieve the foregoing objects, according to one aspect of the present invention, a Chinese word inputting apparatus for inputting an alphabetic string and for converting the alphabetic string into a Chinese character string may comprise:

(a) input means for inputting the alphabetic string,

the alphabetic string having an unknown portion, and for inputting a symbol indicating an unknown alphabetic string in place of the unknown portion of the alphabetic string;

(b) Chinese word syllabifying means for dividing the alphabetic string input by the input means into Chinese syllable segments;

(c) alphabetic string generating means for determining if each of the Chinese syllable segments is comprised of a standard alphabetic string, so as to output the standard alphabetic string when each of the Chinese syllable segments is comprised of a standard alphabetic string, and to generate and output a possible alphabetic string when the symbol indicating an unknown alphabetic string is used in place of the unknown portion of the input alphabetic string;

(d) dictionary means for storing both Chinese character strings and alphabetic information;

(e) Chinese word converting means for converting the standard alphabetic string into a possible Chinese string and for converting the possible alphabetic string into a possible Chinese character string by referring to the dictionary means; and

(f) output means for outputting the possible Chinese character strings converted by the Chinese word

converting means.

According to another aspect of the present invention, a Chinese word inputting apparatus for inputting an alphabetic string having one or more letters and for converting the alphabetic string into a Chinese character string may comprise:

(a) input means for inputting the alphabetic string, the alphabetic string having an unknown letter, and for inputting a symbol indicating an unknown letter in place of the unknown letter of the alphabetic string;

(b) Chinese word syllabifying means for dividing the alphabetic string input by the input means into Chinese syllable segments;

(c) alphabetic string generating means for determining if each of the Chinese syllable segments is comprised of a standard alphabetic string, so as to output the standard alphabetic string when each of the Chinese syllable segments is comprised of a standard alphabetic string, and to generate and output a possible alphabetic string when the symbol indicating the unknown letter is used in place of an unknown letter of the input alphabetic string;

(d) dictionary means for storing both Chinese character strings and alphabetic information;

(e) Chinese word converting means for converting the

standard alphabetic string into a possible Chinese character string and for converting the possible alphabetic string into a possible Chinese character string by referring to the dictionary means; and

(f) output means for outputting the possible Chinese character strings converted by the Chinese word converting means.

According to another aspect of the present invention, a Chinese word inputting apparatus for inputting an alphabetic string and for converting the alphabetic string into a Chinese character string may comprise:

(a) input means for inputting the alphabetic string including an initial letter of the alphabetic string and a symbol indicating tone data appended to the initial letter of the alphabetic string;

(b) Chinese word syllabifying means for dividing the alphabetic string input by the input means into Chinese syllable segments;

(c) dictionary means for storing both Chinese character strings and alphabetic information;

(d) Chinese word converting means for converting the alphabetic string into a possible Chinese character string, and for converting the initial letter and the tone data of the alphabetic string into a possible

Chinese character string by using the tone data to reduce a number of possible corresponding Chinese characters; and

(e) output means for outputting the possible Chinese character strings converted by the Chinese word converting means.

According to another aspect of the present invention, a Chinese word processing method for inputting an alphabetic string and for converting the alphabetic string into a Chinese character string may comprise steps of:

(a) inputting the alphabetic string, the alphabetic string having an unknown portion, and inputting a symbol indicating an unknown alphabetic string in place of the unknown portion of the alphabetic string;

(b) dividing the alphabetic string into Chinese syllable segments;

(c) determining whether each of the Chinese syllable segments is comprised of a standard alphabetic string so as to output the standard alphabetic string when each of the Chinese syllable segments is comprised of a standard alphabetic string, and to generate and output a possible alphabetic string when the symbol indicating an unknown alphabetic string is used in place of the unknown portion of the input alphabetic string;

(d) converting the standard alphabetic string into a possible Chinese character string and converting the possible alphabetic string into a possible Chinese character string by referring to dictionary means containing both Chinese character strings and alphabetic information; and

(e) outputting the possible Chinese character strings.

According to another aspect of the present invention, a Chinese word processing method for inputting an alphabetic string and for converting the alphabetic string into a Chinese character string may comprise steps of:

(a) inputting the alphabetic string, the alphabetic string having one or more letters including an unknown letter, and inputting a symbol indicating an unknown letter in place of the unknown letter of the alphabetic string;

(b) dividing the alphabetic string into Chinese syllable segments;

(c) determining whether each of the Chinese syllable segments is comprised of a standard alphabetic string so as to output the standard alphabetic string when each of the Chinese syllable segments is comprised of a standard alphabetic string, and to generate and output a possible

alphabetic string when the symbol indicating an unknown letter is used in place of the unknown letter of the alphabetic string;

(d) converting the standard alphabetic string into a possible Chinese character string and converting the possible alphabetic string into a possible Chinese character string by referring to dictionary means containing both Chinese character strings and alphabetic information; and

(e) outputting the possible Chinese character strings.

According to another aspect of the present invention, a Chinese word processing method for inputting an alphabetic string and converting the alphabetic string into a Chinese character string may comprise steps of:

(a) inputting the alphabetic string including an initial letter of the alphabetic string and a symbol indicating tone data appended to the initial letter of the alphabetic string;

(b) dividing the alphabetic string into Chinese syllable segments;

(c) converting the alphabetic string into possible Chinese character strings by using the initial letter and the tone data to reduce a number of possible Chinese-character strings; and

(d) outputting the possible Chinese character strings.

