

一般ユーザーにインタビューする対話エージェント

鳥澤 健太郎

北陸先端科学技術大学院大学 情報科学研究科
〒923-1292 石川県能美市旭台 1-1
E-mail: torisawa@jaist.ac.jp

あらまし Web等にある大量の文書から、自動的な知識獲得手法によって獲得された知識を用いて、多様な対象に関して一般ユーザーにインタビューを行なう対話エージェントについて述べる。このシステムの狙いは、多様な情報を一般ユーザーのインタビューによって収集すること、ならびに、従来、blog等のCGMに存在しなかったCGM記述上のガイドラインを自然な形で、ユーザー、すなわち消費者に提示し、質の高いCGM-likeな情報を提供させることである。

キーワード 対話エージェント, 知識獲得

A Dialog Agent That Interviews Users

Kentaro Torisawa

School of Information Science, 1-1 Asahidai, Nomi, Ishikawa, 923-1292 Japan
E-mail: torisawa@jaist.ac.jp

Abstract This paper describes a dialog agent that can interview users about various objects by using the knowledge acquired from a large amount of texts, including Web documents, in an automatic manner. The objective of this research is to let the dialog agent provide the users a certain guideline in writing CGM, such as blogs, in a natural way. Such a guideline does not exist in existing CGM and we are expecting that it can enable users to write high-quality CGM.

Keyword Dialog agent, knowledge acquisition

1. はじめに

本稿では、多様な対象に関して、一般ユーザーにインタビューを行なう対話エージェント AINT(Dialog Agent that INTERviews you)について述べる。本来、対話エージェントは自然言語処理の究極のゴールであると長年にわたり見なされて来た。しかしながら、ここ10年ほどの自然言語処理技術の進歩には瞠目すべきものがあるにもかかわらず、対話エージェントに関する自然言語処理技術ははかばかしい進捗が得られていないというのも事実であろう。実際、自然言語処理分野での様々な学会、あるいは研究会でも対話エージェントに関する発表は少数派と言ってよいであろう。

本研究を行なうにあたっては、このような現状に一石を投じるという狙いもあった。つまり、最新の自然言語処理技術を用いて、従来の対話エージェント研究にはない視点から、その開発を行なうということである。ここでキーとなる概念は対話エージェントの領域依存性である。主としてアカデミックな文脈でこれま

でに開発されてきた対話エージェントは多かれ少なかれ領域依存なものである。これは、これらの対話エージェントが精緻かつ網羅的な知識を必要とするタスク、例えば航空券の予約を実現することを狙っており(多数の論文があるが、例えば[1,2,3])、また、そのような網羅的な知識を記述することが、「航空券の予約」のような、比較的小さな領域を仮定しないかぎり難しいという事情による。

一方で、インターネット上には、いわゆる「人工無能」と呼ばれる対話エージェントが存在する。これは航空券の予約といった人間にとって「有意味な手続き」を実行するのをあきらめた代わりに、広い範囲の入力に対して一見有意味に思われる自然言語のレスポンスを返すというものである。このエージェントは非常にshallowな方法論によって開発されているため、領域依存性はそれほど強くない。しかしながら、人間にとって有意味な手続きを実行しない以上、その存在意義は不明瞭であり、これまではいわゆる一種の娯楽手段



AINT: 何してたんですか? 昨日は。
 ユーザー: スカールクニンっていうレストランに行ってきたよ
 AINT: そのスカールクニンでなにを食べたの? きたさんは ミーゴレンを食べたよと話している、なかむらさんは グリーンカレーを食べたよって言ってますけれど。
 ユーザー: ナシゴレンを食べたよ
 AINT: スカールクニンの住所は? きたさんは 金沢市伏見だと思ふ。って言うていて、なかむらさんは そのとおりです。と話していますが。
 ユーザー: 伏見幼稚園のそばだね
 AINT: そのスカールクニンで食べた感想は? きたさんは なんか甘いような辛いような不思議な感じ。なんか店の人がバリ島に凝ってるそうで、内装もこってるし。って言うているんだけど、なかむらさんは そう。ちょっと面白い。って言うてますけれど。
 ユーザー: うん、いいんじゃないでしょうか
 AINT: スカールクニンの営業時間は? きたさんは 昼と夜やってるという以外はわかりません。と答えており、なかむらさんは 夜11時ぐらいまでやってる以外はわからない。 [注: 右上に続く]

