

オンラインゲームにおけるコミュニケーション支援のための Webを用いた情報抽出

高松 雅彦^{†1} 荒木 健治^{†1}

本研究では、ユーザにとってオンラインゲームの対話文内において未知語となる語の情報抽出を行う。オンラインゲーム上の対話文には、未知語、インターネット用語、スラングなどが多く含まれるため、新たにゲームを始めるユーザにとって発話内容の意図が理解しにくく、コミュニケーションが阻害される原因となる。そこで本稿では、ユーザが指定する任意の語について関連語とその語意をチャットログ内と Web から抽出し、ユーザに提示することでユーザ間のコミュニケーション支援を可能とするシステムを提案する。実験の結果、ユーザの発話内容の理解度の大幅な向上が見られ、支援システムとしての有効性を明らかにした。

Information Extraction from the Web for Supporting Communications on Online Games

MASAHIKO TAKAMATSU^{†1} and KENJI ARAKI^{†1}

In this paper, to extract information about unknown words in the sentence for the user's interaction on online games. Since the statement on the online game conversation includes many unknown words, Internet terminology and slang, it is difficult to understand what the intent of the utterance for the beginners, and causing the fails of the communication. In this paper, we extract related words and their meanings are specified by the user for any words from Web and chat logs, we propose a supporting system that enables communication between users. The evaluation experiments shows that significantly thier understanding of what the user's utterance improves the effectiveness as a support system is confirmed.

^{†1} 北海道大学大学院情報科学研究科

Graduate School of Information Science and Technology, Hokkaido University

1. はじめに

近年、テレビゲームが一家に一台といわれるほど普及している。テレビゲームは登場以来、ハード、ソフト共に多様化し社会に浸透し続けている。また、インターネットの普及により ADSL や光ケーブルといった技術により、音楽や映像など、大量の情報のやりとりが可能になり、様々な面で恩恵を受けている。

こうした中、ゲーム機をインターネットに繋いで遊ぶオンラインゲームが登場し始め、ゲームの形態がオフラインからオンラインへと変化していく中で、「MMO-RPG」(Massively Multiplayer Online Role Playing Game - 多人数同時参加型オンラインロールプレイングゲーム)が人気を集めている。

MMO-RPG では、ユーザはその世界の住人となり、他のプレイヤーと協力(あるいは対立)してプレイすることができる。非常に多数の、見ず知らずの人間が同じ仮想世界上でプレイするため、一般的なネットワークゲームと比べても他のプレイヤーとのコミュニケーションを楽しむ要素が強い。オンラインゲームを使用するユーザは、オンラインゲームに交流や気分転換を求めているという調査報告がなされている¹⁾。

オンラインゲームにおいて、ユーザの要望を把握したりシステムのエラーを発見するために、電子掲示板やチャットの文章が利用されることがある。その内容を把握し時間と共に変化する話題を抽出しようという研究も行われている²⁾。しかし、オンラインゲームにおけるユーザ発話は、未知語、インターネット用語、スラングなどが多く含まれる(表 1)。このことは、新たにゲームを始めるライトユーザにとって、他のユーザの発話内容の意図が理解しにくくコミュニケーションが阻害される原因となる。

チャット発話中の未知語をもとに Web 検索を行い、関連語となっている語について調べたとしても、検索結果のページ内から、チャット発話中に出現したキーワードを元に、重要単語を見つけ、キーワードに関する情報を収集する必要がある。適切なキーワードを用いて、Web 検索を行わなければ、適切な検索結果が得られない。また、大量の Web ページを見なければならず、目的の情報が書かれた Web ページを発見することができない可能性が高いという問題がある。

このような問題に対して、これまで関連語を抽出するための研究が多数提案されている。Ando らの研究³⁾では、大規模コーパスから構文パターンを作成し、定型表現を用いて下位概念の自動抽出を行い、連想辞書との比較を行うことで、下位概念の抽出を行っている。この手法では、大規模なコーパスやシソーラスといった語彙資源の集合を必要とする。それら

