DEIM Forum 2011 A4-3

特定トピックに関するブログ記事集合の観点分類におけるWikipediaの利用

横本　大輔[†1]　林　東權[†2]　牧田　健作[†2]　宇津呂武仁[†1]　河田　容英[†3]
福原　知宏[†4]　神門　典子[†5]　吉岡　真治[†6]　中川　裕志[†7]　清田　陽司[†7]

†1 筑波大学大学院システム情報工学研究科　〒305-8573 茨城県つくば市天王台 1-1-1
†2 筑波大学理工学群工学システム学類　〒305-8573 茨城県つくば市天王台 1-1-1
†3 (株) ナビックス　〒141-0031 東京都品川区西五反田 8-3-6
†4 独立行政法人 産業技術総合研究所 サービス工学研究センター　〒135-0064 東京都江東区青梅 2-3-26
†5 国立情報学研究所　〒101-8430 東京都千代田区一ツ橋 2-1-2
†6 北海道大学大学院 情報科学研究科　〒060-0808 北海道札幌市北区北 8 条西 5 丁目
†7 東京大学 情報基盤センター　〒113-0033 東京都文京区本郷 7-3-1

あらまし　本論文では，特定トピックに関して詳細な記述を含むブログ記事集合に対して，Wikipedia エントリを知識源として，特定トピックにおける観点ごとにブログ記事を分類する枠組みを提案する．この枠組みにおいては，Wikipedia 中において特定トピックのキーワードが出現するエントリを収集し，特定トピックにおける観点の候補とする．さらに，Wikipedia エントリ中の関連語の情報を利用して，ブログ記事を各観点に分類する．提案手法により，特定の検索クエリについて収集されたブログ記事における観点の分布を，素早く俯瞰することが容易になることを示す．
キーワード　ブログ分析，トピック，Wikipedia，観点分類，ファセット

Utilizing Wikipedia as a Knowledge Source in Categorizing Topic related Blogs into Facets

Daisuke YOKOMOTO[†1], Dongkwon LIM[†2], Kensaku MAKITA[†2], Takehito UTSURO[†1], Yasuhide KAWADA[†3], Tomohiro FUKUHARA[†4], Noriko KANDO[†5], Masaharu YOSHIOKA[†6], Hiroshi NAKAGAWA[†7], and Yoji KIYOTA[†7]

†1 Grad. Sch. of Systems and Information Engineering, University of Tsukuba,　Tsukuba, 305-8573, Japan
†2 College of Engineering Systems, School of Science and Engineering, University of Tsukuba, Tsukuba, 305-8573, Japan
†3 Navix Co., Ltd.　Tokyo 141-0031, Japan
†4 Center for Service Research, National Institute of Advanced Industrial Science and Technology, Tokyo, 135-0064, Japan
†5 National Institute of Informatics, Tokyo 101-8430, Japan
†6 Graduate School of Information Scienceand Technology, Hokkaido University, Sapporo, 060-0808, Japan
†7 Information Technology Center, University of Tokyo, Tokyo 113-0033, Japan

Key words　blog analysis, topic, Wikipedia, sub-topic categorization, facets

1. はじめに

近年，世界中でブログサービスやブログツールが普及し，各地域の人々がそれぞれインターネット上で個人の意見や評判を発信することが可能になった．それに伴い，様々な情報がブログに記載され，商用ブログ検索サービスを利用することでそれらの情報を取得することができるようになった．

