

ブログの実態に関する調査研究

～ ブログコンテンツ量の推計とブログの開設要因等の分析 ～

平成 21 年 3 月

総務省 情報通信政策研究所

はじめに

ブログは個人やグループなどにより運営され、時系列的に更新されるウェブページのことであり、CGM(Consumer Generated Media:消費者生成型メディア)の代表例として、この数年急激に普及したメディアである。

米国テクノラティ社の2007年4月の発表によると、日本語による発信量が世界一と報告されており、我が国におけるブログの利活用が非常に活発であることが伺えるが、その実態に関する包括的な調査はない状況であった。

本調査研究は、我が国のブログについて、その発展の経緯を、コンテンツ量の面から把握するとともに、ブログ開設者の利用の傾向等に関する分析を行い、定性データからも把握しようとする初めての試みである。また、ブログが我が国で本格利用されるようになってから、まだ5、6年ほどであるが、ブログサービスを悪用した迷惑なブログの利用の増加も指摘されており、その実態についても把握する。

本調査研究の成果が、ICTの利活用による社会・経済の活性化のために、個人のICTのさらなる利活用の促進や、インターネット上における様々な課題を検討する際に、何らかの参考になれば幸いである。

最後に、本調査研究の実施においては、川浦康至東京経済大学教授、川上善郎成城大学教授、荒木万寿夫青山学院大学准教授、アライド・ブレインズ株式会社の内田斉氏、石川哲也氏、伊藤将雄氏並びに細川裕樹氏、葉玉匡美弁護士及びアジャイルメディア・ネットワーク株式会社の徳力基彦氏他、多くの方々にご協力いただいた。

心よりお礼を申し上げたい。

2009年3月

情報通信政策研究所 調査研究部

主任研究官 佐伯 千種

研究官 岩間 健宏

目 次

第1章 序論	1
1-1 本調査の目的	1
1-2 本調査の内容	1
1-2-1 ブログのコンテンツ量推計調査の内容	1
1-2-2 ブログ開設者の意識調査の内容	2
1-2-3 スパムブログの調査の内容	2
1-2-4 報告書の構成	2
1-3 ブログについて	3
1-3-1 ブログとは	3
1-3-2 ブログサービス事業者数と登録者数等	3
1-3-3 ブログの機能	5
1-3-4 海外のブログの状況	7
第2章 ブログのコンテンツ量推計調査	10
2-1 ブログのコンテンツ量推計調査の概要	10
2-1-1 調査対象範囲	10
2-1-2 調査の対象項目	11
2-1-3 ブログコンテンツ量の推計モデル	13
2-2 実施調査の手順と結果	15
2-2-1 実施調査の手順	15
2-2-2 クローラー調査手順と結果	16
2-2-3 ブログ開設者の運用履歴アンケート調査の概要と結果	23
2-3 ブログのコンテンツ量推計結果	27
2-3-1 国内のブログコンテンツ量(2008年1月現在)	27
2-3-2 ブログコンテンツ量の推移	28
2-4 ブログコンテンツの種類と付加機能の状況等	33
2-4-1 アクティブブログの平均像	33
2-4-2 マルチメディアファイルの推移	33
2-4-3 コメント数、トラックバック数の推移	35
2-4-4 記事の更新曜日・時間	36
2-4-5 更新記事数によるブログの分布	36
第3章 ブログ開設者の意識調査	38
3-1 調査の目的と概要	38
3-1-1 調査の目的と仮説設定	38
3-1-2 アンケート調査の概要	42
3-2 集計結果	43

3-2-1	回答者属性	43
3-2-2	インターネットの利用状況	46
3-2-3	ブログの利用状況	50
3-2-4	各ブログの運営状況	64
3-3	開設者の性別・年代との関連	69
3-3-1	属性別のブログ利用状況	69
3-3-2	属性別の運営状況	78
3-4	ブログ開設者のグループ分け	81
3-4-1	ブログ開設者のグループ分けの実施	81
3-4-2	グループ別のクロス分析	87
3-4-3	各グループの特徴	99
3-4-4	指向性・ブログ内容によるブログ開設者分類仮説の検証	100
第4章	ブログの利用における問題点	101
4-1	スパムブログ	101
4-1-1	スパムブログの調査概要	102
4-1-2	調査結果	103
4-2	スパムトラックバック、スパムコメント	105
4-3	ブログの「炎上」	106
4-4	さいごに	108
参考資料		109
<参考1>	ブログのコンテンツ量推計の計算式と推計手順	110
<参考2>	ブログのコンテンツ量推計におけるデータ補正の考え方	113
<参考3>	ブログのコンテンツ量推計のケース別推計結果と検証	117
<参考4>	クラスタ化手法の選択	131

第1章 序論

1-1 本調査の目的

近年、我が国のインターネット利用においては、ブロードバンドの普及に伴い、これまで非対称的に情報を受信してきた利用者が主体的に情報を作り出し、発信する CGM (Consumer Generated Media:消費者生成型メディア) が注目されている。中でもブログは、特別な知識を必要とせずに情報発信が可能であることや、無料のブログサービスが数多く提供されるようになったことから急速に普及し、UGC (User Generated Contents ; ユーザ生成型コンテンツ) を生み出す代表的なメディアとして成長してきた。

総務省が 2006 年 3 月末に集計したブログの登録者数は 868 万人を超え¹、2007 年 4 月の米国テクノラティ社の調査では、日本語によるブログ記事が世界のブログ全体の 37% を占め、英語 (36%) を超えて世界第一位の記事数であるとの報告もある。

しかしながら、我が国のブログについては、その数や発展の経緯、利用者の意識や要因などを包括的に調査・分析した例がなく、ブログの実態は断片的なデータにとどまっている。そこで、本調査は、ブログの現状と発展経緯について、その実態をコンテンツ量に関する定量データとブログ開設者の定性データの両面から把握し、分析することを目的としている。

1-2 本調査の内容

本調査の内容は大きく分けると、ブログ数、記事数、データ量等のブログのコンテンツ量を定量的に推計するための調査(以下、「ブログのコンテンツ量推計調査」という。)、ブログ開設者の目的・意識を分析する調査(以下、「ブログ開設者の意識調査」という。)及び近年増加していると言われるスパムブログの実態を把握する調査(以下、「スパムブログの調査」という。)の3つで構成される。

1-2-1 ブログのコンテンツ量推計調査の内容

第1のブログのコンテンツ量推計調査に関しては、次の3つを主な調査項目としている。

第1点目は、国内のブログのコンテンツ量の推計である。ブログの現状として、2008年1月現在のブログ数、記事数、データ量等について推計した。

第2点目は、ブログの量的な推移の分析である。ブログの実態を把握する上では、現在のブログのコンテンツ量に加え、新規開設ブログ数や記事数の推移を元に、ブログが時系列的にどのように推移してきたかを動的に分析する必要がある。そこで、ブログ草創期から今日までの各時点でのコンテンツ量の推移を推計している。

¹ 2006年4月総務省発表http://www.soumu.go.jp/s-news/2006/060413_2.html

第3点目は、ブログのコンテンツの種類や付加機能の利用状況等の把握である。当初のブログはテキストベースに簡単なコメントとトラックバックの機能がついただけのものであったが、最近では画像や動画等のファイルを添付することができるようになってきている。そこで、ブログ普及期から今日までの各時点におけるブログの画像・動画等のファイルの添付状況や、コメント・トラックバック数の推移などについて分析する。

1-2-2 ブログ開設者の意識調査の内容

2つめの調査の柱であるブログ開設者の意識調査においては、ブログの開設経験者に焦点をあててアンケート調査を実施し、開設者のブログの利用状況を把握するとともに、ブログの開設・更新の動機において重要な要因の抽出や、ブログ開設要因の傾向によるブログ開設者のグループ分類を行い、ブログを開設する目的や態様の多様性などについて分析している。

1-2-3 スпамブログの調査の内容

本調査の実施に先立ち、有識者ヒアリング等を実施したところ、ブログに関する様々な問題点が抽出された。本調査では、特にいわゆるスパムブログと呼ばれる迷惑なブログの状況を把握・分析する。

1-2-4 報告書の構成

本報告書では、1-3において、ブログについて概観したあと、第2章においてブログのコンテンツ量推計調査の調査概要と推計結果について、第3章において、ブログ開設者の意識調査の調査概要と分析結果について説明する。第4章では、ブログの利用における問題点として、スパムブログの状況について分析するとともに、スパムコメントやスパムトラックバック、ブログの「炎上」について、考察する。

1-3 ブログについて

1-3-1 ブログとは

ブログとは、個人やグループなどにより運営され、時系列的に更新されるウェブページの総称である。「Web」と「Log」（日誌）を一語に綴った「weblog」（ウェブログ）という言葉が誕生し、それを略して「blog」（ブログ）と呼ばれている。

ブログの明確な定義はなく、記事に対して閲覧者がコメントすることや、閲覧者自身のブログへのリンクを作成するトラックバックのような双方向機能が特徴として挙げられることが多いが、我が国では、ブログ向けのツールを使って作られたウェブページを指すことが多い。

本報告書では、特にサーバソフトの設置をする必要もなく、希望者がすぐにブログを開設できるような仕組みをASP（Application Service Provider）形式により提供しているサイトを「ブログサイト」、ブログサイトを運営し、ブログサービスを提供している事業者を「ブログサービス事業者」と呼ぶこととする。

1-3-2 ブログサービス事業者数と登録者数等

我が国ではISP事業者やポータルサイト事業者の多くがブログサービスを提供するほか、特定カテゴリーや地域をテーマにしたもの、個人運営のもの、他のサービスの付帯サービスとして提供しているものなど、規模の小さなブログサービス事業者も多い。また、サービスの開始・停止や名称変更等も頻繁に行われるため、ブログサービス事業者数を把握することは容易ではないが、総務省が平成17年に実施した調査によれば、ブログサービス事業者は115社と報告されている²。

これらの事業者の中には有料ブログサービスを提供するものもあるが、大半が無料でブログサービスを提供し、広告収入を収益源としている。

【ブログサービス事業者の分類】

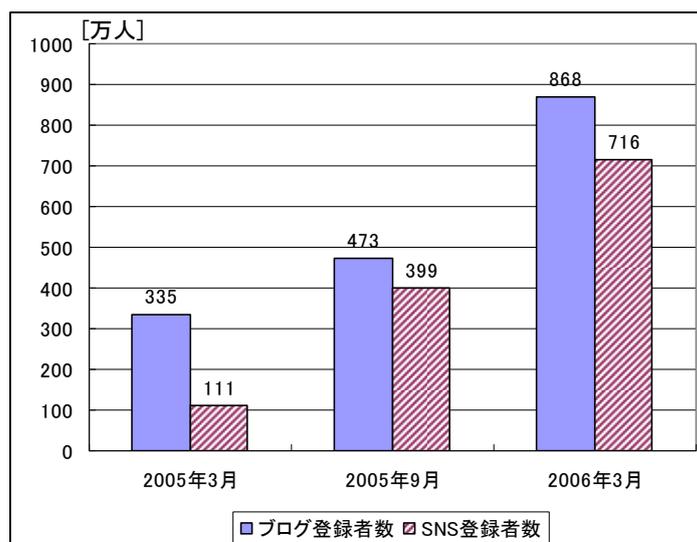
ISPを主力とする事業者	インターネット接続会員に対するサービスメニューの一つとしてブログサイトを運営 ----- 【例】ココログ、ウェブリブログ、So-net blog など
ポータル・ECなどネット関連事業を総合的に展開する事業者	ポータル、通信販売（EC）など数多くのインターネット関連の事業の一つとしてブログサイトを運営 ----- 【例】Yahoo!ブログ、楽天広場 など
その他周辺サービス系の事業者	インターネット広告事業のための媒体の一つとしてブログサイトを運営

² 「ブログ・SNSの現状分析及び将来予測」（平成17年5月 総務省）

	【例】Ameba ブログ、ヤプログ！など
ブログ専門系の事業者	ブログサイトまたはコミュニティサービスの運営を専門とし、関連システムやサービスを提供する事業を展開
	【例】ドリコムブログ、はてなダイアリーなど
個人運営者	個人などが運営
	【例】マイふれす など

ブログの登録者数に関しては、総務省が 2005 年から 2006 年にかけて、3 回にわたってブログと SNS の登録者数を公表しており、これによると、2006 年 3 月のブログ登録者数は、868 万人となっている³（図表 1-3-1）。

図表 1-3-1 ブログ及び SNS 登録者数の推移



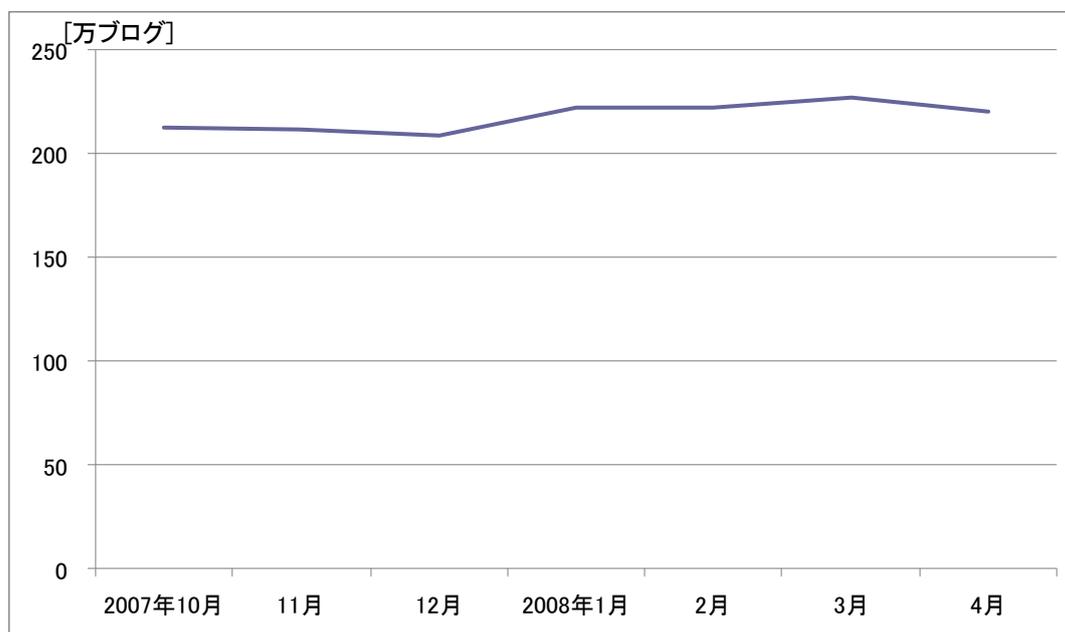
（出典）「平成 18 年版 情報通信白書」をもとに作成

また、個人が運営する「ブログファン」(<http://www.blogfan.org/>) では、国内の主要 21 ブログサイトについて、毎月のアクティブユーザ数（1 箇月（1 日から月の末日まで）のうちに更新されたブログのアカウント数（登録件数））を独自に調査し、発表している。

ブログファンが公表している最近のアクティブユーザ数の推移を見ると、全体としてはほぼ横ばいで推移している（図表 1-3-2）。

³ http://www.soumu.go.jp/s-news/2006/060413_2.html。2006 年 3 月のブログの登録者数は、ブログサービス事業者 53 社の登録者数を単純合計。

図表 1-3-2 主要 21 ブログサイトの月間アクティブユーザ数の推移



出所：ブログファン(<http://www.blogfan.org/service/>)

1-3-3 ブログの機能

ブログの機能は、(1)ブログの開設に関する機能、(2)記事に関する機能、(3)閲覧者と開設者とのコミュニケーションに関する機能、(4)他サービスとの連携に関する機能と、大きく4つに分けることができるが、以下に挙げる各機能の提供状況は、ブログサイトやブログソフトによって異なっている。

(1) ブログの開設に関する機能（ブログ単位の機能）

テンプレート	記事を書き込みする際に使用するテンプレートを選択することにより、ブログのレイアウトやデザインをブログ開設者の好みに合わせることができる。
ブログのジャンル設定	開設するブログに対して、ブログサイトによりあらかじめ分類されているジャンル(またはカテゴリ)を設定する。ブログ単位のジャンル設定を取り入れていないブログサイトも多い。
プロフィールの表示	ブログ開設者のニックネーム・年齢・性別・地域・自己紹介などを登録し、ブログの一部として表示する。
ブログパーツ	ブログのサイドバー(空きスペース部分)に、時計・ニュース・地図・ランキングなどを小さな画像やテキストで表示する。
閲覧制限機能 (限定公開機能)	ブログ開設者がブログ(または個々の記事)に対し、パスワードを設定し閲覧者を制限することができる。同じブログサイト内での会員IDにより制限する機能を取り入れているブログサイトもある。
RSS情報の提供 ⁴	最新のブログの状況を要約したRSS情報をサイドバーで提供し、RSS情報を利用する閲覧者に効率よくブログの更新を伝えることができる。

⁴ RSS (Rich Site Summary) とは、Webページを要約するための共通のデータ・フォーマット。URL・タイトル・記事本文の要約・更新時刻などを記述することができる。利用者はソフトウェアやWebサイトにあるRSSリーダー機能を用いることにより、情報が自動的に配信される。

(2) 記事に関する機能 (記事単位の機能)

タイトルの入力	記事を書き込みする際に入力する記事のタイトル。タイトルと本文はセットで管理される。
記事のカテゴリ設定	書き込みする際に記事ごとにカテゴリ(またはジャンル)を設定することができる。ブログサイトによりあらかじめ分類されているカテゴリと開設者が独自に作成するカテゴリがある。
本文の入力	書き込みする記事の本文(テキスト部分)。HTML タグも使用できる。
画像・音声・動画の書込み	画像・音声・動画ファイルを記事に添付して書き込みすることができる。ブログサイトにより、ファイルの種類、数量、容量の制限がある。
携帯電話からの書込み	携帯電話からメールで記事や画像ファイルを添付することができる。

(3) 閲覧者と開設者のコミュニケーションに関する機能

コメント	記事に対する感想や意見をブログ閲覧者が当該ブログ内で述べることができる。コメントを受け付けるか否かの選択、表示の事前承認など、開設者側で設定できる付加機能がある。
トラックバック	ある他人のブログの記事に閲覧者本人のブログへのリンクを作成することができる。トラックバックを受け付けるか否かの選択、表示の事前承認など開設者側で設定できる付加機能がある。

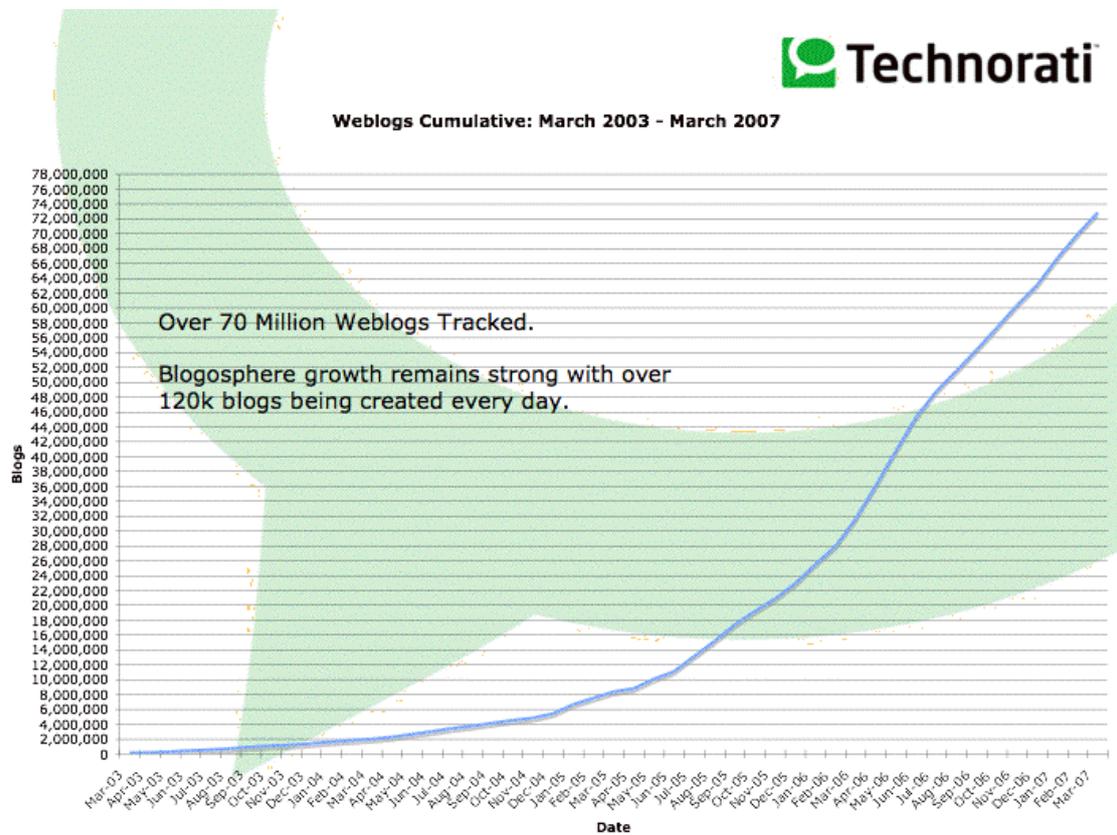
(4) 他サービスとの連携に関する機能

アフィリエイト	ブログサイトが提携または許可しているアフィリエイト広告を載せることにより、クリック数や商品販売に結びついた実績に応じた成功報酬を開設者が受け取ることができる。
外部の動画配信の掲載	動画再生ページを示すタグ(embed タグなど)をHTMLに貼り付けることにより、YouTube など外部の動画配信サービスに置かれた動画を視聴することができる。

1-3-4 海外のブログの状況

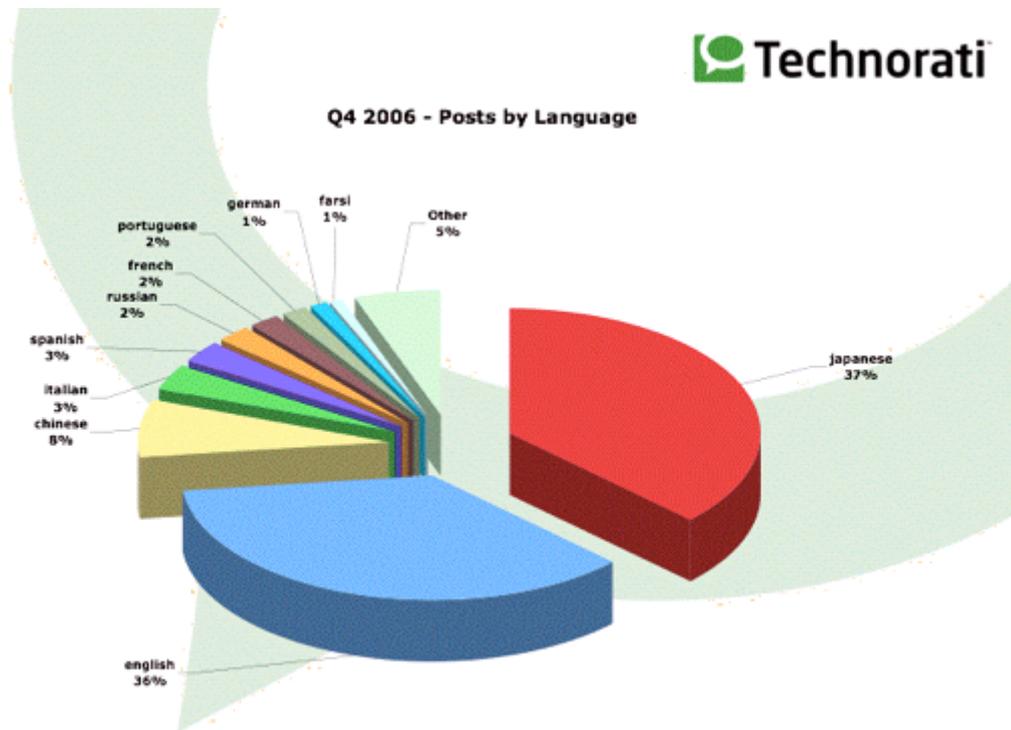
米国 Technorati 社の調査によると、2007年3月時点で同社が追跡するブログ数は、世界全体で7000万以上(図表1-3-3)、日本語による記事は全体の37%と、英語(36%)を超えて世界一位の発信量と報告されている(図表1-3-4)。

図表1-3-3 Technoratiにおける追跡ブログ数の推移



【出典】：<http://www.sifry.com/alerts/archives/000493.html>

図表 1-3-4 Technoratiによる言語別記事割合



【出典】：<http://www.sifry.com/alerts/archives/000493.html>

米国でのブログ活用状況については、Pew Internet & American Life Projectの2006年の調査⁵によると、米国のインターネット利用者の8.2%（1200万人）がブログを運営している（図表 1-3-5）。

図表 1-3-5 米国におけるネットユーザ数・ブログ閲覧者数・ブログ開設者数

Bloggers by the Millions	
Number of American adults who use the internet	147 million
Number of American adults who read blogs	57 million
Number of American adults who keep a blog	12 million

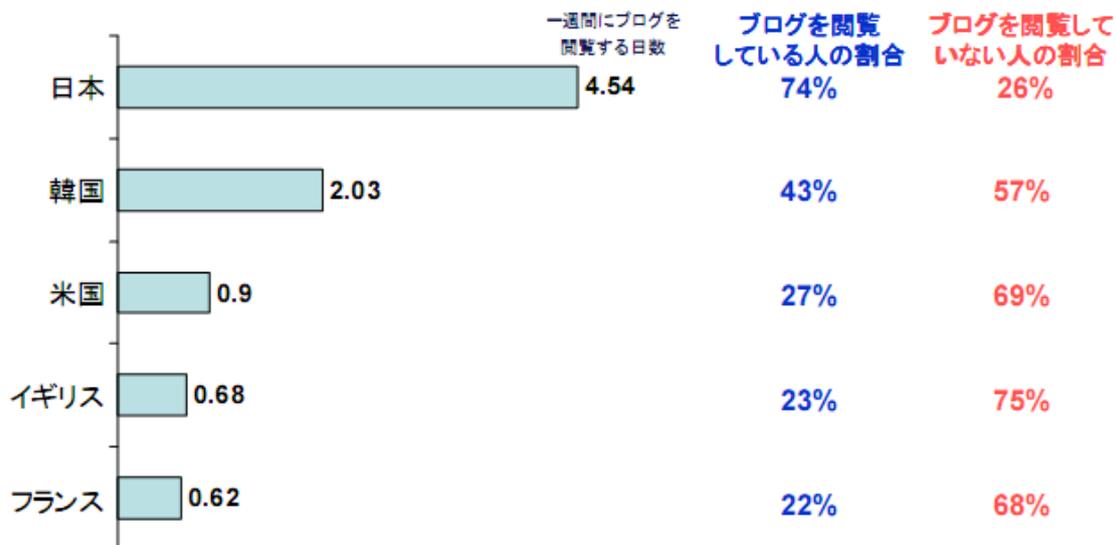
Source: Pew Internet & American Life Project Surveys. January 2006; Feb-April 2006; Nov-Dec 2005 and Feb-April 2006.

ブログの閲覧についてみると、エデルマン・ジャパンの調査（2007年3月公表）によれば、日本では、インターネット利用者の74%が、ブログを閲覧しているのに対して、韓国では43%、米国では27%、イギリスでは23%、フランスでは22%となって

⁵ Pew Internet & American Life Project, 2006, "Bloggers: A portrait of the internet's new storytellers" <http://www.pewinternet.org/pdfs/PIP%20Bloggers%20Report%20July%2019%202006.pdf>

おり、ブログの閲覧頻度は日本が突出して高くなっている（図表 1-3-6）。

図表 1-3-6 ブログ閲覧の頻度の国際比較（エデルマン・ジャパン）



【出典】：エデルマン・ジャパン「日本におけるブログ利用動向調査結果」

http://www.edelman.jp/data/news/JIBS_j.pdf

第2章 ブログのコンテンツ量推計調査

本章では、ブログのコンテンツ量推計調査の概要と結果について説明する。2-1で調査概要を説明し、2-2で実施調査の手順と結果を、2-3でコンテンツ量の推計結果をまとめて分析する。

2-1 ブログのコンテンツ量推計調査の概要

2-1-1 調査対象範囲

本調査の対象となるブログは、日本国内で開設・公開されているブログである。サーバが海外にある場合もあり、「日本国内のブログ」を厳密に定義することは難しいが、サービスやコンテンツが主に日本語で提供されているブログを「国内ブログ」とみなし、調査対象とした。

なお、次のようなものは、後述のクローラーによる情報収集ができないこと等の理由により、本調査の対象から外れている。

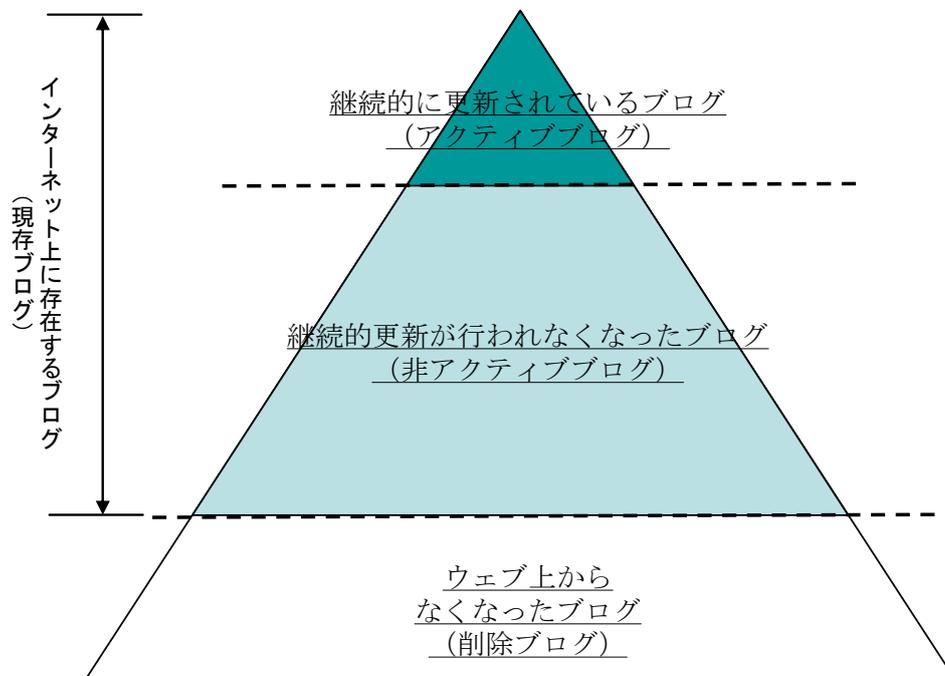
- ・ 閲覧制限を設定しているなど、コンテンツが一般に公開されていないもの
- ・ 会員制の SNS
- ・ パソコンからの閲覧や開設ができないモバイル専用ブログサイト（いわゆる「モブログ」）

2-1-2 調査の対象項目

本調査では、ブログを「アクティブブログ：1箇月に1回以上の頻度で継続的に更新されているブログ」、「非アクティブブログ：継続的な更新が行われなくなったブログ」、「削除ブログ：ウェブ上からなくなったブログ」の3つの状況に分類し（図表 2-1-1）、それぞれの数、各ブログに書き込みされた記事の数、ブログ記事のデータ量を推計する。また、これらの項目について2001年以降の推移についても推計を試みる（図表 2-1-2）。

さらに、ブログの記事を構成する画像や動画などのファイル数、記事に対して書き込みされたコメント数、トラックバック数の推移などについても把握する。

図表 2-1-1 ブログの推計対象



図表 2-1-2 調査対象項目一覧

対象項目	種別、対象時期など
ブログ数	次のブログ数（2001年1月～2008年1月の推移） ・継続的に更新されているブログ（アクティブブログ） ・継続的更新が行われなくなったブログ（非アクティブブログ） ・ウェブ上からなくなったブログ（削除ブログ） ・毎月の新規開設ブログ
記事数	上記4種類のブログの記事数（2001年1月～2008年1月の推移）
データ量	ブログ記事のデータ量（2001年1月～2008年1月の推移）
マルチメディアファイル数	1記事当たりの平均ファイル数とブログ全体でのファイル総数（2003年前半～2007年後半の半年単位での推移及び2008年1月の値）
コメント数	1記事当たりの平均コメント数（2003年1月～2008年1月の推移）
トラックバック数	1記事当たりの平均トラックバック数（2003年1月～2008年1月の推移）

2-1-3 ブログコンテンツ量の推計モデル

ブログの数については、ブログサービス事業者の公表するブログ登録者数やブログ検索事業者の公表している収集ブログ数などのデータがあるが、国内でこれまでに開設されたブログの網羅的なリストはなく、ブログ数などのコンテンツ量を直接知ることのできる資料は存在しない。

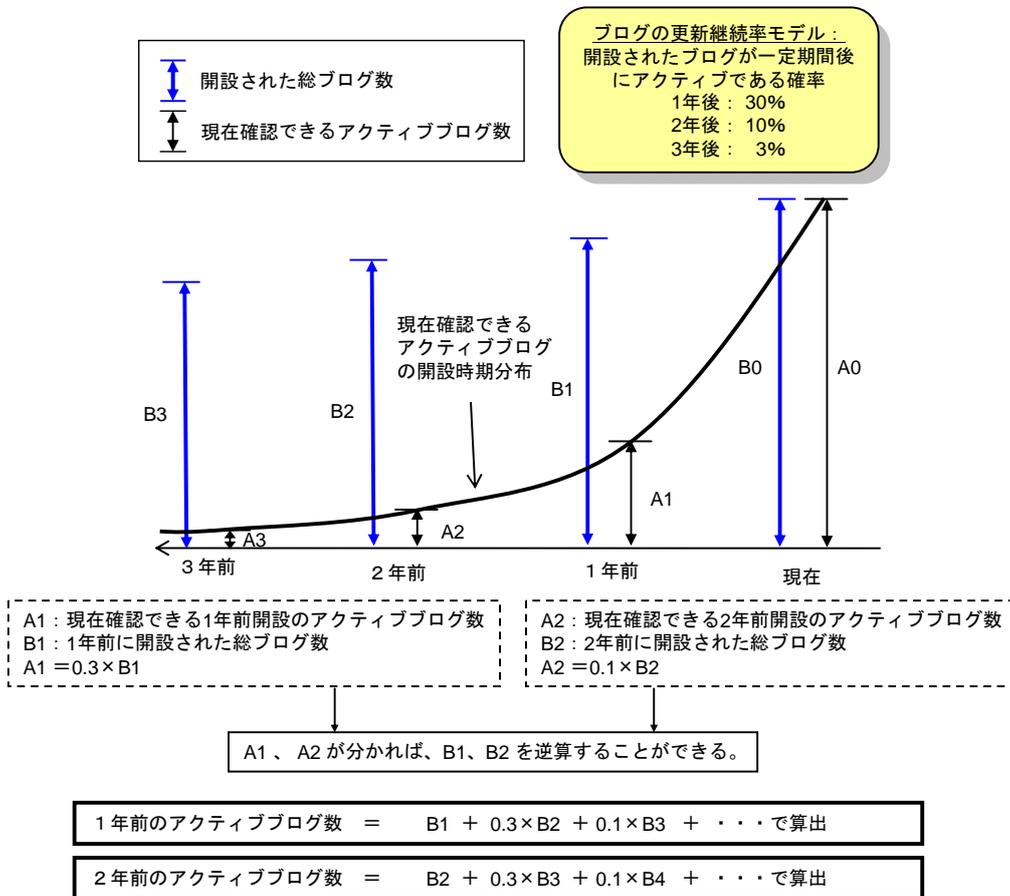
ブログサービス事業者が提供するすべてのブログサイトを一定期間観測すれば、継続的に更新されているアクティブなブログの数を把握することはできるが、この方法では更新されなくなった非アクティブブログや既にネット上から削除されてしまった削除ブログについて把握することはできない。

しかし、ブログの標準的な動態について次のようなモデルを設定し、必要な情報を入力することにより、国内でこれまでに開設されたブログの総数とブログ草創期から今日までのブログ数の推移について、非アクティブブログや削除ブログも含めて推計することが可能になる（推計の計算式と推計手順については、参考1を参照）。

- ① 国内のブログサイトの更新状況を一定期間観測してアクティブブログを収集し、ブログの記事に掲載されている書込み日時等のタイムスタンプのうち、最も古い日付をそのブログの開設日とみなすことで、現在のアクティブブログの開設時期別の数と分布を知ることができる。観測と情報収集は、クローラー⁶等のシステムを使って機械的に実施することができる。
- ② 一方、過去のある時期に開設されたブログは、当初はすべてアクティブブログであるが、時間が経過するにつれ、記事が更新されなくなったり削除されたりして、アクティブブログの割合は次第に低下していくと考えられる。①の方法で現在確認できるのは、過去の各時期に開設されたブログのうち、現在もアクティブなブログに限られる。
- ③ このとき、過去の各時期に開設されたブログが開設時から一定期間の後もアクティブである割合（更新継続率）がわかれば、現在観測できるアクティブなブログの数を、開設時期から現在までの経過期間分の更新継続率で割り戻すことにより、過去の各時期に開設されたブログ数を推計することが可能である。各時期での開設ブログ数を最初のブログ登場時から累計すれば、過去のある時点までに開設されたブログの総数を推計することができる。さらに過去の各時期に開設されたブログ数と更新継続率を用いて、過去の任意の時点でどれだけアクティブブログが存在したかを推計することもできる（図表 2-1-3）。

⁶ クローラーとは、リンクを辿ってウェブコンテンツにアクセスし、各コンテンツの情報を自動収集するシステム

図表 2-1-3 ブログコンテンツ量推計のモデル



④ また、過去のある時期に開設されたブログが一定期間後に削除される割合(削除率)がわかれば、過去の各時期に開設されたブログ数に削除率を乗じて合計することで、その後のある時期における削除ブログ数を推計することができる。最初のブログ登場時からの各時期の削除ブログ数を累計すれば、ある時点までにネット上から削除されたブログの総数を推計することができる。

⑤ 以上によって、過去の任意の時点までに開設されたブログ数の累計、過去の任意の時点で存在したアクティブブログ数、その時点までに削除されたブログ数の累計を推計できれば、これらから過去の任意の時点に存在した非アクティブブログ数を算出することができる。

2-2 実施調査の手順と結果

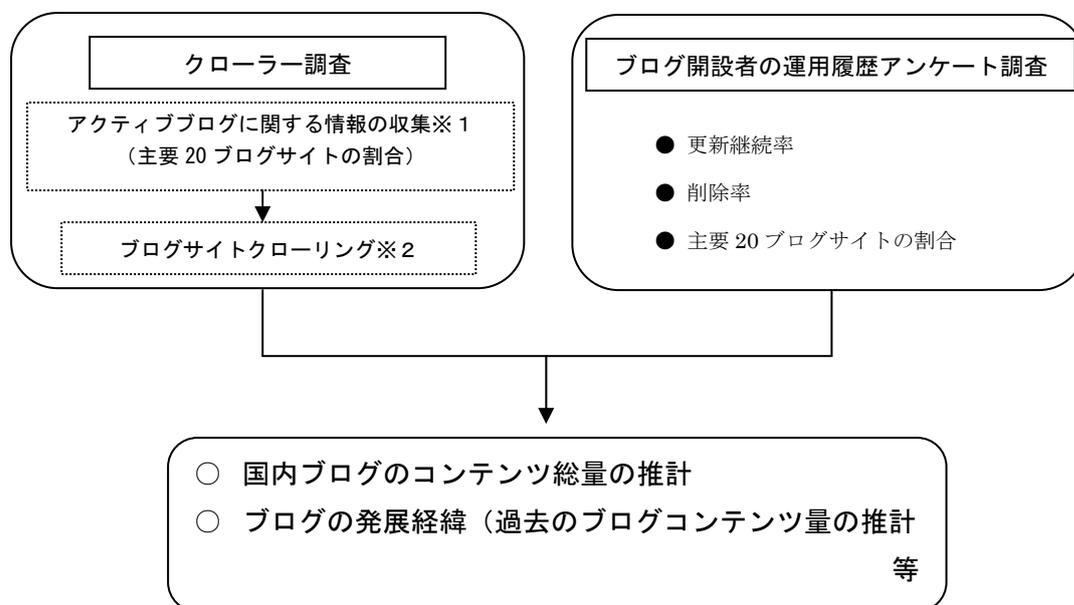
2-2-1 実施調査の手順

ブログのコンテンツ量を 2-1-3 で説明したモデルで推計するためには、現在のアクティブブログに関する情報、更新継続率、削除率等のデータが必要である。本調査では、クローラーによる調査でアクティブブログに関する情報を取得し、ブログ開設者の運用履歴アンケート調査により、更新継続率、削除率等に関するデータを取得することとする（図表 2-2-1）。

なお、アクティブブログに関する情報の取得のため、ブログサービス事業者の提供するすべてのブログサイトをクローラーで悉皆調査することは困難であり、実際には公表登録者数の多いブログサイトのうち、20 を対象としてクローラー調査を行って情報を取得し、これら 20 のブログサイト（以下「主要 20 ブログサイト」という。）を利用するブログが国内全体のブログに占める割合を用いて、国内のブログコンテンツの総量を推計する。

主要 20 ブログサイトのブログが国内のブログに占める割合に関するデータは、ブログの新着記事情報を公開しているpingサーバ⁷又はブログ開設者の運用履歴アンケート調査より得る。

図表 2-2-1 調査の構成



※1 主要 20 ブログサイトと ping サーバを対象

※2 主要 20 ブログサイトを対象

⁷ pingサーバとは、ブログ開設者がブログに新しい記事を書き込みした際、その所在等の情報を提供できる公開サーバである。ブログ開設者は新着記事情報をpingサーバに提供することにより新着記事の存在を広く知らせることができる。

2-2-2 クローラー調査手順と結果

(1) クローラー調査手順

本調査では、アクティブブログに関するデータの情報収集を行うが、クローラーによる情報収集は、新着記事情報の取得とブログコンテンツ構成ファイル等の収集の2つの手順に大別される（図表 2-2-2、2-2-3）。

[新着記事情報の取得]

① 主要 20 ブログサイトの新着記事情報の取得

あらかじめ選定した主要 20 ブログサイトが常時提供している新着情報ページをクローラーで 1 箇月間、定期的に取り得し、掲載されている新着情報を収集する。

② ping サーバからの記事更新情報の取得

主要 ping サーバが公開しているブログの新着記事情報を 1 箇月間、定期的に取り得し、ブログの更新情報を収集する。ping サーバへの新着記事情報提供はブログ開設者の任意なので、アクティブブログ全ての情報を収集することはできないが、主要 20 ブログサイト以外のブログサイトのブログや、開設者が自主設置したサーバでのブログの更新情報を収集することができるため、国内ブログ全体の中での主要 20 ブログサイトのブログの割合を推定するデータとする（図表 2-2-4）。

③ アクティブブログリストの作成

①、②で取得した新着記事情報を解析してブログごとに整理し、調査期間の 1 箇月間に記事の更新が行われたアクティブブログのリストを作成する。

④ 各ブログの RSS 情報の取得

調査期間の 1 箇月間の記事の更新データとして、③で随時作成・更新しているアクティブブログリストを用い、各ブログの RSS 情報をクローラーにより取得する。取得したデータは、更新日時の集計・分析に用いる。

[ブログコンテンツ構成ファイル等の収集]

⑤ 月次表示ページ⁸の HTML ファイルの取得

③で作成したアクティブブログリストの中から、主要 20 ブログサイトのアクティブブログ全体から 5% をランダムサンプリングで抽出し、各ブログのトップページからリンクされている URL を辿っていく方法でクローリングを実施して、ブログ開設月から 2008 年 1 月までの月次表示ページの HTML ファイルを取得する。さらに、取得したフ

⁸ ブログの記事の表示形式はブログサイトにより異なるが、過去の記事については、該当年月に投稿された複数の記事をまとめて閲覧することができる。年月単位にまとめられた複数の記事の表示ページを月次表示ページと呼ぶ。

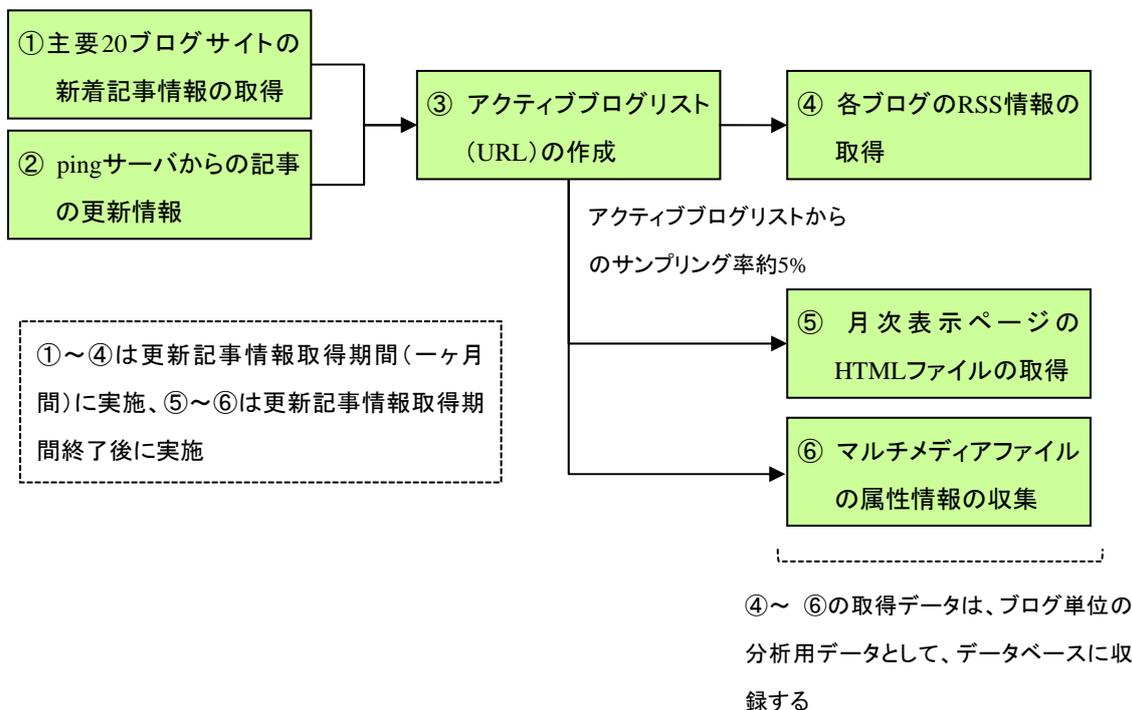
ファイルの内容を解析し、各ページに含まれる記事数、コメント数、トラックバック数などを判定して、これらの情報をデータベースに収録する。

また、月次表示ページの HTML ファイルを過去にさかのぼり、初めての記事が存在した年月をブログ開設月とし、現在のアクティブブログの開設時期分布を作成するための情報として用いる。

⑥ マルチメディアファイルの属性情報の収集

⑤で取得した HTML ファイルからマルチメディアファイル（画像ファイルおよび音声/動画ファイル）の掲載数を解析した後、月次表示形式の記事表示ページを再度クロールし、マルチメディアファイルのデータ量等、必要な属性情報のみを収集する。収集した属性情報は、各月次表示ページごとに整理し、データベースに収録する。

