

OSS デスクトップの普及を阻害する Web コンテンツの分布状況

Distribution of non-standard web content that creates impediments to the widespread use of OSS desktops

伊藤 宣博*¹
Norihiro Itoh

今給黎 道明*¹
Michiaki Imagiire

清水 浩行*²
Hiroyuki Shimizu

飯尾 淳*²
Jun Iio

吉野 公平*³
Kohei Yoshino

瀧田 佐登子*³
Satoko Takita

Abstract: WWW(World Wide Web)の世界的な普及によって、その上で公開される Web コンテンツも膨大な量になっている。コンテンツの作成に当たっては、HTML(Hyper Text Mark-up Language)、スタイルシート、JavaScript などの標準仕様化された技術が用いられる。しかしながら、それを実装する Web ブラウザは、製品ごとに標準仕様への準拠度や解釈の相違、さらには独自仕様の追加等によって完全な互換性を保たないため、特定のブラウザのみに対応したコンテンツが産み出され続けている。本稿では、昨年実施した Yahoo! Japan カテゴリに登録される 20 万サイトという大規模なコンテンツ調査の結果から、ブラウザ間で非互換問題を引き起こしている Web コンテンツの日本における分布状況を明らかにし、利害関係者ごとの対策を探る。

Keyword: OSS web browser, Web contents interoperability, HTML/XHTML, CSS, JavaScript, Exhaustive survey

1 はじめに

WWW (World Wide Web)は、1989 年の誕生から 18 年が経過し、現在ではインターネットにおいてなくてはならないサービスとなっている。英 Netcraft 社が、2006 年 11 月に行った調査^[1]によれば、全世界におけるサイト数が 1 億を越えたことが明らかになっている。同時に、それらのサイト上で公開される Web コンテンツの量も膨大なものとなっている。

この膨大なコンテンツの多くは、HTML (Hyper Text Mark-up Language)、CSS (スタイルシート)、JavaScript といった技術を用いて作成されている。これらの技術は、W3C (World Wide Web Consortium)等の標準化組織によって仕様化され、広く公開されている。したがって、これらの仕様に基づいて作成されたコンテンツは、どの Web ブラウザで見ても同じように表示されるべきである。

しかしながら、現実にはブラウザ間の完全な互換性が保たれていないため、あるコンテンツがすべてのブラウザ製品で同じように閲覧できない場合が発生する。互換性が保たれない理由としては、ブラウザごとに標準仕様への準拠度がまちまちであったり、または仕様のあいまいさによって解釈の異なる実装がなされたりする等が挙げられる。

この問題をさらに複雑にしているのは、現在のデファ

クトスタンダードとして圧倒的なシェアを持つ Microsoft 社の Windows オペレーティングシステム(OS)の存在である。この OS には、一体化した Web ブラウザである Internet Explorer (以下、IE と略記)が存在しており、Windows とともに大きなシェアを持っている。そのため、作成したコンテンツの検証に当たり、IE を基準に、さらには IE のみで検証といったことが当たり前に行われている。また、このブラウザは、標準仕様に基づかない機能や OS に特化した機能を多数持っているため、IE でのみ利用可能なコンテンツが産み出される原因となっている。

このことが、特定企業の OS への囲い込み要因のひとつとなっており、Linux 等のオープンソース(以下、OSS と略記)デスクトップ環境の普及を妨げている。すなわち、IE でしか見ることのできないコンテンツがあるため、Linux 等へ移行できないという状況が作り出されている。

このような状況を踏まえて、独立行政法人情報処理推進機構(以下、IPA と略記)のオープンソースソフトウェア活用基盤整備事業の一環として、「OSS デスクトップの普及に資する Web コンテンツ互換性向上に関する調査」を実施した。これは、Yahoo! Japan カテゴリに登録されているサイトから 20 万サイトを抽出し、そのコンテンツに含まれる各種の非互換性を産み出す要因(以下、非互換要因と表記)の含有度を調査し、その結果を分析したものである。