しかし，特定のトピックについて検索を行った場合でも，そ

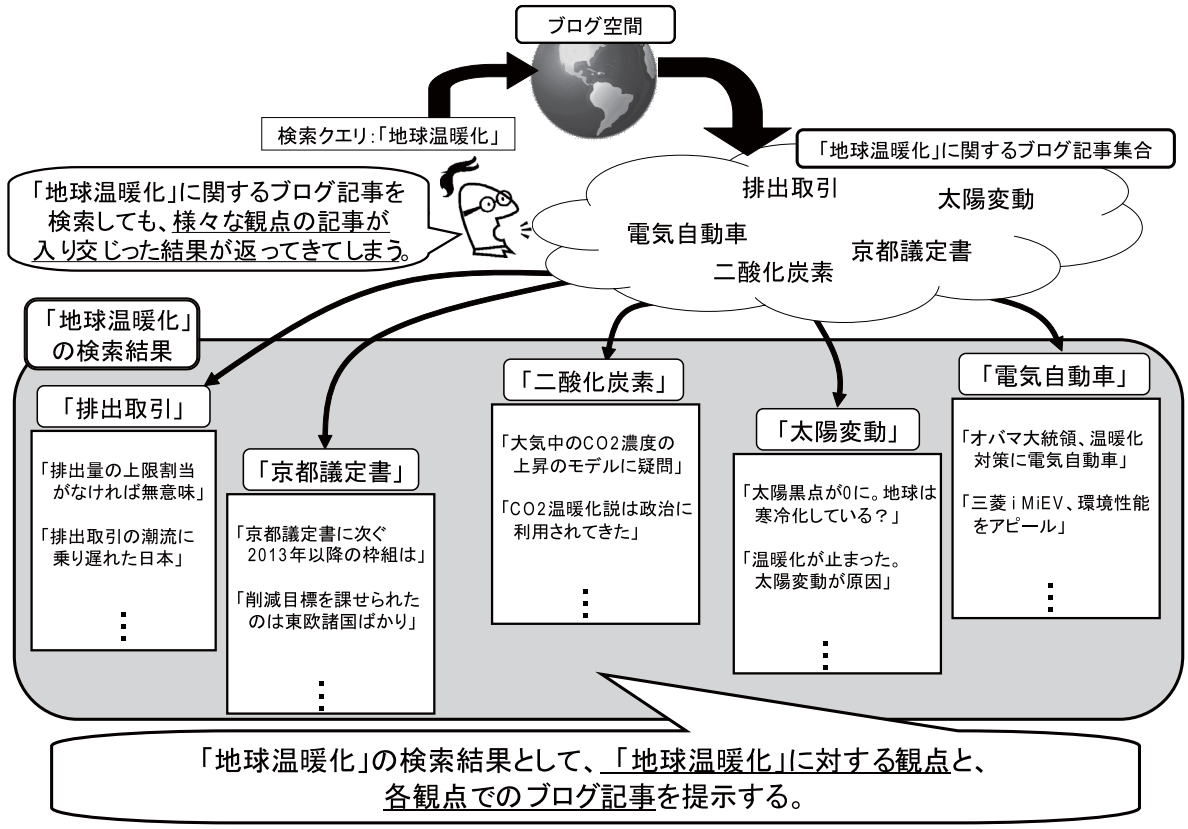

図 1　観点に基づくブログ記事集合の分類

の検索結果には様々な観点が混在している．例えば「地球温暖化」というトピックを検索クエリとして検索すると，温暖化の原因として二酸化炭素を話題にしているブログ記事や，温暖化対策の一つである排出取引について書いているブログ記事，地球温暖化の原因は二酸化炭素などではなく太陽活動の変化である，と述べているブログ記事など,「地球温暖化」について様々な観点で書かれたブログ記事が得られる．このように，検索結果には様々な観点が混在しているため，検索結果を単なるリストとして提示するだけでは，検索結果にどのような観点が含まれているのか知ることができない．そこで本論文では，この問題に対して，ファセット検索の考え方 [9] を導入する．

具体的には，本論文では，特定のトピックに関して収集されたブログ記事集合を，観点に基づいて分類する手法を提案する (図 1)．本論文では，Wikipedia を知識源とし，観点となる Wikipedia エントリと類似したブログ記事を収集することで分類を行うという方式を用いる．そして，提案手法の評価実験を行い，実際に，観点に密接に関連するブログ記事を選定した結果を示す．

以下に本論文の構成を述べる．2. では，本研究において分類の対象とするブログ記事集合の収集方法について述べ，3. では Wikipedia エントリとブログ記事の類似度について述べる．さらに，4. では特定のトピックに関して収集したブログ記事集合を対象として観点を付与する方式を提案し，5. で評価を行う．6. では関連研究と本研究の比較を行い，最後にまとめを行う．