図表 2-2-2 調査で取得・収集するデータと収集の流れ

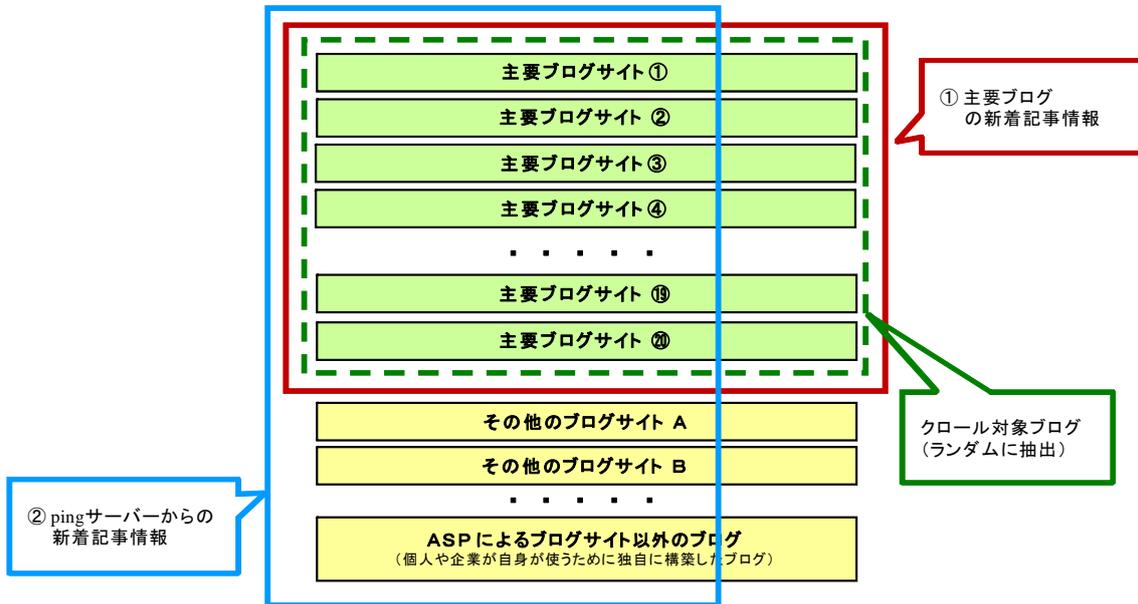


図表 2-2-3 調査で収集・取得するデータ

取得データ	取得方法	目的・用途
① 主要 20 ブログサイトの 新着記事情報	更新記事情報取得期間中 (一ヶ月間)、主要 20 ブログサイトを観測し、新着情報ページの RSS または HTML ファイルから取得する。	各ブログのクロールを実施するために必要となるアクティブブログリストを作成するため。
② ping サーバからの 記事の更新情報	更新記事情報取得期間中 (1 箇月間)、ping サイトを観測し、新規のエントリー情報を取得する。	各ブログのクロールを実施するために必要となるアクティブブログリストを作成するため。
③ アクティブブログ リスト (URL)	取得した①②のデータをもとに、主要 20 ブログサイト毎にアクティブブログリスト (ユニークブログのトップページの URL) を作成する。	作成したアクティブブログリストの URL を用い、ブログ毎に④～⑥のクロールを実施する。
④ 各ブログの RSS 情 報	更新記事情報取得期間中 (1 箇月間)、③で作成したアクティブブログリストを利用し、クロールにより取得する。	更新記事情報取得期間における更新状況の集計・分析に用いる。 <ul style="list-style-type: none"> 記事数 更新日時

<p>⑤ 月次表示ページの HTML ファイル</p>	<p>③で作成したアクティブブログリストから対象をサンプリング抽出し、URL を辿っていく方法でクロールを実施して、ブログ開設月～2008 年 1 月までの月次表示ページの HTML ファイルを取得する。</p>	<p>ブログ開設月～2008 年 1 月までのエントリーの状況に関する集計・分析に用いる。</p> <ul style="list-style-type: none"> • 記事数 • コメント数 • トラックバック数 • マルチメディアファイル数
<p>⑥ マルチメディアファイルの属性情報</p>	<p>⑤で取得した HTML ファイルの解析結果と合わせ、月次表示形式の記事表示ページを再度クロールし、マルチメディアファイルの容量等必要な情報のみ収集する。</p>	<p>マルチメディアファイルの利用状況の集計・分析に用いる。</p> <ul style="list-style-type: none"> • マルチメディアファイル容量

図表 2-2-4 pingサーバによる主要 20 ブログサイトのブログの割合の推定モデル



(2) クローラー調査結果

ア) 新着記事情報の取得結果

2008年1月から2月にかけて1箇月間、主要20ブログサイトを対象に新着記事情報を取得して、ブログごとに整理してリストを作成した結果、確認されたアクティブブログ数は約221万であった(図表2-2-5)。

なお、pingサーバへの送信は任意のため、すべての情報を網羅しているわけではないが、主要20ブログサイト以外を利用するブログの新着情報も集まるpingサーバで確認されたアクティブブログ数は約128万、そのうち主要20ブログサイトを利用していると確認されたアクティブブログ数は約99万となっており、主要20ブログサイトを利用するアクティブブログの割合は77.5%であった。

図表 2-2-5 新着記事情報の収集結果

実施時期： 2008年1月20日～2月18日 (30日間)

A..主要20ブログサイト

更新確認アクティブブログ数	2,213,580
更新確認記事数(件)	22,043,689

B.pingサーバ

更新確認アクティブブログ数	1,283,344
うち、主要20ブログサイトを利用するもの	994,374
主要20ブログサイトを利用するものの占める割合	77.5%

イ) ブログコンテンツ構成ファイルの収集結果

確認できた主要20ブログサイトのアクティブブログの約5%にあたる109,949のブログをランダムサンプリングで抽出し、クローラーで収集したファイル等を分析・集計した結果、情報収集対象となった個々のブログの開設時から2008年1月までの累積記事数は合計約2,031万件であった(図表2-2-6)。

また、2008年1月現在のアクティブブログの開設時期別分布は図表2-2-7のようになっている。

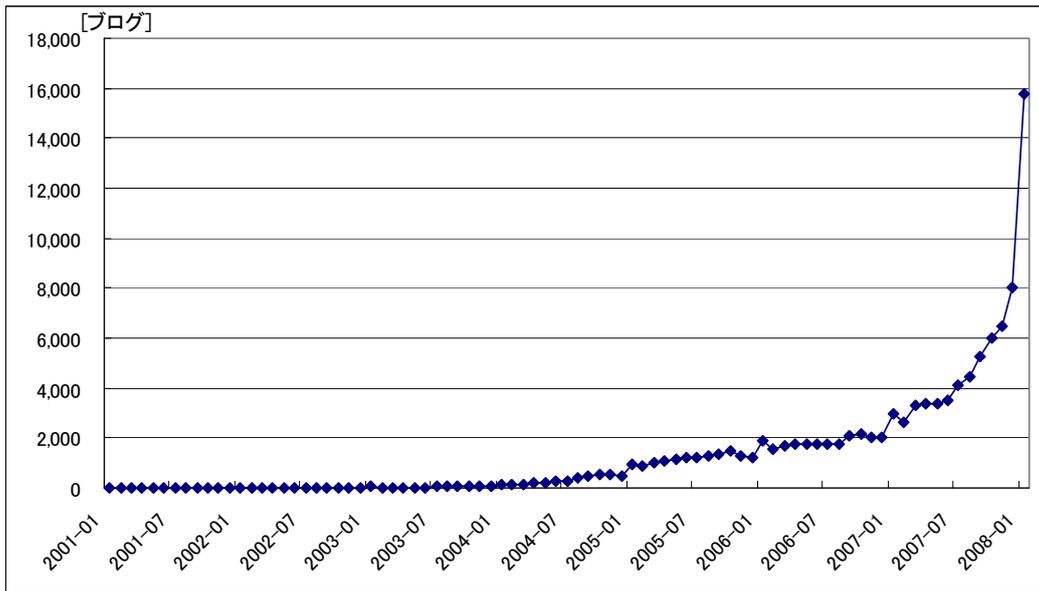
図表 2-2-6 ブログコンテンツ構成ファイル等の収集結果

実施時期：2008年2月21日～3月3日

調査対象のブログ数	109,949
※主要20ブログサイトのアクティブブログ数の5%	
確認記事数(件)	20,314,454

※個々のブログの開設時から 2008 年 1 月までの累計	
確認マルチメディアファイル数	11,105,794
<ul style="list-style-type: none"> ● 画像ファイル数 ● 音声/動画ファイル数 	11,102,111
	3,683
※個々のブログの開設時から 2008 年 1 月までの累計	

図表 2-2-7 2008 年 1 月現在のアクティブブログの開設時期別分布



2-2-3 ブログ開設者の運用履歴アンケート調査の概要と結果

(1) アンケート調査の概要

ブログのコンテンツ量を 2-1-3 で説明したモデルで推計するためには、ブログの更新継続率、削除率等を算出するためのデータが必要であるが、これらの情報はクローラー調査では取得できない。そこで、ブログ開設者経験者を対象とするアンケート調査を行い、ブログの開設時期、継続更新期間、削除時期などの運用履歴に関する設問に対する回答を元に、更新継続率等の推定値を算出する⁹。

なお、本調査は第 3 章のブログ開設者の意識調査のために実施するアンケート調査と併せて行っている。

図表 2-2-8 開設者履歴アンケート調査の概要

調査方法	調査会社登録パネルに対するオンラインアンケート方式
調査実施期間	平成 20 年 2 月 15 日（金）～18 日（月）の 4 日間
回答件数	2,351 件
調査対象	15 歳以上のブログ開設経験者（ブログを現在利用している者及び利用したことがある者）
割付方法	「平成 18 年通信利用動向調査」において、情報発信を行っていると回答した者の男女年代別比率の割付に基づいて回答を得た
主な調査項目	開設ブログごとの利用状況（利用ブログサービス、開設年月、更新停止年月、削除年月等）

(2) 推定値算出の考え方と結果

更新継続率と削除率については、ブログ利用者層の拡大とともに変動していることが予想されたため、時期別に算出した。また、主要 20 ブログサイトを利用するブログの割合については、アンケート結果を見る限り、長期にわたって安定しており、大きな変動は見られないことから、一定と仮定して推計を行うこととする。

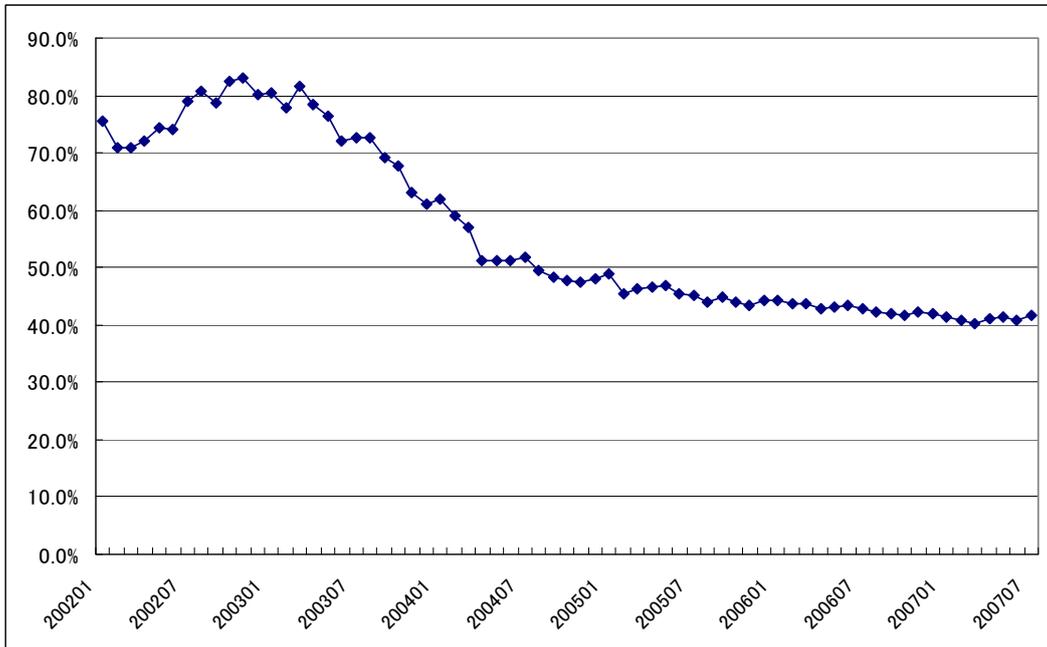
ア) 更新継続率の推定値

ブログの開設時期、更新停止年月等のアンケート回答データを用いて、各月開設ブログの 12 箇月更新継続率（ある月に開設されたブログが 12 ヶ月後にもアクティブである割合）を開設月別に計算した結果が図表 2-2-9 のグラフである。

これによると 2003 年頃までの、早い時期に開設されたブログでは、12 ヶ月更新継続率は 7 割前後と高いが、2004 年にかけて 5 割近くまで低下し、その後は 5 割から 4 割の間で推移している。

⁹ 推定の補正については、参考 2 を参照。

図表 2-2-9 12ヶ月更新継続率の推移



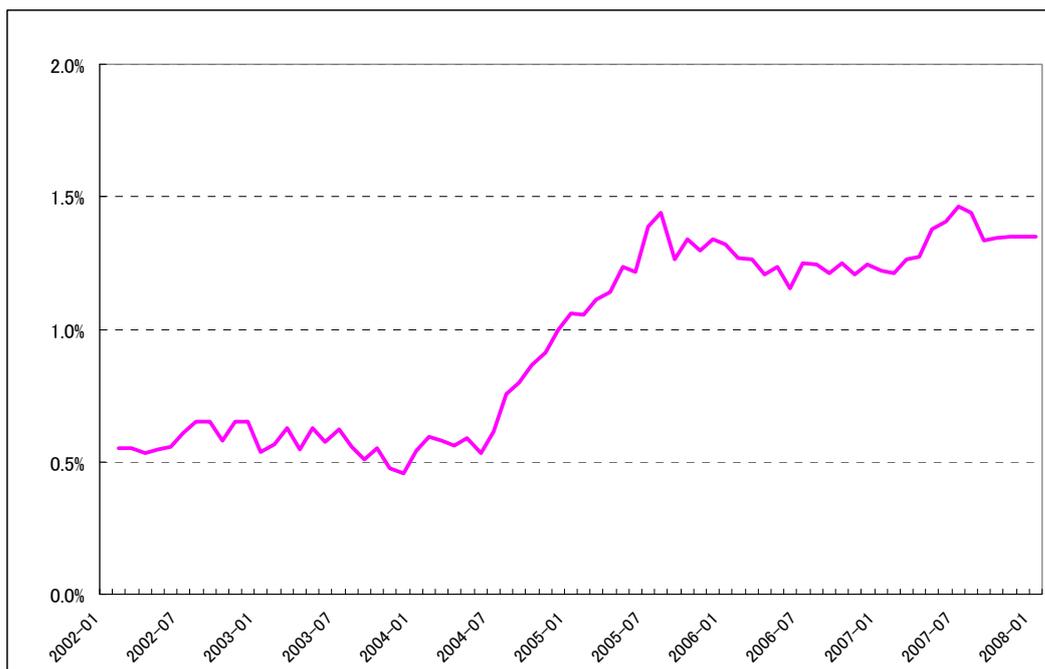
注：補正を行った更新継続率の13か月移動平均値を示した。

イ) 削除率の推定値

過去の各月にアンケート回答者がネット上で公開していたブログ数と、アンケートで回答された各月の削除ブログ数をもとに、各月のブログ削除率（ある月にネット上に存在したブログが、当月中に削除される割合）を算出したものが図表 2-2-10 である。

これによると 2004 年半ばまで削除率は 0.5%で推移していたが、2004 年後半から 2005 年にかけて 1.5%近くまで上昇し、その後おおむね横ばいで推移している。

図表 2-2-10 ブログ削除率の推移



注：アンケート回答から算出したブログ削除率の13か月移動平均値を示した。

ウ) 主要 20 ブログサイトの割合の推定値

ブログの開設時期と利用ブログサイトのアンケート回答を元に、ブログ開設年別の主要 20 ブログサイトの利用割合を算出したのが図表 2-2-11 である。主要 20 ブログサイトのサービスが出揃ったのは 2005 年であるため、それ以前のデータは参考値であるが、おおむね 80%前後で推移している傾向がみられる。

図表 2-2-11 主要 20 ブログサイトの利用割合

情報源	2001年	2002年	2003年	2004年	2005年	2006年	2007年	2008年1月
pingサーバの新着記事情報								77.5%
ブログ開設者アンケート	80.7%	79.6%	79.9%	80.2%	82.3%	83.9%	85.3%	85.8%

注：ブログ開設者アンケートでは、ブログ開設時期別に主要 20 サイトのシェアを集計した。

上記のとおり、2008 年 1 月における主要ブログ 20 サイトの利用割合については、アンケート調査結果 (85.8%) と、ping サーバが提供する新着記事情報から算出した数値 (77.5%) に差がみられる。本アンケート調査は、個人ユーザのみを対象としている点に注意が必要である。現在では、ブログは企業や NPO 等のサイトでも幅広く開設されるようになっており、その多くはブログサービス事業者のブログサイトを利用せず、ブログソフトウェアを独自にサーバにインストールして開設したり、あるいはサーバホ

스팅サービスで提供されるブログ機能を用いたブログと考えられる。これらを考慮すると、国内のブログ全体に占める主要 20 ブログサイトのシェアは、アンケート調査結果の数値（85.8%）より若干低くなると考えられることから、ping サーバで実測した数値（77.5%）を利用することとする。

2-3 ブログのコンテンツ量推計結果

本節では、2-2 の実施調査で得たデータを利用して、2-1-3 のモデルにより推計した 2008 年 1 月現在のブログのコンテンツ量と、2001 年 1 月から現在までのコンテンツ量の推移についてまとめる¹⁰。

2-3-1 国内のブログコンテンツ量（2008 年 1 月現在）

推計結果によれば、2008 年 1 月現在でインターネット上で公開されている国内ブログの全体の数は約 1690 万、記事数は約 13 億 5000 万件、データ量は約 42 テラバイトであった（図表 2-3-1）。データ量のうち、画像・動画ファイルのデータ量を除いたテキスト情報のみのデータ量は約 12 テラバイトであり、このデータ量は書籍 1 冊の原稿の情報量の約 2700 万冊分に相当する¹¹。

1 箇月に 1 回以上記事が更新されているアクティブなブログは約 300 万件と推計され、インターネット上で公開されている国内ブログ全体数の 2 割弱（18%）が、継続的に更新されているブログであるとみることができる。

また、上記のインターネット上に現存するブログの他に、過去に開設されたがネット上から削除されたブログもある。これらの削除ブログを含め、2001 年 1 月以降、これまでに開設された累計では、ブログ数は約 2240 万、記事数は約 17 億 9000 万件、データ量は約 54 テラバイトに達すると推計される。

図表 2-3-1 国内のブログコンテンツ量（2008 年 1 月現在）

	ブログ数 [万ブログ]	記事数 [百万件]	データ量 [テラバイト]	テキストデータ量 [テラバイト]
インターネット上に公開されているブログ	1,690	1,347	42	12
うちアクティブブログ	308	568	20	5
過去に開設されたブログ（削除済みを含む）	2,243	1,787	54	16

¹⁰ 更新継続率、削除率、主要 20 ブログサイトの利用割合などの推定値の設定により、いくつかのケースが想定されるが、各ケースの場合における推計値については参考 3 を参照。

¹¹ 書籍 1 冊の平均ページ数を 260 ページ、1 ページ当たり平均文字数を 900 文字、1 文字 2 バイトとして算出した書籍 1 冊あたりのデータ量 457 キロバイトを用いて換算。

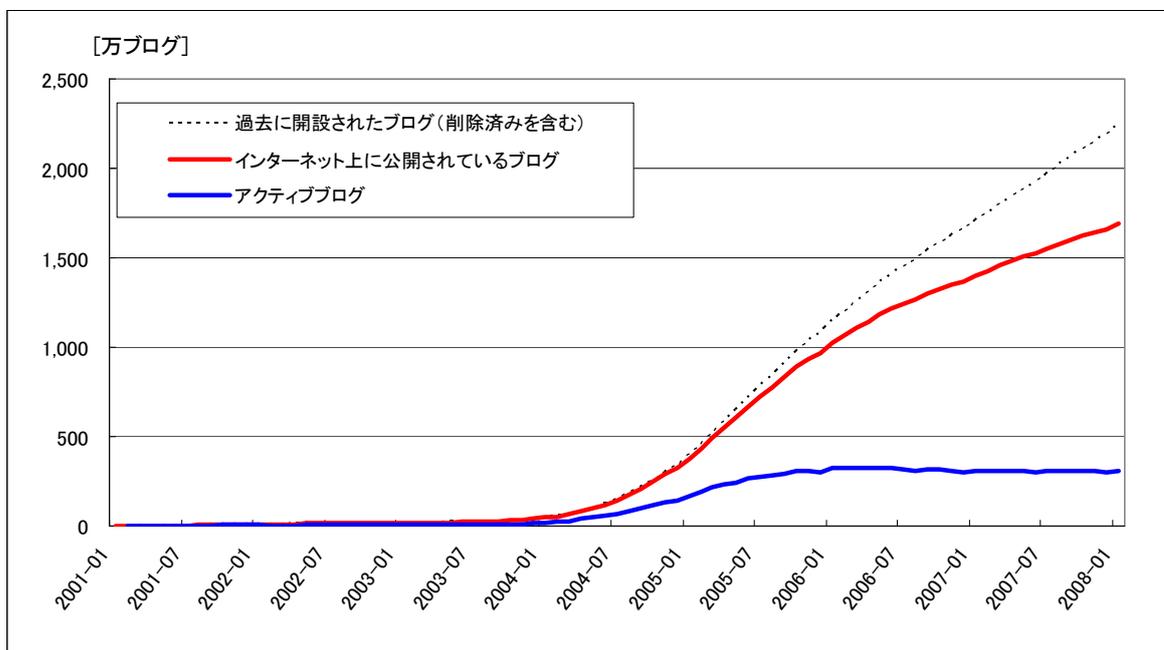
2-3-2 ブログコンテンツ量の推移

(1) ブログ数の推移

ア) 国内ブログ数の推移

インターネット上で公開されているブログ数の推移の推計結果を見ると、2003 年前半までは数万から十数万の規模で推移していたが、2004 年から 2005 年にかけて急増し、2006 年 1 月に 1,000 万に達した後も緩やかに増加していることがわかる（図表 2-3-2）。

図表 2-3-2 国内のブログ数の推移

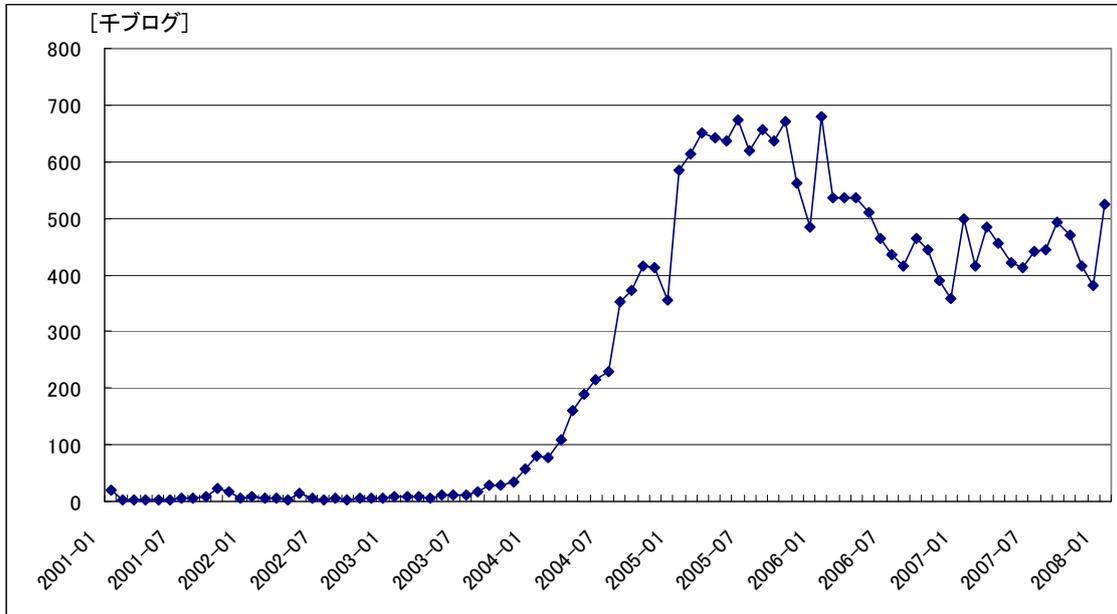


イ) 新規開設ブログ数の推移

過去の各月に新規に開設されたブログ数の推移の推計結果を見ると、主要なブログサイトがサービスを開始した 2003 年の後半から急速に増加し、さらにタレントやいわゆるアルファブロガーによる有名ブログの登場による認知度の向上等も相まって、2005 年はじめから 2006 年のはじめにかけて、月間新規開設ブログ数が概ね 60 万を超えるピーク期となった（図表 2-3-3）。その後は 2006 年にかけて、いったん新規ブログ開設数が月間 40 万程度まで減少したものの、2007 年中も、引き続き 40 万から 50 万程度の水準で推移している。

また、新規開設ブログ数の推移を月別に見ると、毎年 1 月にはブログ開設数が急増し、その後、夏期には横ばいないし減少傾向となるが、9 月か 10 月頃に一時的に増加し、年末にかけて、また減少するという傾向がある。

図表 2-3-3 毎月の新規開設ブログ数の推移

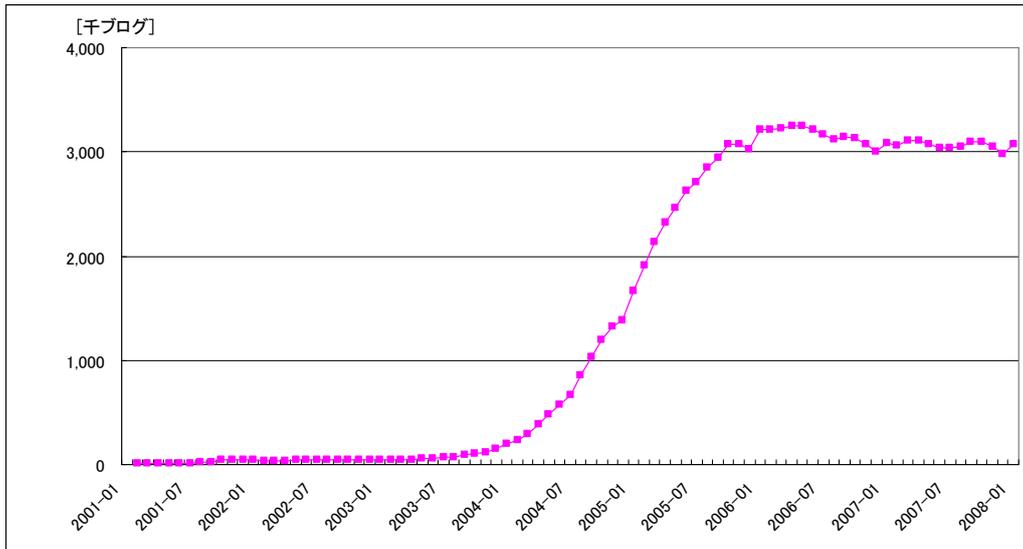


ウ) アクティブブログ数の推移

毎月 1 回以上の継続的な更新が行われているアクティブブログ数の推移の推計結果をみると、2003 年後半から 2005 年にかけて急増し、2006 年前半は約 320 万を超える高い水準となり、その後も 300 万程度でほぼ横ばいで推移している(図表 2-3-4)。2005 年前後の、ブログの認知度向上によるブームの時期を過ぎても、アクティブブログ数はほとんど減少せず、ピーク時と変わらない規模を維持している状態といえる。

ただし、これらのアクティブブログは常に同じものというわけではなく、新規開設で新たに加わるブログがある一方、逆にアクティブブログから休眠状態へ移行するブログ等もあり、常に入れ替わっている。前節で見たとおり、ある時点でアクティブだったブログのうち、1 年後にも継続的に更新され続けている割合は、最近では 4 割から 5 割程度で、残りは休眠状態になったり、削除されていくと推測される(図表 2-2-9 参照)。

図表 2-3-4 アクティブブログ数の推移



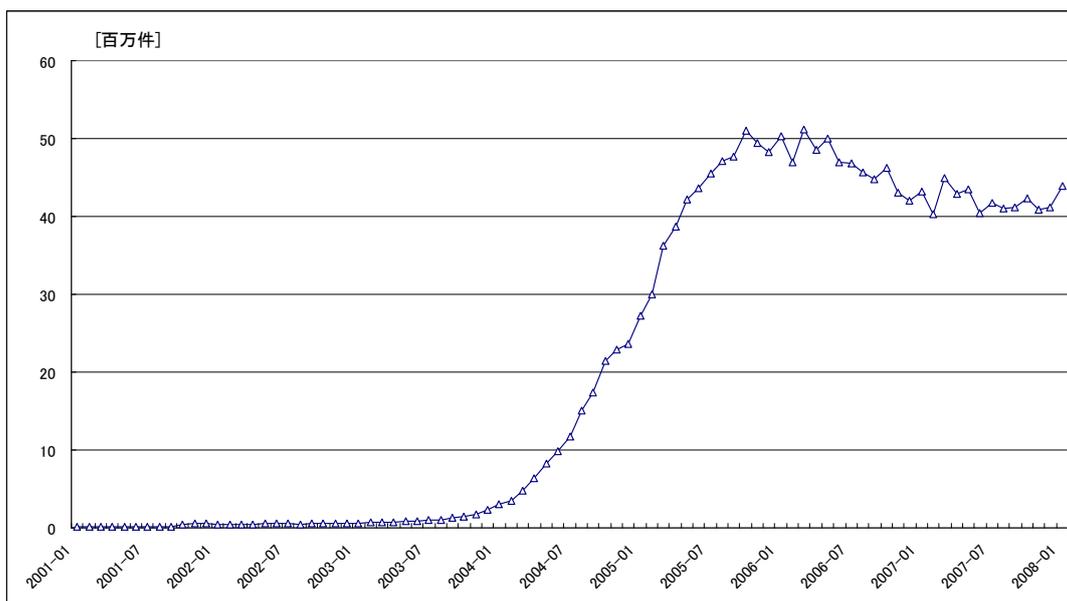
(2) ブログの記事数とデータ量の推移

ア) ブログの新規記事数とデータ量の推移

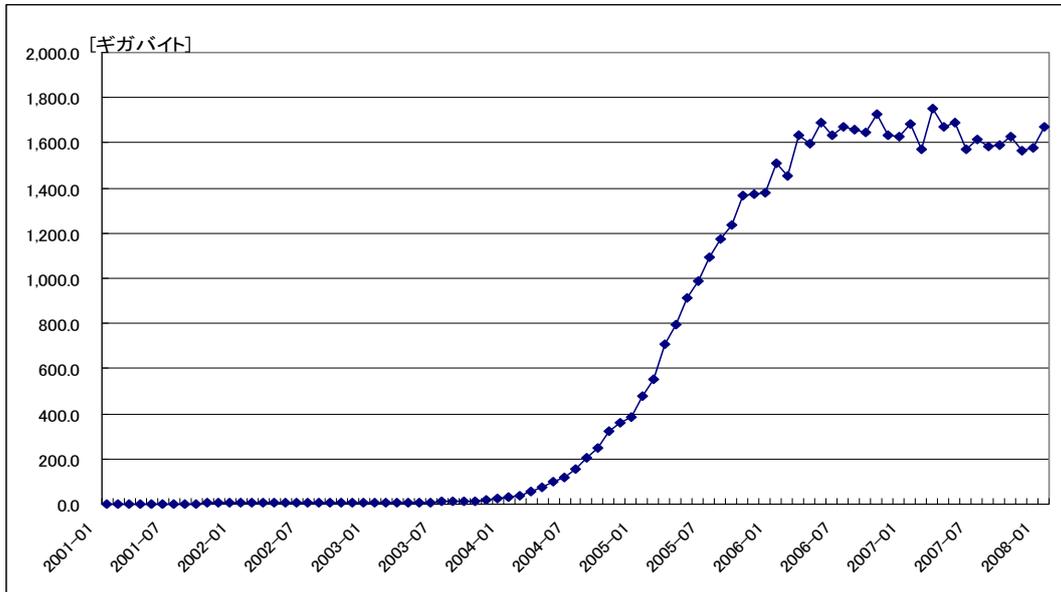
毎月新たに書き込みされる記事数の推移の推計結果を見ると、2005 年後半から 2006 年前半にかけてのピーク時には、毎月 5,000 万件前後の記事が書き込みされている（図表 2-3-5）。その後はやや減少したものの、2007 年中も月間平均 4,000 万件以上の記事が書き込みされ続けており、極めて活発な情報発信が続いている。

また、毎月新たに書き込みされる記事のデータ量の推移の推計結果を見ると、2006 年以降は毎月 1.6 テラバイト前後のデータが追加されている（図表 2-3-6）。

図表 2-3-5 毎月の新規記事数の推移



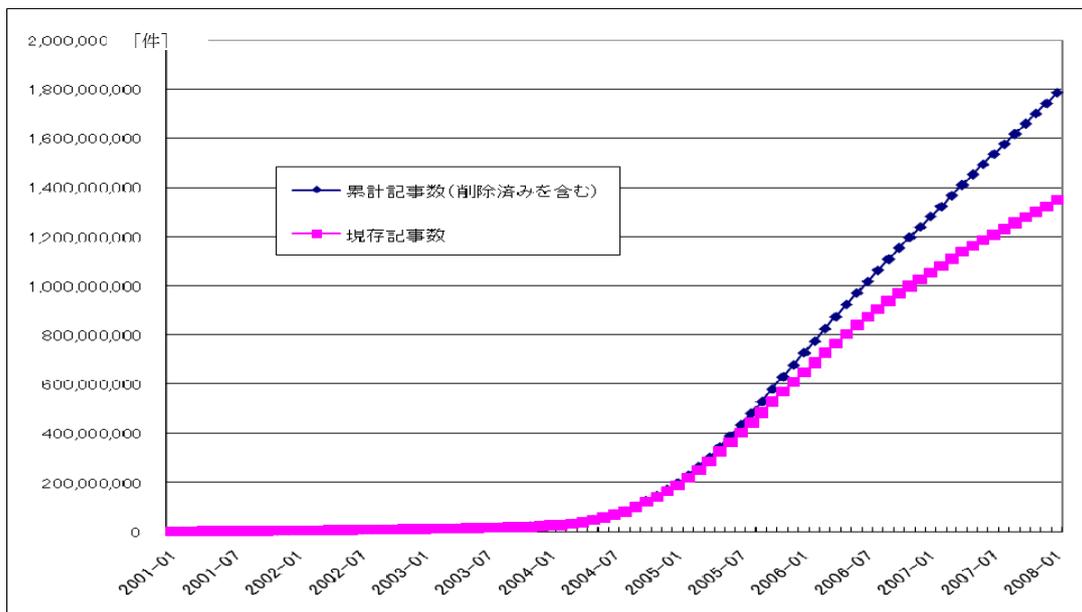
図表 2-3-6 毎月の新規記事データ量の推移



イ) ブログの記事数とデータ量の推移

活発な記事の書込みの結果、ブログ全体での記事数、データ量は増加を続けている。インターネット上で公開されているブログの記事数の推移の推計結果を見ると、2004 年から 2005 年半ばにかけて特に急速に増加し、その後はやや増加ペースが落ちたものの、増加を続けている（図表 2-3-7）。

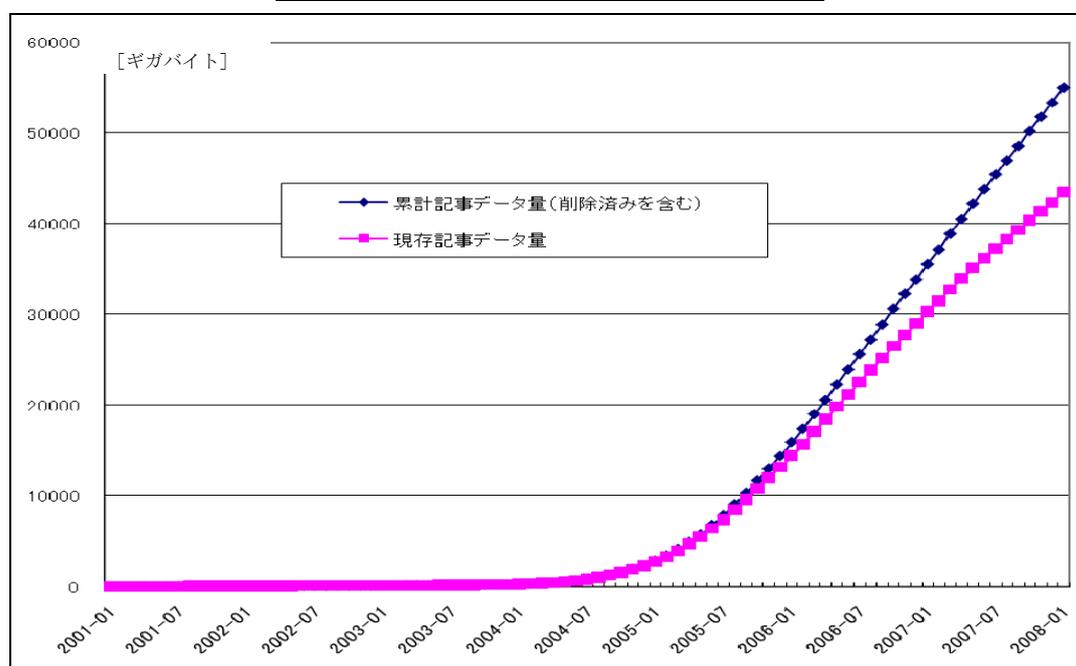
図表 2-3-7 国内ブログの記事数の推移



ブログのデータ量は、ブログ数の増加、記事数の増加、そして後述する画像ファイル等の増加によるデータ量の増加が重なった結果、急速な増加を続けている。

2008年1月までに国内のブログに書き込みされた記事の情報量の合計は約54テラバイトに達するが、ブログ全体のデータ量は、2004年から増加に拍車がかかり、2006年以降は、毎月1.6テラバイト前後のデータが追加されている（図表2-3-8）。

図表 2-3-8 国内ブログのデータ量の推移



2-4 ブログコンテンツの種類と付加機能の状況等

本節では、クローラー調査で収集した主要 20 ブログサイトのコンテンツデータを集計・分析した結果についてまとめる。

2-4-1 アクティブブログの平均像

一般的なブログは、日付で整理されている複数の記事を主な内容としている。ブログ開設者が書き込みをする記事ごとのコンテンツには、文章部分のテキスト情報と、文章に関連する写真等のマルチメディアファイルがある。また、記事には閲覧者又は開設者によるコメントやトラックバックなど、ブログの付加機能を活用した情報が加えられることが多い。

クローラーにより取得したアクティブブログのデータを集計した結果、2008年1月の1箇月間に書き込みされた記事数と、1記事当たりの平均ファイル数、コメント数、トラックバック数は図表 2-4-1 の通りである。

図表 2-4-1 アクティブブログの平均的な記事数等（平成 2008 年 1 月の 1 箇月間）

1 ブログ当たり平均記事数	14.2
1 記事当たり画像ファイル数	0.60
1 記事当たり音声/動画ファイル数	0.00017
1 記事当たりコメント数	1.27
1 記事当たりトラックバック数	0.05

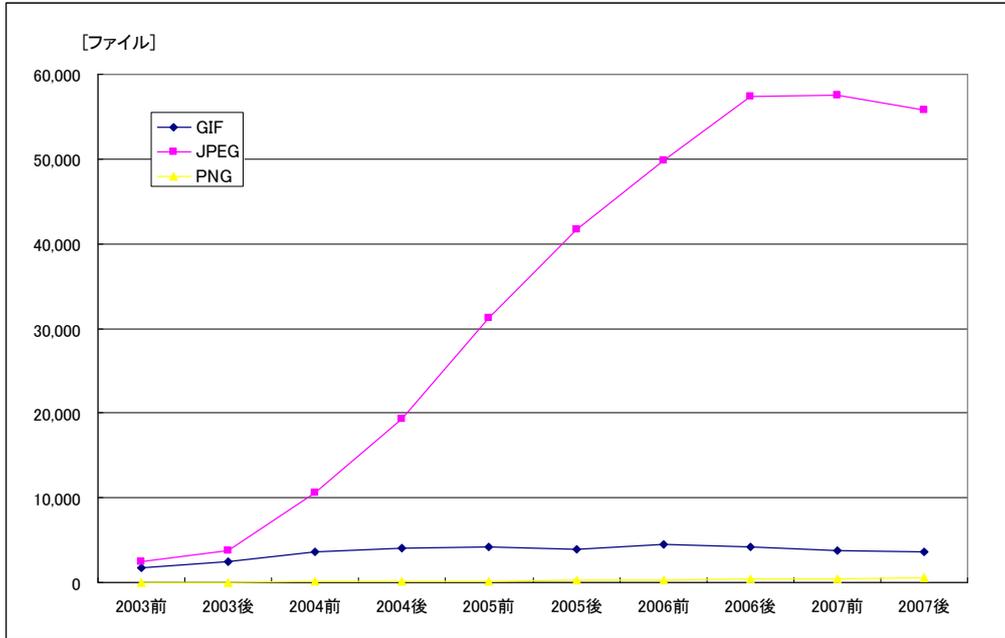
2-4-2 マルチメディアファイルの推移

ブログはもともと文字等のテキスト情報がコンテンツの中心だったが、ブロードバンド環境が広く普及したこと等により、写真画像などのファイルがブログに掲載されることが多くなっている。クローラーにより取得したアクティブブログのデータの集計結果から、ブログの画像ファイル、音声/動画ファイルの掲載状況を経年別に見ると、2004 年前半から 2006 年にかけて急速に増加している（図表 2-4-2、2-4-3）。

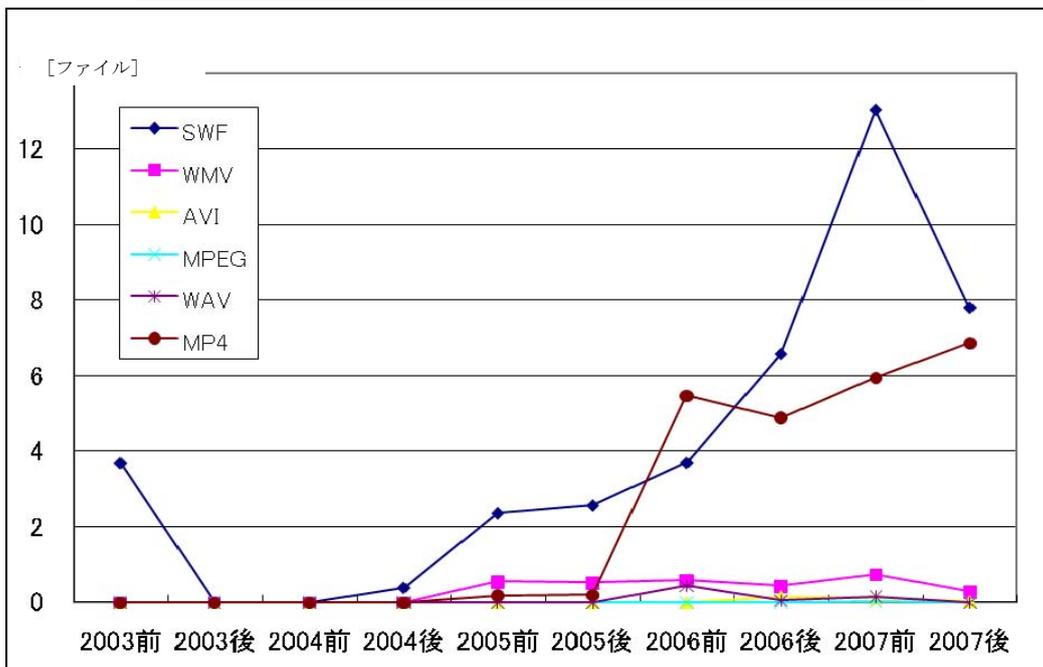
画像ファイルについて見ると、主に写真をアップロードする際に用いる JPEG ファイルは、2006 年前半には一つの記事当たりの平均掲載数は 0.5 を超え、増加傾向が顕著である。一方、アイコン等の軽い画像の掲載に用いることが多い GIF ファイルの平均掲載数については、2004 年から 0.036 と、100 記事当たり 3.6 程度で横ばいである。

音声/動画ファイルについて見ると、ブログサービス事業者によっては掲載を制限していることもあり、増加傾向はあるものの、画像ファイルに比べると掲載数は少ない。

図表 2-4-2 記事 10 万件当たりの画像ファイル掲載数



図表 2-4-3 記事 10 万件当たりの音声/動画ファイル掲載数

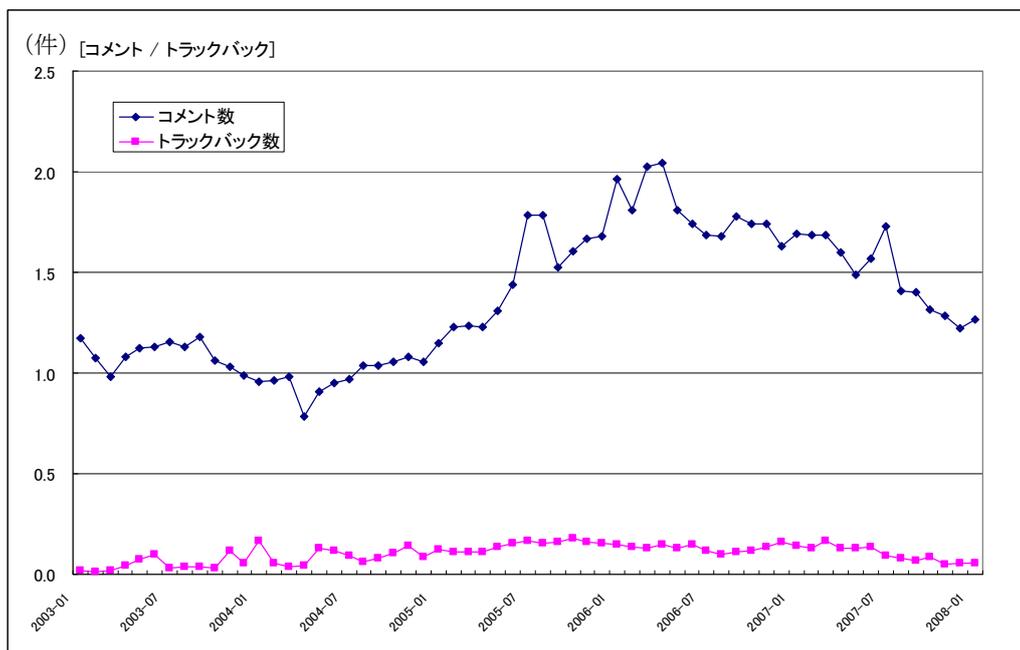


2-4-3 コメント数、トラックバック数の推移

クローラーにより取得したアクティブブログについて、1記事当たりのコメント数を見ると、2003年以降、概ね1.0から2.0の範囲で推移し、1記事当たりのトラックバック数を見ると、概ね0.1前後で推移している（図表2-4-4）。

