

Wikipedia 記事中の意見文クラスタリング

井上 雄介^{†1} 太田 学^{†1}

本稿では、Wikipedia 記事中の意見文をクラスタリングするとともに、各意見の信頼性評価を試みる。我々は、意見文クラスタリングには独自のクラスタリング手法が必要だと考え、既存の文書クラスタリング手法を基に考案した意見クラスタリング手法を提案する。そして各意見の情報源などをもとに各意見の信頼性を分析する。実験では既存の文書クラスタリング手法との比較を行い、提案したクラスタリング手法の有用性を評価する。さらに、各意見の記述量、信頼性を用いた、記事の中立性評価への応用について検討する。

Opinion Clustering for Wikipedia Articles

YUSUKE INOUE^{†1} and MANABU OHTA^{†1}

In this paper, we propose opinion clustering for Wikipedia articles and evaluate the credibility of each opinion. Assuming that a unique method is necessary for opinion clustering, we propose such a method based on an existing document clustering method, and we analyze the credibility of each opinion based on its reliable sources. For experiment, we evaluated the utility of the proposed method by compared to the document clustering method. Moreover, we examine its applicability to evaluation of neutrality of Wikipedia articles using amount of description and the credibility of each opinion.

1. はじめに

オンライン百科事典 Wikipedia の記事の中には多様な意見が記述されている場合がある。Wikipedia 記事中の意見文をクラスタリングすることで、その記事で記述される対象に関する

る意見文を、分かりやすくユーザに提示できる。一方で、我々は意見文クラスタリングには独自のクラスタリング手法が必要であると考え。そこで、本稿では Wikipedia 記事中に存在する意見文をクラスタリングする手法を提案する。また、Wikipedia は誰でも編集可能であることから、記事によっては意見が偏っている可能性がある。このような場合、閲覧者に対して一方的な見解のみを示されることになる。Wikipedia で公開されている「Wikipedia : 中立的な観点」^{*1} というページでは、全ての記事は中立でなければならないとされている。そこで、我々はクラスタリングした意見の信頼性評価、中立性評価への応用の可能性についても検討する。

本稿の構成は以下の通りである。2 節で関連研究について述べ、3 節で提案する意見文クラスタリング手法、4 節で各意見の信頼性評価手法の概要を説明し、5 節で評価実験と考察、6 節でまとめについてそれぞれ述べる。

2. 関連研究

本研究の関連研究として、WISDOM、意見マイニング、Wikipedia 記事の信頼性評価を取り上げて説明する。

2.1 NICT 情報分析システム WISDOM

赤峯ら¹⁾ は Web 情報の信頼性検証のための情報分析システム WISDOM の開発を行っている。このシステムでは、キーワードに関する検索結果、さらに Web 情報についての発信者の種類と数、肯定的な表現と否定的な表現の割合などをユーザに提示することで、中立性も含めた情報の信頼性の検証を支援している。公開されているシステムでは、「電気自動車は環境に良い」や「裁判員制度」のように、分析したい内容を入力すると、検索結果、主要・対立文、意見・評価、原因・対策、レポートのタブが示される。意見・評価の中では、肯定・否定の分類に加えて要望・提言という分類も行っている。

このように WISDOM は Web ページ全般の情報を対象にするが、本研究は Wikipedia の記事を対象とする。また、本研究では、意見を肯定、否定という評価極性に基づく分類ではなく、意見が述べられている対象やテーマに基づいて分類する。

2.2 意見マイニング

飯田ら²⁾ は、Web 文書から意見を抽出し、それらをレーダーチャートの形式で要約、視覚

^{†1} 岡山大学大学院自然科学研究科
Graduate School of Natural Science and Technology, Okayama University

^{*1} <http://ja.wikipedia.org/wiki/Wikipedia:%E4%B8%AD%E7%AB%8B%E7%9A%84%E3%81%AA%E8%A6%B3%E7%82%B9>

化する意見抽出分類手法を提案した。飯田らは要約処理を以下の4つのサブタスクに分けて実行した。

- (1) 属性, 属性値, 主観評価となりうる表現の収集
- (2) テキスト中の属性と属性値 (もしくは主観評価) の対の抽出
- (3) 抽出した対が肯定的な表現であるか, 否定的な表現であるかの判定
- (4) (3) で判定した結果を用いた意見の要約

