

ブログの実態に関する調査研究の結果

～ 国内ブログの総数は約1,690万(2008年1月現在)。活発な情報発信が続く ～

平成20年7月

総務省情報通信政策研究所(IICP)調査研究部

～目次～

1	調査研究の概要	1
2	国内におけるブログのコンテンツ量（推計値）	
	（1）現在のコンテンツ量（2008年1月現在）	3
	（2）ブログ数の推移	4
	（3）記事数の推移	5
3	ブログコンテンツの多様化	
	（1）画像・動画ファイル等の活用状況	8
	（2）コミュニケーション機能等の活用状況	9
4	スパムブログの状況	10
5	ブログの開設動機の傾向	11
【付録】		
1	ブログの実態に関する調査概要	13
2	ブログコンテンツ量の推計モデル	14

1 調査研究の概要

調査の目的

- 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。
- CGM(消費者発信型メディア)の1つとしてこの数年急激に普及したメディアであり、日本語による発信量が世界一と言われている。
- 我が国のブログについて、その数や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う。

調査の概要

ブログのコンテンツ量推計調査

- 国内のブログコンテンツ総量
- ブログコンテンツ量の推移
- アクティブブログの構造

スパムブログの調査

- スパムブログの割合
- スパムブログの内容

ブログ開設者の意識調査

- ブログ開設者の利用状況
- ブログ開設者の開設動機等

各種定量データとブログ開設者の定性的なデータから
ブログの実態を把握、分析

2 国内におけるブログのコンテンツ量（推計値）

(1) 現在のコンテンツ量（2008年1月現在）

- 2008年1月現在、インターネット上で公開されている国内のブログ^(※1)の総数は約1,690万（記事総数は約13億5,000万件。）、データ総量は42テラバイト（うち、テキスト情報は約12テラバイト^(※2)）。
- このうち、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の数は約300万で、ブログ総数の2割弱に相当。
- 2001年1月以降に開設されたブログの総数（既に削除済みのものを含む。）は累計で約2,240万（記事総数は約17億9,000万、データ総量は54テラバイト）^(※1)。

※1 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サイト等を利用して開設されているブログを対象として調査。クローラー（リンクをたど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）調査で得られた2008年1月現在の主要20ブログサイト（公表データによる登録者数が上位のサイトから抽出）におけるアクティブなブログの開設時期別数とブログ開設者のアンケート調査により算出したブログの「更新継続率」・「削除率」に基づき、過去（2001年1月～）に開設されたブログ数等を推計。

※2 書籍1冊の原稿の情報量の約2,700万冊分に相当（書籍1冊当たりのデータ量を457キロバイトとして換算）

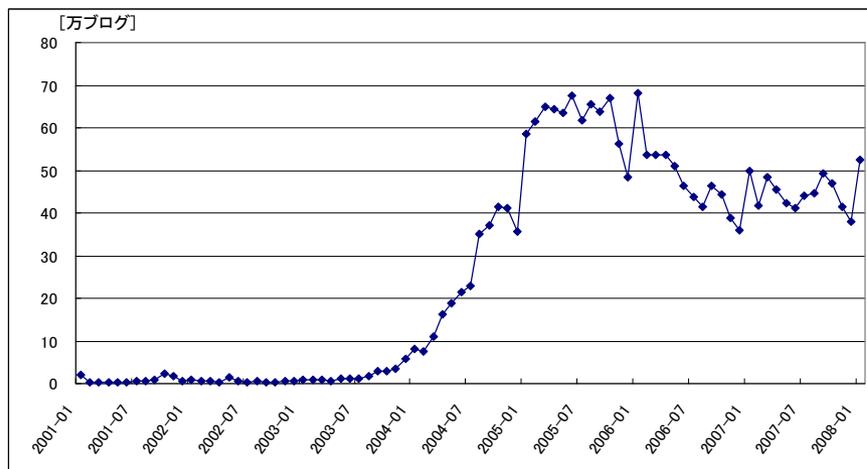
ブログ総数、記事総数、データ総量の推計（2008年1月現在）

	ブログ総数 [万ブログ]	記事総数 [百万件]	データ総量 [テラバイト]	テキストデータ総量 [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