本稿では、その成果を元に、日本における Web コン

*1 株式会社アルゴ 21

*2 株式会社三菱総合研究所

*3 有限責任中間法人 Mozilla Japan

コンテンツのブラウザ間非互換問題の実態を明らかにしていく。同時に、このような問題を減らしていくために、ベンダーやクリエイターといった利害関係者ごとの対策を探っていく。

2 関連研究

本調査の内容に言及する前に、これまでに行われてきた関連活動および関連研究について述べる。言うまでもなく、ブラウザ間におけるコンテンツの非互換問題は、これまでも数多く指摘されてきた。さらに、不具合コンテンツ情報を報告する仕組みを提供することで、情報収集を図るといった試みもなされてきた。

国内外において、これまでに実施されてきたコンテンツの非互換調査および関連活動には、The Web Standard Project^[2]の様々な活動や、もじら組によるWeb標準化レポート^[3]、日本トップウェブサイトプロジェクト^[4]などがある。問題をバザールモデルで収集しようという試みとしては、Firefox1.5から導入された「問題のあるウェブサイトを報告」機能^[5]や、TouchUpWebプロジェクト^[6]へのフィードバックなどがある。また、コンテンツ作成用の各種書籍^[7,8,9]でも、ブラウザ間の互換性確保の重要性は指摘されており、技術的な情報も数多く存在する。さらに、ソフトウェア工学的な観点からも、コンテンツだけでなくウェブサービスの互換性を分析するための試み^[10]や、互換性テストの検討^[11]など、各種の研究が行われている。

このようにブラウザ非互換問題に対する活動は脈々と続いてきたものの、これまで行われた多くの活動では散発的な問題の指摘にとどまっていた。その原因はさまざまに考えることができるが、ブラウザ製品の普及や改善が目的である等、根本的には実態把握が活動の主眼となっていないことであろう。このことが、非互換問題の分布状況を把握するため、我々が網羅的なコンテンツ調査を行うことになった理由でもある。

なお類似の調査として、米国政府のウェブサイトがWeb標準に準拠している割合の調査事例^[12]や、英国におけるウェブサイトの何パーセントがWeb標準に従っているかを測定した最近の調査事例^[13]がある。これらはいずれもW3C Validator^[14]を利用してチェックしたものであり、より現実的な問題発見を目的とした我々のアプローチとはやや方向性が異なる。

3 調査方針と手順の概要

我々は、調査を行うにあたって、二つの基本方針を設定した。ひとつは、非互換要因の抽出と、それらに対する改善方法を示すことである。これらは、非互換要因を含むコンテンツの生成に関わる利害関係者に対する情報として提供する。二つには、非互換要因を含むコンテンツの実態を明らかにすることである。調査対象をインターネット及びイントラネットとし、インターネット上のコンテンツはチェックツールと非互換要因ごとに

用意されたルールセットによる機械的な調査を行い、イントラネットについては目視やインタビューなどの調査手法を用いることとした。これらの方針に則って、以下のように作業手順を決定した。

3.1 非互換要因の収集と分析

Webコンテンツに含まれる非互換要因を調査するためには、はじめに、どのような非互換要因が存在するのかを明確に定義する必要がある。そこで、我々が立ち上げた先行プロジェクトであるTouchUpWebプロジェクトの成果を利用することが考えられた。このTouchUpWebシステムは、Firefoxブラウザ上におけるコンテンツ表示の不具合に対して、コミュニティによって作成・蓄積された修正スクリプトをサーバから利用者に配信することで、これを改善するものである。蓄積された修正スクリプトを分析することで、不具合をもたらした非互換要因を特定することが可能となる。また、インターネット上に公開されている情報や、一般に市販されているブラウザごとの表示や動作の違いを記載した書籍に対する文献調査を行うことで、主要な非互換要因の抽出を行った。

調査の結果、172種の非互換要因が抽出された。ただし、これらはすべての非互換要因を網羅しているわけではないことに注意が必要である。すべてを洗い出すためには、それぞれのブラウザのバージョンごとの仕様やバグをすべて把握する必要があるが、これは非現実的な作業である。

