トピックス

拡大を続けるウェブサイト
～第４回WWWコンテンツ統計調査結果速報～

通信経済研究部研究官　　葛西　義昭

１　はじめに

日本のインターネットの商用利用は平成４年末から平成５年にかけて拡大をはじめた。これ以前は研究開発を主目的として、限定的にしか利用されていなかったインターネットが、平成４年12月のIIJの設立、平成５年12月の日本インターネット協会の設立等の個別の動きの中で、急速に拡大をはじめたのである。当時の資料をみると、インターネットの拡大は「爆発的」と表現されている。それから６年しか経っていない平成11年には社会の急速な情報化がいっそうすすみ、インターネットは一部の研究者が利用する研究開発用のネットワークではなく、企業にとっても、個人の生活においても主要な役割を果たすコミュニケーション手段に成長した。

日本の企業へのインターネット普及率は、平成７年度の段階では11.7%、平成８年度には50.4%、平成９年度には68.2%と急速に増加しており、平成10年度には80%に達した。また、平成10年度には世帯への普及率も10%を超え、インターネット利用人口は1700万人と推計されている。

このようなインターネットの利用急増を反映し、高度情報通信社会推進本部の「高度情報通信社会に向けた基本方針」をはじめとして、政府も積極的にインターネットに対する政策を発表している。

しかしながら、インターネットの拡大速度はあまりにも急であり、インターネットに関する統計は、あまり充実していないのが実情である。

このような現状を背景として、当通信経済研究部では平成10年２月、平成10年８月、平成11年２月、平成11年８月の４回にわたり「WWW（World Wide Web）コンテンツ統計調査」を実施してきた。本稿では、平成11年８月に行なった最新のWWWコンテンツ統計調査の結果を報告する。

２　WWWコンテンツ統計調査の概要

インターネットの機能は、メール、ニュースほか様々であるが、通信経済研究部で実施しているWWWコンテンツ統計調査は、シンクタンクであるアライド・ブレインズ社と共同開発した統計用サーチ型ロボットエンジン（通称：Loki）を使用して得られた結果をもとに、日本国内（jpドメイン内）のWWWに関するサーバ数、ページ数、データ量を推計するものである。

WWWはデータにリンクを張って、ネットワーク上にあるあらゆる情報ファイルにアクセスできる広域情報システムを可能にする。Lokiは、このWWWの特性を利用してURL（Uniform Resource Locater）によって表記されたリンクをたどってWWWページを自動的に巡回し、各WWWページの情報を収集するソフトウェア・プログラムである。実際にWWWページにアクセスして調査を行なう点にLokiを使った本調査の特徴がある。

同種の統計調査のうちでは米国NEC Research Instituteが米国サイエンス誌に昨年発表したものが有名であるが、この調査を含め、大半の調査では少数のサンプルから全体を推計する手法をとっている。Lokiを使用したWWWコンテンツ統計調査はjpドメイン上のリンクを辿って実際にWWWページを調査するという、フィールドワーク的な手法を取ったことにより、現在行なわれているインターネットのコンテンツ調査のなかで、最も精度の高い結果が期待できるものとなっている。

理論的には、Lokiを稼動させることにより、すべてのWWWサーバ、ページの情報を取得することが可能であるが、現実にはインターネットへのアクセス速度に制約があることと、WWWページは日々拡大していることから、jpドメイン内のみの検索という限定をつけても、WWWページに隈なくアクセスし、総数を把握することは困難である。

したがって、WWWコンテンツ統計調査では、実際に調査したデータに基づいて推計を行なっている。Lokiによって調査しているデータのうちWWWサーバ数については、調査中に発見したサーバの毎日の新規発見率を用いて推計を行う。これは、調査期間中の毎日のサーバの新規発見率を基にサーバ総数を推定するものである。また、WWWページの推計に際しては、調査期間中に発見したファイルのうち既発見ファイルの占める比率から回帰式を導き出し、（図表１参照）この式から、「既発見URL率が100%になる時点＝すべてのWWWページを巡回し終えた時点におけるファイル数」と仮定してその構成を推計する。

３　WWWコンテンツ統計調査の結果

以下では、今回のWWWコンテンツ統計調査結果を報告するとともに平成10年２月、８月、平成11年２月に実施した調査結果を比較することにより、WWWコンテンツ拡大の状況を示すこととする。