って言うてますけれど。
 ユーザー: 昼は12時からやってたと思う
 AINT: スカールクニンのアクセスは? きたさんは 近くに駅はないね。バスはあるとおもうけどと答えてますが。
 ユーザー: 確かにバス停あるよ
 AINT: スカールクニンの駐車場は? きたさんは 店の前に数台止められる感じと答えており、なかむらさんは うんって言うてますけれど。
 ユーザー: 4台ぐらい止められる
 AINT: スカールクニンの電話は? なかむらさんは 不明と答えてますが。
 ユーザー: 右に同じ
 AINT: どうやってスカールクニンに行ったの? きたさんは 家から近いから歩いて行ったよ。と話している、なかむらさんは 自転車って言うてますけれど。
 ユーザー: 車
 AINT: そのナシゴレンというのは料理のこと?
 [注: 過去に入力された名詞に関して質問開始]
 ユーザー: そうだよ
 AINT: どこでナシゴレンを作ったの?
 ユーザー: 当然、スカールクニンでしょ。
 AINT: そのナシゴレンを食べた感想は?
 ユーザー: 割と複雑な味、多分調味料いろいろつかってるんだね。
 AINT: どうやってナシゴレンを食べたの?
 ユーザー: そりゃ、箸で
 AINT: どうやってナシゴレンを作ったの?
 ユーザー: ちょっとわからないけど、専用調味料をうってるのは見たことあるね。後チャーハンといっしょじゃないか
 AINT: そのチャーハンというのは作品、それとも料理のこと? [注: 過去に入力された名詞に関して質問開始]

図1: AINTの実行時画面と対話例 (実際のデモは上記対話の抜粋が以下のURLから閲覧できる。
<http://www-tori.jaist.ac.jp:8000/project/demoJ.mp4>)

としかみなされておらず、また、人工無能の開発自体が自然言語処理の研究とはみなされてこなかった。

本研究の狙いは、以上に述べた両極端の試みの中間を狙う。最新の自然言語処理技術、特に形態素解析、構文解析は非領域依存であると言っても差し支えないレベルにあるが、これを利用しつつ、新たな種類の「有意義な手続き」を設定し、この手続きを実行するタスクをこなすことのできる、領域依存性の強くない対話エージェントの枠組みを提案することである。具体

的には、ユーザーにインタビューを行なうことで、有用な情報を収集し、その情報を他のユーザーに提示するというタスクを設定し、領域依存性の低い対話エージェントを開発する。図1に本研究で実際に開発した対話エージェント AINT の実行時画面、及び対話例を挙げるが、これから分るように、対話エージェントはユーザーに対して、ユーザーの入力に現れたレストランや料理といった多様な対象に関して質問を發し、ユーザーから情報を収集している。また、質問を發する

と同時に他のユーザーからやはりインタビューで収集した情報を提示する、一種の掲示板機能を持っている。重要なことは、このような対話を行なう対話エージェントは領域が限定されているならば、比較的簡単に構築することができるものの、これを多様な対象に関するインタビューに拡張するのは難しいということである。本研究の狙いはまさにその多様な対象に関するインタビューを行なうこと、つまりは領域依存性をなるべく緩和した対話エージェントを開発することにある。

以下では、まず、インタビューを行なうというタスクについて述べ、そのタスクと領域依存性の関係について説明する。ついで、本対話エージェントの動作について述べ、最後に不完全ながら行なった評価実験について述べる。

2. タスクの設定：インタビュー

以上に述べたように、本研究ではユーザーの視点から見た「有意味な手続き」として、一般ユーザーから有用な情報を、インタビューによって収集、蓄積するという手続きを想定する。まず、ここで当然の疑問は、そのような対話エージェントはそもそも有用と言えるのか、というものである。我々の意図は、対話エージェントを一種の CGM (Consumer Generated Media) として利用することである。つまり、対話エージェントがあるユーザーからインタビューによって得た情報を、やはり対話エージェントが他のユーザーに提示したり、あるいは、そもそも収集された情報をブログのような形で文書として他のユーザーに提示するという狙いを狙う。ブログやレビューサイトを典型的な例とする CGM は、一般にどのような内容を書くか、ユーザーに対してなんら制約を課さない。これは多様な視点からの情報を記述できるという点において望ましいが、一方でその質を保証できないといった欠点もある。例えば、レストランに関するブログの記事で、ユーザーが実際に食べた料理の名前を書いていないとすれば、その記事の有用性は落ちると言わざるを得ない。通常、ブログを書くのは、「書き手」としての訓練を経ない人々であり、実際、重要な情報の欠落したブログ記事、レビューは良く目にする。本研究では、システム側から質問するという形で、そのような記述すべき内容を自然な形でサジェストし、CGM の質の向上を狙う。また、CGM を書くのは、通常アクティブなユーザーに限られているが、インタビューという能動的な働きかけを行なうことで、アクティブでないユーザーから情報を収集することも狙う。例えば、カーナビなどに本研究で開発したような対話エージェントを実装し、そのカーナビで行った場所について、必ずしもアクティブでないユーザーから有用な情報を入手、他の

