

用語の属性を考慮した上位，下位概念辞書の構築

Construction of a Hyponymy Dictionary Taking Account of Attribute Words

福田悟志 難波英嗣 竹澤寿幸
Satoshi Fukuda Hidetsugu Nanba Toshiyuki Takezawa

広島市立大学大学院情報科学研究科
Graduate School of Information Sciences, Hiroshima City University

1. まえがき

本研究では，特許公開公報から構築されたシソーラスにおける上位，下位概念に対して，上位語と下位語に共通する属性(ファセット)を発見する手法を提案する。テキストデータベースからシソーラスを構築する代表的な手法として，「A などの B」などの定型表現に着目して，用語の上位，下位概念を自動的に抽出するものが挙げられる。しかし，どのような側面でこれらの概念が上位，下位関係であるのかを裏付けることは難しい。そこで本研究では，定型表現を用いて抽出した概念が要素技術である場合，それに付随する属性に着目する。例えば，「テンキーなどのキーボード」という文における「キーボード」の場合，「操作性」や「使い勝手」などが属性である。また，「テンキー」の場合，「操作性」や「操作ミス」などが属性であり，このことから，「キーボード」と「テンキー」は，「操作性」という共通の属性を持つ上位，下位関係であると示すことができる。また，本論文では，本研究の手法を用いることで，誤った上位，下位関係を効率的に除去することができることを示す。一般的な上位，下位概念の定義から，下位概念に位置する要素技術は，上位概念のそれを特定の領域に適用させるために発展，改良したものと考えられる。そのため，下位概念の属性の種類は，上位概念より限定されると考えられる。このことから，付随する属性の種類を見た時，下位概念の方が多ければ誤った上位，下位関係であるとみなすことができる。また，要素技術間に共通する属性がない場合，それらを裏付ける関係性がないとみなすことができるため，誤った関係と判断できる。

2. 特許公開公報からの要素技術と属性の抽出方法，および誤った上位，下位概念の検出手段

2.1. 構造解析におけるタグの定義

本研究では，福田ら[福田 2013]の手法を用いて，機械学習による特許公開公報の構造解析を行う。以下に，本研究で使用する構造タグを示す。

- **TECHNOLOGY** : 要素技術を示す。
- **EFFECT** : 効果(新しい機能の追加，新しく得られた物質，問題点の抑制や解決したこと)を示す。
- **ATTRIBUTE, VALUE** : 例えば，「処理速度(ATTRIBUTE)が向上(VALUE)」のように，「属性」と「属性値」の対で表現する。

2.2. 要素技術－属性リストの構築

まず，特許公開公報に含まれる 3 つの項目【発明が解決しようとする課題】【課題を解決するための手段】【発明の効果】に対して，2.1 節で定義したタグを付与する。次に，【発明が解決しようとする課題】および【課題を解決するための手段】の項目から，TECHNOLOGY タグが付与された箇所を抽出し，【発明の効果】の項目からATTRIBUTE タグが付与された箇所を抽出する。その後，抽出された要素技術と属性をそれぞれ対応付ける。

2.3. 誤った上位，下位概念の検出手段

本研究では，以下の 2 種類による判断基準から，誤った上位，下位概念を検出する。

- (1) 2 つの概念間に共通する属性があるかどうか調べる。
- (2) 特定の要素技術における属性の種類を調べ，上位，下位概念間における属性の異なり数を比較する。

3. 実験

評価データは以下の手順で作成した。

- (1) 特許公開公報から「A などの B」などの定型表現により構築した上位，下位概念リストから，「キーボード」が上位概念にある下位概念の用語を抽出する。
- (2) その中から要素技術を表す用語を抽出する。

上記により得られた 310 語から，上位概念「キーボード」と正しい上位，下位関係である下位概念を人手で判断した。その結果，9 語の正しい下位概念を発見した。また，以下に示す比較手法により，本手法の有効性を確認する。

- **提案手法**: 2.3 節で述べた検出手段を用いる。
- **ベースライン**: 手順(1)で獲得した下位概念を全て正しいものと判断。

上記で述べた比較手法による実験結果を表 1 に示す。評価尺度には，精度，再現率を使用した。表 1 より，本手法を用いることで，全体の 1/3 にあたる 109 語を誤った下位概念として検出することができた。最後に，本手法により自動的に構築した上位，下位概念辞書の例を表 2 に示す。例えば，「キーボード」と「テンキー」には，「操作性」と「利便性」という属性が共通している事が分かる。

表 1 誤った上位，下位関係の検出による実験結果

手法	再現率	精度
提案手法	1.000	0.044 (9/204)
ベースライン	1.000	0.029 (9/310)

表 2 本手法により構築した上位，下位概念辞書の例

上位概念	下位概念	属性
キーボード	フラットキーボード	操作性，操作感
キーボード	タッチキーボード	キー操作性
キーボード	テンキー	操作性，利便性

4. おわりに

本研究では，上位，下位概念に共通する属性に着目し，誤った上位，下位関係を検出する手法を提案した。また，共通する属性を発見することで，特許シソーラスにおける新たな知識獲得を行うことができることを示した。

参考文献

[福田 2013] 福田悟志，難波英嗣，竹澤寿幸 (2013) "論文と特許からの技術動向情報の抽出と可視化". 情報処理学会論文誌データベース，Vol.6, No.2, pp.16-29.

外部発表情報

福田悟志，難波英嗣，竹澤寿幸，橋田浩一 (2014) "用語の属性を考慮した上位，下位概念辞書の構築". 言語処理学会第 20 回年次大会，pp.967-970.