According to yet another aspect of the present invention, a language processing apparatus for processing a language represented by at least a first character form and a second character form may comprise:

(a) input means for inputting a character string having an unknown portion, represented by the first character form, the character string including a symbol in place of the unknown portion;

(b) character string generating means for determining whether the symbol indicating an unknown portion is used in the character string input by the input means for replacing the symbol by a grammatically possible character string of the first character form to create a possible character string;

(c) dictionary means for storing character strings of the first character form and corresponding character strings of the second character form;

(d) converting means for converting the possible character string represented by the first character form obtained by the character string generating means into a corresponding possible character string represented by the second character form by referring to the dictionary means; and

(e) outputting the corresponding possible character string represented by the second character form converted by the converting means.

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention will be described in more detail in conjunction of appended drawings, wherein:

Fig. 1 is a perspective view illustrating a Chinese word inputting apparatus according to embodiment 1 of the present invention;

Fig. 2 is a block diagram of the Chinese word inputting apparatus according to embodiment 1 of the present invention;

Fig. 3 shows an example of a Pinyin-Chinese dictionary;

Fig. 4 shows input examples using a symbol "*";

Fig. 5 shows an input example using a symbol "?";

Fig. 6 shows input examples using initials with tone data;

Fig. 7 is a flow chart showing an operation of the Chinese word inputting apparatus according to embodiment 1 of the present invention;

Fig. 8 is a detailed flow chart showing an operation of the Chinese word inputting apparatus according to embodiment 1 of the present invention;

Fig. 9 is a table showing standard Pinyin alphabetic strings;

Fig. 10 is a table showing standard Pinyin alphabetic strings;

Fig. 11 is a table showing standard Pinyin alphabetic strings;

Fig. 12 shows an example where an input Pinyin alphabetic string is divided into syllable segments;

Fig. 13 shows a list of possible Pinyin alphabetic strings corresponding to "do*", generated by the Pinyin alphabetic string generating unit;

Fig. 14 shows a list of possible Pinyin alphabetic strings corresponding to "ji*", generated by the Pinyin alphabetic string generating unit;

Fig. 15 shows methods for setting a word boundary by the word boundary setting unit;

Fig. 16 shows a combination of Pinyin alphabetic strings;

Fig. 17 shows a list of possible Chinese words corresponding to "do*", generated by the Chinese word converting unit;

Fig. 18 shows a list of possible Pinyin alphabetic strings corresponding to "ch?ng", generated by the Pinyin alphabetic string generating unit;

Fig. 19 shows a list of possible Chinese words

corresponding to "ch?ng", generated by the Chinese word converting unit;

Fig. 20 shows an example of a dictionary separately retaining initial information and tone data as a search key;

Fig. 21 shows a an example of a dictionary retaining an initial and tone data as a search key;

Fig. 22 shows input examples using "*", "?", and "an initial and tone data" in Pinyin alphabetic strings;

Fig. 23 shows an input example using "?" in a hiragana character string;

Fig. 24 shows an example of possible hiragana character strings displayed, corresponding to "か?";

Fig. 25 shows an example of a Japanese hiragana-kanji dictionary;

Fig. 26 shows an input example using "*" in a Roman letter string;

Fig. 27 shows an example of a Roman letter-kanji dictionary;

Fig. 28A shows a conventional example of inputting Pinyin alphabetic strings;

Fig. 28B shows a conventional example of inputting Pinyin alphabetic strings;

Fig. 28C shows a conventional example of inputting Pinyin alphabetic strings;

Fig. 29 shows a list of words beginning with "invent" displayed by a conventional retrieval tool called "The American Heritage Dictionary"; and

Fig. 30 shows a list of words including one arbitrary letter between "m" and "t", displayed by "The American Heritage Dictionary".

DESCRIPTION OF THE PREFERRED EMBODIMENTS

Embodiment 1.

This embodiment provides a Chinese word inputting apparatus and a Chinese word processing method which may be operated by an operator having limited knowledge of the Pinyin alphabetic form.

Fig. 1 is a perspective view showing a Chinese word inputting apparatus according to embodiment 1 of the invention. In Fig. 1, reference numeral 100 is a processing unit, 101 is a keyboard, 102 is a display unit, 103 is a disk unit, 104 is a printer, 105 is a mouse, and 106 is a mouse pad.

Fig. 2 is a block diagram of the Chinese word inputting apparatus according to embodiment 1 of the present invention.

In Fig. 2, reference numeral 1 denotes an input unit. The input unit 1 includes the keyboard 101 which

is provided with at least an alphabetical key, a numerical key, a conversion key, and a definition key.

Reference numeral 2 denotes a controller comprising a character string converter 3 and a dictionary 4. When a character string to be converted or a control command is input via the input unit 1, data processing described hereinafter is performed according to a control program.

Fig. 3 illustrates an embodiment of the dictionary 4. The dictionary 4 in Fig. 3 is a portion of a Pinyin-Chinese dictionary indicating correspondence of a Pinyin alphabetic string to a Chinese word.