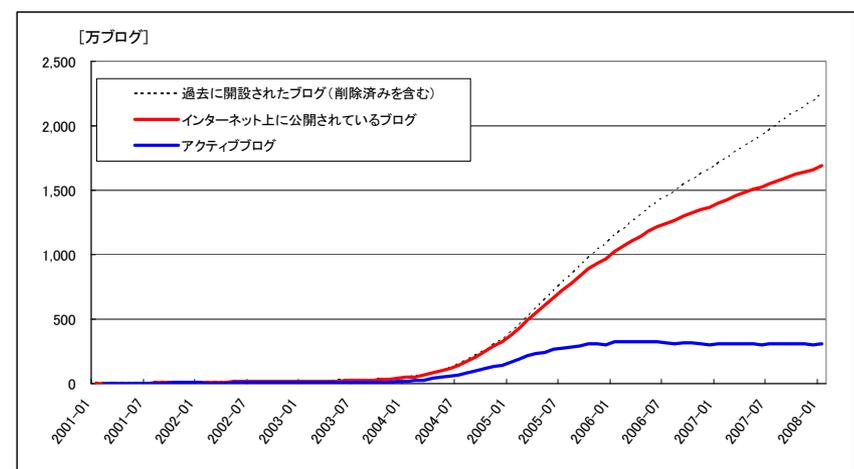
(2) ブログ数の推移

- 毎月新たに開設される新規ブログ数は、主要なブログサービスが開始された2003年から急速に増加し、近年は毎月40万から50万程度で推移。
- インターネット上で公開されているブログ数は、特に2004年から2005年頃にかけて急増し、その後も引き続き増加傾向。
- 一方、アクティブブログ数は同様に2004年から2005年にかけて急増したが、その後は300万でほぼ横ばいで推移。
- 新規にブログが開設される一方で、更新されなくなった既設ブログも多い。