表 1 オンラインゲームにおける発話例
Table 1 Utterance examples on the online game

売】地獄 7M @ 1 耳ヨロ ~ ^^ ポタ有がお秘密鍵のみ募集 ^^ 沼秘密ありませんか? lv27 エンチャです 1 ポタあり。廃坑秘密にポタ持ち s (職なんでもお k) 募集 ~ 耳よろです ^^ 2 連予定】 滝秘密 PT から。持ちさん@2 BIS さん@1 耳お願いします ^^ オガ秘密募集 鍵 s B I S s 耳ヨロ ^^ 滝秘密ないですか? マジアチャです ^^ 2 連呪い秘密にてポタ持ち s 募集 1 個 7 万で売ってください。紋章用にいくつか買います

の語彙資源は一般的な語彙の集合であり、オンラインゲームの用語に特化したコーパスや辞書は存在しないため、有効な結果は得られないと考えられる。

佐藤らの研究⁴⁾では、与えられた専門用語から、4種類のクエリを作成する。そこから、重要文の抽出を行い、Webのヒット数を用いてフィルタリングを行い、関連する専門用語の抽出を行っている。

芳鐘らの研究⁵⁾では、ある用語と関係する用語を、同義関係、階層関係、関連関係の3種類のタイプごとに、シラバスから収集する手法の検討を行っている。

これらの研究では、名詞を対象として語の抽出を行っているが、チャットで用いられる単語は名詞に限られるわけではないので不十分であると考えられる。

そこで本稿では、大量の文章群を使わずに、単一のオンラインゲームのチャットログから関連語候補を抽出し、Webを用いて関連語候補を選定しスニペットを抽出する。次に情報をユーザに提示することでユーザ間のコミュニケーション支援を可能とするシステムを提案する。

2. システム概要

本システムの処理の流れを図1に示す。発話文中の任意の1語をユーザが選択しその語を入力(検索語)として、関連語候補を文内文脈情報を利用した手法と共起頻度を利用した手法により抽出し、それぞれの手法でWebから詳細な情報を取得し文類似度に基づいて統合し、その結果をユーザに提示する。チャットログに含まれる半角英数字と半角カタカナは、それぞれ全角英数字と全角カタカナへ自動的に変換している。

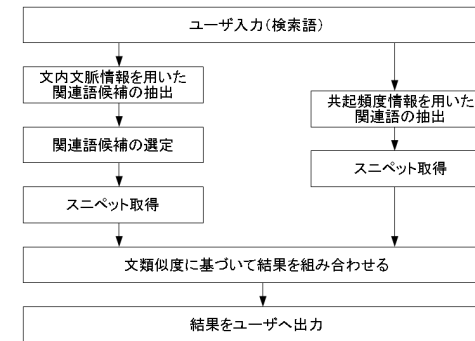


図 1 システム概要
Fig. 1 Flow of this system

表 2 「ポタ」を含む発話の例
Table 2 An example of the utterance including "Pota"

ポタ/出し/たい/ので/PT/組ん/で/もらえる/方/い/ます/か? 1/ポタ/有り/。/廃坑/秘密/に/ポタ/持ち/s/募集/~ /耳/よろ/です/^/^ ポタ/持ち/さん/@/2 ポタオガ/秘密/鍵/募集 赤目/秘密/ポタ/持ち/、/BIS s /募集/です/。 2 /連/ポタ/あり/オガ/秘密

2.1 文内文脈情報を用いた関連語抽出

2.1.1 オフライン処理

本手法は関連性のある単語は文法的に類似した表現で使われるという考えに基づいている。ユーザの検索語を入力とする。まず、検索語を含む文を、オンラインゲームのチャットログ内から全て抽出する。検索語を含む文を全て抽出することにより、抽出の範囲を広げ抽出のパターンを拡張している。

