別記様式第３号　　課程博士用

平成 17 年 2 月 28 日

豊橋技術科学大学長　殿

審査委員長　中川 聖一 印

論文審査及び最終試験の結果報告書

このことについて，下記の結果を得ましたので報告いたします。

学位申請者	武田 善行	学籍番号	第９８３４２５号
申請学位	博士（工学）	専攻名	電子・情報工学専攻
論文題目	全文字列の統計量に基づく索引語分析に関する研究		
公開審査会の日	平成 17 年 2 月 23 日		
論文審査の期間	平成 17 年 1 月 26 日～平成 17 年 2 月 28 日	論文審査の結果	合格
最終試験の日	平成 17 年 2 月 23 日	最終試験の結果	合格

論文内容の要旨

本論文では，大規模なテキストに対し，そこに含まれるすべての文字列を対象に分析を行うことを出発点として，情報検索に適する索引語の選択対象を単語から全文字列へ広げる手法について述べている．まず，索引語を通常の単語から選ぶことが暗黙のうちに可能性を制限することであることを論じたのち，索引語をあらゆる文字列から選ぶという考え方を示している．選ぶ範囲を拡大したときに有効なものを選び出すことができるかどうか明らかでないという問題に対しては，それを解決する具体的な方法を示し，それが実際に情報検索で効果があることを示している．また，その選択方法で生じた計算量の増大という問題に対し，他の問題にも広く応用できる対処法を明らかにしている．

第1章は序論であり，本研究の目的と背景を説明している．ここにおいて，すべての文字列を分析することの価値と，その難しさについて議論されている．第2章では，全文字列と索引語に適する文字列とを対比しながら大規模な統計的な分析を行っている．分析は，日本語ばかりでなく中国語についても行われ，また，新聞ばかりでなく技術的な文書についても行われている．第3章では，統計分析の結果を利用して，コンピュータが人間にとって索引語と思われるような文字列を特定する手法を示している．この手法は，辞書を使用しないという条件で実現されている．第4章では，コンピュータが選別した索引語が人間による選別結果より検索性能が高いことを示している．第5章では，大規模な文字列の統計的な分析のために文字列を効率的に同じ分布をもつ同値類に分類する手法を提案している．その方法は計算量がコーパスの大きさに比例し，実際に大規模なテキストで動作の確認を行っている．第6章は，結論であり，本論文のまとめと今後の課題について述べている．

審査結果の要旨

すべての文字列を対象とするというアプローチは，表明することは容易だが実際に実行することは難しいと考えられていた．それゆえ，このようなアプローチはコンピュータの能力が不十分であった過去には実行されていないアプローチであり，新規性があり，今後，重要性が高まる提案と判断できる．全文字列に対する統計分析を具体的に示した結果は，日本語や中国語などの単語の境界が明らかでない言語について，長い文字列を処理対象とするという方法の実例を示したものであり，これを参考に多くの新しい研究を生み出しうる，基礎的で学術価値の高い結果と考えられる．また，多くの言語処理の研究において，最初に形態素解析システムを走行させ単語を切り出して処理するという方法は広く行われているが，その処理は未知語を処理できないという重要で難しい問題をもつ．この問題に対し，ひとつの解決方法を示したことは，多くの自然言語処理の研究に対して参考となる波及効果の高い結果である．さらに，このアプローチを実現するときには，コンピュータ処理に時間がかかるという問題があるが，統計的な性質が等価な同値類を作成することで対処することを示している．これは，すべての文字列を対象としているのにもかかわらず，その同値類が単語の数と比較できる程度までに縮小されるという結果であり，単語だけを処理対象としていた研究者にも文字列の分析を興味の対象とさせ得る結果であり，論文で示された問題提起を多くの研究者にも受け入れやすくする効果を持つ結果と考えられる．実行が難しい問題を提起した上で，実際に実行し，その問題を実際に解く方法を示しているため，工学的な応用性，発展性からも高く評価できる．これらの結果は学術論文２編で公表されている．よって，本論文は博士（工学）の学位論文に相当するものと判定した．

審査委員

中川聖一 印　　増山繁 印　　青野雅樹 印

梅村恭司 印　　印　　印

（注）　論文審査の結果及び最終試験の結果は「合格」又は「不合格」の評語で記入すること。