

Weblog 情報を融合したコンテンツ・ブラウジング

是津 耕司^{†,††} 日野洋一郎^{††} 中島 伸介[†] 門林理恵子[†] 呉 受妍[†]
林 正樹^{†††} 田中 克己^{††}

†† 京都大学大学院情報学研究科社会情報学専攻 〒606-8501 京都市左京区吉田本町

† 独立行政法人 情報通信研究機構 〒619-0289 京都府相楽郡精華町光台 3-5

††† NHK 放送技術研究所 〒157-8510 東京都世田谷区砧 1-10-11

E-mail: †{zetsu,snakajima,rieko,sooyeon}@nict.go.jp, ††{hino,tanaka}@dl.kuis.kyoto-u.ac.jp,
†††hayashi.m-hk@nhk.or.jp

あらまし Weblog (blog) は、今日の web コンテンツの中で“個人の関心”を最も良く示しているコンテンツであると考えられる。我々は、blog をコンテンツ・ブラウジングに取り込むことで、“human-aware”な新たなコンテンツ・ブラウジング体験 (contents browsing experience) をユーザに提供することを目指す。本論文では、blog を融合することでコンテンツ・ブラウジングがどう変わるのかについて、我々の取り組みを基に議論する。

Content Browsing Methods Integrating Weblog Information

Koji ZETTSU^{†,††}, Yoichiro HINO^{††}, Shinsuke NAKAJIMA[†], Rieko KADOBAYASHI[†], Sooyeon OH[†], Masaki HAYASHI^{†††}, and Katsumi TANAKA^{††}

†† Department of Social Informatics, Graduate School of Informatics, Kyoto University
Yoshida-Honmachi, Sakyo-ku, Kyoto, 606-8501 Japan

† National Institute of Information and Communications Technology
3-5 Hikaridai, Seikacho, Soraku-gun, Kyoto 619-0289 Japan

††† NHK Science & Technical Research Laboratories
1-10-11 Kinuta, Setagaya-ku Tokyo 157-8510, Japan

E-mail: †{zetsu,snakajima,rieko,sooyeon}@nict.go.jp, ††{hino,tanaka}@dl.kuis.kyoto-u.ac.jp,
†††hayashi.m-hk@nhk.or.jp

Abstract Weblog (blog) is a kind of web contents which represent people's interests straightforwardly. We propose content browsing methods integrating blog information, in order to provide “human-aware” content browsing experience. In this paper, we discuss how blog information changes current content browsing experience.

1. はじめに

近年、blog (weblog の略) と呼ばれる web コンテンツが、急速な勢いで増え続けている。ビジネス利用が増えてきた web ページに代わり、個人の意見やコメントを自由に公開できる手段として、blog が活用されている。Blog を支える基盤技術やコンテンツ形式は、従来の web ページと同じである。しかし、様々な blog 用ツールを介したユーザの情報利用形態が、従来の web ページとは大きく異なる。以下に、その特徴を示す：

今考えていることを、今公開する。

Blog 作成ツールを使ってテキストを記述し blog サイトに投稿すると、即座に web に公開される。面倒な HTML への変換や blog サイトへの転送などは、blog ツールが自動で行ってくれる。この手軽さが、“今ユーザが考えていることを、今公開する”という利用形態を生み出している。Blog コンテンツは、ユーザのアイデアを直接かつタイムリーに表す一方、情報が断片的で信頼性に欠けるという問題もある。

関心のある人にもみ伝える。

メーリングリストや BBS などは、自分の書いた情報に関心が有るか無いか分からない人にまで、一方的に情報を公開してし

まう。一方、blog の場合、RSS (Rich Site Summary) などに基づいて定期的に公開される (フィードされる) blog 記事を、関心を持つユーザだけが購読 (syndicate) する。読み手に強制しない情報伝達形態が、個人的意見や評判など、ユーザの個性が反映された情報の公開を促進している。

誰が関心を持っているかが分かる。

通常、Blog 記事には作者のユーザ ID が付けられており、誰がその記事を書いたのかが分かる。Web ページと同様に、ある blog 記事から他の blog 記事をリンクで参照することができ、“トラックバック” という仕組みにより、参照先の記事から参照元の記事を辿ることができる。これにより、ある記事がどんなユーザからどのように関心を持たれているのかを、記事の発信者側から確認することができる。