この手法ではまず, (1) で収集した表現から評価表現辞書を作成している。属性とは, ある対象のある側面を表す表現を指しており, 評価対象が「車」の場合は, 「インテリア」や「エンジン」がそれに当たる。属性値は属性の値を指す表現であり, 例えば, 属性「室内」に対し「きれい」や「広い」などがそれに当たる。主観評価は属性値に該当しない書き手の個人的な感想などを指す。例えば, 「気に入らない」, 「好きだ」, 「許せない」などの表現がある。次に, (2) で抽出した属性-属性値を評価表現辞書を用いて, 表現を〈車1, 内装, 高い〉や〈車2, デザイン, 平凡〉のように3つ組で構成して抽出する。文献2) では, (2) の段階までを実装し, 実験, 評価を行っている。

飯田らの研究は, Web 文書を対象としているが, 本研究では Wikipedia の記事を対象とする。また, 本研究では意見文のある種のテーマに従って分類することを目的とする。

2.3 Wikipedia 記事の信頼性評価

Adler ら³⁾⁻⁴⁾ や Hu ら⁵⁾ は Wikipedia の編集履歴を用いることで記事の信頼性を評価している。記事の残存率から著者の信頼性を求め, 記事の中で信頼性の高い著者の書いた部分ほど信頼性が高いとしている。Adler らはその手法を実装し, WikiTrustModule として公開している。さらに, 鈴木ら⁶⁾ は評価の対象とする著者を編集量の多い著者だけに限定した手法を提案しており, 計算量の削減と精度の向上を実現している。

本研究では, Wikipedia 記事の意見文クラスタリングを行った上で各意見の信頼性評価を試みる。またその信頼性評価は, 脚注などによって示される意見の出所などの情報源に基づいて行う。

3. Wikipedia 記事の意見文クラスタリング

本節では, Wikipedia 記事の意見文クラスタリングについて述べる。

3.1 文書クラスタリング

まず本研究で比較対象とする文書クラスタリングについて述べる。まず Wikipedia 記事から人手で抽出した各意見文に対して, 文ベクトルを生成する。初期状態では各意見文を全て

クラスタとみなして, 各クラスタの間の距離を計算する。距離の近いクラスタを併合し, クラスタ間の距離を再計算する。設定したクラスタ数に到達するまでクラスタの併合を繰り返すことでクラスタリングを行う。

以下で意見文のベクトル表現について説明し, 各意見文間の距離を定義する。そして, 本研究で用いたクラスタリング手法について説明する。

3.1.1 意見文ベクトル

抽出した意見文集合中の i 番目の意見文を以下のようなベクトルで表す。

$$\begin{pmatrix} w_i(t_1) \\ w_i(t_2) \\ \vdots \\ w_i(t_n) \end{pmatrix} \quad (1)$$

ただし, t_1, t_2, \dots, t_n は特徴語である。また, $w_i(t_k)$ は特徴語 t_k の重みで, 以下に示す TF-IDF により計算する。

TF-IDF による重みづけ

i 番目の意見文中の特徴語 t_k の重みを以下の式で定義する。

$$w_i(t_k) = tf_i(t_k) \times idf(t_k) \quad (2)$$

$$idf(t_k) = \log \frac{n_i}{df(t_k)} \quad (3)$$

ここで $tf_i(t_k)$ は意見文における特徴語 t_k の出現頻度, n は Yahoo! により検索可能なページの総数で, これを 250 億と見積もり, $df(t_k)$ はそのうち特徴語 t_k の出現する文書数, すなわち検索結果数である。

各意見文間の距離

2つの意見文 s_i と s_j の距離 $dis(s_i, s_j)$ を, ユークリッド2乗距離を用いて以下のように定義する。

$$dis(s_i, s_j) = \sum_{k=1}^n (w_i(t_k) - w_j(t_k))^2 \quad (4)$$

クラスタリング手法

本研究ではワード法を用いてクラスタリングする。その手順を以下に示す。

- (1) 各意見文 s_1, s_2, \dots, s_i を初期クラスタとする。

- (2) ユークリッド 2 乗距離を用いてクラスタ間距離を計算する。
 (3) 最小距離を与えるクラスタ A とクラスタ B を併合し、クラスタ C を生成する。
 (4) クラスタ間の距離を再計算する。
 • クラスタ C とクラスタ X の距離 R_{CX} は以下で計算される。
 • $R_{CX} = \alpha R_{XA} + \beta R_{XB} + \gamma R_{AB}$ ただし、 $|X|$ をクラスタ X の要素数として
 $\alpha = \frac{|X|+|A|}{|X|+|C|}, \beta = \frac{|X|+|B|}{|X|+|C|}, \gamma = \frac{|X|}{|X|+|C|}$
 (5) (3) へ戻る。