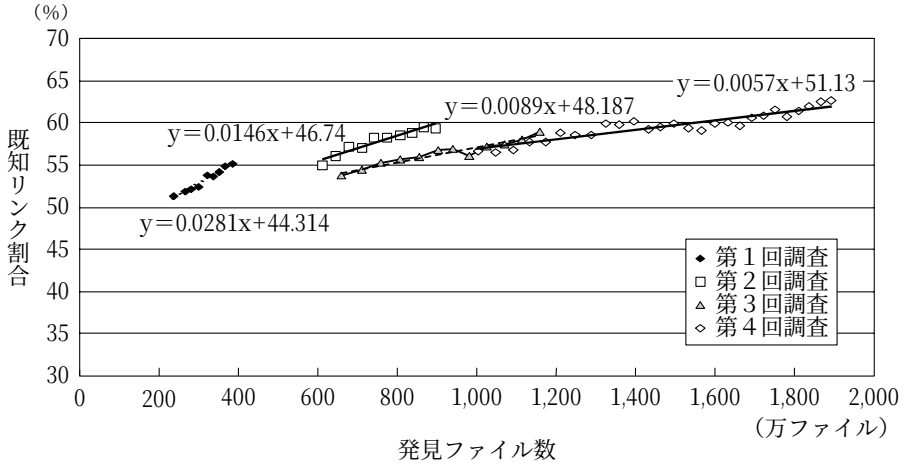
3.1　WWWサーバ数

今回の調査で発見されたWWWサーバの数は約82,000台であった。この結果によりjpドメインには約85,000台のWWWサーバが存在するものと推計している。

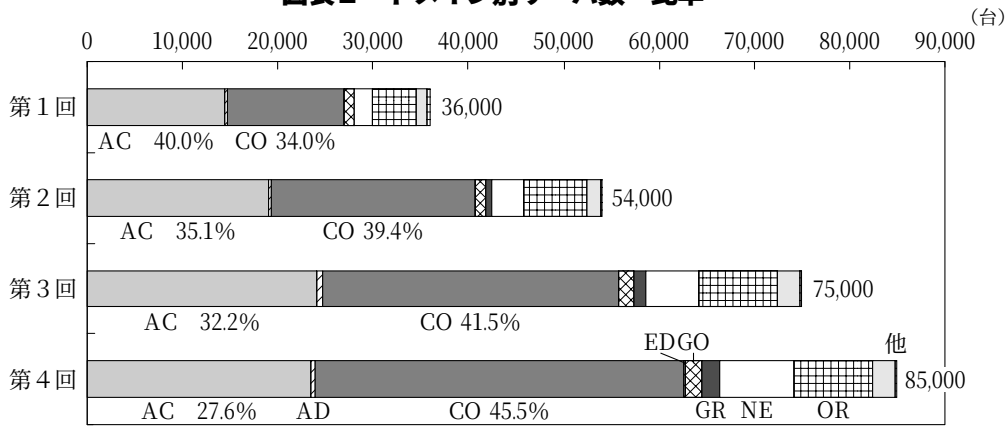
これまでの第一回：36,000台、第二回：54,000台：第三回：75,000台という結果とあわせてグラフにしたものが図表２である。

WWWサーバ数全体については、これまでの３回の調査傾向と変わらず、増加を続けているが、増加の割合はこれまでより低くなっており、これ

図表１　既知ファイルへのリンク割合の推移



図表２　ドメイン別サーバ数・比率



までの調査では半年毎に約20,000台増加していたのに対し、今回は10,000台の増加に留まっている。

次にドメイン別に見ていくと、以下のような状況が判明している。acドメインについては、サーバ総数に占める割合が第三回調査時に32.2%であったところが今回は27.6%に減少し、さらに実サーバ数が約24,100台から約23,400台程度に減少しているとの結果になった。推計値であることを考慮すると、この結果のみをもってacドメインのサーバ数が減少したということはできないが、少なくともacドメインのサーバ数は横ばいであるといえる。