これらの特徴から、blog は、今日の web コンテンツの中で“個人の関心”を最も良く示しているコンテンツであると考えることができる。我々は、blog をコンテンツ・ブラウジングに取り込むことで、“human-aware”な新たなコンテンツ・ブラウジング体験 (contents browsing experience) をユーザに提供することを目指す。本論文では、blog を融合することでコンテンツ・ブラウジングがどう変わるのかについて、我々の取り組みを基に議論する。

2. ユーザアノテーション・メタデータの提示モデル

コンテンツのブラウジングでは、コンテンツ自身だけではなく、そのメタデータをいかに効果的に閲覧できるようにするのが重要である。コンテンツのメタデータには、コンテンツの書誌情報や内容記述情報など、コンテンツの提供者側が用意するメタデータと、それに対するユーザの反応やコメント、感想などを記述したユーザアノテーション・メタデータの 2 種類がある。あるコンテンツに関して書き込まれる blog 記事は、ユーザアノテーション・メタデータに分類される。提供者側メタデータを提示するだけでは、ユーザは、コンテンツに対して他のユーザがどのような関心を示しているのかが分からない。ユーザの観点で楽しみながらコンテンツを閲覧するためには、ユーザアノテーション・メタデータの提示が必要である。

ユーザアノテーション・メタデータの提示モデルは、以下のように分類することができる：

- (1) エージェント・ユーザ対峙型モデル
エージェント (キャラクターやロボット) が、ユーザに向かってメタデータを提示・説明する。
- (2) エージェント・ユーザ共視聴モデル
エージェントが、ユーザと同じ側に立って、一緒にメタデータを見る。例えば、ロボットを傍らに置き、一緒にテレビを見る形態がこれに該当する。
- (3) 演芸視聴モデル
コンテンツとメタデータを、2 人以上のエージェントが、

エージェント間の対話という形式で提示する。ユーザは、この対話による演出を視聴している。例えば、蓬萊ら [1] は、漫才形式による web ページの提示方法を提案している。

(4) 観客視聴モデル

エージェントは、コンテンツを見ている観客のように、コンテンツのほうを向いてメタデータを提示している (喋っている)。ユーザは、この観客エージェントを傍らで見ている。

本論文では、上記の観客視聴モデルに基づく blog 情報の提示方法を提案する。

3. Blog ユーザの関心の動向の把握

Web 上には膨大な量の多種多様なコンテンツが存在しており、その中からユーザにとって有用なコンテンツを探し出すのは至難の業である。キーワードによる web ページ検索では、僅かなキーワードでも大量の検索結果が返されてきてしまい、重要な情報にたどり着くことすらできない。こうした背景から、コンテンツの“社会的位置づけ”を表す情報を検索に取り入れる試みが、盛んに行われている。例えば、web ページ間のリンク構造に基づいて、多くのページからリンクされている (リンク支持率の高い) ページを重要とみなしたり [2]、あるページ集合の中でポータルの存在のページ (authority/hub) を発見したりすること [3] が行われている。

我々は、様々な blog ユーザのコンテンツに対する関心の動向を一覧性高く把握する方法として、BlogSafari と呼ばれるブラウジング・ツールを開発している。BlogSafari では、以下のことを目指している：

- 自分が関心のあるコンテンツに、他のユーザがどのような関心を示しているのかを把握し、そのコンテンツの社会的位置づけを把握する。
- あるコンテンツに関心を持つユーザの中でリーダー的な存在のユーザを発見し、そのユーザの記事から希少かつ重要な情報を得る。
- 同じコンテンツに関心を持つ他のユーザを発見し、コミュニケーションを図る。

3.1 BlogSafari の概要

図 1 に、BlogSafari の実行画面を示す。BlogSafari は、blog ユーザ (あるいは個人が所有する blog サイト) を動物に見立て、ある blog ユーザのあるコンテンツへの関心を、動物の捕食行動として表す。すなわち、あるコンテンツを動物の餌として表し、ある動物がその餌に近づくことで、blog ユーザのコンテンツへの関心を視覚的に表現する。blog ユーザのコンテンツに対する関心は、そのユーザが投稿した blog 記事にコンテンツへのリンク参照や記述 (例えばそのコンテンツの名前など) が含まれているかどうかを判断する。ユーザの特徴や関心の変化に応じて、動物の見た目や振る舞いを変化させる。