本研究では決められたクラスタ数でクラスタリングを停止する。

3.2 意見クラスタリング

本稿では 3.1 節で述べた文書クラスタリング手法に基づく意見文クラスタリングを提案する。本稿で提案する意見文クラスタリングでは、意見文を意見が述べられている対象やテーマに基づいて分類することを目的としている。そして、意見が述べられている対象やテーマは、クラスタリング対象の記事に関連の深い特徴語であると考えた。そのため意見文クラスタリングでは、クラスタリング対象の記事に関連の深い特徴語の重みを大きくすることで、関連の深い特徴語を含む文を同じクラスタに分類できるようにした。本研究では、WISDOM¹⁾*1 に記事のタイトルを入力し、関連キーワードとして挙げられた語を、関連の深い語として利用する。WISDOM では、分析したい内容を入力した際に、その内容に関する関連キーワードを提示する。「赤ちゃんポスト」に関連の深い語として得られた語を表 1 に示す。この中で、クラスタリングの対象となる、抽出した意見文中に含まれていた語を太字で示す。

4. 各意見に対する信頼性分析

3.1, 3.2 節のクラスタリングで得られた各意見に対して、脚注を用いて信頼性の分析を行う。Wikipedia 記事の脚注には書籍名、URL、新聞記事、補足説明など様々な情報が書き込まれている。そこで、脚注は種類によってその信頼性が異なると考え、本研究ではまずどの種類の情報であるかを判断し、その種類や情報源の信頼性によって意見の信頼性を判断する。

以下で本研究における脚注の分類について説明する。

- (1) 書籍名
 書籍名が挙げられている場合は数のみを意見の信頼性に反映する。本研究では、ISBN やページが示されている脚注を書籍名と判断する。

*1 <http://wisdom-nict.jp/>

表 1 「赤ちゃんポスト」の関連語
Table 1 Related words of “Baby hatch”

重要キーワード このとりのゆりかご:394 匿名:248 命:153
人・主体 蓮田太二理事長:24 病院の蓮田太二理事長:8 新生児:119 男児:42 3 歳児:12 乳児:47 子供:278 親:492 母親:30 里親:14 捨て子:36 養子:10 淑徳大教授:4 座長:10 蓮田晶一院長:8
組織・団体 厚生労働省:9 慈恵病院:483 熊本市の慈恵病院:166 熊本県の検証会議:4
場所 ドイツ:12 児童相談所:10 柏女霊峰:6
赤ちゃん 赤ちゃんの命:9 赤ちゃんポスト設置:19
その他 育児放棄:18 事情:45 報告書:32 置き去り:18 罪:24 戸籍:20 ケース:30 賛否両論:14 最終報告書:11

- (2) URL
 URL の場合は信頼するドメインとその他に分けて、信頼性への反映方法を変える。信頼するドメインの種類と例を以下に示す
- 政府系機関
 例) go.jp, gov
 - 教育機関
 例) ac.jp, edu
 - 新聞社
 例) asahi.com mainichi.jp
- (3) 新聞記事
 URL が付されていない場合、信頼する新聞社の新聞記事の場合は、全て信頼性が高い情報源であると考えられる。
- (4) その他
 その他の場合は、意見の信頼性判断の根拠としない。

文献 7) の結果から以下の式を用いて意見の信頼性のスコアを計算する。

$$R = 2 * N_1 + N_2 + N_3 \tag{5}$$

式 (5) において、

- N_1 : 参考文献, 書籍名の脚注の数
- N_2 : 信頼する URL, 新聞記事の脚注の数
- N_3 : 信頼性の不明な URL の数である。

5. 評価実験と考察

評価実験では複数の意見が書かれている記事に対して、3 節で説明した手法を用いてクラスタリングを行う。まず本研究で考える理想的なクラスタリング結果を示すために、記事を手手で分類した。次に、3.1 節の文書クラスタリング手法と 3.2 節の提案手法でクラスタリングする。さらに、4 節で述べた信頼性評価手法を用いて、クラスタリングの結果得られた各意見の信頼性を示す。実験には、「赤ちゃんポスト」と「代理母出産」の 2 つの記事を用いた。