The character string converter 3 comprises a Chinese word syllabifying unit 5, a Pinyin alphabetic string generating unit 9, a word boundary setting unit 6, and a Chinese word converting unit 7.

The Chinese word syllabifying unit 5 divides an input alphabetic string into syllable segments based on the principle of Chinese syllables. The syllable segments are further divided into an initial consonant or consonant cluster (hereinafter referred to as initial consonant(s)), and finals, or initial consonant(s) plus finals.

The Pinyin alphabetic string generating unit 9 analyzes the content of a subdivided syllable segment and generates a possible, standard Pinyin alphabetic string.

The word boundary setting unit 6 sets the boundary of words according to the Pinyin alphabetic string which has been divided into syllable segments.

The Chinese word converting unit 7 converts a possible Pinyin alphabetic string into a corresponding possible Chinese character string.

The display system 8 displays a converted Chinese character string, and includes the display unit 102.

An operation of this embodiment will now be described with reference to Fig. 1. The keyboard 101 is used as the input unit 1. Processing for the controller 2 is performed by the processing unit 100. Processing for the character string converter 3, including the Chinese word syllabifying unit 5, the Pinyin alphabetic string generating unit 9, the word boundary setting unit 6, the Chinese word converting unit 7 is performed by a central processing unit of the processing unit 100. The dictionary 4 is stored in a main memory or cache memory of the processing unit 100. A portion of the dictionary 4 can be stored in the disk unit 103. The display system 8 includes the display unit 102.

Fig. 4 illustrates an input example where an operator of the system shown in Fig. 1 is uncertain of the spelling of more than two, consecutive letters within

a Pinyin alphabetic string he desires to enter. In this embodiment, the symbol "*" is used to represent more than two, consecutive uncertain letters. Symbols other than "*" can also be used.

In an input example 1(1) in Fig. 4, the operator intends to obtain a Chinese word "東京". The corresponding, correct Pinyin alphabetic string is "dongjing" as shown in (2). The operator, when he cannot remember the correct Pinyin alphabetic string, enters "do*ji*", using "*" to the uncertain portions as shown in (3).

An input example 2 in Fig. 4 shows a case where the operator enters "b*j*" as shown in (6), though he desires to enter a Chinese word "北京" as shown in (4), but is not sure of the spelling of the corresponding Pinyin alphabetic string.

As hereinafter described, a Chinese syllable is comprised of a combination of initial consonant(s) and finals. The above-mentioned "*" will be used when a portion or whole of the finals is unknown.

Fig. 5 is an input example where a Chinese word inputting operator is not sure of the spelling of one letter within the Pinyin alphabetic string he desires to enter. In this embodiment, the symbol "?" is used to represent one, uncertain letter. Symbols other than "?"

can also be used. In an input example 3(1) in Fig. 5, the operator intends to obtain a Chinese word "長城". The correct Pinyin alphabetic string corresponding to (1) is "changcheng" as shown in (2). The operator, when he cannot remember the correct spelling of the Pinyin alphabetic string, enters "ch?ngch?ng", using "?" to each of the unknown letters within the Pinyin alphabetic string.

Fig. 6 is an example where a Chinese word inputting operator enters initials and tone data. The numeral indicates the tone data such as, for example, 0 to 4 (or 1 to 5) after the initial letter of a Pinyin alphabetic string corresponding to a desired Chinese character. Five kinds of other symbols can also be used to indicate the tone data. The tone data is rendered effective where the operator knows only an initial of a Pinyin alphabetic string and is not sure of the spelling of the remainder of the string, or where he desires to omit inputting an entire Pinyin alphabetic string. In an input example 4(1) in Fig. 6, the operator desires to obtain a Chinese word "日本". The corresponding, correct Pinyin alphabetic string is "riben" as shown in (2), where "ri" and "ben" correspond to each of the two Chinese characters respectively. In this case, the operator, when he knows only the initials of the Pinyin alphabetic string and is

not sure of the remainder of the string, or he desires to omit inputting the entire Pinyin alphabetic string, enters "r4b3", using the initials and the tone data.

In an input example 5 in Fig. 6, the operator enters "h2klglsl" as shown in (6), while he desires to enter a Chinese word "航空公司" ("hangkonggongsi" in the Pinyin alphabetic string).

Fig. 7 and Fig. 8 are flow charts illustrating an operation of the disclosed embodiment.

The operation will be first explained with reference to the input example 1 in Fig. 4, where more than two, uncertain alphabetic strings exist.

At step 21, a Pinyin alphabetic string input from the keyboard 101 is read by the controller 2 of the processing unit 100. It is assumed herein that the input Pinyin alphabetic string is "do*ji*".

At step 22, the Chinese word syllabifying unit 5 divides the input Pinyin alphabetic string into Chinese syllable segments.

Figs. 9 to 11 show lists of standard Pinyin alphabetic strings, indicating a correlation of initial consonant(s) and finals in forming a Chinese syllable. As shown in Figs. 9 through 11, a Chinese syllable is comprised of a combination of initial consonant(s) and succeeding finals. There are 21 initial consonant(s) and

37 finals, but as illustrated in Figs. 9 through 11, some combinations of initial consonant(s) and finals do not constitute syllables. A predetermined symbol "*" may be used in place of an uncertain portion of a final, however, initial consonant(s) must always be input correctly. Accordingly, division of an input Pinyin alphabetic string into syllable segments is performed so that each syllable segment begins with an initial consonant(s) in the Pinyin alphabetic string.

