

要旨

電子的に保存された大量のテキストデータを十分に活用するにはコンピュータによる自動処理が欠かせない。高度な処理には医学用語の意味情報が必要不可欠であるが、存在する全ての文字列表現の意味を予め記述するのは非常に労力を要するものである。

本研究では(1)医学用語（特に疾患名）の内部構造の表現方法と(2)内部構造情報を用いた自動 ICD コーディング方法を提案した。(2)について、実験の結果、既存手法では解けなかった問題の一部が解けた。

更に、疾患名に対して内部構造表現を自動付与する解析器を作成した。評価実験の結果、83.7%の精度で語を解析可能であり、提案した内部構造表現に実用性があることが示された。