ユーザーに提示するといった応用も考えられよう。

また、インタビューの重要な特徴として、対話の主導権がインタビューする側、つまりエージェント側にあるということが挙げられる。これは、ユーザー主導の対話にくらべて、ユーザーからの入力の変換を縮減させるということにつながり、実現に必要な知識の量を押し下げるのが可能となる。これによって、対話エージェントの頑健性を確保するのがより容易になるものと考えている。

3. 領域依存性とインタビューというタスク

技術的に見ると、以上に述べたインタビューというタスクは、これまで開発が行なわれて来た対話エージェントの多くが行なうタスクと比べると、そもそも情報の流れが逆である。つまり、通常対話エージェントでは主として詳細な情報がエージェントからシステムに提示されるのに対して、本研究で開発する対話エージェントにおいては、詳細な情報は主としてユーザーからエージェントに流れる。我々は、この情報の流れの向きが、対話エージェントの領域依存性を考える上で非常に重要であると考えている。より具体的には、そもそも、これまでのエージェントからユーザーに情報が流れるという想定でのエージェントは詳細な情報を前もって内部にもっており、ユーザーが入力し得る多様な自然言語表現とその詳細な情報を結びつける必要があった。例えば、航空券の予約を行なう対話システムでは、飛行機の時刻表、ルート、運行状況等の詳細な情報がシステム内に存在し、対話時には、ユーザーの多様な入力をそのような情報の特定の部分に対応づける必要が出てくる。このような対話エージェントが詳細かつ網羅的な知識を必要とする理由は、このシステム内の情報がそれなりに複雑な構造をしており、ユーザーの入力と、その複雑な構造やその構造に対する操作への自然言語での言及方法をシステム内に知識として持っていなければならないからである。

であるならば、単に領域依存性を排除するという目的を達成したいのであれば、システム内に複雑な情報を必要としないタスクを設定すれば良いということになる。本研究で設定したインタビューというタスクは、前もって複雑な情報をシステムに組み込む必要がなく、そのような点において望ましい。もちろん、対話エージェントがインタビューにおいて質問を発するためには、一定量の知識が必要となるが、これにはデータベース、知識ベースへのアクセスのような手続的知識は必要なく、あくまで言語の枠内での知識となるため、昨今、盛んに研究されている自然言語文書からの(言語的)知識獲得手法である程度対応可能である。

