様式８

論　文　内　容　要　旨

報告番号	甲　先　第　454　号	氏　名	松本　拓真
学位論文題目	ダブル配列辞書の時間効率		

内容要旨

文字列は計算機システムにおける最も基本的なデータ表現方法の一つであり，大量の文字列データを扱う現代において，文字列の集合を効率よく管理することは非常に重要である．文字列集合を管理するデータ構造の多くは，主にハッシュテーブルや，トライと呼ばれるラベル付きグラフに基づいている．ダブル配列はラベル付きグラフを表現するデータ構造の一つであり，ラベル付きグラフの状態遷移表を1 次元に圧縮した構造として表現される．ダブル配列の特徴は，入力文字列に基づく検索を高速に実行できる反面，構築速度が遅く，メモリを比較的多く消費するというボトルネックを持つ．その特徴から，数百MB から数GB 程度のデータに対するキーの追加や削除などの更新を必要としない静的辞書として利用されることが多い．本研究では，ダブル配列のボトルネックの改善に取り組む．具体的には，構築アルゴリズムの根幹であるXCHECK と呼ばれる計算の高速化と，更新が容易でありノード数の少ないグラフ表現であるパトリシアトライをダブル配列で表現する方法を提案する．XCHECK の高速化では，ダブル配列の要素あたり1bit のデータを追加する代わりに，構築速度の数倍から数十倍の改善を実現した．また，パトリシアトライを用いた動的キーワード辞書の実装では，自然言語からなるデータセットでダブル配列に基づく従来の技法と性能比較すると，メモリと検索速度を改善した上，XCHECK の高速化と合わせることで同程度の構築時間を維持した．