毎月の新規開設ブログ数の推移



国内のブログ総数の推移

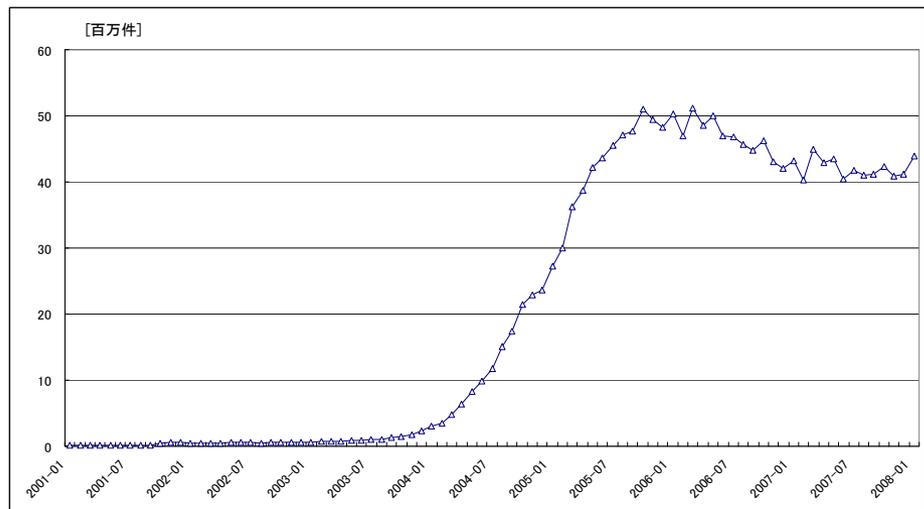


(3) 記事数の推移

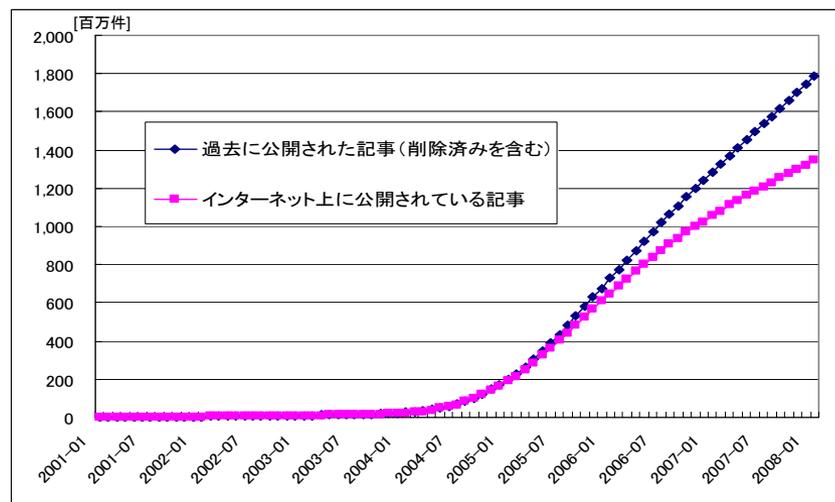
- 毎月新たにブログに書き込みされる記事(※)数については、2004年から2005年にかけて急増し、近年も毎月4,000万から5,000万件程度で推移。活発な情報発信が続く。
- ネット上で公開されている記事の総数は2004年から急速に増加。近年においても増加傾向が続く。

