

第5週B4ゼミ

特許分類コード体系に基づくオントロジーの構築
－ 情報分野におけるケーススタディー －

出典

- 難波 英嗣、乾 孝司、岩山 真、櫻井 孝、稲田 浩一、藤井 敦
- 特許分類コード体系に基づくオントロジーの構築 – 情報分野におけるケーススタディー –
- 言語処理学会 第20回年次大会 発表論文集、2014
- pp. 400-403

はじめに

- 人手でオントロジーを構築し、更新することは非常にコストがかかる
- 自動でオントロジーを構築する手法が提案されているが、人手にとって代わるレベルではない
- 特許分類コード体系のひとつであるFタームに着目し、学術論文など他のジャンルの文献にも利用可能なオントロジーの構築を目指す

Fタームとは

- 特許を目的・効果・構成などの様々な観点から分類することを目的とした分類体系
 - 技術分野を示すテーマコードと観点の集合から構成される

BA11	• 翻訳方式
BA12	• • 直接翻訳
BA13	• • 間接翻訳
BA14	• • • トランスファー方式
BA15	• • • • 意味解析
BA16	• • • • • 文脈解析
BA17	• • • ピボット方式

Fタームに基づくオントロジーの構築

- 手順1) Fタームからの知識抽出
- 「同義」「上位・下位」「属性・定義域・値域」「全体・部分」
 - 一部自動、残りを人手で判断
- (手順2) 手順1 で得られた知識への他の知識体系のマッピング

関係 1	属性：方式 定義域：機械翻訳 値域：直接翻訳， 間接翻訳
関係 2	上位：間接翻訳 下位：トランスファー方式
関係 3	上位：間接翻訳 下位：ピボット方式

上位、下位概念辞書の自動構築

- 上位・下位概念辞書の構築

- Hearst[Hearst 1992]の定型表現法
- AなどのB

- Fタームの例

言語処理技術

- 文解析
- • 形態素解析

「形態素解析」の上位語

- 40 処理 > 形態素解析
- 23 自然言語処理 > 形態素解析
- 20 構文解析 > 形態素解析
- 20 言語処理 > 形態素解析
- 17 技術 > 形態素解析
- 14 構文解析処理 > 形態素解析

上位、下位概念辞書の自動構築

- BA01 とBA02 の間に上位・下位関係が成立していることがわかった
- 同じドットレベルのBA03についてもBA01との間に上位、下位関係がある

BA01 • X

BA02 • • Y ※ BA01 の下位概念

BA03 • • Z

BA04 • • • W

属性辞書の自動構築

- 属性関係にある語は「AのB」と表記される
- [名詞句]の[名詞句]にマッチする箇所のすべて抽出
- 全体・部分関係にある語も上記を満たす
 - パソコンのキーボード など
- 属性になりうる用語(目的、対象)
 - 種類が多いとは考えにくい
- 高頻度で出現するものは属性である可能性が高い
 - 目的・用途, 利用分野・応用・適用分野, 手段, 効果¹, 方式, 対象, 時期, 入力・出力, 構成・構造・機構・形状・利用技術・配置, 機能

属性辞書の自動構築

- Fタームコードが00で終わるものを参照
 - テーマ中で最上位のもの
- 説明語句の中に選定した用語が含まれている
 - テーマコードに対して属性関係にあると推測

Fタームから抽出された知識へのマッピング

属性語の1つ「効果」に着目

- 福田らの手法[福田 2013]によりTECHNOLOGY、EFFECT ATTRIBUTE、VALUEをタグ付け

PM 磁束制御用コイルを設けて
<TECHNOLOGY>閉ループフィードバック制御
</TECHNOLOGY> を施すため、
<EFFECT><ATTRIBUTE> 電力損失
</ATTRIBUTE> を <VALUE> 最小化
</VALUE></EFFECT>できる。

- 属性(ATTRIBUTE)と属性値(VALUE)の対を「効果表現リスト」

まとめ

- オントロジーを構築する手法を提案し、現在構築中
- 所感
- 目指したい適合率、再現率は？
 - いい結果は得られそうな気がする
- オントロジーに似た構成のものはまだまだ存在するのでは？