【35】

氏　名（本籍）	法隆大輔（東京都）
学位の種類	博　士（農　学）
学位記番号	博甲第4498号
学位授与年月日	平成19年9月30日
学位授与の要件	学位規則第4条第1項該当
審査研究科	生命環境科学研究科
学位論文題目	**農業関連文書のテキストマイニングにおける語彙資源の利用に関する研究**

主　査	筑波大学教授（連係大学院）	農学博士	二宮正士
副　査	筑波大学教授	農学博士	永木正和
副　査	筑波大学教授（連係大学院）	博士（農学）	平藤雅之
副　査	筑波大学准教授（連係大学院）	博士（農学）	竹澤邦夫

論文の内容の要旨

テキストマイニングは，大量のテキストデータを要約，分類，あるいは統計的な処理をすることによって，そのテキストデータの中にどのようなことが書かれているか簡潔に提示するための一連の技術である。電子計算機の普及とともに，多くのテキストデータが蓄積されるようになったことから，テキストマイニングの技術が重要になっている。テキストマイニングでは，分析の過程において辞書やオントロジーなどの語彙資源を参照するため，語彙資源の充実がテキストマイニングの発展に欠かせない。農業に関連した分野においてもテキストマイニングを適用した事例が現れつつあるが，農業に関連した分野の語彙資源は非常に少ないのが実状で，これを充実させていく必要がある。しかし，語彙資源の整備を行う際には，(1) 作業に多くの労力が必要である，(2) 何がその分野の用語かの判断が難しい，(3) 語彙資源を最新の状態に保つことが難しい，(4) 用語の表記が一定しない，といった問題がある。そこで，本研究では，これらの点について，実用上の観点からの解決に取り組んだ。

農業関連の語彙資源の有無がテキストマイニングの結果に与える影響を評価するため，農業用語辞書を用いたテキストの自動分類の実験を行った。農業用語辞書がある場合とない場合でそれぞれ自動分類を行い，結果を比較した。適合率の平均でみると，農業用語辞書がある場合の方がない場合よりも高かった。この実験により，農業関連の語彙資源の効果が定量的に示された。また，この実験に関連して，ネットワークを通して形態素解析の機能を提供するシステムの開発を行った。

農業関連の語彙資源の効果が示されたことから，農業関連の語彙資源を整備する方法を検討した。既存の固有表現抽出の技術を用いれば，コーパス中からこうした用語を自動で抽出できる。しかし，抽出された用語の中には，用語の断片や無意味な文字列も含まれるため，さらにそこから選別を行う必要がある。このような場面での選別の方法として，同じ分野の独立な2つのコーパスから抽出した用語の間で共通する語のみを取り出す方法を開発した。「畜産」，「果樹」，「野菜」の3分野を対象として2種類のコーパスを用意し，固有表現抽出でそれぞれから用語の抽出を行い，そこから両者に共通する語を取り出し，さらに一般用語を取り除いた。この選別によって得られた用語のリストでは，専門分野の書籍の索引に見られる語，すなわち専門家が集めた用語の含まれる割合が選別の前より上昇した。

経営支援システム等での利用を想定して，農業技術体系データの電子化を行った。農業技術体系データは，作業の種類ごとに作業内容，作業時期，使用する資材，必要な時間などの情報を記載したものである。農業技術体系データの電子化を進める際に明らかになったのは，資材の名称に例えば「マニュアスプレッタ」と「マニュアスプレッター」のような表記の多様性が多数存在することである。このような表記の多様性を放置すれば，別々の資材として扱われるという不都合が生じる。そこで，類似した表記を提示するアプリケーションを作成し，人間による表記の調整を支援することで問題の解決を図った。また，作物の栽培データにおける表記の多様性についても検討した。「燐酸」と「P205」のような表記の多様性がみられたが，オントロジーを利用して表記の多様性を吸収するシステムを構築したところ，支障なく運用できた。

以上のように，(1) の問題については，農業に関連した用語を整備する際に，抽出した用語を選別する手順を自動化する方法を開発した。また，表記の多様性解消を支援するアプリケーションを開発し，労力の軽減が可能であることを示した。(2) の問題については，用語そのものを判断の対象とするのではなく，分野に関連した文書を収集することで，判断基準の置き換えを図った。(3) の問題については，最新の文書さえ集めることができれば，そこから抽出と選別を行うことが可能であることを示した。また，ネットワークを通して形態素解析の機能を提供する事例から，最新の語彙を提供する方法を示した。(4) の問題については，アプリケーションによって表記の多様性の解消を支援する，あるいはオントロジーの形で多様な表記を蓄積する，という方法について実証した。

審 査 の 結 果 の 要 旨

インターネット上でテキスト情報が爆発的に増加する中，テキストを知識として効率的に活用するためのテキストマイニング技術の重要性は，ますます高まっている。本研究は，そのようなテキストマイニング技術の基盤として必須でありながら，あまり研究対象とし注目されてこなかった語彙資源の重要性を指摘した上で，それを整備するに当たって存在する問題点の解決方法を提案するものである。すなわち，語彙資源の整備を行う際に直面する，(1) 語の選別は人力に依存し作業に多くの労力が必要，(2) 抽出した語がどの分野の用語かの判断困難，(3) 語彙資源を最新の状態に保つことが困難，(4) 用語の表記が一定しない，といった問題に対して，農業関連文書を対象に，具体的解決方法を提示することに成功した。

このように，実用上極めて重要にもかかわらず，これまで十分に研究対象とならなかった語彙資源整備に着目し，具体的成果を示せたことは高く評価できる。また，あえて問題解決を一般化せず，農業関連用語という特定分野を対象に研究を進めたことも，現場で本当に使える技術開発という側面から正しい選択といえる。篤農家から新規参入者への経験的知識の伝達や自動化ロボットの制御といった新分野でもテキストマイニングの重要性が注目される中，本研究の成果は大いに生かせるものであり今後の発展に期待したい。

よって，著者は博士（農学）の学位を受けるに十分な資格を有するものと認める。