BlogRadio: Blog情報の感情マイニングと可聴化に基づくWeb閲覧補完

郡　宏志†　　竹原　幹人††　　大島　裕明††　　小山　聡††　　田中　克己††

† 京都大学工学部情報学科
〒 606-8501 京都府京都市左京区吉田本町
†† 京都大学大学院情報学研究科 社会情報学専攻
〒 606-8501 京都府京都市左京区吉田本町
E-mail: †{kori,takehara,ohshima,oyama,tanaka}@dl.kuis.kyoto-u.ac.jp

あらまし　ユーザは，Web ページを閲覧している際に興味のある内容に対する評判や評価といった，他のユーザによる主観的な情報を同時に知ることは出来ない．一方で，近年，ユーザが Web 上で他のコンテンツに対する評価を容易に発信する手段として Blog が注目されており，Blog 記事の数は爆発的に増加し続けている． そこで，本稿では Web 閲覧中のユーザの興味のある内容を補完する情報を Blog から抽出し提示することにより，閲覧中のコンテンツの位置づけ・評判などの世の中の反応を理解しながら Web を閲覧するシステムを提案する．

キーワード　Blog，感情マイニング，可聴化，情報補完

BlogRadio: Complementing Web Browsing by Emotion Mining and Sonification of Weblog Information

Hiroshi KORI†, Mikihito TAKEHARA††, Hiroaki OHSHIMA††, Satoshi OYAMA††, and Katsumi TANAKA††

† Informatics of the faculty of Engineering, Kyoto University
Yoshida-honmachi, Sakyo-ku, Kyoto, 606-8501 Japan
†† Graduate School of Infomatics, Kyoto University
Yoshida-honmachi, Sakyo-ku, Kyoto, 606-8501 Japan
E-mail: †{kori,takehara,ohshima,oyama,tanaka}@dl.kuis.kyoto-u.ac.jp

Abstract　Users cannot learn subjective information, such as reputation and evaluation by other users, about the content in which they are interested while browsing web page. On the other hand, in recent years Weblog is becoming a popular way to publish evaluation on other contents in the Web easily, and the number of Weblog entries are increasing extremely. In this paper we propose a system in which users can browse web pages while learning reputation of these pages in the world, such as trust, reputation, and location of the contents browsing. From Weblog entries, our system extracts information that complements the content in which users are interested, and displays it to them.

Key words　Blog, emotion mining, sonification, information complement

1. はじめに

現在，インターネットというメディアを通して様々な情報が得られるようになった．従来までの Web から得られる情報というのは，比較的大規模な団体の公的な情報が主であった．しかし，ネットワークインフラの整備とインターネット接続サービスのコストの低下により，Web 上で，個人であっても容易に情報を発信することが可能となった．それに伴い，コンテンツに対する評判や評価といった，ユーザの主観に基づいて記述されたコンテンツが，Web 上で増加してきている．特に，最近は Blog という，書き手の独自の視点で，他のコンテンツに対して自分の考えなどのコメントをつけて Web 上で発信すると

いう新たな情報発信方法が広がりつつある．こういった性質を持つ Blog は，個人の主観的意見を最もよく反映した Web コンテンツであるため，ユーザアノテーション・メタデータに分類される．

従来の Web 閲覧環境では，ユーザは閲覧ページの内容に対する他のユーザの反応やコメント，感想などといった閲覧コンテンツのユーザアノテーション・メタデータを同時に知ることは出来ない．そこで，もし Web ページを閲覧しながら同時にその内容に関する Blog 記事を閲覧することが出来れば，ユーザは，閲覧コンテンツに対するユーザアノテーション・メタデータ，すなわち，閲覧コンテンツの評価や社会的位置づけなどを知ることが出来る．こうして，Web ページ閲覧による客観的な情報の閲覧と，Blog 記事視聴による主観的な情報の閲覧のシームレスな統合を実現することが可能となる．さらに，その主観的情報も，単に他ユーザの意見を寄せ集めるだけではなく，より一覧性高く整理された形でも見られることが望ましい．また一方で，文章データを音声データに変換する技術である音声合成も近年非常に精度が上がっており，人間の発声により近く，明瞭性の高い音声品質を実現している．

そこで，本論文では，ユーザが Web ページを閲覧している状況において，ユーザの興味を反映した内容に関して記述している Blog 記事の集合を整理し，さらに，各記事の内容を音声の形で可聴化し，提示する BlogRadio という Web 閲覧システムを提案する．本システムを使用することにより以下のような点が実現可能となる．

- 客観的情報と主観的情報のシームレスな閲覧
- 主観的情報を整理された形で一覧性高く概観

2. 基本的事項及び関連研究

2.1 基本的事項

2.1.1 Blog の概要

Blog とは，もともと Weblog と呼ばれ，アメリカにおいて 1999 年以降に急速に発達した Web コンテンツである．その急速に発達した理由として，Blog サイトを構築するためのツールが充実しているという理由が挙げられる．代表的なものとしては，MovableType [1] などがあり，現在では goo ブログ [2] などといった Blog のホスティングサービスもさかんである．こうして，ユーザは HTML 文書のソースを直接書くといった操作を通さずに手軽にコンテンツを作成することができる．