2. 特定のトピックに関するブログ記事の収集

本研究においては，初期トピック t_0 に対して，関連するブログ記事集合を収集した結果に対して，観点の分類を行う．そこで，本節ではまず，初期トピック t_0 を含むブログ記事の収集方法を述べる．

初期トピック t_0 を含むブログ記事の収集においては，Yahoo!Japan 検索 API(注1)を利用し，初期トピック t_0 をクエリとして，日本語ブログホスト大手 8 社(注2)のドメインに限って検索を行った．検索の際には，複数のドメインを一度に指定して検索し，1,000 件の記事を取得する．次に，ブログ記事検索後，検索結果の URL をブログサイト単位にまとめる．その結果，一つの検索クエリあたり約 200 前後のブログサイトが取得される．次に，各ブログサイトをドメイン指定し，初期トピック t_0 を検索クエリとすることにより，各ブログサイト中において初期トピック t_0 を含むブログ記事を収集し，ブログ記事集合 $P(t_0)$ を作成する．

3. Wikipedia エントリとブログ記事の類似度

3.1 Wikipedia エントリの関連語 idf ベクトルの生成

3.1.1 Wikipedia 関連語

トピック名がタイトルである Wikipedia エントリ e を知識源として，トピック名に密接に関連する Wikipedia 関連語を収

(注1)：http://www.yahoo.co.jp/

(注2)：fc2.com, yahoo.co.jp, yaplog.jp, ameblo.jp, goo.ne.jp, livedoor.jp, Seesaa.net, hatena.ne.jp

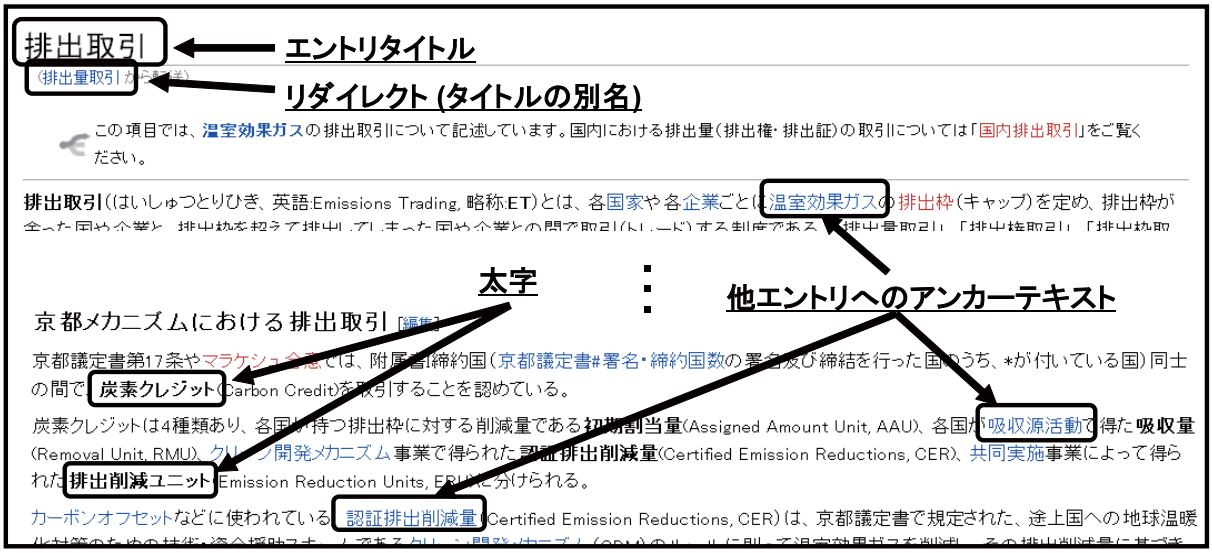

図 2　Wikipedia エントリおよび Wikipedia 関連語の例

集する．Wikipedia エントリ「排出取引」の場合について，エントリのスナップショットの抜粋，および，Wikipedia 関連語の例を図 2 に示す．この例の場合は，エントリタイトル「排出取引」の別名であるリダイレクト「排出量取引」で検索を行った結果，エントリタイトル「排出取引」の下に，リダイレクト「排出量取引」が表示され，エントリとしては「排出取引」の本文が提示されている．その他，エントリ中の太字「炭素クレジット」，および，他エントリへのリンクのアンカーテキスト「吸収源活動」,「カーボンオフセット」,「認証排出削減量」が提示されている．本稿においては，各エントリのリダイレクト，各エントリ本文中の太字，および，本文中における他エントリへのリンクのアンカーテキストを Wikipedia 関連語として収集する．Wikipedia エントリ e について収集された関連語集合を $R(e)$ とする．