※ ブログサイト等の作成フォームにより1回の書き込みで生成されるブログのコンテンツの単位をいう。

毎月の新規記事数の推移



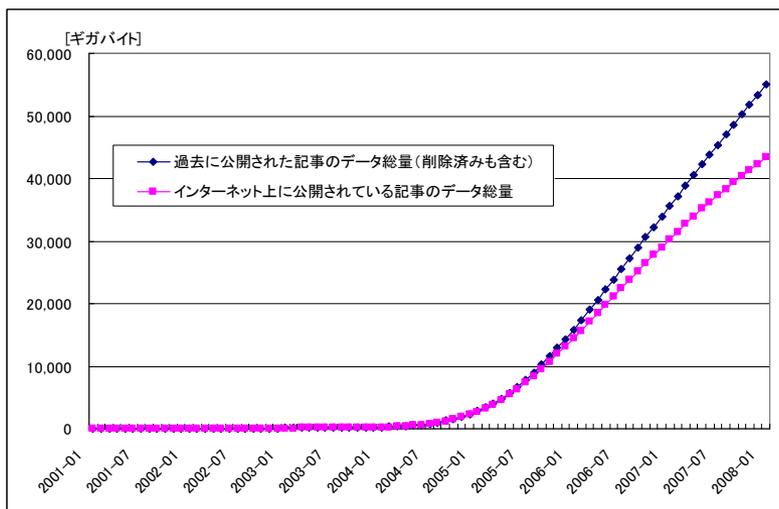
国内ブログの公開記事総数の推移



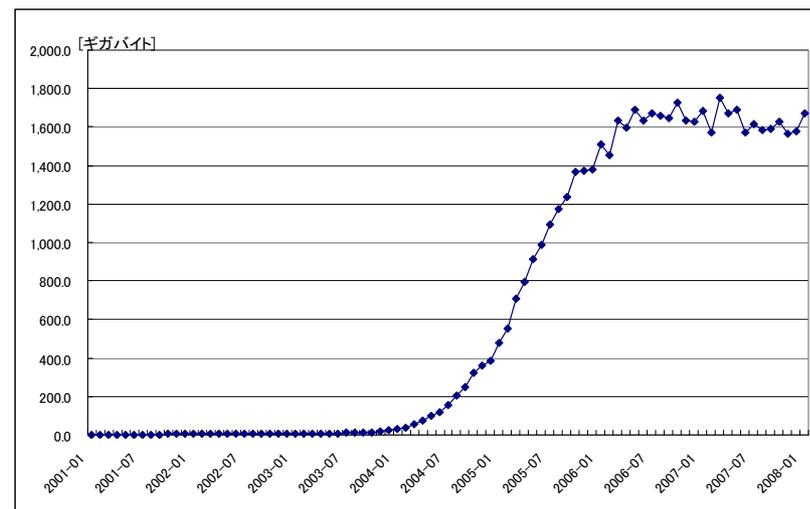
<参考> 拡大するブログのデータ総量

- ブログのデータ総量は、ブログ数の増加、記事数の増加、そして画像・動画ファイル等の増加により急速に増加。
 - 近年では新規の記事書込みに伴い、毎月1.6テラバイト前後のデータが追加されている。

国内ブログのデータ総量の推移



毎月の新規記事データ量の推移



3 ブログコンテンツの多様化

(1) 画像・動画ファイル等の活用状況

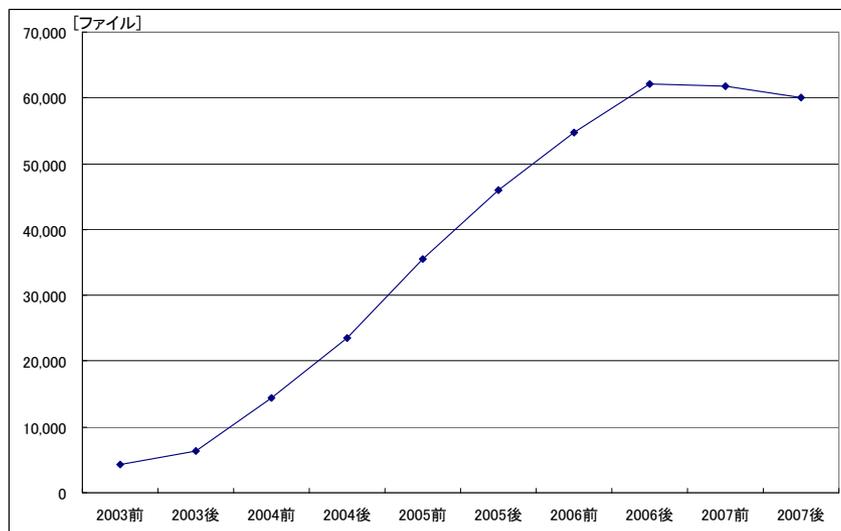
○ 1記事当たりの画像ファイル(※1)、音声・動画ファイル(※2)の掲載数は、2004年から2006年にかけて急速に増加。画像ファイル等によりブログコンテンツが豊かになる傾向。

- ・ 2006年以降、画像ファイルは1記事当たり0.6ファイルで推移

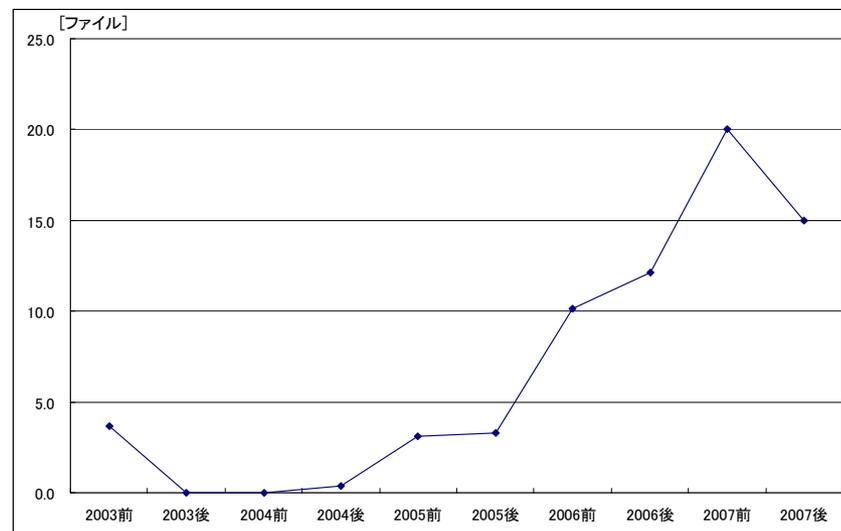
※1 画像ファイルには、主に写真をアップロードする際に用いるJPEGファイル、アイコン等の軽い画像の掲載に用いることが多いGIFファイルなどがある。