抽出された非互換要因は、大きく二つの観点から分類される。第一の観点は要因が含まれる箇所による分類であり、図1のように分類をおこなった。このうち、「HTML/XHTMLとCSS」というのは、個別では問題ではないが、組み合わせによって要因化するものを表している。図1からは、このような組み合わせによる非互換要因は少なく、単独で発現する要因が多くを占めていることがわかる。

第二の観点は標準規格との比較による分類であり、図2のように分類を行った。「ブラウザ独自機能」が60%を占めており、件数にして103件となっている。

さらに各要因に対して、「修正の可否」の観点から4段階、「影響度」の観点から5段階の分類を行った。このうち、「影響度」の観点から分類ごとの影響度を見たのが、以下に示す二つのグラフである(図3, 図4)。図3では、JavaScript/DOMの多くを対象外とした以外は、非互換要因が含まれる箇所別では影響度ごとに大きな差異は生じていない。図4の方を見ると、影響度「大」と「小」で「ブラウザ独自機能」が占める割合が高いことがわかる。

3.2 ルールセットの整理と増強

抽出された172種の非互換要因から、コンテンツの判定に用いるためのルールセットを作成した。ルール

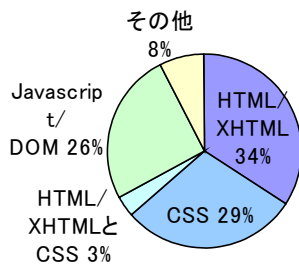


図 1 非互換要因の内訳(a)

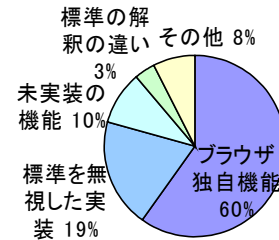


図 2 非互換要因の内訳(b)

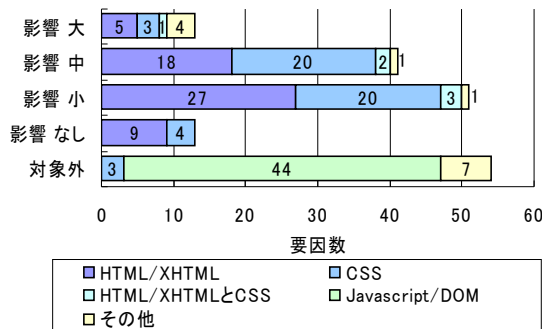


図 3 影響度別の分類(a)

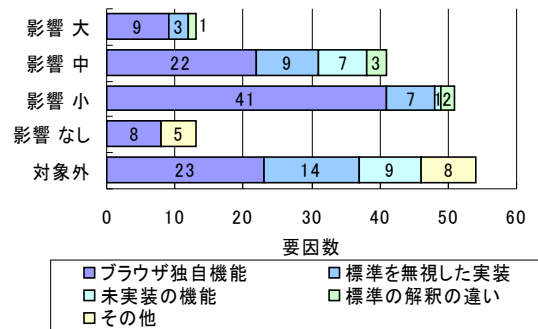


図 4 影響度別の分類(b)

セットは JavaScript による非互換要因判定モジュール (図 5) であり、調査の目的に適さないものや問題があるものを除き、要因ひとつに対してひとつの判定モジュールが作成された。調査期間内に二度のコンテンツチェックが行われており、1 回目の実施結果を踏まえたルールセットの増強が行われた。これには、1 回目で保留された判定モジュールの修正分と、新たに要因として追加された項目を含んでいる。チェック中にも非互換要因の抽出は継続して行われており、それらの中には今回の調査から外れることになった項目もある。調査期間内に作成された判定モジュールの数は、142 件である。

3.3 チェックツールの開発

今回の調査では、調査期間内に順次提供される判定モジュールを追加組み込み可能なチェックツールが必要であったが、既存のものをそのまま利用することは困難であったため、目的にあったクローリングツールを開発し、それを利用することとした。また、期間内の調査作業完遂のため、同時に各種の支援スクリプトを作成し、利用した。