3.1.2　Wikipedia エントリの関連語 idf ベクトル

Wikipedia エントリ e を表現するベクトルとして，収集されたそれぞれの関連語 $r \in R(e)$ を次元とし，値をその関連語の重み $w(r)$ とするベクトル $\vec{I}(e)$ を定義する．ここで，関連語 $r \in R(e)$ の重み $w(r)$ は，ブログ記事との類似度を測る際の関連語 r の重要度に基づいて設定する．例として,「排出取引」エントリの関連語としては,「排出枠」や「温室効果ガス」など,「排出取引」というエントリとの類似度を測る際に重要となる関連語が収集される．しかし一方で,「社会」,「日本」など,「排出取引」との関連が弱く，類似度を測る際に重要でない関連語も収集されてしまう．これらの重要度の違いを考慮するために，本稿では，逆文書頻度 (inverse document frequency, idf) を用いる．本研究では，Wikipedia の全エントリを文書集合 W として，関連語 r の逆文書頻度 $\mathrm{idf}(W,r)$ を定義する．

$$\mathrm{idf}(W,r) = \log \frac{|W|}{\left|\{e \in W | r \in R(e)\}\right|}$$

そして，エントリ e の関連語 r $(\in R(e))$ の重み $w(r)$ として，この逆文書頻度 $\mathrm{idf}(W,r)$ を用いる．

$$w(r) = \mathrm{idf}(W,r)$$

この重み $w(r)$ を用いて，Wikipedia エントリ e の関連語集合 $R(e)$ に対して，関連語 idf ベクトル $\vec{I}$ を以下のように定義する．

$$\vec{I}(e) = \left(w(r_1),\ \ldots,\ w(r_n) \right)$$

3.2　ブログ記事のターム頻度ベクトルの生成

Wikipedia エントリ e，および，2. で収集したブログ記事集合 $P(t_0)$ 中の各ブログ記事 p $(\in P(t_0))$ の組に対して，Wikipedia エントリ e の関連語 r_i $(\in R(e),\ i = 1, \ldots, n)$ を次元とする p のターム頻度ベクトル $\vec{G}(p,e)$ を次のように定義する．

$$\vec{G}(p,e) = \left(freq(p,r_1),\ \ldots,\ freq(p,r_n) \right)$$

ただし，$freq(p,r_i)$ は，Wikipedia エントリ e の関連語 r_i $(\in R(e),\ i = 1, \ldots, n)$ のブログ記事 p における出現頻度である．

3.3　Wikipedia エントリとブログ記事の類似度の定式化

Wikipedia エントリ e とブログ記事 p の類似度 $Sim(e,p)$ は，3.1 節で定義した Wikipedia エントリの関連語 idf ベクトル $\vec{I}$ と，3.2 節で定義したブログ記事のターム頻度ベクトル $\vec{G}(p,e)$ の内積として，次のように定義する．

$$Sim(e,p) \ = \ \vec{I}(e) \cdot \vec{G}(p,e) \ = \ \sum_{r \in R(e)} w(r) \times freq(p,r)$$

4.　特定トピックに関するブログ記事集合の観点への分類

4.1　観点候補の収集

初期トピック t_0 に対して，収集したブログ記事に付与する観点の集合 $F(t_0)$ を作成する．具体的には，まず，本文中に，初期トピック t_0 が出現する Wikipedia エントリを f_0 とする．次に，f_0 のうち，ブログ記事集合 $P(t_0)$ において，エントリタイトル $t(f_0)$ の文書頻度が 30 以上となるものを選定し，観点集合 $F(t_0)$ を構成する．

$$F(t_0) \ = \ \left\{ f \ \middle|\ \mathrm{df}(\ P(t_0),\ t(f)\ ) \ \geq 30 \right\} \qquad (1)$$