※2 音声/動画ファイルには、再生方式により、SWFファイル、WMVファイル、AVIファイル、MPEGファイルなどがある。

記事10万件当たりの画像ファイル掲載数



記事10万件当たりの動画ファイル掲載数



(2) コミュニケーション機能等の活用状況

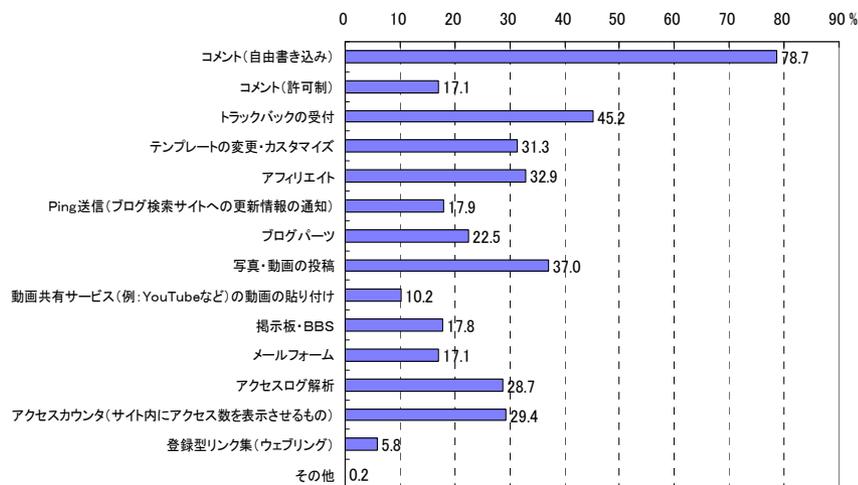
- ブログ開設経験者を対象としたアンケートによると、ブログの付加機能については、「コメント(※1)」が最も多く全体の約8割が利用している。その他、「トラックバック(※2)」については約半数、「アフィリエイト(※3)」、「写真・動画の投稿」、「テンプレートのカスタマイズ」が各3割程度となっている。
 - ・ 2003年以降、1記事当たりのコメント数は概ね1.5前後、トラックバック数は概ね0.1前後で推移。

※1 書き込みされた記事に対して閲覧者が記事に対する意見や感想を当該記事内で述べる事ができる機能。コメントを事前許可制に設定する機能もある。

※2 書き込みされた記事の閲覧者が、自身のブログへのリンクを当該記事内に作成することができる機能。

※3 ブログサイトが連携又は許可している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機能。

利用しているブログの付加機能(複数回答)



4 スパムブログの状況

- 検索エンジンからのアクセスを増やすために、様々なキーワードを大量に埋め込んだ広告誘導のブログや、他のブログから掲載内容をコピーして作成されたブログなどの、いわゆる「スパムブログ」の増加が指摘。
- 2008年1月現在のアクティブブログについて調査したところ、ブログ数では12%がスパムブログに該当。
- スパムブログの出現率については、ブログサイト間で大きな差がみられる。
- スパムブログの内容は、販売誘導が38.3%、アフィリエイト収入が17.1%、アダルト・出会い系サイトへの誘導が7.0%。

【スパムブログの主な種類】

●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、読者には意味がとまらない文章が掲載されている。

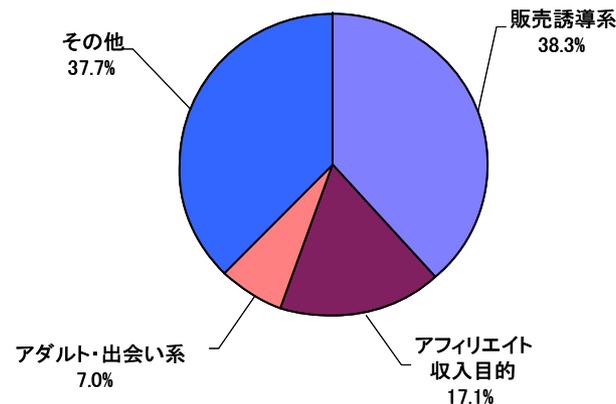
●キーワード抽出型

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)。ブログのタイトルと内容が無関係で、広告ばかりが掲載されている。