Fig. 12 illustrates an example where "do*ji*" is divided into syllable segments. "do*ji*" is divided into syllable segments, "do*" and "ji*", which are hereinafter referred to as a syllable portion 1 and a syllable portion 2, respectively. Division of a Pinyin alphabetic string into syllable segments has conventionally been performed to take advantage of the features of the Chinese language.

At step 23, the Pinyin alphabetic string generating unit 9 receives each of the divided syllable segments and determines whether each of the divided syllable segments is a standard Pinyin alphabetic string or not. If it is, the answer is YES, and the operation proceeds to step 25. If it includes other symbol such as "*" or "?" besides the Pinyin alphabetic string, the answer is NO, and the operation proceeds to step 24.

Since the syllable portion 1 "do*" includes the other symbol "*" besides the Pinyin alphabetic string, the answer is NO, and the operation goes to step 24. A further detailed operation flow of the Pinyin alphabetic string generating unit 9 at step 24 will be explained with reference to Fig. 8.

At step 51 in Fig. 8, it is determined if "*" is included in a syllable segment. If it is not, the answer is NO, and the operation branches off to step 53. The operation at steps following step 53 will be hereinafter described.

Since the syllable segments 1 "do*" includes "*", the answer is YES, and the operation proceeds to step 52.

At step 52, the Pinyin alphabetic string generating unit 9 replaces "*" in a syllable segment by a grammatically possible Pinyin alphabetic string, and generates a list of possible Pinyin alphabetic strings. The Pinyin alphabetic string generating unit 9 retains a table of initial consonant(s) to finals as shown in the lists of the standard Pinyin alphabetic strings in Figs. 9 through 11. The Pinyin alphabetic string generating unit 9 retrieves the table according to the already-known portion of the input alphabetic string, or a combination of the initial consonant "d" and the finals beginning with "o" therefrom, and searches for a grammatically

possible Pinyin alphabetic string. The maximum number of syllables made from a combination of initial consonant(s) and finals are 21 x 37. However, as shown in Figs. 9 through 11, there are cases where a combination of initial consonant(s) and finals does not constitute a syllable. Accordingly, the number of the possible Pinyin alphabetic strings is not as large as 21 x 37.

Fig. 13 illustrates a list of possible Pinyin alphabetic strings generated by the Pinyin alphabetic string generating unit 9. The possible Pinyin alphabetic strings corresponding to "do*" are "dong" and "dou". After the completion of step 52, the operation proceeds to step 25 in Fig. 7.

At step 25, it is determined if the syllable segment is a final one of the Pinyin alphabetic string, and if the answer is YES, the operation proceeds to step 26. Since the syllable segment is not a final one in this case, the answer is NO, and the operation returns to step 23.

At step 23, the Pinyin alphabetic string generating unit 9 receives the succeeding syllable segment and determines whether it is comprised solely of a Pinyin alphabetic string or not. Since the syllable portion 2 "ji*" is not comprised of a Pinyin alphabetic string solely, the same processing as described before is

performed, and a list of possible Pinyin alphabetic strings illustrated in Fig. 14 is obtained.

For each of the syllable segments of an input Pinyin alphabetic string, the operation from steps 23 through 25 is repeated, and after completion, the operation goes to step 26. In the case of "do*ji*", as the operation for both the syllable segments 1 and 2 has been completed, the operation proceeds to step 26.

With reference to the operation after step 26, it is not the essential point of the present invention whose operating advantage resides in replacing an uncertain character by a predetermined symbol. The operations of step 26 and thereafter will therefore be made only roughly.

At step 26, the word boundary setting unit 6 sets the word boundary. The word boundary will now be described with reference to Fig. 15. In the case of a word which is comprised of three syllable segments, there are provided the following three methods for setting its word boundary:

- (1) Setting all three syllable segments to be one word
- (2) Setting the adjoining two syllable segments among the three to be one word
- (3) Setting each syllable segment to be one word

In this embodiment, at step 26 and step 27, the word

boundary setting method is adopted wherein the longest length of a combination of syllable segments is set to be one word boundary. The operation is then performed at steps 26 through 28, and a list of possible Chinese words is generated. At step 29, when it is judged that the most suitable Chinese word is not included in the list, the operation returns to step 26 again. At step 26, the previous word boundary setting is reviewed, the combination of syllables which is one syllable less than the previous one is set to be one word, and a possible word is searched again.

When a word is comprised of three syllable segments as shown in Fig. 15, for example, the combination of three syllable segments is first set to be one word, and for which a suitable Chinese word is searched. When the suitable word is not found, then, a combination of two syllable segments is set to be one word, and a possible word is searched again. In this case, when the suitable word is not found again, each syllable segment is set to be one word. When each syllable segment is set to be one word and there are no inputting errors, the suitable Chinese word may be found without fail.

In the case of "do*ji*", the two syllable segments are first set to be one word at step 26. Combinations of the Pinyin alphabetic strings are shown in Fig. 16. The

number of combinations of the syllable segment 1 and the syllable segment 2 becomes 14 (2×7).

