Vol.2014-MBL-70 No.44
Vol.2014-UBI-41 No.44
2014/3/15

口コミデータを活用するデータベースシステムの実現

杉本 祐介[1,a)]　土井 千章[2]　中川 智尋[2]　太田 賢[2]　稲村 浩[2]　水野 忠則[3]　菱田 隆彰[3]

概要：近年，インターネット上では，Facebook や Twitter における日記や amazon.com や食べログにおける口コミなど，ユーザからの投稿を利用したサービスが数多く普及している．これらのサービスに寄せられる投稿の中には，楽しい，きれいといった感情や，関連性の高いものとの結びつきを示す特徴語が含まれており，解析によってそれらを明らかにすることで，有用なデータを導き出すことが可能であると考えられる．そこで本研究では，口コミに含まれる感情および特徴語に着目し，口コミに含まれる感情および特徴語の活用を行うデータベースシステムの開発を行う．

キーワード：口コミ, 感情, 特徴語

Implementation of the database system to applying user reviews

Sugimoto Yuusuke[1,a)]　Doi Chiaki[2]　Nakagawa Tomohiro[2]　Ohta Ken[2]　Inamura Hiroshi[2]
Mizuno Tadanori[3]　Hishida Takaaki[3]

1. はじめに

近年，インターネット上では，Facebook や Twitter における日記や amazon.com や食べログにおける口コミなど，ユーザからの投稿を利用したサービスが数多く普及している．これらのサービスが普及したことで，我々は，家族や友人，あるいは有名人などの近況や，売れ筋の商品や料理の美味しい飲食店などの情報を気軽に取得することができるようになった．これらのサービスの利用者数は，携帯電話やスマートフォンなどの普及に伴い，増加の一途を辿っており，ユーザから寄せられる投稿の量も膨大なものになっている．ユーザから寄せられる投稿の中には，楽しい，きれいといった感情や，関連性の高いものとの結びつきを示す特徴語が含まれており，解析によって導き出した感情や特徴語を基に，データの分類や傾向の分析などを行う研究が数多く行われている．

店頭，あるいはインターネット上で商品を購入する際，インターネット上のその商品のレビューが判断材料に使われることは多く，レビューの内容が商品の売り上げに影響することも珍しくない．大手サイトにおいては，人気商品の場合 1 つの商品に数十件のレビューがつくことも珍しくなく，色々な人の意見を参考に購入を検討することができる．だが，参考にするレビューが多くなれば多くなるほど，読まなければならない量が多くなり，全体の評価がかえって分かりにくくなることは少なくない．これを打開するため，一部のサイトでは評価の高いレビューを優先的に表示する，といったことをしているが，レビューの評価を行わないユーザが多いことや，参考にするレビューが減るといった理由で視点の偏りが起こることは否めないのが現状である．

観光業を収入源としている観光地において，いかに効率よく観光客を呼び込むか，ということは死活問題である．そのため，各観光地では，遊園地などのアミューズメントや大型ショッピングモールの誘致，その観光地の名物などを生かした町おこしなど，色々な手段を使って観光客の呼び込みを行っている．ところが，これらの手段を講じるに

[1] 愛知工業大学大学院 経営情報科学研究科
Graduate School of Business Administration and Computer Science, Aichi Institute of Technology
[2] NTT ドコモ 先進技術研究所
Research Laboratories, NTT DOCOMO, Inc.
[3] 愛知工業大学 情報科学部
Faculty of Information Science, Aichi Institute of Technology
[a)] yuusuke.sugimoto@gmail.com

は多くの投資が必要であり，また投資に見合った効果が出るとも限らないため，余裕のない観光地ではなかなか実行に踏み切れないのが現状である．

これらの現状を踏まえ我々は，インターネット上における口コミが持つ効力に着目し，口コミを活用するためのシステムの開発を進めている．それらの研究を基に，本稿では，口コミの内容と口コミに含まれる感情の関係性の調査，および口コミデータを活用するためのデータベースシステムの開発について述べる．本論文では，第 2 節では関連研究について，第 3 節では口コミの内容と口コミに含まれる感情の関係性の調査結果，第 4 節では今回開発したデータベースシステムについて，第 5 節ではまとめを述べる．

