

全文検索システム構築の検索要件定義におけるオントロジーの適用と考察

日本ユニシス株式会社

國府田 一輝

kazuki.kokufuda@unisys.co.jp

検索要件定義における問題点

全文検索システム構築の検索要件定義のフェーズにおいて、顧客が検索対象としたい文字列の情報を入手する必要がある。しかし、開発者は業務の専門家ではなく、知識がないために開発者と顧客との間の知識と語彙に差異が生じ、顧客が要求する検索対象文字列を開発者が漏れ無く入手することは難しい。

手法・ツールの適用による解決

検索対象領域に関するオントロジーを用いて、検索対象文字列を決定する方法を研究した。オントロジーを用いることで、ヒアリング作業で獲得漏れをした文字列をシステムにより提供する。また、検索対象範囲に関する知識を体系化する。

オントロジーの構築

① センテンスから動詞・名詞を抽出

② 動詞を整理し抽象化

③ 格フレーム辞書の作成

④ オントロジーの構築

① テキストデータから得たセンテンス(単文)から、名詞・動詞を抽出

「ある大学生が江東区に住所を移した。」

形態素解析

名詞: "大学生", "江東区", "住所"
動詞: "移し"

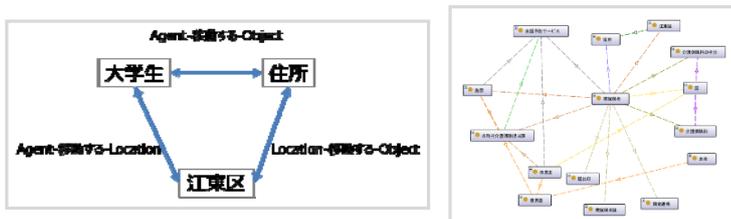
② 類似した概念を表す動詞をまとめ、抽象化

| 動詞(抽象化前) | 動詞(抽象化後) |
|----------------|----------|
| 入所する, 転入する, 移す | 移動する |

③ 文の意味構造を、動詞を中心とした格構造によって表現

| <Agent>が | <Goal>に | <Object>を | 移動する |
|----------|-------------|-----------|------|
| 大学生が | 江東区に | 住所を | 移す |
| 被保険者が | 市町村介護保険担当課に | 介護予防サービスを | 申しでる |
| 介護事業者が | 市町村介護保険担当課に | 介護予防サービスを | 申しでる |
| 介護保険施設が | 市町村介護保険担当課に | 介護予防サービスを | 申しでる |
| 被保険者が | 関に | 介護保険料を | 渡す |
| 介護事業者が | 関に | 介護保険料の半額を | 渡す |

④ 格フレーム辞書からオントロジーを構築



適用実験

- 介護業界の文書を検索対象範囲とし、評価用オントロジーを構築
- 問い合わせ: 「介護予防サービスを申しでる」
・被保険者が(Agent)市町村介護保険担当課に(Location)
・介護事業者が(Agent)市町村介護保険担当課に(Location)
⇒ 動作主格: "被保険者" "介護事業者"
場所格: "市町村介護保険担当課"
- 問い合わせ: 「生活保護を申しでる」
⇒ 未知語: "生活保護"

まとめ

- 検索対象文字列の候補の提示や語彙の不足の表現が行えるところに、オントロジーの有用性が認められた。
- オントロジーの構築にコストや知識を必要とする点や、概念に設定するプロパティ(被保険者の概念に対し、年齢や要介護度を付与するなど)の充実が課題