表 1　ブログ記事・観点組の評価結果 (その 1)

初期トピック	観点数	観点	正解率 (%)	正解と判定されたブログ記事・観点組数 / 評価対象のブログ記事・観点組数
喫煙	20	受動喫煙	87.5	7 / 8
		禁煙	100.0	5 / 5
		禁煙ファシズム	50.0	3 / 6
		その他	63.6	21 / 33
		合計	69.2	36 / 52
臓器移植	18	臓器の移植に関する法律	100.0	9 / 9
		免疫抑制剤	46.2	6 / 13
		宇和島徳洲会病院	100.0	4 / 4
		その他	68.4	26 / 38
		合計	68.8	45 / 64
地球温暖化	26	京都議定書	75.0	9 / 12
		再生可能エネルギー	62.5	5 / 8
		環境税	100.0	4 / 4
		その他	64.8	35 / 54
		合計	67.9	53 / 78
医療事故	14	医師	56.2	9 / 16
		医療訴訟	54.5	6 / 11
		日本医療機能評価機構	100.0	4 / 4
		その他	48.6	18 / 37
		合計	54.4	37 / 68
高齢化社会	12	少子化	80.0	8 / 10
		社会保障	77.8	7 / 9
		年金	71.4	7 / 5
		その他	17.1	6 / 35
		合計	47.5	28 / 59

4.2　ブログ記事への観点の付与手順

次に，特定トピックに関連するブログ記事集合中の各ブログ記事に対して，観点を付与する手順の詳細を以下で述べる．以下では，2. において，初期トピック t_0 を含むブログ記事を収集して作成した集合 $P(t_0)$ 中の各ブログ記事に対して観点を付与する．

まず観点を付与する際の条件として，観点 f を付与するブログ記事の候補は，$P(t_0)$ のうち f のタイトルが文字列として出現するブログ記事の集合 $P_f(t_0)$ に限定する．

次に，各観点 f ($\in F(t_0)$) に対して，集合 $P_f(t_0)$ 中のブログ記事 p のうち，p と f の間の類似度の上位 20 位までのブログ記事を収集する．

$$p_1, p_2, \ldots, p_i, \ldots, p_j, \ldots, p_{20}$$

(ただし $i < j$ ならば $Sim(f, p_i) \geqq Sim(f, p_j)$)

そして，各ブログ記事 p に対して，観点集合 $F(t_0)$ 中で類似度最大となる観点が f 自身である場合のみ，

$$f = \underset{f' \in F(t_0)}{\operatorname{argmax}} Sim(f', p)$$

ブログ記事 p に観点 f を付与することとする．

これにより，ブログ記事 p および付与された観点 f の組 $\langle p, f \rangle$ を作成し，評価の対象とする．

5.　評　　価

5.1　評 価 方 法

前節で観点が付与されたブログ記事 p に対して，ブログ記事・観点組 $\langle p, f \rangle$ の評価を行った．評価を行う際には，ブログ記事 p 中の記述内容が初期トピック t_0 および観点 f の双方に関連している場合に，ブログ記事・観点組 $\langle p, f \rangle$ は正解とし，初期トピック t_0 もしくは観点 f の少なくとも一方には関連していない場合には不正解とした．評価尺度としては，以下の正解率を用いた．

$$\text{正解率} = \frac{\text{正解と判定されたブログ記事・観点組数}}{\text{評価対象のブログ記事・観点組数}}$$

5.2　評 価 結 果

初期トピックとして,「喫煙」,「臓器移植」,「医療事故」,「高齢化社会」,「アルコール依存症」,「リストラ」」,「地球温暖化」,「スマートフォン」,「プリウス」の 9 トピックを対象として行った評価結果を表 1，表 2 に示す．また，正解と判定されたブログ記事・観点組の例を表 3 に，不正解と判定されたブログ記事・観点組の例を表 4 に，それぞれ示す．

表 3 に示すように、初期トピック「喫煙」，観点「受動喫煙」について，正解と判定されたブログ記事では,「受動喫煙症」や「健康増進法」など,「受動喫煙」と密接に関わる語が出現し，

表 2　ブログ記事・観点組の評価結果 (その 2)