開発当初は判定モジュールの仕様に流動的な部分があったが、最終的に TouchUpWeb プロジェクトの成果を流用しやすく、またコンテンツデータを DOM として扱うことが容易な、JavaScript を採用することが決定した。このことから、チェックツール本体は、判定モジュールとなる JavaScript の実行環境を備え、かつ、連続実行可能であることが必要とされた。

```

(
function() {
  try {
    var elms = document.getElementsByTagName("bgsound");
    if (elms.length != 0) return "NG";
    return "OK";
  } catch (ex) {
    return "EXCEPTION";
  }
}
);

```

図 5 非互換要因判定モジュールの例

この実現のために、OSS のツールである JSpider の利用検討から始まり、Mozilla XULRunner など 4 つのプロトタイプを作成を行って検討を重ねた結果、最終的には Firefox の拡張機能 (アドオン) として提供する形となった。開発に難航した部分は、正規化されていない HTML を DOM に変換する部分であった。

3.4 クローリングによるデータ収集と分析

インターネットに対する網羅的な実態調査については、日本語のコンテンツを中心に調査した。具体的には Yahoo! Japan カテゴリ全体を対象として、登録されている 27 万弱のサイトから JP ドメインのものを中心に抜粋した 20 万サイトを対象に、チェックツールを適用することにした (図 6)。コンテンツ調査における一連の作業手順は、図 7 のように行った。

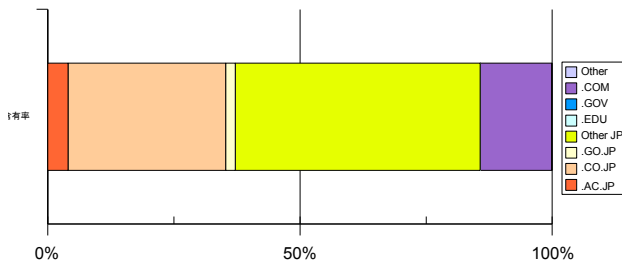


図 6 選別サイトのドメイン構成

3.5 イン트라ネットに対する状況調査

イントラネットの場合、一般には公開情報として明らかになっていない、イントラ特有の複雑な問題が多数存在すると考えられるため、自動チェックツールがうまく機能すると考えにくい面がある。そこで、イントラネットの調査は人手による調査と、当事者である情報システム部に対するインタビュー調査により、状況を分析した。作業手順は調査対象によって若干異なるが、基本的には以下の手順で実施した。この調査は、5 組織を対象とした。

- ① システム化方針のインタビュー
- ② システムの選択
- ③ システム概要のインタビュー
- ④ ログイン ID の選択とログイン
- ⑤ メニュー項目の選択
- ⑥ 各機能へのアクセス
- ⑦ 依存性チェック

また、コンテンツの作成を企業活動として行っているコンテンツクリエイターに対して、インタビューを実施した。インタビューの目的は、特定ブラウザでの非互換要因の認識や、コンテンツの作成方法及び検証方法をインタビューで明らかにしたうえで、非互換コンテンツが作成される背景にある現実的な問題点を浮彫りにすることである。

このインタビューでは、個人としてのクリエイター、小規模サイト構築のクリエイター、大企業のサイト構築を手がけているクリエイターに対して、下記の項目の質問を設定して聞き取りを行った。

- ① コンテンツを制作する際に気を付けていること
- ② コンテンツの動作・表示確認を行うブラウザ
- ③ コンテンツ作成時の留意点 (コスト、標準準拠等)
- ④ 上記に対して顧客からのニーズ、顧客の意識
- ⑤ コンテンツ検証に関するテスト方法
- ⑥ ユーザからのクレーム等の有無
- ⑦ ブラウザ毎に動作や表示が異なることに対する意識と意見
- ⑧ IE7 の登場で問題が出たかどうか
- ⑨ 今後の対応方針
- ⑩ 使用しているオーサリングツール