図 1 では、ブタ、ヒヨコ、ニワトリなど各々の動物が、ひとりひとりの blog ユーザを表している。また、“拉致”、“CVS”、

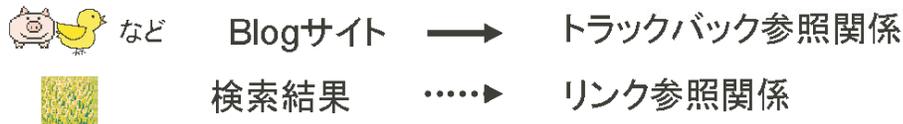
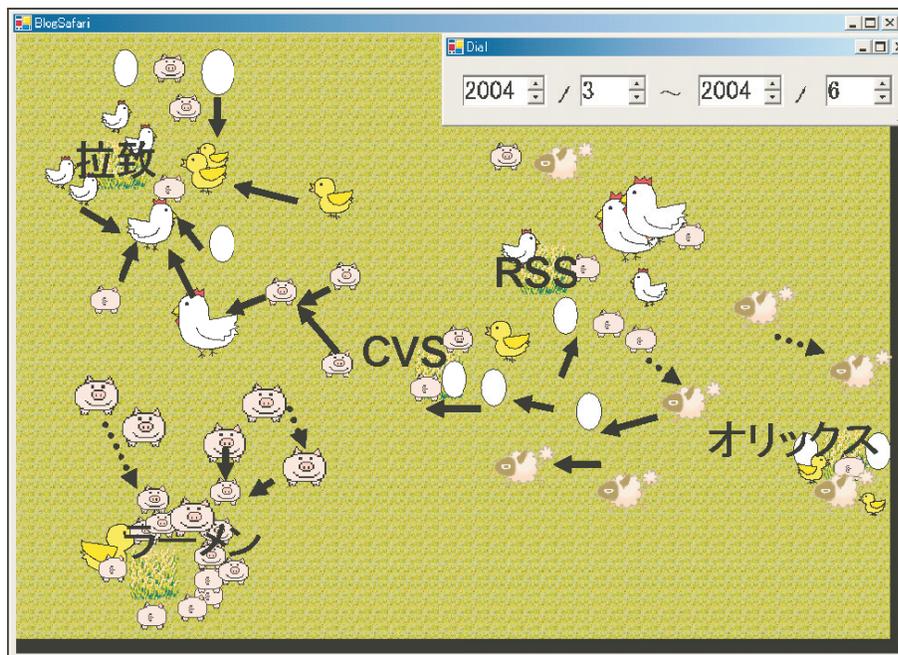


図1 BlogSafariの実行イメージ

“ラーメン”などの文字列で表されている餌（牧草）は、その文字列を質問として検索された web ページ集合を表している。例えば、“CVS”のそばにいる動物は、キーワード“CVS”で検索された web ページに関心を持っている blog ユーザ（web ページにリンクを張っているなど）を表している。餌は、検索結果だけでなく、個々の web ページや blog 記事を指定することも可能である。

BlogSafari を使って、以下のようなことが把握できるようになる：

どのコンテンツに皆に関心を持っているか？

図1において、“拉致”という餌の周りには動物がたくさん集まっている。このことは、“拉致”というキーワードで検索される web ページに対し、多くの blog ユーザが関心を示していることを表している。すなわち、web を通して、“拉致”に対する社会的な関心の高さが分かる。また、ユーザの特徴に応じて種類の動物を変えることで、あるコンテンツに対しどのようなユーザが関心を持っているかを把握することができる。

中心人物は誰か？

あるユーザの記事から他のユーザの記事へのリンクやトラックバックを矢印で示すことにより、誰が誰に関心を持っているかを把握することができる。多くの動物から参照されている動物は、そのユーザの記事が他のユーザから頻繁にリンクされていることから、議論の中心人物であると考えられる。こうしたユーザの記事は、重要な情報を含んでいると推測される。

希少な意見の発見



図2 エントリー投稿頻度と動物の大きさの関係

同じ種類の動物の群れの中に異なる種類の動物が混ざっていると、それらは同じコンテンツに関心を持つ例外的なユーザを表している。例えば、図1では、“ラーメン”に関するコンテンツの周りには、多数のブタに混じって一羽だけヒヨコが存在している。こうした例外的なユーザは、コンテンツに対する希少な意見を述べていることが多く、コンテンツを特徴付ける上では重要な情報であると考えられる。