コメント・トラックバックは、ブログ普及期から近年でも、閲覧者と開設者のコミュニケーション機能として活用されていることが伺える。

図表2-4-4 記事当たりのコメント数およびトラックバック数



2-4-4 記事の更新曜日・時間

2008年1月の1箇月間に更新された記事の更新日時を分析すると、曜日別の更新比率は日曜が15.0%ともっとも高く、金曜が13.7%ともっとも低くなっており、適度に分散している状況が伺える（図表2-4-5）。同様に時間帯別の更新比率では、5.0%を超える時間帯は18時台～0時台であり、23時台がピークとなっていることから、夜から深夜にかけての余暇の時間帯に更新している状況が伺える。

図表 2-4-5 曜日別・時間帯別の更新記事数シェア

更新曜日	比率	更新時間	比率
月	14.4%	0時台	6.0%
火	14.1%	1	3.9%
水	14.3%	2	2.6%
木	14.3%	3	1.8%
金	13.7%	4	1.3%
土	14.1%	5	1.3%
日	15.0%	6	1.6%
合計	100.0%	7	2.4%
		8	2.8%
		9	3.1%
		10	3.6%
		11	3.8%
		12	4.1%
		13	3.9%
		14	3.8%
		15	4.0%
		16	4.3%
		17	4.8%
		18	5.1%
		19	5.1%
		20	6.1%
		21	7.2%
		22	8.1%
		23	9.4%
		合計	100.0%

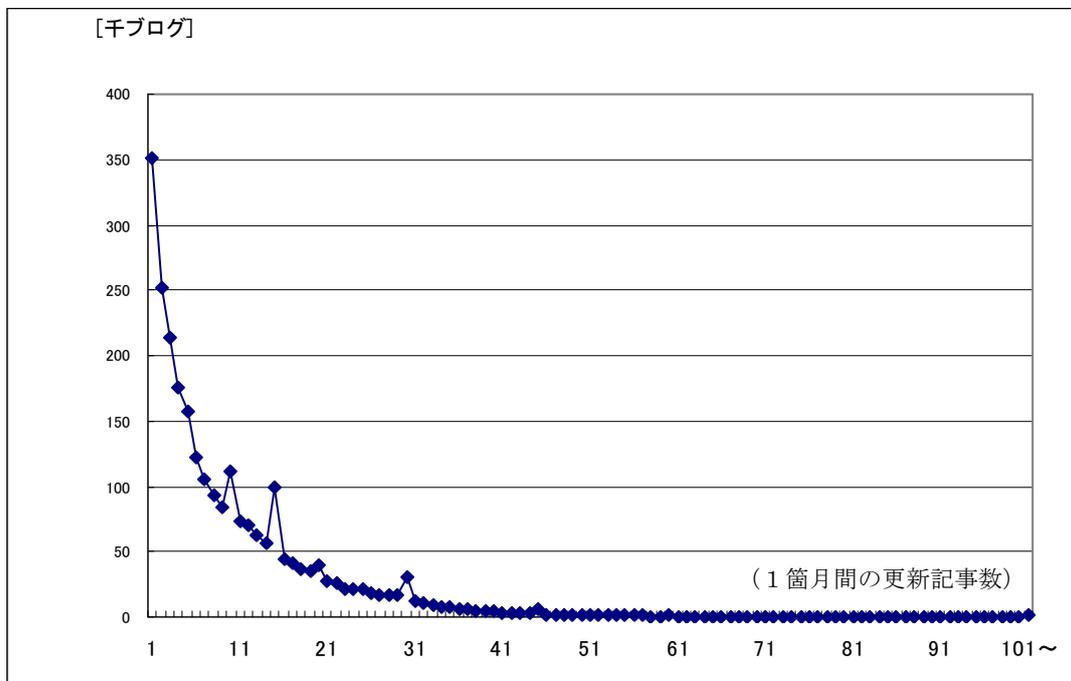
2-4-5 更新記事数によるブログの分布

2008年1月の1箇月間に更新された記事数によるブログの分布を見ると、月間更新記事数1～10のブログが格段に多く、更新記事数が増えるごとにブログ数は逡減している（図表2-4-6）。

月間更新記事数が50を超えるブログも見受けられるが、これらについては、1日に何度も更新する熱心なブログと、後述する機械的に記事を大量に更新するいわゆる「ス

パムブログ」が混在していると考えられる。

図表 2-4-6 1 箇月間の更新記事数によるブログの分布



第3章 ブログ開設者の意識調査

3-1 調査の目的と概要

3-1-1 調査の目的と仮説設定

ブログ開設者の意識調査では、ブログの利用状況、開設目的・想定する読者、ブログの内容などについて把握するとともに、ブログの開設において重要な要因の抽出や、ブログ開設要因の傾向によるブログ開設者のグループ分けを行うことを目的とする。

ブログ開設者の意識を明らかにするため、調査実施に先立って、ブログ開設者の利用スタイルを分類するための枠組みを仮説として設定し、仮説をアンケート調査によって検証するという方法をとった。本項ではアンケート調査で検証する仮説の詳細を説明する。

ブログ開設者の利用スタイルの仮説設定にあたり、先行研究として川浦・川上らの『ウェブログの心理学』¹²を参考にした。彼らは、ブログ向けのツールが登場する以前の1997年に、個人のホームページのひとつであるウェブ日記を更新するネットユーザに対して、以下の4つのタイプを設定してアンケート調査を実施している。

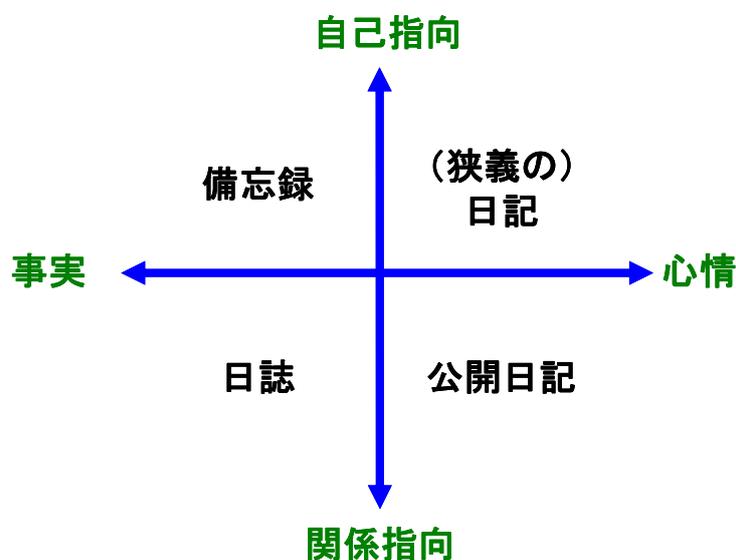
- ① 日々の生活の記録を自分のために覚書として残す（備忘録）
- ② 日々得た情報を他の人に提供できる（日誌）
- ③ 他の人に自分という人間を知ってもらえる（公開日記）
- ④ 自分で自分を理解することができる（狭義の日記）

そして、この4つのタイプと更新動機や効用などの関連要因との関係を分析（正準判別分析）し、ウェブ日記を、「書き手の指向性」と「日記の内容」の二軸によって整理している。書き手の指向性とは、書き手の意識が自分自身に向かっているか、それとも読者ないし読者との関係に向かっているかに関する軸である。また内容とは、書かれていることがらであり、心情と事実に大別される。

川浦・川上らの分析によれば、ウェブ日記の4分類は書き手の指向性と日記の内容という二軸からなる平面上に配置され、図示すると、図表3-1-1のようになる。

¹² 山下清美、川浦康至、川上善郎、三浦麻子 2005年 『ウェブログの心理学』 NTT出版

図表 3-1-1 指向性・内容の二軸によるウェブ日記の位置づけ



本調査では、この分類が、ブログが登場する以前のウェブ日記についてのものではあるが、ブログが広く利用されるようになった現在においても有効であると考え、指向性と内容の2軸を踏襲してブログの利用スタイルに関する仮説を検討した。

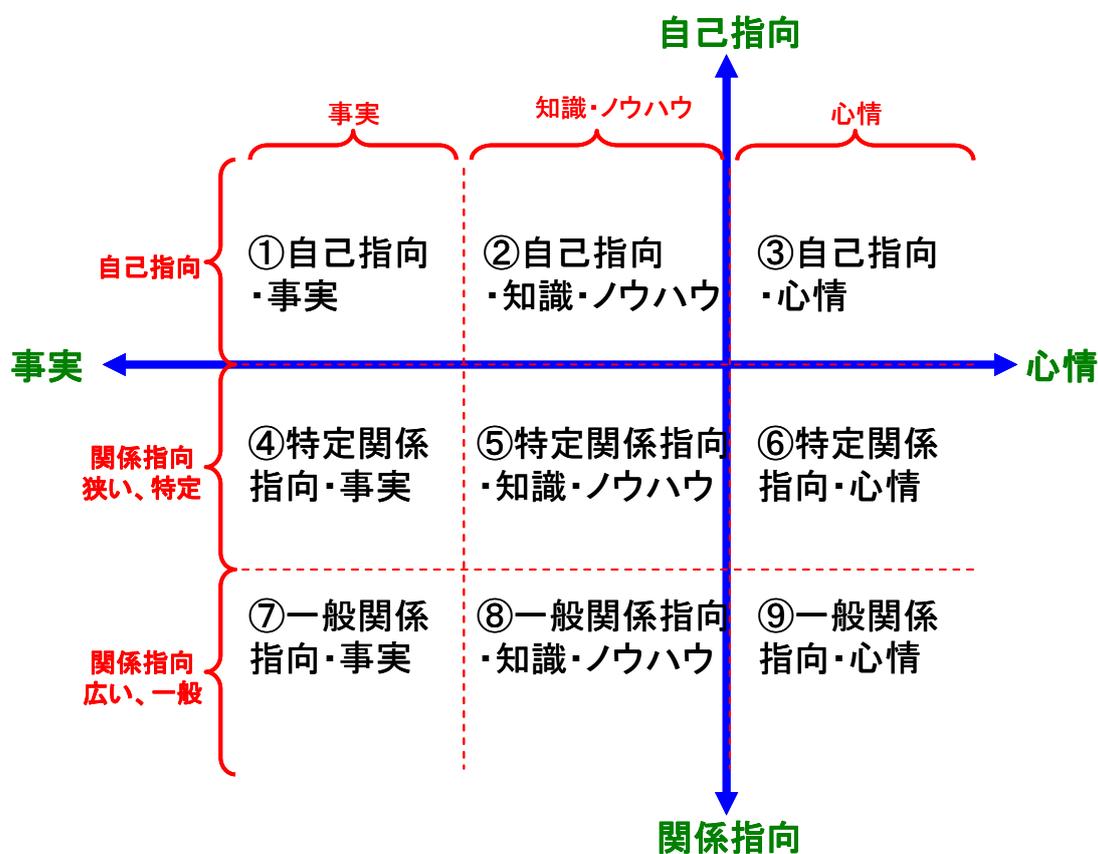
ただし、現在のブログは、1997年当時には考えられなかったような様々な用途・目的で利用されていることも事実であり、このブログ利用の多様性を踏まえて、2軸を以下のように拡張した。

コミュニケーションツールとしてのブログが幅広く利用されるようになった要因・背景の一つとして、インターネットの普及に伴い、家族・友達から、地域コミュニティ、さらには社会全体など想定読者が拡大・多様化していることが考えられる。こうした状況においては、指向性のうち、関係指向に包含される内容が多様になっていると想定される。したがって本調査では、関係指向の細分化を図り、関係指向のうち、読み手として特定の読者や集団を想定しているものを「特定関係指向」、特に具体的な読み手を置かず、世間一般を読み手と想定するものを「一般関係指向」と分類し、「自己指向」と合わせて指向性を3つに分類した。

ブログの内容についてみると、1997年当時と比べて、現在のブログは非常に多様な用途に利用されており、内容も多様化していることが考えられる。特にアフィリエイトの広がりにより、ブログに自分の経験に基づく知識やノウハウを公開する動機も強まっていると考えられる。そこで事実のうち自分の体験に基づいて整理された知識・ノウハウを独立させ、心情と合わせて内容を3つに分類した。

以上の拡張を元に、ブログの利用スタイルに関する仮説を模式図で表現すると、図表3-1-2のようになる。この仮説によれば、ブログの利用スタイルは、指向性の3分類と内容の3分類を掛け合わせた9つに分類できると考えられる。

図表 3-1-2 本調査におけるブログ利用スタイルの仮説



9つに分類した場合、それぞれのブログ利用スタイルで具体的に想定されるブログ利用法を図表 3-1-3 に示す。

図表 3-1-3 9つのブログ利用スタイルで想定される具体的な利用例

ブログ利用スタイル	利用法の例
① 自己指向・事実	世間や身の回りの出来事を自分のために記録しておく。備忘録、作業ログなど
② 自己指向・知識・ノウハウ	自分の知識や、仕事・生活で得たノウハウを、自分の為に記録する。知恵袋、ノウハウメモなど
③ 自己指向・心情	自分の感じたこと、気持ちを自分の為に記述する。内省的エッセイ、反省録など
④ 特定関係指向・事実	友人・コミュニティなどに、共通の関心事項について最新動向などの情報を提供する。好きなタレントの活動情報ブログ、地域生活情報ブログなど
⑤ 特定関係指向・知識・ノウハウ	友人・コミュニティなどに、共通の関心事項についての自分の専門的知識やノウハウを提供する。新米ママさん向け子育て体験記など
⑥ 特定関係指向・心情	友人・コミュニティなどに、共通の関心事項について、感想や思っていることなどを伝える。友人に考えや気持ちを伝えるブログ、地域向けレストラン評価ブログなど
⑦ 一般関係指向・事実	世間や身の回りの出来事を発信する。自分なりの視点でまとめた時事情報ブログ、

実	自分や周囲のニュースをまとめたブログなど
⑧ 一般関係指向・知識・ノウハウ	自分の専門知識やノウハウをわかりやすく体系的に整理し発信する。法律解説ブログや投資指南ブログなど
⑨ 一般関係指向・心情	自分の感じたこと、考え方、意見などを発信する。環境保全等の啓蒙ブログや時事批評ブログ、商品レビューブログなど

3-1-2 アンケート調査の概要

今回のアンケートでは、ブログ利用スタイルに関するこの仮説の検証を軸として、その他インターネットの利用状況や、各種ネットサービスの利用状況、開設するブログの利用状況などを質問した。調査の実施概要は、図表 3-1-4 に示すとおりである。

図表 3-1-4 アンケート調査実施概要

調査方法	調査会社登録パネルに対するオンラインアンケート方式
調査実施期間	平成 20 年 2 月 15 日（金）～18 日（月）の 4 日間
回答件数	2,351 件
調査対象 割付方法	15 歳以上のブログ開設経験者（ブログを現在利用している者及び利用したことがある者）を対象に、「平成 18 年通信利用動向調査」において、情報発信を行っているとは回答した者の男女年代別比率の割付に基づき回答を得た
主な調査項目	<ul style="list-style-type: none">・ 回答者属性（性別、年代、職業、居住地域 等）・ ブログの利用全般に関する質問（ブログ開設・更新の動機、ブログの想定読者と記述内容 等）・ 開設ブログごとの利用状況（利用ブログサービス、開設年月・更新停止年月・削除年月、ブログのテーマ等）

アンケートへの回答数は合計 2,351 件であった。なお、開設ブログ毎の利用状況については、開設したブログの合計やブログ移転の動機などの共通項目を質問した後に、開設したブログ毎にそれぞれの状況について最大 3 つまで回答を得ている。このためブログ毎の質問項目については、最大 4,295 件の回答が得られており、それらの回答数を母数として集計・分析を行っている。

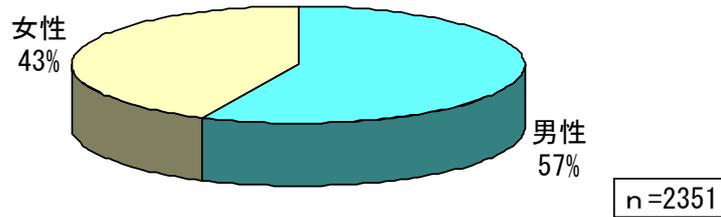
3-2 集計結果

3-2-1 回答者属性

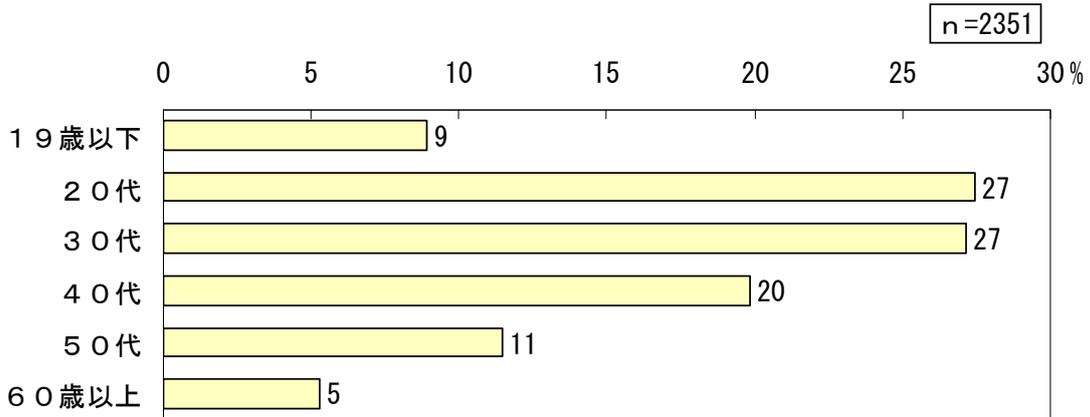
(1) 性別・年代

回答者の性別及び年代別の構成は、以下の通りであった。(図表 3-2-1、3-2-2)。

図表 3-2-1 性別



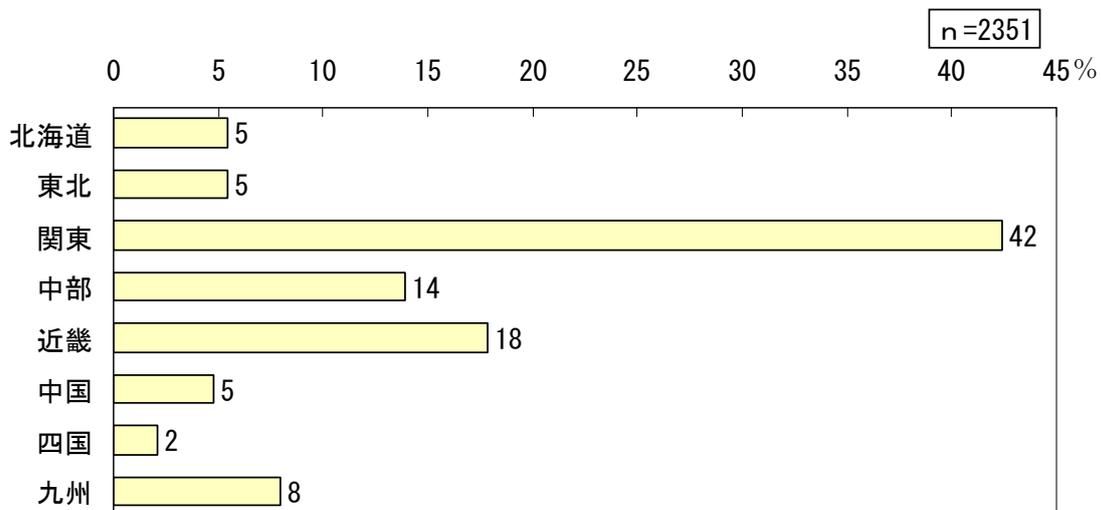
図表 3-2-2 年代別



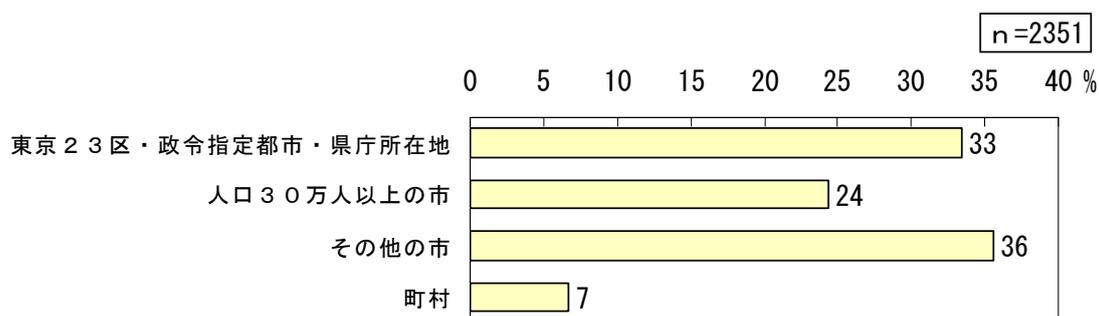
(2) 居住地

回答者の居住地は、関東・近畿・中部の3地域の合計で7割以上を占めている(図表 3-2-3)。都道府県内でも町村よりは、市レベル以上の回答者の割合が高く、東京23区、政令指定都市、県庁所在地及び市の合計で9割以上を占めている(図表 3-2-4)。

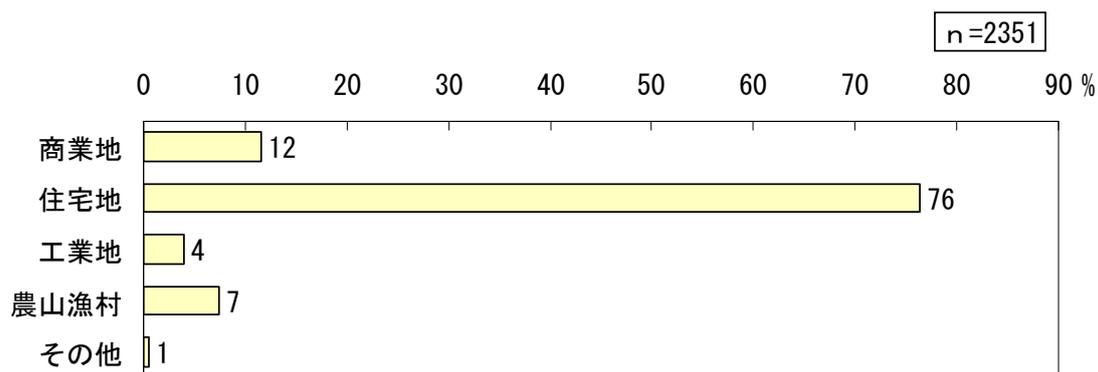
図表 3-2-3 居住する地域ブロック別



図表 3-2-4 居住地の市町村規模別



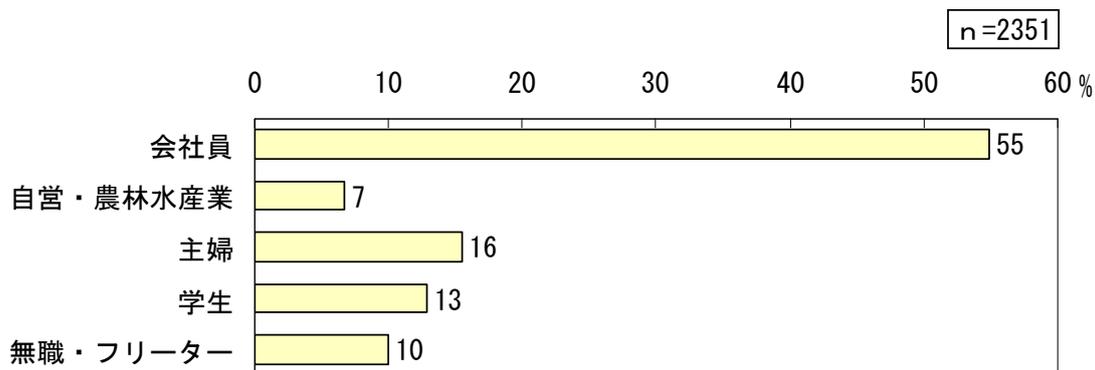
図表 3-2-5 居住地の性格別



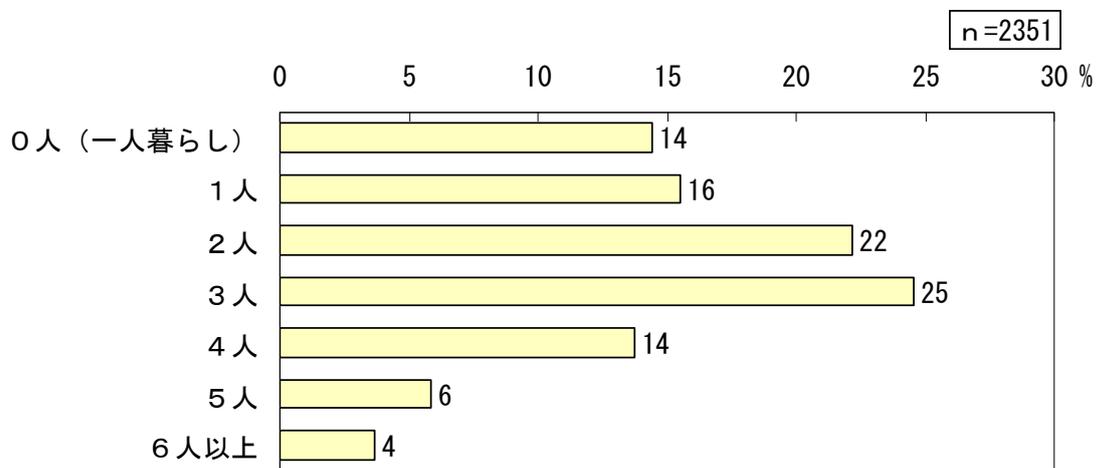
(3) 職業・同居者数

回答者の職業は会社員の割合が圧倒的に高く、次いで主婦、学生の割合が高い（図表 3-2-6）。同居者数は 2, 3 人の割合が高く、47%を占める（図表 3-2-7）。

図表 3-2-6 職業別



図表 3-2-7 同居者の数別



3-2-2 インターネットの利用状況

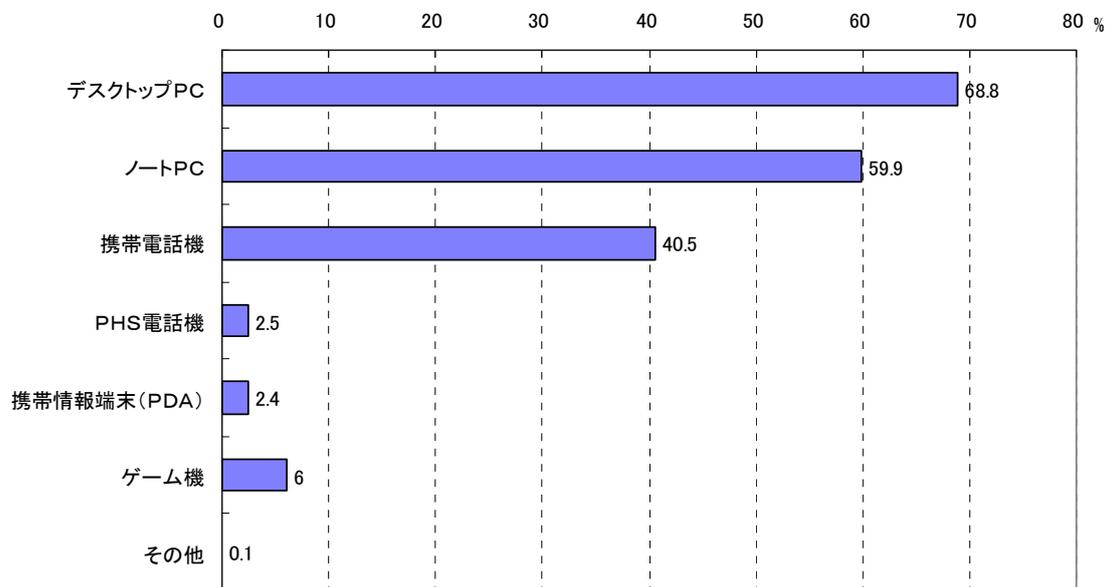
(1) インターネットの利用環境

次にブログ開設者のインターネット利用状況について見る。

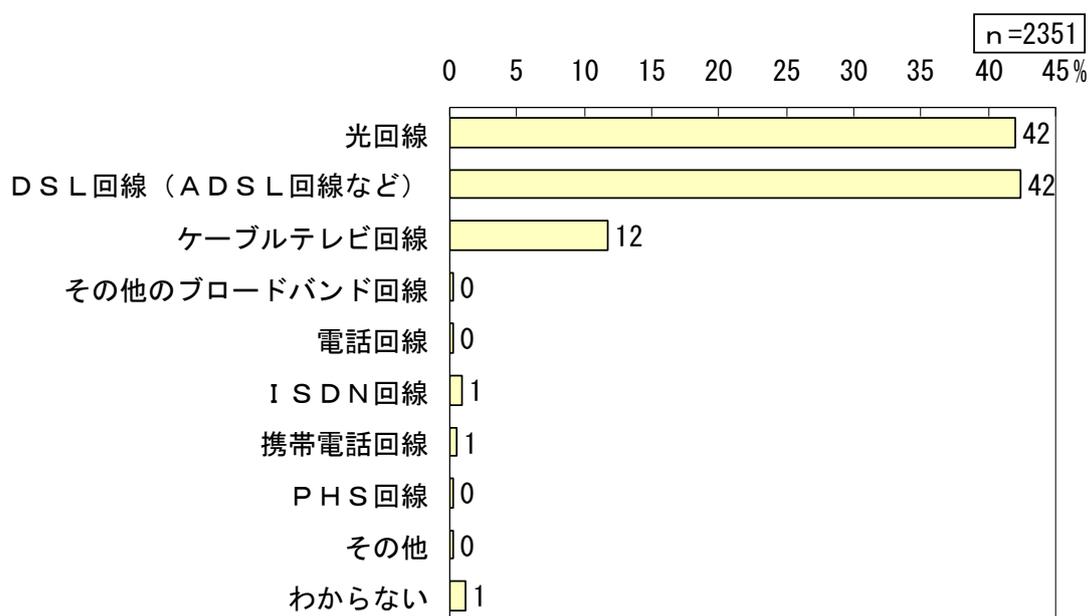
インターネット利用端末は、デスクトップPCとノートPCの割合が約6割ないし7割と高く、携帯電話は約40%であった(図表3-2-8)。

インターネット利用回線は、光回線とDSL回線が圧倒的であり、ほとんど全ての回答者がブロードバンド利用者である(図表3-2-9)。

図表 3-2-8 インターネット利用端末 (複数回答)



図表 3-2-9 主なインターネット利用回線

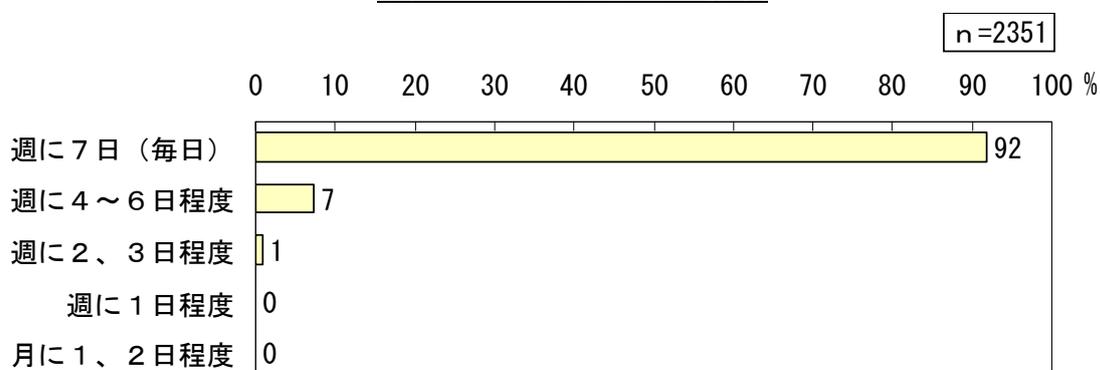


(2) インターネットの利用頻度・利用時間

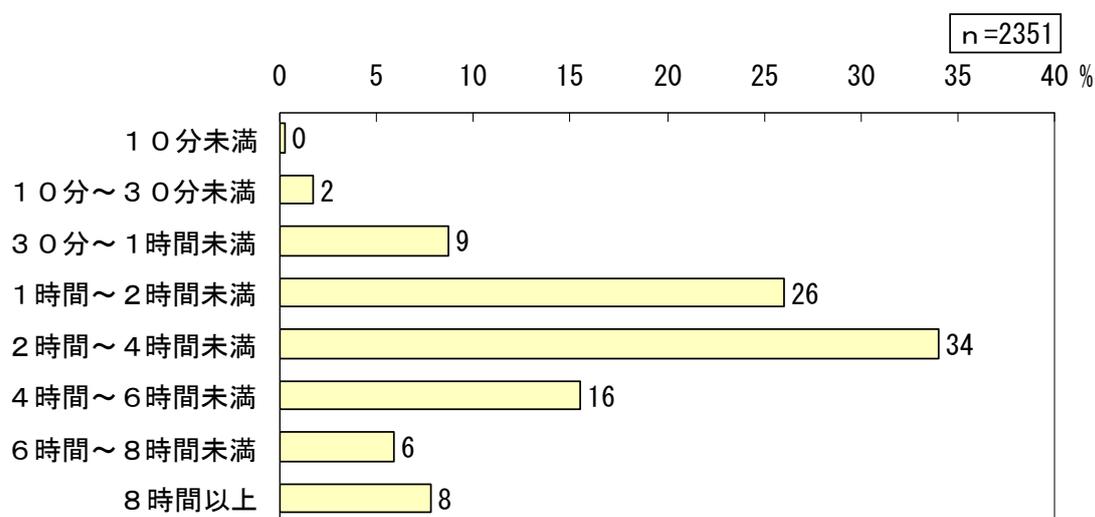
インターネット利用頻度では、9割以上が「毎日」と回答しており、ブログ開設者はネットを日常的に利用しているといえる (図表 3-2-10)。

一日の平均的なネット利用時間は、個人的な利用が 2 時間~4 時間未満の割合が最も高くなっている一方、仕事等の利用では 10 分未満の割合が約 4 割と最も高くなっている (図表 3-2-11、3-2-12)。仕事でのネット利用時間で 10 分未満の割合が高いのは、職業が主婦や学生、無職・フリーターの回答者の割合が 4 割と高いためと考えられる (図表 3-2-6 参照)。10 分未満以外では 1 時間~2 時間未満程度の割合が高い。

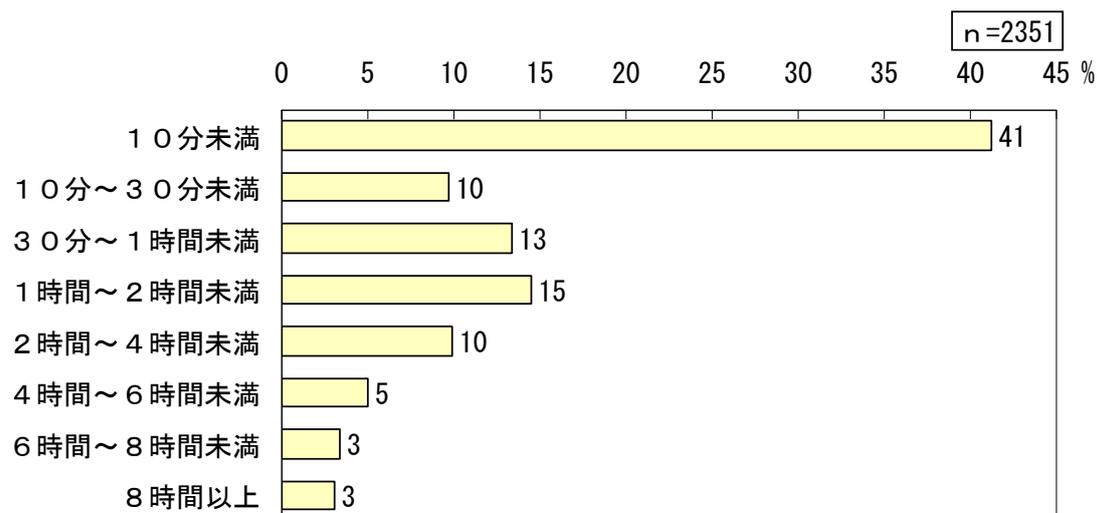
図表 3-2-10 ネット利用頻度



図表 3-2-11 個人での一日の平均的なネット利用時間



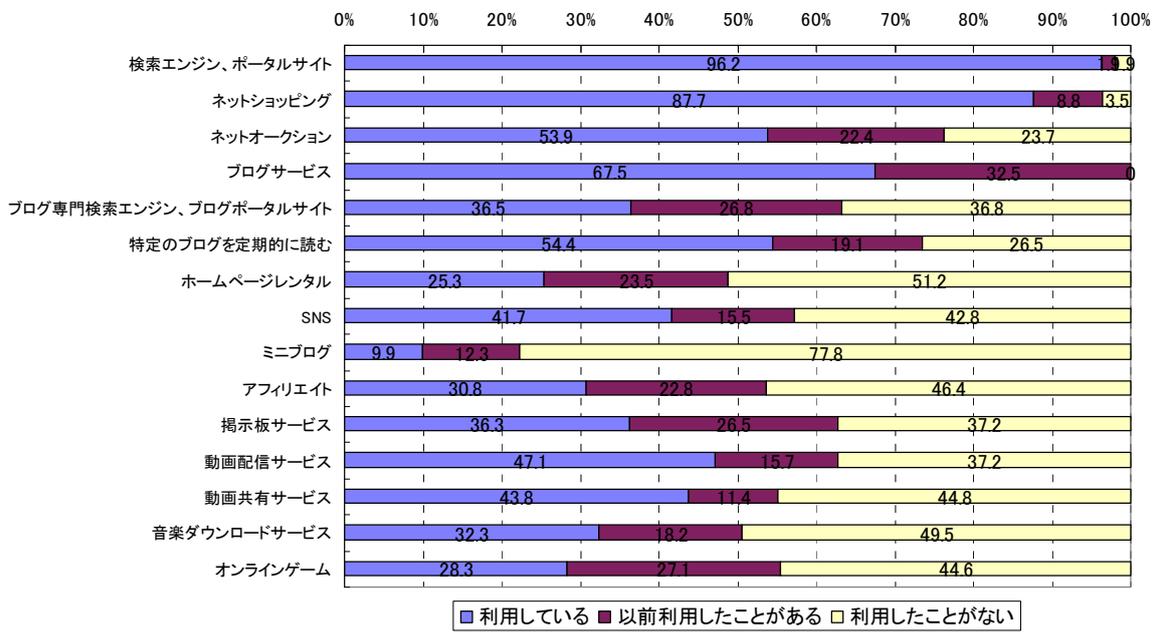
図表 3-2-12 仕事等での一日の平均的なネット利用時間



(3) 利用しているインターネットサービス

ブログ開設者の他のネットサービスの利用状況を見ると、検索エンジンやネットショッピングについては、9割以上の回答者に利用経験がある（図表 3-2-13）。ネットオークションも7割以上が利用しており、ソーシャルネットワーキングサイト（SNS）、掲示板サービス、動画共有サービス、オンラインゲーム、アフィリエイトについても、半数以上は利用経験がある。

図表 3-2-13 ネット上の各種サービスの利用状況



3-2-3 ブログの利用状況

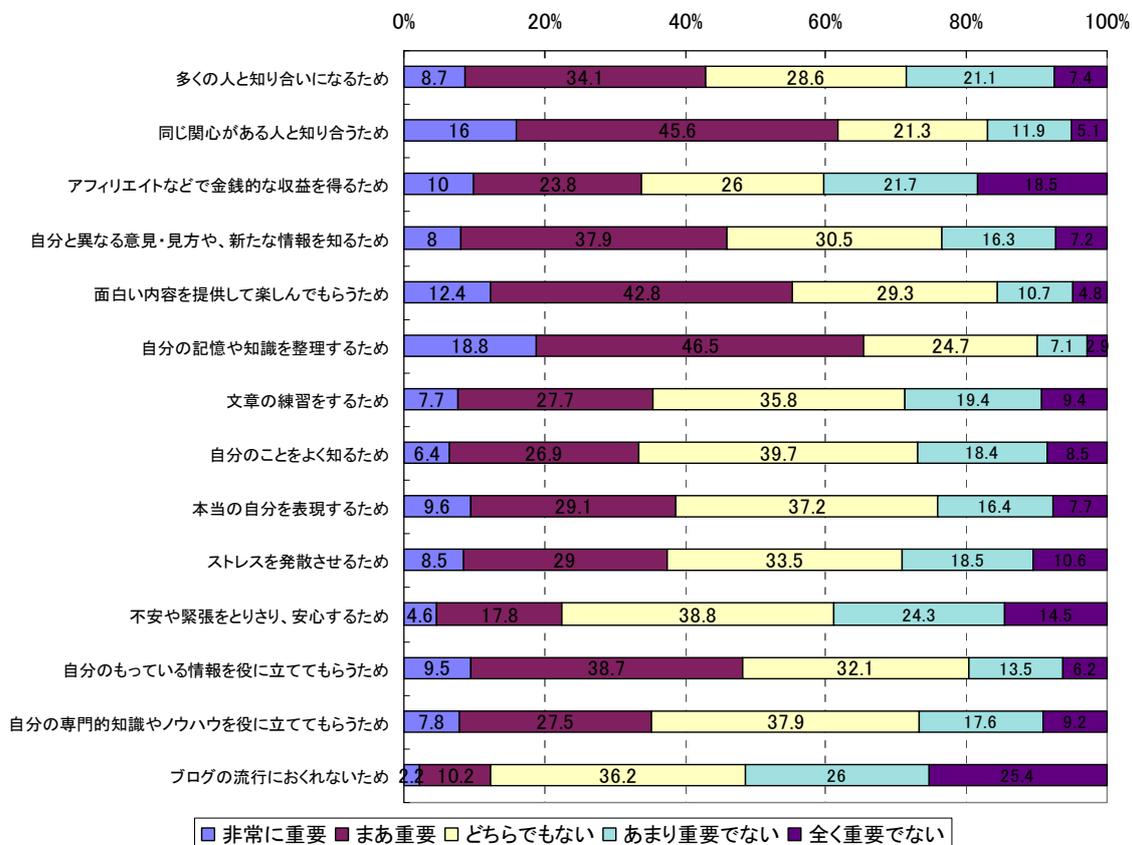
(1) ブログ更新の動機・ブログの利用スタイル

ブログ開設者のブログ更新動機として考えられる理由を挙げ、動機としてどの程度重要かたずねたところ、「非常に需要」、「まあ重要」を合わせた比率で見ると、「自分の記憶や知識を整理するため」、「同じ関心がある人と知り合うため」の2つがそれぞれ6割を超えて第1位、第2位となり、重視される傾向が見られた。ブログは主に、コミュニケーションツールであると同時に、個人の知識や情報の整理ツールとして利用されているといえる（図表 3-2-14）。

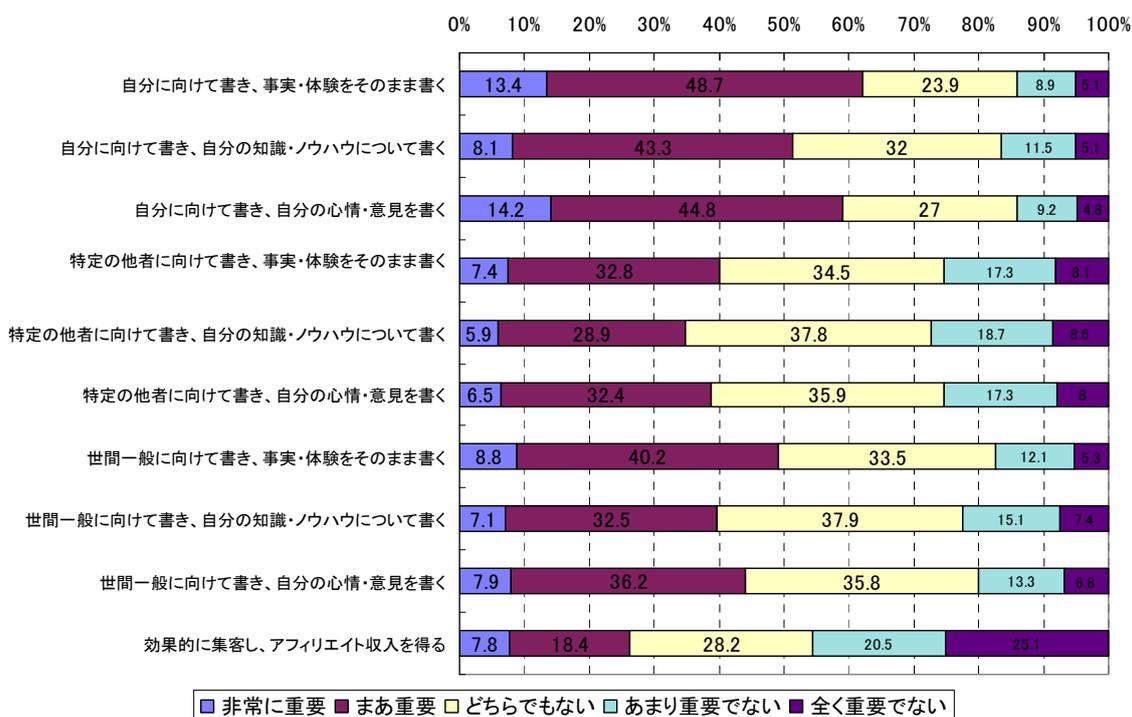
また、3-1-1 で述べたブログ利用スタイルの仮説を検証するための直接的な質問項目として、ブログを自分自身・特定の他者・世間一般のいずれに向けて書いているか（想定読者）、自分の体験・知識・心情（ブログの内容）のいずれを重視して書いているかを、9通りの組み合わせとして質問した（図表 3-2-15）。

これらの回答は、多変量解析を使って、回答者がブログをなぜ・どのように利用しているかを分析し、ブログ開設者像を分類するために、3-4 で詳しく検討する。

図表 3-2-14 ブログを更新する動機（各項目について5件法）



図表 3-2-15 ブログの利用スタイル（各項目について5件法）



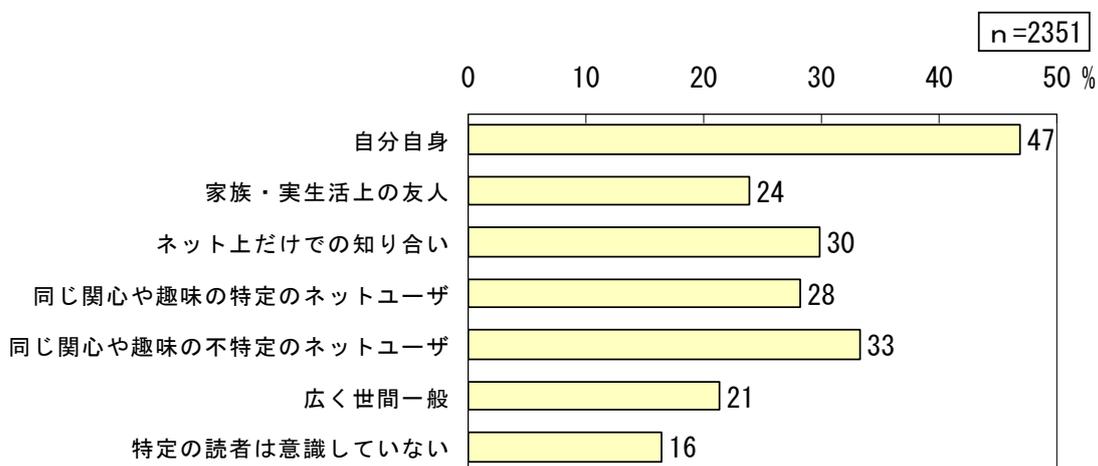
(2) 想定読者、ブログの内容、匿名／実名でのブログ更新

ブログの想定読者は、複数回答が多く見られる。「自分自身」を選択する回答の割合が高く（47%）、「特定の読者は意識していない」という回答の割合が最も低かった（図表 3-2-16）。