Blog コンテンツの内容は，社会的な問題などを扱ったものや，興味のあるニュース記事や Web サイトに独自の論評を加えるもの，個人的な日記やカメラつき携帯電話で撮った写真を載せたものなど，多岐にわたる．さらに，その内容は書き手が個人として自由に書くものであるため，非常に主観的なものであると言える．また，Blog サイトの形式としては，そのトップページに最も最近に書かれた「エントリ」と呼ばれる記事を複数表示するという形式をとる．通常は，Blog サイトの管理者のみがエントリを追加することができる．さらに，現在は RSS（RDF Site Summary または Rich Site Summary）と呼ばれる XML で記述された Blog サイトの要約を公開している場合が多く，本研究では RSS を有するものを Blog と扱うこととする．

2.1.2 Bulkfeeds

本システムでは，Blog の検索エンジンとして，Bulkfeeds [3] の全文検索を利用している．Bulkfeeds では，主に特定の Blog ポータルサイトの更新情報を元に検索対象となる RSS を登録しており，2005 年 1 月 9 日現在で約 58 万件の RSS が登録され、約 430 万件の記事がインデクスされている．本システムでは，Blog 感情辞書の構築及び，Web コンテンツの内容を補完する Blog エントリの取得においてこの検索エンジンを利用している．

2.2 関 連 研 究

2.2.1 可聴化に関する研究

まず，可聴化の研究の目的として最も多いのは，主に視覚的には認知しがたいデータのパターンや不規則性を明らかにするというものである．例えば，Maria Barra [4] らによる研究では，HTTP サーバの状態をモニタするのにユーザの好きな音楽にサーバのイベントによる音を混ぜるといった方法を提案している．この研究は，可聴化により情報の気付かせを狙ったものと位置づけられる．

他の目的では，視覚的に障害のあるユーザのために Web コンテンツを可聴化するといったものも存在する．例えば，Lori Stefano Petrucci [5] らは HTML ドキュメントをテキストだけで構成されたコンテンツに変換する方法を提案し，それらをテキストスピーチや点字により出力することにより，視覚的障害者の Web に対するアクセシビリティを向上させることを試みている．このように，以上の研究は，可聴化によりコンテンツの情報の絶対量は増加しない．

一方で，コンテンツにさらなる情報を可聴化により付加する方法も提案されている．金らは，ユーザが Web ページを閲覧している際に，未巡航 Web ページ群に対するメタデータを，アンカーをクリックする前に検索して提示するシステムを提案している [6]．その際の検索されたメタデータは可聴化され，効果音という形で表現される．本システムとは，情報の可聴化によりユーザの閲覧している Web ページに対して情報の絶対量を増加させているという点で類似している．しかし，この研究が可聴化により提供するのは，ユーザの未巡航ページに関する情報であるのに対し，本システムが提供するのは，閲覧ページを含むユーザの興味を反映した内容に関する情報である点が異なる．また，どちらもメタデータを提供するが，この研究の提供するのはコンテンツの内容記述情報といったコンテンツの提供者側が提供するメタデータであるのに対し，本システムが提供するのは，コンテンツに対するユーザの反応やコメント，感想などといったユーザアノテーション・メタデータである点も異なる．

2.2.2 POC

畦地らは，POC（Public Opinion Channel）[7] という，インターネット上の自動放送システムを提案している．このシステムでは，インターネット上のコミュニティから意見を収集し，それらを元に番組を作成し，コミュニティに送信する．POC

では，他ユーザの投稿文を収集し，ユーザに提示するというのが本システムと類似している点であるが，ユーザの得られる情報というのは POC というシステムに投稿された情報だけである．それに対して，本システムでは，ユーザは Web ページを閲覧しており，なおかつ Web 上の Blog から情報を収集可能である．従って，こちらの方がユーザが得られる情報の量は多いと考えられる．また，ユーザは他のユーザの意見を求める際に，システムに対して能動的に検索結果を要求する必要がない点も本システムの特長である．しかし，意見文を対話番組化する手法は，今後の研究において大いに参考になると考えられる．詳細は 7.1.1 節に譲る．

2.2.3 感情マイニングに関する研究

感情マイニングに関する研究としては，熊本らによる Web ニュース記事から喜怒哀楽を抽出するという研究がある [8]．この研究では，新聞データベースにおける「悲しい」,「うれしい」,「怒る」,「喜ぶ」という単語と，他の単語との共起度を元に Web ニュース記事を読んだ時にユーザに生じるであろう感情を推定する手法を提案している．本研究とは，対象が Blog と Web ニュースとで異なる点と，感情の評価方法が異なる．本研究では各感情ごとに独立した値を求めているのに対し，熊本らは「怒る⇔喜ぶ」「悲しい⇔うれしい」という軸上に記事を投影しているという点が異なる．本研究では，感情を表す単語とその他の単語との共起を用いるアプローチを参考とする．

3. BlogRadio

3.1 BlogRadio システム概要

3.1.1 システムの目的

本論文では，以下 2 点を実現するために，BlogRadio というシステムを提案する．

- Web ページ閲覧による客観的な情報の閲覧と，Blog 記事閲覧による主観的な情報の閲覧をシームレスに統合する
- 主観的情報を個々に閲覧するだけでなく，より整理された形でマクロ的視点により一覧性高く概観する．