一方、coドメインについては、サーバ総数に占める割合が34.0%から39.4%、41.5%、45.5%と増加を続けており、実数も第一回調査時と第四回調査を比較すると12,200台程度から38,700台と３倍を超える増加を見せている。サーバ総数は第一回から第四回までにおよそ49,000台増加しているが、増加した台数の半数以上がcoドメインにおける増加であることから、日本のWWWは、企業のインターネット導入に牽引されて成長していると考えることができる。

grドメイン、neドメインについても、割合、実数ともに増加傾向にある。両ドメインは任意団体、インターネットサービスプロバイダーが主に利用しているドメインであることから考えて、この結果は、企業だけでなく、個人利用の局面においてもインターネットが普及してきていることを表すものと考えられる。

３.２　ページ数、ファイル数、構成

今回のjpドメイン内の総ページ数に関する調査結果は、3,850万ページであった。なお、これまでの調査では第一回1,020万ページ、第二回1,790万ページ、第三回2,950万ページであった。この一年間の増加ページ数は2,060万ページ、さらにこの一年間における一日あたり増加ページ数は5.6万ページである。

インターネット上には、web上のページを表現するHTMLファイルのほかに、画像、動画、音声、文書（PDFファイル等）などのファイルが置かれており、その総ファイル数は今回の調査結果では約8,570万ファイルであった。過去の調査結果では、第一回1,890万ファイル、第２回3,650万ファイル、第３回5,820万ファイルという結果になっている。この半年の増加は2,750万ファイル、この一年の増加は4,920万ファイルということになる。単に比率でみていくとそれぞれ前回の1.9倍、1.6倍、1.5倍と増加割合が落ちているように感じられるが、実数を確認するとファイル総数の伸びは衰えていないことがわかる。この結果をサーバ数の変化と重ね合わせて考えると、サー

バは整理されているが、利用者のインターネット利用は一層進んでいることが表れているものと考えられる。

ファイル数の構成比については図表３のような結果となった。（図表３－１）が明らかに示しているとおり、調査のたびに全てのファイル形式でファイル数が増加している。

ファイル数の構成比（図表３－２）の内容を確認してみると、今回の調査で初めてHTMLファイル数を、画像ファイル数が上回った。HTMLファイルの増加を画像ファイルの増加が上回るという結果から、WWWページは、単純に量的な面での拡大をつづけているだけではなく、質的にも画像の多い見やすいページへと進化していることが伺える。

次にドメイン別にファイル数の変化を示すのが

図表３

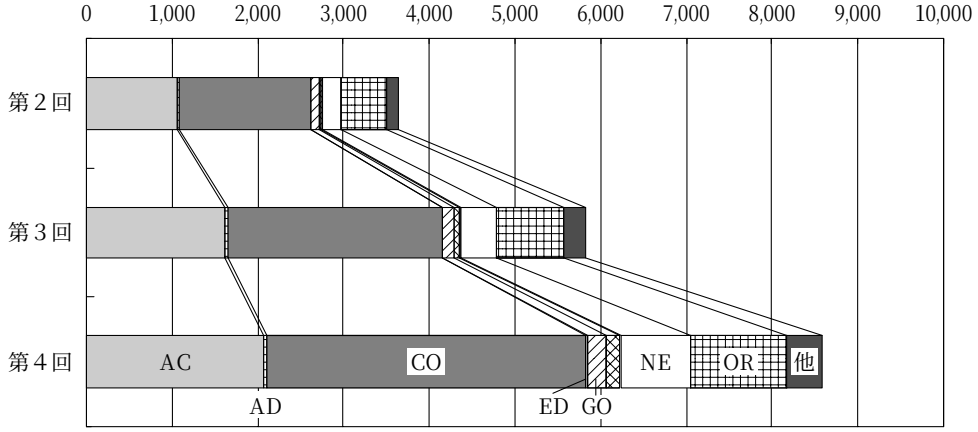
図表３－１　実数（推計）　［万file］

	第１回調査	第２回調査	第３回調査	第４回調査
総ファイル数	1890	3648	5822	8574
HTML	1023	1784	2953	3845
画像	827	1775	2726	4469
動画	2	4	5	7
音声	3	10	11	25
文書・データ	25	61	116	173
不明・他	11	14	11	54

