

画像特徴量の性質継承に基づく Web からの概念階層抽出

服部 峻†

† 東京工科大学 コンピュータサイエンス学部

〒 192-0982 東京都八王子市片倉町 1404-1

E-mail: †hattori@cs.teu.ac.jp

あらまし 上位下位 (is-a) 関係や部分全体 (has-a) 関係といった概念階層は、様々な自然言語処理システムにとって非常に重要な基本的知識である。人海戦術的に人手で概念階層 (辞書) の構築が進められている一方で、Web などの大規模な文書データベースから自動的に知識抽出する研究も数多く行われている。後者のアプローチの一つとして、各概念の性質として has-a 関係や振る舞いを表す言語的記述を Web から抽出した上で、対象概念から下位概念候補への性質継承の度合いに基づく抽出手法を提案して来た。本稿では、既存手法のように各概念の意味的・言語的な has-a 関係の性質継承の度合いを計算するだけでなく、各概念を表す (典型) 画像が持つ特徴量の性質継承の度合いを計算することによって、Web テキストだけでなく Web 画像も活用して概念階層を自動構築する手法を提案する。

キーワード 上位下位関係, 部分全体関係, 概念階層, 性質継承, Web マイニング, テキスト解析, 画像解析.

Concept Hierarchy Extraction from the Web based on Property Inheritance of Image Features

Shun HATTORI†

† School of Computer Science, Tokyo University of Technology

1404-1 Katakura, Hachioji, Tokyo 192-0982, Japan

E-mail: †hattori@cs.teu.ac.jp

1. はじめに

概念間の上位下位 (is-a) 関係や部分全体 (has-a) 関係といった意味的な階層関係は、情報検索における検索質問の拡張・修正 [1]~[3], 五感情報に基づくオブジェクト名サーチ [4], [5], 質問応答 [6] や機械翻訳, セマンティック Web などにおける知識の共有・再利用, テキストマイニングによるオブジェクトの情報抽出 [7]~[9] など、様々な自然言語処理システムにとっても非常に重要な基本的知識である。オブジェクトの外観情報抽出の場合、例えば「カワセミ」の外観情報は、「青い鳥」「美しい鳥」などと「カワセミ」の上位語である「鳥」と、或いは、「青い羽」「長い嘴」などと「カワセミ」の構成要素である「羽」や「嘴」と外観修飾句のペアで記述されることが多く、対象のオブジェクト名の上位語や部分語の知識を利用できれば、より効率的に外観情報を抽出することが可能になる。

WordNet [10] や Wikipedia [11] において、概念階層の構築が人海戦術的に進められている一方で、多種多様なユーザにより文書が記述され蓄積されて行く Web や Blog などの大量の文書コーパスをテキストマイニングすることで自動的に構築・拡

張する研究も数多く行われている。文書コーパスから上位下位関係を抽出する従来手法の多くは、「 x such as y 」などの構文パターンに合致する記述が文書コーパス中に大量に含まれるならば、概念 y は概念 x の下位概念である」という仮説に基づいているが、この仮説が真であるとしても、「構文パターンに合致する記述が文書コーパス中に大量に含まれる」ことは、「概念 y が概念 x の下位概念である」ことの十分条件でしかなく必要条件ではない。従って、高い適合率を実現するために厳密な構文パターンだけを用いると、その構文パターンに合致する記述が文書コーパス中で少なくなってしまうため、本当は上位下位関係を持つ概念間に対して上位下位関係が無いと誤判定する危険性が増し、再現率を損なってしまう。逆に、高い再現率を実現するために曖昧な構文パターンを用いると、そもそもの仮説を満たさなくなり、適合率が著しく悪化してしまう。

オブジェクト名サーチなどの基本的知識として利用するためには、対象語 (クラス名) に属する下位語を出来る限り網羅的に抽出する必要がある。従って、厳密な構文パターンに基づく従来手法では不十分であり、再現率を改善しつつ適合率も高く維持するためには、上位下位関係の緩い構文パターンを用いて