また，Blog 記事を整理する際に用いる尺度としては，Blog 記事の書き手の感情を利用する．それは，他ユーザの反応をより汎用的な尺度で評価するためである．その為，各 Blog 記事の書き手の感情パラメータを求める．感情パラメータとは，Blog 記事に込められた書き手の感情の強さの度合いを各感情ごとに定量化したものである．Blog 記事の感情パラメータを求めることにより，ユーザはよりプリミティブなレベルで他ユーザの反応を知ることができる．

3.1.2 システム全体像

図 1 に BlogRadio のシステム全体像を示す．ユーザは Web ページを閲覧している．システムは，ユーザの Web ページ閲覧アクションに基づいて質問を生成して Blog 記事を検索し，その検索結果を取得する．こうして，ユーザの興味を判定し，それに関係する内容が記述されている Blog 記事を得る．得られた Blog 記事から，あらかじめ作成しておいた Blog 感情辞書を用いて各 Blog 記事の感情パラメータを求める．そして，得られた感情パラメータに基づいて Blog 記事を各感情に分類し，整

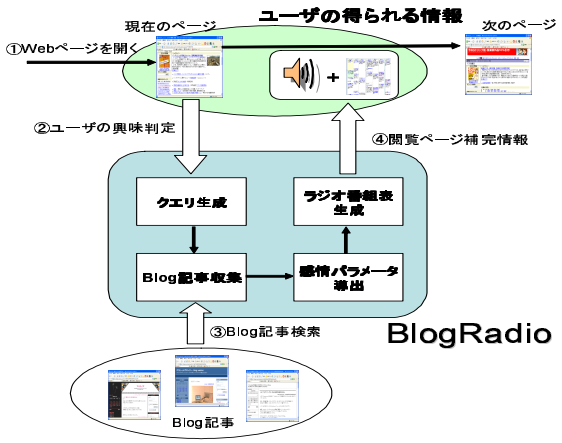

図 1 BlogRadio システム全体像

理した結果をラジオ番組表メタファを用いてユーザに提示する．また，Blog 記事の内容というのは非常に主観的であるので，その内容の幅も非常に広いものとなっている．したがって，各記事を気に入るかどうかはユーザに強く依存するため，ユーザは視聴する記事を自由に選択するというスタイルをとる．よって，ユーザはシステムとの簡単なインタラクションにより，視聴する Blog 記事を選択可能であるとする．こうして，ユーザは実際にチャンネルを選択してラジオを聴くような感覚で，ラジオの番組表を用いて視聴する Blog コンテンツを自由に選択することができる．

こうして，ユーザは以下の 3 つの情報を得ることが可能となる．

- Web ページ
  ユーザは，Web ページを閲覧している．本システムは，ユーザが Web ページを閲覧することによりニュース等といった，ある内容に関する客観的な情報を取得しているという状況を想定している．
- ラジオ番組表メタファによる Blog 記事の検索結果
  ユーザは，Blog 記事の検索結果を記事に込められている書き手の感情，そして記事の書かれた時刻という 2 つの次元により整理された形で取得することが可能である．
- 音声による Blog 記事
  ユーザは，Blog 記事の内容を，音声という形で視聴することが出来る．

3.1.3 システム使用例

本システムにおけるコンテンツ・ブラウジングの例としては以下のようなシナリオを想定している．

（1） ユーザは本システムを利用してニュースサイトなどといった Web ページを閲覧している．Web ページのあるリンクアンカーをクリックし，新たなページを開く．

（2） システムは，新たに Blog 記事を検索し，それらを分類する．ユーザはシステムにより提示された Blog 記事の感情分類の結果を見る．そこで，まずどういった感情をもったユーザが多いか，閲覧している内容がいつ話題になったか，そして，時間軸に沿って他ユーザの感情がど

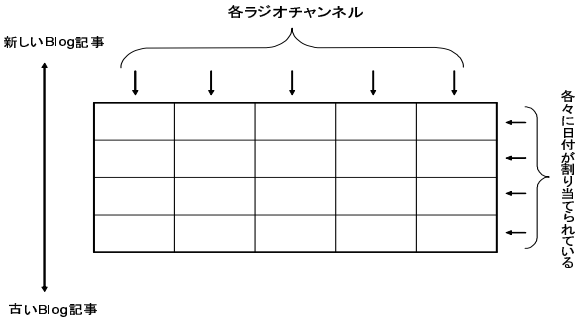

図 2　ラジオ番組表メタファのモデル

のように変化していったか，等を理解する．

（3） ユーザは感情レベル，時間レベルでの世の中の反応を知った後，システムによりその感情の理由を知らされる．ユーザは興味のある感情，あるいは日にちの Blog 記事を選択し，Web ページを閲覧しながら，Blog 記事の内容を音声により視聴する．

（4） もしも，視聴している Blog 記事の内容に興味が無ければ，同じラジオチャンネル内の他の Blog 記事を，もしくは，他の感情のラジオチャンネルを選択する．

以上を繰り返す．

3.2　ラジオ番組表メタファ

ユーザは，検索結果としての Blog エントリを整理し提示する際に，ラジオ番組表メタファを利用する．このラジオ番組表メタファで扱う情報は，感情情報及び時間情報である．