3.2 動物メタファによる blog 情報の表現

3.2.1 動物の大きさによる投稿頻度の表現

あるコンテンツに対し blog ユーザがどの程度頻繁に記事を投稿しているかは、そのコンテンツに関心を持つユーザの中での存在感の大きさを表していると考えられる。そこで我々は、blog ユーザの投稿頻度を動物の大きさに反映させることを行う。図2に示すように、blog ユーザが頻繁に記事を投稿していれば動物も大きくなり、逆に投稿頻度が低くなると動物は痩せて小さくなっていく。直感的には、コンテンツに対応する“餌”をどれだけ懸命に食べているかということを表している。

3.2.2 動物の成長度合いによる信頼性の表現

動物が子供から大人へ成長する度合を用いて、blog ユーザの

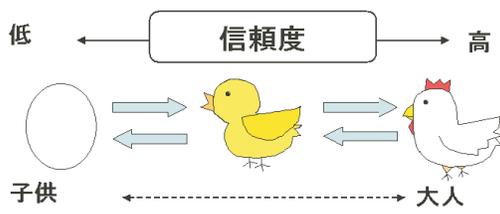


図3 信頼度と動物の成長度合いの関係

信頼性を表現する。図3に示すように、信頼度が高くなるほど動物は大人に成長し、信頼度が低くなるほど子供に戻っていく。あるblogユーザの信頼性は、そのユーザが投稿した記事に対する、他のblogユーザからのリンク参照数に基づいて評価する。すなわち、他のblogユーザから多く参照されているほど、信頼性が高いと判断する。

4. 文化遺産コンテンツに対するblog情報の提示

4.1 従来のブラウジング閲覧方法の問題点

近年、人々の美術館・博物館離れが深刻になってきている。学術的に価値の高い文化遺産をより多くの人々に見てもらおうと、近年、文化遺産のデジタル・コンテンツをオンラインで閲覧できるようにする試みが盛んに行われている。今日、こうしたサービスの多くは、アーカイブされたコンテンツとその解説を、一つ一つカタログ形式で閲覧できるようにしている。こうした閲覧方法は、調査や教育目的の利用に適している。

一方、実世界では、人々は必ずしも調査や教育だけを目的として美術館や博物館を訪れるわけではない。文化遺産に関する知識があまりない人でも、娯楽の一種として美術館や博物館を訪れることが多い。しかし、こうした人々にとって困るのは、どの作品をどのような観点で見たらよいかの分からない、陳列された作品を漠然と眺めるだけで終わってしまうことである。こうした人々は少なくなく、いかにして彼らに関心を持たせるかが大きな課題である。同様の問題が、オンライン上の文化遺産コンテンツにも当てはまる。

初心者ユーザを取り込むべく、仮想美術館などでは、従来からガイドエージェントなどによる閲覧の誘導が行われてきた[4]。しかし、ユーザの閲覧目的や知識レベルに応じて、様々な閲覧シナリオを、システム側で事前に用意しなければならないことが問題であった。また、エージェントが直接ユーザに話しかける提示方法では、ユーザ自身がエージェントを受け入れられないと、最後まで興味を持続して説明を聞けなくなるという問題がある。

その他にも、仮想美術館に同時にアクセスしてきた人どうしが、オンライン・チャットなどを通してコミュニケーションを行う方法が提案されてきた[5], [6]。コンテンツに対する意見交換をリアルタイムで行うことが可能であるが、反面、初対面のユーザとのコミュニケーションを遠慮したり、お互いの興味や知識レベル合わないと会話が成立しないなど、人的要因が障害となり十分活用されるに至っていない。

4.2 傍観者メタファに基づくblog情報の“視聴”

我々の基本的なアイデアは、ある文化遺産コンテンツに関

連する様々な話題をblogから抽出し、文化遺産に詳しくないユーザでも、そのblog情報からコンテンツに対する関心の糸口を発見できるようにすることである。あるblog記事は、その記事を書いたblogユーザ個人にとって関心のあることを表している。そのため、閲覧ユーザは、コンテンツに対するblog記事を、第三者の立場で傍観することになる。このことを考慮し、コンテンツに対するblog情報の提示方式として、傍観者メタファを提案する。傍観者メタファによるコンテンツ閲覧のイメージを、図4に示す。blog情報を提供するキャラクタ達は、ユーザに背を向けたままコンテンツの前で喋り続け、ユーザはそれを傍聴している。ユーザは、自分の興味のある内容を喋っているキャラクタを見つけると、その話に割り込んでキャラクタとインタラクションを始め、blogやコンテンツの情報を取得する。傍観者メタファの特徴は、従来のように、コンテンツに対する既定の説明をシステムから一方的に押し付けるのではなく、コンテンツに関する様々な話題を一覧しながら、コンテンツのある“見かた”をユーザ自身が自由度高く選択できるようにすることである。

