理工学専攻修士論文要旨

提 出 年 度： 2005 年度
提　 出　 日： 2006 年 1 月 30 日
専 修 コ ー ス：知能情報コース
学 生 番 号： 35604182
学 生 氏 名： 安田智成
研究指導教員：原田実教授

（論文題目）
未知語に対する語意説明のインターネットからの獲得と電子辞書への自動登録

（内容の要旨）

原田研究室では、EDR 電子化辞書の情報を元に日本語文章の意味解析を行うシステム、SAGE を開発してきた。しかし、既存の EDR 電子化辞書は日々発生している新語には対応しておらず，SAGE では未知語と判定してしまい解析精度を下げる原因となっている．そこで、新語を EDR 電子化辞書に登録することで，SAGE の解析精度向上，また意味解析の応用システムの精度向上を提案する．

本研究では，インターネットから新語と推察される語に関する，見出し，語意説明，出典の３つの情報を新語データソースとして収集する．これらの語意説明を SAGE で解析し意味グラフ化したものを基に新語の基本情報や概念としての挿入位置を決定し、EDR 電子化辞書に新語として登録する「新語登録システム」を開発することにした．

新語登録処理の具体的な手順は，(１)新語の語意説明の意味グラフ化，(２)基準概念の判定，(３)概念体系上の挿入位置決定，の 3 ステップである．

(１)新語の語意説明の意味グラフ化では，新語データソースに登録されている新語が EDR 電子化辞書に登録済みでなければ，その語意説明を SAGE で解析し，意味グラフへと展開する．

(２)基準概念の判定では，新語の語意説明の中からその新語の特徴を最も端的に捉えている語を基準概念とする．基準概念とは新語を概念登録する場合の上位概念である．その決定方法には定義パターンと呼ぶ特定の日本語表現を用いた．意味グラフ中にこの定義パターンが存在すれば，その定義パターンノードから「modifier」格で隣接するノードを基準概念として抽出する．定義パターンが存在しなければ，文末のノードを基準概念として抽出する．また，基準概念として抽出された語句と並列の関係を表す「and」格もしくは「or」格で隣接する語句があった場合，その語句も基準概念として抽出することとする．

(３)概念体系上の挿入位置決定は，次のステップで行う．(a)基準概念が概念体系上で下位概念を持たない場合は，新語を基準概念の下位概念として配置する．(b)基準概念が概念体系上で下位概念を持っている場合は，まず新語と基準概念の下位概念の見出し同士での表層の包含関係を利用する．新語見出し⊃子概念見出しであった場合には新語が子概念より特殊化された存在であるとして当該子概念の下位概念に決定する．一方，新語見出し⊂子概念見出しであった場合は，新語は子概念より一般化された存在であるとして，基準概念と当該子概念の間の概念として決定する．もう一つの方法としては，新語と子概念の語意説明の意味グラフ間の双方向でのグラフ類似度(竹原 Metis 2005)の値を比較することで概念挿入位置を決定する．

様々なインターネット上の専門サイトから抽出した新語データソース 194 セットに対して評価を行ったところ，正解と判定できるものが 171(88.4%)，不正解が 15(7.7%)，その他(取得した語意説明文が適切でなかった)が８(4.1%)となり，これらの新語の登録は SAGE 解析精度の向上に貢献したので本手法の有効性を示すことができたと言える．