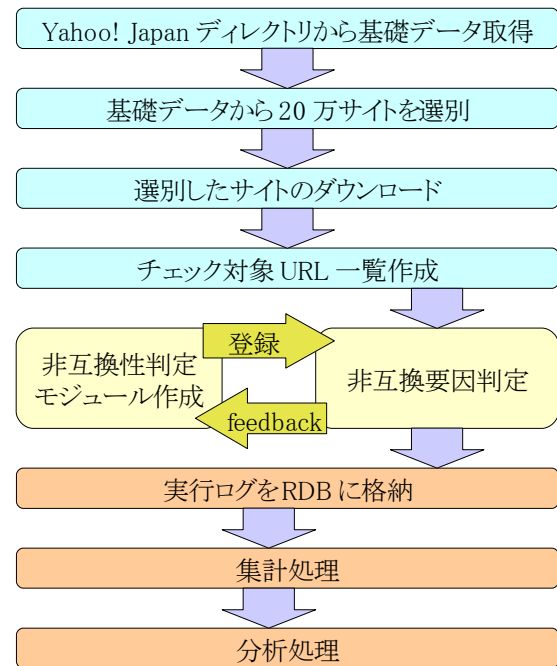


図 7 判定調査作業手順

インターネット調査およびイントラネット調査、双方のデータ分析において、オーサリングツールに起因する非互換要因、Web 標準の軽視等がそれぞれの調査結果上に顕著に表れていることから、コンテンツの供給者であり、また最も顧客に近いところにいる、クリエイターの立場を、インタビューを通して理解することが必要であると判断したからである。この調査は、個人経営を含む 3 社のクリエイターを対象として実施した。

3.6 まとめと提言

調査の結果は、特定プラットフォームに依存しないコンテンツの制作を推進するため、コンテンツクリエイターやオーサリングツールベンダーに対する指針の形で、提言書としてまとめることとした。

4 チェックツール

インターネット上にある非互換要因を含むコンテンツの実態を明らかにするため、非互換要因判定モジュールを連続適用する、図 8 に示すようなチェックツールの開発を行った。これは先に述べたように、Firefox の拡張機能として実現している。本ツールでは、判定モジュールを安全に実行するためのサンドボックスを持ち、そこに必要な API を提供する形をとった。また、判定モジュールの開発にも利用できるよう、各種の支援機能を提供している。チェックツールが持つ主な機能は、以下の通りである。

- ① コンソール機能
- ② 手動実行機能
- ③ 自動実行機能
- ④ モジュール管理機能

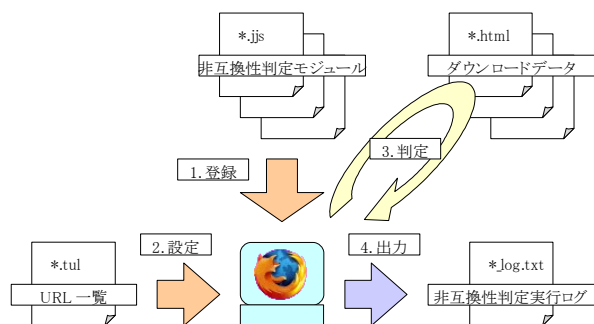


図 8 チェックツールの動作イメージ

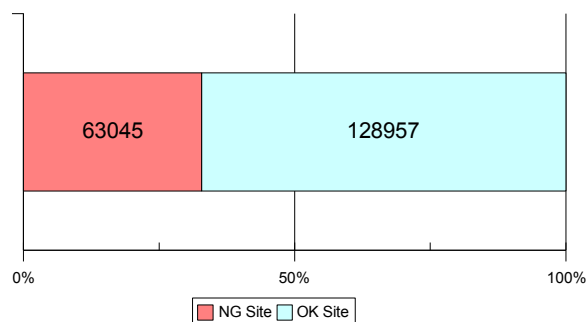


図 9 非互換要因検出率

- ⑤ ログ出力機能
- ⑥ スクリプト表示機能
- ⑦ スクリプト入力機能
- ⑧ ソースコード表示機能
- ⑨ 原イメージ表示機能
- ⑩ ウィルスチェック機能
- ⑪ モジュール支援 XPCOM API
- ⑫ 判定モジュール実行用サンドボックス