候補語を網羅的に（但し多くのノイズも一緒に）収集した上で、構文パターンに合致する文書頻度とは異なる評価尺度、つまり、上位下位関係の十分条件ではなく、必要（十分）条件を用いてフィルタリングする必要がある。そこで、「概念 y は概念 x の下位概念である」という上位下位関係の必要条件として、オブジェクト指向方法論におけるクラス間でのデータメンバ（属性）とメソッド（振る舞い）の継承関係 [12] や、属性分析法における事象階層構造での属性遺伝 [13] に基づき、「概念 y が概念 x の性質を全て継承する」という「性質継承」を仮定する。

そこで、これまでの研究において、上位下位関係の必要十分条件として概念間の「性質継承」を仮定し、対象概念（クラス名）が与えられた場合に、その対象概念と下位概念候補各々の典型的な性質を求めた上で、対象概念から下位概念候補への性質継承の度合いを評価することによって、対象概念に属する下位概念（具体的なオブジェクト名）を Web から抽出する手法 [14] を提案して来た。既存手法では、Web テキストだけを用いて、各概念の性質として has-a 関係や振る舞いを表す言語的記述を Web から抽出した上で、対象概念から下位概念候補への性質継承の度合いを計算している。さらに、二つの概念間の直接的な関係を評価するだけでなく、これらの周辺に在る概念との関係も考慮したり、各概念の典型的な性質を抽出する手法においても、各概念と各性質との間の直接的な関係を評価するだけでなく、対象概念の上位概念からの性質継承や対象概念の下位概念集合からの「性質集約」も考慮したりすることで改善を図っている。

本稿では、既存手法のように各概念の意味的・言語的な has-a 関係の性質継承の度合いを計算するだけでなく、各概念を表す（典型）画像が持つ特徴量の性質継承の度合いも計算し、これら二つを統合して上位下位関係の評価とすることによってロバスト性の向上を図り、Web テキストだけでなく Web 画像も活用して概念階層を自動構築する手法を提案する。

2. 関連研究

本章では、概念間の上位下位関係や部分全体関係を、Web や新聞記事などの大量の文書コーパスからテキストマイニングによって自動的に抽出する関連研究について紹介する。

2.1 上位下位関係

新聞記事や Web などの大量の文書コーパスをテキストマイニングすることで、概念間の上位下位関係を自動抽出する手法がこれまでに数多く提案されている。Hearst [15] は、“ x such as y ” や “such x as y ” といった上位下位関係の構文パターンを用意しておき、文書コーパスから構文パターンに合致する記述を収集することで、概念間の上位下位関係を獲得する手法を提案し、新しい構文パターンを発見する手法についても述べている。従来研究の多くはこの流れを汲み、概念間の上位下位関係を抽出するための様々な構文パターンが提案されている [16]~[18]。しかしながら、前章でも述べたように、“ x such as y ” などの構文パターンに合致する記述が文書コーパス中に大量に含まれることは、「概念 y が概念 x の下位概念である」ことの必要条件ではないため、構文パターンを網羅的に用意したとして

も、構文パターンに合致する記述が文書コーパス中に運良く十分大量に含まれない限り、上位下位関係を持つ概念間に対して上位下位関係が無いと誤判定してしまう危険性が根本的に残る。

上位下位関係の構文パターンを利用するだけでなく、国語辞典や百科事典における見出し語とその説明文の構造をモデル化し、構文解析などによって上位下位関係を獲得する手法も提案されている。鶴丸 [19] らは、国語辞典を利用し、見出し語とその語義文に現れる定義語との間に階層関係を付けることで、シソーラスを自動構築している。桜井 [20] らは、Web から用語説明を自動生成した上で上位語を決定している。大石 [21] らは、Web を事典的に利用するために構築された Cyclone コーパスを用いて、見出し語とその説明文の方向性を考慮した確率的な出現頻度モデルと局所的な構文情報に基づく統計モデルによって、単語の階層関係を統計的に自動識別している。一方、森本 [22] らは、専門用語の構成規則に基づいて、複合用語を基本構成用語（語基）に分解し、用語の各語基の包含関係を比較することで、専門用語間の階層関係を解析している。