得られた文の全てを、MeCab⁶⁾を用いて形態素解析を行う。形態素解析結果から、検索語の直前(PW)に出現する語を抽出し出現回数でソートする。PWが検索語と共起しやすく、出現回数が多いほど単語間の連結が強いと考えられるためである。同様にして、検索語の直後(NW)に出現する語を抽出し出現回数でソートする。

例えば、検索語が「ポタ」であるとき、「ポタ」を含む文を全て抽出する(表2)。表2で

表 3 「秘密」、「オガ」を含む発話の例
Table 3 An example of the utterance including "Secret" and "Oga"

新規/オガ/秘密/募集
コボ/秘密/PT/募集/中
鍵/さん/急募/オガ/秘密
新規/オガ/秘密/募集
ボタ/有り/2/連/オガ/秘密

は、PW に「秘密」、「連」が抽出され、NW に「出し」、「有り」、「持ち」、「オガ」、「あり」が抽出される。この際、PW・NW の語が記号、句読点、数字である場合、以降の処理で抽象度の高い語が抽出されてしまうため除外している。

出現頻度上位 M 位までの語について、それらの語を含む文をチャットログ内から全て抽出する。得られた文から、PW の直後に出現する単語 (NPW) を出現回数でソートし、出現頻度上位 N 位までの語を抽出する。同様に、NW の直前に出現する単語 (PNW) を出現回数でソートし、出現頻度上位 N 位までの語を抽出する。これは、検索語と PW、または検索語と NW に関連があり、かつ PW と NPW、NW と PNW に関連があるとき、検索語と PNW、または検索語と NPW に関連性が成り立つということを仮定している。

$$(\text{検索語}) R_1(PW) \cap (PW) R_2(NPW) \Rightarrow (\text{検索語}) R_3(NPW) \quad (1)$$

$$(\text{検索語}) R_4(NW) \cap (NW) R_5(PNW) \Rightarrow (\text{検索語}) R_6(PNW) \quad (2)$$

ここまでの過程で、NPW、PNW においてそれぞれ最大 $M \times N$ 語が抽出される。また、 $M \times N$ 語のうちの重複するものを重複度とする。重複度の高い語が、出現パターンが文法的に類似し関連性の高い語として関連語候補に抽出される。

表 3 では、「秘密」の NPW として「募集」、「PT」が抽出され、「オガ」の PNW として「急募」、「新規」、「連」が抽出される。

関連語候補で重複度の高い単語上位 5 位において、Web を用いて関連語の選定を行う。最終的に抽出された関連語候補が助詞・助動詞などの非自立語であった場合、それらは単独では意味を成さないため除外している。

チャットログ内に同一内容の発話が出現することがあるが、重複度の算出に出現頻度が影響するため、重複する内容であってもそれぞれの発話で単語の出現頻度を増加させている。

2.1.2 オンライン処理

検索語と関連語候補をクエリとして、Web 検索を行う。検索エンジンとして、Yahoo! JAPAN が API を提供している Yahoo! 検索 Web サービス⁹⁾ を用いた。

「検索語 and 関連語候補」をクエリとして生成する。Web の検索結果から得られるヒッ

ト件数を用いて、関連語を選定する。各クエリでヒット件数が最大となったときの検索語を関連語とする。そのときの Web 検索結果より得られる情報からスニペットを取得し保存する。

2.2 共起頻度情報を用いた関連語抽出

2.2.1 オフライン処理

ユーザが入力した検索語を含む文を、チャットログ内から全て抽出する。検索語を含む文を全て抽出することにより、抽出の範囲を広げ抽出のパターンを拡張している。

得られた文の全てを、MeCab を用いて形態素解析し、検索語と同文中に出現する単語の品詞情報を取得する。取得した品詞情報のうち、名詞、形容詞、形容動詞、副詞、動詞、未知語に該当する単語の出現頻度を求め、それら以外の品詞、非自立語は除外する。これは「が」や「は」などの助詞、記号などと多く共起しても、それら自体が単独で意味を成さず、また検索語以外の語とも共起しやすいため、適切な結果が得られないためである。取得した単語群ごとに、共起頻度でソートする。