5.1 クラスタリング

人手で分類した結果

本研究では、まず意見文を手手で分類する。意見が書かれている対象やテーマに基づいて分類した。実験に用いた「赤ちゃんポスト」の記事から抽出した意見文を、人手で分類した結果を表 2 に示し、以下で各クラスタについて説明する。

- Cluster1 には「新生児は早急かつ安全に保護されてしかるべき」や「児が生存できる」のように、子供の生存に関する意見文を分類した。
- Cluster2 には育児放棄や捨て子に関する意見文を分類した。ここで、「塩崎恭久官房長官は」の文には、2 つの意見が併記されており、本来 2 つに分けるべき文である。しかし、本研究では一つの意見として扱うため、先に記述されている意見により Cluster2 に分類した。
- Cluster3 には、赤ちゃんポストの運用に関する意見文を分類した。運用に関する意見文は一つだけだったので、一つの意見文で一つのクラスタとしている。
- Cluster4 には、「法的な裏付けがない」や「超法規的措置」のように、法的な観点からの意見文を分類した。
- Cluster5 には、赤ちゃんポストが匿名で新生児を置いていけるものであることに対する意見文を分類した。「子供の親を知る権利を侵害している」という意見文は、匿名であ

表 2 「赤ちゃんポスト」の記事に対して人手で分類した結果
Table 2 Result of manual classification of “Baby hatch”

クラスタ	意見文	信頼性
Cluster1	<ul style="list-style-type: none"> ● 設置に際しては、しばしば「捨て子」を容認するのといった議論にも発展するシステムではあるが、それ以上に「捨て子」が依然として存在している以上、それらの新生児は早急かつ安全に保護されてしかるべきだという議論もあり、道徳と人道の双方の観点からの議論がある。 ● 中絶では児は生存できないが、このシステムにより児が生存できるための選択肢が増える。 	
Cluster2	<ul style="list-style-type: none"> ● 新生児の殺害・虐待・育児放棄を防ぐ。 ● 育児放棄、捨て子を助長する。 ● 塩崎恭久官房長官は「法解釈以前に、子供を捨てない策を考えなければいけないのではないか」、高市早苗少子化担当相も「もう少し議論を深める必要がある」と、慎重な意見を述べた。 	1 1
Cluster3	<ul style="list-style-type: none"> ● 柳沢伯夫厚労相は「設置自体には違法性はないが、病院側の運用には慎重さが求められる」と発言。 	1
Cluster4	<ul style="list-style-type: none"> ● 預けられた子供たちは法的な裏付けがない実験的システムの犠牲者になりかねない。 ● 新生児の死体遺棄は年間数件であり、超法規的措置を取らざるを得ない必要性が感じられない。 	1
Cluster5	<ul style="list-style-type: none"> ● 子供の親を知る権利を侵害している。 ● 全国に乳児院や養護施設等があり、匿名でなければいけない理由が曖昧である。 ● 安倍晋三首相は 2 月 23 日、首相官邸での記者団の質問に対して、「子供を産むからには親として責任を持ってもらうことが大切で、そういうお子さんに対応する施設もあるし、匿名で子供を置いていけるものを作るには大変抵抗を感じる」と慎重論を唱えた。 	1 1
Cluster6	<ul style="list-style-type: none"> ● 児童養護施設に問題を丸投げしている（養護施設では大学や専門学校への進学が事実上断念され、また虐待の問題は恒常化している）。 	

るために発生する事象について述べられているので、このクラスタに分類した。

- Cluster6 には、本来親が負うべき責任を児童養護施設に負わせることになるという、責任の所在に関する意見文を分類した。このテーマに関する意見文は一つだけだったので、Cluster3 と同様に一つの意見文で一つのクラスタとしている。

文書クラスタリング

3.1 節で述べた文書クラスタリング手法を用いて、「赤ちゃんポスト」の記事から抽出した意見文に対して、クラスタリングを行った結果を表 3 に示す。クラスタ数は、人手で分類した結果に合わせて 6 とした。

意見文クラスタリング

3.2 節で説明した提案手法を用いて、意見文クラスタリングを行う。実験では、表 1 に示した「赤ちゃんポスト」の関連語の重みを 10 倍としている。「赤ちゃんポスト」の記事から

表3 「赤ちゃんポスト」の文書クラスタリング結果
Table 3 Result of document clustering of "Baby hatch"