構文パターンに依存しない抽出手法により、概念間の上位下位関係抽出の再現率の改善を図っている研究も多数あり、本論文と非常に関連がある。小淵 [23] らは各語に対して意味素の集合を割り当て、Sanderson [24] らや KnowItAll [25] では文書コーパス中での各語の出現の仕方に基づいて素性を割り当て、語概念間で包含関係が認められる場合に上位下位関係があると判定する。山本 [26] らは、あらゆる形容詞や形容動詞を索引として予め定めた共起ベクトルを各抽象名詞を修飾する頻度に基づいて求め、抽象名詞間の共起ベクトルの包含関係をオーバラップ相関係数や補完類似度で評価して、抽象名詞の階層構造を自動構築している。一方、性質継承に基づく抽出手法では、対象概念の典型的な性質の上位 n 件のみを索引として採用し、典型的でない多数の性質については性質継承の度合いを評価する際に関知しない点異なる。また、山本らは一回でも共起することを重視して各索引に重み付けしているが、Web においては複数回共起しなければ重視すべきでないと考える。新里 [27] らは、箇条書きや表などの HTML タグの繰り返しパターンにより下位語候補を抽出し、DF や IDF などの統計量や表題に基づいて上位語の候補を絞り、各名詞が持つ動詞との係り受け関係を特徴ベクトル化し類似度を計算している。これらの研究と同様に、提案している性質継承に基づく抽出手法も、本質的にはベクトル間の類似度計算に準じているが、索引として用いる性質語を部分全体（has-a）関係や振る舞い表現に限定し自動抽出している点、注目している概念間の関係だけでなく、上位概念や同位概念、下位概念といった周辺概念も考慮して性質継承の度合いを評価する点異なる。

2.2 部分全体関係

鶴丸 [28] らは、国語辞典に基づくシソーラスの構築に関して、同義関係によるグループ化を利用した極大語の処理、及び、上位下位関係や同義関係との融合による部分全体関係の拡張可能性について論理的に考察している。Sundblad [29] らは、自然言語の質問文コーパスに対して単純なパターンマッチングを行うことで、上位下位関係だけでなく部分全体関係を収集している。

3. 性質継承に基づく下位概念抽出

本章ではまず、概念間の上位下位 (is-a) 関係の十分条件として構文パターンや文書構造を仮定する手法とは異なり、必要十分条件として概念間の「性質継承」を仮定し、対象概念とその下位概念候補との間の性質継承の度合いを評価することによって、上位下位関係を Web から抽出する手法について概説する。

次に、各概念の性質として、Web から抽出した部分全体 (has-a) 関係や振る舞いを表す言語的記述だけを用い、対象概念から下位概念候補への性質継承の度合いを評価して上位下位関係を Web から抽出する既存手法を定義する。

最後に、各概念を表す (典型) 画像が持つ特徴量もその性質として用い、各概念の言語的な性質継承の度合いと画像的な性質継承の度合いとを統合することによってロバスト性の向上を図り、Web テキストだけでなく Web 画像も活用して上位下位関係を Web から抽出する手法を提案する。

3.1 概要

「概念 y は概念 x の下位概念である」という上位下位 (is-a) 関係の必要 (十分) 条件として、「概念 y が有する性質の集合 $P(y)$ は、概念 x が有する性質の集合 $P(x)$ の全てを包含する (概念 y が概念 x の性質を全て継承する)」という概念間の「性質継承」を仮定する。

$$\text{isa}(y, x) = 1 \Leftrightarrow P(y) \supseteq P(x) \text{ and } y \neq x,$$

$$P(c) = \{p \in P \mid \text{has}(p, c) = 1\}.$$

但し、 P は性質の全体集合を表し、 $\text{has}(p, c)$ は概念 c が性質 p を有するか否かの二値 $\{0, 1\}$ を表す。

$$\text{has}(p, c) = \begin{cases} 1 & \text{if 概念 } c \text{ が性質 } p \text{ を有する,} \\ 0 & \text{otherwise.} \end{cases}$$