2. 関連研究

ユーザの投稿を解析することで特定の単語を導出する研究や，導出した単語を活用する研究は広く行われている．

徳久らの研究 [9] では，あらかじめ定義した感情表現を元に，Web テキストから感情生起要因（その感情表現を用いるのに至った要因）を収集し，収集した感情生起要因を用いてユーザの感情を推定している．徳久らは，構築した感情生起要因コーパスを用いて同コーパスから抽出したテストデータを評価し，高い精度の感情推定が実現されていることを確認した．

高村らの研究 [7] では，各単語を電子，各単語が持つ感情極性を電子のスピンの向きとみなすことで，エネルギー関数による感情推定を行っている．高村らは，WordNet[1], Penn TreeBank[2] の語釈文や表現を学習に利用し，General Inquirer[6] の語彙をテストデータとして評価を行い，高い精度の感情極性分類が行えることを確認した．

中山らの研究 [4] では，あらかじめ定義したシードを元に感情語やパターンを収集し，収集した感情語やパターンを新しいシードとすることで，より精度の高い辞書を構築する手法を提案している．

椎田らの研究 [5] では，口コミサイトに投稿された飲食店に対する口コミから特徴語を取得し，PLSI（Probabilistic Latent Semantic Indexing）を適用して飲食店および特徴語の同時分類を行っている．椎田らは，食べログに投稿された口コミに対して上記手法を適用することで，ジャンルや雰囲気，予算などの属性で飲食店を分類できていることを確認した．

これらの手法をインターネット上などから取得した口コミに適用することで，口コミに含まれる感情や特徴語などの情報を抽出することが可能である．今回は，これらの研究で用いられている解析手法を参考に，インターネット上の口コミの調査を行う．

表 1 感情語の一例

カテゴリ	単語数	一例
喜び	411	喜ぶ，嬉しい，楽しい，良い
悲しみ	231	悲しい，痛い，がっかり，残念
受容	141	許容，許す
嫌悪	192	嫌い，苦手，後悔，辛い
恐れ	399	恐ろしい，心配，不安，問題
怒り	67	怒る，文句，クレーム
驚き	271	驚く，感心，びっくり
期待	265	期待，希望，欲しい

3. 価格比較 Web サイトにおける口コミの調査

3.1 調査対象

価格比較 Web サイトである価格.com[11] および coneco.net[10] の 2 つの Web サイトに投稿された口コミを利用した．利用した口コミ数は，価格.com のものが約 4000 件，coneco.net のものが約 2000 件である．今回は口コミに含まれる情報のうち，タイトルと本文を調査対象とし，coneco.net ではその 2 つに加え，長所と短所について記述された部分についても調査対象とした．

3.2 調査方法

今回の調査は，以下に示すような方法で行った．

(1) ベースとなる感情語辞書を作成
(2) 感情語辞書に含まれる各単語について，日本語 WordNet[13] より類似した単語を取得，登録
(3) 各口コミサイトより口コミを取得
(4) 取得した口コミのうち，調査対象となる部分に MeCab[12] を利用して形態素解析を行う
(5) 形態素解析によって得られた単語の原形と感情語辞書の各単語のマッチングを行う

1 および 2 については第 3.3 項，3 については第 3.4 項，4 および 5 については第 3.5 項においてそれぞれ詳しく述べる．

3.3 感情語辞書の作成

感情語辞書を作成するにあたり，まず最初にベースとなる感情語辞書を手作業によって作成した．ベースとなる感情語の選定には [14] を参考にし，感情語の分類には Plutchik の感情の輪を参考にした．Plutchik の感情の輪は，感情を 8 つの基本感情（喜び，悲しみ，受容，嫌悪，恐れ，怒り，驚き，期待）に分類し，これら自身あるいはこれらの組み合わせによって感情を表現するモデルである．今回利用した感情語辞書ではそれに倣い，感情語を喜び，悲しみ，受容，嫌悪，恐れ，怒り，驚き，期待の 8 つのカテゴリに分類した．

ベースとなる感情語辞書を作成後，辞書に含まれる各単語について日本語 WordNet に問い合わせ，返ってきた単語を同じカテゴリに属する単語として追加することで，感情語の充実を図った．感情語の一例を表 1 に示す．

表 2 形態素解析の結果（一部）

表層形	品詞	原形
期待	名詞	期待
通	名詞	通
、	記号	、
とても	副詞	とても
楽しかっ	形容詞	楽しい
た	助動詞	た
です	助動詞	です
。	記号	。

表 3 調査結果

カテゴリ	出現数（価格）	出現数（coneco）
喜び	12425	7265
悲しみ	953	699
受容	217	213
嫌悪	1324	1015
恐れ	3209	2394
怒り	323	136
驚き	2941	1682
期待	1666	1122