2.2.2 オンライン処理

検索語とオフライン処理より得られた結果から、検索クエリを生成する。各品詞ごとの共起頻度の多い順から上位 10 語と検索語をそれぞれ組み合わせ、検索クエリを作成し Web 検索を行う。「検索語 and 共起頻度の多い語」をクエリとして生成する。それぞれのクエリにおいて、Web 検索結果より得られる情報からスニペットを取得し保存する。

2.3 異なる抽出手法の出力の組み合わせ

文内文脈情報を用いた関連語抽出と共起頻度情報を用いた関連語抽出の出力結果の組み合わせを行う。文内文脈情報を用いた手法は、関連性として単語間のつながりのパターンを重視し、共起頻度情報を用いた手法は、関連性として同一文内の共起パターンを重視する。文内文脈情報を用いて得られた全てのスニペットと、共起頻度情報を用いて得られた全てのスニペット各々に対して、trigram を利用し文の類似度を計算する。trigram は、検索対象を単語単位ではなく文字単位で分解し、後続の 2 文字を含めた状態で出現頻度を求める方法で用いた。

たとえば、文 A 「友人とオンラインゲームをプレイする」と文 B 「オンラインゲームを友人とプレイする」。この二つの文には、それぞれ 15 種類の trigram が含まれる (図 2)

文 A と文 B では語順が入れ替わっているが、trigram に分解して要素を比較すると、それぞれ 15 種類ある trigram のうち、11 種類が共通であることがわかる。このような手法で文同士の類似度を測ることで、もっとも類似している文を抽出することができる。

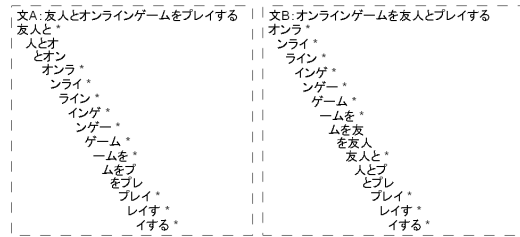


図 2 trigram の例
Fig. 2 The example of the trigram

それぞれの文の類似度の平均値を算出し、平均値を上回る類似度を有するスニペットのペアをユーザに提示する。

3. 評価実験

実験対象として本実験では、オンラインゲーム「RED STONE」⁸⁾のチャットログ(10,275 発話, 864KB)をシステム上に実装した。このチャットログの発話における品詞の内訳を表 4 に示す。形態素解析に MeCab^{*1}, 辞書に IPA 辞書^{*2}を用いている。

テストデータとしては、ブログ等から人手で抽出した RED STONE で使用される用語 323 語のうち、本実験で用いるチャットログ内に出現する 72 語を使用した。

前処理としてチャットログ内に含まれるユーザの発話時刻、ユーザ名の削除を行った。これは本実験では、ユーザの発話のみを用いるためである。

3.1 精度評価実験

ベースラインとして Yahoo! 検索⁷⁾, 他研究との比較として事典検索システム Cyclone¹⁰⁾との比較を行った。Cyclone は様々なページから言葉に関する説明を自動的に収集し、Web を事典のように使うためのシステムである。文章表現や HTML レイアウトに基づいて Web ページから用語説明箇所を抽出し、さらに複数の用語説明を分野や語義に基づいて分類することで組織化し、情報の質を高めている。

Yahoo 検索, Cyclone は検索結果が複数ページにわたることがあるが、本実験では検索結果の 1 ページに出現する情報のみを実験に用いている。

表 4 チャットログの発話の品詞内訳

Table 4 A part of speech of the utterance of the chat log

品詞	出現回数
名詞	66,264
未知語	54,888
記号	40,666
助詞	16,561
動詞	14,540
助動詞	11,180
接頭詞	2,115
形容詞	1,993
副詞	555
感動詞	354
接続詞	278
連体詞	43
フィラー	10

