

High speed and compact methods for string retrieval by a double array

Associate Professor Masao Fuketa

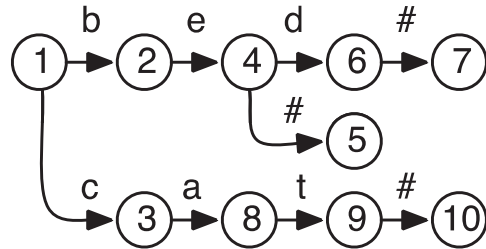


Figure 1 Example of a trie

BASE	1	1	2	1	-1	3	-2	1	6	-3
CHECK		1	1	2	4	4	6	3	8	10
	#	a	b	c	d	e	t			
CODE	4	6	1	2	5	3	7			

Figure 2 Example of a double array

Content:

Retrieving strings is used in many applications and a very important technology. Retrieval speed and saving memory are required for string retrieval. A trie (Figure 1) is one of data structures to retrieve strings, and a double array (Figure 2) is one of retrieval methods by using the trie. As large string sets are frequently used due to development of Internet, compact data structures such as LOUDS are used. The speed of the compact data structure is very slow compared with the double array.

Hence, my research is to save memory with maintaining the high speed of the double array. By constructing the double array for each depth of trie, values of BASE and CODE are determined for each depth, and then reduction of number of bits representing BASE values and saving memory of double array are achievable. Moreover, a method to reduce number of bits representing CHECK values is under investigation.

Furthermore, this research applies to similar string retrieval and DNA sequence retrieval .

Keywords: trie, data compression, dababase

E-mail: fuketa@is.tokushima-u.ac.jp

Tel. +81-88-656-7564

Fax: +81-88-655-4424