傍観者メタファによるコンテンツ・ブラウジングの例として、以下のようなシナリオが考えられる：

- (1) あるユーザが、仮想美術館をウォークスルーしながら、様々な文化遺産のコンテンツを閲覧している。
- (2) それぞれのコンテンツの前にキャラクタ達が立って、何やら話している。あるキャラクタはコンテンツの前で独りで話しており、あるキャラクタは他のキャラクタとコンテンツの前で議論している。
- (3) あるコンテンツの前に差し掛かったとき、その前で2体のキャラクタが議論しているのを見つけた。近づいてみると、何やらタレントの名前を言っている。
- (4) 興味を持ったので、これらのキャラクタの会話に割り入り、「何について話しているのですか？」と尋ねてみる。
- (5) 一方のキャラクタがこちらを振り返り、「仏像の顔がタレントのXXに似ているという話です」と、会話の概要を答えてくれた。
- (6) 会話の内容に興味を持ったので、そのコンテンツについて更に詳しく調べてみることにした。

このシナリオの中で、blog情報は以下のように利用される：

- 2, 3では、展示されているコンテンツに関するblog記事を検索し、その記事を書いたblogユーザに対応するキャラクタをコンテンツの前に表示する。Blog記事を音声に変換し、キャラクタに喋らせる。もし、blog記事に他のblogユーザからコメントが付けられていれば、そのコメントを書いたblogユーザのキャラクタを横に表示してコメントの内容を喋らせる。
- 5では、2人のキャラクタが喋っているblog記事の内容を要約し、ユーザに返答する。

図5に、提案手法を実現するシステムの構成を示す。

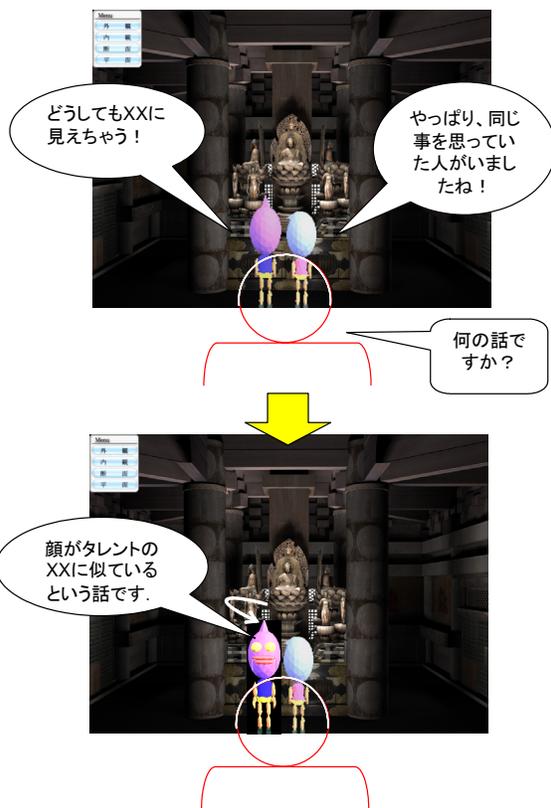


図 4 傍観者メタファによる文化遺産コンテンツ・ブラウジング

Blog 検索

コンテンツの属性情報を受け取り、そのコンテンツについて述べている blog 記事を検索する。また、それらに付与されたコメント記事やトラックバック記事も取得する。

キャラクター生成

検索結果の blog 記事集合から、登場するキャラクターの台詞や演出（喜怒哀楽やアクションなど）を記したキャラクター情報（スクリプト）を生成する。

文化遺産コンテンツ DB, 3D ウォークスルー生成

文化遺産コンテンツ DB に基づいて、3D のウォークスルー画面を生成する。

ブラウザ

3D ウォークスルー画面に、キャラクターによる blog 視聴の画面を合成して表示する。その際、ユーザのウォークスルーに同期して、閲覧中の文化遺産コンテンツの属性情報を Blog 記事検索モジュールに渡す。検索結果に対して生成されたキャラクター情報を受け取り、ウォークスルー画面に表示する。また、ユーザの割り込みに応答し、blog 記事を表示する。

