

## 審査の結果の要旨

氏 名 王 悦

本論文は、医療・生物学文献におけるタンパク質名および遺伝子名に対する認識器を開発するために作成された、タンパク質および遺伝子名をタグ付けした複数のコーパスに対して、コーパス間の不整合を検出し、その情報を利用して複数のコーパスを統合することで、より大規模かつ複数分野にまたがる整合性の高いコーパスを作成する手法を提案し、さらにそれを用いることによって単一のコーパス上での訓練した場合よりも高精度かつ頑健な認識器を作成できることを示したものである。

本論文は八章からなり、第一章では本論文における研究の動機付けとなる背景として、タンパク質および遺伝子名をタグ付けしたコーパスにおける問題点および、それを改善あるいは解決するための関連する従来手法について概観している。第二章は、本論文の貢献および概要に関してまとめたものである。第三章では、本論文で扱うタンパク質名および遺伝子名に関する複数のコーパスについてそれぞれの概略を紹介している。また第四章においては、本論文が用いているBANNERとよばれるタンパク質名ならびに遺伝子名の認識器について解説を行っている。第五章においては、コーパス間の不整合の性質について上記BANNERを用いて解析を行うとともに、タンパク質名・遺伝子名のタグの境界のゆれに関して、エントロピーの概念を用いて、各コーパスの比較を行っている。第六章においてはそれらの比較解析をさらに進め、複数あるコーパスのうち、最も整合性が高くタグ付け対象も広いGENEIAコーパスに特に着目し、GENEIAコーパスおよび他のコーパスとの比較解析を行うことで、各コーパス間の不整合の性質をより詳細に解析している。それらの結果に基づいて第七章では、検出されたコーパス間の不整合を排除した上で異なるコーパスを統合することにより、タンパク質および遺伝子名の認識器の精度および頑健性が向上することを示した。第八章では、本論文で提案した手法と有効性を総括し、さらに今後考察すべき課題についての展望が示されている。

なお、本論文の研究は、Jing-Dong Kim氏、Rune Saetre氏、Sampo Pyysalo氏、大田朋子氏、ならびに辻井潤一氏との共同研究であるが、論文提出者が主体となって立案、分析、検証を行ったもので、論文提出者の寄与が十分であると判断する。

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。