3.4 口コミの取得

価格.com に投稿された口コミは，口コミの一覧表示を行うページから直接取得した．口コミの一覧表示は，1 ページあたり 15 件の口コミが表示される方式になっており，今回はページ番号をランダムに生成することで無作為に口コミを取得した．価格.com では，点数のみで内容を記入しないタイプの口コミも行えるようになっているため，内容が記入されていない口コミについては取得対象外とした．

coneco.net に投稿された口コミは，coneco.net に用意されている API を利用して取得した．coneco.net の口コミ検索 API に対し，全カテゴリの口コミの一覧を要求することで，口コミについての情報が 20 件ずつ取得できる方式になっているため，価格.com と同じく，ページ番号をランダムに生成することで無作為に口コミを取得した．

3.5 口コミの解析

口コミの解析は，MeCab を利用して形態素解析を行うことで文章を単語ごとに分割し，その結果求められた各単語の原形が感情語辞書に登録されている単語と一致した場合，対応するカテゴリの出現数を+1 する，という形で行った．ここで，形態素解析とは，文法や辞書などを元に，文を形態素（文字が意味を持つ最小の単位）に分解する技術のことであり，MeCab はそれに特化したオープンソース形態素解析エンジンである．

例文を使って解析時の動作の例を示す．例文には，“期待通り、とても楽しかったです。” という文を用いる．この文に対して形態素解析を行うと，表 2 に示すような結果を得ることができる．次に，ここで得た単語の原形が感情語辞書に登録されていないかを調べる．今回の例では，“期待” という単語が期待カテゴリに，“楽しい” という単語が喜びカテゴリにそれぞれ登録されているため，期待カテゴリと喜びカテゴリの出現数が 1 ずつプラスされることになる．

3.6 調査結果

先に述べたような調査を取得した口コミに対して行った結果，表 3 に示すような結果を得ることができた．喜びのカテゴリに属する単語の出現数が突出しているが，これは感情語辞書に登録された単語の出現率に差があるためであると考えられる．この表から分かる通り，インターネット上の口コミには多くの感情情報が含まれており，口コミの要約やカテゴライズ，感情を基にしたレコメンドなど，様々な用途への利用が可能であると考えられる．そのため，今回開発するシステムでは，感情情報を標準で扱えるような仕組みを組み込むことにした．

4. データベースシステムの実現

4.1 提案データベースシステムの概要

前節までにおける調査により，口コミに含まれる感情情報が有用であることが確認できた．これらの情報を利用することで，感情によるレビューの要約やカテゴライズなど，素のデータだけでは行えないような手法を適用することが可能になる．また，口コミには，感情情報以外に特徴語が含まれており，感情情報と同じように利用することが可能であると考えられる．

データベースシステムにおいてこれらの情報を利用するには，感情や特徴語の取り扱いを考慮した設計を行う必要がある．また，投稿された口コミからの各情報の抽出や，抽出した情報を利用した検索，といった機能を実装する必要があり，一般のデータベースシステムだけでこれら全ての要求を満たすのは困難である．そのため我々は，感情や特徴語の取り扱いに対応したデータベースシステムを開発することにし，その第 1 段階として，感情の取り扱いに対応したデータベースシステムの開発を行うことになった．

口コミを活用するためには，口コミを収集，集約，解析，分析するといった作業が必要である．本研究では，これらをまとめて行えるようなデータベースシステムとして，感情・地方特化型自律システム AHLE（Autonomous Human probes system with Local and Emotion functions）を開発した．AHLE は，以下に示す 4 つの要素から構成されている．

(1) 口コミの格納を行うライフログデータベース
(2) データの解析を行う解析エンジン
(3) ライフログデータベースとクライアント間の橋渡しを行うインターフェース
(4) 口コミの取得や表示を行うクライアント