At step 27, the Chinese word converting unit 7 generates a list of possible Chinese words for the combinations of possible Pinyin alphabetic strings, referring to the dictionary 4. For each of the 14 combinations of Pinyin alphabetic strings as shown in Fig. 16, the dictionary 4 is referred to, and it is examined whether the corresponding Chinese words exist or not. When the corresponding Chinese words exist, they are included in the list of possible Chinese words. Fig. 17 illustrates a Chinese word list generated. The Pinyin alphabetic strings shown in parenthesis are only for reference, and are not actually displayed.

At step 28, the Chinese word converting unit 7 selects and displays a suitable Chinese word on the display unit 102. In this situation, to select the most suitable word, priority should be placed on either a relationship with words which have been previously referred to, or words frequently used. Alternatively, the entire list or a portion of the list of possible Chinese words is displayed on the display unit 102, among which the user may select the most suitable one at step 29.

At step 29, the user judges whether a possible

Chinese word displayed on the display unit 102 is the intended one or not. While each of the possible Chinese words is displayed on the display unit 102, the user continues judging until he identifies and selects the most suitable word displayed, or until no more possible Chinese words of the list appear on the screen. When the most suitable word is not found, the operation returns to step 26.

In this example, since "東京" is the intended Chinese word, the user selects it from the display, and the operation proceeds to step 30.

At step 30, the word boundary setting unit 6 determines whether the last syllable segment is included in the word boundary being processed. When the answer is NO, the operation returns to step 26, and processing is performed again for the remainder of the syllable segments at subsequent steps so as to find a suitable Chinese word. At step 30, when the answer is YES, the operation is finished. In this example, since the last syllable segment is included, the operation is finished.

Word processing performed in the situation where an uncertain letter in the Pinyin alphabetic string is indicated by a symbol such as, for example, "?", will now be described with reference to the input example 3 in Fig. 5. The description of similar processing to the

example using the symbol "*" will be omitted.

At step 21, "ch?ngch?ng" is input, and is divided into syllable segments "ch?ng" and "ch?ng" at step 22. At step 23, it is examined whether a symbol besides the Pinyin alphabetic string is included or not. In this case, "?" is included, the operation proceeds to step 24.

Now, detailed description of the processing at step 24 will be made with reference to Fig. 8.

At step 51, the Pinyin alphabetic string generating unit 9 determines whether "*" is included in the syllable segment or not. The answer is NO, and the operation proceeds to step 53. At step 53, the Pinyin alphabetic string generating unit 9 determines whether "?" is included or not. If the answer is NO, error processing is performed at step 55.

Since "?" is included in "ch?ng", the answer is YES, and the operation shifts to step 54.

At step 54, the Pinyin alphabetic string generating unit 9 retrieves the table which has been described before, and replaces "?" in the syllable segment by one grammatically possible Pinyin alphabetic writing, and generates a list of possible Pinyin alphabetic strings.

Fig. 18 illustrates a list of possible Pinyin alphabetic strings generated by the Pinyin alphabetic string generating unit 9. The Pinyin alphabetic strings

of "chang", "chong", "cheng" are listed, corresponding to "ch?ng".

When step 54 is finished, the operation proceeds to step 25 in Fig. 7.

The second syllable segment, "ch?ng" is the same as the first syllable segment. Accordingly, a list of possible Pinyin alphabetic strings is as shown in Fig. 18.

Processing at steps 26 through 30 is the same as previously described before.

Fig. 19 illustrates a list of Chinese words displayed when the two syllable segments are set to be one word.

Now, the operation of inputting an initial with tone data will be described.

A processing performed when inputting an initial with tone data will be described with reference to the input example 4 in Fig. 6.

At step 21, the Pinyin alphabetic string input via the keyboard 101 of the input unit 1 is entered to the controller 2. In this example, the input Pinyin alphabetic string is "r4b3".

At step 22, the Chinese word syllabifying unit 5 divides the input Pinyin alphabetic string into Chinese syllable segments. In dividing into syllable segments,

the Chinese word syllabifying unit 5 recognizes an initial appended by tone data and divides the Pinyin alphabetic string into "r4" and "b3" according to the system of the present invention wherein an initial and tone data is set to be one syllable segment.

At step 23, the Pinyin alphabetic string generating unit 9 receives the divided syllable segment from its beginning, and determines whether it is comprised solely of the Pinyin alphabetic string or not. "r4", which is an initial and tone data, is determined to be a Pinyin alphabetic string. Therefore, the answer is YES, and the operation proceeds to step 25.

At step 25, it is determined whether the syllable segment is a final one or not. In this case, the answer is NO, and the operation returns to step 23.

At step 23, the Pinyin alphabetic string generating unit 9 receives the succeeding syllable segment "b3" and determines if it is comprised solely of the Pinyin alphabetic string. Since "b3" is the Pinyin alphabetic string, the operation is forwarded to step 25.

At step 25, when it is determined that all syllable segments have been examined, the operation goes to step 26.

At step 26, the word boundary setting unit 6 sets the word boundary. In the case of "r4b3", the two

syllable segments are first set to be one word.