言い換えると、「概念 y は概念 x の下位概念である」ならば「概念 x と概念 y が共有する性質の数が概念 x の有する性質の数と等しい (かつ、概念 y の有する性質の数よりも小さい)」という次の関係が成り立つ。

$$\text{isa}(y, x) = \begin{cases} 1 & \text{if } \sum_{p \in P} \text{has}(p, y) \cdot \text{has}(p, x) = \sum_{p \in P} \text{has}(p, x), \\ 0 & \text{if } \sum_{p \in P} \text{has}(p, y) \cdot \text{has}(p, x) < \sum_{p \in P} \text{has}(p, x). \end{cases}$$

以上により、 $\sum \text{has}(p, y) \cdot \text{has}(p, x) / \sum \text{has}(p, x)$ の値が 1 と等しいか否かに基づいて上位下位関係の有無を判定できるが、概念 c と性質 p の任意のペアに対して二値 $\{0, 1\}$ で $\text{has}(p, c)$ を正確に求められることが必要不可欠である。これは容易ではなく、各概念に対して Web から典型的な性質を抽出する手法を用いると基本的には連続値 $[0, 1]$ しか利用できない。そこで概念 y が概念 x の下位概念であるか否かを表す二値の近似として、その相応しさを表す連続値を、概念 x から概念 y へ性質が継承されている度合い $\sum \text{has}(p, y) \cdot \text{has}(p, x) / \sum \text{has}(p, x) \cdot \text{has}(p, x)$ によって評価する。

x : target concept
 y : hyponym candidate
 $P(c)$: property vector

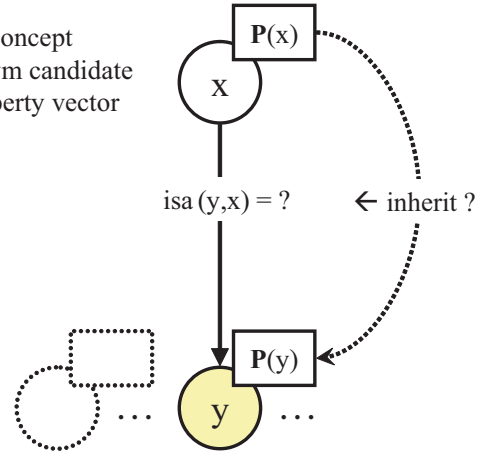


図 1 性質継承に基づく下位概念抽出

3.2 手法

対象概念 x が与えられた場合に、その下位概念候補集合 $C(x)$ を Web から網羅的に収集した上で、対象概念 x の典型的な性質を求め、その上位 n 件のみを下位概念候補 y が継承している度合いに基づいてランキングすることで、対象概念 x の下位概念を Web から抽出する手法 (図 1) について述べる。

Step 1. 下位概念候補の収集 :

対象概念 x の下位概念を出来る限り洩れなく含む下位概念候補集合 $C(x)$ を収集する必要があるが、あらゆる語概念を候補としてしまうと、候補数およびノイズ割合が大きくなり過ぎるため不適切である。本稿では、「 y は x である」という上位下位関係を表すような構文パターンを用いて構成した検索クエリを Yahoo!ウェブ検索 API に与え、最大 1000 件の検索結果 (タイトルおよびスニペット) から形態素解析で y の部分に当たる名詞句を切り出し、その全てを下位概念候補とする。

Step 2. 各概念の典型的な性質の抽出 :