表 5 精度評価実験結果

Table 5 Experimental evaluation of accuracy

	正解数	不正解数	正解比率
Yahoo!Japan	6	66	8.3%
cyclone	7	65	9.7%
本システム	25	47	34.7%

3.2 実験結果および考察

実験結果を表 5 に示す。表 5 を見ると、本システムが他システムに比べオンラインゲーム用語に関する情報を、4 倍程度多く抽出できていることがわかる。テストデータの単語が一般的な既知語である場合があり、他システムでは一般的に用いられる語意を抽出してしまうことが多々見られたが、本システムではチャットログから関連性の高い語を抽出し検索クエリを自動的に生成することで、特定の分野における情報を抽出することが可能である。2 手法の出力結果のいづれにも適切な情報が含まれる場合、文類似度を用いて選定を行うことで、不適切な情報を淘汰することができる。また検索クエリに用いる単語数を増やすことで、より特定の分野に絞った情報を抽出することができると考えられる。

しかし関連語候補を抽出する際に、形態素解析の誤りにより適切なクエリが生成されず、Web から情報を抽出できないことがあった。

RED STONE のチャット発話を、MeCab を用いて形態素解析し正誤判定を行う予備実験

*1 <http://mecab.sourceforge.net/>

*2 <http://mecab.sourceforge.net/src>

表 6 予備実験：形態素解析の正誤判定

Table 6 A preliminary experiment: error rate of morphological analysis errata

	誤りではない	誤り	誤り率
被験者 A	143/200	57/200	28.5%
被験者 B	125/200	75/200	37.5%
被験者 C	133/200	67/200	33.5%
平均	134/200	66/200	33.2%

を行った。評価者は RED STONE のプレイ経験がある 3 名、テストデータとして 200 文を用いて人手で評価した (表 6)。その結果、30% 程度の形態素誤りが含まれることが確認された。

形態素解析の誤りは、検索クエリの生成だけでなく関連語候補を抽出する際にもノイズを含んでしまうため、精度を低下させる原因となる。これは、オンラインゲームにおける発話に話し言葉が多く含まれるためであると考えられる。また、ユーザの発話に単語を羅列した表現が多いことや漢字、アルファベット、記号や顔文字の混在した表現が見られることも形態素解析の誤りとして挙げられる。

システムの出力例を表 7、表 8 に示す。

表 7 で「Lv」とは「レベル」の略であり、ロールプレイングゲームなどにおいて、キャラクターやユニットの強さの段階を表す数値である「Lv」に関する抽出結果で「Lv (MMORPG 用語)「レベル」の略」というオンラインゲームに限定した語意を得ることができた。

表 8 で「秘密」とは「秘密ダンジョン」の略、もしくは「秘密ダンジョン探検パーティー」の略として用いられる用語である。オンラインゲームとは無関係の語意を抽出してしまい、誤りとしている。関連語候補の抽出で、適切な候補を抽出することができず、以降の処理においても適切な結果を得ることができなかった。秘密と関連性の高い語として「PT」の略である「パーティー」が関連語候補の上位に挙がっている。検索語「秘密」は名詞であり、「あり」は動詞であるので、検索語と関連語の品詞情報を一致させる等、関連語候補の抽出に品詞情報を用いることで精度の向上が見込めると考えられる。

3.3 印象評価実験

オンラインゲームをプレイしながらユーザに本システムを使用し、本システムの印象評価実験を行った。

まず、ユーザがオンラインゲームをプレイし、ゲームチャット発話の理解度を評価する。同時に本システムを使用しチャット内に出現する未知語の関連語検索を行う。ここでいう

表 7 検索語「Lv」の出力例

Table 7 The output example of search word "Lv"

