

## 審査の結果の要旨

氏名 ソンブーンウィワット クンワディー

本論文は「**Research on Language Specific Crawling and Building of Thai Web Archive** (言語特定クローリングとタイウェブアーカイブの構築に関する研究)」と題し、英文8章から構成されている。ウェブ上から特定の言語で表記されたページを効率よく取り込む(クローリング)方式を提案し、ウェブの実データを用いた実験を行い、提案する方式の有効性を論じると共に、タイ語で記述されたウェブアーカイブを構築、特定言語ウェブの特性について論じている。

第1章は、「**Introduction** (序章)」であり、本論文の背景および目的について概観し、本論文の構成を述べている。

第2章は、「**Related Work** (関連研究)」と題し、ウェブアーカイビング、目的に特化したクローリング、ウェブ構造の解析手法に関し、関連研究をまとめている。

第3章は、「**Language Specific Crawling** (言語特定クローリング)」と題し、特定言語で記述されたページのみを効率良く収集するための手法を提案している。本提案手法では、特定言語で記述されたページ探索戦略として、該当言語ページ内のリンク先の言語種別およびリンクホップ数、特定言語が使われているドメインのサーバ情報を利用することで、単純な幅優先探索、あるいは、特定言語ページへのリンクのみを利用する場合に比べ短時間に効率よく当該言語ページの収集が可能であることを示している。提案方式を詳述すると共に、当該手法の正しさ、および効率について、ウェブから収集されたデータセットを用いたシミュレーション実験を行い、その有効性を示している。

第4章は、「**Thai Web Archive** (タイウェブアーカイブ)」と題し、第3章で提案した言語特定クローリングを実装し、実際に2007年に5回のタイウェブスナップショットの収集を行い、その結果得られたタイウェブアーカイブの基礎データを示している。

第5章は、「**Thai Web Community** (タイウェブコミュニティ)」と題し、タイウェブスナップショットからコミュニティを抽出し、コミュニティ内のタイ言語で記述されたページ比率を解析すると共に、コミュニティ内のリンク情報からクローリングされていないページの特性を解析し、タイに関連する内容でありながらタイ言語以外で記述されたページの収集が可能であることを示している。その結果、提案している言語特定クローリングに当該言語のコミュニティ解析情報を加え、当該言語を用いる国、地域に関連したページの収集について議論、検討を行っている。

第6章は、「**Characteristics of Thai Web** (タイウェブの特徴)」と題し、収集したタイウェブスナップショットを用い、タイウェブグラフのリンク構造解析を行い、その形状を明らかにしている。また、日本語など他言語のウェブグラフのリンク構造と比較を行い、言語間の構造の差異を論じている。さらに、ウェブアーカイブのコンテンツ解析として、

タイウェブアーカイブから社会的に話題となり得る語を含むページを抜き出し、その特性について議論、検討を行っている。

第7章「Conclusions and Future Work (結論)」では、本論文の成果と今後の課題について総括している。

以上これを要するに、本論文は、言語特定クローリング手法に関し、ページ内リンク情報、特定言語ドメイン内のサーバ情報に着目することにより、当該言語ページの収集の効率化を図るものであり、シミュレーション解析により有効性を明らかにすると共に、実ウェブからタイウェブアーカイブを構築し、ウェブコミュニティ解析を用いることで、未収集のページのみならず当該言語以外で記述されている関連ページをも発見可能であることを明らかにしており、電子情報学上貢献するところが少なくない。

よって本論文は博士（情報理工学）の学位請求論文として合格と認められる。