3.2.1　メタファの単純化されたモデル

ラジオ番組表メタファのモデルは，図 2 のように表される．すなわち，より上方のセルに対して，より新しい時刻印が割り当てられる．各セルの単位は日付である．セルの横の行は，同じ日付で同期されている．また，セルの縦の列の単位が各感情に対応する．よって，例えば「悲しい」内容と判定された Blog 記事は全て同じ列に並べられる．以上から分かる通り，各セルには，複数の Blog が割り当てられ得る．例えば,「悲しい」内容と判定された Blog 記事が同じ日に複数存在するなら，セル内の Blog 記事も複数となる．逆に，ひとつも Blog 記事の割り当てられないセルも存在する．

3.2.2　実際のメタファ

実際のラジオ番組表メタファは，以下の 2 つの単位で構成される．

- ラジオチャンネル
  各ラジオチャンネルは，各々のセルの縦の列に該当する．すなわち，ひとつのチャンネルにひとつの感情が割り当てられる．
- チャプター
  各ラジオチャンネルには，Blog 記事が並べられており，それらのひとつひとつをチャプターとする．前述した通り，チャンネル内の各 Blog 記事は，書かれた日付により同期された形で，時系列に並べられている．もしも，分類された感情以外のパラメータで大きい値が存在するならば，その感情はその Blog 記事の副次的な感情を表すとする．すると，そのチャプターの背景色を変化させることにより，副次的な感情で強いものがあることをユーザに示す．

このラジオ番組表メタファを利用して，主に以下のような情報が理解可能となる．

- トピックがいつ話題となったか
  各セルには，その割り当てられた日付に書かれた Blog 記事が入っている．すなわち，記事が多く入っている時期がそのトピックが話題となり，多く取り上げられた時期であることを表す．例えば，2 日前のセルに Blog 記事が多く集中して入っていれば，そのトピックは，2 日前に話題となったことが分かる．
- トピックにどの感情を持った人が多いか
  各ラジオチャンネルには，その感情の内容を表す Blog 記事が入っている．すなわち，そのトピックに対してどのような感情を持った人が多いかを各チャンネルに割り振られている Blog 記事の数により，把握できる．例えば,「怒り」を表すチャンネルに多くの記事が集中的に入っていれば，そのトピックに対して怒っている人が多いということが分かる．

また，以上の 2 つを組み合わせて，いつの時期にどの感情を持った人が多いかを見ることにより，1 週間前にはそのトピックに対して怒っている人が多かったようだが，昨日からは喜んでいる人の方が多くなってきているようだ，といったことが分かる．

4.　Blog 検索のクエリ生成方法

本システムにおいて，ユーザは自ら検索という行動を起こすことなく，システムがユーザの背後で独自に検索を行う．検索の目的は，ユーザの興味を反映した内容について記述している Blog 記事を検索することである．これを Blog 検索エンジンを用いて行う．また，検索はユーザがリンクナビゲートする度に行うので，生成するクエリはユーザのインスタントな興味を反映するものでなければならない．そのようなクエリを生成する方法について述べる．

4.1　アンカーキーワードの抽出

リンクナビゲート時に，ある Web ページから張られた順リンクを辿って他の Web ページを閲覧した際に，ユーザはアンカー文字列を参考にして，次に自分が閲覧するページを決定する．よって，リンクナビゲート時のアンカー文字列から，ユーザのより直接的な興味を抽出可能であると考えることが出来る．そこで，本システムでは，Blog 検索のためにユーザのクリックしたアンカー文字列に含まれる文字列を抽出する．リンクナビゲート時のアンカー文字列に含まれる各単語を，アンカーキーワード $(A_1, A_2, \cdots A_n)$ とする．

4.2　特徴キーワードの抽出

ユーザが自らの意思で選んだアンカー文字列がユーザの直接的な興味を表すのに対して，ユーザがリンクナビゲートした結果の Web ページは，ユーザがリンクアンカーをクリックした

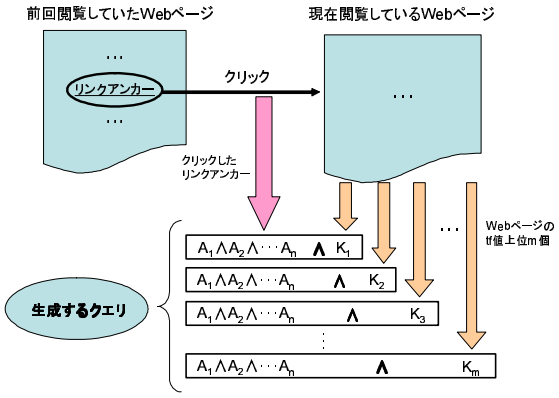

図 3　クエリ生成

結果，表示されたページである．したがって，開いた Web ページは，アンカーキーワードほどユーザの興味を直接表してはいないものの，ユーザの興味をある程度は表していると考えることが出来る．その為，ユーザが閲覧している Web ページも考慮してクエリを生成する．そこで，ユーザが閲覧している Web ページを特徴付けるために，Web ページ内の各単語の出現頻度に基づいた特徴ベクトルを作成する．Web ページの特徴ベクトルの要素としては，ページ内に出現する各単語の出現回数である tf 値を用いる．Web ページ内に出現する単語のうち，その Web ページにおいて出現頻度が高い単語は，その Web ページを特徴付けていると考えることが出来る．そこで，Web ページに出現する単語を，その出現頻度が大きいものから順に m 個抽出する．それらを $(K_1, K_2, \cdots K_m)$ とし，Web ページの特徴キーワード群とする．