関連語候補 (() 内は重複度) ・ PT (3) ・ レベル (3) ・ BISS (2) ・ 募集 (2) ・ レベ (2)
Web を用いた関連語の選定 ・ 検索語「LV」と関連の高い語は「レベル」です
Web より得られた詳細情報 ・ Lv (MMORPG 用語)「レベル」の略。 M ... 低レベルの秘密ダンジョン探検パーティーなどにおいてこの用語で募集がかかることがある。 ... (MMORPG 用語) レベルが高いキャラがパーティーに加わることによって、 ... ・ Lv (MMORPG 用語)「レベル」の略。 M ... もしくは「旧レッドアイ秘密ダンジョン探検パーティー」の略。 適正レベルは 55 ~ 65。 ... 低レベルの秘密ダンジョン探検パーティーなどにおいてこの用語で募集がかかることがある。

表 8 検索語「秘密」の出力例

Table 8 The output example of search word "Secret"

関連語候補 (() 内は重複度) ・ PT (3) ・ ひみ (3) ・ 宅 (2) ・ あり (2) ・ 火力 (2)
Web を用いた関連語の選定 ・ 検索語「秘密」と関連の高い語は「あり」です
Web より得られた詳細情報 ・ 【A】「営業秘密」は法律用語であり、義務や権利の行使が明確に定められている一方、「秘密情報」は契約で任意に規定できる ... 営業秘密、機密情報、秘密情報、守秘情報などいろいろな言い方がありますが、営業秘密は、不正競争防止法の中で使用される用語であり ... ・ そして、使徒の、エヴァンゲリオン、ゲンドウの秘密が次第に明らかになっていく。 聖書からの引用や哲学用語が多く用いられている、 登場人物の殆どが何らかの秘密を抱えている、 最初から無数の伏線が張られているといった特徴があり、そのため物語自体が非常に難解である。

表 9 印象評価実験結果
Table 9 An experimental result

	項目 A	項目 B	項目 C	項目 D	項目 E
ユーザ A	3	5	4	2	3
ユーザ B	2	3	5	3	5
ユーザ C	2	5	4	1	5
ユーザ D	2	4	4	3	4
ユーザ E	1	3	4	3	4
ユーザ F	2	4	4	3	5
ユーザ G	2	4	4	3	4
ユーザ H	1	3	4	2	5
ユーザ I	3	5	4	3	4
ユーザ J	3	4	4	2	4
ユーザ K	2	5	4	2	4
ユーザ L	1	3	3	1	4
ユーザ M	1	4	5	3	4
平均	1.9	4.0	4.1	2.4	4.3
	システム使用前 の用語の理解度	システム使用後 の用語の理解度	発話内容への理 解度の貢献度	処理時間	本システムを使 いたい

「未知語」とは、そのユーザにとって意味が特定できない、不明な語を示す。入力ユーザが任意に選択した単語で、複数回の試行を行っている。

印象評価のための被験者は、10代男性1名、10代女性2名、20代男性7名、20代女性2名、30代男性1名の計13名である。このうち、全てのユーザに2年以上のコンピュータ使用経験があり、5名のユーザにオンラインゲームのプレイ経験があった。

3.4 実験結果及び考察

アンケートによる5段階評価を行った。アンケートの評価項目は、システム使用前の用語の理解度(A)、システム使用後の用語の理解度(B)、発話内容の理解度への貢献度(C)、処理時間(D)、本システムを使いたいかどうか(E)である。各項目の結果を表9に示す。

システムの使用前後で、用語の理解度が平均2.1ポイント向上した。理解度の変化を図3に示す。また、発話内容の理解度への貢献度、本システムを使いたいかどうかの項目で、高い評価を得ることができた。このことはコミュニケーション支援システムとして本システムが有効であることを示していると考えられる。

