Faculty of Engineering
Tokushima University

ダブル配列による高速かつ省メモリな文字列検索手法

[キーワード:トライ，データ圧縮，データベース]　准教授　泓田　正雄

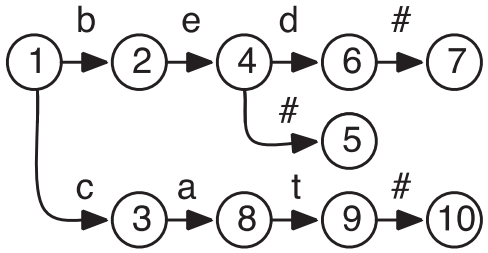

図1　トライの例

	1	2	3	4	5	6	7	8	9	10
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

図2　ダブル配列の例

内容:

文字列検索は，多くのアプリケーションで使われており，非常に重要な技術である．文字列検索には，検索スピードとコンパクトなメモリ使用量が求められている．キー検索をするためのデータ構造であるトライ（図1）を用いた実装方法の一つであるダブル配列（図2）は，高速性とコンパクト性をあわせ持つ手法である．インターネットの発達により，大規模な文字列集合を扱うことが多くなり，LOUDSなどのさらにコンパクトなデータ構造が使われる場合があるが，検索速度はダブル配列より遅くなっている．

そこで，ダブル配列の高速性を維持したまま，メモリ使用量を少なくする研究を行っている．トライの深さごとにダブル配列を構築することにより，階層ごとにBASEやCODEの値を決定することができるので，BASE配列を表現するバイト数の削減することができ，ダブル配列のメモリ使用量を少なくすることができる．また，CHECK配列を表現するバイト数の削減する手法についても研究中である．

さらに，高速な類似文字列検索や，DNAの塩基配列の検索などの応用を考えている．

分野:メディア情報学科・データベース

専門:情報検索，自然言語処理

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

Tel.　088-656-7564

Fax:　088-655-4424