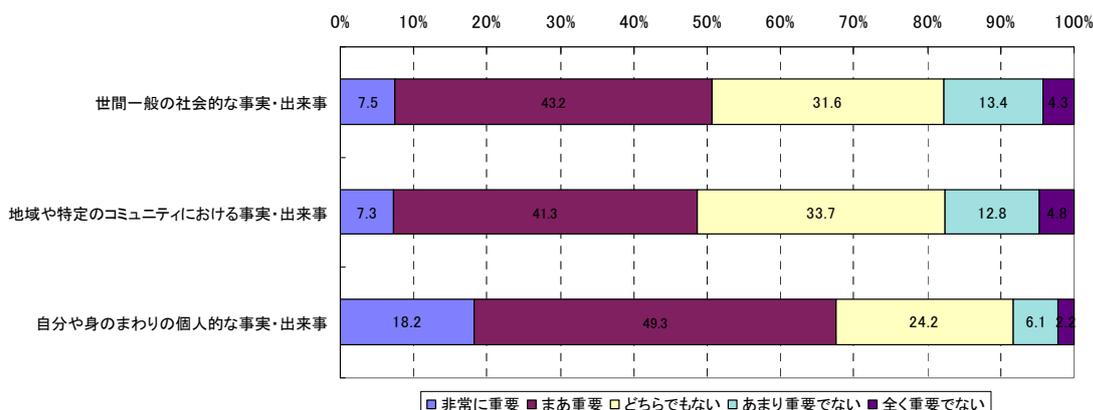
ブログの内容では、「自分の身のまわりの個人的な事実・出来事」が最も重視されており、身近な体験がブログの題材として好まれていると言える（図表 3-2-17）。

ブログを匿名で書いているか実名で書いているかについては、匿名、あるいは実名のわからないように配慮したハンドルネームで書いているという回答の割合が高い（図表 3-2-18）。

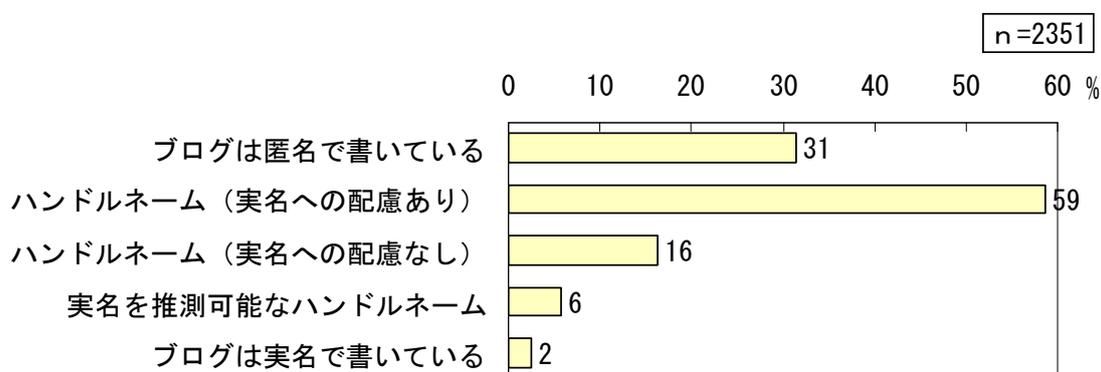
図表 3-2-16 ブログの想定読者（複数回答）



図表 3-2-17 ブログの内容（5件法）



図表 3-2-18 ブログ運営における匿名/実名（複数回答）



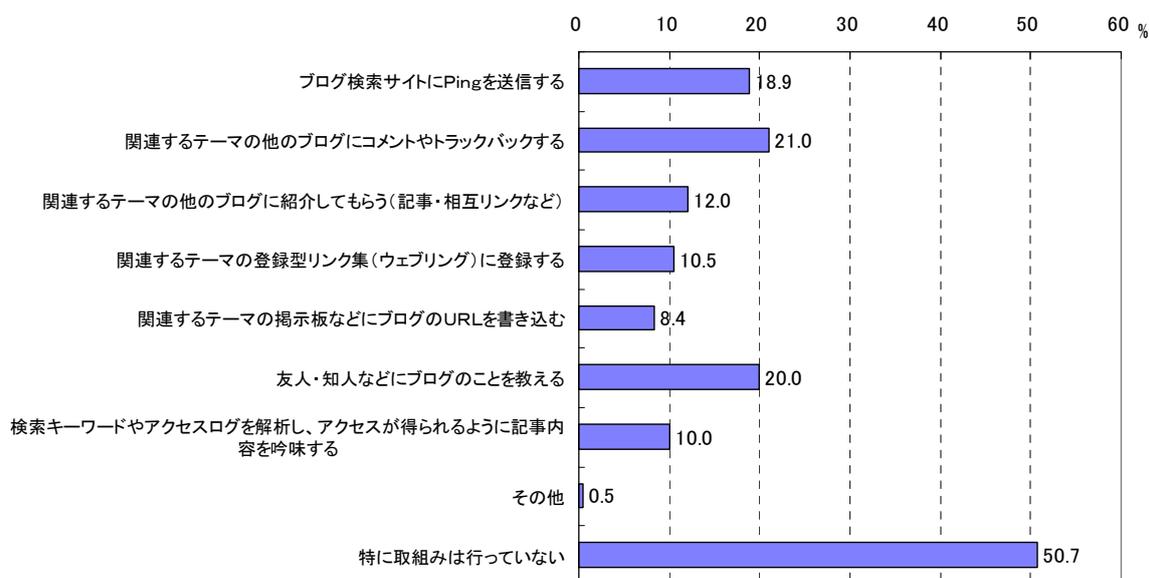
（3）ブログの認知向上のための取組み、ブログへの反応

ブログの認知向上やアクセス向上のための取組みは、半数程度は特に実施していない。残り半数は ping 送信や同種のブログへのコメント・トラックバック、知人・友人に教えるといった工夫をする割合が比較的高い（図表 3-2-19）。

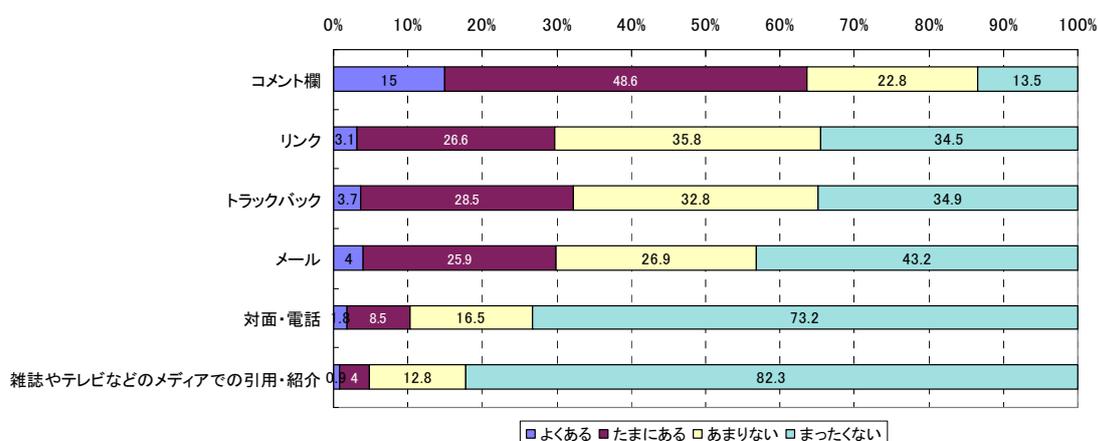
ブログに寄せられる反応は、形態としては「コメント欄」の割合が最も高く、6割程度の回答者が、反応が「よくある／たまにある」と回答している。一方、トラックバックによる反応はリンクやメールと同程度であった（図表 3-2-20）。

ブログに寄せられる反応の内容としては、共感、励まし、感想・意見の割合が高く、概ね肯定的な反応を経験している者の割合が高い（図表 3-2-21）。

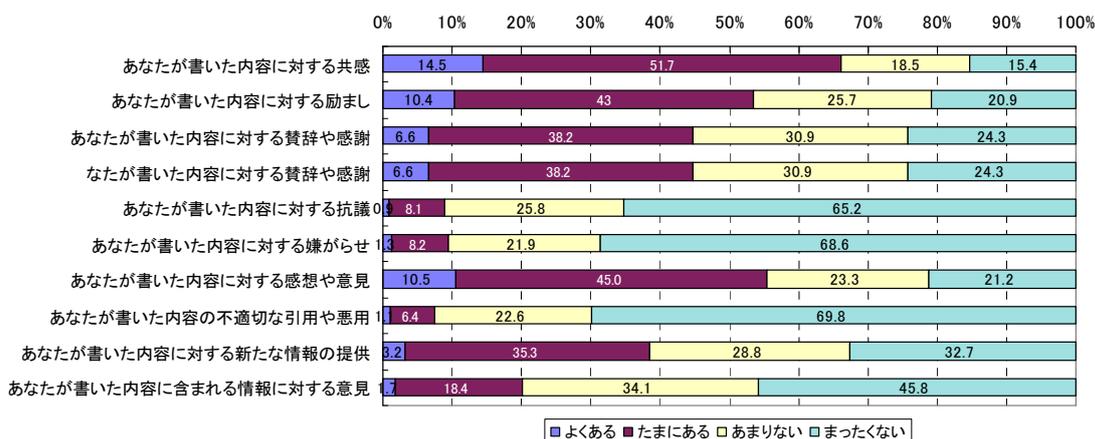
図表 3-2-19 ブログの認知向上・アクセス向上のための取組み（複数回答）



図表 3-2-20 ブログに寄せられる反応の形態（各項目につき 4 件法）



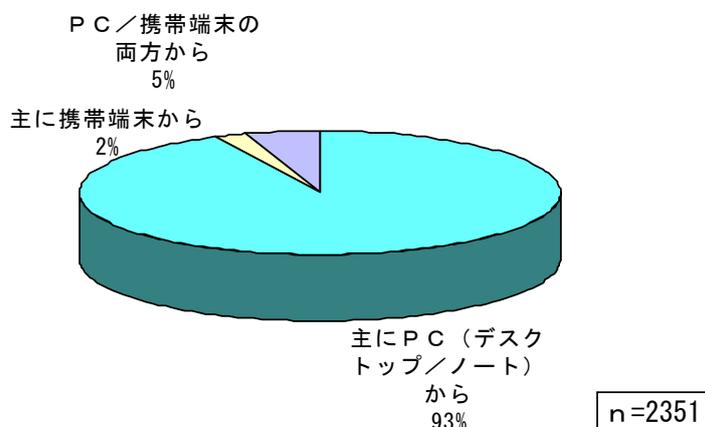
図表 3-2-21 ブログに寄せられる反応の内容（各項目につき 4 件法）



(4) ブログの更新手段

ブログの更新手段では、「主に PC から」行う割合が 9 割以上であった(図表 3-2-22)。

図表 3-2-22 ブログの更新手段

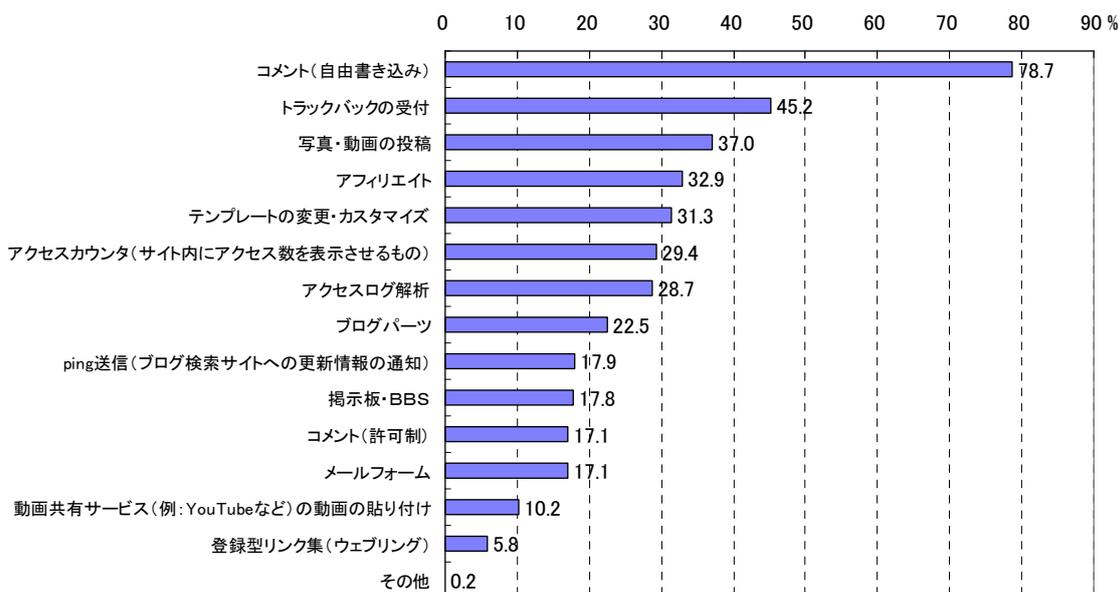


(5) ブログの付加機能の利用状況

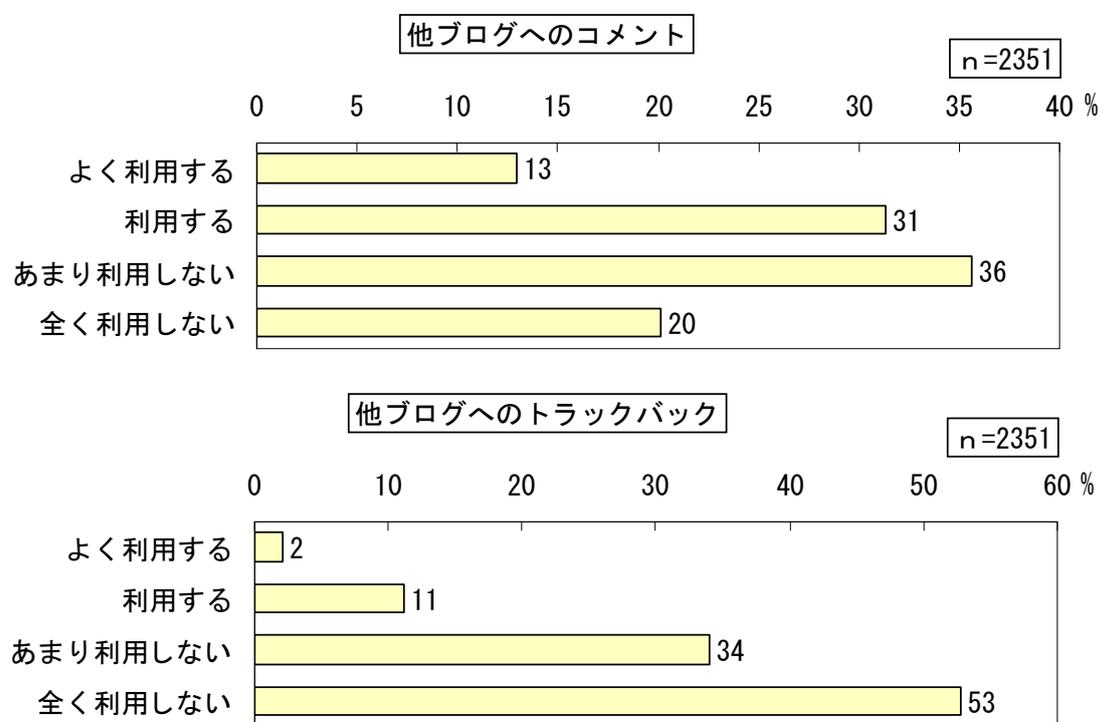
ブログで利用している付加機能としては、コメントの割合が最も高く、約8割が開設者の承認が不要で自由に書き込みができるコメント機能を利用している。トラックバックの利用率は約半分である。その他の機能では、写真・動画の投稿が4割弱、アフィリエイト、テンプレートのカスタマイズが3割程度と高くなっている（図表 2-2-23）。

他のブログを閲覧する利用者としての質問では、コメントを「よく利用する/利用する」との回答が44%であるが、トラックバックを「よく利用する/利用する」との回答は13%であった（図表 3-2-24）。

図表 3-2-23 利用しているブログの付加機能（複数回答）



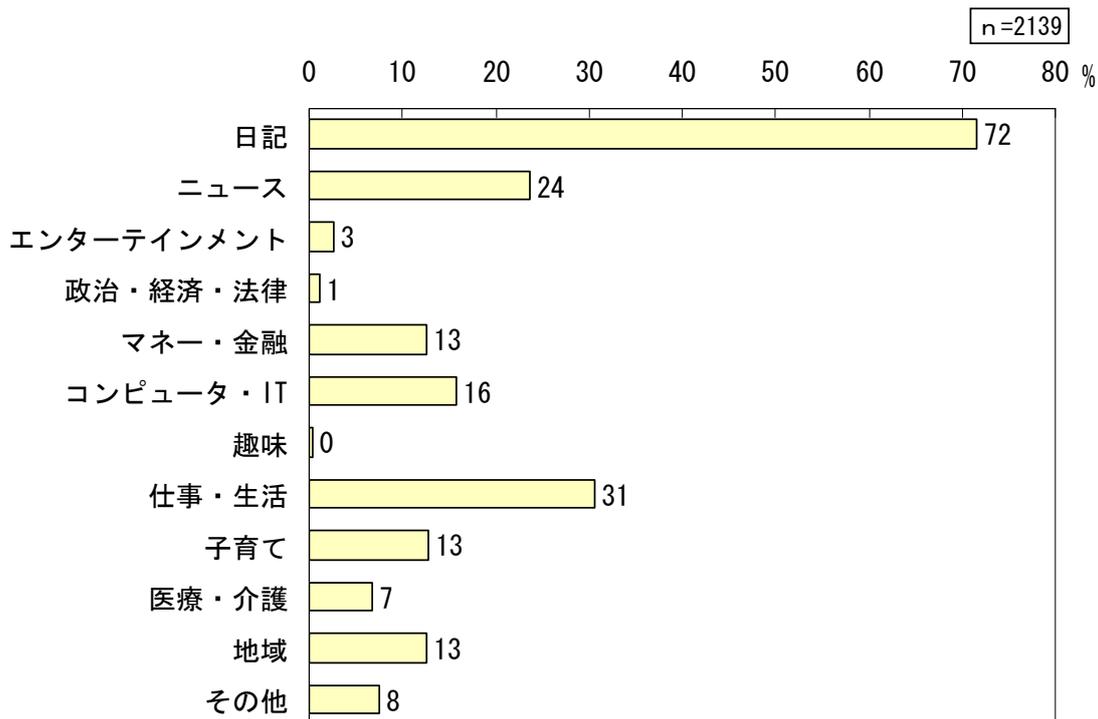
図表 3-2-24 他のブログに対するコメントやトラックバックの利用



(6) よく閲覧するブログのテーマ

ブログ閲覧者としてよく見るブログのテーマとしては、日記の割合が最も高く、より内容が特定のテーマでは、仕事・生活、ニュース、コンピュータ・IT、マネー・金融、子育て、地域の割合が高い (図表 3-2-25)。

図表 3-2-25 よく閲覧するブログのテーマ (複数回答)

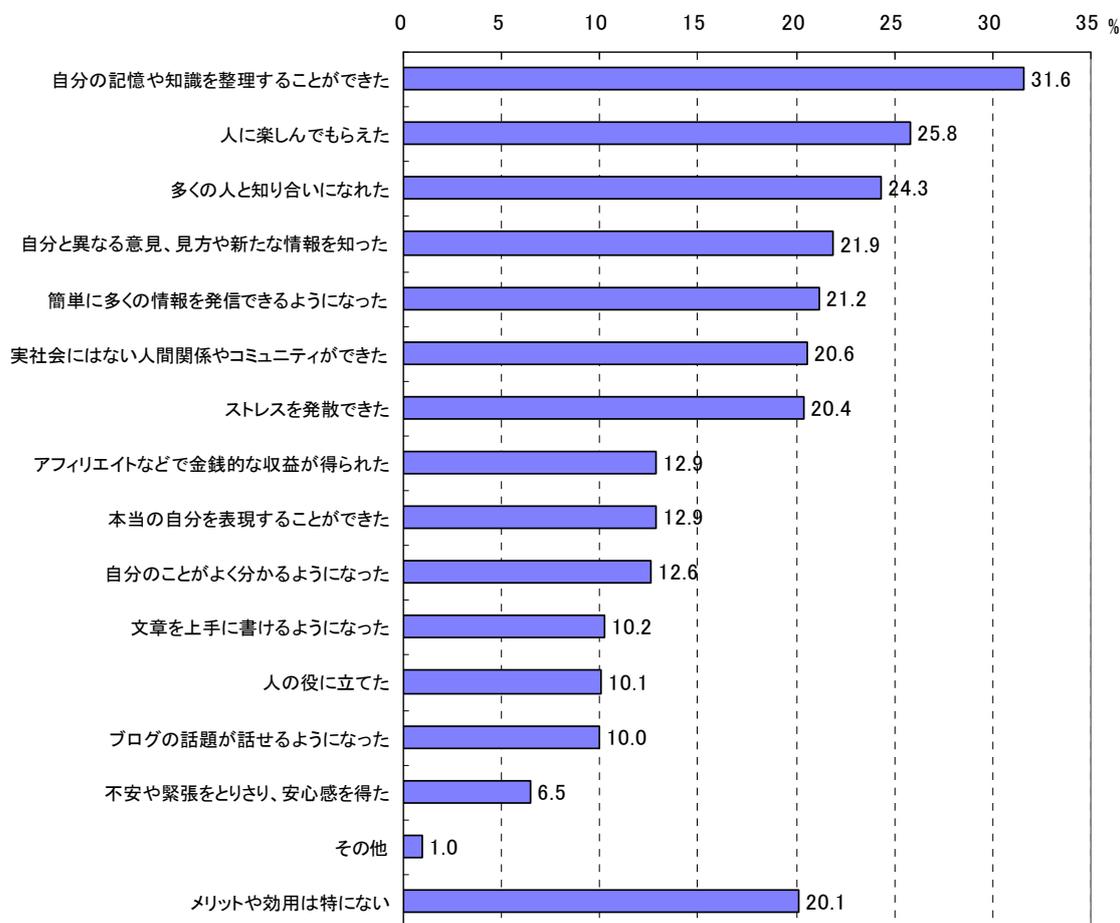


(7) ブログにより得られたメリット、コストの評価

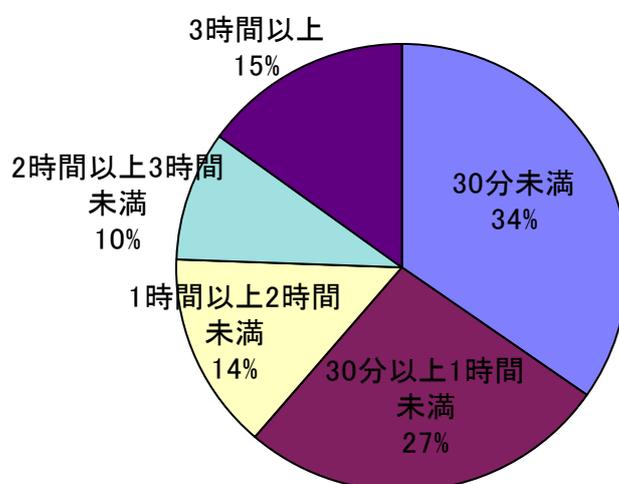
ブログを作成することで得られたメリット・効用は、ブログ更新の動機 (3-2-3 (1)、図表 3-2-14 参照) と概ね類似した傾向がある。メリット・効用が最も割合の高かった「自分の記憶や知識を整理することができた」は、更新動機で最も高かった「自分の記憶や知識を整理するため」と対応している。これに次いで高かった「人に楽しんでもらえた」、「多くの人と知り合いになれた」、「自分と異なる意見、見方や新たな情報を知った」も更新の動機で類似した項目が高くなっている。また、「簡単に多くの情報を発信できるようになった」、「実社会にない人間関係やコミュニティができた」は、ブログの機能を活用したことによりメリットが実感されたことを示している (図表 3-2-26)。

多くのブログサービスは基本的なブログ機能が無償で利用できるため、ブログ開設者が負担するコストは実質的にブログ更新の時間的コストに等しいと考えられる。ブログ更新にかかる週当たりの時間は、約 6 割が週 1 時間未満であった (図表 3-2-27)。

図表 3-2-26 ブログ作成により得られたメリット・効用（複数回答）



図表 3-2-27 週当たりのブログ更新時間

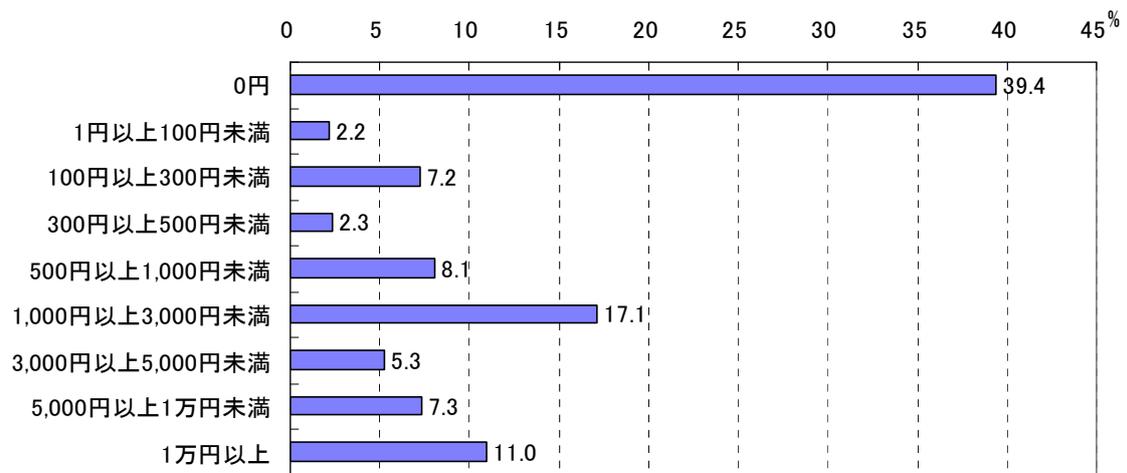


ブログ利用のメリットと、ブログ更新にかかる時間を踏まえ、ブログの利用により得られるメリットと、かかるコストを仮想的に金額で評価してもらった。その結果、メリット・コストいずれについても「0円」と答える割合が約4割と最も高いものの、0円以外の回答では「1000円以上3000円未満」の割合が最も高かった(図表 3-2-28、3-2-29)。

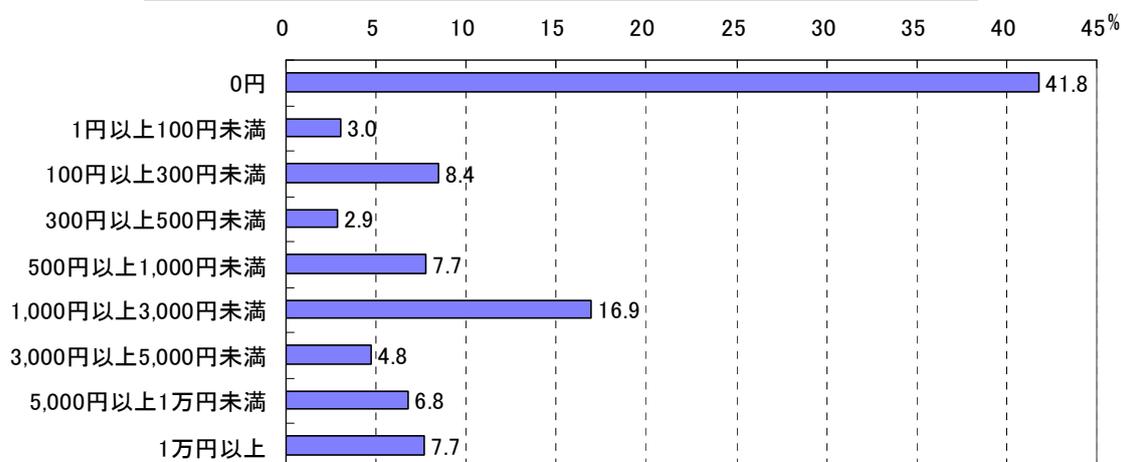
ブログ利用のメリットの平均額は 25,592 円であり、コストの平均額は 3,281 円であった。いずれも非常に高額の評価を行った少数の回答者により、金額が高く振れており、中央値で見ると、メリットが 300 円、コストが 100 円であった。平均的にはメリットがコストを上回っており、ブログの利用継続に繋がっていると言える。

メリットとコストの評価を元に、月当たりの支払意向額を尋ねたところ、基本的にはメリットとコストの評価と同様の傾向であり、「0円」が 35%と最も高く、それ以外では、「100円以上300円未満」の割合が最も高かった。100円以上300円未満は有料サービスのブログの平均的な月額利用料と概ね等しい金額となっている(図表 3-2-30)。

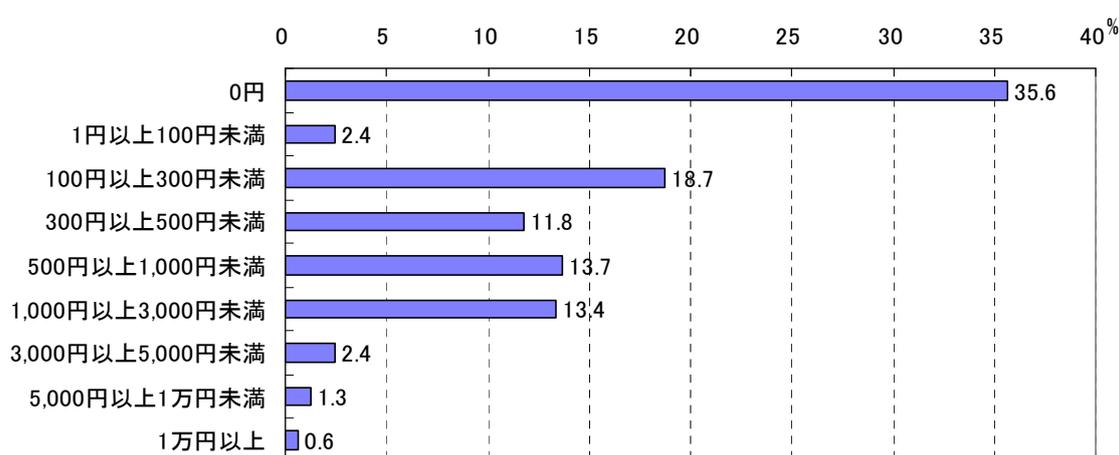
図表 3-2-28 ブログ利用で得られるメリットの月当たり経済的評価



図表 3-2-29 ブログ利用にかかるコストの月当たり経済的評価



図表 3-2-30 ブログサービスが有料だった場合の月当たり支払意向額

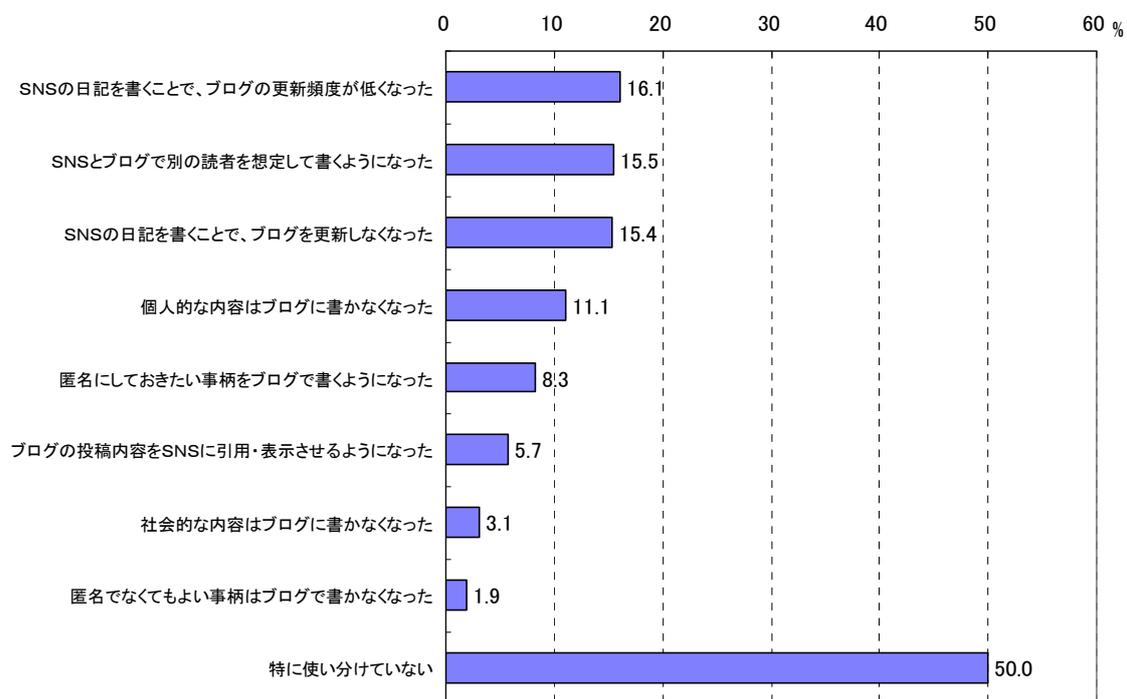


(8) ブログと SNS の関係・ブログを止めた理由

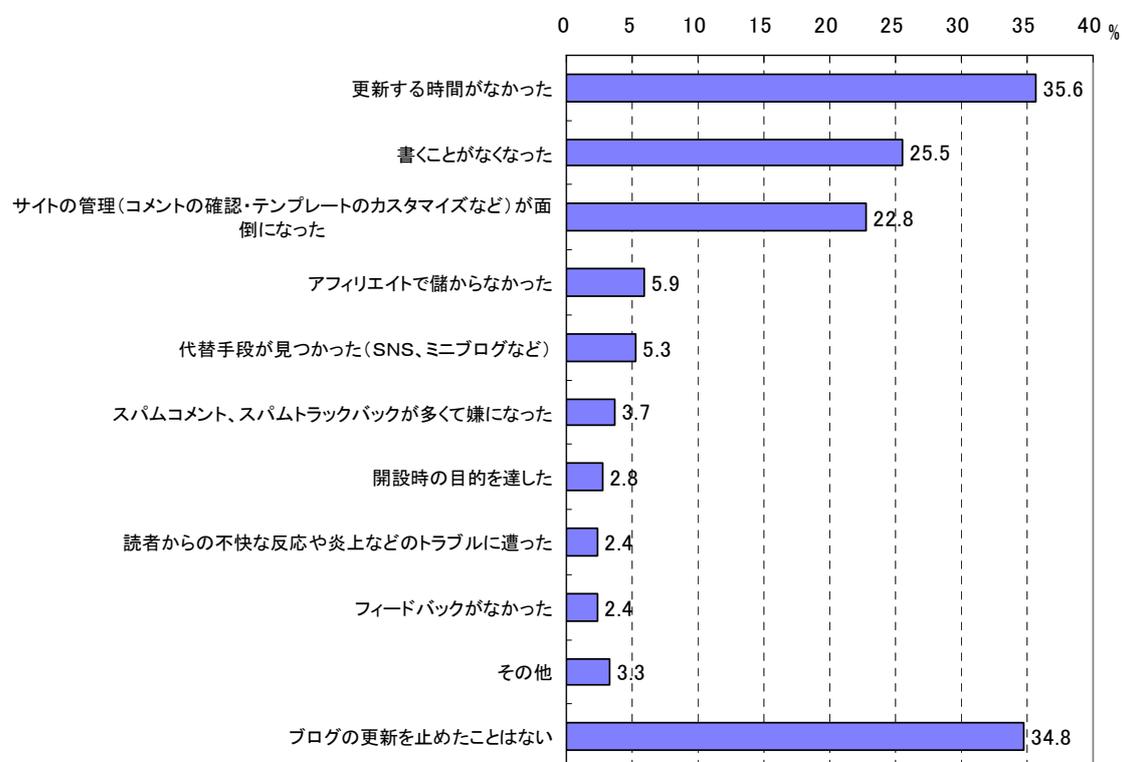
ブログとほぼ同時期に普及し、同様に日記的な機能をもつ SNS との関係について、SNS とブログの両方の利用経験がある者（ブログ開設者全体の約 6 割、3-2-2(3)、図表 3-2-13 参照）の回答を見ると、ほぼ半数は特に使い分けを意識していない。SNS により「ブログの更新頻度が下がった」、または「更新を止めた」と答えている者が併せて 3 割程度いる一方、「別の読者を想定して書くようになった」など、使い分けている者も、約 15%いる（図表 3-2-31）。

ブログの更新を止めたことがあると回答した者は約 65%で、ブログを止めた理由は「更新する時間がなくなった」、「書くことがなくなった」、「サイトの管理が面倒になった」という個人的な理由の割合が高い（図表 3-2-32）。

図表 3-2-31 ブログと SNS の使い分け（複数回答）



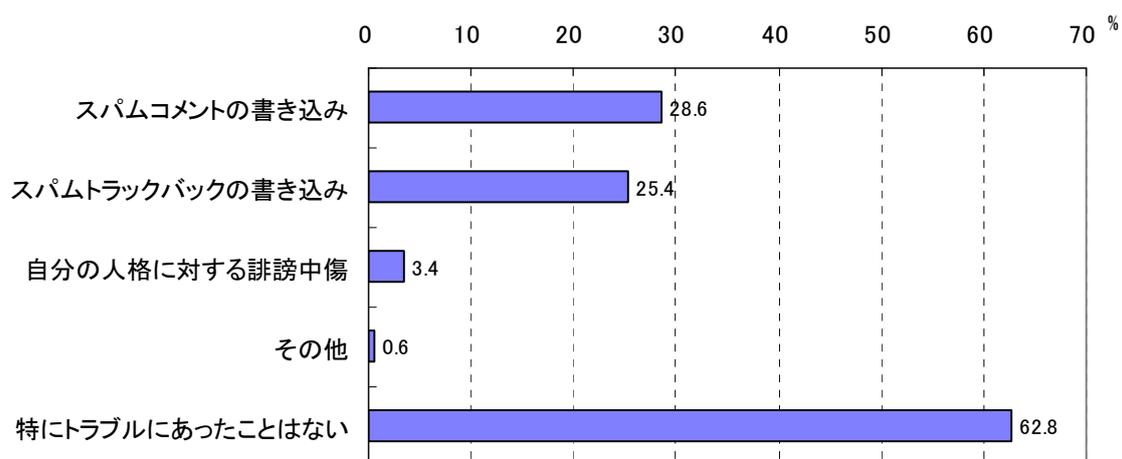
図表 3-2-32 ブログ更新を止めた理由（複数回答）



(9) ブログ上のトラブル

ブログを開設して遭遇したトラブルは、「特にない」という回答の割合が最も高いものの、スパムコメントにあった経験がある、スパムトラックバックにあった経験がある、とそれぞれ3割弱の開設者は回答している。自分の人格に対する誹謗中傷（いわゆる「炎上」を含む）の経験は3.4%であった（図表 3-2-33）。

図表 3-2-33 ブログで遭ったトラブル（複数回答）



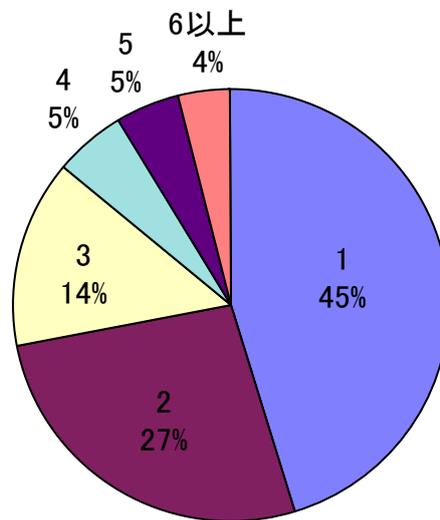
3-2-4 各ブログの運営状況

(1) 立ち上げブログ数・複数ブログ間の関係

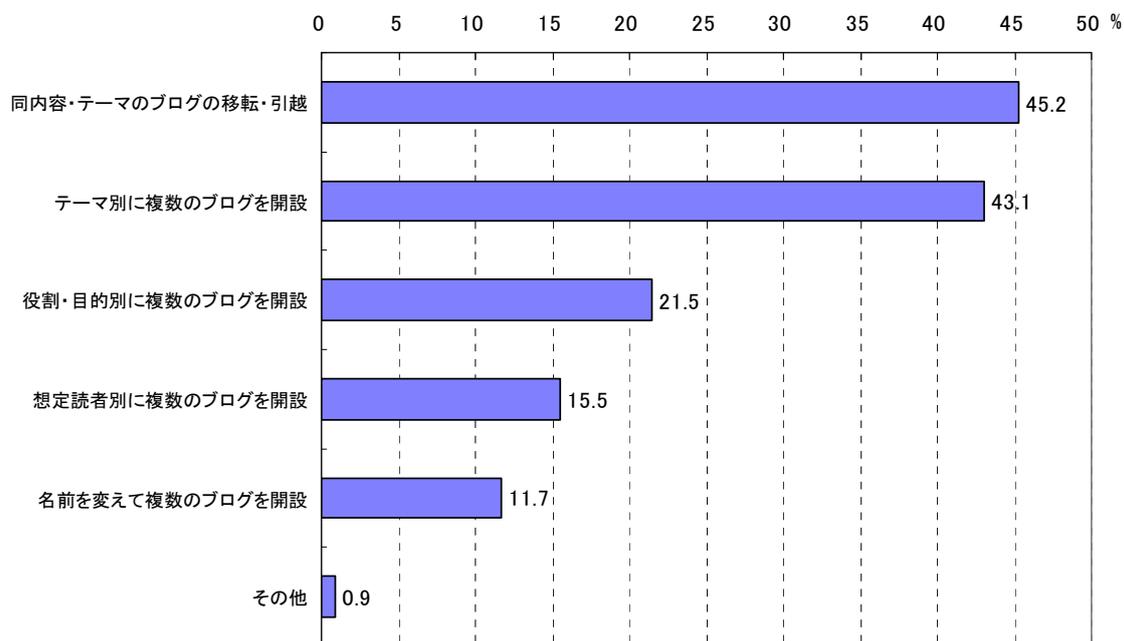
各ブログの運営の状況では、基本的に立ち上げたブログ毎に立ち上げ時期や更新状況などについて質問しているが、最初に共通の設問として、これまで立ち上げたブログ数を尋ねた。1つが約45%と最も割合が高いが、半数以上は複数のブログを立ち上げている(図表3-2-34)。

立ち上げたブログ間の関係は、移転・引越という回答と、テーマ別に分けるという回答の割合が高い(図表3-2-35)。

図表 3-2-34 これまで立ち上げたブログ数



図表 3-2-35 立ち上げた複数のブログ間の関係（複数回答）

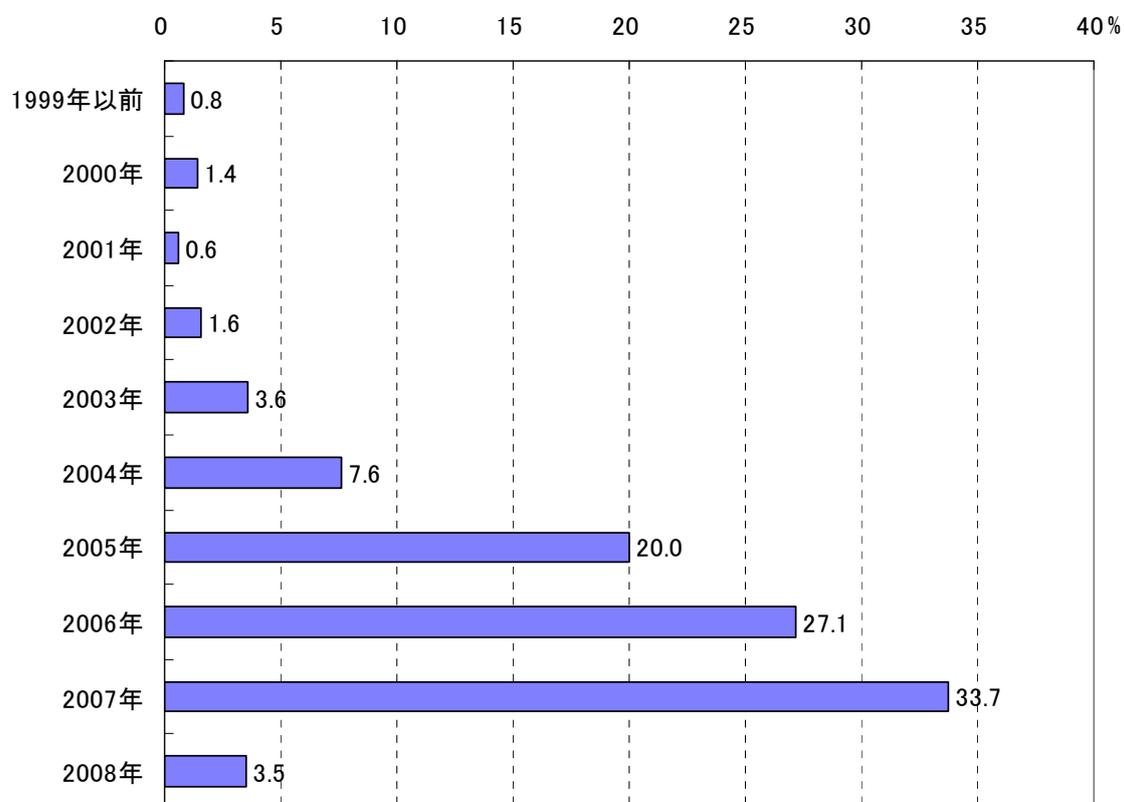


(2) ブログ開設時期・利用ブログサイト

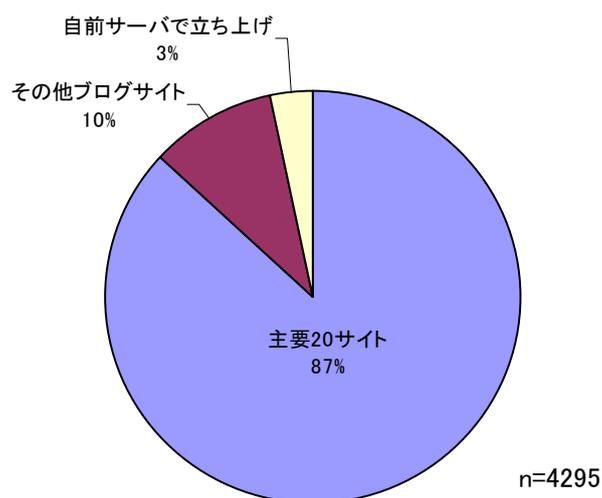
ブログ開設時期は、2005年以降の割合が圧倒的に高い。なお2008年が少ないのは調査が2008年2月実施のためである（図表 3-2-36）。

アンケートで回答されたブログの約87%が、主要20ブログサイトを利用して開設されたものであり、それ以外のブログサイトや自前でサーバを立ち上げてブログを利用する割合は13%であった（図表 3-2-37）。