●コピー型

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

スパムブログの内容



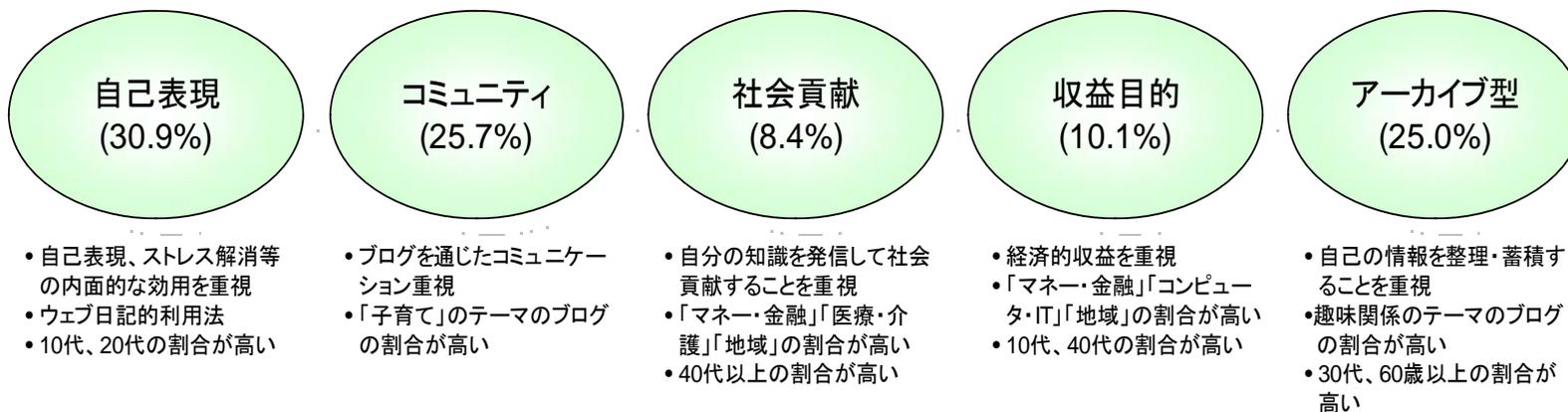
その他：分類が困難なもの、意図が不明なものが含まれる

5 ブログの開設動機の傾向①

- ブログ本来の情報発信機能の他、コメント・トラックバックのコミュニケーション機能、アフィリエイトなどの収益機能が活用され、ブログを開設する目的・態様も多様。
- ブログ開設の動機として①自己表現(30.9%)、②コミュニティの形成(25.7%)、③社会貢献(8.4%)、④収益目的(10.1%)、⑤アーカイブ型利用(25.0%)の5つが重視される傾向(※)。

※ ブログ開設経験者アンケートのブログ開設・更新の動機とブログの想定読者と記述内容に関する設問の回答に基づいて、7つの主なブログ開設要因を抽出し、7つの開設要因の傾向により、5つのグループに分けた。