4.3　クエリ生成

本システムは，ユーザの興味を反映した内容に関して記述している Blog 記事を検索するために，各 Web ページから抽出した特徴キーワードとアンカーキーワードを組み合わせ，AND 検索による検索を行う．ユーザがクリックしたアンカー文字列は，ユーザの直接的な興味を表し，その結果，リンクナビゲートした結果の Web ページは，ユーザの間接的な興味を表す．したがって，Blog 検索を行う場合，アンカーキーワードは必須の要素であると考えることが出来る．そこで，図 3 のように $(A_1, A_2, \cdots A_n, K_1), (A_1, A_2, \cdots A_n, K_2), \cdots, (A_1, A_2, \cdots A_n, K_m)$ をクエリとして，m 回 AND 検索を行う．こうして得られた検索結果が，ユーザの興味を反映した内容に関して記述している Blog 記事であると考える．

5.　感情パラメータの導出

本システムでは，Blog 記事の感情パラメータを求めるため，熊本ら [8] の単語同士の共起を用いたアプローチを参考としながら，Blog 検索エンジンを利用して Blog の感情辞書を構築し，その辞書に基づいてナイーブベイズ法を利用して各 Blog 記事の感情パラメータを求め，書き手の感情を判定する．辞書構築のため，解析した Blog エントリは，計 48328 件である．以下

喜び	不安	悲しい	苦しい	怒り
歓喜	おぼつかない	かなしい	困難	憤り
ジョイ	心もとない	哀しい	しにくい	立腹
悦び	案ずる	悲哀	難しい	腹立ち

表 1　各感情対応語の一部

喜び	不安	悲しい	苦しい	怒り	計
213704	351943	107861	441249	131617	1246374

表 2　辞書中の各感情の単語延べ総数

単語	喜び	不安	悲しい	苦しい	怒り
勝利	61	32	5	42	9
会社	84	230	53	253	91
政治	22	45	6	54	27
涙	120	79	142	99	70
プログラミング	0	2	2	14	1

表 3　感情辞書の一部

に感情辞書の構築方法及び，各 Blog 記事の感情パラメータ導出方法を提示する．尚，本システムで抽出する感情は，{ 喜び，不安，悲しい，苦しい，怒り } の 5 つである．

5.1　感情辞書構築方法

Blog の感情辞書を構築するため，Blog から各感情に強く関係のある単語を抽出する．まず,「喜び」,「不安」,「悲しい」,「苦しい」,「怒り」 という各単語と各々の類義語を，ここでは「感情対応語」とする．類義語に関しては，言語工学研究所 [9] によるシソーラス検索の検索結果における「同義語」を参考とした．表 1 に 5 種類の感情対応語の一部を挙げる．

次に，各感情に対応する感情対応語をクエリとして順次 Blog 検索エンジンに投入し，その検索結果を取得する．そして，検索結果の各 Blog 記事におけるクエリ，すなわち感情対応語の出現箇所の周辺に共起する単語の出現回数をカウントしていく．また，Blog 記事の中でもとりわけ日記に近い内容の Blog 記事ではひとつの記事の中に複数の感情が含まれている場合が多い．したがって，ノイズを少しでも減らすために Blog 記事中のクエリとの単純共起ではなく，出現箇所の周辺の共起を解析対象とした．

以上の操作により作成した辞書中の各感情における単語の延べ総数を表 2 に，作成した Blog 感情辞書の一部を表 3 に示す．

5.2　ナイーブベイズによる感情パラメータ導出

次に，作成した感情辞書を元に各 Blog 記事の感情パラメータを求める．本システムでは記事の各感情パラメータを求めるのに，固定されたドキュメント d が各感情クラス c に分類される確率を利用する．すなわち，$\Pr(c|d)$ を求める．$\Pr(c|d)$ を各感情クラス c について求め，その値が最大の感情クラス c にその Blog 記事を分類する．$\Pr(c|d)$ はベイズ理論により以下のように表される．

$$\Pr(c|d) = \frac{\Pr(c)\Pr(d|c)}{\sum_{\gamma} Pr(\gamma)Pr(d|\gamma)} \quad (1)$$

すると，まず分母の部分は感情クラス c に依存しないため，

無視できる．また，各感情の Blog 中の分布を表す $\Pr(c)$ は判定するのが非常に困難であるために，一様であると仮定する．次に，タームの出現回数も考慮する多項モデルによるナイーブベイズによる分類によると，$\Pr(d|c)$ は以下のように表される．$n(d,t)$ をドキュメント d 中のターム t の出現数，$\theta_{c,t}$ を感情クラス c に属するドキュメントにターム t が出現する確率とすると，

$$\Pr(d|c) = \Pr(L = \ell_d|c) \left(\begin{array}{c} \ell_d \\ \{n(d,t)\} \end{array} \right) \prod_{t \in d} \theta_{c,t}^{n(d,t)} \qquad (2)$$

where

$$\ell_d = \textstyle\sum_t n(d,t)\ , \left(\begin{array}{c} \ell_d \\ \{n(d,t)\} \end{array} \right) = \frac{\ell_d!}{n(d,t_1)!n(d,t_2)!\cdots}$$