At step 27, the dictionary 4 illustrated in Fig. 20 is referred to, and a list of possible Chinese words is generated. A form of the table illustrated in Fig. 21 can also be used as the dictionary 4. The items of "the Pinyin alphabetic strings" in Fig. 20 and Fig. 21 are only for reference, and are not actually included in the dictionary. The advantage of the dictionary illustrated in Fig. 20 is that, since it separately retains initial information and tone data as a search key, the initial information can be used alone even when only the initials of the Pinyin alphabetic string corresponding to each of Chinese characters have been input without tone data. However, it is required that the input alphabetic string be processed into the alphabetic string comprised of the initials alone, when the alphabetic string is input in the form of an initial with tone data. The dictionary illustrated in Fig. 21 retains an initial and tone data as a search key, the input alphabetic string can be used for a dictionary search without being processed. A Chinese word corresponding to "r4b3" is "日本" according to the dictionary 4, and it is included in the list of Chinese words.

At step 28, the Chinese word converting unit 7 selects a suitable Chinese word from the list of Chinese

words and displays it on the display unit 102. In this case, "日本" is displayed.

When inputting an initial with tone data, the tone data can be used to reduce the number of possible Chinese words to be displayed, and a conversion result of "日本" can be obtained.

As described above, when the user has limited knowledge of the Chinese Pinyin system, and is not sure of the correct spelling of a Pinyin alphabetic string to be entered, symbols such as "*", and "?" can be used to replace an uncertain portion of the Pinyin alphabetic string. Abbreviated input of the Pinyin alphabetic string is also possible, using an initial and tone data. A Chinese word input apparatus with good operability for the user can be provided accordingly.

In the system of the present invention, an input Pinyin alphabetic string is divided into syllable segments, each of which corresponding to one Chinese character, and the above-mentioned processing is performed. Consequently, for conversion of a word comprised of more than two uncertain letters into a Chinese character string, three kinds of inputting systems using "*", "?", and "an initial and tone data" may be applied individually or simultaneously.

Fig. 22 illustrates examples using three kinds of

inputting systems individually or simultaneously.

An input example 6 of Fig. 22 illustrates a Pinyin alphabetic string which is to be converted into two Chinese characters with the symbol "*", indicating that one unknown letter string is used in place of the uncertain Pinyin alphabetic string corresponding to one Chinese character of the two.

An input example 7 illustrates a Pinyin alphabetic string which is to be converted into two Chinese characters with the symbol "?", indicating that one unknown letter is used in place of the uncertain Pinyin alphabetic writing corresponding to one Chinese character of the two.

An input example 8 shows a Pinyin alphabetic string which is to be converted into three Chinese characters with the symbol "*" indicating an unknown letter string and the symbol "?" indicating an unknown letter used together in place of the uncertain portions of the Pinyin alphabetic string corresponding to two Chinese characters of the three.

An input example 9 shows a Pinyin alphabetic string which is to be converted into four Chinese characters with two initials of the Pinyin alphabetic string respectively corresponding to the first and second Chinese characters of the four input, and the first

initial is appended by tone data.

As described hereinbefore, the present invention can provide a Chinese word inputting apparatus and a Chinese word processing method which enable easy, efficient input by an operator having limited knowledge of the Chinese Pinyin system e.g., where the operator's knowledge of the finals of a Chinese syllable is insufficient.

Further, the present invention can provide a Chinese word inputting apparatus and a Chinese word processing method that are compatible with conventional Pinyin inputting systems, and which enable easy and efficient input for both conventional Pinyin inputting systems and systems in accordance with embodiments of the present invention.

Embodiment 2.

The second embodiment of the present invention is directed to a Japanese word processing apparatus. A Japanese sentence can be represented by a combination of three kinds of character forms including hiragana, katakana, and kanji. The Chinese system for inputting a Pinyin alphabetic string is similar to systems for inputting a hiragana character string, katakana character string, or Roman letter string in Japanese. In this embodiment, it is assumed that only nouns are converted,

and the following description is made with reference to examples of converting hiragana into kanji, or converting a Roman letter or Roman letters into kanji.

Fig. 23 illustrates an input example where Japanese hiragana is converted into kanji. In Fig. 23, (1) shows the case where a Japanese word inputting operator intends to obtain a Japanese kanji "海外". (2) denotes a hiragana character string corresponding to (1). The operator, when he cannot remember the hiragana character string accurately, enters "かいが?", using "?" to an uncertain character as shown in (3).

In this embodiment of the present invention, a Japanese word processing apparatus is provided which converts a hiragana character string into a corresponding kanji character string in a similar manner as in the Chinese word inputting apparatus in accordance with the first embodiment of the present invention. Fig. 24 shows a list of possible hiragana character strings displayed when converting "?" into a corresponding hiragana character string. In listing possible hiragana character strings, 74 characters including hiragana characters from "あ" through "ん", "か", "き", "こ", "け", "け", "こ", "け", "け", etc. are possibly displayed in view of the grammar for the Japanese language. The dictionary stores hiragana and its corresponding kanji. Fig. 25 illustrates an example

of a Japanese hiragana-kanji dictionary.