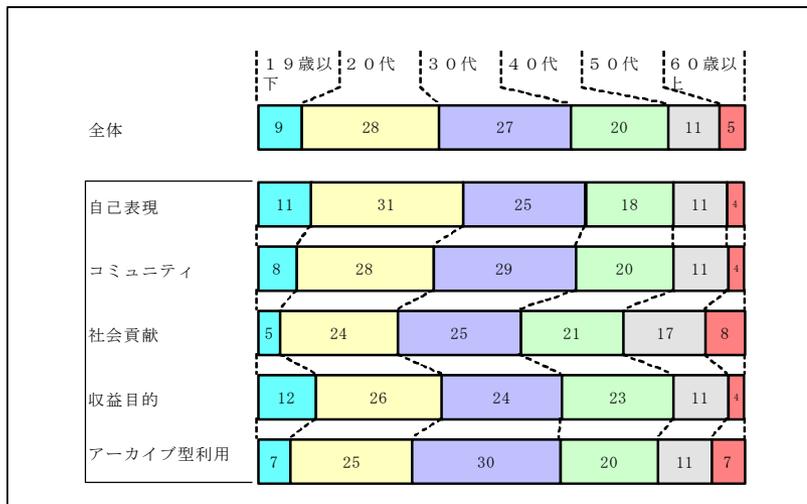
ブログ開設動機による分類



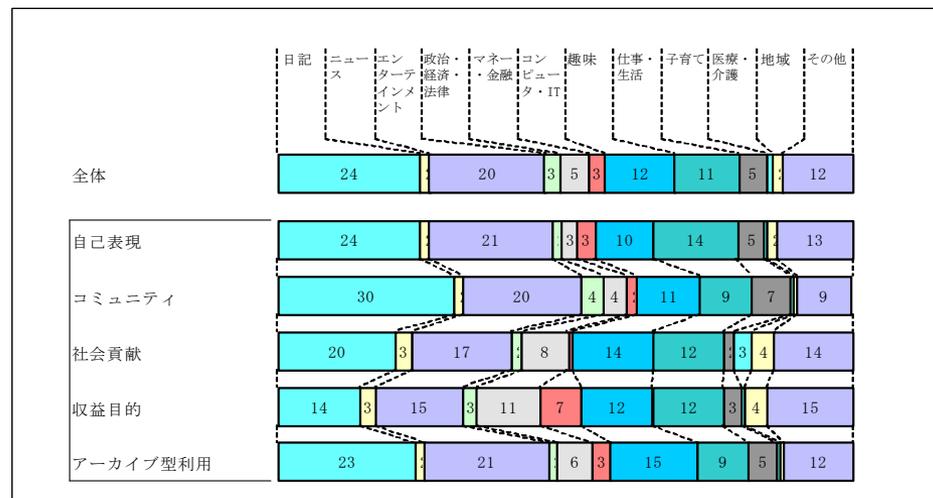
5 ブログの開設動機の傾向②

- 「自己表現重視」は全体の30.9%、自己表現やストレス解消等の内面的な効用重視。ウェブ日記的利用法で10代、20代の割合が高い傾向。
- 「コミュニティの形成重視」は全体の25.7%、ブログを通じたコミュニケーションを重視。「子育て」のテーマのブログの割合が高い傾向。
- 「社会貢献重視」は8.4%、自己の知識を発信して社会貢献することを重視。テーマでは「マネー・金融」、「医療・介護」、「地域」の割合が高く、40代以上の割合が高い傾向。
- 「収益目的重視」は10.1%、経済的収益を重視し、テーマでは「マネー・金融」、「コンピュータ・IT」、「地域」の割合が高く、10代、40代の割合が高い傾向。
- 「アーカイブ型利用」は、25.0%、自己の情報を整理・蓄積することを重視。趣味関係のテーマのブログの割合が高く、30代、60歳以上の割合が高い傾向。

ブログ開設動機別の年代別傾向



ブログ開設動機別のテーマ別傾向



【付録1】ブログの実態に関する調査概要

1 ブログ開設経験者アンケート調査概要

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	<ul style="list-style-type: none"> 回答者属性(年代、性別、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)

2 ブログコンテンツ量の調査概要

(1) アンケート調査での情報収集

1のブログ開設経験者を対象にしたアンケート調査により、ブログの「更新継続率」と「削除率」を算出した。「更新継続率」とは、ある時点でのアクティブブログのうち、1箇月後にも月1回以上の頻度で更新されているブログの割合である。また、「削除率」とは、ある時点にネット上に存在するブログのうち、1箇月以内に削除されるものの割合である。更新継続率、削除率共に2001年1月以降の各月ごとに算出した。