各概念 c の有する典型的な性質 p を Web 上の情報資源を活用して抽出する。既存手法では、オブジェクト指向に則り属性名や振る舞い (メソッド) 表現を想定し、Web テキストだけを用いて、「 c の p 」という構文パターンを用いて構成した検索クエリを Yahoo!ウェブ検索 API に与え、最大 1000 件の検索結果から形態素解析で p の部分に当たる名詞句を切り出している。その上で、概念 c の有する典型的な性質としての言語的記述 p の相応しさの度合いを以下の式によって計算している。

$$\text{has-txt}(p, c) := \frac{\text{sf}(["c \text{ の } p"])}{\text{sf}(["c \text{ の } "])} \in [0, 1]$$

但し、 $\text{sf}(["c \text{ の } "])$ は最大 1000 件の検索結果の数、 $\text{sf}(["c \text{ の } "])$ はその検索結果におけるフレーズの出現頻度を表している。

本稿では、Web テキストだけでなく Web 画像も活用し、各概念の性質として、部分全体 (has-a) 関係や振る舞いを表す言語的記述だけでなく、各概念を表す (典型的な) 画像が持つ特徴量も用いる。まず、Yahoo!画像検索 API にフレーズ検索クエリ $["c"]$ を与え、最大 100 件の検索結果を得る。Web 画

像検索は Web 文書検索に比べて、依然として精度が良くない場合があるため、VisualRank アルゴリズム [30] でリランキングを行い、その上位 k 件の画像集合 $I_k(c)$ から色特徴量および SIFT 特徴量 [31] を抽出する。概念 c の有する典型的な性質としての色特徴量 p の相応しさの度合い $\text{has-imgC}(p, c)$ は、上位 k 件の画像において占有する画素面積の割合の平均で計算する。

$$\text{has-imgC}(p, c) := \frac{\sum_{i \in I_k(c)} \text{prop}(p, i)}{k} \in [0, 1]$$

但し、 $\text{prop}(p, i)$ は、Web 画像 i において色特徴量 p を占有する画素面積の比率を返す関数である。一方、SIFT 特徴量 p の相応しさの度合い $\text{has-imgS}(p, c)$ は、上位 k 件の画像が有する割合で計算する。

$$\text{has-imgS}(p, c) := \frac{\sum_{i \in I_k(c)} \text{cont}(p, i)}{k} \in [0, 1]$$

但し、 $\text{cont}(p, i)$ は、Web 画像 i が SIFT 特徴量 p を含有していれば 1、含有していなければ 0 を返す関数である。

Step 3. 性質継承の度合いの評価：

対象概念 x の典型的な性質として部分全体関係や振る舞いを表す言語的記述を用い、典型的な性質としての相応しさ上位 n 件の集合 $P_n(x)$ を下位概念候補 y が継承している度合い $\text{inherit-txt}_n(y, x)$ は次式で定義されている。

$$\text{inherit-txt}_n(y, x) := \frac{\sum_{p \in P_n(x)} \text{has-txt}(p, y) \cdot \text{has-txt}(p, x)}{\sum_{p \in P_n(x)} \text{has-txt}(p, x) \cdot \text{has-txt}(p, x)}$$

一方、対象概念 x の典型的な性質としての色特徴量および SIFT 特徴量を用い、上位 n 件の典型的な性質集合 $P_n(x)$ を下位概念候補 y が継承している度合い $\text{inherit-imgC}_n(y, x)$ および $\text{inherit-imgS}_n(y, x)$ も以下のように同様に定義する。

$$\text{inherit-imgC}_n(y, x) := \frac{\sum_{p \in P_n(x)} \text{has-imgC}(p, y) \cdot \text{has-imgC}(p, x)}{\sum_{p \in P_n(x)} \text{has-imgC}(p, x) \cdot \text{has-imgC}(p, x)}$$

$$\text{inherit-imgS}_n(y, x) := \frac{\sum_{p \in P_n(x)} \text{has-imgS}(p, y) \cdot \text{has-imgS}(p, x)}{\sum_{p \in P_n(x)} \text{has-imgS}(p, x) \cdot \text{has-imgS}(p, x)}$$