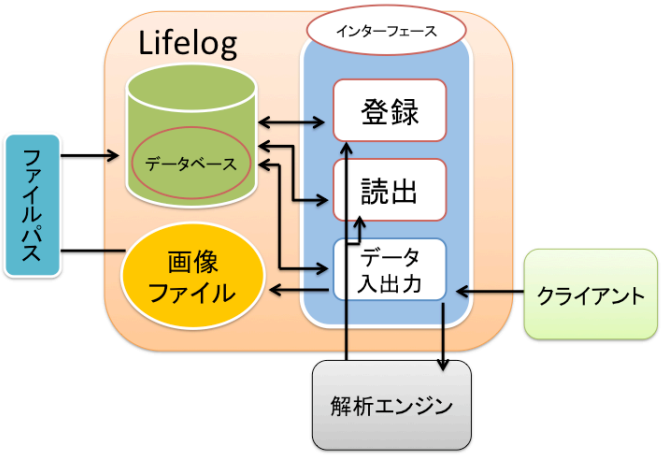

図 1 AHLE の概要図

AHLE の概要図を図 1 に示す．ここで図中において，口コミではなくライフログという表記がされているが，これは，AHLE が口コミだけではなく，ライフログの活用を見据えているためである．

4.2 データモデル

複数種類の口コミをデータベースに集約するためには，格納するデータ項目を決定する必要がある．本システムの開発においては，格納するデータ項目の決定に中村らの論文 [3] を参考にした．中村らは，ライフログに含まれる各データを 5W1H の観点から分類し，異なるライフログを 1 つのデータモデルに落とし込めるようにしている．

本研究では，このデータモデルをベースとして，いくつかの変更を行ったデータモデルを利用している．ライフログが特定のユーザに対するものである場合，中村モデルでは object として対象のユーザを取り扱っているが，対象をユーザに限定しないようにするため，提案モデルでは target として取り扱っている．

場所情報については，中村モデルでは location として取り扱っているが，このモデルではこの場所情報が何のものなのか，といったことは考慮されていない．そこで提案モデルでは，投稿を行った場所および対象の場所の 2 つに分類し，それぞれの場所情報を取り扱えるようにした．

ライフログの内容については，中村モデルでは content 内に元データをそのまま格納し，ref_schema 内で指定した外部スキーマを基に content 内のデータの解釈を行う形を取っている．中村モデルでは，このような実装形態を取ることで，取得元のデータを利用するようなサービスがそのまま流用できるようになっている．今回は，ライフログの解析を容易にするため，ライフログが持つ主要な項目を抽出して格納する形を取ることにし，外部スキーマの採用は見送ることにした．

また，前節における調査において，口コミに含まれる感情情報が特に有用であることが分かったため，内容部とは別に，各感情値を格納するためのフィールドを用意することにした．中村らが提案したデータモデルおよび今回利用したデータモデルの対比を表 4 に示す．

表 4 データモデルの対比

観点	中村モデル	提案モデル	備考
WHEN	date	date	日付
	time	time	時刻
WHO	user	user	ユーザ
	party	party	同行したユーザ
	object	——	対象のユーザ
	——	target	対象
WHERE	location	——	場所
	——	fromLocation	投稿を行った場所
	——	toLocation	対象の場所
HOW	application	application	アプリケーション
	device	device	利用したデバイス
WHAT	content	——	ログの内容
	ref_schema	——	外部スキーマ
	——	category	カテゴリ
	——	description	本文
	——	picture	画像
	——	evaluation	評価値
	——	emotion	感情値
WHY	——	——	(未採用)

4.3 ライフログデータベース

ユーザから投稿された口コミなどを格納する部分である．この部分では，前項において示したようなデータ項目を口コミなどから抽出した後，データベースへの格納を行っている．AHLE では，表 5 に示す項目を口コミから抽出し，データベースに格納している．

データモデルにおける user, party および target については，名前，性別，年齢，現住所および出身地の 5 つの情報が格納できるようになっている．

fromLocation および toLocation については，GPS などによる位置情報および別のテーブルにおいて定義した場所情報の 2 つの情報が格納できるようになっている．場所情報については，1 か所に複数の店舗などが密集している場合などに，どの店舗に対するライフログなのか，といった識別に利用する．

emotion については，前節における調査において感情を 8 つに分類したため，各極性の値をそれぞれ格納することで，感情ベクトルとして扱えるようにした．解析エンジンを用いてデータの導出を行う際に感情ベクトルを利用することで，ライフログの要約やカテゴライズ，ユーザの感情に合わせたレコメンドなどを可能にすることが期待できる．