図表 3-2-36 ブログ開設時期



図表 3-2-37 利用しているブログサイト等



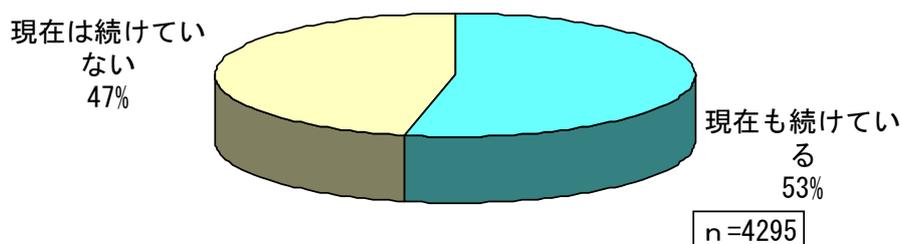
(3) ブログの更新状況・現在の状態

ブログの更新状況では、約半数のブログが現在も更新中であると回答している（図表

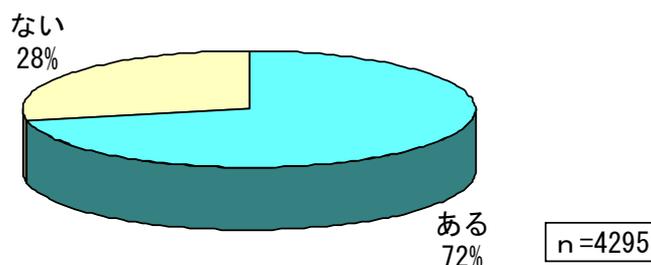
3-2-38)。

またブログがネット上にあるかという質問には、7割以上があると回答している（図表 3-2-39）。

図表 3-2-38 ブログの更新状況



図表 3-2-39 現在もネット上にあるか

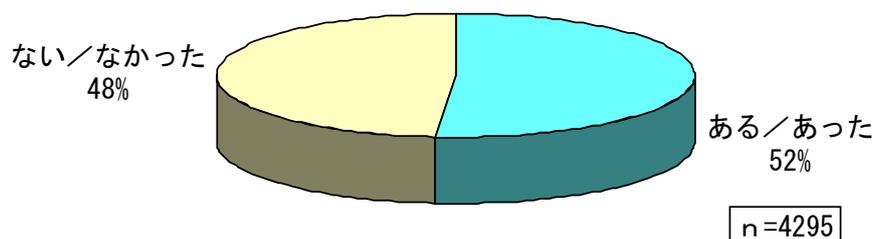


(4) ブログのテーマ

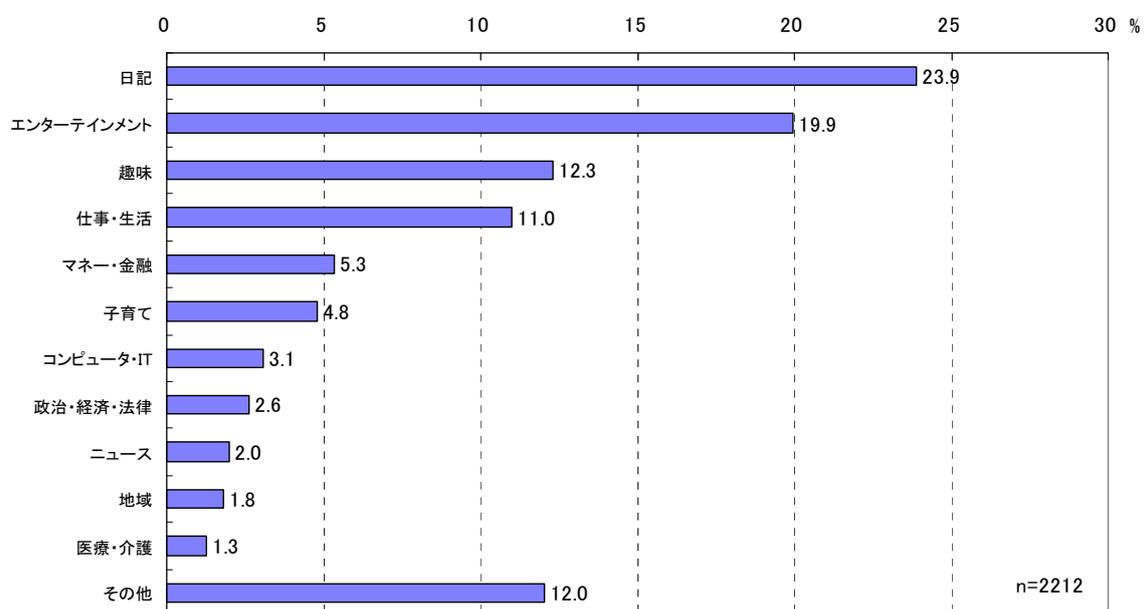
開設したブログについて、特定のテーマがあるかどうか、又はあったかを尋ねたところ、約半数がある、又はあったと回答している（図表 3-2-40）。

ある、又はあったと回答があったブログについて、最も近いテーマを選択してもらった結果、日記の割合が最も高く、より内容が特定のテーマでは、エンターテインメント（芸能やスポーツ、音楽、映画など）や趣味（車、バイク、旅行など）、仕事・生活の割合が高い（図表 3-2-41）。

図表 3-2-40 ブログのテーマの有無



図表 3-2-41 ブログのテーマ



3-3 開設者の性別・年代との関連

本節では、性別・年代によるブログ利用状況の違いを分析するため、クロス集計結果の検討を行う。性別・年代以外の属性によるクロス集計も行ったが、ブログの利用は主に性別・年代により大きな違いが見られるため、本節の分析では性別と年代を基本として分析を行う。

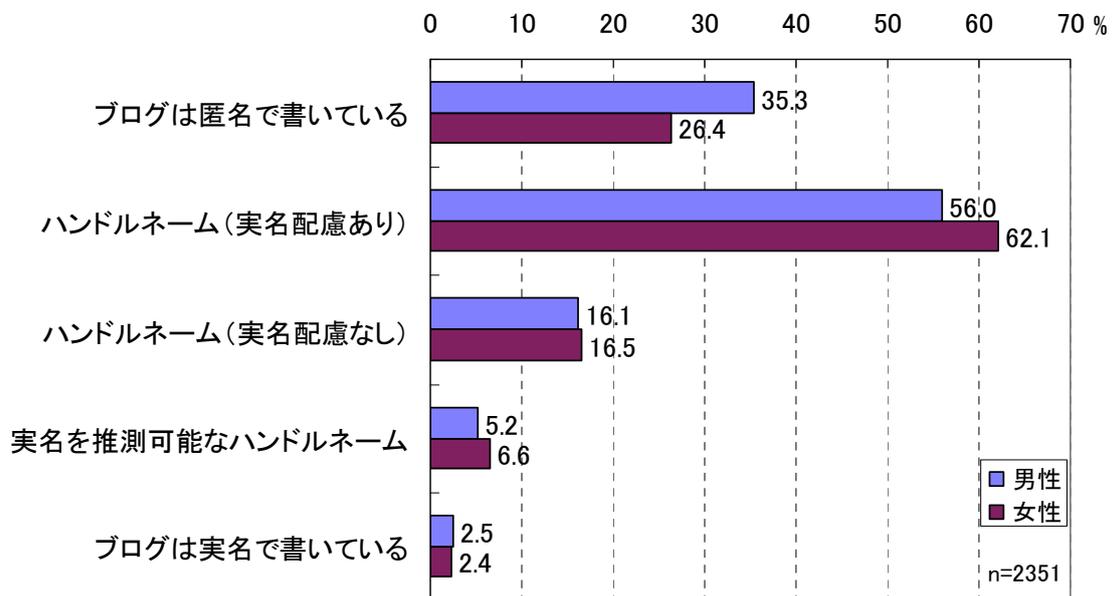
3-3-1 属性別のブログ利用状況

(1) 匿名／実名でのブログ運営

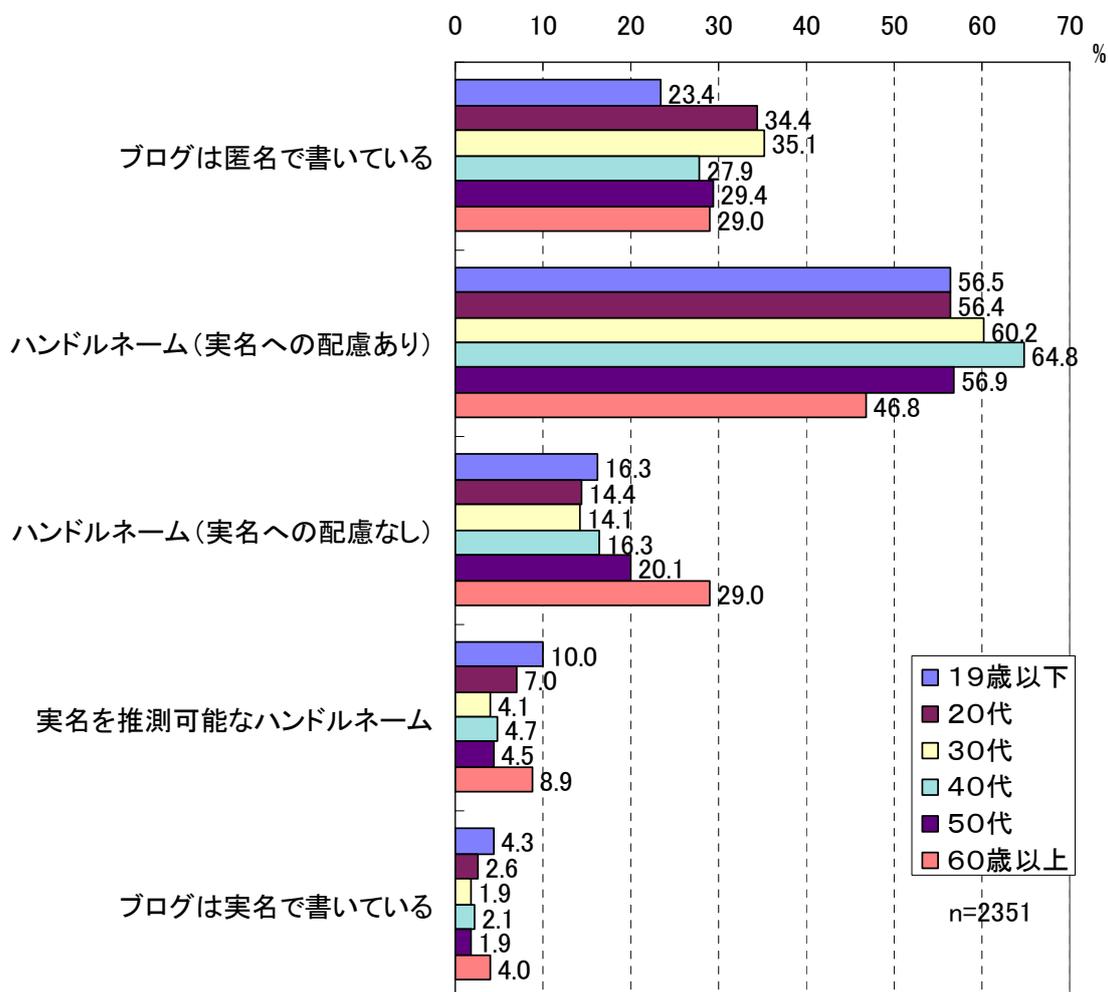
匿名／実名のどちらでブログを運営しているかについて、性別・年代別に見ると、いずれも匿名の割合が非常に高く、実名を公開する割合は性別・年代に関係なく低い。実名に配慮したハンドルネームと答えた割合は女性の方が高く、匿名で書いていると答えた割合は男性の方が高い（図表 3-3-1、3-3-2）。

年代別では、匿名の割合が 20 代～50 代の世代が高く、10 代・60 歳以上は低く、実名の割合は逆に 20 代～50 代が低く、10 代・60 歳以上代の割合が高い傾向がみられる。

図表 3-3-1 ブログ運営での匿名／実名（性別、複数回答）



図表 3-3-2 ブログ運営での匿名／実名 (年代別クロス 複数回答)

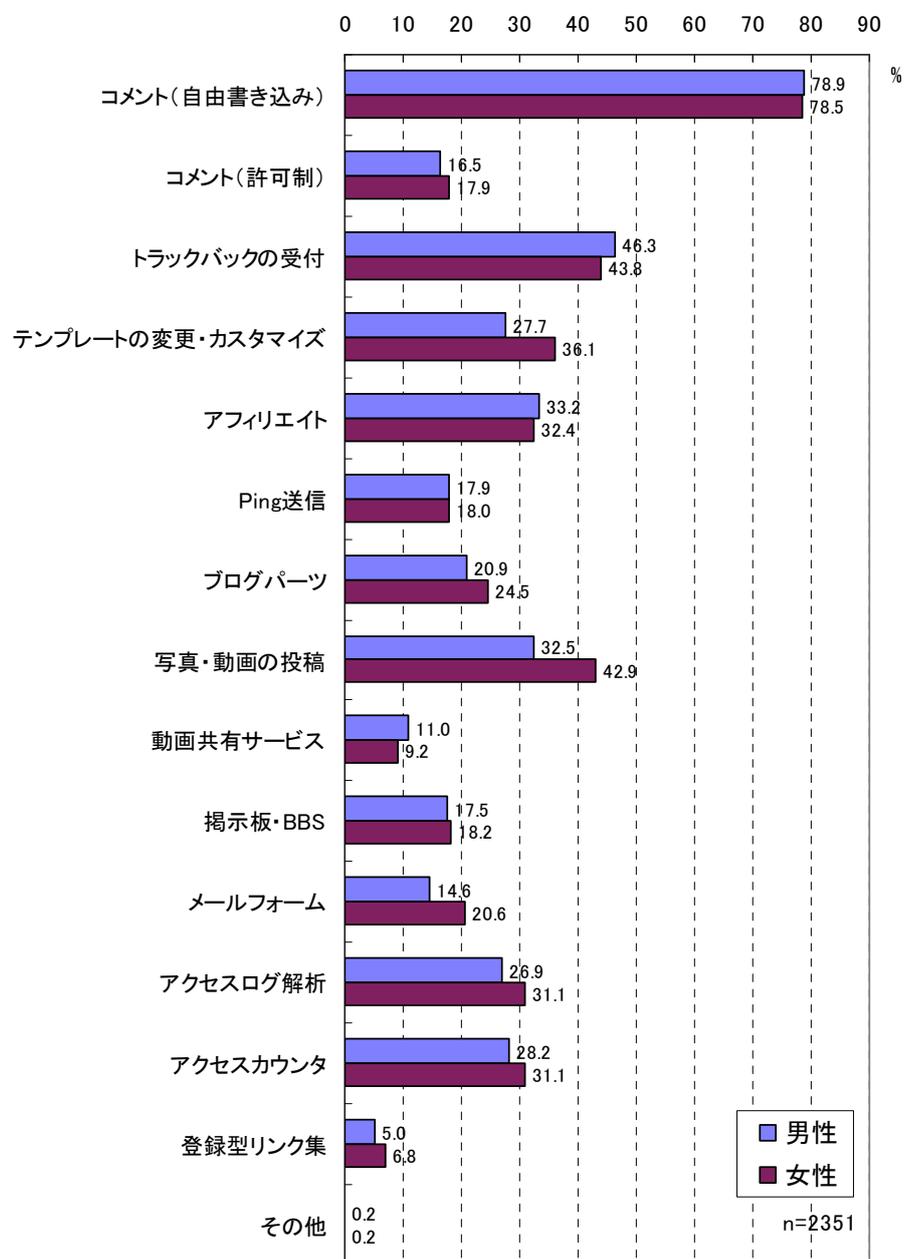


(2) 利用している付加機能

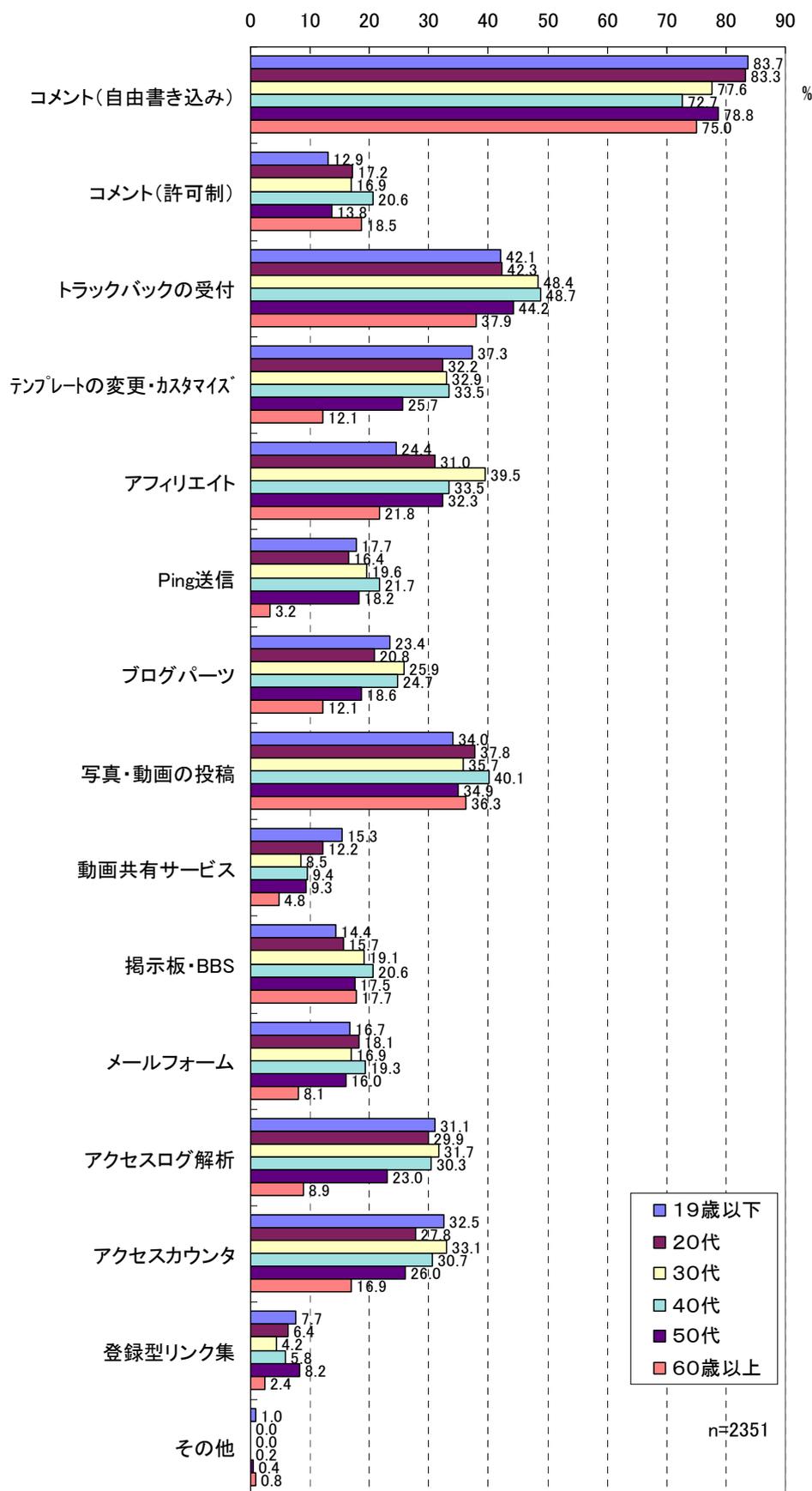
ブログの付加機能の利用状況を見ると、男性よりも女性の方が利用割合の高い機能が多い。特にテンプレートの変更・カスタマイズや、写真・動画の投稿、メールフォームは女性の方が5%以上、男性より利用率が高くなっている(図表3-3-3)。

年代別では、高齢になるほど利用率が低くなる傾向があるものの、コメントや写真・動画の投稿ではあまり年代による違いは見られない。テンプレートの変更・カスタマイズ、動画共有、ブログパーツは若い年代ほど利用が盛んな傾向がある(図表3-3-4)。

図表 3-3-3 利用している付加機能 (性別クロス 複数回答)



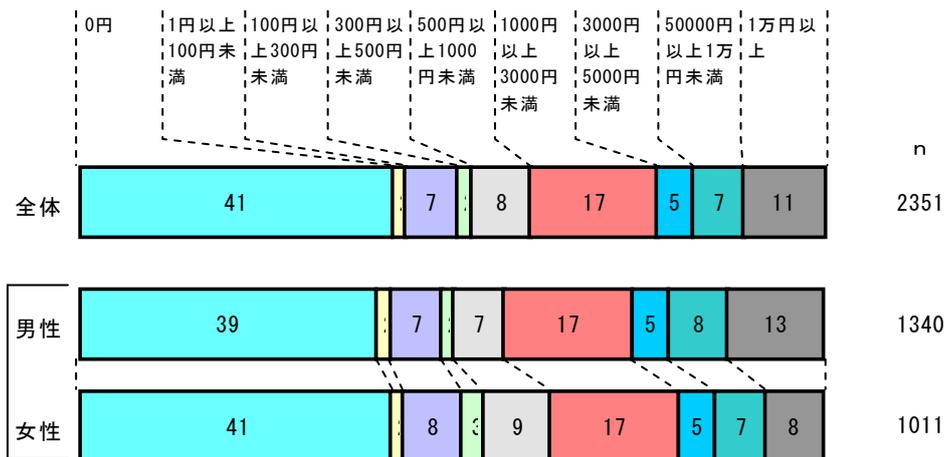
図表 3-3-4 利用している付加機能 (年代別、複数回答)



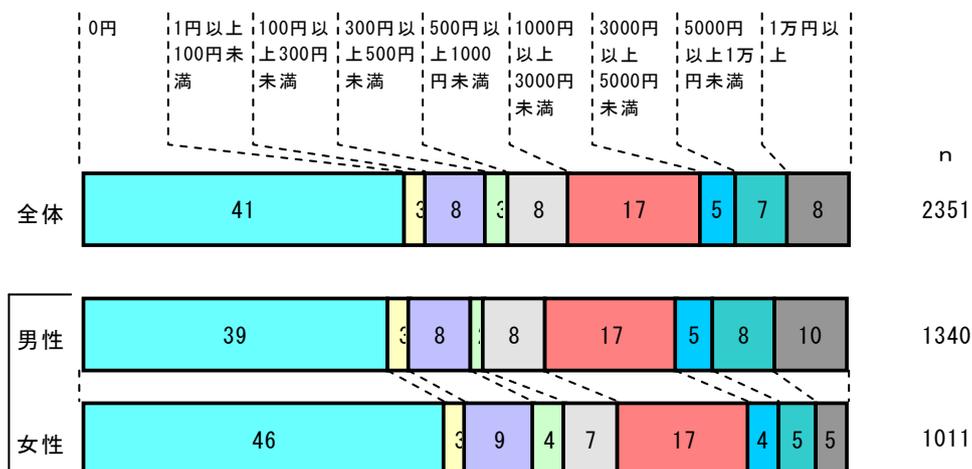
(3) ブログ利用のメリット、コスト、支払意向額

ブログから得られるメリットやコストを見ると、男性の方がメリット・コストともに高く見積もる傾向がある。またブログが有料だった場合の支払意向額についても、男性の方が高い傾向がある（図表 3-3-5、3-3-6、3-3-7）。

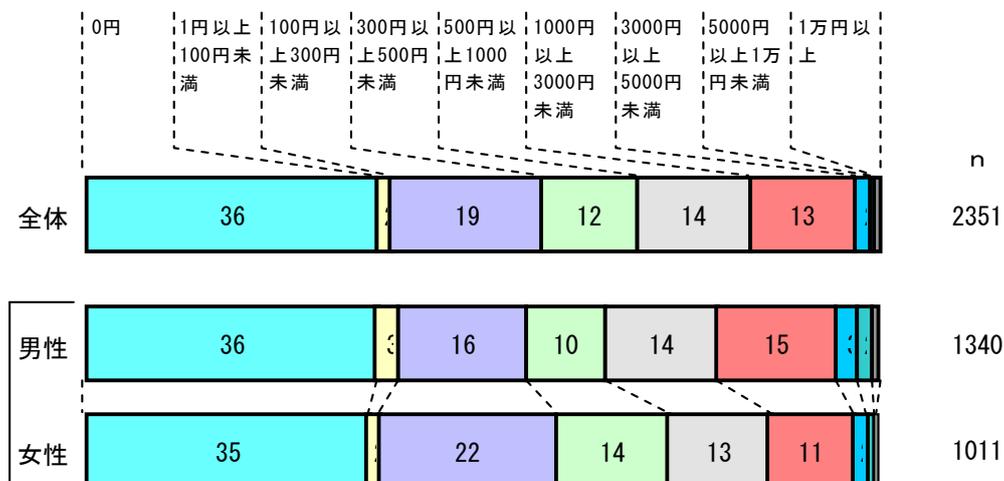
図表 3-3-5 ブログ利用のメリット（性別クロス）



図表 3-3-6 ブログ利用のコスト（性別クロス）

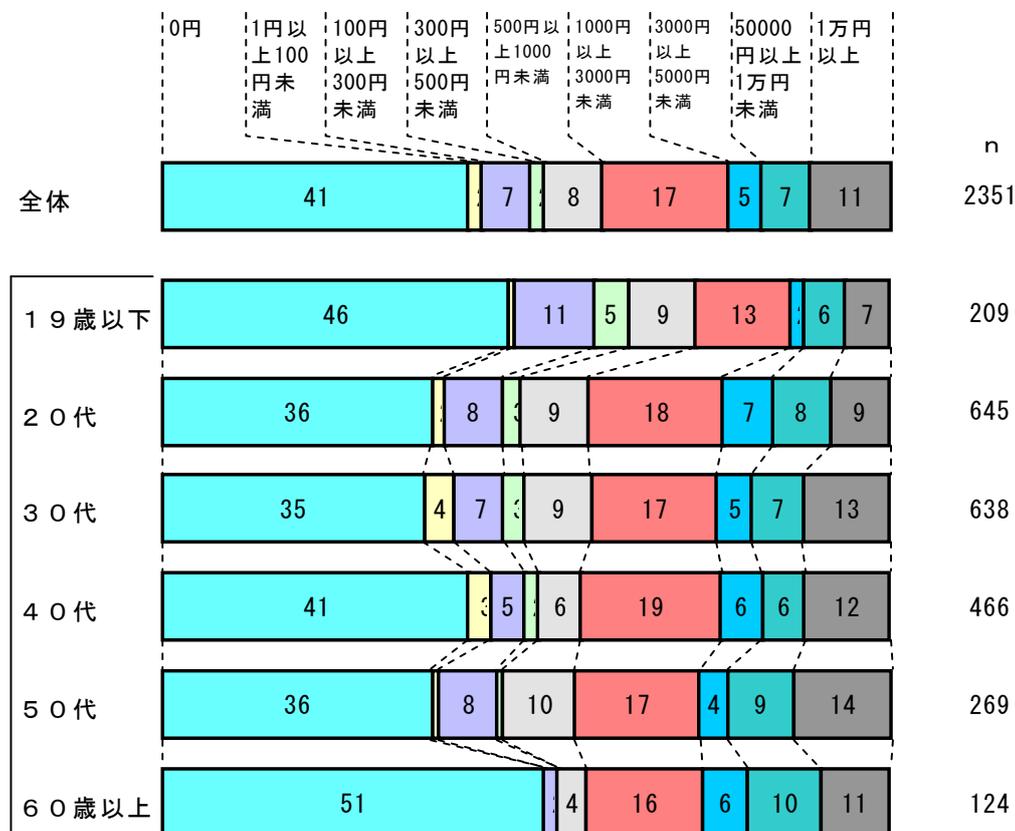


図表 3-3-7 ブログサービスへの支払意向額（性別クロス）

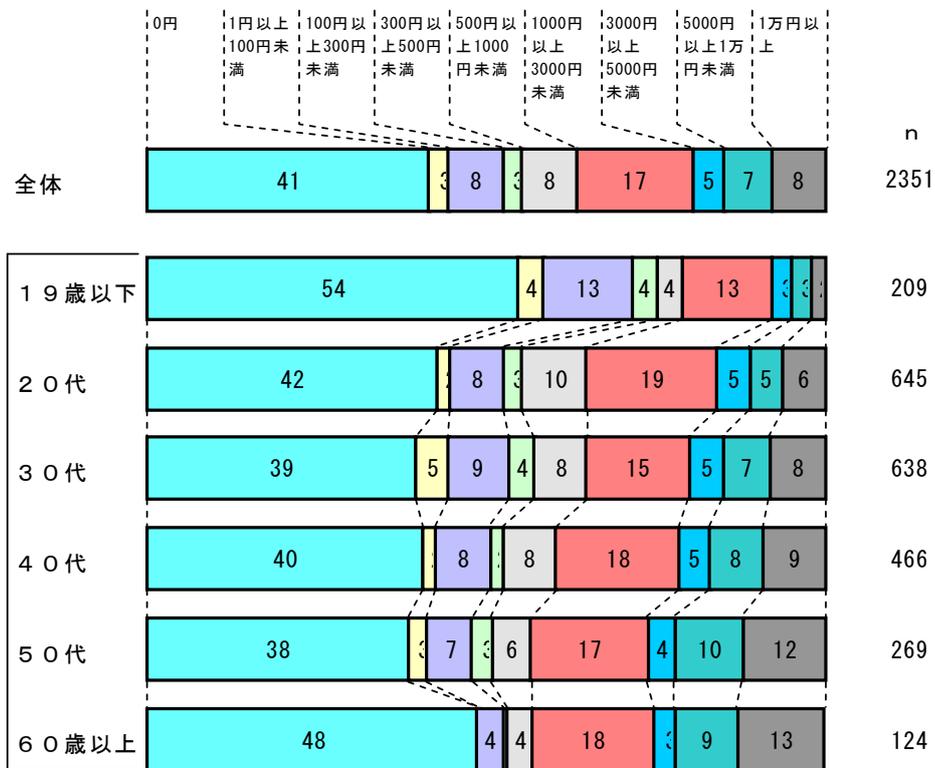


ブログサービスへの支払意向額は、年齢別に見ると、上の年代になるほど高くなる傾向がある。支払意向額はメリットとコストの評価との関連が薄く、年収など他の要因が影響していると考えられる（図表 3-3-8、3-3-9、3-3-10）。

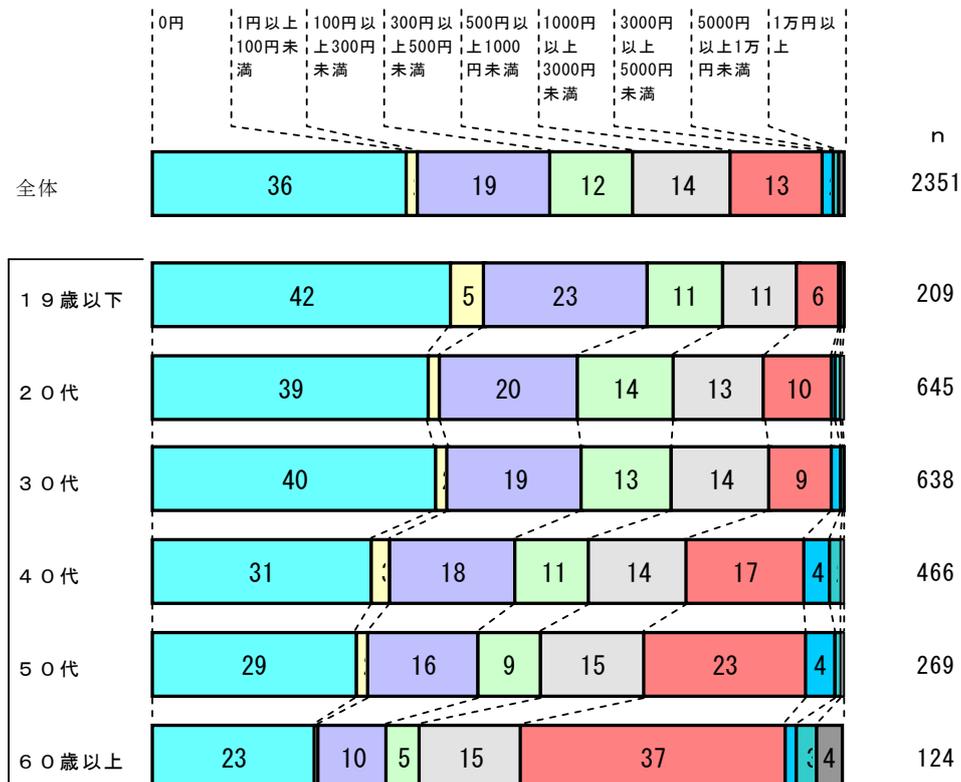
図表 3-3-8 ブログ利用のメリット（年代別クロス）



図表 3-3-9 ブログ利用のコスト（年代別クロス）



図表 3-3-10 ブログサービスへの支払意向額（年代別クロス）

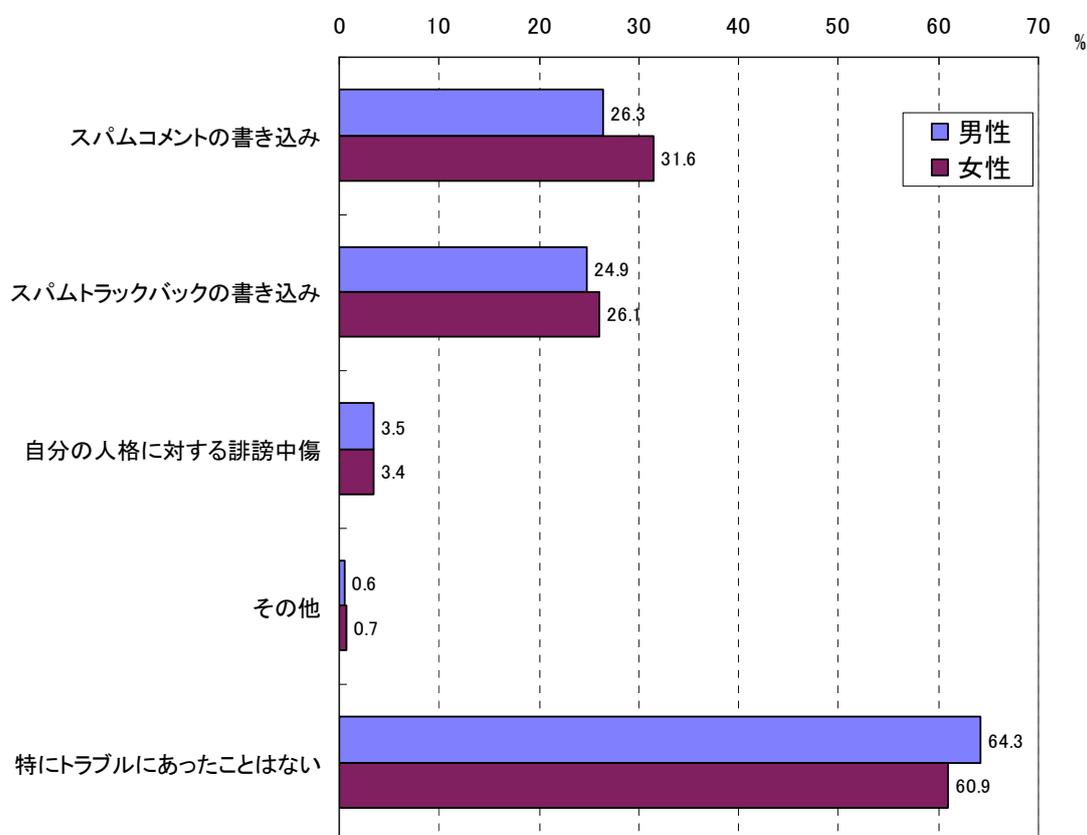


(4) ブログであったトラブル

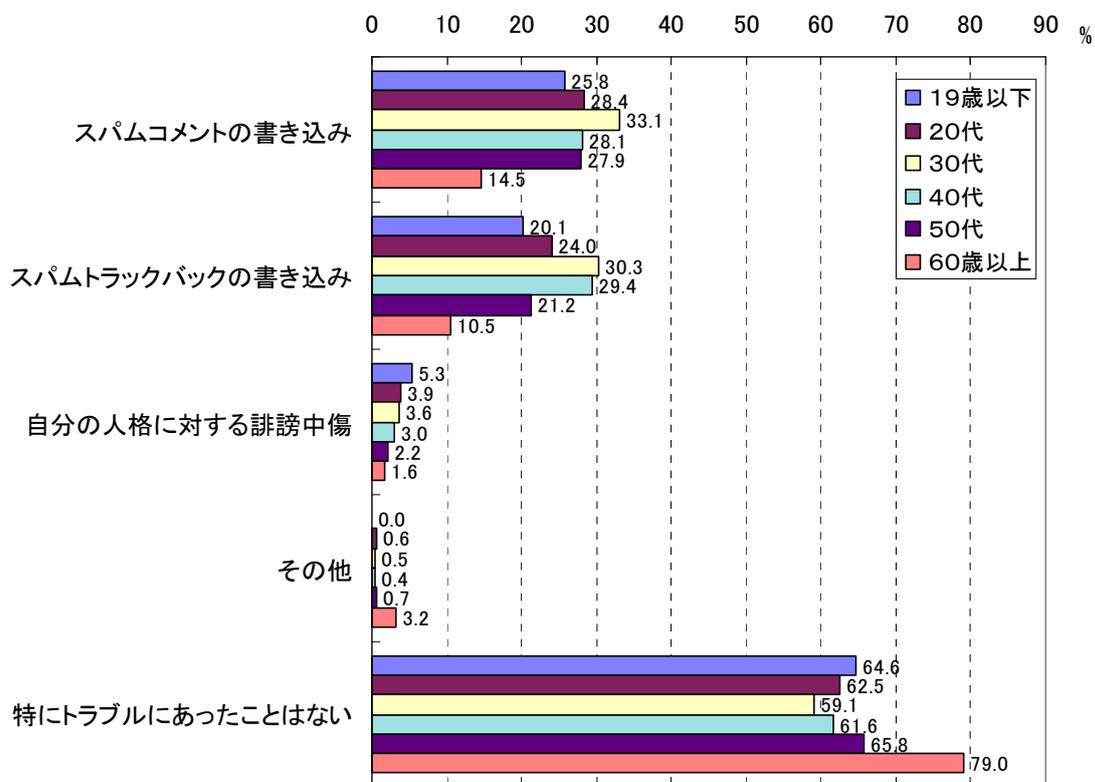
ブログでトラブルにあった経験について性別に見ると、男女共に6割以上の回答者は「特にトラブルにあったことはない」と回答している。トラブル経験率は女性の方が3%程度高く、スパムコメントやスパムトラックバックを受けたと答える割合も女性の方が高い(図表3-3-11)。

年代では見ると、スパムコメント・スパムトラックバックについては、30代40代の経験率が高い一方で、「自分の人格に対する誹謗中傷」については、若い世代ほどトラブルにあう割合が高い傾向がある(図表3-3-12)。

図表3-3-11 ブログであったトラブル(性別クロス)(複数回答)



図表 3-3-12 ブログであったトラブル（年代別クロス）（複数回答）



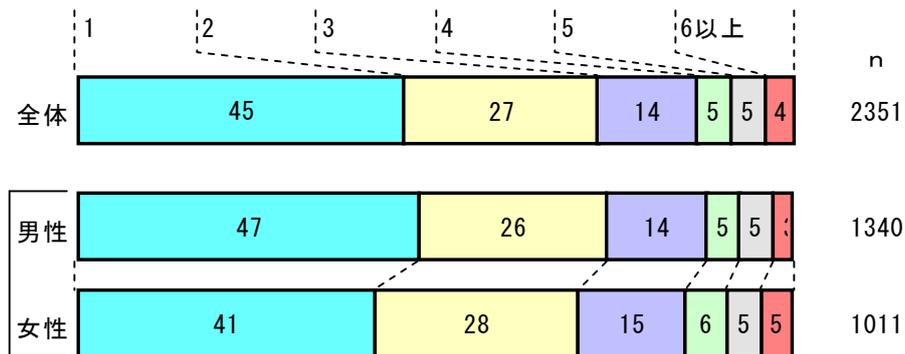
3-3-2 属性別の運営状況

(1) 開設ブログ数

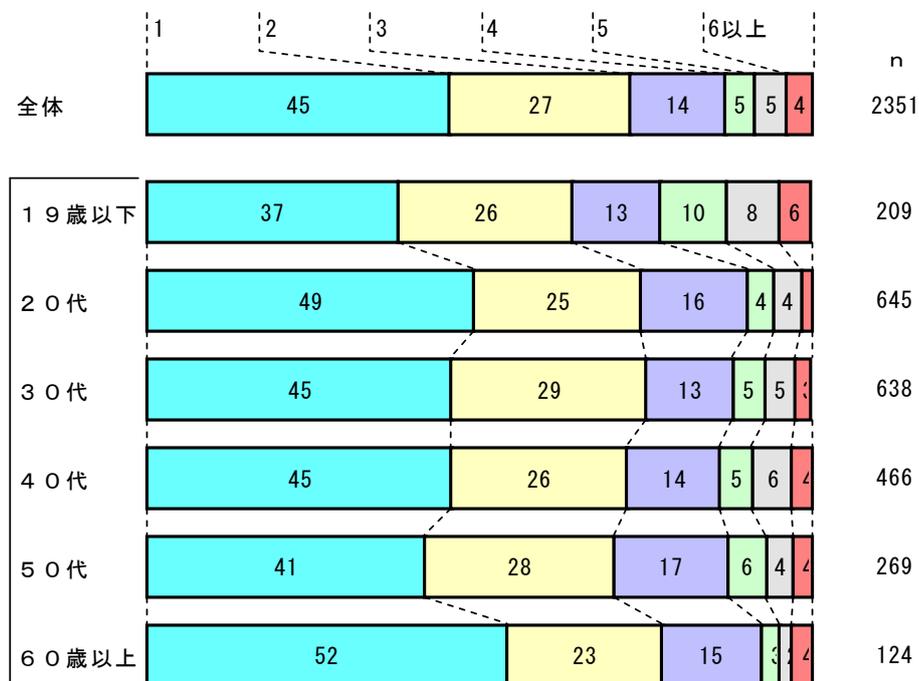
男女共に半数以上が複数のブログを立ち上げているが、女性の方がブログを複数立ち上げる割合が高い（図表 3-3-13）。

年代別には、19歳以下・50代がブログを複数立ち上げる割合が高い（図表 3-3-14）。

図表 3-3-13 開設ブログ数（性別クロス）



図表 3-3-14 開設ブログ数（年代別クロス）



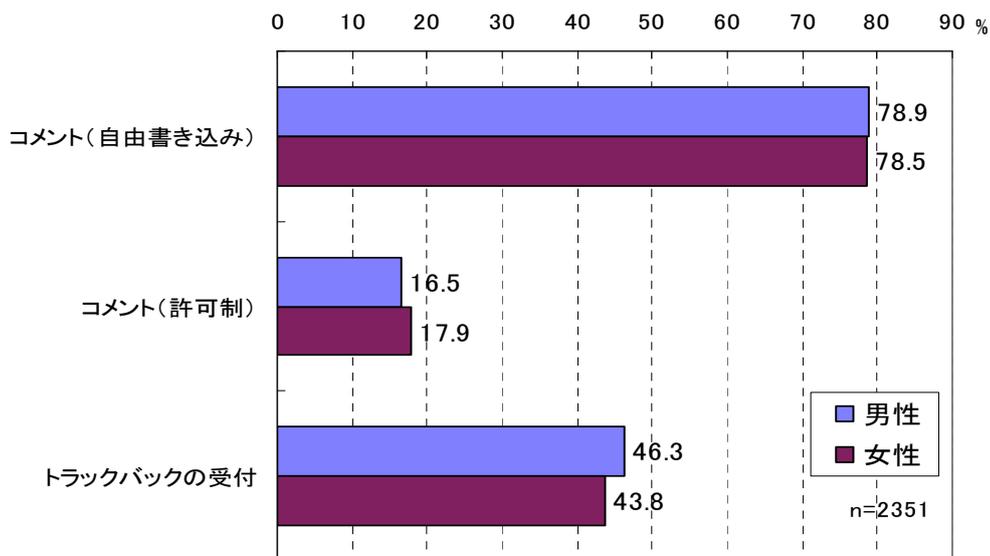
(2) コメントとトラックバックの受付状況

コメント(自由書き込み)は、性別による違いはなくとも8割弱の利用率であるが、

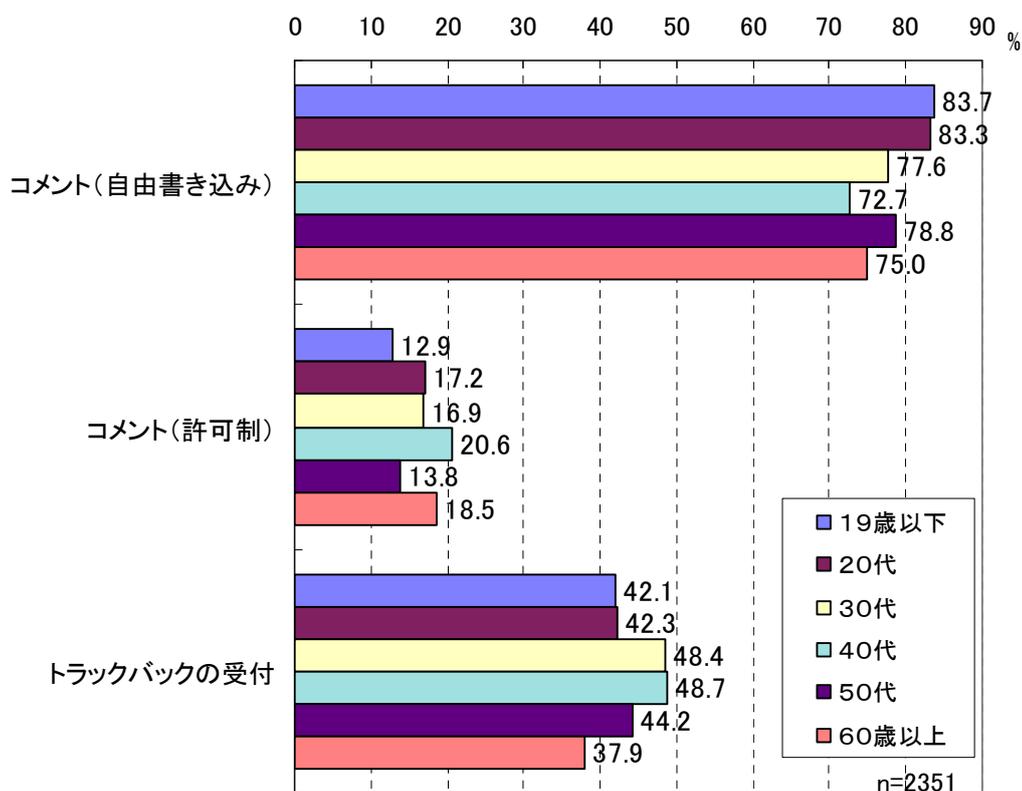
トラックバックの受付は男性の方が若干高く、46.3%となっている（図表 3-3-15）。

年代別では、コメントは、若い年代ほど利用率が高い傾向があるものの、60 歳以上でも 75%が利用している。トラックバックの受付は 30 代・40 代の利用率が他の世代より若干高くなっている（図表 3-3-16）。