図表３－２　ファイル数の構成比　［%］

	第１回調査	第２回調査	第３回調査	第４回調査
HTML	54.1	48.9	50.7	44.9
画像	43.8	48.7	46.8	52.1
動画	0.1	0.1	0.1	0.1
音声	0.1	0.3	0.2	0.3
文書・データ	1.3	1.7	2.0	2.0
不明・他	0.6	0.4	0.2	0.6

図表４　ドメイン別ファイル数の推移



図表４のグラフである。直近３回の調査結果を比較してみると、すべてのドメインでファイル数が増加しており、ドメイン別のファイル数構成比率はさして変化していないことが確認できる。サーバ数が横ばいであるacドメインでもファイル数の増加が進んでいることから、acドメインのサーバ数の伸びが止まったことが、acドメインにおけるWWWの重要性は低くなったことを表すわけではないことが、ここでも検証されている。また、この結果が、WWWの拡大は単純な量的拡大だけでなく質的変化を伴っていることを推測する根拠となりうる。

3.3　データ量

WWWコンテンツの総データ量は、図表５のとおり第一回306GB、第二回664GB、第三回1,024GB、第四回1,889GBという結果となった。サーバ数、ページ数とは違い、その増加割合、増加量ともに前回の調査結果を上回っている。直近１年間のデータ量増加は1,225GB、半年間のデータ量増加は865GBであるから、この１年間は１日あたり3.3GB、特にこの半年に限っては１日あたり4.7GBのデータが増加していることとなる。

図表５

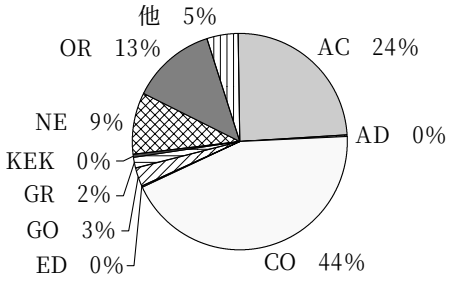
図表５－１　実数（推計）　［GB］

	第１回調査	第２回調査	第３回調査	第４回調査
総データ量	306	664	1024	1889
HTML	46	86	150	211
画像	141	306	409	745
動画	40	78	113	280
音声	11	29	39	88
文書・データ	53	151	300	546
不明・他	15	14	14	19

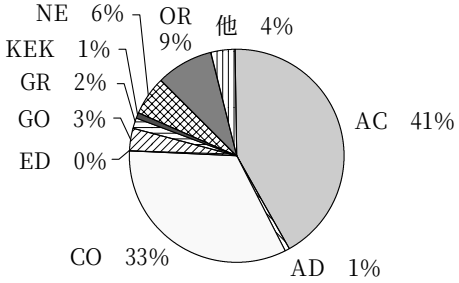
図表５－２　データ量の構成比　［%］

	第１回調査	第２回調査	第３回調査	第４回調査
HTML	15.1	12.9	14.6	11.2
画像	46	46.2	39.9	39.5
動画	13	11.7	11.1	14.8
音声	3.6	4.4	3.8	4.6
文書・データ	17.3	22.7	29.3	28.9
不明・他	4.9	2.1	1.3	1.0

図表６　ドメイン別ファイル数



図表７　ドメイン別データ量



さらにドメイン別の結果で注目すべき点はacドメインのデータ量である。acドメインはcoドメインに比べてファイル数は少なかった（図表６）が、データ総量は多い（図表７）ことが判明した。理由として、音声・動画などのマルチメディアファイルのデータ量がかなり多いことが挙げられる（音声、動画のデータ量はacドメインのデータ量の30%を占めるのに対しcoドメインでは10%）。また、画像については、ファイル数はcoドメインの方が圧倒的に多いのに、データ量はacドメインとcoドメインでほぼ互角であることも判明した。

全体に、acドメインの方が「重いマルチメディアファイル」を多用しており、逆に、coドメインの方は「ビジュアルだが軽いページ」を指向していると言える。

４　今後の課題

ここまで報告してきた調査結果から総合的に判断すると、日本のインターネットは、単調な増加を続けている段階から、質的な成長を遂げる段階にきているといえる。今後は、インターネットの質的転換を捉えるため、WWWコンテンツ統計調査の一環としてホームページの内容を分析することが必要であり、この分析手法を確立することが急務であると考える。