表 5 ライフログデータベースに格納するデータ項目

項目名	備考
logNo	ライフログ番号
datetime	投稿日時
userName	投稿者名
userGender	投稿者の性別
userAge	投稿者の年齢
userPresentAddr	投稿者の現住所
userBirthAddr	投稿者の出身地
partyName	同行者名
partyGender	同行者の性別
partyAge	同行者の年齢
partyPresentAddr	同行者の現住所
partyBirthAddr	同行者の出身地
targetName	対象者名
targetGender	対象者の性別
targetAge	対象者の年齢
targetPresentAddr	対象者の現住所
targetBirthAddr	対象者の出身地
fromLocation	投稿時の位置情報
fromAccuracy	上記位置情報の精度
fromPlaceNo	投稿時の場所番号
fromPlaceName	投稿時の場所名
toLocation	対象の位置情報
toAccuracy	上記位置情報の精度
toPlaceNo	対象の場所番号
toPlaceName	対象の場所名
application	投稿に使用したアプリケーション名
device	投稿に使用したデバイス名
category	カテゴリ
description	本文
picturePath	投稿された画像の格納場所
overallEvalValue	評価値
joyValue	感情値（喜び）
trustValue	感情値（受容）
fearValue	感情値（恐れ）
surpriseValue	感情値（驚き）
sadnessValue	感情値（悲しみ）
disgustValue	感情値（嫌悪）
angerValue	感情値（怒り）
anticipationValue	感情値（期待）

4.4 解析エンジン

ライフログデータベースなどから取得した情報を元に，クライアントが必要としているデータを導き出す部分である．本研究ではまだ開発に着手できていないが，観光者から投稿されたライフログを元にしたおすすめスポットやおすすめルートの導出，ユーザの趣味嗜好に合わせたレコメンデーションなどの観光地推薦機能 [8] を実装予定である．なお，このエンジンは，インターフェースに内包される形，独立したエンジンとして複数のインターフェースから利用される形のどちらにおいても運用が可能である．

4.5 インターフェース

ライフログデータベースとクライアントの間に立ち，ライフログデータベースから取得したデータの加工，クライアントから投稿されたデータの整形などを行う部分である．AHLE では，外部からのライフログデータベースへのアクセスに対応するため，ライフログデータベースへの問い合わせを代理で行い結果を XML 形式で返す API を作成し，外部から利用できるようにしている．現在，AHLE で利用可能な API を表 6 に示す．

ライフログデータベースから取得したデータの詳細な解析を行いたい場合や，文字情報以外に投稿された画像などを取り扱いたい場合は，個別のインターフェースを作成し，それを利用することで，この部分の機能を拡張することが可能である．独自のインターフェースを作成する場合，AHLE で用意している API へのアクセスおよびクライアントからのアクセスができる形（クライアントに内包される形でも良い）であれば実装形態は不問である．

4.6 クライアント

実際にユーザからライフログを取得し，ユーザが求めている情報の表示を行う部分である．インターフェースを経由してデータの送信や取得が行える形であれば実装形態は問わないため，Web アプリケーションやスマートフォンアプリを始め，自由な形での実装が可能である．

5. まとめ

本研究では，口コミの活用を目的として，口コミに含まれる感情についての調査および口コミを活用するためのデータベースシステムの開発を行った．口コミからの感情情報の抽出には MeCab による形態素解析および感情語辞書とのマッチングを用い，その際に用いる感情語辞書には，Plutchik の感情の輪を参考に，感情を 8 つに分類したものを利用した．価格.com や coneco.net に投稿されているレビューに対して調査を行った結果，十分な量の感情情報が含まれていることを確認することができた．

口コミに含まれる感情情報や特徴語といった情報を利用することで，素のデータだけでは導出できないデータの導出が可能になると考えられる．そのため，本研究では，感情情報の取り扱いに対応したデータベースシステム，AHLE の開発を行った．AHLE では，標準で感情ベクトルを扱うことができるようになっており，今後，感情情報に対応した解析エンジンを実装することで，ライフログの要約やカテゴライズ，ユーザの感情に合わせたレコメンドなどが可能になると考えられる．