技術的な課題を以下に挙げる：

- あるコンテンツについて述べている blog 記事の特定。リンクなど、blog の構造的情報だけでなく、あるコンテンツに関する一連の議論を示す blog 記事集合の特定など、内容情報まで考慮する必要がある。6 節で、さらに詳しく述べる。
- Blog 記事からのキャラクターの台詞や演出の生成。

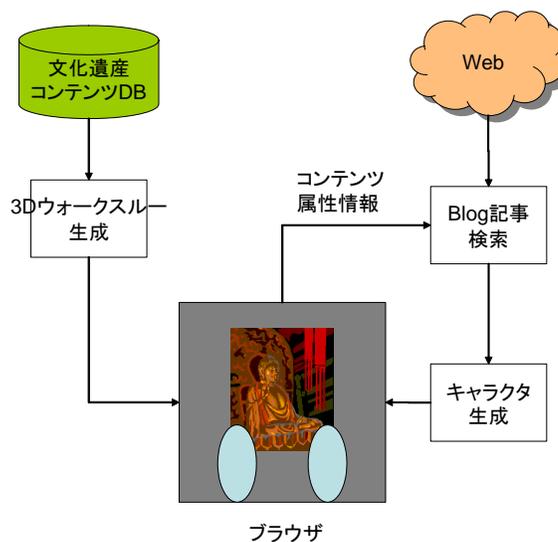


図 5 傍観者メタファによるコンテンツ・ブラウジングのシステム構成

例えば [7] では、web ページの内容をキャラクターの対話に変換し、TV のように受動的に視聴する方法を提案している。本研究では、複数の blog 記事の意味構造を考慮した変換が必要になる。

- キャラクターとのインタラクションによる blog 情報の提示。
- ユーザの割り込みに反応し、質問応答形式で blog 情報を提供する。
- 3D ウォークスルー画面とキャラクターによる blog 情報視聴画面の合成。

5. ブラウジングからの blog 生成

従来、blog 記事の生成は、blog 用の編集ツールを使って行われていた。あるコンテンツに関するコメントを書こうとすれば、blog 記事編集ツールを起動し、コンテンツを特定する情報（名前や URL など）とコメントを書いて、blog サイトに投稿する。汎用的なツールを使うため、ユーザは、コンテンツのブラウジングと blog 記事の編集を、別個の作業として行わなければならない。これに対し、コンテンツ・ブラウジングを blog 生成と一体化させ、コンテンツに対するユーザの関心をより直接的に記録することが考えられる。

例えば、web ページのブラウジングでは、ブックマークがコンテンツに対するユーザの関心を最も直接的に反映している。MyClip^(注1)では、web ページのブックマークに連動した blog の生成を行っている。ユーザが web ページをブックマークする際、そのページに対するコメントを一緒に記述すると、ページへのリンクを含んだ blog 記事が生成される。また、中島らのコンテキスト・ブックマーク [8] では、ユーザがある web ページをブックマークする際に、閲覧履歴を特徴付けるキーワードを一緒に保存する。Blog の生成は行っていないが、類似の技術を提供している。

(注1) : MyClip, <http://clip.myblog.jp/>

一方，4節で述べた文化遺産コンテンツのブラウジングでは，ウォークスルー中のユーザが体験したことを blog に記録することが考えられる．例えば，単にコンテンツに付与したコメントだけでなく，コンテンツの前で他のユーザと議論した内容や，ウォークスルー中に発言した内容なども，blog 生成の対象となりうる．ただし，あるコメントや発言が，何に対して述べられたものなのかを，ウォークスルーの文脈を考慮して判定しなければならない（例えば，コンテンツのある部分なのか，コンテンツ全体なのか，コンテンツのおかれた空間なのかなど）．