ただし，式 (2) の第 1 項及び第 2 項は各クラスをランキング付けする場合は無視できるので，第 3 項のみを考慮すればよい．第 3 項は感情辞書により求まるので，こうして各クラスの $\Pr(d|c)$ が求まり，$\Pr(c|d)$ も求まる．すると，各感情パラメータ $Param(c)$ は $\Pr(c|d)$ を，評価したターム数により正規化することにより，以下のようにして求まる．

$$Param(c) = \frac{\Pr(c|d)}{\sum_{t \in d} n(d,t)} \qquad (3)$$

本手法でいくと，Blog 記事のタイトル及び本文がドキュメント d に，ターム t が単語に対応する．こうして，各 Blog 記事の感情パラメータを求めることができる．

5.3 感情パラメータに基づいた副感情抽出

感情というのは非常に複雑であり，どんな Blog 記事でも単純な 5 つの感情に分類可能であるとは考えにくい．したがって，各記事に対して，分類された感情以外に副次的に強い感情が存在するならば，ラジオ番組表メタファにおいてそれを各記事に該当するチャプターの背景色という形でユーザに提示する．

分類された感情以外のパラメータ $Param(c)$ で非常に大きい値が存在するならば，該当する感情クラス c はその Blog 記事の副次的な感情を表すと考えることが出来る．そのため，各感情パラメータ $Param(c_i)$ の中で最大値に対する比がある閾値以上という条件を満たす感情クラス c は，その Blog 記事の副次的な感情であると判断する．そう判断された場合には，その Blog 記事にはラジオ番組表メタファにおいて記事の該当するチャプターに対して感情クラス c に対応する背景色を付けることとする．

6. プロトタイプシステム

本章では以上までに述べたことをもとにプロトタイプシステムの設計及び実装を行い，本研究の手法の有効性を検証する．プロトタイプシステムの実装は，Microsoft 社の VisualStudio C#.NET により行った．

本システムの実装は 2 段階となっている．まず，感情パラメータを導出するために事前に Blog 検索エンジンを用いて感情辞書を構築しておく．次に，ローカルに保存してあるその辞書を用いた BlogRadio のプロトタイプシステムを実装する．いずれは，この Blog 感情辞書も静的なものではなく，動的に更新されるものへとしていきたい．

6.1 設　　計

6.1.1 感情辞書構築

Blog 感情辞書を構築するために，RSS 検索エンジン Bulkfeeds [3] を用いて各感情対応語を含む Blog 記事を収集した．ただし，検索結果の数が 4000 件を超えるものに関しては 4000 件までを取得した．その後，各感情対応語の周辺部分を茶筌 [10] を用いて形態素解析を行い，辞書を構築するために，出現する各単語の基本形を取得し，その出現数をカウントした．ただし，辞書に含める品詞は（名詞，動詞，形容詞，副詞）とした．

6.1.2 Blog の検索

本システムでは，クエリ生成のためにユーザのクリックしたアンカー文字列と閲覧する Web ページを用いる．いずれも，茶筌 [10] による形態素解析を行い，出現する単語の品詞の中でも名詞だけをキーワードとして評価した．また，ストップワードを設定し，それらを結果から除去した．Web ページの特徴キーワードとしては，tf 値が上位 8 位までの単語を使用した．

こうして生成したクエリを用いて，RSS 検索エンジン Bulkfeeds [3] により，Blog 記事を検索した．検索結果は各クエリにつき最新の記事 10 件まで取得した．また，近年では，メーリングリストやニュースの RSS 配信も行われているので，明らかにそれらであると考えられるものについては，検索結果から除去した．

6.1.3 感 情 判 定

感情判定のため，検索結果の Blog 記事に対しやはり形態素解析を行い，出現する単語の基本形を用いてナイーブベイズ法により各記事の感情パラメータを求めた．ただし，式（2）における $\prod_{t \in d} \theta_{c,t}^{n(d,t)}$ は，非常に微小な値となるので，log をとり，

$$\log(\prod_{t \in d} \theta_{c,t}^{n(d,t)}) = \sum_{t \in d} n(d,t) \log \theta_{c,t} \qquad (4)$$

とした．また，$\theta_{c,t} = 0$ の場合，$\log \theta_{c,t} = -\infty$ となるため，それを防ぐパラメータスムースという操作を行った．

感情パラメータを利用した副感情抽出に関しては，log をとったことにより，$Param(c) < 0$ であることを考慮し，

$$\frac{Param(c)}{Param(c_{max})} < 1 + \alpha \quad (\alpha > 0) \qquad (5)$$

が成り立つ際に，その Blog 記事には感情クラス c に対応する副次的な感情が存在すると判断した．

6.2 実　　装

以上の設計をもとに，BlogRadio のプロトタイプシステムを実装した．図 4 にブラウザの，図 5 にラジオ番組表のユーザインタフェースを示す．今回は Web ブラウザとラジオ番組表は別のウィンドウを用いて実装した．

