

表記ゆれ及び階層関係が取り扱い可能な身体部位名称辞書モデルと検索システム

工学部 情報工学科 西原 典孝, 島田 英之

Keywords: 部位名称検索, 部位名称辞書, 電子カルテ

1. 開発目的

電子カルテや各種の医学資料の検索やデータ分析などを行なおうとする際、一番重要な語句は、臓器名などの部位名称である。しかし部位名称は複合名詞表現を取るため、様々な表記ゆれが実際の文章中には現れる(胃体小弯, 胃体部小弯, 胃体の小弯部など)。本研究では様々な表記ゆれに対応でき、かつ部位名称間の上位・下位関係も記述できるような辞書構成法とそれを用いた検索法を提案する。

2. 部位名称辞書の構成法

本辞書モデルは以下の3つの組み合わせで構成される。

- (1) 記述辞書: 人間が作る辞書ベースであり、部位名称を構成要素の組み合わせで記述。
- (2) 展開辞書: 記述辞書から各項目を自動的に展開し生成される辞書。
- (3) 検索法: 展開辞書を使用した部位名称の検出法。

表1が記述辞書の一例である。部位名称を構成要素とその組み合わせ方として表記する。記述辞書から自動生成された展開辞書の一例が表2である。展開辞書では、構成要素の全組み合わせによる部位名称の生成とその接続関係、および上位概念の生成を行っている。

検索システムの基幹部である検索エンジン自体の機能は、文章中からの部位名称の抽出にある。展開辞書から正規表現を生成し、部位名称をパターンとして探し出す。

表1 記述辞書 (一部抜粋)

基本部位名称	上位概念	左右識別表現	局所部位指定
肺	胸部	右, 左	-上葉, -中葉, -下葉
肺野	肺	*	*:上葉-, 中葉-, 下葉-
肺尖	肺上葉	*	0
咽頭	頭頸部, 気道	0	上+ (-鼻部), 中+ (-口部), 下+ (-喉頭部)

表2 自動生成された展開辞書 (一部抜粋)

部位名称	部位本体	上位概念	左右	局所部位指定
右肺	肺	胸部	右	*
右肺上葉	肺	右肺, 肺上葉	右	-上葉
右肺中葉	肺	右肺, 肺中葉	右	-中葉
右肺下葉	肺	右肺, 肺下葉	右	-下葉
右肺野	肺野	肺野, 右肺	右	*
右上葉肺野	肺野	右肺野, 上葉肺野, 右肺上葉	右	上葉-

「実際の癌診断カルテからの検出例」(文面上の語句と検出部位名称)

前立腺の右葉 (前立腺右葉), 右肺門部中葉 (右中葉肺門), 鎖骨上窩のリンパ節 (鎖骨上リンパ節), 右の甲状腺 (甲状腺右葉), 胃体部小弯 (胃体小弯), 両側肺門・縦隔リンパ節 (右肺門リンパ節, 左肺門リンパ節, 縦隔リンパ節)

3. 応用の可能性

本辞書モデルでは、人間が作成する記述辞書とそこから自動生成される展開辞書を区分し、少ないコストで多様な複合表現や複雑な階層構造に対応した部位名称検索用データを得ることができる。その展開辞書と連動した検索エンジンを与え、実際のカルテデータで部位名称検出に有効なことを確認した。近年活発に研究されてきている医療情報分野の電子カルテや投影レポートを対象にしたデータマイニングの研究などに対する有用なツールになるものと考えられる。