また、既に述べたように、インタビューにおいては、詳細な情報が主としてユーザーから対話エージェント

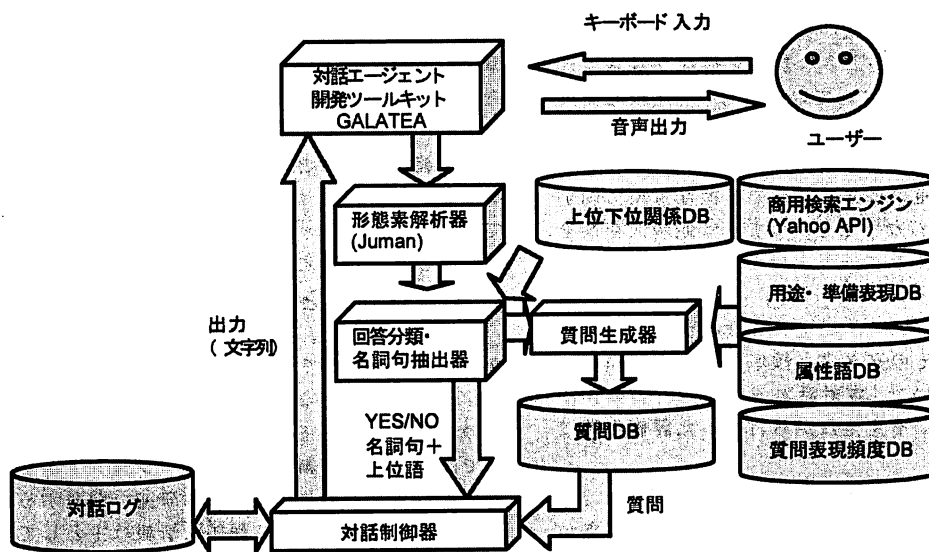


図2. 対話エージェント AINT の構造

へと流れる。この情報を他のユーザーに提示するためにはやはり、一定の整理を行なう必要があるが、この整理法の最も単純なものとしては、実際に対話エージェントが発した質問文を整理の際のラベル、あるいは構造として使うことが考えられる。(また、図2の対話例を注意深く読むと分るように、実際に現状の対話エージェントではそのようになっていっている。)つまり、インタビューというタスクにおいては、ユーザーから情報を収集するための質問とエージェント内に保持する情報の構造とが同一視できる。これは対話エージェントの保持すべき情報の処理があくまで言語の枠内で行なえるということの意味するが、これも我々の対話エージェントの領域依存性、あるいは実現を容易にする理由である。

まとめると、本研究では領域依存性の緩和を狙うがこれは我々の設定したインタビューというタスクにおいては、より容易になる。これは、インタビューとその結果の他のユーザーへの提示があくまで、言語の枠内の処理だけだすみ、また、必要とされる知識も言語的なものにとどまるため、大量の文書からの自動的な知識獲得によってある程度抽出できる。次節では、このような視点を踏まえつつ、対話エージェントの具体的な実現方法について述べる。

4. 対話エージェントの設計と実現

4.1. 概要

本研究で開発した対話エージェント AINT の基本的動作は非常に単純である。とにかく、ユーザーの入力

文中にある名詞句をピックアップし、それに関する質問を生成し、その回答をログとして記録する。また、過去において、同一の対象に関して、同一の質問を他のユーザーに発したことがあれば、それに対する他のユーザーの回答をログから取り出し、ユーザーに提示する。AINT の目的はユーザーから多様な対象に関して、有用な情報を収集することであり、質問の生成にあたってはこの有用性という点が肝要となる。本研究においては、ある対象に関して有用な情報とは、その対象を「利用する」際に重要な情報であるとみなし、対象の利用法、準備法[4]、属性語[5,6]に関する質問を生成する。

AINT の構造を図2に示す。まず、ユーザーとのやり取りは、擬人化エージェント開発ツールキット GALATEA

(<http://hil.t.u-tokyo.ac.jp/~galatea/index-jp.html>) を介して行なわれる。これは音声入力、音声出力を処理すると同時に、対話エージェントを擬人化した存在である人の頭部を表示し、それを音声入出力とシンクロさせて動作させることが可能なツールである。AINT では、多様な対象に関する対話を想定しているため、現状の限られたボキャブラリでの音声認識では不十分であることが予想される。従って、本研究では出力のみ音声で行なわせることとした。

ユーザーからのキーボード入力は、まず、形態素解析器(juman)をかけられ、解答分類・名詞句抽出器を適用される。このモジュールの出力は2種類ある。一つは、入力文中にある名詞句のリストであり、これは後

に質問文を生成するのに使われる。この名詞句は簡単な語彙統語パターン[7, 8, 9], 例えば、「XっていうY」によって、その名詞句の上位語が認識できる場合には、上位語が付与されている。(e.g., 「スカールクニンっていうレストラン」, 図1参照) また、もう一つの出力は、入力の意味するところを YES, NO, 「いずれでもない」の3者に分類したものである。これは、後ほど説明するように、対話において必要な確認をシステムが発した際のユーザーのレスポンスを処理するのに使われるが、実装としては、入力文を単語ベクトルに変換し、それを SVM で分類することによって実現されている。ついで、入力文から抽出された名詞句は、まず、質問生成器に与えられ、後述するような方法によって、それらの名詞句に対する一連の質問文が生成され、質問 DB(データベース)に格納される。対話制御器は有限状態オートマトン風の内部状態を持っており、この状態とユーザーの入力、及び、質問 DB の内容、それから過去の対話ログからユーザーへのレスポンスを生成し、また、ユーザーの入力を直前の質問と関連づけた上で対話ログに書き込む。

以上のアーキテクチャで重要な事は、ユーザーの入力の「理解」は最小限にとどめられているということである。デモを見た研究者からも誤解をうけるが、ユーザーの入力は、YES, NO といった必要最低限のラベルと、そもそもエージェントが発した質問に対して答えているという前提だけによって、処理されている。このような単純な仕組みであっても、一見対話が行なわれているように見えるのは、あくまで、エージェントの発する質問を参照点としつつ、エージェントが主導的に対話を組み立てているからであり、タスクの設定によっては非常に簡単な仕組みによって、有意味な手続きを実行しつつ、対話らしきことが出来る可能性を示唆している。