初期トピック	観点数	観点	正解率 (%)	正解と判定されたブログ記事・観点組数 / 評価対象のブログ記事・観点組数
プリウス	17	電気自動車	22.2	4 / 18
		トヨタ自動車	35.3	6 / 17
		ハイブリッドカー	75.0	9 / 12
		その他	21.7	5 / 23
		合計	34.3	24 / 70
スマートフォン	37	Android	100.0	4 / 4
		W-ZERO3	66.7	2 / 3
		ウィルコム	44.4	4 / 9
		その他	26.6	17 / 64
		合計	33.8	27 / 80
アルコール依存症	15	飲酒運転	84.6	11 / 13
		精神疾患	16.7	2 / 12
		麻薬	8.3	1 / 12
		その他	28.1	9 / 32
		合計	33.3	23 / 69
リストラ	10	就職活動	20.0	3 / 15
		雇用	50.0	4 / 8
		退職勧奨	100.0	1 / 1
		その他	3.8	1 / 26
		合計	18.0	9 / 50

Wikipedia エントリ「受動喫煙」との類似度が高くなったため，正しい観点が付与されている．同様に，初期トピック「臓器移植」，観点「臓器の移植に関する法律」について，正解と判定されたブログ記事でも,「自民」や「A 案」,「D 案」など,「臓器の移植に関する法律」と密接に関わる語が出現している．

一方，ブログ記事へ観点を付与した結果における誤りは，表 4 のように分類できる．以下，この分類の一覧を示す．

(a)　ブログ記事が初期トピックと関連のある場合．

(a1)　人手でブログ記事に付与した参照用観点が，観点集合 $F(t_0)$ に含まれる．

(a2)　人手でブログ記事に付与した参照用観点は，観点集合 $F(t_0)$ には含まれないが，Wikipedia に存在

(a3)　人手でブログ記事に付与した参照用観点が Wikipedia に存在しない

(a4)　ブログ記事は，初期トピックに関連するが，特定の観点の付与は困難

(b)　ブログ記事が初期トピックと関連のない場合．

この場合，提案手法によりブログ記事に付与された観点が，ブログ記事に適合している度合いが大きい場合と小さい場合がある．

このうち，まず，表 4(a) に,「ブログ記事が初期トピックと関連のある場合」の例を示す．

「(a1) 参照用観点が観点集合 $F(t_0)$ に含まれる」場合の例では,「禁煙のために日常生活をどのように改善すればよいか」について書かれたブログ記事に対して,「ニコチン依存症」という観点が付与されていた．このブログ記事では,「ニコチン」,「喫煙」,「依存症」等，初期トピック「喫煙」に関する他の観点との間で共有される関連語が多く出現し，観点「ニコチン依存症」との類似度が高くなっていた．この問題に対する対策の一つとして，観点集合 $F(t_0)$ 中の各エントリの本文の集合を文書集合とみなして逆文書頻度を測定し，この値を各関連語の重みとする，という手法が考えられる．

「(a2) 参照用観点は観点集合 $F(t_0)$ には含まれないが，Wikipedia に存在」の例では，ブログ記事中に出現した Wikipedia エントリタイトルを観点候補とすることにより,「腎移植」を観点候補とすることが可能である．今後は，この方式の精緻化が必要である.「(a3) 参照用観点が Wikipedia に存在しない」の例では，ブログ記事中の文字列の中から，観点名として適切な用語を抽出する必要がある．類似の観点を共有するブログ記事が一定数以上存在する場合には，外部知識を用いず，主として，クラスタリング対象の文書集合の情報のみを用いる先行手法 [1,8] との併用が効果的であると考えられる．

「(a4) 初期トピックに関連するが，特定の観点の付与は困難」の例では，ブログ記事中に,「喫煙についてのブログ著者の意見」が書かれているが，特定の観点を付与することは困難であった．今後は，このような「特定の観点の付与が困難」なブログ記事の同定に特化した方式を導入する必要がある．