6.3 考　　察

6.3.1 聴覚と視覚を用いたインタフェース

Web ページを閲覧しながら，Blog 記事を聴覚により聞くと

図 4　Web ブラウザのウィンドウ

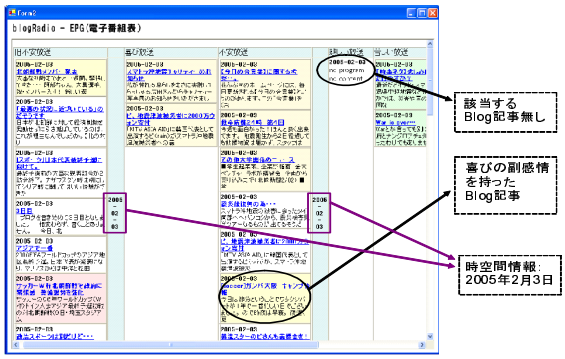

図 5　ラジオ番組表のウィンドウ

いう行為は，違和感なく行うことが出来た．ただし，それはユーザが Web ブラウジングを行っている際にどれだけ積極的に情報を取得しようとしているかにも関連すると考えられる．例えば，ニュース記事をチェックしたり，趣味に関する話題を Web で閲覧したりといった，ユーザがあまり積極的に情報を得ようとはしていず，どちらかというと受動的に気軽な気持ちで Web ページを閲覧している場合は，音声情報というのはそれほど邪魔にはならなかった．これが，例えばレポートを書くための資料を収集している等といった，より積極的に情報を取得しようとしている状況であるならば，音声情報は，あまり頭には入らないであろうと考えられる．

もうひとつの問題に，Blog 記事の本文の全文読み上げだと，全て聞くのに時間がかかり過ぎ，ユーザが途中で飽きてしまうといった問題もあった．従って，内容の要約などを行い，記事自体の長さを短縮したり，記事に何らかの手を加えることによりユーザを楽しませるコンテンツに変換するといった工夫が必要であると考えられる．

6.3.2　ユーザが視聴する Blog を自由に選択するスタイル

今回のシステムでは，ユーザは自由に視聴する記事を選択できるため，好みの記事を選択することが出来た．また，ラジオ番組表メタファにより，ユーザは素早く自分が視聴したいコンテンツにたどり着くことが出来た．しかし反面，自由度が非常に高かったので，実際にラジオを視聴する時のように，興味のある番組がなかったのでとりあえず何らかの番組を流している内にユーザの興味が広がっていくといった現象があまりなかった．このように，ユーザの興味を自然と広げられるように，ある程度ユーザの自由度を制限することも考慮するべきであると考えられる．

6.3.3　ラジオ番組表メタファ

時間情報及び感情情報により構成されるラジオ番組表メタファは，話題の時間的推移及び他ユーザの感情の分布を見るには有効であった．ただし，感情の判定に関してはまた後述するが，常に妥当であると考えられる結果が出たわけではなかった．時間情報からは特に，ユーザの知らないサブトピックに関する内容がある時期話題になっていること等も分かった．例えば，「iPod」に関するページを閲覧していた際には 2 月の 2 日，3 日にドラえもんのデザイン入り「iPod mini」について言及している Blog 記事がラジオ番組表に多く出現しており，それがその時期に話題になっていたことが分かった．

6.3.4　Blog の検索について

ユーザが新しいページを開くたびに Blog 検索を行うという仕組みは，ラジオ番組表を概観するのを重視している場合にはユーザが飽きにくい反面，多くの Blog 記事を視聴したいという場合に時間が足りないといったことが多かった．

また，毎回検索を行うと，ラジオ番組表が表示されるまでに検索による時間がかかり，それがユーザのストレスにつながることが多かった．また，ユーザがクリックしたアンカー文字列が，「記事全文」などといった，あまり特徴的なキーワードで構成されていない場合に，あまりいい検索結果が得られなかった．このような場合は，リンクナビゲート時にユーザの興味があまり変化していないことを表している．以上のように，ユーザがある話題に関するページを連続して閲覧し続けている場合は，番組表をあまり更新しないといった工夫も必要であると考えられる．

6.3.5　感 情 判 定

本研究で提案した手法を用いて，感情判定の簡単な実験を行った．ランダムに抽出した Blog 記事 120 件に対し，今回提案したナイーブベイズ法により感情分類を行った．その結果を表 4 に示す．まず，この 120 件の内書き手の感情が不明瞭であると判断されるものが 40 件存在し，これらは適合しているかの判断が非常に困難であった．したがって，これらを適合していないと仮定した場合の適合率を表中における適合率とし，これらを除いて計算した場合の適合率を，明瞭な記事の適合率とした．この結果によると，感情が不明瞭である記事を含めた適合率は 42%にとどまっているが，それらを除いた場合の適合率及び再現率はいずれも 63%であった．したがって，書き手の感情が明瞭に表れている記事に対してはまずまずの結果が得られていると考えられる．これら 40 件のような書き手の感情が不明瞭な記事の扱いは，これからの課題とする．

表 5，表 6 に，実際に感情判定を Blog 記事に対して行った結果を示す．以上のように，この Blog 記事は，「喜び」に分類される．

	喜び	不安	悲しい	苦しい	怒り	全体
適合率 (%)	66	33	55	43	45	42
明瞭な記事の適合率 (%)	80	54	55	73	71	63
再現率 (%)	40	86	71	60	63	63