図表 3-3-15 コメント・トラックバックの受付状況（性別クロス）



図表 3-3-16 コメント・トラックバックの受付状況（年代別クロス）

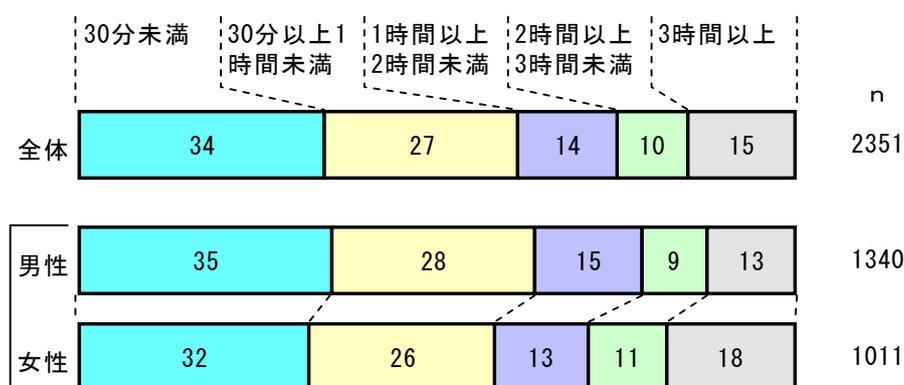


(3) ブログ更新時間

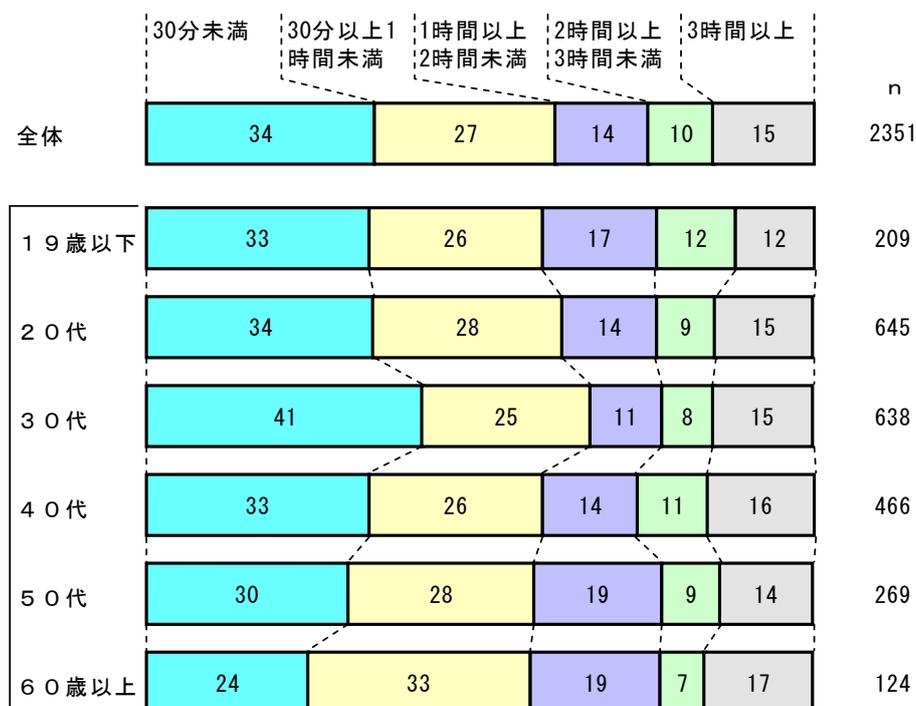
ブログ更新時間は、男性よりも女性の方が週に1時間以上更新に時間をかける回答の割合が高い（図表3-3-17）。

年代別では30代が平均的なブログ更新時間が短い傾向があり、約66%が週に1時間未満となっている。逆に19歳以下と40代以上で週に1時間以上更新に時間をかける割合がやや高く、40%強が週に1時間以上となっている（図表3-3-18）。

図表 3-3-17 ブログ更新時間（週）（性別クロス）



図表 3-3-18 ブログ更新時間（週）（年代別クロス）



3-4 ブログ開設者のグループ分け

3-4-1 ブログ開設者のグループ分けの実施

前節では、性別・年代という基本属性に従ってブログ開設者の特徴を検討した。本節では、ブログの利用動機と利用スタイルに関する設問の回答を基に、因子分析・クラスター分析の多変量解析を行い、ブログをなぜ・どのように利用しているかに関する特徴に基づくブログ開設者のグループ分類を行う。

3-1-1 で述べたとおり、本調査ではブログ利用の動機とスタイルについて、指向性（自己指向／特定関係指向／一般関係指向）と、ブログ内容（事実／知識・ノウハウ／心情）の2つを軸に、分類できるという仮説をもっている。

したがって、アンケートで質問した「ブログの開設・更新の動機」と「ブログの想定読者と記述内容」の2問を変数に因子分析を行った。2設問の詳細を図表3-4-1に示す。質問はいずれも全アンケート回答者（2351件）が5件法（「非常に重要／まあ重要／どちらでもない／あまり重要でない／まったく重要でない」）によって回答している。

図表 3-4-1 対象設問の項目一覧

【ブログの開設・更新の動機】	【ブログの想定読者と記述内容】
多くの人と知り合いになるため	自分に向けて書き、事実・体験をそのまま書く
同じ関心がある人と知り合うため	自分に向けて書き、自分の知識・ノウハウについて書く
アフィリエイトなどで金銭的な収益を得るため	自分に向けて書き、自分の心情・意見を書く
自分と異なる意見・見方や、新たな情報を知るため	特定の他者に向けて書き、事実・体験をそのまま書く
面白い内容を提供して楽しんでもらうため	特定の他者に向けて書き、自分の知識・ノウハウについて書く
自分の記憶や知識を整理するため	特定の他者に向けて書き、自分の心情・意見を書く
文章の練習をするため	世間一般に向けて書き、事実・体験をそのまま書く
自分のことをよく知るため	世間一般に向けて書き、自分の知識・ノウハウについて書く
本当の自分を表現するため	世間一般に向けて書き、自分の心情・意見を書く
ストレスを発散させるため	効果的に集客し、アフィリエイト収入を得る
不安や緊張をとりさり、安心するため	
自分のもっている情報を役に立ててもらうため	
自分の専門知識やノウハウを役に立ててもらうため	
ブログの流行におくれないため	

上記設問について、スクリープロットを元に因子数を7つと仮定し、因子負荷量の低い質問を除外した。全ての項目で十分な因子負荷量が得られた段階で、7因子での全体の分散に対する寄与率は66.4%であった（カイ2乗=1592.3、df=84、p<0.001）。この結果を元に、7因子に命名を行った（図表3-4-2）。

図表 3-4-2 3 回目の因子分析の結果 (7 因子)

因子名	回転後の因子行列(a)	因子						
		1	2	3	4	5	6	7
メンタル効用指向	自分のことをよく知る	0.78	0.05	0.09	0.07	0.18	0.01	0.13
	本当の自分を表現する	0.76	0.10	0.13	0.11	0.14	-0.04	0.09
	不安や緊張をとり安心する	0.67	0.15	0.04	0.14	0.13	0.06	0.08
	ストレスを発散させる	0.65	0.12	0.07	0.08	0.14	0.00	0.00
	文章の練習をする	0.61	0.03	0.07	0.06	0.13	0.08	0.11
特定他者指向	特定の他者に向けて書き、事実・体験をそのまま書く	0.15	0.79	0.11	0.16	0.11	0.05	0.10
	特定の他者に向けて書き、自分の知識・ノウハウについて書く	0.08	0.77	0.16	0.15	0.06	0.12	0.28
	特定の他者に向けて書き、自分の心情・意見を書く	0.18	0.77	0.19	0.17	0.08	0.03	0.03
社会一般指向	世間一般に向けて書き、事実・体験をそのまま書く	0.12	0.12	0.79	0.15	0.10	0.07	0.10
	世間一般に向けて書き、自分の知識・ノウハウについて書く	0.19	0.16	0.79	0.16	0.08	0.08	0.08
	世間一般に向けて書き、自分の心情・意見を書く	0.06	0.20	0.71	0.13	0.07	0.13	0.34
	同じ関心がある人と知り合う	0.10	0.17	0.11	0.84	0.05	-0.02	0.12
コミュニティ効用指向	多くの人と知り合いになるため	0.14	0.17	0.15	0.76	0.09	0.10	0.08
	異なる意見や情報を知る	0.20	0.12	0.17	0.49	0.04	0.13	0.18
	自分に書いて書き、事実・体験をそのまま書く	0.21	0.05	0.05	0.06	0.81	-0.02	-0.03
自己指向	自分に書いて書き、自分の知識・ノウハウについて書く	0.20	0.12	0.10	0.06	0.69	0.06	0.26
	自分に書いて書き、自分の心情・意見を書く	0.38	0.07	0.09	0.06	0.64	-0.08	-0.06
	金銭的な収益を得る	0.05	0.04	0.07	0.13	0.00	0.89	0.10
経済的収益効用指向	効果的に集客し、アフィリエイト収入を得る	0.02	0.11	0.13	0.03	-0.03	0.88	0.08
	自分の専門知識を役立ててもらおう	0.16	0.16	0.20	0.17	0.07	0.13	0.84
知識・情報発信効用指向	自分の情報を役立てる	0.17	0.18	0.22	0.20	0.04	0.08	0.74

因子抽出法: 最尤法 回転法: Kaiser の正規化を伴うハリマックス法

なお、ブログの想定読者と記述内容は、ブログの内容（事実・体験／知識・ノウハウ／心情・意見）よりも、想定する読者（自分自身／特定の他者／世間一般）の影響が強く反映された因子構造となった。

各因子の特徴を概説する。

- ① メンタル効用指向：自分のことを知る・表現する、ストレスを発散するなど、ブログ運営において心理的なメリットを重視する因子
- ② 特定他者指向：友人など特定の読者を意識してブログを書く因子
- ③ 社会一般指向：世間一般を読者として意識してブログを書く因子
- ④ コミュニティ効用指向：ブログを通じて多くの人や同じ関心の人と出会うことを重視する因子
- ⑤ 自己指向：自分自身を読者として意識してブログを書く因子
- ⑥ 経済的収益効用指向：アフィリエイト等ブログ運営を通じて収益を得ることを重視する因子
- ⑦ 知識・情報発信効用指向：ブログを通じて自分の知識や情報を人に役立ててもらおうことを重視する因子

7 因子について算出された因子得点を元に、回答者のクラスタ分析を行い、ブログ開

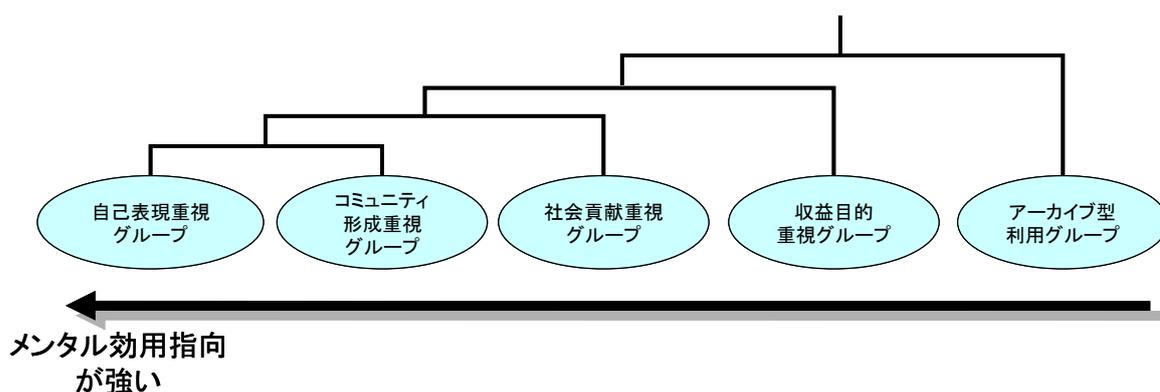
設者のグループを抽出した。クラスタ分析に当たり、クラスタ化の手法による偏りに注意するため、Ward 法、最遠隣法、K 平均法の 3 手法でクラスタ分析を実施し、結果の比較を行った。クラスタ数については、因子数が 7 つであることから、4~7 と仮定し、各場合についてクラスタ分析を実施した。クラスタ分析はできるだけ手法による偏りのない方法を検討の上、Ward 法を用いて分析を行った。

Ward 法で 4~7 分類のクラスタ分析を行い、各ケースにおけるクラスタの特徴を解釈し、結果の比較を行った（詳細は参考 4 参照）。この結果、5 分類の場合が最も各クラスタの特徴を解釈しやすかった。5 分類の因子得点の特徴を元に、各クラスタを以下のように命名した。

- ① 自己表現重視グループ
- ② コミュニティ形成重視グループ
- ③ 社会貢献重視グループ
- ④ 収益目的重視グループ
- ⑤ アーカイブ型利用グループ

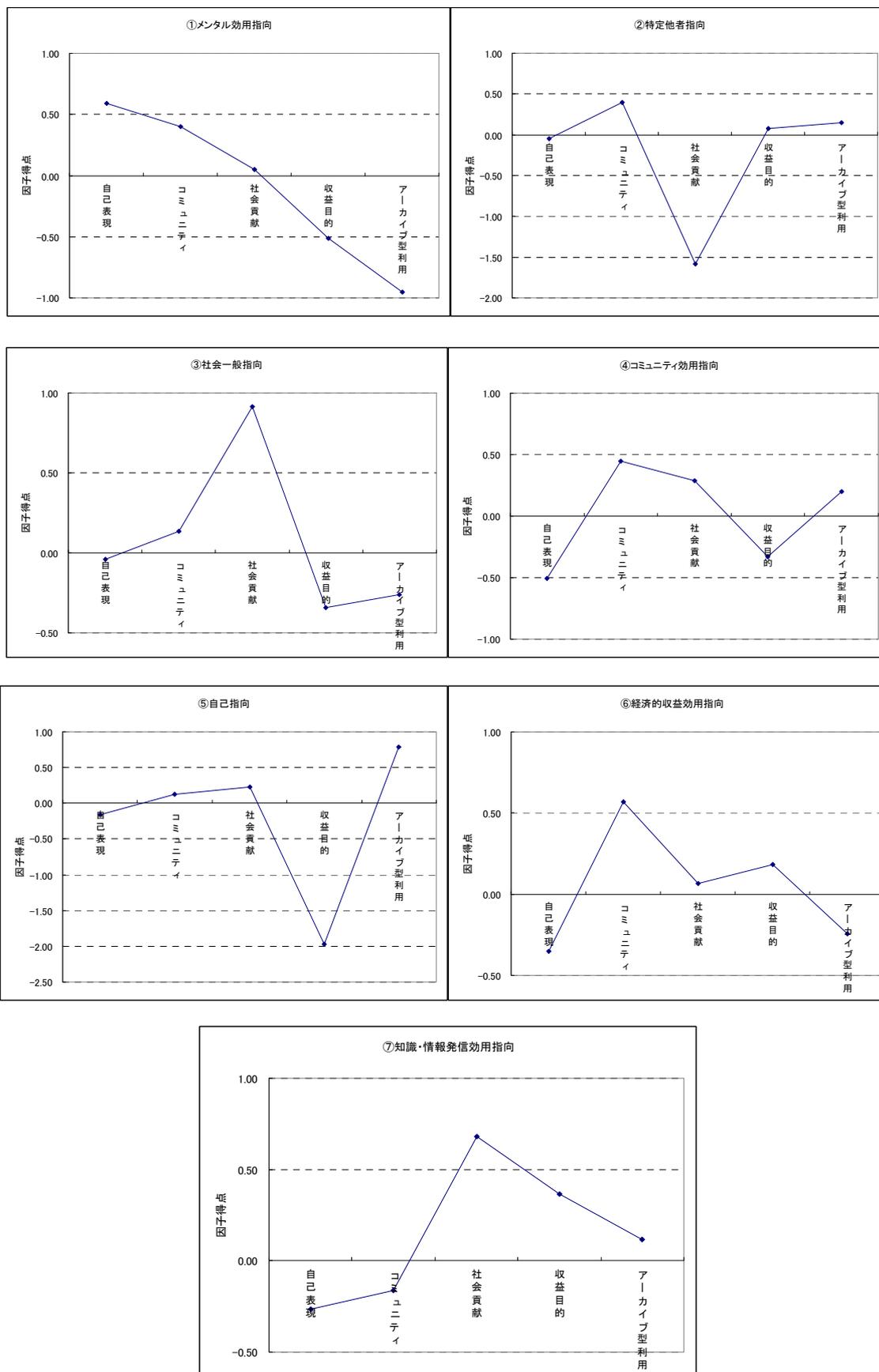
デンドログラム（階層型クラスタにおいて、近い回答者同士をクラスタ化していく過程を示した樹状図）を確認すると、5 つのクラスタは、第 1 因子であるメンタル効用指向の影響の強さの順に分かれていることが分かる。

図表 3-4-3 デンドログラムにおける 5 分類の位置付け



5 分類の場合の 7 因子に対する因子得点は、図表 3-4-4 の通りである。

図表 3-4-4 5 分類の場合の 7 因子に対する因子得点



7 因子の得点を元に、各グループの特徴を概説する。なお、グループ名がそれぞれ長くなるため、クロス分析結果のグラフではカッコ内の略称を用いて表記する。

① 自己表現重視グループ（自己表現）

自己表現重視グループは、メンタル効用指向が高く、自己の心情・意見や事実・体験等を記述し、自己実現やストレス解消などの内面的な効用を重視するグループである。自分のためにブログを書くため、コミュニティ効用指向や経済的収益効用指向が低い傾向がある。

② コミュニティ形成重視グループ（コミュニティ）

コミュニティ形成重視グループは、コミュニティ効用指向が高く、ブログを通じて同じ関心を持つ多くの人と知り合うことを強い利用動機とするグループである。また経済的収益効用指向も高く、アフィリエイトなどブログを通じた収入に高い関心を持つグループでもある。

③ 社会貢献重視グループ（社会貢献）

社会貢献重視グループは、社会一般指向と知識・情報発信効用指向が高く、世間一般に対して広く自分のもつ知識や情報を役立ててもらい、いわば社会貢献的な意識がブログ更新の動機となっているグループである。社会貢献の他、メンタル効用やコミュニティ効用、経済的収益効用も比較的重視する傾向があり、ブログから様々なメリットを享受していることが想定される。

④ 収益目的重視グループ（収益目的）

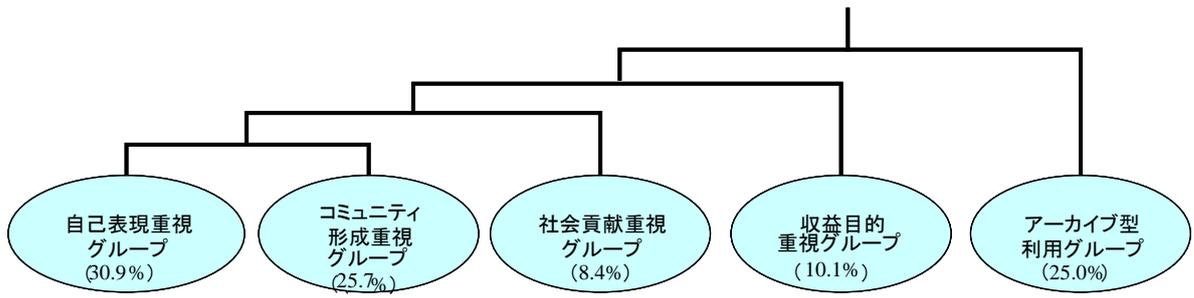
収益目的重視グループは、知識・情報発信効用指向と経済的収益効用指向が高いグループである。経済的収益効用指向だけで見ると、コミュニティ重視型が最も高く、収益目的重視グループは 2 番目であるが、収益目的重視グループはそのほかのメンタル効用・コミュニティ効用を重視する傾向が弱く、自分のもつ知識や情報を同じ関心や趣味ともつネットユーザの役にたててもらい、経済的な収益を得ることが相対的に強いブログ利用動機となっていると考えられる。

⑤ アーカイブ型利用グループ（アーカイブ型）

アーカイブ型利用グループは、自己指向が非常に高い他、知識・情報発信効用指向とコミュニティ効用指向が比較的高いグループである。このタイプは、自分用の情報アーカイブとしてブログを利用し、それを公開することで生まれるコミュニケーションや他者への貢献を効用としていると想定される。

以上 5 分類が回答者全体に占める割合は、図表 3-4-5 の通りである。

図表 3-4-5 5分類の回答者に占める割合



3-4-2 グループ別のクロス分析

3-4-1の5分類を軸に、アンケートの他の設問とクロス分析を行い、各グループの特徴や傾向について、より詳しく分析する。

(1) 基本属性

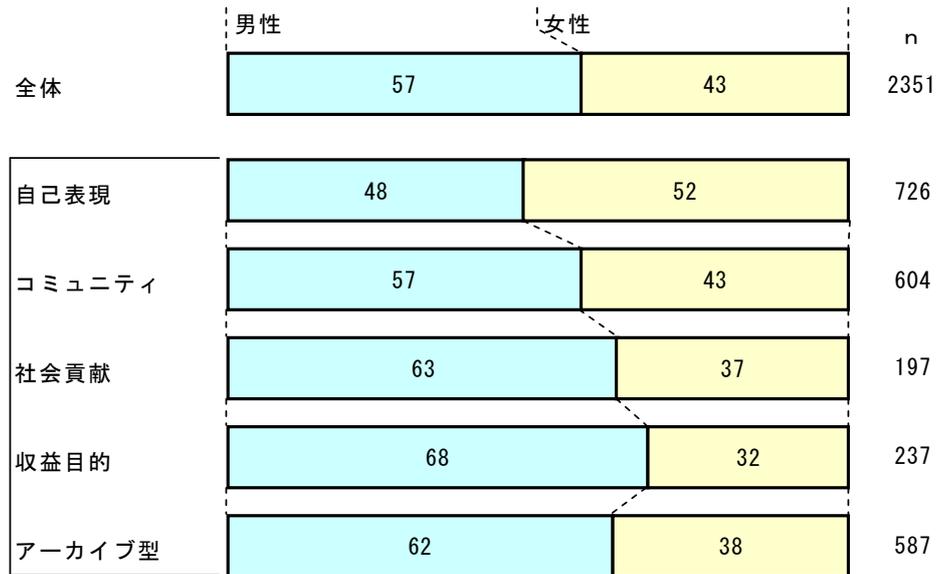
基本属性を見ると、まず年齢では、自己表現重視グループは若い年代の割合が高いのに対して、社会貢献型はより年上の年代の割合が高い（図表3-4-6）。性別では収益目的重視、社会貢献重視、アーカイブ型利用のグループで男性の割合が高く、自己表現重視グループで女性の割合が高い（図表3-4-7）。コミュニティ形成重視グループはいずれも全体平均と近く、特に目立った特徴はない。

基本属性からは、社会貢献重視グループは年長の男性が世間一般に役に立つ情報を発信するという開設者像、自己表現重視グループは若い女性が自分のメンタルな安定を目的に、自分のためのブログを更新するという利用者像が想定される。

図表 3-4-6 年齢（グループ別）

	19歳以下	20代	30代	40代	50代	60歳以上	n
全体	9	28	27	20	11	5	2351
自己表現	11	31	25	18	11	4	726
コミュニティ	8	28	29	20	11	4	604
社会貢献	5	24	25	21	17	8	197
収益目的	12	26	24	23	11	4	237
アーカイブ型	7	25	30	20	11	7	587

図表 3-4-7 性別（グループ別）

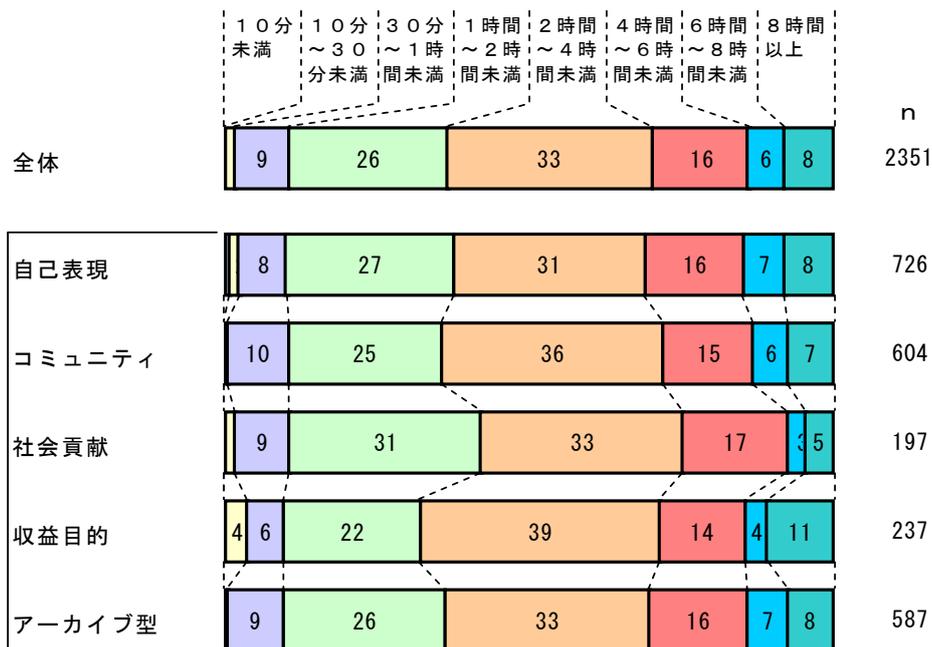


(2) ネット利用の特徴

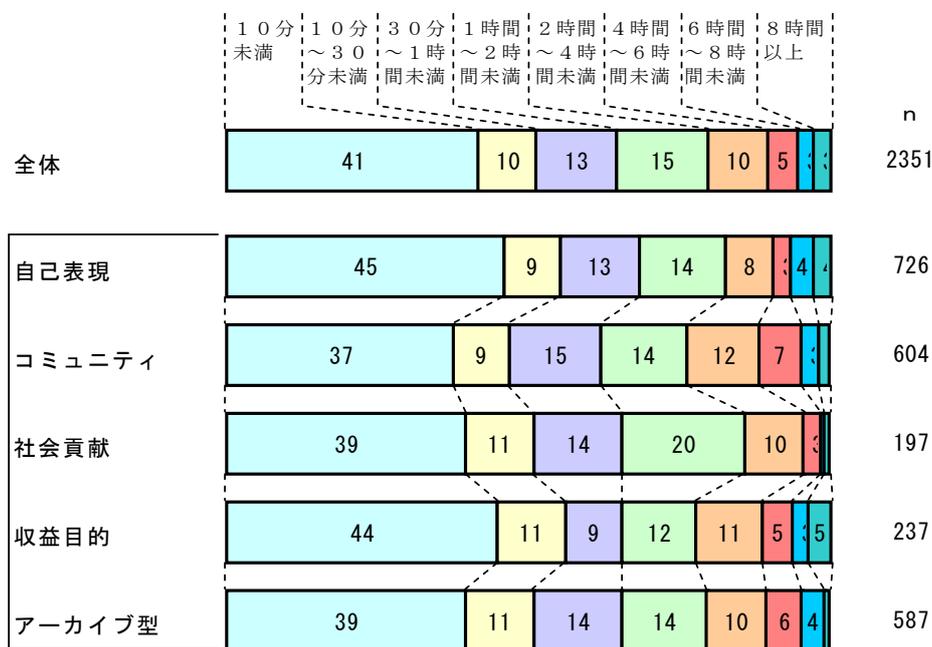
各グループのネット利用の特徴について、個人目的又は仕事目的のネットの利用時間とブログ以外のネットサービスの利用状況を見る。

ネット利用時間について、社会貢献重視グループは個人利用の時間が短く、自己表現重視グループは、仕事での利用時間が短い傾向がある（図表 3-4-8、3-4-9）。仕事のネット利用時間は、概ね男性の割合の高いユーザタイプで長くなっており、性別の影響が大きいと見られる。

図表 3-4-8 一日のネット利用時間（個人）（グループ別）



図表 3-4-9 一日のネット利用時間（仕事）（グループ別）



ブログ以外のネットサービスの利用状況を見ると、コミュニティ形成重視グループは、動画配信やオンラインゲームなどのサービスの利用率が高い傾向がみられる（図表 3-4-10、3-4-11）。また SNS やミニブログなど、ブログ以外のコミュニケーションサービスの利用率も高く、コミュニケーション指向の高いネット利用者像が伺える（図表 3-4-12、3-4-13）。

図表 3-4-10 動画配信サービス (グループ別)

	利用している／ たまたま利用して いる	利用したことが ある (今は利用 していない)	利用したことが ない	n
全体	47	16	37	2351
自己表現	45	16	39	726
コミュニティ	51	16	33	604
社会貢献	51	14	35	197
収益目的	48	16	36	237
アーカイブ型	45	15	40	587

図表 3-4-11 オンラインゲーム (グループ別)

	利用している／ たまたま利用して いる	利用したことが ある (今は利用 していない)	利用したことが ない	n
全体	28	27	45	2351
自己表現	26	28	46	726
コミュニティ	35	28	37	604
社会貢献	23	25	52	197
収益目的	26	28	46	237
アーカイブ型	27	25	48	587

図表 3-4-12 SNS (グループ別)

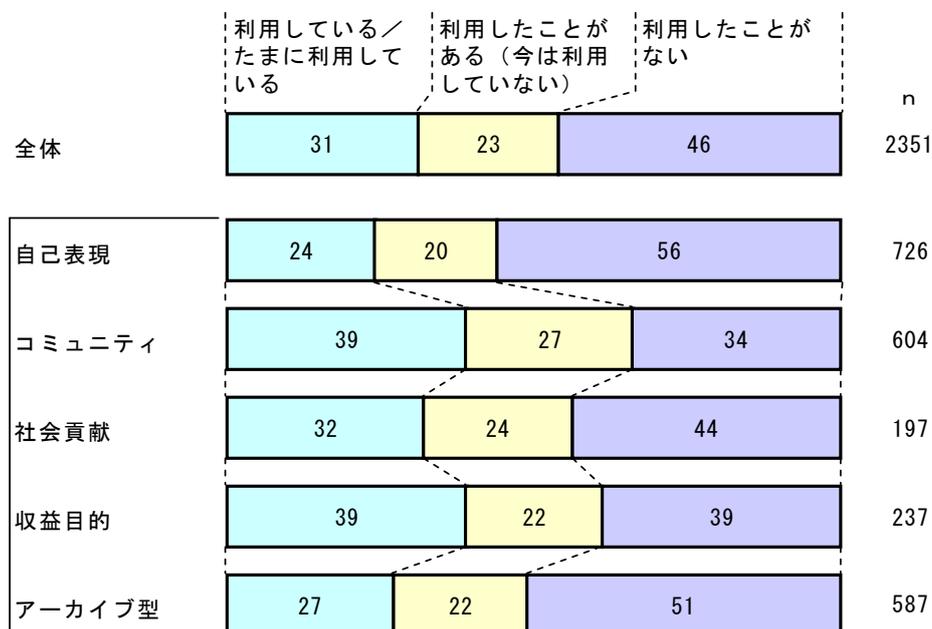
	利用している／ たまたま利用して いる	利用したことが ある (今は利用 していない)	利用したことが ない	n
全体	42	15	43	2351
自己表現	42	16	42	726
コミュニティ	42	19	39	604
社会貢献	38	15	47	197
収益目的	40	14	46	237
アーカイブ型	43	12	45	587

図表 3-4-13 ミニブログ (グループ別)

	利用している／ たまたま利用して いる	利用したことが ある (今は利用 していない)	利用したことが ない	n
全体	10	12	78	2351
自己表現	9	11	80	726
コミュニティ	15	20	65	604
社会貢献	9	10	81	197
収益目的	7	8	85	237
アーカイブ型	8	9	83	587

アフィリエイトについては、経済的収益効用指向の高さに比例して、コミュニティ形成重視グループと収益目的重視グループで利用率が高い (図表 3-4-14)。

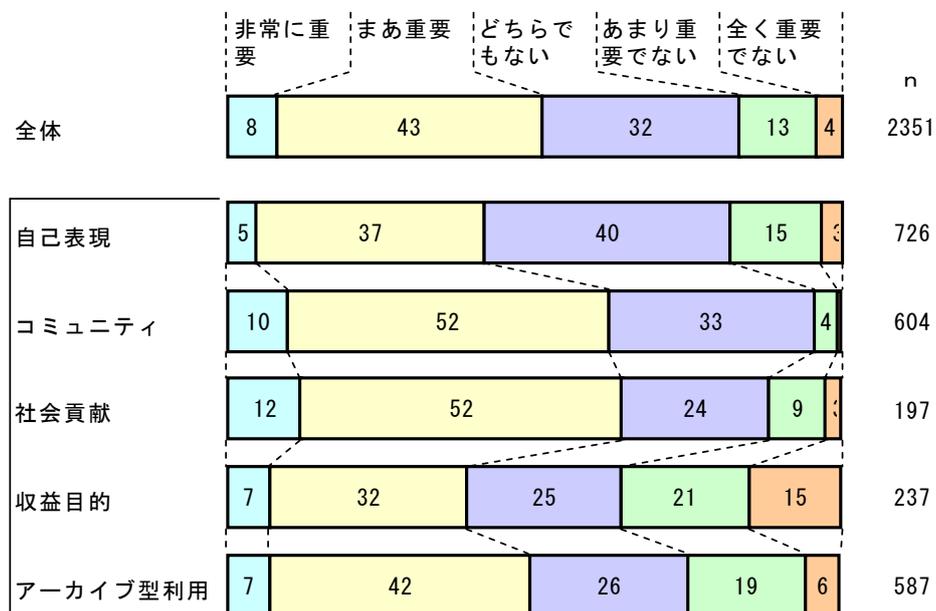
図表 3-4-14 アフィリエイト（グループ別）



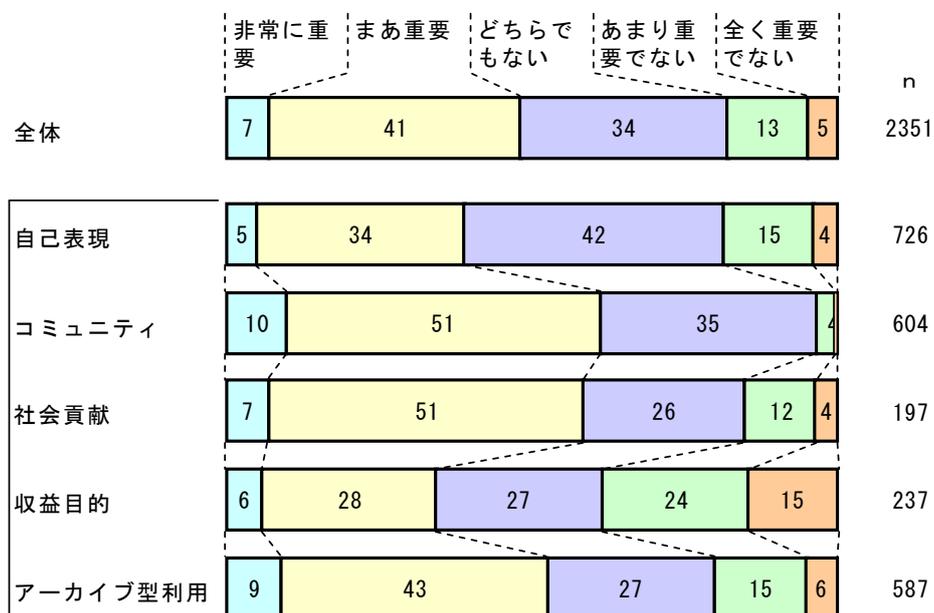
（3）ブログ利用の特徴

ブログに書く内容について、世間一般の社会的な事実や出来事、地域やコミュニティの事実や出来事、個人的な事実や出来事の3つをそれぞれの程度重視しているかについては、グループタイプに関わらず個人的な事実や出来事が最も重視されている（図表 3-4-15、3-4-16、3-4-17）。収益目的重視グループはいずれの内容についても3割以上の回答者が「あまり重要でない」、「全く重要でない」と答えている。

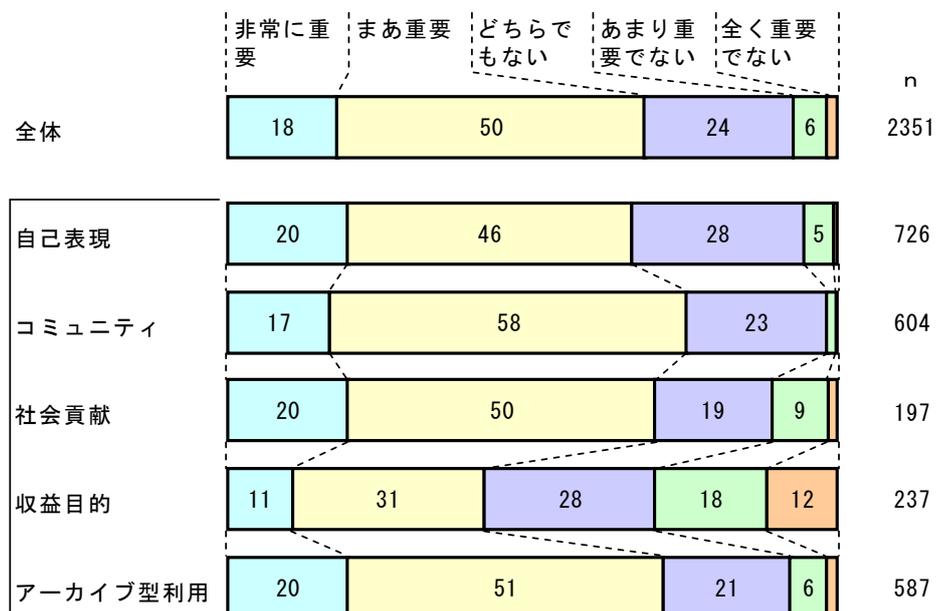
図表 3-4-15 世間一般の社会的事実や出来事（グループ別）



図表 3-4-16 地域やコミュニティの事実や出来事（グループ別）

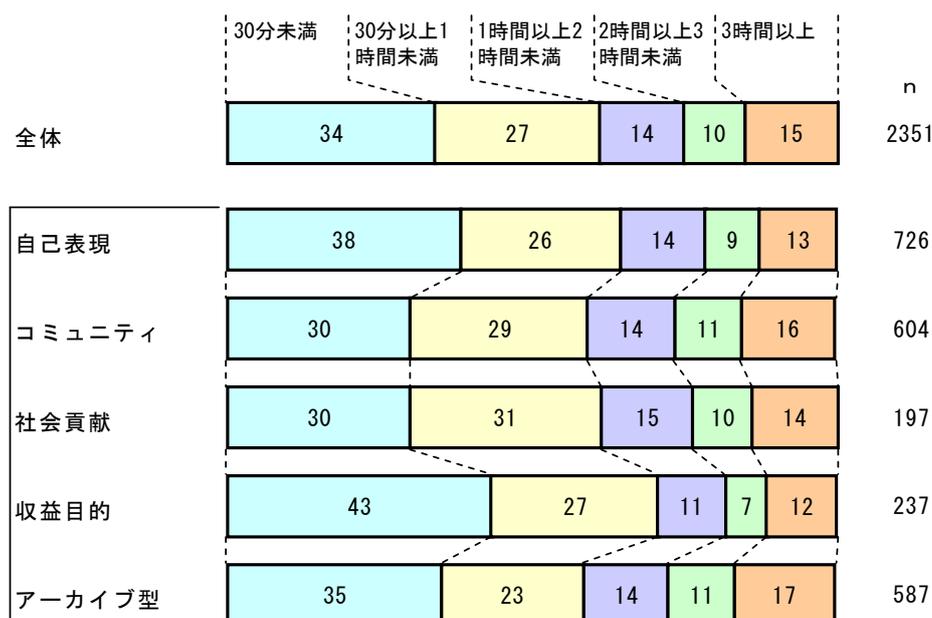


図表 3-4-17 個人的な事実や出来事（グループ別）



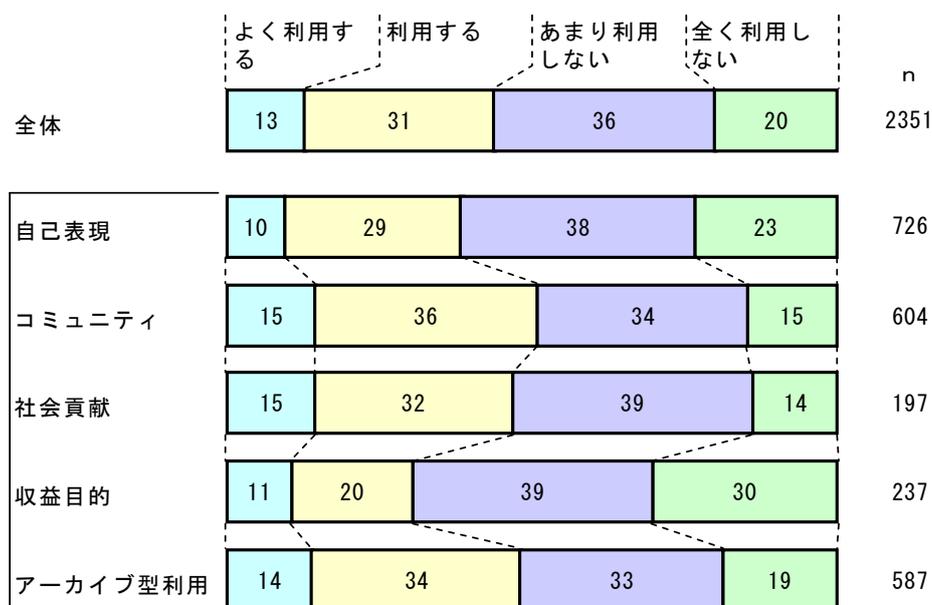
ブログの更新時間では、コミュニティ形成重視、アーカイブ型利用、社会貢献重視のグループの約 4 割の回答者が 1 時間以上と答えている一方、収益目的重視グループや自己表現重視グループの約 4 割の回答者が 30 分未満と答えており、グループによってブログ更新にかかる時間に大きな差が見られる（図表 3-4-18）。

図表 3-4-18 ブログの平均的な週間更新時間（グループ別）

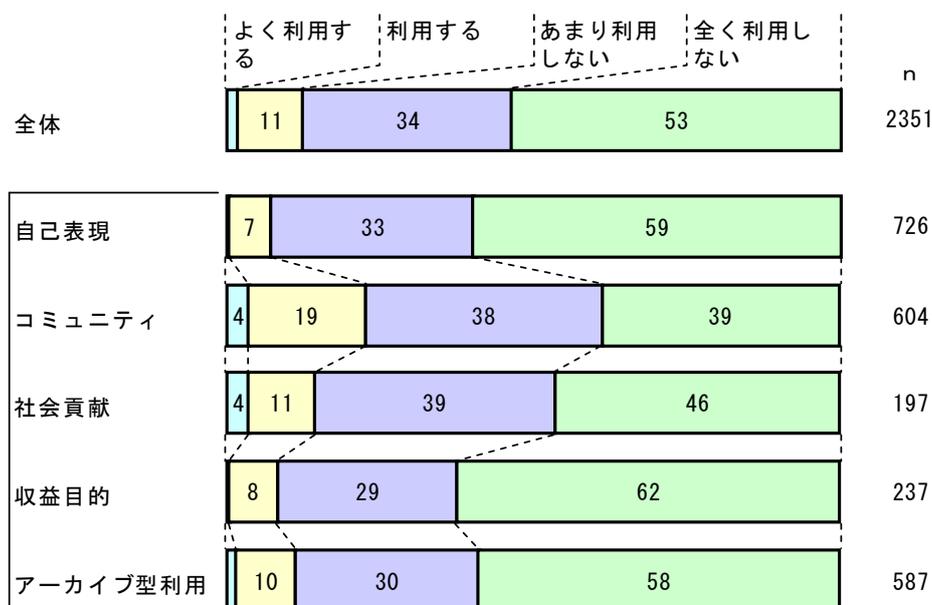


他ブログへのコメントやトラックバックは、コミュニティ形成重視グループの利用率が高い。コミュニティ指向因子の強さと相関が見られ、社会貢献重視グループとアーカイブ型利用グループでも比較的利用率が高い傾向がある（図表 3-4-19、3-4-20）。

図表 3-4-19 他ブログへのコメント（グループ別）



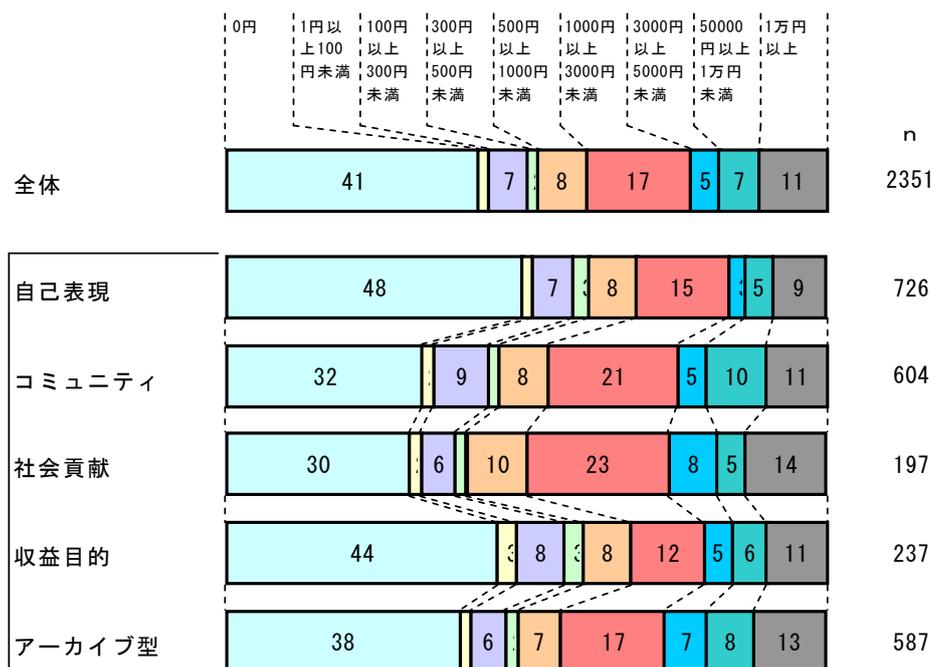
図表 3-4-20 他ブログへのトラックバック（グループ別）



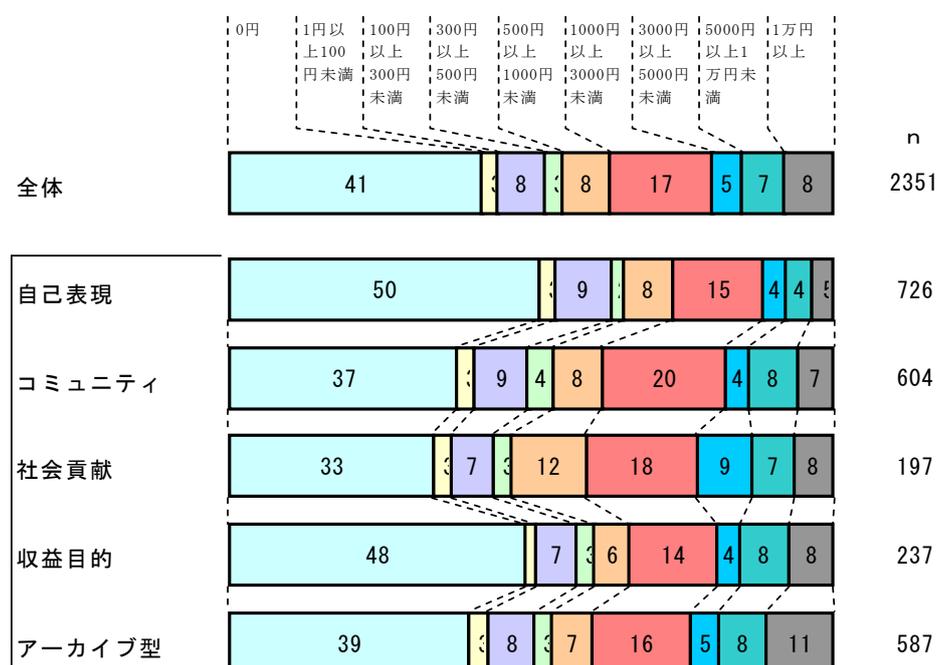
ブログ利用のメリットの経済的評価額とブログサービスが有料だった場合の支払意向額については、社会貢献重視グループとコミュニティ形成重視グループが平均より比