一方で、ユーザーの入力を簡単な解析だけで済ませているが故の問題点はもちろん多々ある。例えば、ユーザーが他の質問に対して発した入力文にすでに答えが書いてあるにも関わらず、その答えを得るための質問を発することも当然ながらある。(e.g., 「昨日は何をしてたのですか?」「スカールクニンというレストランでナシゴレン試した。」「そのスカールクニンというレストランで何を食べたのですか?」)これは、例えばユーザーの質問文に現れた動詞を記録しておき、それを利用した質問文を発するのを避けるといった簡単なヒューリスティックによってある程度避けることができるが、そもそも「言い換え」(試す⇔食べる)や動詞の省略が起きた場合には無力であるため、現在はあえて、そのようなヒューリスティックは実装していない。これは今後言い換え、省略の解析等の研究が

進んだ場合に対応する予定である。

4.2. 質問生成器

AINT の核となる技術は、ユーザーの入力文中にある多様な名詞句に対して質問を生成する手法である。まず、質問文は、前述したように、大きく、対象の利用に関するもの、(e.g., 「レストランで何を食べたの?」)、利用する上での準備 (e.g., 「レストランにどうやって行ったの?」)、属性語(e.g., 「レストランの住所は?」)の3種に別れる。具体的な質問は、入力文中の名詞句の上位語に対して、そのコアとなる表現(「食べる」「行く」「住所」等)を、前もって自動的な知識獲得によって作成しておいたデータベースから取り出し、変形を施すことで生成される。入力文中の名詞句 (e.g., 「スカールクニン」) に関してではなく、その上位語 (e.g., 「レストラン」) に関して質問を生成するのは、往々にして入力文中の名詞句は出現頻度が少ないため、知識獲得手法が適切な知識を獲得できていない可能性が高いからである。より具体的には、質問は以下のプロセスによって生成される。

ステップ1: それまでに与えられた入力文中の名詞をすべて取り出す。

ステップ2: 各名詞句に、もし、語彙統語パターンによって上位語が付与されている場合には、その上位語と名詞句、そうでなければ、Web から抽出された上位下位 DB ([10]の手法で獲得された約40万対の上位下位関係を格納)を参照して可能な上位語をすべて名詞句と組み合わせる

ステップ3: すべての名詞句のヒット件数を商用検索エンジン(現在は Yahoo API)で調べる。もし、その値が上限(現在は300万)を越えていれば、質問の対象とはしない。これはあまりにヒット件数の多い名詞はインタビューの対象とならないほど自明の存在である可能性が高いからである。

ステップ4: ヒット件数の大小、名詞句の字種(カタカナかどうか)、入力文がどのようなタイプの質問に答えたものか、などを考慮したヒューリスティックス、乱数によって質問の対象となる名詞句を選ぶ。字種に注目するのは、カタカナで書かれた名詞句はブランド名、料理名など質問の対象として適切であることが多いからである。

ステップ5: 選んだ名詞句に付与されている上位語が適切であるか、あるいは複数個ある場合には、適切なものをユーザーに選んでもらう。まず、最初にユーザーに対して、エージェントが認識している名詞句と上位語の候補が正しく、上位下位関係になっているかどうかを問い合わせ、(e.g., 「そのナシゴレンというのは料理のこと?」)それに対して否定的な答えが帰って来た場合や、そもそも上位語の付与に失敗した場合

には上位語を直接入力してもらう。(e.g., 「ナシゴレンというのは何のこと?」) なお, 上位下位関係 DB によって複数の上位語が付与されている場合には, その内からユーザーに適切な上位語を選択してもらう。

ステップ6: Web 文書から[4, 6]にある手法によって獲得された, 対象の利用を表す用途表現(「レストランで食べる」), 対象を利用する上での準備を表す準備表現(「レストランに行く」) 属性表現(「レストランの住所」)等を上位語に対して DB から取り出し, 質問を生成する。この際, 用途表現, 準備表現に関しては, 「何を」「どうやって」等の疑問代名詞と用途表現, 準備表現とを組み合せ, 疑問文に変形した上で Web 文書中の出現頻度をチェックし, 頻度が極端に小さい質問文は候補から除外する。生成された質問は, 質問 DB に格納され, 適宜ユーザーに対して発せられる。