最後に、以上で定義した三種類の性質継承度を次式によって統合し、下位概念候補 y が対象概念 x の下位概念である相応しさの度合い $\text{isa-PI}_n(y, x)$ を評価する尺度として用いて、下位概念候補集合 $y \in C(x)$ の全てをランキングした結果を返す。

$$\begin{aligned} \text{isa-PI}_n(y, x) := & (1 - \alpha - \beta) \cdot \text{inherit-txt}_n(y, x) \\ & + \alpha \cdot \text{inherit-imgC}_n(y, x) \\ & + \beta \cdot \text{inherit-imgS}_n(y, x) \end{aligned}$$

但し、 $\alpha \in [0, 1]$ および $\beta \in [0, 1]$ は、三種類の性質継承度を統合するための線形パラメータである。

4. 評価実験

本章では、本稿で提案した各概念を表す（典型）画像が持つ特徴量の性質継承に基づく下位語抽出手法を検証するため、収集した下位語候補群を構文パターン（に合致する Web 文書頻度）だけに基づいてランキングする手法、及び、これまでの著者の提案手法である各概念の意味的・言語的な has-a 関係の性質継承だけに基づく下位語抽出手法との比較実験を行う。

表 1 のように「動物」に属する五種類の概念（カテゴリー）を表す上位語を三種類の下位語抽出手法の入力として、適合率・再現率グラフを描くと図 2 のようになる。但し、本稿で提案した画像特徴量（とテキスト的な has-a 関係）の性質継承に基づく下位語抽出手法における各パラメータは、 $n = 10$, $k = 100$, $\alpha = 0.5$, $\beta = 0$ である。再現率が 0.0 から約 0.35 と低い区間では、構文パターンに基づく下位語抽出手法の適合率が最も高いが、それ以降の再現率が高い区間では、性質継承に基づく下位語抽出手法が常に上回っており、再現率を改善しつつ適合率も高く維持できている。次に、MAP (Mean Average Precision) で比較すると、構文パターンに基づく下位語抽出手法では 0.566、Web テキストの has-a 関係の性質継承に基づく手法では 0.507、そして、本稿で提案した Web 画像の特徴量の性質継承に基づく手法では 0.637 である。構文パターンという十分条件だけでなく性質継承という必要十分条件を考慮することによって約 12.5% の改善、Web テキストだけでなく Web 画像も活用することによって約 25.6% の改善が得られている。