較的高い（図表 3-4-21、3-4-22）。この 2 つのタイプは因子得点においても、メンタル効用、コミュニティ効用、経済的収益効用のいずれも相対的に高い傾向があり、様々な意味でブログから効用を得ていることが、評価が高くなる理由と考えられる。逆に効用が全体に低く、メンタル効用か経済的収益効用などに限定される自己表現重視グループと収益目的重視グループは評価が低くなっている。

図表 3-4-21 ブログ利用のメリットの経済的評価（グループ別）



図表 3-4-22 ブログサービスが有料だった場合の支払意向額（グループ別）



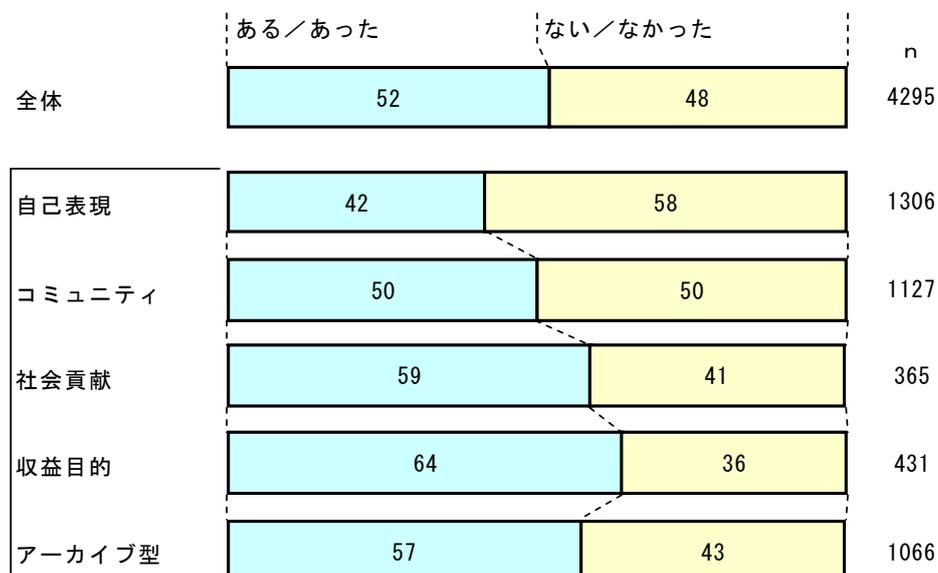
（４）運営するブログの特徴

ブログテーマの有無は、収益目的重視、社会貢献重視のグループで特定のテーマを有する割合が高い傾向がある。テーマを絞って役立つ情報を提供することで、収益を高めようとしていると考えられる（図表 3-4-23）。

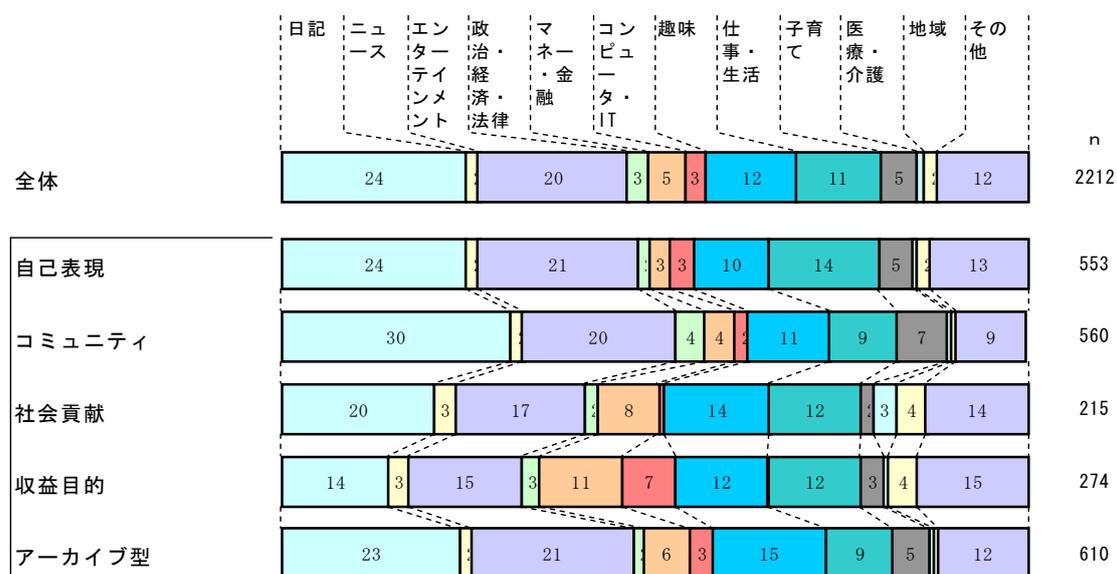
具体的なブログテーマはグループにより傾向が異なる。自己表現重視グループは特定のテーマがない割合が高いことが特徴で、テーマがあるブログは、全体の傾向と比べて特に目立ったものはない（図表 3-4-24）。コミュニティ形成重視グループは、テーマとして「日記」が特に高い他、「子育て」の割合が高い。社会貢献重視グループは、特定のテーマを持つブログの割合が高く、テーマとしては、「マネー・金融」、「地域」、「医療・介護」の割合が特に高い。

収益目的重視グループは「日記」の割合が最も低く、特に「マネー・金融」や「コンピュータ・IT」など専門的なテーマの割合が高い。アーカイブ型利用グループでは、車・バイク、旅行など「趣味」に関するテーマの割合が高い傾向があり、ブログが事実や経験を記録する公開備忘録として利用されていることが伺える。

図表 3-4-23 ブログに特定のテーマがあるか（全回答ブログ／グループ別）



図表 3-4-24 ブログテーマ（全回答ブログ／グループ別）



3-4-3 各グループの特徴

5 グループ中最も大きな割合を占める自己表現重視グループは、女性、10代・20代の割合が高い傾向があり、自己実現やストレス解消など内面的な効用を重視するグループである。

コミュニティ形成重視グループは、性別、年代に明確な特徴はないが、ブログ利用のメリットを高く評価しており、ブログの更新時間が長いという点で、活発にブログを利用するグループであるといえることができる。社会貢献重視グループは、40代以上の男性の割合が高い傾向があり、特定のテーマについて自分の知識を役立ててもらうことを重視するグループである。

収益目的重視グループは、男性、10代・40代の割合が高い傾向があり、専門性の高いテーマをもつ点が特徴である。アーカイブ型利用グループは、男性、30代以上の割合が高く、趣味関係のテーマの割合が高い点が特徴である。

クロス分析に基づき、各グループの特徴をまとめたのが図表 3-4-25 である。

図表 3-4-25 グループごとの特徴

	自己表現重視グループ	コミュニティ形成重視グループ	社会貢献重視グループ	収益目的重視グループ	アーカイブ型利用グループ
回答者に占める割合	30.9%	25.7%	8.4%	10.1%	25.0%
性別	女性の割合が高い	—	男性の割合が高い	男性の割合が高い	男性の割合が高い
年代	10代、20代の割合が高い	—	40代以上の割合が高い	10代、40代の割合が高い	30代以上の割合が高い
ブログ更新において重視すること	自己実現・ストレス解消など内面的効用を重視	ブログを通じた出会い・コミュニケーションを重視	社会一般に自分の知識を役立ててもらうことを重視	ブログから得られる収益を重視	自分用に整理した情報を役立ててもらうことを重視
ブログの主なテーマ	特定テーマを持たない割合が高い	日記、子育てなど	マネー・金融、医療・介護、地域など	マネー・金融、コンピュータなど	趣味関係
ブログ利用のメリット評価		利用メリットを高く評価	利用メリットを高く評価		
その他		ブログの更新時間が長い	ブログの更新時間が長い		ブログの更新時間が長い

3-4-4 指向性・ブログ内容によるブログ開設者分類仮説の検証

最後に、本調査の仮説である、指向性（自己指向／特定関係指向／一般関係指向）と、ブログ内容（事実／知識・ノウハウ／心情）の2軸によるブログ開設者の分類と、5グループの関係について考察する。

分析の結果では、ブログの内容は重要な要因としては抽出されず、5グループの分類に対する説明力は弱い。また、指向性軸との関連についてみると、「自己指向」、「特定他者指向」、「社会一般指向」の3つの因子が抽出されたが（図表3-4-2参照）、コミュニティ形成重視や社会貢献のように、これらのどの指向も強いグループもあり、ブログ開設者の分類において、「指向性軸」ではうまく分類できない状況になっている。

5グループの分類においてもっとも強い因子は「メンタル効用指向」となっており（図表3-4-3参照）、「メンタル効用指向」とその他の「非メンタル効用指向」で大きく分類することも考えられるが、例えばコミュニティ形成重視グループは、「メンタル効用指向」とともに、「コミュニティ効用指向」や「経済的収益効用指向」の得点も高い。また社会貢献重視グループは、「コミュニティ効用指向」と「知識・情報発信効用指向」とともに、「メンタル効用指向」も高いグループとなっている（図表3-4-4参照）。

「指向性軸」や「効用軸」だけでは簡単に説明しきれないことが、現在の我が国のブログの多様性を示していると考えられ、今後の多様な観点からの分析の可能性を示唆しているとも言える。

第4章 ブログの利用における問題点

第2章及び第3章では、ブログコンテンツの発展の推移や、ブログ開設者の利用目的の多様化などについて分析したが、近年、ブログサービスを悪用した迷惑なブログの利用の増加が指摘されている。ブログの迷惑利用の態様は、おおまかに、①他人のブログコンテンツや人気キーワードを悪用し、アフィリエイトなどで収益を得たり、特定のサイトに誘導することを目的とした迷惑なブログ（スパムブログ）、②他人のブログに広告などを書き込む、スパムコメント、スパムトラックバックの2つに分類できる。

また、スパムコメントではないが、ブログ開設者に批判的な書き込みが殺到し、ブログが閉鎖に追い込まれる「炎上」という事態に発展することがある。本章では、このようなブログ利用における負の側面について分析する。

4-1 スパムブログ

スパムブログについての決まった定義はないが、他人のブログコンテンツや人気キーワードを悪用して検索エンジンやブログサイトのポータルからのアクセスを増やし、アフィリエイトで広告収入を得たり、特定のサイトに誘導したりすることを目的としているブログ（記事）などを指すのが一般的である。2005年頃から増え始め、近年、増加傾向が著しいと言われている。

スパムブログの多くは、ブログサイトのアカウントの生成からスパム記事の書き込みまで、プログラムにより自動化して大量に行われるという（2-4-5の図表2-4-6参照）。このようなスパムブログの自動的な作成方法の例としては、次のようなものがあげられる。

●ワードサラダ型

文章をフレーズ単位で機械的に組み合わせて自動生成しているブログ（記事）。文法的には正しいが、閲覧者には意味が通らない文章が掲載されている。

●キーワード抽出型

他ブログやニュースサイトなどから抽出した話題のキーワードを自動的に取得して生成するブログ（記事）。

●コピペ型

ニュースサイトの情報や他ブログの記事だけをコピーして作成されたブログ（記事）。

4-1-1 スパムブログの調査概要

今回、スパムブログの状況を把握するために、独自に設定したスパムブログの判別基準を基に、2008年1月現在のアクティブブログについて、次のような目視による判別調査を実施した。

(1) 調査対象ブログ

ブログのコンテンツ量推計調査におけるクローラー調査で取得したアクティブブログのリストを基に、主要20ブログサイトからアクティブブログをそれぞれ800ずつ、それ以外のものから1,100をランダムにサンプリングした。

(2) 調査方法

調査対象のブログのURLへアクセスし、調査時点でブログに掲載されている複数の記事を対象に、目視によりスパムブログか否かの判別を実施した。

(3) スパムブログの判別基準

以下の条件に当てはまるブログをスパムとした。

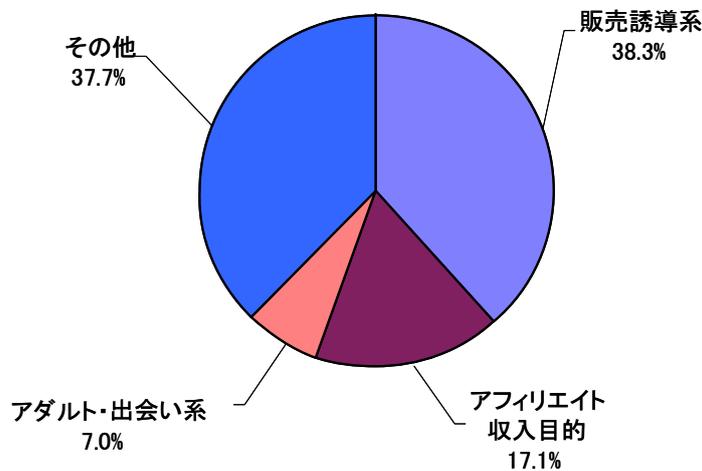
- ・機械的に更新している、又は他のブログの記事を貼り付けることで更新していると見られるもの
- ・すべてが機械的に更新されているとはみられないが、出来事や関心事等の記述がなく、アフィリエイトや広告記事を大量に掲載しているもの
- ・アダルトや出会い系などの特定のサイトに誘導することを目的としているとみられるもの

4-1-2 調査結果

スパムブログ調査の結果では、2008年1月現在、アクティブなブログの12%程度はスパムであることがわかった。なお、スパムブログの出現率について、調査対象の主要20ブログサイト間でみると、0.9%から62.3%の大きな差が見られた。

スパムブログの内容としては、販売誘導(38.3%)、アフィリエイト収入(17.1%)、アダルト・出会い系サイトへの誘導(7.0%)を意図したものが見られる(図表4-1-1)。

図表 4-1-1 スパムブログの目的別分類

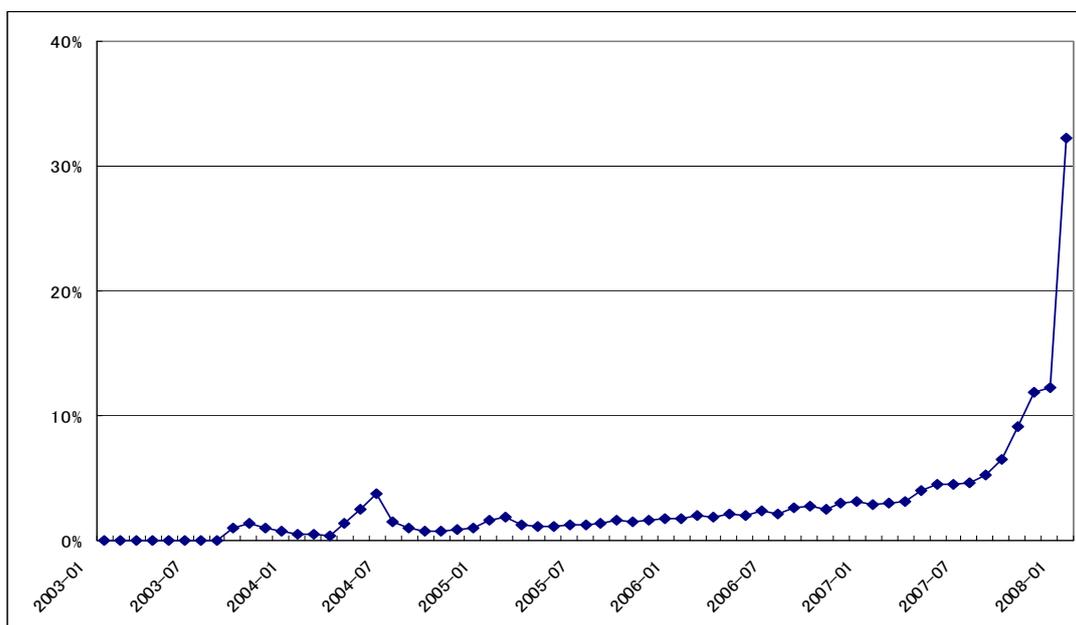


その他: 分類が困難なもの、意図が不明なものが含まれる

また、クローラーにより取得した主要20ブログサイトの記事数データを分析して、2008年1月中に書き込みされた記事全体におけるスパム記事数をみると、その割合は、32.2%となった。スパムブログが記事を頻繁に更新することによって、検索されたり、新着記事情報に登場する機会を多くしようとしていることが伺える(図表4-1-2)。

なお、2008年1月現在のアクティブブログの各月ごとの記事数全体におけるスパム記事の割合は、2008年1月が突出して高く、それ以前は、急激に落ち込んでいる。これは、記事掲載1箇月以内に、スパムブログの多くがブログサービス事業者によって削除され、また新たなスパムブログが毎月、大量に生成されていることを示すものと考えられることができる。

図表 4-1-2 各月ごとの記事数全体におけるスパム記事の割合



スパムブログの問題点としては、まずブログ閲覧者にとって、スパムブログが大量に自動生成され、必要な情報になかなかとり着けない、という点があげられる。スパムブログが頻繁に露出することにより、ブログの情報メディアとしての信頼性を損なうことにもなりかねない。

また、ブログサービス事業者の負担の増加もあげられる。ブログサイトによっては、サービス規約を改正し、巡回してスパムブログを削除したり、ブログアカウント取得時の認証コードを必要とするなどの対策をとっているところも出てきているようである。ただ、上述のように、スパムブログはプログラムにより大量に手間をかけずに作成することができるので、スパムブログのアカウントの利用を停止しても、また別のアカウントが生成され、スパム行為が繰り返されることが多い。最近では、スパムブログを判定するフィルタリングソフトなどが開発されている。

スパムブログによって多額の利益を得ている業者もいるというが、悪質な業者に対しては、適切な対応をとるとともに、迷惑なブログサイトのURLやアフィリエイトを不用意にクリックしないなど、悪質業者がスパムブログから利益を生み出せない土壌を作ることも肝要であるといえよう。

4-2 スパムトラックバック、スパムコメント

ブログのコンテンツの内容に関係なく、プログラムによって機械的に、または無作為に行われる「スパムトラックバック」、「スパムコメント」による迷惑行為も、スパムブログと同様にブログの課題の一つとなっている。

スパムブログによる迷惑の多くは、ブログ閲覧者が検索エンジンやブログポータルサイトから閲覧した際に発生するのに対し、スパムトラックバックおよびスパムコメントは、ブログ開設者への直接的な迷惑行為となる。そのため、利用者数の拡大を図るブログサイトにとっても、以前から大きな問題となっていた。あるブログサイトの管理者によると、2007年2月時点で毎日書き込みされる全5万件のコメントのうち、90%以上がスパムであったという。

なお、今回調査の対象とした主要20ブログサイトについて、提供サービス状況を確認したところ、そのほとんどがスパムコメント、スパムトラックバックについて、以下のような対策をとっている。

- ブログ管理者が承認した後に、トラックバックの結果またはコメントの内容を表示する機能の導入
- URL等の情報によりあらかじめ用意してあるブラックリストの掲載情報に合致したトラックバックを禁止
- IPアドレス、キーワード、文字コード等の条件によるトラックバックやコメントの制限

また、スパムのパターンを特定するフィルタリングを導入しているブログサービス事業者も多く、一定の成果をあげているようである。ブログ開設者の意識調査の結果においても、ブログのコミュニケーション機能が重視される傾向が明らかになっており、ソーシャルメディアとしてのブログ機能が有効に活用されるためにも、機械的に行われる悪質なスパムコメント・スパムトラックバックへの対策の充実が期待される。

4-3 ブログの「炎上」

ブログの利用における問題点の一つとして、ブログの「炎上」と呼ばれる現象が挙げられる。ブログの炎上とは、ブログ記事の内容に対する批判的な意見や、ブログ開設者自身への誹謗中傷が、コメントやトラックバックにより多数寄せられ、ブログが機能不全に陥ることをさす。同様の現象は、ニュースグループにおいて意見の対立が感情的な批判になる「フレーミング」など、インターネットの初期から存在していたが、有名人のブログでもしばしば炎上が発生するため、社会的に認知されている。

今回のアンケートの結果では、ブログ上で生じるトラブルとして、「自分の人格に対する誹謗中傷（いわゆる「炎上」を含む）」を回答した者は、3.4%であった（図表 3-2-33 参照）。

炎上にあった経験のある回答者の割合を性別・年代別に見ると、性別での差異はほとんど見られなかったが（図表 4-3-1）、年代別に見ると、20代以下の若年層ほど経験率が高くなる傾向が見られた（図表 4-3-2）。

なお、ブログ開設者のインタビューにおいて、ブログが炎上した理由について尋ねたところ、その理由は様々であったが、意見の対立する事項について書いた記事が、自分が想定していない読者の目に触れ、批判的なコメントがあった時に、それについて感情的な対応をしたりすることなどが挙げられた。炎上に対する対処としては、①コメントを削除、停止する、②ブログアカウントを削除する、③謝罪する、といった行動がとられることが多いが、コメントの削除や停止をしてもそれを上回るコメントが殺到し、追いつかないということもある。

ブログの炎上自体は頻繁に起こりうるものではないが、ブログを公開しているということは、不特定多数の目に触れるということでもあり、ブログを開設する側にも情報を発信する立場としてのリテラシーが求められると言えよう。

また、警察庁の公表するサイバー犯罪等に関する相談件数をみると、インターネット上の名誉毀損・誹謗中傷等に関する相談件数が、近年、増加傾向が顕著である¹³。このような誹謗中傷でブログが炎上し、悲劇的な結果を招くケースも報道されている。

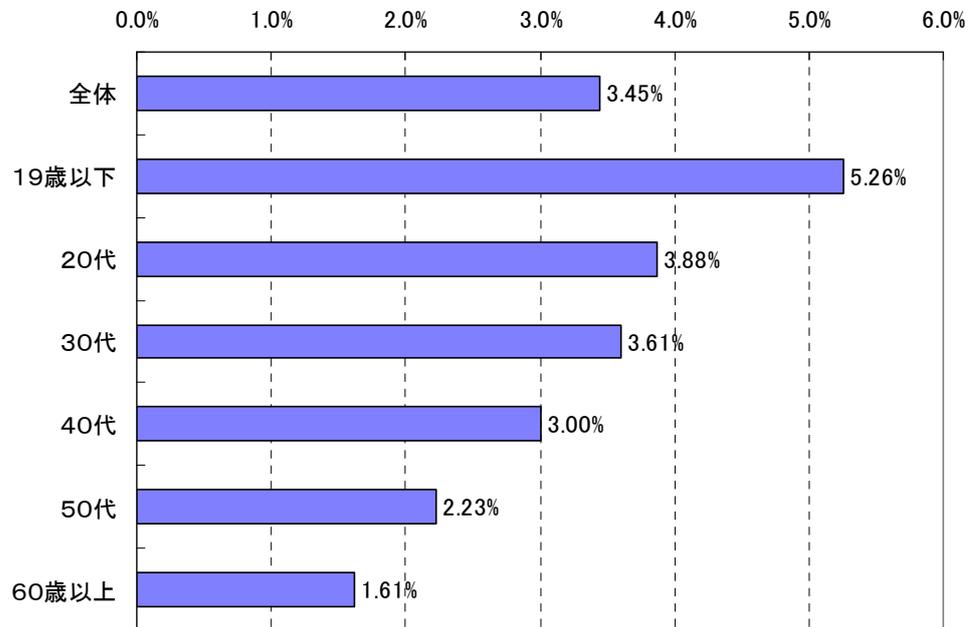
インターネット上におけるコミュニケーションが拡大する中、犯罪行為に対する的確な対応とともに、ネット社会におけるエチケット（ネチケット）の確立が課題と言える。

¹³警察庁ホームページ (<http://www.npa.go.jp/cyber/statics/h20/pdf43.pdf>)

図表 4-3-1 性別による「自分の人格に対する誹謗中傷」の経験率



図表 4-3-2 年代による「自分の人格に対する誹謗中傷」の経験率



4-4 さいごに

CGM（消費者生成型メディア）として会員制サービスのSNSとともに注目されているブログであるが、本調査研究において、国内のブログが2006年にかけて急拡大し、その後も大幅な規模の縮小には至っておらず、アクティブブログ数や新規記事数等はピーク時とほぼ同程度の規模を維持しており、活発な情報発信が続いていることが明らかになった。

また、ブログの利用においても、日記的な自己表現のみならず、コミュニティの形成、社会貢献、収益といった多様な目的で利用され、様々な分野における自身の知識・経験などを生かしたコンテンツが発信されている利用実態が浮かび上がった。

多くの閲覧者を集めるいわゆるアルファブロガーも既に数多く登場し、他メディアとの連携も広がっている。ブログの伝搬力に着目した企業が販促活動に利用するなど、ブログの社会・経済活動への影響力は今後も高まっていくと考えられる。

他方、ブログを悪用して、本来の利用に支障を及ぼすケースも増加してきていることも明らかになっている。このようなブログの負の側面についても引き続き注視していく必要があるだろう。

参考資料

参考1 ブログのコンテンツ量推計の計算式と推計手順

参考2 ブログのコンテンツ量推計におけるデータ補正の考え方

参考3 ブログのコンテンツ量推計のケース別推計結果と検証

参考4 クラスタ化手法の選択

<参考1> ブログのコンテンツ量推計の計算式と推計手順

本研究で実施したブログのコンテンツ量推計の具体的な計算式ならびに計算手順を以下に示す。

① 過去の各月における開設ブログ数の推計

現在観測できる、過去の各月に開設されたアクティブブログ数をもとに、別途設定するブログの1箇月更新継続率を用いて、過去の各月に開設されたブログ総数を推計する。算出した各月開設ブログ数を累計し、各月の累計ブログ総数を推計する。この累計ブログ総数とは、その月までに開設されたブログすべての総数であり、その月におけるアクティブブログ、非アクティブブログ、その月までの削除ブログのすべてが含まれる。推計式は、以下のとおりである。

$$B_i = C_i / L_i^{M_i}$$

$$B_x = \sum_i^x B_i$$

i : ブログが開設された月

x : 推計対象の月

B_x : x 月までに開設された累計ブログ総数

B_i : i 月に開設されたブログ総数

C_i : 現在確認できる i 月に開設されたアクティブブログの数

L_i : i 月開設ブログの1箇月更新継続率

M_i : i 月から 2008 年 1 月までの月数

② 各月におけるアクティブブログ総数の推計

①で推計した各月開設ブログ総数と各月開設ブログの1箇月更新継続率から、各月におけるアクティブブログ総数を推計する。推計式は以下のようになる。

$$A_x = \sum_i^x (B_i \times L_i^{M_x})$$

i : ブログが開設された月

x : 推計対象の月

A_x : x 月におけるアクティブブログ総数

B_i : i 月に開設されたブログ総数

L_i : i 月開設ブログの1箇月更新継続率

M_x : i 月から x 月までの月数

③ 各月の削除ブログ総数、累計削除ブログ総数、非アクティブブログ数の推計

①、②で推計した各月の累計ブログ総数、アクティブブログ総数、および別途設定する一箇月あたりのブログ削除率から、各月の削除ブログ総数、累計削除ブログ総数、非

アクティブブログ総数を推計する。推計式は以下のようになる。

$$D_x = G_x \times Z_x = \left(B_x - \sum_i^{x-1} D_i \right) \times Z_x$$

$$TD_x = \sum_i^x D_i$$

$$U_x = B_x - A_x - TD_x$$

i : ブログが開設された月
 x : 推計対象の月
 D_x : x 月に削除されたブログ数
 D_i : i 月に削除されたブログ数
 G_x : x 月にネット上に存在するブログ数
 Z_x : x 月のブログの一箇月削除率
 B_x : x 月までに開設された累計ブログ総数
 TD_x : x 月までの累計削除ブログ総数
 U_x : x 月の非アクティブブログ総数
 A_x : x 月のアクティブブログ総数

④ 各月新規記事総数、記事総数、記事データ量等の推計

クローラー等によって現在観測できる各月累計ブログ数、各月新規記事数及びデータ量と、②で推計した各月アクティブブログ総数から、実際の各月新規記事総数、現存記事総数、現存記事データ量を推計する。推計式は、以下のようになる。

$$E_x = A_x \times CE_x / C_x$$

$$K_x = \sum_i^x E_i \times G_x / B_x$$

$$V_x = K_x \times EV_x$$

i : ブログが開設された月
 x : 推計対象の月
 E_x : x 月の新規記事総数
 A_x : x 月におけるアクティブブログ総数
 CE_x : 現在確認できるアクティブブログによる x 月での新規記事数
 C_x : 現在確認できる x 月でのアクティブブログ総数
 K_x : x 月のブログの総記事数
 E_i : i 月における新規記事数
 G_x : x 月にネット上に存在するブログ総数
 B_x : x 月までに開設された累計ブログ総数
 V_x : x 月の現存ブログの総データ量
 EV_x : 現在確認できる x 月の記事の平均データ量

⑤ 累計ブログ総数等の推計

以上の結果から、2008年1月現在の累計ブログ総数（削除ブログも含む）、現存ブログ総数、アクティブブログ総数、非アクティブブログ総数、削除ブログ総数を算出する。

推計式は、次のようになる。

$$A = C$$

$$B = \sum_i B_i = \sum_i (C_i / L_i^{M_i})$$

$$TD = \sum_i D_i$$

$$U = B - A - TD$$

$$G = A + U$$

$$K = \sum_i E_i \times G / B$$

$$V = K \times EV$$

i : ブログが開設された月

x : 推計対象の月

A : 2008年1月におけるアクティブブログ総数

C : 現在確認できるアクティブブログ総数

B : 2008年1月までに開設された累計ブログ総数

B_i : i月に開設されたブログ総数

C_i : 現在確認できるi月に開設されたアクティブブログの数

M_i : i月から2008年1月までの月数

L_i : i月開設ブログの一箇月更新継続率

TD : 2008年1月までに削除されたブログ数の累計

D_i : i月に削除されたブログ数

U : 2008年1月における非アクティブブログの総数

G : 2008年1月における現存ブログの総数

K : 2008年1月における現存ブログの記事総数

E_i : 2008年1月までの各月の新規記事数

V : 2008年1月の現存記事の総データ量

EV : 2008年1月における記事の平均データ量

＜参考2＞ブログのコンテンツ量推計におけるデータ補正の考え方

ブログのコンテンツ量推計の基本的な手順は参考1で説明したが、実際の推計過程では、クローラー調査で見つかった想定外の現象の影響や、アンケート回答の偏りによる影響を回避するため、推計データにいくつかの重要な補正を行っている。ここでは、それらの補正の考え方について解説する。

1 調査実施日付に基づくクローラー調査結果データの補正

今回実施したクローラー調査では、ブログ新着記事情報の収集を2008年1月20日から2月18日までの30日間実施した。ブログコンテンツ量推計は、2008年1月を「調査時点」とみなして、2008年1月の新規開設ブログ数を起点に過去の推移を推計しているが、実際には調査期間が上記のように20日間ずれているため、クローラーが捕捉した「1月の新規開設ブログ数」には、「1月1日から19日までに新規開設され、1月20日以降には更新されなかったブログ」が含まれておらず、実際の「1月の新規開設ブログ数」よりも小さい値となる。

推計に当たっては、上記の点を補正するため、「1月1日から19日までに新規開設され、1月20日以降は更新されなかったブログ」の数を、クローラー調査結果データを元に推計し、クローラーが捕捉した1月の新規開設ブログ数に加えてコンテンツ量推計に用いている。

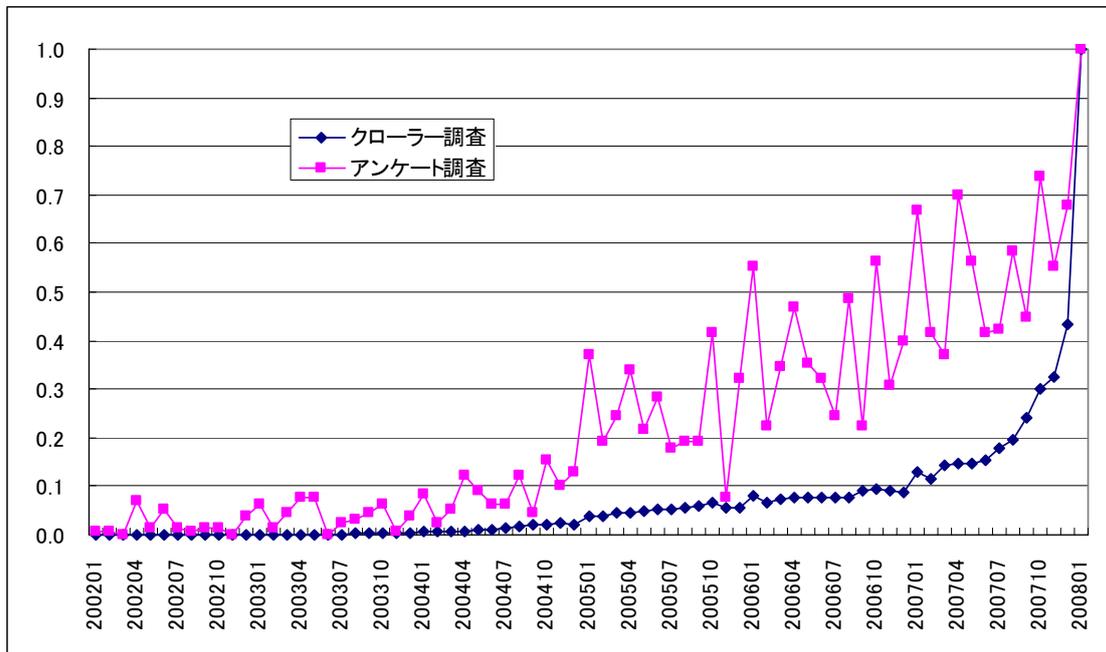
2 アンケート回答の重みづけ補正

今回実施したブログ開設者アンケートは、民間調査パネルを用いたウェブアンケート調査である。第2章に示したように、回答数は回答者の年齢・性別による割付を行ったが、この他に、ブログの開設時期・更新終了時期による回答率の違いが生じていると考えられる。この回答の偏りは、推計に用いるブログの更新継続率の算出結果に影響し、結果としてブログのコンテンツ量推計結果に大きな影響を与えるため、客観性の高いクローラー調査のデータを用いてアンケート回答の偏りを補正するウェイトづけを行った。

補正の考え方は、次のとおりである。

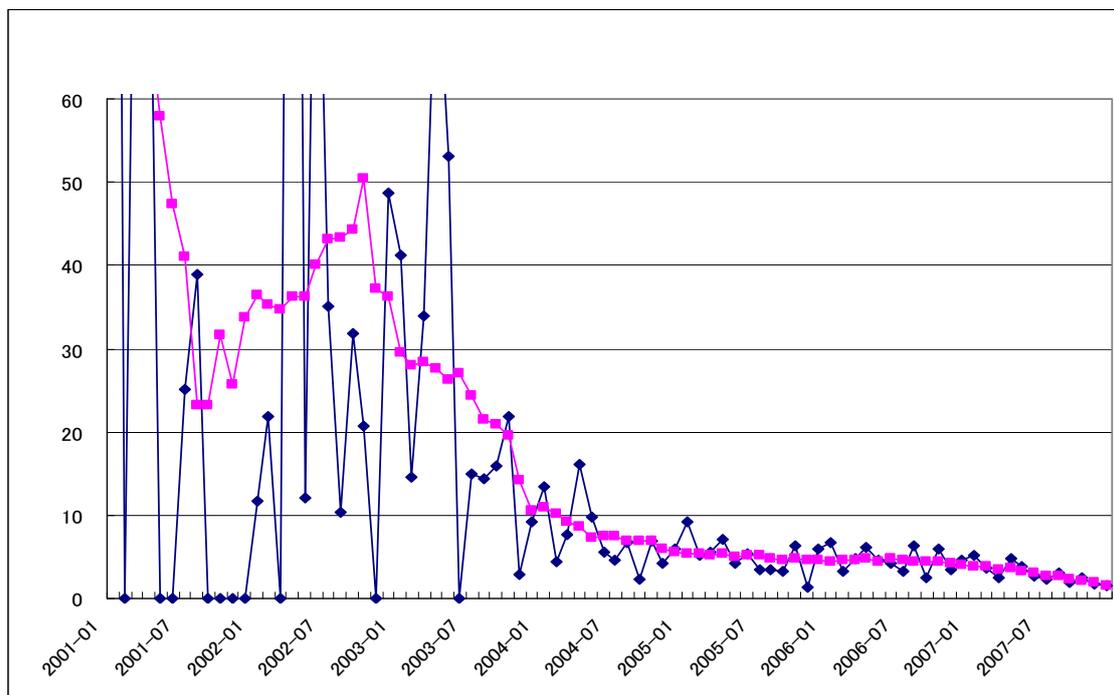
- ① アンケート回答データから集計した、2008年1月現在のアクティブブログの開設月分布と、クローラー調査で補足したアクティブブログの開設月分布（前項の補正を行ったもの）を比較すると、図表参2-1のように大きな違いがある（それぞれ2008年1月現在のアクティブブログの開設数を1として比較している）。クローラー調査結果は、いわば現実のブログの実測データであり、両者の違いは、アンケート回答データの偏りによるものと考えられる。

図表参2-1 クローラー調査とアンケート調査のアクティブブログ開設時期分布



② そこで、図表参2-1の両データの比を、アンケート回答データの補正ウェイトとして用いる。すなわち、アクティブブログの開設月について、アンケート回答データの分布が、クローラー調査結果データの分布に一致するように、アンケート回答データに重みづけを行う。この考え方で算出したブログ開設月別の補正ウェイト値を図表参2-2に示す。補正ウェイト値は、ブログ開設時期が古い回答者ほど大きい値となるが、これは、ブログ開設時期が古い（ブログ継続期間が長い）アクティブブロガーほど、アンケートに回答しやすい傾向があることを示している。ただし、既にブログの更新をなくなっている非アクティブブロガーでは、ブログ開設時期とブログ更新継続期間の長さとは対応しないため、このような回答傾向があるとは考えにくい。そのため、図表参2-2の補正ウェイトによる重みづけは、現在もブログの更新を続けている回答者についてのみ行うこととした。

図表参 2-2 アンケート回答補正ウェイトの算出結果



※横軸は、回答されたブログの開設年月を表す。

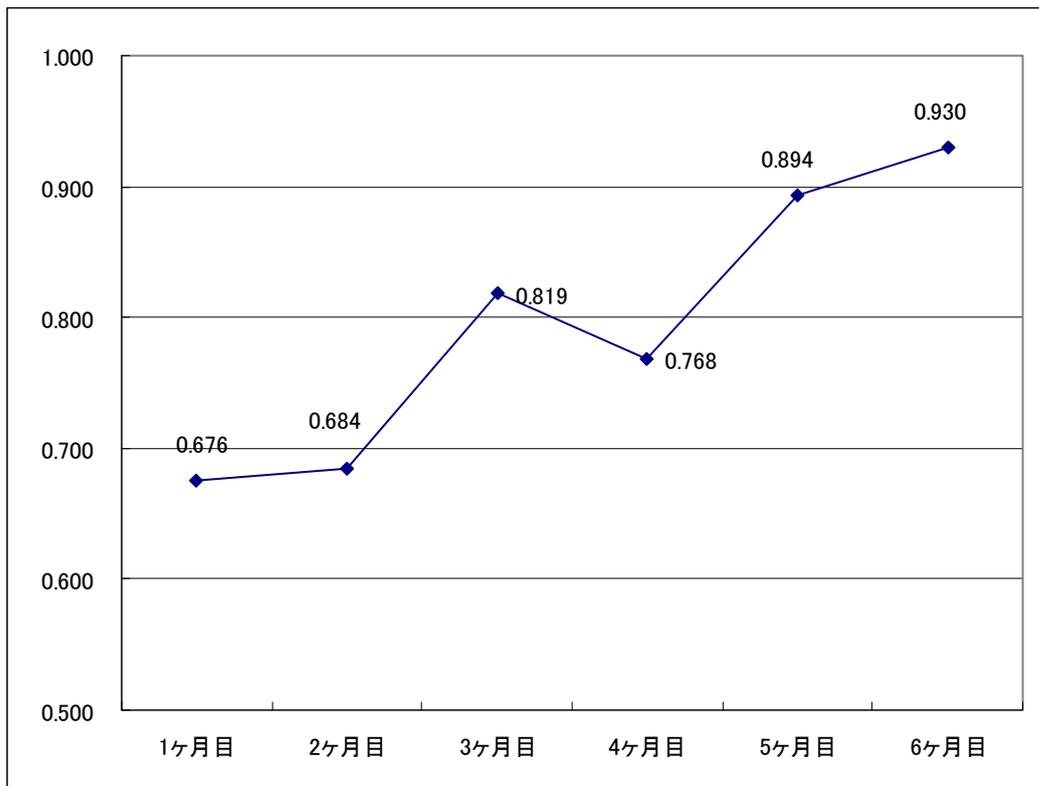
3 ブログ開設直後の短期更新継続率の設定

図表参 2-1 のクローラー調査結果を見ると、最近 6 ヶ月間については、見かけ上、新規開設ブログ数が急激に増加しているように見える。これは、実際に新規開設ブログ数が急増しているのではなく、ブログ開設直後には特に更新継続率が低く、多くのブログが短期間で休眠したり、削除されていることを示すものと考えられる。

コンテンツ量推計の中で、ブログの更新継続率はアンケート回答データから算出したものを用いているが、ブログ開設直後の 6 ヶ月間については、上記のクローラー調査結果から算出した特別な更新継続率（短期更新継続率）を用いて計算を行うこととした。

図表参 2-3 に、今回のコンテンツ量推計で算出した短期更新継続率を示す。これによると、更新継続率は、ブログ開設後 1 ヶ月目が最も低く、新規開設ブログの約 1 / 3 は、翌月には更新されなくなっている。開設 2 ヶ月目以降、更新継続率は次第に上昇するが、6 ヶ月後にも更新されているブログは、新規開設ブログ全体の 24.3%にとどまる。

図表参 2-3 ブログ開設直後の1ヶ月更新継続率の推移



＜参考3＞ブログのコンテンツ量推計のケース別推計結果と検証

推計に用いるブログの更新継続率、ブログの削除率、主要20ブログサイトの利用割合については、前提となる仮説や推定値の設定によりいくつかのケースが想定される。ここでは、各ケースの場合における推計値と、それぞれのケースでのコンテンツ量推計結果を示し、各ケースの推計結果のうち、どれが現実に即したものであるかを、本研究以外の調査データ等を参照して検証する。

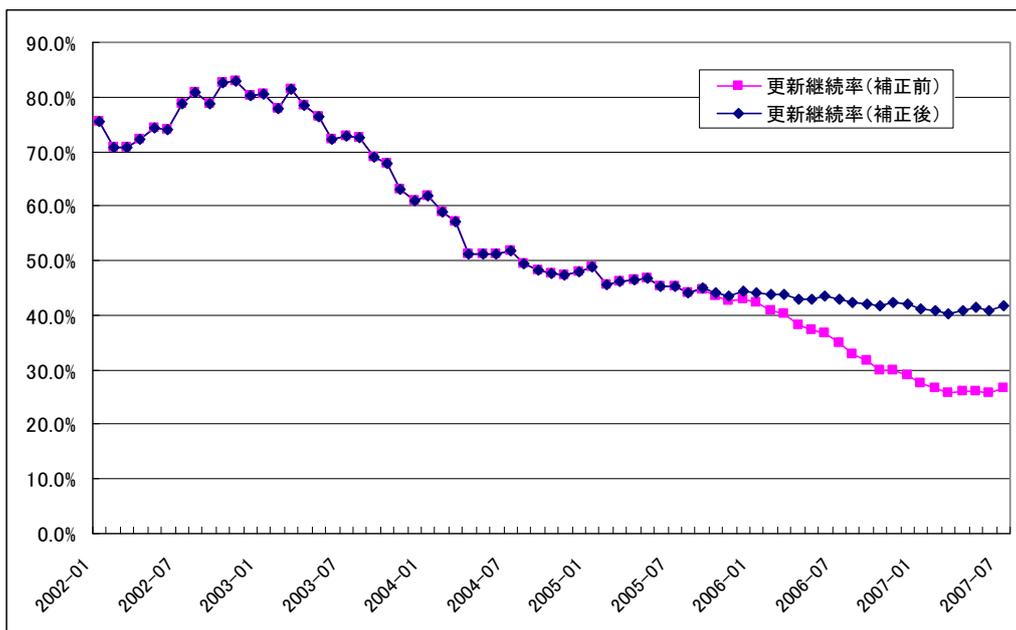
1 推定値によるケース設定

① ブログの更新継続率

ブログの更新継続率は、ブログの導入期から成長期にかけて、利用者層がイノベータ層からアーリーマジョリティ層に広がる過程では、更新継続率や削除率は大幅に変化して不安定であると考えられるが、ブログが成長期から安定期に移行したと考えられる2006年以降は、利用者層はマジョリティ層が中心となり、更新継続率や削除率は概ね一定で推移していると考えられる。しかし、アンケートの回答から算出した時期別の更新継続率は、図表参3-1に示すとおり、2006年以降急速に低下していた。

これは、アンケート回答者、特にすでにブログ更新をやめてしまった非アクティブなブログ開設者の回答の偏りの補正ができていないためと考えられるが、非アクティブブログはクローラー調査の対象外であるため、補正の手がかりとなるデータが存在しない。そこで、2006年以降については更新継続率は原則一定と仮定してアンケート回答のウェイト補正を行い、更新継続率を算出し直した。

図表参3-1 2つのケースにおける更新継続率の推移



上記の仮定が妥当か否かは、推計結果の妥当性を評価し、検証する必要がある。

そこで、ここでは、ブログの更新継続率について、

- ・ 2006年以降、更新継続率は一定とする補正を行った場合（高継続率ケース）
- ・ 更新継続率の補正を行わない場合（低継続率ケース）

両方の値を用いて推計し、推計結果の比較検証を行う。

②ブログの削除率

ブログ削除率については、アンケート回答から直接集計した値を推計に用いているが、更新継続率の推計の際に行ったアクティブブログに関するウェイトづけを行って算出したブログ削除率の推移が図表参3-2の「補正あり」である。