調査期間内の開発であったため、なにより時間的な制約があったが、Firefox が持つ XUL (XML ベースでユーザインタフェースを定義) と XPCOM (各種機能を提供するコンポーネント) を利用することで、試行錯誤しつつ開発を進めることができた。開発のベースに Firefox 2.0 が使えたことで、再起動の手間が減ったことも幸いした。

5 網羅的調査の結果

Web サイトは、ほとんどの場合、トップページからのリンクによる階層構造になっており、かつトップページに含まれるリンク数は、通常 10 よりもかなり多いと考えられる。したがって、仮に調査対象を 1 階層増やしただけで、処理時間は $<(200,000 + \text{各ページのリンク数の総和}) * 1 \text{ サイトあたりの処理時間}>$ に増大する。これは、20 万サイトの場合、数百万ページを扱う計算となる。1 ページの調査に 10 秒をかけた場合、1 年以上かかると想定される。さらに調査階層を増やした場合については論外である。このような理由から、調査対象サイトとして選別した 20 万サイトについて、トップページのみが今回の調査対象となっている。

調査対象サイトは、Yahoo! Japan カテゴリから取得したものであるため、既にサイトの変更や消滅などの理由で、調査時点で有効な URL でなくなってしまうものも含まれていると考えられた。この情報は実際にサイトにアクセスして初めて検出されるものであるため、仮に無効サイトの入れ換えを行った場合、新サイトがまた無効であった場合には入れ換え作業の繰り返しが発生することが予想された。そこで、この分の扱いについては、処理効率などを考慮した結果、無効サイト扱いとしてカウントすることとし、サイトの入れ換えは行

わないこととした。

また、調査対象サイトへのアクセスに対して、セキュリティ面の懸念があげられた。具体的には、サイトがウィルスに感染している場合や悪意を持ったスクリプトを含んでいる場合、調査作業になんらかの影響を与えることが危惧された。こうしたリスクを 100% 除外することは難しいが、あらかじめ検出可能であるものについては無効サイト扱いとし、調査対象から外していくという方針とした。このため、オンラインで利用可能なウィルスチェッカーを検討し、その機能のエッセンスをダウンロード処理に組み込むこととした。

5.1 最終チェック対象サイト数

調査対象 20 万サイトのうち、上記のダウンロード状況やチェック時の障害により、7,998 サイトを無効サイトとして除外した。うち、サイトやコンテンツ自体がすでに存在していない等の理由により除外されたものは 7,635 件であり、これは全体の 3.0% にあたる。また、ウィルスチェックの結果、陽性となったサイトは 9 サイトである。

最終的に、残りの 192,002 サイトが、非互換要因判定モジュールによるチェックの対象となった。

5.2 チェック対象非互換要因判定モジュール

172 種に及ぶ非互換要因を抽出したが、それらの中には自動的に検出することが困難であるなどの理由から、非互換要因判定モジュールの作成を行わなかったものがある。このため、クローリング調査に用いた非互換要因判定モジュールは 142 件となった。

5.3 非互換要因検出率

チェック対象となった 192,002 サイトを母数として、何らかの非互換要因が検出されたサイト数は 63,045 サイトであった (図 9)。検出率として見ると 32.8% となり、3 分の 1 近くのサイトが、今回作成したルールセットによって、何らかの非互換要因を含む可能性がある判断された。ただし、これらのコンテンツすべてが、何らかの非互換要因を含み、異なるブラウザでの閲覧に支障をきたすことを意味していることではないことに注

意が必要である。また、ひとつのコンテンツに複数の非互換要因が検出される場合もあり、検出された非互換要因の総数は、109,168となった。

5.4 非互換要因別検出件数

非互換要因別で見ると、検出数にかなりばらつきがある。かなりの検出数がある非互換要因の項目がある一方、今回の調査範囲ではまったく検出されなかった非互換要因も24種あった。検出件数の上位10件を表1に示す。