表 1 実験対象の上位語の下位語候補数と正解数

対象語	下位語候補数	正解数
哺乳類	104	38
鳥類	42	11
魚類	53	8
爬虫類	44	13
両生類	31	8

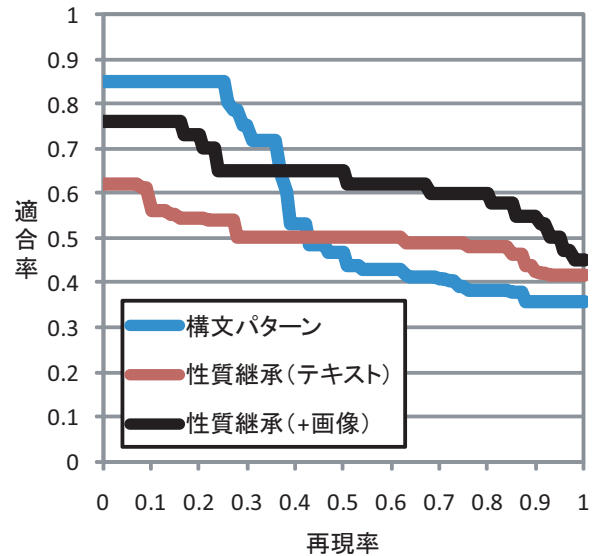


図 2 適合率・再現率グラフによる比較

5. おわりに

上位下位 (is-a) 関係や部分全体 (has-a) 関係といった概念階層は、様々な自然言語処理システムにとって非常に重要な基本的知識である。人海戦術的に人手で概念階層 (辞書) の構築が進められている一方で、Web などの大規模な文書データベースから自動的に知識抽出する研究も数多く行われている。これまでの研究において、後者のアプローチとして、各概念の性質として has-a 関係や振る舞いを表す言語的記述を Web から抽出した上で、対象概念から下位概念候補への性質継承の度合いに基づく抽出手法を提案して来た。本稿では、既存手法のように各概念の意味的・言語的な has-a 関係の性質継承の度合いを計算するだけでなく、各概念を表す (典型) 画像が持つ特徴量の性質継承の度合いも計算し、これら二つを統合して上位下位関係の評価とすることによってロバスト性の向上を図り、Web テキストだけでなく Web 画像も活用して概念階層を自動構築する手法を提案した。