(2) ブログ総数の推計

クローラー調査では、ブログサービス事業者が提供する20の主要ブログサイトとブログ検索サイトを対象に、2008年1月から2月にかけての1箇月間で更新が行われたアクティブブログを捕捉し、主要ブログサイト内の各ブログの記事のクロウリング(リンクをたどってデータを集めること)を行った。各ブログの記事のうち最も古い記事の日付をそのブログの開設日とみなし、開設年月別の新規開設ブログ数を集計した。このデータを基に、これまでに開設されたブログ総数を推計した。

(3) マルチメディアファイル等の推移

クローラーで取得したアクティブブログのデータを基に集計

3 スпамブログの調査概要

(1) 調査対象ブログ

2の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、20の主要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。

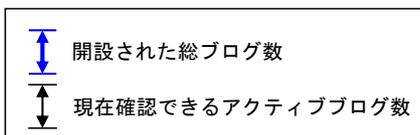
(2) 調査方法

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スパムブログか否かの判別を実施した。

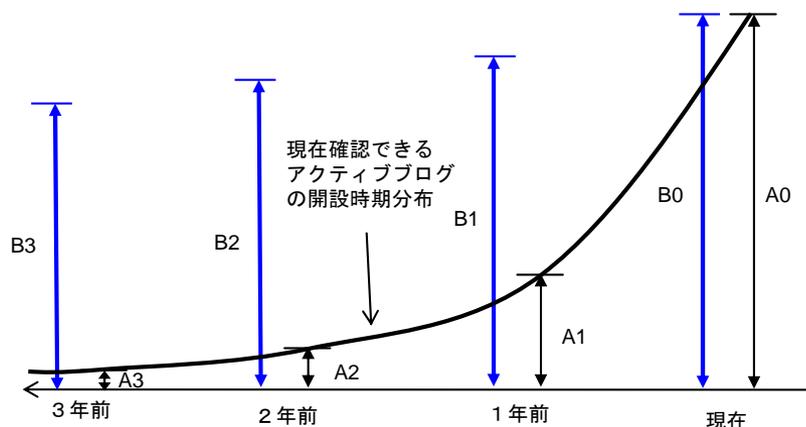
(3) スпамブログ判別基準

- ・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見られないが、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、出会い系の記事を掲載しているもの

【付録2】 ブログコンテンツ量の推計モデル



ブログの更新継続率モデル：
開設されたブログが一定期間後
にアクティブである確率
1年後：30%
2年後：10%
3年後：3%



A1：現在確認できる1年前開設のアクティブブログ数
B1：1年前に開設された総ブログ数
 $A1 = 0.3 \times B1$

A2：現在確認できる2年前開設のアクティブブログ数
B2：2年前に開設された総ブログ数
 $A2 = 0.1 \times B2$

A1、A2、A3・・・から、B1、B2、B3・・・を算出

現在までに開設されたブログ総数 = $B0 + B1 + B2 + B3 + \dots$

クローラーにより、ブログサイトを一定期間観測すれば、アクティブブログの数を把握することはできるが、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。そこで、次のようなモデルを設定し、国内で現在までに開設されたブログの総数を推計した。

- ① 現在観測できるアクティブブログを収集し、ブログの記事の最も古い日付をそのブログの開設日とみなし、現在観測できるアクティブブログの開設時期別の数と分布を把握
($A0, A1, A2, A3, \dots$ を把握)
- ② 現在観測できるアクティブなブログの開設時期別数を、開設時期から現在までの更新継続率で割り戻し、過去の各時期に開設された総ブログ数を推計 ($A0, A1, A2, A3, \dots$ を更新継続率を用い、 $B0, B1, B2, B3, \dots$ を推計)
- ③ 各時期に開設された総ブログ数を累計することで、現在までに開設されたブログ総数を推計 ($B0, B1, B2, B3, \dots$ を累計)