最も検出件数の多かったID 45は、コンテンツの表示位置に関するものであり、コンテンツ全体の位置がずれるという非互換要因である。また、5, 6, 8位のID 128, 75, 61は、枠線や水平線の色に関するものであり、IEとFirefoxでは色が異なる程度である。また、10位のID 47は、ページ内の画像にツールバーが表示されるIE独自の機能だが、右クリックのメニュー等で代用可能である。以上のことから、これらの非互換要因は閲覧に大きな影響は与えないと考えられる。

残りの項目には少々留意が必要である。2位のID 137は、テーブルのheight属性の変更のみをチェックすることが困難なため、JavaScript内に".height"が含まれるか否かで判別しており、目的外のコンテンツも検出されてしまっていると考えられる。

表1 検出件数の多い非互換要因(上位10件)

順位	ID	タイトル	検出件数
1	45	<body>タグの topmargin, rightmargin 属性	9,273
2	137	要素.height	8,169
3	89	color 指定の先頭の"#"	7,662
4	35	<marquee>タグ	5,654
5	128	<iframe>タグの border, bordercolor, allowtransparency 属性	4,931
6	75	<hr>タグに指定した color プロパティ	4,697
7	126	<frameset>タグの border, bordercolor, frameborder, framespacing 属性	3,659
8	61	<th><td>タグの bordercolor, bordercolorlight, bordercolordark 属性	3,571
9	77	text-decoration blink	2,993
10	47	<meta>タグの imagetoolbar 属性	2,851

表2 ブラウザ判別の利用状況

ID	タイトル	検出件数
107	ブラウザ判別	1,598
130	CSS 行頭の"*html"	303
131	CSS 行頭の"*html*"	4
132	CSS プロパティ直前の"_"	1

3, 4位のID 89, 35はチェックモジュールの不備と考えられる。前者はXHTMLの場合のみ差異が生じるが、チェックモジュールではHTMLも対象としていた。後者は<marquee>タグの一部の属性値に非互換要因が含まれているが、チェックモジュールではすべての<marquee>タグを対象としていた。

7位のID 126は、framespacing属性が与える影響が問題となる。framespacing属性はフレーム間の隙間を設定する属性だが、数ピクセル程度の小さな値が設定されることが多いと考えられ、この属性に対応しているか否かが、大きな影響を与えるとは考えにくい。

9位のID 77はIEが対応していないテキストの点滅に関するものである。TouchUpWebシステムの運用経験から、この属性はFirefoxに向けて作成されたサイトではなく、IE向けに作成したサイトに何らかの理由によって紛れ込んでしまっていることが多い。そのため、Firefoxで閲覧すると様々なテキストが点滅し、閲覧が困難になってしまうものである。

5.5 ブラウザ判別の利用状況

調査では、非互換の原因を明らかにすると同時に、非互換を回避しようとする取組み状況を調査するため、ブラウザに応じた動作・表示をさせるための仕組みの有無をチェックするモジュールも追加した。表2にその結果を示す。

ID 107のブラウザ判別は、JavaScriptの非互換に対処するために用いられる手法のひとつである。20万件の1%にも満たない検出件数ではあるが、クロスブラウザを考慮したコンテンツの存在を表している。また、JavaScriptを用いていないコンテンツ、単純なJavaScriptのみを利用しているコンテンツ、他の判別手法を用いているコンテンツがあることを考慮すると、実数はもう少し増えると考えられる。

一方、ID 130~132はCSSハックとも呼ばれ、スタイルシートをブラウザに応じて適用させるテクニックである。CSSハックは他にも様々な手法があるが、今回調査した手法は3手法あわせて300件強であり、あまり使われていない技術であることが判明した。

これらの結果から、明示的にクロスブラウザを考慮して作成されているコンテンツは全体の1%弱ということがわかった。ここからも、非互換問題に対する意識の低さを垣間見ることができる。