表 4　感情判定実験結果

タイトル：皆さんお疲れ様！そしてありがとう。
土・日２日間のイベントが終了し、今日の撤収作業でほぼ片付き、アタクシのバイトも本日をもって終了しました。とは言っても実行委員でもあるので、招集がかかっていた金曜日から今日までのバイト代はいただかない事にしました。（中略）はじけた人達と辛かったり楽しかったりした思い出を語りながら喜びを分かち合えるのは、本当に幸せな事だと思います。田舎だからこそ、なせる事なのかなと。我が町には、春以外はお祭りイベントがあるので、これからもこの達成感と喜びを味わうため、精一杯協力していければと思います。ホントに楽しかったです。かかわった全ての方に「ありがとう」と言いたいです。

表 5　感情判定を行った記事

喜び	不安	悲しい	苦しい	怒り	分類結果
-7.96	-8.49	-8.42	-8.55	-8.52	喜び

表 6　上記の記事の感情パラメータ

7.　今後の課題とまとめ

7.1　今後の課題

7.1.1　ラジオチャンネル構成法

現時点では Blog 記事を各感情に分類した結果を提示したものを使用しているが，今後は実際のラジオコンテンツに近づけるために Blog 記事群の番組化を行うことを検討している．その番組化に伴い，チャンネル構成法も変化させたい．

- Blog より抽出した感情情報
- 各 Blog 記事の有する時間情報
- Blog 記事の書き手の情報

以上のデータを主に用いて，他のユーザの主観的意見を紹介したりといった番組の自動生成を試みたい．その際に，POC [7] における番組化手法は大いに参考になると考えられる．

7.1.2　Blog 記事検索手法

Web ページ閲覧アクションよりユーザの興味を判定し，そこから Blog 記事検索を行う手法も今後，より検討する必要がある．現時点での手法は，第 4 章で説明した通りであるが，この手法は,「Blog コンテンツ」の検索というものに最適なものではない．今後はラジオチャンネル構成法と合わせて Blog 特有の話題構造なども考慮しながら，より Blog 記事の検索に適合した手法を検討する．

7.1.3　感情判定法

現在分類に使用している品詞以外に，感動詞や「！」「♪」などといった感情に関連すると考えられる文末表現，そして，顔文字等といった，主観的な内容のコンテンツに多く登場する情報も利用して感情判定の適合率を上げることを考えたい．また，プロトタイプシステムより，今回扱った，比較的基本的であると考えられる５つの感情だけではバリエーションが不十分であることが分かったので，これら以外の感情抽出や，分類法の再考なども試みたい．

7.2　ま と め

本研究では，主観的情報の閲覧とその一覧性の高い表示，及び客観的情報の閲覧をシームレスに統合する BlogRadio というシステムを提案した．そのシステムの概要は，ユーザが Web ページを閲覧している状況において，ユーザの興味を反映した内容に関して記述している Blog 記事の集合を整理し，さらに各記事の内容を音声の形で可聴化し，提示するというものである．こうして，ユーザは客観的な情報を Web ページから，他のユーザによる主観的な情報を Blog から取得する．

また，本研究で提案した内容をもとにプロトタイプを実装し，その有効性を検証した．その結果，ユーザは，閲覧コンテンツとそれに対するユーザの反応を同時に知ることが可能となり，さらにユーザの反応の分布を整理された形で概観することも可能となった．ただし，問題点も幾つか見つかり，それらを明らかにするとともに，今後の研究の方針についても言及した．今後は，現在の問題点の解決策について検討していくとともに，現在のシステムを今回明らかにした方針に基づいて，発展させて行きたい．

謝　　辞

本研究の一部は,《知的資産》文部科学省科学技術振興費プロジェクト「異メディア・アーカイブの横断的検索・統合ソフトウェア開発」（代表：田中克己）および，平成 16 年度科研費特定領域研究 (2) 「Web の意味構造発見に基づく新しい 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」によるものです．ここに記して謝意を表すものとします．

文　　献

[1] Movable Type
http://www.movabletype.org/.
[2] goo ブログ
http://blog.goo.ne.jp/.
[3] Bulkfeeds
http://bulkfeeds.net/.
[4] M. Barra, T. Cillo, A. D. Santis, U. F. Petrillo, A. Negro, V. Scarano, T. Matlock and P. P. Maglio: “Personal webmelody: Customized sonification of web servers”, WWW Posters (2001).
[5] L. S. Petrucci: “Websound: a generic web sonification tool allowing hci researchers to dynamically create new access modalities”.
[6] 金星庸, 角谷和俊, 田中克己：“質問センサーによる未巡航ウェブページ群情報の“ 可聴化 ””, 情報処理学会研究報告, **Vol.2003, No5**, pp. 147–154 (2003).
[7] 畦地真太郎, 藤原伸彦, 角薫, 平田高志, 矢野博之, 西田豊明：“パブリック・オピニオン・チャンネル”, 人工知能学会誌, **Vol.15**, pp. 69–73 (2000).
[8] 熊本忠彦, 田中克己：“Web ニュース記事からの喜怒哀楽抽出”, 第 165 回自然言語処理研究会.
[9] 言語工学研究所
http://www.gengokk.co.jp/.
[10] 形態素解析システム茶筌
http://chasen.naist-nara.ac.jp/.