クラスタ	意見文	信頼性
Cluster1	● 設置に際しては、しばしば「捨て子」を容認するのといった議論にも発展するシステムではあるが、それ以上に「捨て子」が依然として存在している以上、それらの新生児は早急かつ安全に保護されてしかるべきだという議論もあり、道徳と人道の双方の観点からの議論がある。	
Cluster2	● 新生児の殺害・虐待・育児放棄を防ぐ。 ● 育児放棄、捨て子を助長する。 ● 新生児の死体遺棄は年間数件であり、超法規的措置を取らざるを得ない必要性が感じられない。 ● 預けられた子供たちは法的な裏付けがない実験的システムの犠牲者になりかねない。	1
Cluster3	● 中絶では児は生存できないが、このシステムにより児が生存できるための選択肢が増える。	1
Cluster4	● 全国に乳児院や養護施設等があり、匿名でなければいけない理由が曖昧である。	
Cluster5	● 子供の親を知る権利を侵害している。	1
Cluster6	● 児童養護施設に問題を丸投げしている（養護施設では大学や専門学校への進学が事実上断念され、また虐待の問題は恒常化している）。 ● 柳沢伯夫厚労相は「設置自体には違法性はないが、病院側の運用には慎重さが求められる」と発言。 ● 安倍晋三首相は 2 月 23 日、首相官邸での記者団の質問に対して、「子供を産むからには親として責任を持ってもらうことが大切で、そういうお子さんに対応する施設もあるし、匿名で子供を置いていけるものを作るには大変抵抗を感じる」と慎重論を唱えた。 ● 塩崎恭久官房長官は「法解釈以前に、子供を捨てない策を考えなければいけないのではないか」、高市早苗少子化担当相も「もう少し議論を深める必要がある」と、慎重な意見を述べた。	1 1 1

抽出した意見文に対して、意見クラスタリングを行った結果を表4に示す。クラスタ数は人手で分類した結果に合わせて6とした。

考 察

まず、各手法でクラスタリングした結果を見て、人手で分類した各クラスタと同じテーマの意見文が最も多く分類されているクラスタと対応させる。よって表2、表3、表4のクラスタ1~6には、そのような対応関係がある。次に、表2と表3、そして表2と表4で各クラスタを比較して、すべての意見文の内、正しくクラスタリングできた文の割合を適合率とし、表5に示す。「赤ちゃんポスト」の記事についての実験結果では、文書クラスタリングの結果と比べ意見文クラスタリングの結果は、人手で行った分類結果に近いものとなった。これは、表4のCluster2において、関連語として得られた「育児放棄」の重みを大きくしたことで、2つの意見文のみを同じクラスタに分類できたことが理由の一つとして挙げられる。さ

表4 「赤ちゃんポスト」の意見クラスタリング結果
Table 4 Result of opinion clustering of "Baby hatch"

クラスタ	意見文	信頼性
Cluster1	● 設置に際しては、しばしば「捨て子」を容認するのといった議論にも発展するシステムではあるが、それ以上に「捨て子」が依然として存在している以上、それらの新生児は早急かつ安全に保護されてしかるべきだという議論もあり、道徳と人道の双方の観点からの議論がある。	
Cluster2	● 新生児の殺害・虐待・育児放棄を防ぐ。 ● 育児放棄、捨て子を助長する。	1
Cluster3	● 中絶では児は生存できないが、このシステムにより児が生存できるための選択肢が増える。 ● 児童養護施設に問題を丸投げしている（養護施設では大学や専門学校への進学が事実上断念され、また虐待の問題は恒常化している） ● 柳沢伯夫厚労相は「設置自体には違法性はないが、病院側の運用には慎重さが求められる」と発言。 ● 塩崎恭久官房長官は「法解釈以前に、子供を捨てない策を考えなければいけないのではないか」、高市早苗少子化担当相も「もう少し議論を深める必要がある」と、慎重な意見を述べた。	1 1
Cluster4	● 新生児の死体遺棄は年間数件であり、超法規的措置を取らざるを得ない必要性が感じられない。	
Cluster5	● 全国に乳児院や養護施設等があり、匿名でなければいけない理由が曖昧である。 ● 子供の親を知る権利を侵害している。 ● 安倍晋三首相は 2 月 23 日、首相官邸での記者団の質問に対して、「子供を産むからには親として責任を持ってもらうことが大切で、そういうお子さんに対応する施設もあるし、匿名で子供を置いていけるものを作るには大変抵抗を感じる」と慎重論を唱えた。	1 1
Cluster6	● 預けられた子供たちは法的な裏付けがない実験的システムの犠牲者になりかねない。	1