5. 評価実験

対話エージェントの評価は非常に難しい。本稿では, 特に生成される質問のクオリティーに関してのみ簡単な評価実験を行なった。まず, 新聞に頻出する名詞句の内, Web 文書約 7600 万文書中で, 「X を楽しむ」「X を使用する」「X を用いる」「X を利用する」「X を使う」などの表現に 10 回以上現れている 4,341 個の名詞句を選びだして, 上記の方法で 12,467 個の質問を生成した。(名詞句は, 実際に質問をされる対象の上位語として想定している。実際の対話では, それらの上位語がその下位語に展開されることになるため, 質問のバリエーションは約 1 万 2 千という数よりもはるかに大きくなる。) それらの質問の内 100 個をランダムに選び出して 2 名の被験者に評価してもらった。より具体的には以下のラベルを各質問に割り振ってもらった。ただし, 質問への解答をするユーザーは既にその質問の対象に関して「利用する」経験をしたものであると仮定した。

- A. 質問文が一般的な状況で質問してもおかしくないものであり, もし答えが得られればそれは有用なものと期待できる。
- B. 質問の対象となっている名詞句の意味が非常に曖昧であるが, その意味が絞り込まれている限りにおいて, 一般的な状況で適切と見なせる。
- C. 特定の状況下では, 質問は適切であり, その答えは有用なものと期待できる。
- D. A,B,C のいずれでもない。

表 1 に結果を示す。文脈に依存せず適切とされた質問はすべての質問の 40% から 50% であるが, 状況依存性の高い質問が 30% 以上あり, ナンセンスな質問である D は両被験者とも 20% 以下としている。これらの数字は必ずしも, 現状のシステムがすぐに実用になるこ

とを意味してはいないが, 今後, 大量の Web 文書等から, 質問が適切な文脈に現れるキーワード等を抽出し, 質問生成時に使うことや, なんらかの文脈に関する質問を適宜ユーザーに発することで, 改善される可能性も示唆している。また, 現状では特に用途表現, 準備表現の学習に約 600 個の名詞に対する正解データ, 及び比較的少量のテキストコーパスを使っているが, これらを今後増加させることで, さらなる精度向上を図る予定である。

ラベル	被験者 1	被験者 2
A	55	43
B	1	4
C	34	35
D	10	18

表 1: 質問の妥当性

6. おわりに

本稿では, 多様な対象に関して, ユーザーにインタビューを行なう対話エージェントについて報告し, タスクによっては, 現状の自然言語処理技術でも領域依存性の低い対話エージェントを構築できる可能性を示した。しかしながら, 性能評価や, インタビュー中での質問の妥当性, 特に文脈に依存した妥当性など問題点は多く, 今後も改善を続ける予定である。

文 献

- [1] E. Levin et al., “The AT&T-DARPA Communicator Mixed-Initiative Spoken Dialog System”, ICSLP2000, Beijing, China, Oct, 2000
- [2] 渡辺卓夫, 大澤一郎, 米澤明憲, “対話領域の独立性を指向した日本語対話理解システム”, 人工知能学会誌, Vol 2, No. 2, pp. 206-213, 1987
- [3] Y. Kiyota et al., “Dialog Navigator: A Spoken Dialog Q-A System based on Large Text Knowledge Base”, ACL 2003 interactive poster and demo session, pp. 149-152, Sapporo, Japan, July, 2003
- [4] 鳥澤健太郎, “対象の用途と準備を表す表現の自動獲得”, 自然言語処理, Vol 13(2), 2006
- [5] K. Tokunaga, et al., “Automatic Discovery of Attribute Words from Web Documents”, IJCNLP 2005, pp. 106-118, 2005
- [6] N. Yoshinaga, et al., “Finding Specification Pages According to Attributes”, WWW06 (poster), pp. 38-39, 2006
- [7] M. A. Hearst, “Automatic Acquisition of Hyponyms from Large Text Corpora”, COLING’92, pp. 539-545, 1992
- [8] 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”, 情報処理学会研究会報告 2003-NL-157, pp.77-82, 2003
- [9] 今角恭祐, “並列名詞句と同格表現に注目した上位下位関係の自動獲得”, 修士論文, 九州工業大学, 2001
- [10] A. Sumida et al., “Concept-instance Relation Extraction from Simple Noun Sequences Using a Search Engine on a Web Repository”, Workshop on the Web Content Mining with Human Language Technologies, Athens, USA, November, 2006