一方，表 4(b) に示す「ブログ記事が初期トピックと関連のない場合」の例では，いずれの場合も，ブログ記事の内容は，

表 3　ブログ記事・観点組の正解例

初期トピック	提案手法により付与された観点 = 人手で付与した参照用観点	ブログ記事の内容
喫煙	受動喫煙	受動喫煙等で化学物質過敏症になった人物. 受動喫煙防止に取り組んでほしい，と県知事にメールを送った
	禁煙	禁煙治療に関する社説に異議. 禁煙のつらさに対する理解が足りない，と主張
	禁煙ファシズム	ヒステリックにタバコ規制へ突き進む，このような社会状況を「禁煙ファシズム」として厳しく批判してきた A 氏へのインタビューを掲載
臓器移植	臓器の移植に関する法律	生命倫理会議が公表した「臓器移植法改定に関する緊急声明」への解説
	免疫抑制剤	腎移植と透析どちらが良いのか，という議論について. 腎移植をしても，免疫抑制剤を飲み続ける必要がある，と説明
	宇和島徳洲会病院	「宇和島徳洲会病院が 3 年ぶりに修復（病気）腎移植再開」という報道について，さまざまなメディアの報道内容を紹介

表 4　ブログ記事・観点組の不正解例

(a)　ブログ記事が初期トピックと関連のある場合

誤りの分類	初期トピック	人手で付与した 参照用観点	ブログ記事の内容	提案手法により 付与された観点
参照用観点が観点 集合 $F(t_0)$ に含まれる	喫煙	禁煙	『タバコ病辞典』の著者 加濃さんによる『禁煙のための日常生活改善ガイド』を紹介	ニコチン依存症
参照用観点は 観点集合 $F(t_0)$ には 含まれないが, Wikipedia に存在	臓器移植	腎移植	病気腎移植について，学会に説明を求める. 腎透析に苦しむ人にとって，病気腎移植は 大きな希望である	免疫抑制剤
参照用観点が Wikipedia に存在しない	スマートフォン	moTweets (アプリケーション名)	メニューも日本語，マルチアカウント対応, リストや会話のやり取りの一覧表示も可能と 今いち押しのアプリケーションです	Pocket PC
初期トピックに関連 するが，特定の 観点の付与は困難	喫煙	「初期トピック全般」	「喫煙は趣味」というにであれば，喫煙者が 自分らの責任で迷惑をかけずに吸えるような 環境を作るべき	受動喫煙

(b)　ブログ記事が初期トピックと関連のない場合

誤りの分類	初期トピック	人手で付与した 参照用観点	ブログ記事の内容	提案手法により 付与された観点
ブログ記事が初期 トピックと関連無	地球温暖化	「初期トピック と関連無」	敦賀市環境審議会の風力発電に関する答申がかたまったようです	風力発電 (ブログ記事との 適合性大)
	臓器移植		衆院議員石川知裕容疑者 (36)，政治資金規正法違反容疑で逮捕	鳩山由紀夫 (ブログ記事との 適合性小)

初期トピックと関連しない内容であった．ただし，初期トピックが「地球温暖化」の場合の例では，提案手法によって，観点集合 $F(t_0)$ 中の観点「風力発電」がブログ記事に付与されており，この観点は当該ブログ記事に最も適合する観点であった．一方，初期トピックが「臓器移植」の場合の例では，提案手法によって付与された観点「鳩山由起夫」は，観点集合 $F(t_0)$ 中では，ブログ記事の内容にやや近いと言えるが，当該ブログ記事にとってより適切な観点は,「地球温暖化」とは無関係な，より政治色の強い観点であった．しかし，観点集合 $F(t_0)$ 中には，そのような政治色の強い観点が含まれていなかったため，結果的に，観点「鳩山由起夫」が付与された．なお，これらのいずれの例においても，初期トピックとブログ記事との間の類似度や，初期トピックと観点との間の類似度に対して下限を設けることにより，観点付与の性能を改善できる可能性がある．今後の課題として，それらの方式に取り組む．

6. 関 連 研 究

本論文に関連して，TREC-2009 におけるブログ検索タスク [6] においては，ファセット検索によるブログサイト検索タスクが導入され,「意見の有無」,「個人的情報・公的情報の別」,「トピックについて専門的あるいは詳細な情報を含むか否か」の 3 種類のファセットをブログサイトに付与するタスクが行わ