6. コンテンツに対する blog 情報の抽出

3節や4節で述べてきたコンテンツ・ブラウジングの実現において，あるコンテンツに関連する記述を，blog からどのように抽出してくるかが重要な課題である．

あるコンテンツをリンク参照している blog 記事は，そのコンテンツに対する関心を最も直接的に表している．Blog には，リンクが張られたことをリンク先に伝える“トラックバック”という仕組みがあり，これに基づいてリンク元の blog 記事を取得することが可能である．しかし，技術的には可能であっても，実際にコンテンツをリンク参照している blog 記事はそれほど多くない．また，blog 記事の内容は断片的であり，個々の blog 記事だけで十分な記述内容を得ることは難しいと考えられる．

これらのことを考慮し，あるコンテンツを参照する適切な blog 記事集合を特定する必要がある．これは，以下の2つのプロセスから成る：

- (1) コンテンツを参照する blog 記事を特定する．
- (2) コンテンツを参照する blog 記事を含み，かつ意味的まとまりを表す blog 記事集合を発見する．

1では，対象となるコンテンツをリンク参照する blog 記事（直接的な参照）だけでなく，コンテンツを識別する情報（名前など）を含む blog 記事（暗黙的な参照）も抽出する．一方，2では，直感的には，1で発見された blog 記事を含み，かつ一連の議論を構成する blog 記事集合を特定する．例えば，[9]のように，ニュースグループやメーリングリストなどのスレッドに相当する“blog スレッド”を特定する方法が考えられる．

7. ま と め

本論文では，blog 情報を融合したコンテンツ・ブラウジングについて，我々の取り組みを中心に説明した．様々なコンテンツに対する blog ユーザの関心の動向を可視化する方法として，BrogSafari を提案した．Blog から得られる様々なユーザの意見やコメントとともに文化遺産コンテンツを閲覧する方法として，傍観者メタファを提案した．コンテンツのブラウジングと blog 生成を一体化することについても述べた．また，コンテンツに対する blog 情報の抽出について考察を述べた．

謝 辞

本研究は，一部平成 16 年度科研費特定領域研究 (2) 「Web の意味構造に基づく新しい Web 検索サービス方式に関する研

究」(課題番号：16016247 代表：田中克己)，および 21 世紀 COE プログラム「知識社会基盤構築のための情報学拠点形成」による．ここに記し謝意を表します．

文 献

- [1] 蓬萊博哉，瀬本明代，田中克己：理解しやすさとユーモアを考慮した Web コンテンツの対話番組変換，日本データベース学会 Letters, Vol. 2, No. 2, pp. 29–32 (2003).
- [2] Brin, S. and Page, L.: The anatomy of a large-scale hypertextual Web search engine, *Proceedings of the 7th international conference on World Wide Web*, pp. 107–117 (1998).
- [3] Kleinberg, J. M.: Authoritative sources in a hyperlinked environment, *Journal of the ACM*, Vol. 46, No. 5, pp. 604–632 (1999).
- [4] Kadobayashi, R. and Mase, K.: Seamless Guidance by Personal Agent in Virtual Space Based on User Interaction in Real World, *Proceedings of The Third International Conference and Exhibition on The Practical Application of Intelligent Agents and Multi-Agent Technology (PAAM98)*, pp. 191–200 (1998).
- [5] Paolini, P., Barbieri, T., Loiudice, P., Alonzo, F. and Gaia, G.: Visiting a Museum Together: how to share a visit to a virtual world, *proceedings of Museums & Web '99* (1999).
- [6] 由良俊介，鶴坂智則，坂村健：デジタルミュージアムマルチメディア MUD のためのブラウザの設計と実装，情報処理学会論文誌，Vol. 40, No. 2, pp. 661–669.
- [7] Tanaka, K., Nadamoto, A., Kusahara, M., Hattori, T., Kondo, H. and Sumiya, K.: Back to the TV: Information Visualization Interfaces Based on TV-Program Metaphors, *Proceedings of IEEE International Conference on Multimedia and Expo 2000 (ICME2000)*, pp. 1229–1232 (2000).
- [8] 中島伸介，黒田慎介，田中克己：閲覧履歴を反映したコンテキスト依存型 Web ブックマーク，情報処理学会論文誌：データベース，Vol. 43, No. SIG5-4 (TOD14), pp. 23–26 (2002).
- [9] 中島伸介，館村純一，日野洋一郎，原良憲，田中克己：リンク構造の時間特性に着目した Weblog 解析に基づくコンテンツの信頼性評価の検討，日本データベース学会 Letters (掲載予定)，Vol. 3, No. 1, pp. 19–22 (2004).