Fig. 26 illustrates an input example where Roman letters are converted into kanji. The Roman letter string corresponding to Chinese characters "海外" is "kaigai" as shown in (2) of Fig. 26. The operator, when he cannot remember the Roman letter string accurately, enters "kaig*", using "*" to an uncertain letter string as shown in (3). When listing possible Roman letter strings, the possible Roman letter strings should be comprised of the letters selected from 26 Roman letters. Fig. 27 shows an example of a Roman letter-kanji dictionary.

As described before, when the hiragana character string or Roman letter string is uncertain, a symbol such as "*" or "?" can be used in place of an uncertain hiragana character string.

In a language such as Chinese or Japanese which is represented by a plurality of character forms, when a character string represented by the first character form to be input is uncertain, a symbol can be used in place of the uncertain character string, and a possible character string represented by the second character form can be obtained. According to the present invention, a language processing apparatus which can be easily operated by the user can be provided. In the present

embodiment, symbols such as "*" and "?" are used in place of the uncertain character strings. However, other symbols can also be used.

Having thus described several particular embodiments of the invention, various alterations, modifications, and improvements will readily occur to those skilled in the art. Such alterations, modifications, and improvements are intended to be part of this disclosure, and are intended to be within the spirit and scope of the invention. Accordingly, the foregoing description is by way of example only, and not intended to be limiting. The invention is limited only as defined in the following claims and the equivalents thereto.



What is Claimed is:

1. A Chinese word inputting apparatus for inputting an alphabetic string and for converting the alphabetic string into a Chinese character string, comprising:

(a) input means for inputting the alphabetic string, the alphabetic string having an unknown portion, and for inputting a symbol indicating an unknown alphabetic string in place of the unknown portion of the alphabetic string;

(b) Chinese word syllabifying means for dividing the alphabetic string input by the input means into Chinese syllable segments;

(c) alphabetic string generating means for determining if each of the Chinese syllable segments is comprised of a standard alphabetic string, so as to output the standard alphabetic string when each of the Chinese syllable segments is comprised of a standard alphabetic string, and to generate and output a possible alphabetic string when the symbol indicating an unknown alphabetic string is used in place of the unknown portion of the input alphabetic string;

(d) dictionary means for storing both Chinese character strings and alphabetic information;

(e) Chinese word converting means for converting the standard alphabetic string into a possible Chinese string



and for converting the possible alphabetic string into a possible Chinese character string by referring to the dictionary means; and

(f) output means for outputting the possible Chinese character strings converted by the Chinese word converting means.

2. A Chinese word inputting apparatus for inputting an alphabetic string having one or more letters and for converting the alphabetic string into a Chinese character string, comprising:

(a) input means for inputting the alphabetic string, the alphabetic string having an unknown letter, and for inputting a symbol indicating an unknown letter in place of the unknown letter of the alphabetic string;

(b) Chinese word syllabifying means for dividing the alphabetic string input by the input means into Chinese syllable segments;

(c) alphabetic string generating means for determining if each of the Chinese syllable segments is comprised of a standard alphabetic string, so as to output the standard alphabetic string when each of the Chinese syllable segments is comprised of a standard alphabetic string, and to generate and output a possible alphabetic string when the symbol indicating an unknown letter is used in place of the unknown letter of the



input alphabetic string;

(d) dictionary means for storing both Chinese character strings and alphabetic information;

(e) Chinese word converting means for converting the standard alphabetic string into a possible Chinese character string and for converting the possible alphabetic string into a possible Chinese character string by referring to the dictionary means; and

(f) output means for outputting the possible Chinese character strings converted by the Chinese word converting means.

3. A Chinese word inputting apparatus for inputting an alphabetic string and for converting the alphabetic string into a Chinese character string, comprising:

(a) input means for inputting the alphabetic string, including an initial letter of the alphabetic string and a symbol indicating tone data appended to the initial letter of the alphabetic string;

(b) Chinese word syllabifying means for dividing the alphabetic string input by the input means into Chinese syllable segments;

(c) dictionary means for storing both Chinese character strings and alphabetic information;

(d) Chinese word converting means for converting the alphabetic string into a possible Chinese character



string, and for converting the initial letter and the tone data of the alphabetic string into a possible Chinese character string by using the tone data to reduce a number of possible corresponding Chinese characters; and

(e) output means for outputting the possible Chinese character strings converted by the Chinese word converting means.

4. A Chinese word processing method for inputting an alphabetic string and for converting the alphabetic string into a Chinese character string, comprising steps of:

(a) inputting the alphabetic string, the alphabetic string having an unknown portion, and inputting a symbol indicating an unknown alphabetic string in place of the unknown portion of the alphabetic string;

(b) dividing the alphabetic string into Chinese syllable segments;

(c) determining whether each of the Chinese syllable segments is comprised of a standard alphabetic string so as to output the standard alphabetic string when each of the Chinese syllable segments is comprised of a standard alphabetic string, and to generate and output a possible alphabetic string when the symbol indicating an unknown alphabetic string is used in place of the unknown portion



of the input alphabetic string;

(d) converting the standard alphabetic string into a possible Chinese character string and converting the possible alphabetic string into a possible Chinese character string by referring to dictionary means containing both Chinese character strings and alphabetic information; and

(e) outputting the possible Chinese character strings.