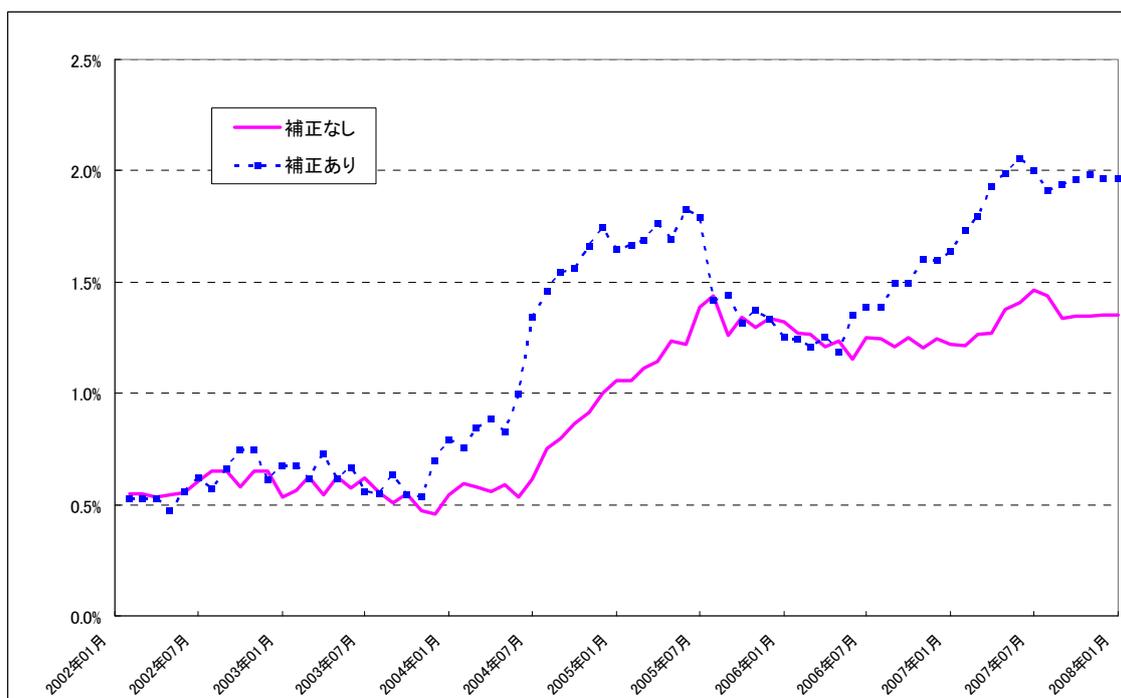
「補正なし」の場合は2006年以降、削除率に大きな変化が見られないのに対し、「補正あり」の算出結果では、2006年以降にブログ削除率が急速に高まっており、上記仮説に反する結果となっている。どちらが妥当な削除率であるかは、推計結果の妥当性を評価し検証する必要がある。

そこで、ここではブログの削除率について

- ・ 更新継続率と同様の補正をかけて算出した場合（高削除率）
- ・ アンケート回答に補正をかけず算出した場合（低削除率）

の2ケースについてそれぞれ推計し、推計結果の比較検証を行う。

図表参3-2 アンケート回答補正によるブログ削除率の差異



注：アンケート回答から算出したブログ削除率の13か月移動平均値を示した。

③主要 20 ブログサイトの割合

第 2 章で述べたとおり、主要 20 ブログサイトの利用割合（シェア）については、ping サーバで把握したアクティブブログでの割合（77.5%）が妥当と考えられるが、アンケート調査では 85.8%という値が得られており、また、他のアンケート調査例でも 8 割前後の値であることが多い。このため、ここでは

- ・ 主要ブログサイトのシェアを 77.5%とする。（低割合：高位推計）
- ・ 主要ブログサイトのシェアを 85.8%とする。（高割合：低位推計）

の 2 ケースについてそれぞれ推計し、結果を示すこととする。

これらの条件の組み合わせは全部で 8 通りとなる。ここでは、以下の 8 通りのケースについてコンテンツ量の推計値を算出し、推計結果の妥当性について評価検証する。

- ・ 高継続率、高削除率、高位推計 （高高高ケース）
- ・ 高継続率、高削除率、低位推計 （高高低ケース）
- ・ 高継続率、低削除率、高位推計 （高低高ケース）
- ・ 高継続率、低削除率、低位推計 （高低低ケース）
- ・ 低継続率、高削除率、高位推計 （低高高ケース）
- ・ 低継続率、高削除率、低位推計 （低高低ケース）
- ・ 低継続率、低削除率、高位推計 （低低高ケース）
- ・ 低継続率、低削除率、低位推計 （低低低ケース）

2 ケース別の国内のブログのコンテンツ量推計結果（2008年1月現在）

（1）ブログ数

図表参3-3に、8ケースそれぞれでの2008年1月のブログ数の推計結果を示す。各ケースのうち、現存ブログ数の推計値が最も大きくなるのは「低継続率・低削除率・高位推計」ケースで、2181万ブログとなる。これは主に、非アクティブブログ数が大きく推計されるためである。一方、最も小さい推計結果となるのは「高継続率・高削除率・低位推計」ケースで、1404万ブログとなる。

図表参3-3 2008年1月現在の国内ブログ数の推計結果

	高継続率				低継続率			
	高削除率		低削除率		高削除率		低削除率	
	高位推計	低位推計	高位推計	低位推計	高位推計	低位推計	高位推計	低位推計
インターネット上に公開されている ブログ総数	1,555	1,404	1,690	1,526	2,005	1,810	2,181	1,969
うちアクティブブログ数	308	278	308	278	307	278	307	278
うち非アクティブブログ数	1,247	1,126	1,383	1,249	1,697	1,533	1,874	1,692
これまでに削除されたブログ総数	689	622	553	499	882	796	706	637
これまでに開設されたブログ総数 (削除済みを含む)	2,243	2,026	2,243	2,026	2,887	2,607	2,887	2,607

（単位：万ブログ）

なお、現存ブログだけでなく、既に削除済みのものを含めた累計では、低継続率・高位推計の組み合わせが最も大きく、2887万ブログと推計される。全体として、更新継続率を低く設定したケースの方が、ブログ数は大きい値となる。

（2）ブログの記事数・データ量

図表参3-4に、2008年1月にインターネット上で公開されている国内ブログの記事数、データ量、HTMLファイルデータ量の推計結果を示す。

やはり、推計結果が最も大きいのは「低継続率・低削除率・高位推計」ケースで、記事数15億3000万件、データ量は48テラバイトとなる。一方、推計結果が最も小さいのは「高継続率・高削除率・低位推計」ケースで、記事数11億1800万件、総データ量36テラバイトとなる。

図表参 3-4 2008 年 1 月現在のブログ記事数の推計結果

	高継続率				低継続率			
	高削除率		低削除率		高削除率		低削除率	
	高位推計	低位推計	高位推計	低位推計	高位推計	低位推計	高位推計	低位推計
記事件数 [100万件]	1,238	1,118	1,347	1,216	1,408	1,271	1,532	1,383
総データ量 [テラバイト]	40	36	42	38	45	41	48	44
HTMLデータ量 [テラバイト]	11	10	12	11	12	11	13	12

(インターネット上に現存するブログの記事)

上記は、インターネット上に存在し公開されているブログ記事の総数だが、これまでに開設されたブログのうち、インターネット上から既に削除されたブログも加えた総量推計結果を図表参 3-5 に示す。ここでは、削除率の高低による推計結果の差はなく、「低継続率・高位推計」ケースが最も大きい推計結果、「高継続率・低位推計」ケースが最も小さい推計結果となる。

図表参 3-5 2008 年 1 月までに公開された全ブログ記事数の推計結果

	高継続率				低継続率			
	高削除率		低削除率		高削除率		低削除率	
	高位推計	低位推計	高位推計	低位推計	高位推計	低位推計	高位推計	低位推計
記事件数 [100万件]	1,787	1,614	1,787	1,614	2,027	1,831	2,027	1,831
総データ量 [テラバイト]	54	49	54	49	61	55	61	55
HTMLデータ量 [テラバイト]	16	14	16	14	18	16	18	16

(既に削除された記事も含む)

3 コンテンツ量の推移の推計結果

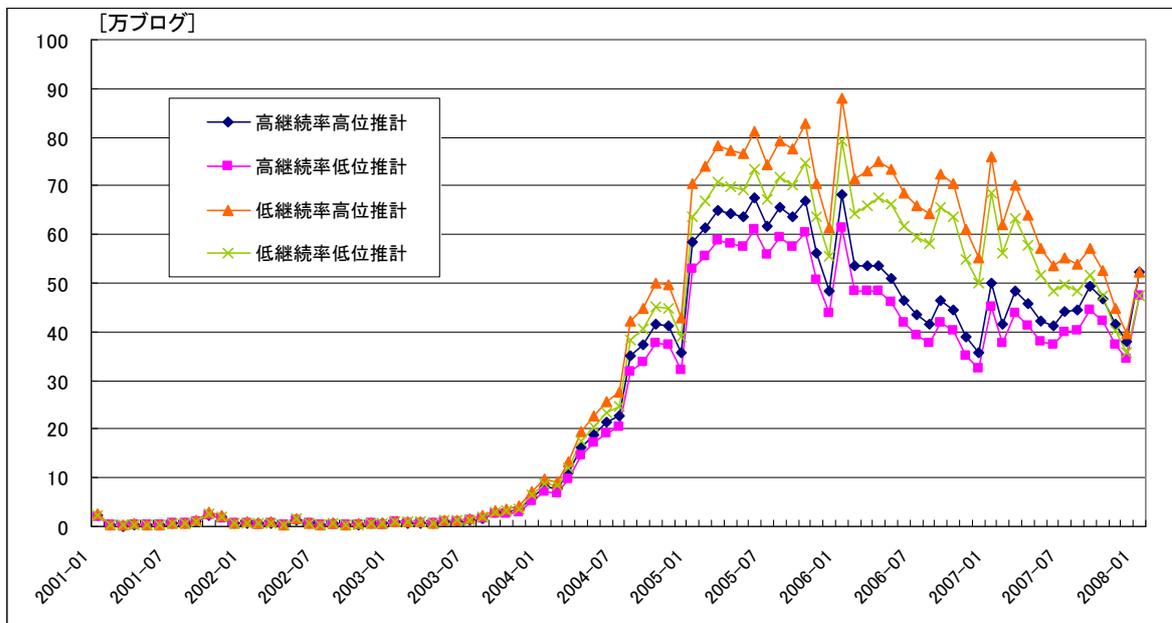
次に、各ケースの設定値を用いて推計した、過去のブログのコンテンツ量の推移を示す。

(1) 毎月の新規開設ブログ数の推移

図表参 3-6 に、2001 年 1 月以降の毎月の新規開設ブログ数の推計結果を示す。

8つのケースいずれの推計結果でも、2001 年から 2005 年までの新規開設ブログ数は類似した推移を示している。しかし、2006 年以降については「高継続率ケース」と「低継続率ケース」で推移が大きく異なる。

図表参 3-6 毎月の新規開設ブログ数の推移の推計



注：新規開設ブログ数には削除率は関係しないため、4系列のみ表示した。

「低継続率ケース」では、2006 年 1 月以降、季節変動を除くとほぼ一貫して毎月の新規開設ブログ数が減少しているのに対し、「高継続率ケース」では 2006 年末に新規開設ブログ数が底を打ち、2007 年には横ばいから徐々に回復に転じているという推計結果になっている。

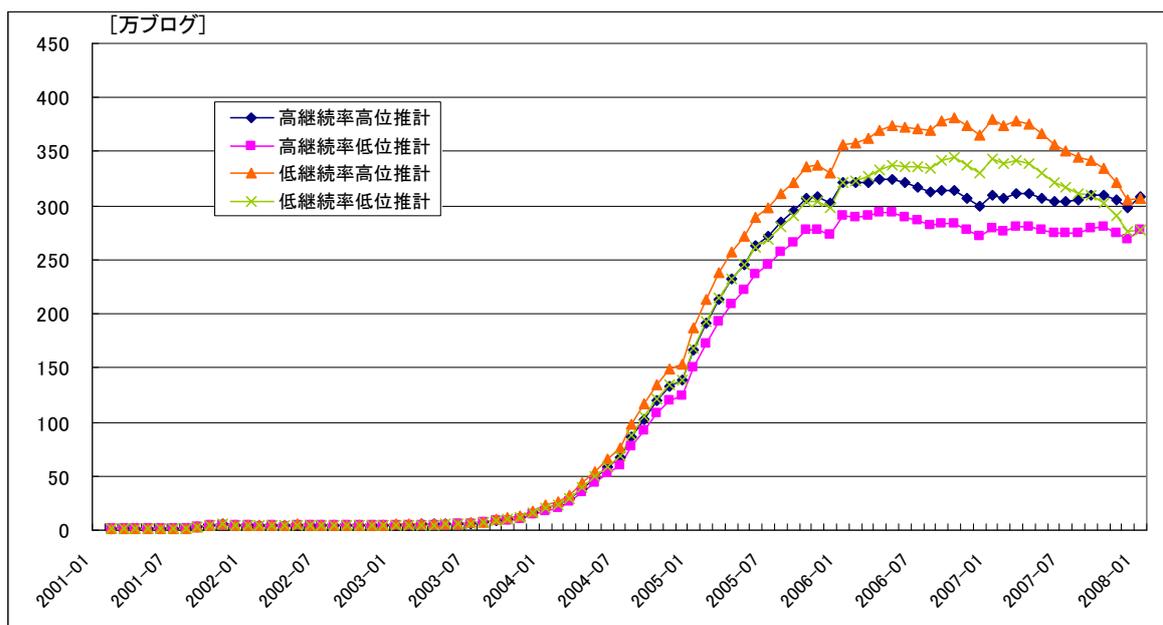
(2) アクティブブログ数の推移

図表参 3-7 に、2001 年以降の毎月のアクティブブログ数に関する推計結果を示す。

新規開設ブログ数と同様、アクティブブログ数も 2003 年後半から増加が加速し、2004 年から 2005 年にかけて急増したという基本的な推移は各ケース共通である。

2006年以降の推移については、ケースによって推計結果に違いがある。「高継続率ケース」では、2006年以降のアクティブブログ総数は300万ブログ前後でほぼ横ばいで推移しているのに対し、「低継続率ケース」では2007年前半に340万～380万ブログに達した後、アクティブブログ数が減少に転じるという推計結果になっている。

図表参3-7 各月におけるアクティブブログ数の推移の推計



注：アクティブブログ数には削除率は関係しないため、4系列のみ表示した。

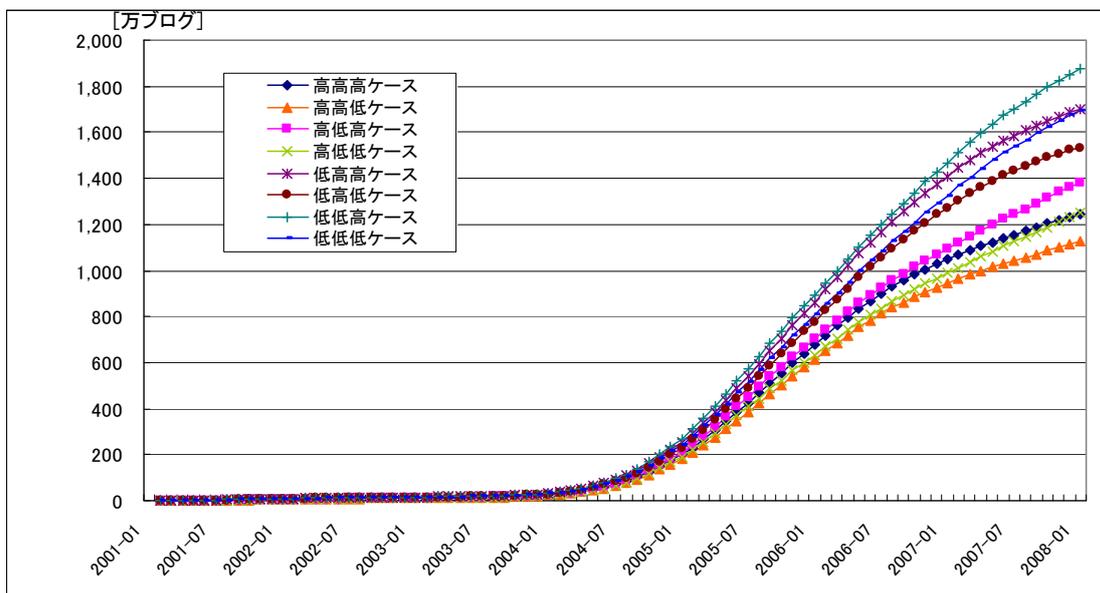
(3) 非アクティブブログ数、累計開設ブログ数の推移

図表参3-8, 3-9に、2001年以降の非アクティブブログ数および累計削除ブログ数の推移を示す。

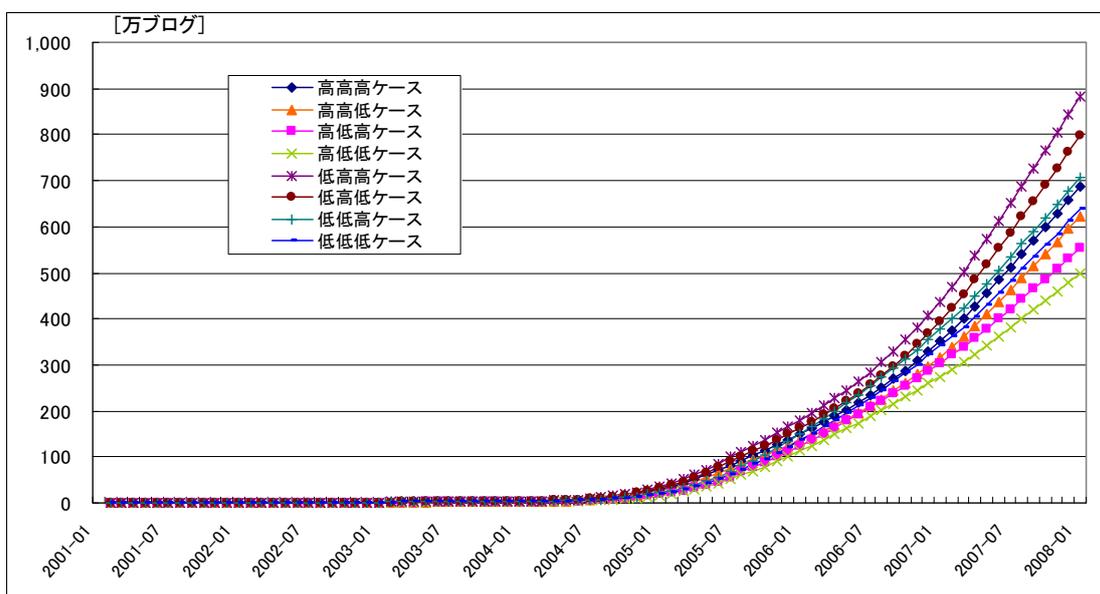
ブログ全体の規模拡大と削除率の上昇によって、毎月の削除ブログ数は増加傾向にあるが、特に、最近のブログ削除率を高く想定している「低継続率ケース」では、2007年に入って削除ブログ数の増加を一段と多く推計している。

一方、非アクティブブログ数も一貫して増加しているが、その増加ペースは2006年以降、やや鈍化している。特に、「低継続率ケース」では、2007年に入ってから非アクティブブログの増加にブレーキがかかった形となっている。

図表参3-8 非アクティブブログ数の推移の推計



図表参3-9 累計削除ブログ数の推移の推計

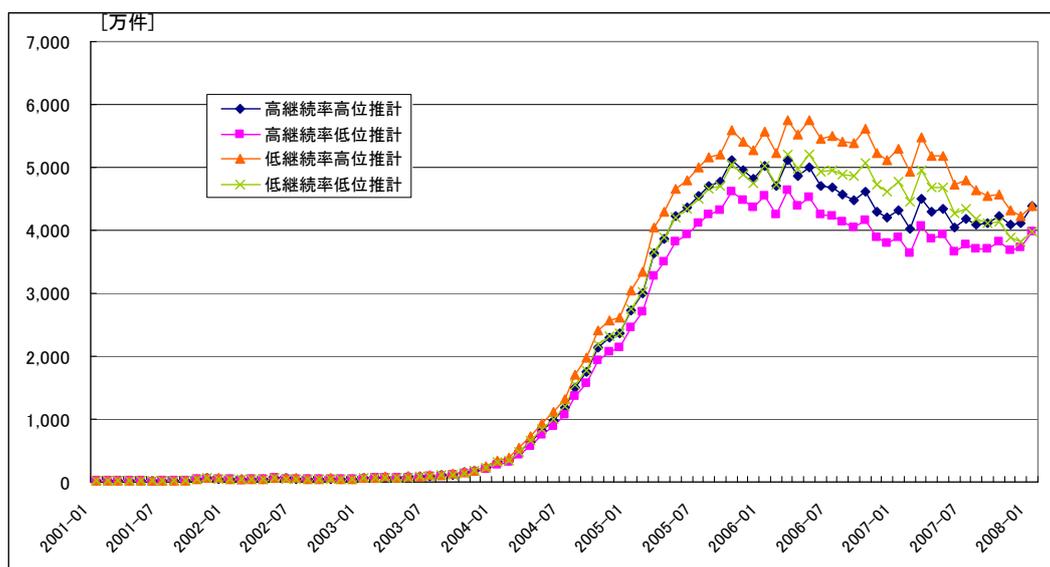


(4) 毎月の新規新規記事数の推移

図表参3-10に、2001年以降の毎月の新規記事数の推計結果を示す。

ケース別に推移を比較すると、「高継続率ケース」では2006年中は次第に新規記事数が減少したものの、2007年には下げ止まっているのに対し、「低継続率ケース」ではむしろ2007年以降、新規記事数の減少が顕著になっている。

図表参3-10 毎月の新規記事数の推移

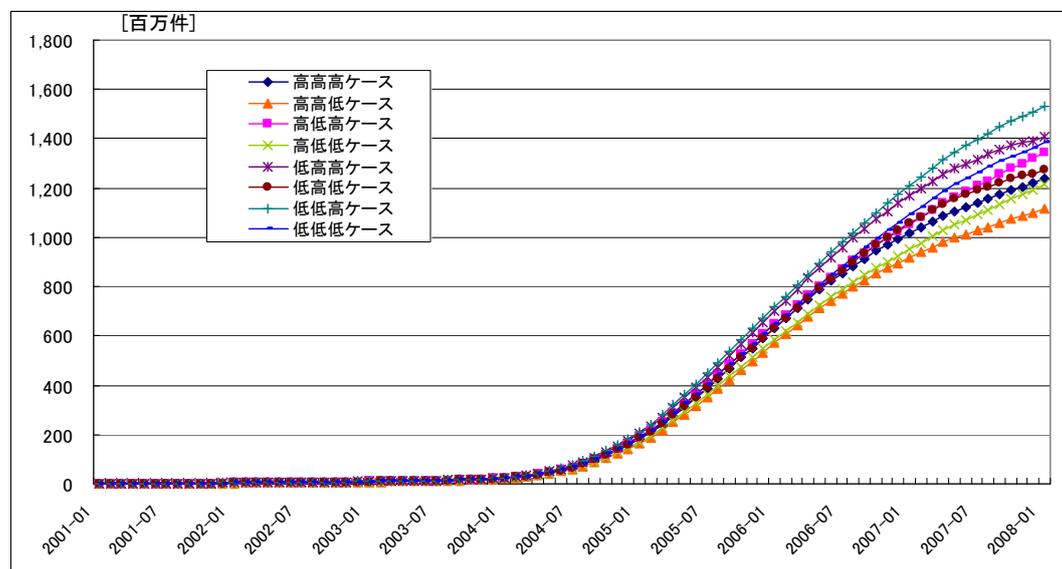


注：新規記事数には削除率は関係しないため、4系列のみ表示した。

図表参3-11に、2001年以降の各月においてインターネット上で公開されていた現存記事数の推計結果を示す。

現存記事数は2004年初頭から急速に増加し、2005年～2006年のブログブーム期に増加のピークとなり、2007年以降伸びが鈍化したという基本的な推移は各ケース共通である。2008年1月現在の推計値が最も多い「低継続率・低削除率・高位推計」と最も少ない「高継続率・高削除率・低位推計」では、推計値には37%の差がある。

図表参3-11 毎月の現存記事数の推移

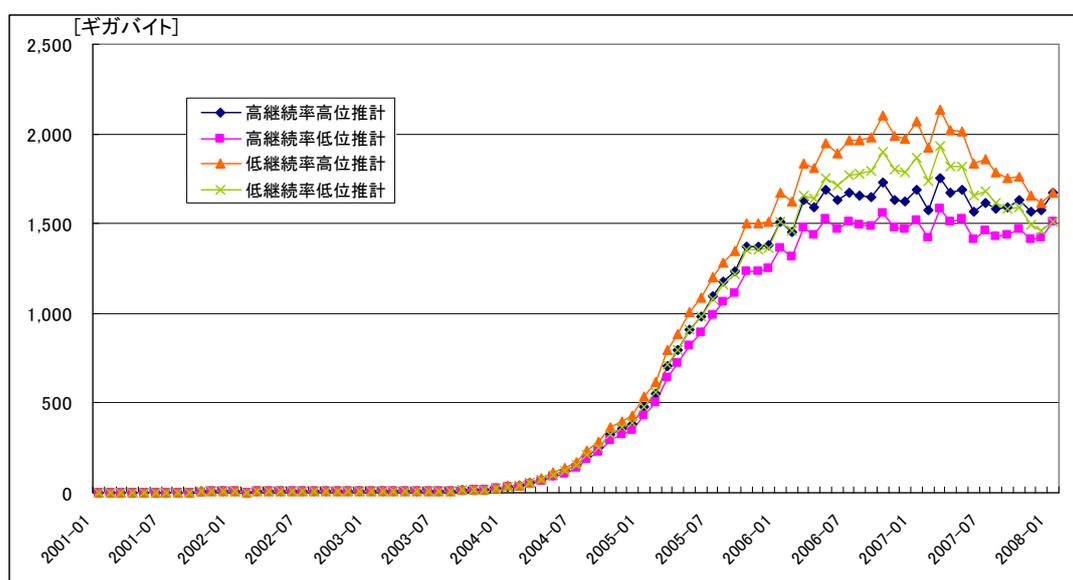


(5) 毎月の新規記事データ量の推移

図表参3-12に、2001年以降の各月における新規記事のデータ量の推計結果を示す。

各ケースの推計結果の特徴は、新規記事数の推移の推計結果と似ており、「高継続率ケース」では2006年、2007年を通じてほぼ横ばいで推移しているのに対し、「低継続率ケース」では2007年1月をピークに減少に転じたという推計結果になっている。

図表参3-12 毎月の新規記事データ量の推移

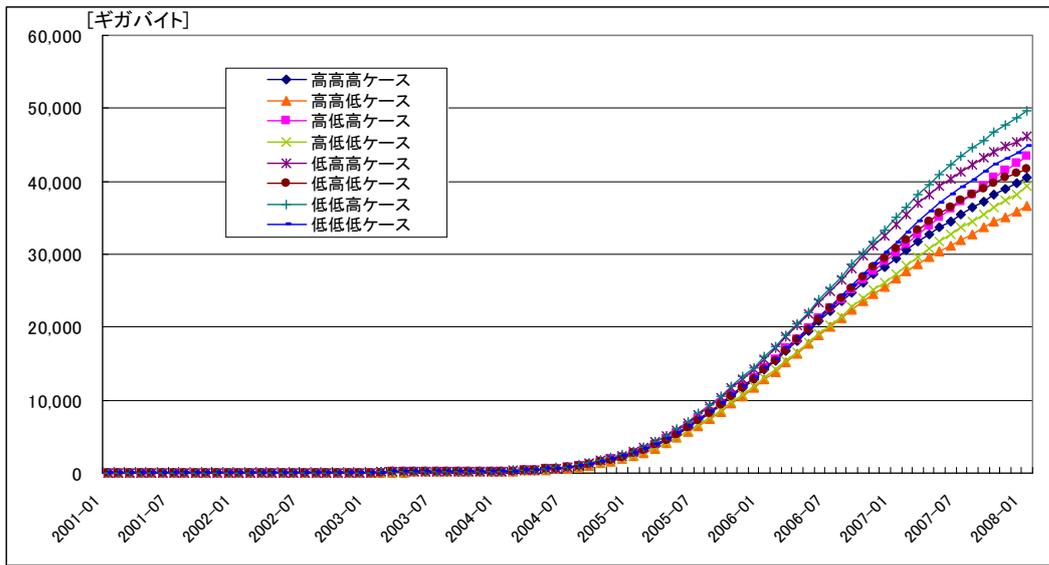


注：新規記事データ量には削除率は関係しないため、4系列のみ表示した。

図表参3-13は、各月においてインターネット上に公開されていた現存記事のデータ量の推計結果である。

いずれのケースでも、現存記事のデータ量は伸び続けているが、「低継続率ケース」では最近になって現存記事データ総量の伸びが鈍化している。一方、「高継続率ケース」ではそれほど顕著な鈍化傾向は見られず、データ総量の伸びが続いているという推計結果になっている。

図表参 3-13 毎月の現存記事データ量の推移



4 推計結果の妥当性の検討

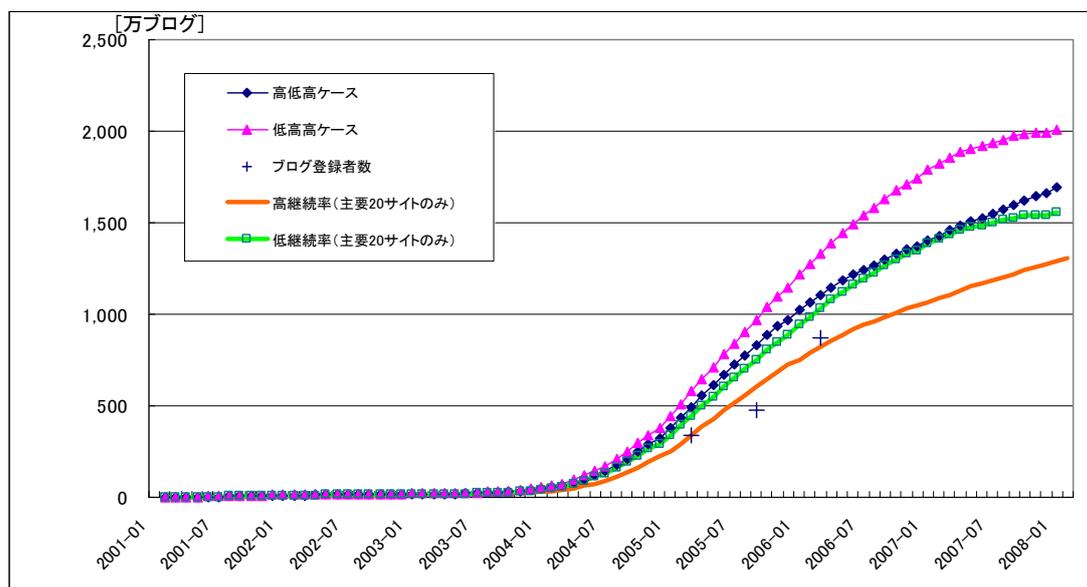
ここでは、本調査以外の情報源から得られるブログの規模に関するデータと今回の推計結果を比較照合し、推計結果の妥当性について検討する。

(1) 過去の総務省公表データとの整合性

総務省では平成 17 年から平成 18 年にかけて、国内のブログ登録者数を 3 回にわたって集計し、発表している。それによると、平成 17 年 3 月から平成 18 年 3 月までの 1 年間に、国内のブログ登録者数は 335 万人から 868 万人へと、1 年間で 2 倍以上の急激な伸びを見せている。

総務省が発表した「ブログ登録者数」は、必ずしもブログ開設者の人数そのものではなく、ブログサイトに登録した ID の総数であり、本調査で言う「現存ブログ数」に近いと考えられる。そこで発表されたブログ登録者数の値と今回の現存ブログ数の推計結果と重ね合わせてみると、いずれのケースでも、推計結果が総務省発表のブログ登録者数を上回る値となっている。

図表参 3-14 現存ブログ数の推計結果と「ブログ登録者数」との比較



これは、総務省が発表したブログ登録者数が商用ブログサービスへの登録者数であり、自主サーバ等によるブログ開設者を含んでいないことや、ブログ登録者の集計が悉皆調査ではなかった（当時の全てのブログ事業者を対象にはしていなかった）ことによると思われる¹⁴。

今回の調査で、クローラー調査の対象とした主要 20 ブログサイトのみの現存ブログ

¹⁴ 総務省調査の集計対象事業者数は、第 1 回が 20 社強、第 2 回が 33 社、第 3 回が 53 社となっている。

数の推計結果と総務省の公表データを比較すると、「高継続率ケース」での推計結果が、ほぼ総務省が公表したブログ登録者数と重なる。この点を考慮すると、本調査における推計結果、特に「高継続率ケース」での推計結果は、総務省の公表データと矛盾せず、十分な整合性を持っていると評価できる。

(2) 最近のアクティブブログ数の推移から見た推計条件の妥当性

設定した 8 ケースのうち、ブログの更新継続率の想定が異なる「低継続率ケース」と「高継続率ケース」では、2007 年以降のコンテンツ量の推移に大きな違いが生じた。

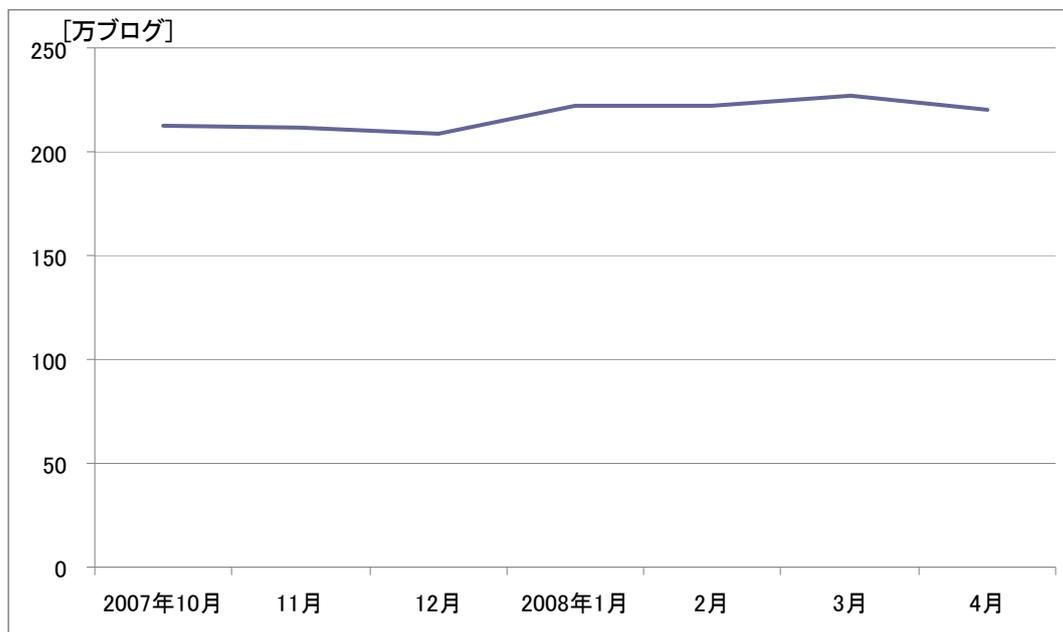
- ① 「高継続率ケース」では 2006 年 1 月以降、いったん新規開設ブログ数が減少した後、横ばいから微増で推移するのに対し、「低継続率ケース」では新規開設ブログ数が一貫して減少を続ける。(図表参 3-6 参照)

- ② 「高継続率ケース」ではアクティブブログ数や新規記事数が 2006 年以降ほぼ横ばいで推移するのに対し、「低継続率ケース」ではこれらが 2007 年以降に大きく減少する。(図表参 3-7, 3-12 参照)

日本国内で、アクティブブログ数を独自に調査・集計している例としては、第 1 章でも述べた「ブログファン」(<http://www.blogfan.org/>)がある。ブログファンでは、国内の主要 21 ブログサイトについて、毎月のアクティブブログ数(更新があったブログ数)を独自に調査し発表している。このデータは悉皆調査でないため、本調査の推計結果と直接比較はできないが、ユーザ数の多い主要ブログサイトを対象としているため、国内のアクティブブログ数の推移をおおむね反映していると考えられる。

ブログファンが公表している最近のアクティブブログ数の推移を見ると、全体としてはほぼ横ばいで推移している。これが正しいとすれば、本調査の推計結果のうち、高継続率ケースがより現実のブログ総数の推移をよく表していると考えられる。

図表参 3-15 主要 21 ブログサイトの月間アクティブブログ数の推移



出所：ブログファン(<http://www.blogfan.org/service/>)

(3) 妥当性検討結果のまとめ

ブログの更新継続率、主要 20 ブログサイトのシェアについて外部のデータを参照して検討した結果、更新継続率については「高継続率ケース」がより現実に近いと考えられる。また、主要 20 ブログサイトのシェアについては、第 2 章でも検討したとおり、個人ブログのみを対象とする場合には、アンケート調査結果がほぼ正しいと考えられるが、企業ブログ等を含めると、ping サーバで実測した値に近付くと推定される。

ブログの削除率については、適切な調査事例が見当たらず、比較検証ができないが、「高継続率ケース」を採用するとすれば、最近のブログ継続率が高い想定となるため、最近のブログ削除率はより低い値をとる想定が妥当と考えられる。

これらを総合すると、推計を行った 8 ケースのうち、「高継続率・低削除率・高位推計」が現実の国内ブログの実態に最も近い推計結果と考えられ、報告書本文ではこのケースの推計結果を掲載した。ただし、継続率以外の要素については不確実性が強く、企業ブログやブログ削除率に関する客観的なデータによる検証が望まれる。

＜参考4＞クラスタ化手法の選択

算出された因子得点を元に、回答者のクラスタ分析を行い、プログユーザのグループを抽出した。クラスタ分析に当たりクラスタ化の手法による偏りに注意するため、Ward法、最遠隣法、K平均法の3手法でクラスタ分析を実施し、結果の比較を行った。クラスタ数については、因子数が7つであることから、4～7と仮定し、各場合についてクラスタ分析を実施した（図表参4-1）。

図表参4-1 クラスタ化の手法別分析結果（左からK平均法、Ward法、最遠隣法）

KMeans_4				Ward_4				最遠隣法_4						
	観測度数	N期待度数	N	残差		観測度数	N期待度数	N	残差		観測度数	N期待度数	N	残差
1	610	587.8		22.3	1	237	587.8		-350.8	1	1136	587.8		548.3
2	669	587.8		81.3	2	1330	587.8		742.3	2	667	587.8		79.3
3	635	587.8		47.3	3	587	587.8		-.8	3	534	587.8		-53.8
4	437	587.8		-150.8	4	197	587.8		-390.8	4	14	587.8		-573.8
合計	2351				合計	2351				合計	2351			

KMeans_5				Ward_5				最遠隣法_5						
	観測度数	N期待度数	N	残差		観測度数	N期待度数	N	残差		観測度数	N期待度数	N	残差
1	736	470.2		265.8	1	237	470.2		-233.2	1	1136	470.2		665.8
2	343	470.2		-127.2	2	726	470.2		255.8	2	667	470.2		196.8
3	534	470.2		63.8	3	604	470.2		133.8	3	285	470.2		-185.2
4	355	470.2		-115.2	4	587	470.2		116.8	4	249	470.2		-221.2
5	383	470.2		-87.2	5	197	470.2		-273.2	5	14	470.2		-456.2
合計	2351				合計	2351				合計	2351			

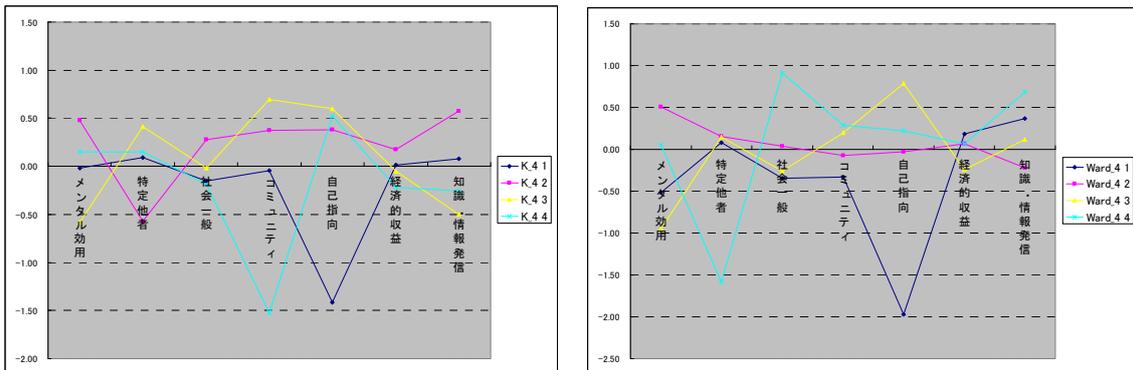
KMeans_6				Ward_6				最遠隣法_6						
	観測度数	N期待度数	N	残差		観測度数	N期待度数	N	残差		観測度数	N期待度数	N	残差
1	338	391.8		-53.8	1	237	391.8		-154.8	1	1136	391.8		744.2
2	271	391.8		-120.8	2	726	391.8		334.2	2	570	391.8		178.2
3	429	391.8		37.2	3	604	391.8		212.2	3	285	391.8		-106.8
4	294	391.8		-97.8	4	430	391.8		38.2	4	249	391.8		-142.8
5	268	391.8		-123.8	5	197	391.8		-194.8	5	97	391.8		-294.8
6	751	391.8		359.2	6	157	391.8		-234.8	6	14	391.8		-377.8
合計	2351				合計	2351				合計	2351			

KMeans_7				Ward_7				最遠隣法_7						
	観測度数	N期待度数	N	残差		観測度数	N期待度数	N	残差		観測度数	N期待度数	N	残差
1	220	335.9		-115.9	1	237	335.9		-98.9	1	1136	335.9		800.1
2	240	335.9		-95.9	2	554	335.9		218.1	2	570	335.9		234.1
3	314	335.9		-21.9	3	604	335.9		268.1	3	285	335.9		-50.9
4	281	335.9		-54.9	4	430	335.9		94.1	4	186	335.9		-149.9
5	567	335.9		231.1	5	197	335.9		-138.9	5	97	335.9		-238.9
6	483	335.9		147.1	6	172	335.9		-163.9	6	63	335.9		-272.9
7	246	335.9		-89.9	7	157	335.9		-178.9	7	14	335.9		-321.9
合計	2351				合計	2351				合計	2351			

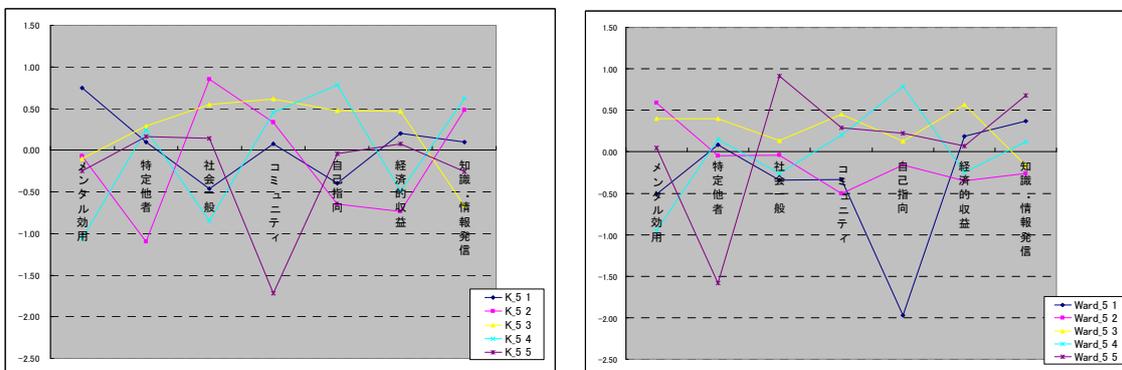
※最近隣法の場合、最大クラスタ以外の度数が全て1となったため、最遠隣法に変更した

最遠隣法は、他の手法と比べてクラスタに含まれる回答者数のばらつきが非常に大きく、偏りの大きい結果となっている。そこで、K平均法とWard法の間で4～7クラスタの場合の、分析結果を比較した（図表参4-2～4-5）。

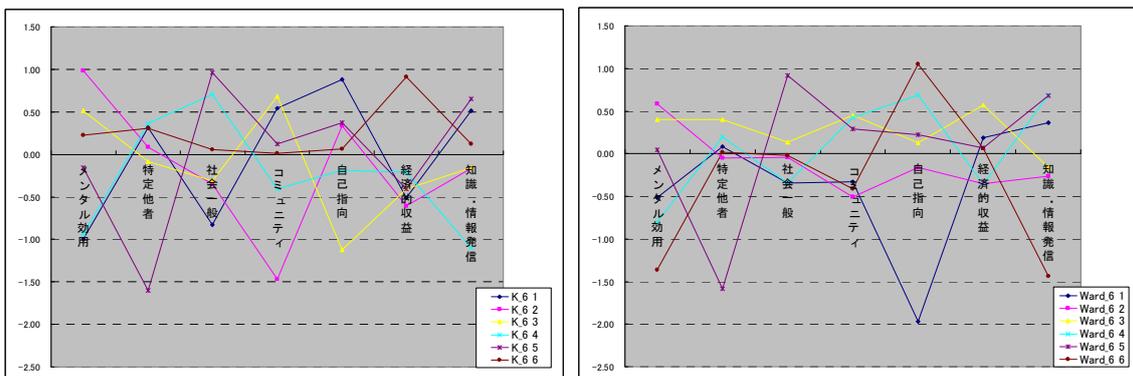
図表参 4-2 4分類でのクラスタ分析結果の比較 (左: K 平均法、右: Ward 法)



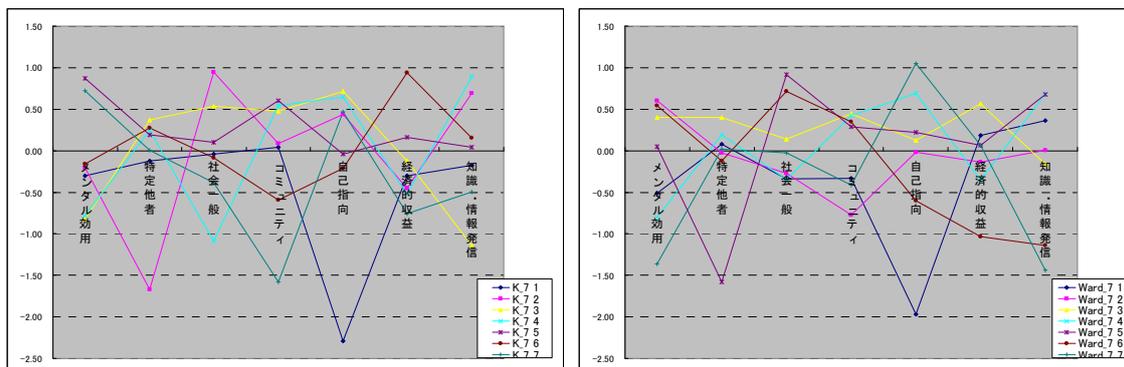
図表参 4-3 5分類でのクラスタ分析結果の比較 (左: K 平均法、右: Ward 法)



図表参 4-4 6分類でのクラスタ分析結果の比較 (左: K 平均法、右: Ward 法)



図表参 4-5 7分類でのクラスタ分析結果の比較（左：K平均法、右：Ward法）



この結果を見ると、K平均法とWard法では、クラスタごとの因子得点の特徴がある程度共通しており、手法によるクラスタの異常な偏りはないと考えられる。またいずれのクラスタ数でも2つの手法に共通性が見られることから、クラスタを分けすぎることによるブレも生じていないと考えられる。したがって、2つの方法・クラスタ数のいずれを選択しても、分析手法上の問題はないといえることができる。

今回の分析では、Ward法のほうがクラスタ規模の大小が分かれており、解釈可能性が高いため、Ward法を用いて分析を進める。また階層化クラスタのWard法であれば、デンドログラムを参考にクラスタの解釈を行うことも可能であり、これもクラスタの特徴を分析する際に使用した。

総務省情報通信政策研究所（調査研究部）

<http://www.soumu.go.jp/iicp/>

〒100-8926 東京都千代田区霞ヶ関 2-1-2
中央合同庁舎第 2 号館 11 階
TEL:03-5253-5496 FAX:03-5253-5497