また、海外と比較を行なって日本のWWWコンテンツの特徴を明らかにすることが必要なため、現在はシンガポール（sg）ドメイン内のWWWコンテンツに対し試験的な調査を実施しているところである。

インターネットは日本で開催が予定されているINET2000、電子商取引の普及に向けた各種政策など、非常に注目を集めており、今後も一層拡大していくことが予想されている。

また、世界的にみてインターネットの普及が圧倒的に先行した米国の状況を見ると、さまざまなかたちで新たなビジネスも生み出されている。

日本はインターネットをめぐる大きなビジネスチャンスを生かすためにインターネットの急激な進展の状況を確実にとらえ、次の戦略を構築していかなければならない。この観点からも、WWWコンテンツ統計調査の重要性はより一層増してくるものと思われる。

以上

参考文献

郵政省編　「平成11年版通信白書」　ぎょうせい
日本インターネット協会編　「インターネット白書 '99」　インプレス
村井純著（1995）「インターネット」　岩波新書
会津泉著（1994）「進化するネットワーク」　NTT出版

参考　（平成11年度通信白書　資料５より）

ドメイン名		表　　記	組織の種別および登録資格等
属性型	AC	***.ac.jp	⒜学校教育法および他の法律の規定による学校（EDドメイン名の登録資格の⒜に該当するものを除く）、大学共同利用機関、大学校、職業訓練校 ⒝学校法人、職業訓練法人
	CO	***.co.jp	株式会社、有限会社、合名会社、合資会社、相互会社、特殊会社、その他の会社および信用金庫、信用組合、外国会社 ※外国会社の場合には、日本において外国会社の登記を行っていること。
	GO	***.go.jp	日本国の政府機関、各省庁所轄研究所、特殊法人（特殊会社を除く） ※特殊法人はGOドメイン名とORドメイン名のいずれかを選択することができる。
	OR	***.or.jp	⒜財団法人、社団法人、医療法人、監査法人、宗教法人、特殊法人（特殊会社を除く）、農業協同組合、生活協同組合、その他CO、AC、ED、GOのいずれにも該当しない日本国法に基づいて設立された法人 ⒝国連等の公的な国際機関、外国政府の在日公館、外国政府機関の在日代表部その他の組織、各国地方政府（州政府）等の駐日代表部その他の組織 ※特殊法人はGOドメイン名とORドメイン名のいずれかを選択することができる。
	AD	***.ad.jp	⒜JPNICの会員が運用するネットワーク ⒝JPNICがインターネットの運用上必要と認めた組織 ※JPNICの会員が運用するネットワークを、ドメイン名登録等に関する規則の第９条における組織とする。
	NE	***.ne.jp	日本国内のネットワークサービス提供者が、不特定または多数の利用者に対して営利または非営利で提供するネットワークサービス ※登録するドメイン名をネットワーク上における利用者の識別子の一部とするために利用すること。 ※ネットワークサービスを提供する組織は、日本に在住する個人または日本に登記のある法人であること。 ※利用者に対して提供するネットワークサービスの内容が明文化されていること。 ※同一の組織が異なるサービス内容を持った複数のネットワークサービスを提供している場合、一つ一つのネットワークサービスを、ドメイン名登録等に関する規則の第９条における１組織とする。
	GR	***.gr.jp	複数の日本に在住する個人または日本に登記のある法人で構成される任意団体※代表者および副代表者は、日本に在住する個人または日本に登記のある法人であること。
	ED	***.ed.jp	⒜保育所、幼稚園、小学校、中学校、高等学校、盲学校、聾学校、養護学校、専修学校および各種学校のうち主に18歳未満を対象とするもの ⒝⒜に準じる組織で主に18歳未満の児童・生徒を対象とするもの ⒜⒜または⒝に該当する組織を複数設置している学校法人、⒜または⒝に該当する組織を複数設置している大学および大学の学部、⒜または⒝に該当する組織をまとめる公立の教育センターまたは公立の教育ネットワーク
地域型	一般地域型	***.tokyo.jp等	⒜AC、CO、ED、GO、OR、NE、GRのいずれかの属性型ドメイン名の登録資格を満たす組織 ⒝病院 ⒜日本に在住する個人
	地方公共団体	***.metro.tokyo.jp等	地方公共団体、地方公共団体の下部組織