5. A Chinese word processing method for inputting an alphabetic string and for converting the alphabetic string into a Chinese character string, comprising steps of:

(a) inputting the alphabetic string, the alphabetic string having one or more letters including an unknown letter, and inputting a symbol indicating an unknown letter in place of the unknown letter of the alphabetic string;

(b) dividing the alphabetic string into Chinese syllable segments;

(c) determining whether each of the Chinese syllable segments is comprised of a standard alphabetic string so as to output the standard alphabetic string when each of the Chinese syllable segments is comprised of a standard alphabetic string, and to generate and output a possible



alphabetic string when the symbol indicating an unknown letter is used in place of the unknown letter of the alphabetic string;

(d) converting the standard alphabetic string into a possible Chinese character string and converting the possible alphabetic string into a possible Chinese character string by referring to dictionary means containing both Chinese character strings and alphabetic information; and

(e) outputting the possible Chinese character strings.

6. A Chinese word processing method for inputting an alphabetic string and for converting the alphabetic string into a Chinese character string, comprising steps of:

(a) inputting the alphabetic string, including an initial letter of the alphabetic string and a symbol indicating tone data appended to the initial letter of the alphabetic string;

(b) dividing the alphabetic string into Chinese syllable segments;

(c) converting the alphabetic string into possible Chinese character strings by using the initial letter and the tone data to reduce a number of possible Chinese character strings; and



(d) outputting the possible Chinese character strings.

7. A language processing apparatus for processing a language represented by at least a first character form and a second character form, comprising:

(a) input means for inputting a character string having an unknown portion, represented by the first character form, the character string including a symbol in place of the unknown portion;

(b) character string generating means for determining whether the symbol indicating an unknown portion is used in the character string input by the input means, for replacing the symbol by a grammatically possible character string of the first character form to create a possible character string;

(c) dictionary means for storing character strings of the first character form and corresponding character strings of the second character form;

(d) converting means for converting the possible character string represented by the first character form obtained by the character string generating means into a corresponding possible character string represented by the second character form by referring to the dictionary means; and

(e) outputting the corresponding possible character



string represented by the second character form converted by the converting means.

8. The language processing apparatus according to claim 7, further comprising syllabifying means for dividing the character string, input by the input means and represented by the first character form, into syllables corresponding to a character string represented by the second character form.

9. The language processing apparatus according to claim 7 or claim 8, wherein the character string represented by the first character form includes a known character string and a symbol indicating an unknown character string in place of an unknown portion of the character string.

10. The language processing apparatus according to claim 9, wherein the symbol indicating the unknown character string corresponds to a single character.

11. The language processing apparatus according to claim 9, wherein the symbol indicating the unknown character string corresponds to a plurality of characters.

12. The language processing apparatus according to claim 10 or claim 11, wherein the language of the language processing apparatus is Chinese, wherein the first character form is Pinyin alphabetic writing, wherein the dictionary means stores Pinyin alphabetic strings and



Chinese character strings corresponding to Pinyin alphabetic strings, and wherein the converting means converts a possible Pinyin alphabetic string into a corresponding Chinese character string by referring to the dictionary means.

13. The language processing apparatus according to claim 12, wherein the character string generating means includes a table for storing combinations of grammatically possible Pinyin alphabetic strings and the character string generating means generates a grammatically possible Pinyin alphabetic string, by using the table.

14. The language processing apparatus according to claim 8, wherein the language of the language processing apparatus is Chinese, wherein the input means inputs an initial letter of a Pinyin alphabetic string and a symbol indicating tone data appended to the initial letter of the Pinyin alphabetic string, wherein the dictionary means stores Chinese character strings corresponding to initial letters of Pinyin alphabetic strings and tone data, and wherein the converting means converts the initial letter of the Pinyin alphabetic string and the tone data into a possible Chinese character string.

15. The language processing apparatus according to claim 10 or claim 11, wherein the language of the language



processing apparatus is Japanese, and wherein the first character form is Japanese hiragana.

16. A language processing method for processing a language that may be represented by at least a first character form and a second character form, comprising steps of:

(a) inputting a character string represented by the first character form, the character string having a known portion including at least one character and an unknown portion including a symbol indicating an unknown character string;

(b) determining whether the symbol indicating an unknown character string is used in the character string input by the step of inputting and generating a grammatically possible character string represented by the first character form when the symbol is used;

(c) converting the grammatically possible character string represented by the first character form generated by the step of generating a grammatically possible character string into a corresponding character string represented by the second character form by referring to a dictionary that contains character strings represented by the first character form and corresponding character strings represented by the second character form; and

(d) outputting the corresponding character string



represented by the second character form converted by the step of converting.

17. The language processing method according to claim 16, further comprising a step of dividing an input character string into a character string represented by the first character form indicating syllables corresponding to a character string represented by the second character form.

18. The language processing method according to claim 16 or claim 17, wherein in the step of inputting the symbol indicating an unknown character string corresponds to a single character.

19. The language processing method according to claim 16 or claim 17, wherein in the step of inputting the symbol indicating an unknown character string corresponds to a plurality of characters.

20. The language processing method according to claim 16, claim 17, claim 18, or claim 19, wherein the step of inputting inputs a Pinyin alphabetic writing as the first character form, and wherein the step of converting converts the Pinyin alphabetic string into a Chinese character string corresponding to the Pinyin alphabetic string.