れた．

文献 [3] は，Web ページの検索結果を分類し，各分類に対して適切な要約文を付与するという手法を提案している．この手法では，分類対象の Web ページの情報のみを利用してクラスタリングを行うため，データが十分に存在しない場合，まとまりのよい分類を行うことが難しくなる．これに対し，本研究の手法では分類対象の情報だけではなく，Wikipedia を知識源として利用しているため，分類対象が少ない場合でも分類を行うことができるという利点がある．

また，文献 [1,8] では，検索された個々の Web ページに対してラベルの付与を行い，付与されたラベルに基づいて分類を行う手法を提案している．これらの手法でも，ラベルを付与する対象のページの情報しか用いていない．これに対し，本研究の手法では，観点となる Wikipedia エントリのタイトルをラベルとしている．このように，ラベルの付与においても，付与対象の情報に加えて，Wikipedia の知識も用いることで，より容易にラベルを付与することができていると考えられる．

その他に観点に基づいて検索結果を提示する研究としては，トピック，ブロガー，リンク先，感想といった観点でブログを閲覧するもの [2] や，Wikipedia の検索に観点を利用するもの [4] などがある．また，本研究の発展として，文献 [7] においては，ブログ記事の時系列の分布，および，ブロガーの分布を考慮して，特定のトピックについて収集されたブログ記事集合における観点分布を提示する方式を提案している．文献 [5] では，韓国語のブログ記事を対象として本論文の手法を適用し，言語に依存せず，ブログ記事への観点付与が可能であることを示している．

7. おわりに

本論文では，特定トピックに関して詳細な記述を含むブログ記事集合に対して，Wikipedia エントリを知識源として，特定トピックにおける観点ごとにブログ記事を分類する枠組みを提案した．この枠組みにおいては，Wikipedia 中において特定トピックのキーワードが出現するエントリを収集し，特定トピックにおける観点の候補とした．さらに，Wikipedia エントリ中の関連語の情報を利用して，ブログ記事を各観点に分類した．提案手法の評価実験を行い，実際に，観点に密接に関連するブログ記事を選定した結果を示した．提案手法により，特定の検索クエリについて収集されたブログ記事における観点の分布を，素早く俯瞰することが容易になることを示した．

文　献

[1] 馬場康夫, 黒橋禎夫. キーワード蒸留型クラスタリングによる大規模ウェブ情報の俯瞰. 情報処理学会論文誌, Vol. 50, No. 4, pp. 1399–1409, 2009.

[2] 藤村考, 戸田浩之, 井上孝史, 廣嶋伸章, 片岡良治, 杉崎正之. マルチファセット型ブログ検索システム BLOGRANGER の開発. 電子情報通信学会技術研究報告, OIS2005-92, pp. 19–24, 2006.

[3] 原島純, 黒橋禎夫. PLSI を用いたウェブ検索結果の要約. 言語処理学会第 16 回年次大会論文集, pp. 118–121, 2010.

[4] C. Li, N. Yan, S. B. Roy, L. Lisham, and G. Das. Facetedpedia: Dynamic generation of query-dependent faceted interfaces for Wikipedia. In *Proc. 19th WWW*, pp. 651–660, 2010.

[5] D. Lim, D. Yokomoto, K. Makita, T. Utsuro, and T. Fukuhara. Utilizing Wikipedia as a knowledge source in categorizing topic related Korean blogs into facets. 言語処理学会第 17 回年次大会論文集, 2011.

[6] C. Macdonald, I. Ounis, and I. Soboroff. Overview of the TREC-2009 blog track. In *Proc. TREC-2009*, 2009.

[7] 牧田健作, 横本大輔, 宇津呂武仁, 福原知宏. トピックに関する話題の時系列分布に着目したブログ分析. 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 フォーラム— 論文集, 2011.

[8] 戸田浩之, 中渡瀬秀一, 片岡良治. 特徴的な固有表現を用いたラベル指向ナビゲーション手法の提案. 情報処理学会論文誌：データベース, Vol. 46, No. SIG 13(TOD 27), pp. 40–52, 2005.

[9] D. Tunkelang. *Faceted Search*. Synthesis Lectures on Information Concepts, Retrieval, and Services. Morgan & Claypool Publishers, 2009.