処理時間の評価は高くない。チャットの発話は数秒毎に行われるため、リアルタイムに結果を表示するには、処理速度の向上が求められる。一方、一部の被験者からは「ゲームをしながらなら多少時間がかかっても気にならない」といった意見が得られた。

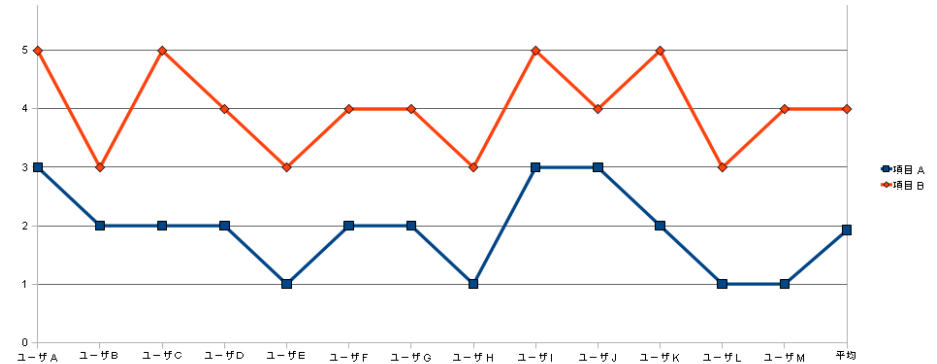


図 3 理解度の変化
Fig. 3 A change of the understanding degree

4. ま と め

本稿では、Web を利用した関連語抽出手法をオンラインゲームのチャットに応用することで、オンラインゲームにおけるユーザ間コミュニケーションの支援を可能とするシステムを提案し、精度評価実験と印象評価実験を行った。実験の結果から、実際に行われたチャットのログから手がかりとなる情報を抽出することにより、用語の限定された語意を含む情報を Web から抽出することが可能であることが確認された。またユーザの印象評価において、用語の理解度が平均2.1ポイント向上、本システムを使いたいという項目で平均4.3ポイントという高い評価を得ることができ、本システムの事典的なコミュニケーション支援システムとしての有効性が確認された。

今後は、さらなる精度の向上、ユーザに扱いやすいインタフェースの開発について検討を行っていく予定である。

参 考 文 献

- 1) 平井大祐, 葛西真記子,
”オンラインゲームの使用が使用者に与える心理的影響”,
ゲーム学会誌, Vol.1, No.1.
- 2) 津田耕平, Ruck Thawonmas, 今野優樹,
”語の影響度の計量による掲示板におけるキーワードの発見: テキストマイニングのオンラインゲームへの応用”,
知能と情報 (日本知能情報ファジィ学会誌) Vol.7, No.2, pp.184-193, 2005 .
- 3) Maya Ando, Satoshi Sekine, and Shun Ishizaki ,
Automatic Extraction of Hyponyms from Newspaper Using Lexicosyntactic Patterns ,
LREC 2004.
- 4) 佐藤理史, 佐々木靖弘,
ウェブを利用した関連用語の自動収集,
情報処理学会研究報告, 2003-NL-153-8, pp. 57-64, (2003)
- 5) 芳鐘冬樹, 井田正明, 野澤孝之, 宮崎和光, 喜多一,
言い換え表現抽出に基づく関連用語収集手法のシラバス検索システムへの応用,
名古屋大学附属図書館研究年報, No.3, pp.15-22 (2005).
- 6) 工藤拓,
MeCab: Yet Another Part-of-Speech and Morphological Analyzer ,
<http://mecab.sourceforge.net/>
- 7) Yahoo!検索,
<http://search.yahoo.co.jp/>
- 8) RED STONE ,
<http://www.redsonline.jp/>
- 9) Yahoo!デベロッパーネットワーク,
<http://developer.yahoo.co.jp/>
- 10) 藤井 敦, 石川 徹也,
World Wide Web を用いた事典知識情報の抽出と組織化,
論文誌 D-II, Vol.J85-D-II, No.2, pp.300-307, Feb. 2002.