

理工学専攻修士論文要旨

提出年度：2005年
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：35604161
学生氏名：前澤 敏之
研究指導教員：原田 実 教授

(論文題目)

意味グラフに対する言い換えエンジンの開発研究

(内容の要旨)

原田研究室で開発している SAGE は意味解析によって文章の意味構造を表現できる「意味グラフ」を生成する。本研究室で開発している質問応答やアンケート分類など意味グラフを入出力とするシステムの精度向上のためには意味グラフの言い換えが有効である。一般に、ある言語表現があったときに、その表現と近似的に等価な意味を持つ言語表現を「言い換え」という。例えば「飛行機を発明したのは誰ですか」と「誰が飛行機を発明しましたか」はお互いに言い換えの関係にある(言い換え表現対)。そこで、本研究では意味グラフに対する言い換えエンジン Serpent の研究開発を行った。意味グラフに対する言い換えとは言い換えの対象となる文(原文)の意味グラフに関してある適合条件を満足する部分グラフ(Hole)を同じ意味を表す、より単純な意味グラフの集合で置き換えた文(換え文)を生成することである。

Serpent はルールベースシステムであり、用いる言い換への規則(言い換え知識)は Key、Head、Body の 3 つの部分からなる。Key は言い換への検索フェーズを高速化するのに用いられる。Head には原文中のある部分グラフが言い換え知識に対応する Hole となるための適合条件が記述される。Body には Hole の集合を置き換える意味グラフの集合に関する情報が記述される。

言い換えエンジン Serpent による言い換え処理は「言い換え知識の検索」、「競合解決」、「言い換え知識の照合」、「言い換への実行」の 4 つのフェーズからなる。「言い換え知識の検索」では言い換え知識に与えられた Key を用いたハッシュをトライと併せて利用することで原文に対して適用できる言い換え知識を高速に検索する。「競合解決」では原文に対して検索フェーズを通過した複数の言い換え知識から言い換え後のノード数が最小となるものを選び出す。「言い換え知識の照合」では競合解決によって選択された言い換え知識の Head を満足する原文の部分グラフを Hole として選び出す。「言い換への実行」では照合フェーズで特定された Hole を Body で表される意味グラフで置き換える。

奈良先端大の言い換え事例 464 文を用いて Serpent の評価を行った。464 文中言い換への発生した 46 文については言い換え前の平均類似度が 67.63%で言い換え後が 75.91%であった。この結果言い換えによって言い換え表現対の類似度が向上したことがわかった。また 464 文に対する言い換への実行時間は 266 ミリ秒であり実用的な実行時間で言い換えられることがわかった。