今後の課題としては，以下に示すような点が挙げられる．

- 特徴語に関する調査，特徴語への対応
- 感情や特徴語を利用した解析エンジンの実装

今後は，これらの問題点を改善しつつ，AHLE への機能追

表 6 利用可能な API

API 名，パラメータ名	備考
registerLifelog.php	XML 形式のデータを POST で送信することでライフログの登録を行う．登録に成功した場合はステータスコード 200 を，失敗した場合は 400 を返す．
getLifelogs.php	DB に登録されているライフログを XML 形式で取得する．
［userName］	指定されたユーザのライフログのみを返す．
［placeNo］	指定場所のライフログを返す．（複数指定可）
［notPlaceNo］	指定場所以外のライフログを返す．（複数指定可）
［category］	指定カテゴリのライフログを返す．（複数指定可）
［notCategory］	指定カテゴリ以外のライフログを返す．（複数指定可）
［order］	指定された列名を使用してソートを行う．
［isDESC］	（order の指定が必須，値不問）ソートを降順で行う．
［limit］	取得する件数の上限を指定する．
［skip］	（limit の指定が必須）先頭から無視する件数を指定する．
registerPlace.php	XML 形式のデータを POST で送信することで場所情報の登録を行う．登録に成功した場合はステータスコード 200 を，失敗した場合は 400 を返す．
getPlaces.php	DB に登録されている場所情報を XML 形式で取得する．
［placeName］	指定された名前の場所情報のみを返す．
［category］	指定カテゴリの場所情報のみを返す．（複数指定可）
［notCategory］	指定カテゴリ以外の場所情報のみを返す．（複数指定可）
［latitude］	
［longitude］	} （3 項目の指定が必須）指定された座標付近の場所情報のみを返す．
［range］	
［order］	指定された列名を使用してソートを行う．
［isDESC］	（order の指定が必須，値不問）ソートを降順で行う．
［limit］	取得する件数の上限を指定する．
［skip］	（limit の指定が必須）先頭から無視する件数を指定する．
executeSQL.php	SQL 文を直接実行する．基本的に他の API では間に合わない際の代用．

加や各種検証を行っていきたい．また，AHLE を用いたサービスの構築を行い，AHLE の実用性についても検討を進める予定である．

参考文献

[1] Fellbaum, C.: WordNet: An Electronic Lexical Database, *MIT Press* (1998).
[2] Mitchell P. Marcus, Beatrice Santorini, M. A. M.: Building a Large Annotated Corpus of English: The Penn Treebank, *Computational Linguistics*, Vol. 19, No. 2, pp. 313–330 (1993).
[3] 中村匡秀，下條　彰，井垣　宏：異なるライフログを集約するための標準データモデルの考察，電子情報通信学会技術研究報告， Vol. 109, No. 272, pp. 35–40 (2009).
[4] 中山記男，江口浩二，神門典子：感情表現の抽出手法に関する提案，電子情報通信学会技術研究報告， Vol. 104, No. 416, pp. 13–18 (2004).
[5] 椎田太輝，手塚太郎，木村文則，前田　亮：確率的潜在意味解析を用いた飲食店・特徴語同時分類結果の飲食店推薦システムへの応用，第 3 回データ工学と情報マネジメントに関するフォーラム（DEIM 2011） (2011).
[6] Stone, P. J.: The General Inquirer: A Computer Approach to Content Analysis, *MIT Press* (1966).
[7] 高村大也，乾　孝司，奥村　学：スピンモデルによる単語の感情極性抽出，情報処理学会論文誌， Vol. 47, No. 2, pp. 627–637 (2006).
[8] 樽井勇之：協調フィルタリングとコンテンツ分析を利用した観光地推薦手法の検討，上武大学経営情報学部紀要，No. 36, pp. 1–14 (2011).
[9] 徳久良子，乾健太郎，松本裕治：Web から獲得した感情生起要因コーパスに基づく感情推定，情報処理学会論文誌， Vol. 50, No. 4, pp. 1365–1374 (2009).
[10] オセニック株式会社：価格比較サイト [coneco.net] コネコネット，オセニック株式会社（オンライン），入手先 ⟨http://www.coneco.net/⟩（参照 2014-01-31）.
[11] 株式会社カカクコム：価格.com，株式会社カカクコム（オンライン），入手先 ⟨http://kakaku.com/⟩（参照 2014-01-31）.
[12] 工藤　拓：MeCab: Yet Another Part-of-Speech and Morphological Analyzer，京都大学 (online)，available from ⟨http://code.google.com/p/mecab/⟩ (accessed 2014-01-31).
[13] NICT 独立行政法人情報通信研究機構：日本語 WordNet，NICT 独立行政法人情報通信研究機構（オンライン），入手先 ⟨http://nlpwww.nict.go.jp/wn-ja/⟩（参照 2014-01-31）.
[14] 日本語表現インフォ制作委員会：感情表現インフォ：ピッタリの気持ちの描写が探せる言葉の辞書，日本語表現インフォ制作委員会（オンライン），入手先 ⟨http://hyogen.info/depa/kanjo⟩（参照 2014-01-31）.