らに、Cluster5においても、関連語として得られた「匿名」という語の重みを大きくしたことで、「全国に乳児院や」という意見文と「安倍晋三首相は」という意見文を同じクラスタに分類できた。

一方、「代理母出産」の実験では、意見クラスタリングの結果は文書クラスタリングのそれと比べてほとんど変化がなかった。ここで、「赤ちゃんポスト」から抽出した意見文は一文当たりの特徴語の数が7.17個で、「代理母出産」では一文当たり14個の特徴語が存在した。さらに、一文当たりの関連語の数は、「赤ちゃんポスト」が1.41個であり、「代理母出産」が1.23個であった。このことから、「代理母出産」の場合は、関連語が各意見文に与える影響が、「赤ちゃんポスト」に比べて小さくなっていることが原因の一つと考えられる。

5.2 信頼性評価

4節で述べた信頼性評価手法を用いて、5.1節でクラスタリングした各意見文に信頼性の

表 5 適合率
Table 5 precision

	赤ちゃんポスト	代理母出産
文書クラスタリング	0.41	0.46
提案手法	0.67	0.46

スコアを付与した。「赤ちゃんポスト」の各意見文の情報源をもとに、式 (5) により算出したスコアを表 2, 表 3, 表 4 の右の列に示す。

考 察

実験では、各意見の情報源のみを用いて評価を行っている。各意見の信頼性が 1 になっているのは、各意見の情報源である脚注が 1 つずつしかなく、その内書籍を示す脚注が 1 つもなかったためである。情報源の種類による信頼性の差を示すことはできなかったが、情報源の有無は示すことができた。今後は、信頼性の中身をより具体的なものにするとともに、各意見に対する文書量を用いて、記事全体の中立性評価に利用することを検討する。

6. ま と め

本稿では、Wikipedia 記事中の意見文クラスタリングを提案した。本研究では、文書クラスタリング手法をもとに、独自の意見文クラスタリングの手法を提案した。実験では、人手で抽出した意見文に対して、提案手法を用いてクラスタリングを行い、文書クラスタリング手法の結果と比較した。さらに、クラスタリングした意見文に対して、どの程度その根拠となる参考情報が付与されているかについて述べた。実験結果より、「赤ちゃんポスト」の記事については、わずかながら効果を確認することができたが、「代理母出産」の記事については、文書クラスタリングと比べて効果を確認できなかった。今後の課題は、記事からの意見文の自動抽出を検討、実装することである。さらに、クラスタリング手法の改善や人手で行ったクラスタリング結果の妥当性の検証も必要である。また、現在考慮していない賛成・反対・中立といった極性の分類を各クラス内で行うことも考えている。さらに、各意見の文書量や信頼性を用いて、Wikipedia 記事全体の中立性評価についても研究を進めていく予定である。

参 考 文 献

1) 赤峯 亨, 宮森 恒, 加藤義清, 兼岩 憲, 角 薫, 乾健太郎, 黒橋禎夫: 情報の信頼性分析に向けた評価データおよびプロトタイプシステム WISDOM, 情報処理学会研

究報告 2007-NL-180(18) (2007).

- 2) 飯田 龍, 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一: 意見抽出を目的とした機械学習による属性-評価値対同定, 情報処理学会自然言語処理研究会 NL-165-4 (2005).
- 3) Adler, B. and de Alfaro, L.: A contentdriven reputation system for the Wikipedia, WWW '07: Proceedings of the 16th international conference on WorldWideWeb, pp.261-270 (2007).
- 4) Adler, B. T., Pye, I. and Raman, V.: Measuring author contributions to the Wikipedia, WikiSym '08: International Symposium on Wikis (2008).
- 5) Hu, M., Lim, E.-P., Sun, A., Lauw, H. W. and Vung, B.-Q.: Measuring article quality in Wikipedia: models and evaluation, *CIKM*, pp.243-252 (2007).
- 6) 鈴木 優, 吉川正俊: Wikipedia におけるキーパーソン抽出による信頼度算出精度および速度の改善, 人工知能学会第 21 回セマンティックウェブとオントロジー研究会 SIG-SWO-A901-1 (2009).
- 7) 井上雄介, 太田 学: 脚注と参考文献を用いた Wikipedia 記事の信頼性評価の一手法, 第 2 回データ工学と情報マネジメントに関するフォーラム (DEIM2010) B10-5 (2010).