文 献

- [1] Mandala, R., Tokunaga, T., and Tanaka, H.: "The Use of WordNet in Information Retrieval," Proceedings of the COLING ACL Workshop on Usage of WordNet in Natural Language Processing, pp.31-37 (1998).
- [2] Hattori, S., Tezuka, T., and Tanaka, K.: "Activity-based Query Refinement for Context-aware Information Retrieval," Proc. of the 9th Int'l Conference on Asian Digital Libraries (ICADL'06), LNCS Vol.4312, pp.474-477 (2006).
- [3] Hattori, S., Tezuka, T., Hiroaki, O., Oyama, S., Kawamoto, J., Tajima, K., and Tanaka, K.: "ReCQ: Real-world Context-aware Querying," Proc. of the 6th Int'l and Interdisciplinary Conference on Modeling and Using Context (CONTEXT'07), LNAI Vol.4635, pp.248-262 (2007).
- [4] 服部 峻, 田中 克己: "外観・状況表現を用いたオブジェクト名検索," iDB フォーラム 2008, 情報処理学会研究報告「データベースシステム」, Vol.2008, No.88, pp.109-114 (2008).
- [5] Hattori, S. and Tanaka, K.: "Object-Name Search by Visual Appearance and Spatio-Temporal Descriptions," Proc. of the 3rd Int'l Conference on Ubiquitous Information Management and Communication (ICUIMC'09), pp.63-70 (2009).
- [6] Fleischman, M., Hovy, E. and Echiabi, A.: "Offline Strategies for Online Question Answering: Answering Questions Before They Are Asked," Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics (ACL'03), pp.1-7 (2003).
- [7] 服部 峻, 手塚 太郎, 田中 克己: "オブジェクトの外観情報の Web マイニング," 電子情報通信学会第 18 回データ工学ワークショップ (DEWS'07) 論文集, L4-6 (2007).
- [8] Hattori, S., Tezuka, T., and Tanaka, K.: "Mining the Web for Appearance Description," Proceedings of the 18th International Conference on Database and Expert Systems Applications (DEXA'07), LNCS Vol.4653, pp.790-800 (2007).
- [9] 服部 峻, 田中 克己: "コンテキストに依存する外観情報の Web からの抽出," 電子情報通信学会第 19 回データ工学ワークショップ (DEWS'08) 論文集, A2-1 (2008).
- [10] Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., and Miller, K. J.: "Introduction to WordNet: An On-line Lexical Database," International Journal of Lexicography, Vol.3, No.4, pp.235-312 (1993).
- [11] Völkel, M., Kröttsch, M., Vrandečić, D., Haller, H., and Studer, R.: "Semantic Wikipedia," Proceedings of the 15th International Conference on World Wide Web (WWW'06), pp.585-594 (2006).
- [12] 青木 利晃, 片山 卓也: "オブジェクト指向方法論のための形式的モデル," 日本ソフトウェア科学会学会誌コンピュータソフトウェア, Vol.16, No.1, pp.12-32 (1999).
- [13] 王 凱軍, 池田 満, 國藤 進: "属性分析法に基づく類似性の分析," 第 18 回人工知能学会全国大会, 2F3-02 (2004).
- [14] 服部 峻, 田中 克己: "性質継承と概念の再帰的適用に基づく Web からの概念階層抽出," 情報処理学会論文誌 (トランザクション) データベース, Vol.1, No.3 (TOD40), pp.60-81 (2008).
- [15] Hearst, M. A.: "Automatic Acquisition of Hyponyms from Large Text Corpora," Proceedings of the 14th International Conference on Computational Linguistics (COLING'92), Vol.2, pp.539-545 (1992).
- [16] Caraballo, S. A.: "Automatic Construction of a Hypernym-Labeled Noun Hierarchy from Text," Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics (ACL'99), pp.120-126 (1999).
- [17] 安藤 まや, 関根 聡, 石崎 俊: "定型表現を利用した新聞記事からの下位概念単語の自動抽出," 情報処理学会研究報告「自然言語処理」, Vol.2003, No.98, pp.77-82 (2003).
- [18] Emmanuel, M., Christian, J.: "Automatic Acquisition and Expansion of Hypernym Links," Computer and the Humanities, Vol.38, No.4, pp.363-396 (2004).
- [19] 鶴丸 弘昭, 竹下 克企, 伊丹 克企, 柳川 俊英, 吉田 将: "国語辞典情報を用いたシソーラスの作成について," 情報処理学会研究報告「自然言語処理」, Vol.1991, No.37, pp.121-128 (1991).
- [20] 桜井 裕, 佐藤 理史: "ワールドワイドウェブを利用した用語説明の自動生成," 情報処理学会論文誌, Vol.43, No.5, pp.1470-1480 (2002).
- [21] 大石 康智, 伊藤 克亘, 武田 一哉, 藤井 敦: "単語の共起関係と構文情報を利用した単語階層関係の統計的自動識別," 情報処理学会研究報告「音声言語情報処理」, Vol.2006, No.40, pp.25-30 (2006).
- [22] 森本 貴之, 藤原 謙: "例外処理を考慮した用語間の階層・関連関係の抽出," 情報知識学会第 8 回研究報告会講演論文集, No.8, pp.17-22 (2000).
- [23] 小淵 洋一, 齊藤 隆: "意味の分割によるシソーラスの自己組織," 情報処理学会研究報告「情報学基礎」, Vol.1992, No.54, pp.17-23 (1992).
- [24] Sanderson, M., and Croft, B.: "Deriving Concept Hierarchies from Text," Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp.206-213 (1999).
- [25] Etzioni, O., Cafarella, M., Downey, D., Kok, S., Popescu, A.-M., Shaked, T., Soderland, S., Weld, D.S., and Yates, A.: "WebScale Information Extraction in KnowItAll (Preliminary Results)," Proceedings of the 13th International World Wide Web Conference (WWW'04) (2004).
- [26] 山本 英子, 神崎 享子, 井佐 原均: "出現状況の包含関係による語彙の階層構造の構築," 情報処理学会論文誌, Vol.47, No.6, pp.1872-1883 (2006).
- [27] 新里 圭司, 鳥澤 健太郎: "HTML 文書からの単語間の上位下位関係の自動獲得," 自然言語処理, Vol.12, No.1, pp.125-150 (2005).
- [28] 鶴丸 弘昭, 前田 英幸, 山本 和博, 日高 達, 吉田 将: "国語辞典に基づくシソーラスの構築に関する一考察," 電子情報通信学会技術研究報告「言語理解とコミュニケーション」, Vol.93, No.367, pp.29-36 (1993).
- [29] Sundblad, H.: "Automatic Acquisition of Hyponyms and Meronyms from Question Corpora," Proceedings of the ECAI'02 Workshop on Natural Language Processing and Machine Learning for Ontology Engineering (2002).
- [30] Jing, Y. and Baluja, S.: "VisualRank: Applying PageRank to Large-Scale Image Search," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.30, No.11, pp.1877-1890 (2008).
- [31] Lowe, D. G.: "Distinctive Image Features from Scale-Invariant Keypoints," International Journal of Computer Vision, Vol.60, No.2, pp.91-110 (2004).