6 調査結果分析

6.1 非互換要因の傾向

非互換要因チェックで検出された要因をルールセットの分類に沿って集計し、非互換分類ごとの傾向を分析した(図10)。この結果、「IE独自のプロパティ」や「IE独自の属性」などIE独自の機能に関する非互換要因が最も多いことがわかった。次に多い要因は「標準の無視」によるものである。<meta>タグでの

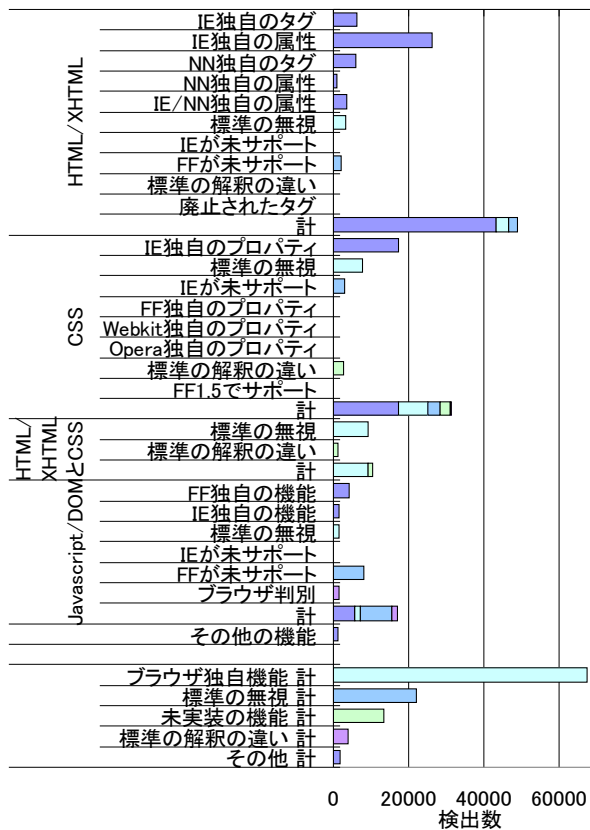


図 10 非互換要因の分布

Generator 記述とのクロス分析の結果、「標準の無視」の非互換要因を持つコンテンツには、特定のオーサリングツールを用いて作成されたコンテンツが多いことがわかった。この要因は本質的にはブラウザの実装が標準に沿っていないことが原因であるが、問題が顕在化したのはオーサリングツールが原因だといえよう。

6.2 影響度による分析

先に示したように、各非互換要因には「大」、「中」、「小」、「なし」、「対象外」の5段階で影響度を設定した。ここでは影響度ごとの検出数を表3に示す。

影響度が大きいために閲覧ができないような、影響度「大」の非互換要因は検出数の3.0%、レイアウトが崩れるような、影響度「中」の非互換要因は16.6%であり、比較的少数であることがわかった。IEとFirefoxで表示が異なっても、多くの場合その差異は小さいことが明らかとなった。

さらに、影響度をサイト数との関係で分析したのが表4である。コンテンツに含まれる非互換要因のうち、最大影響度が「大」のサイトは全サイトの1.70%である。検索エンジンで所望のページを検索し、多くのサイトを次々に閲覧するのが当然になっている近年の状況において、100サイトのうち1~2サイトの閲覧が困難であるということは大きな問題といえよう。

影響度が「大」の非互換要因は閲覧に対して致命的な影響を与えることから、検出数の多い5件を表5に

表 3 影響度ごとの非互換要因検出数

影響度	非互換要因数	検出数	全検出数に占める割合
大	14	3,309	3.0%
中	40	18,146	16.6%
小	51	69,169	63.4%
なし	13	1,178	1.1%
対象外	48	17,366	15.9%
合計	166	109,168	100.0%

表 4 サイトごとの最大影響度の集計(全サイト比)

最大影響度	サイト数	全サイト比
大	3265	1.70%
中	13943	7.26%
小	38498	20.05%
なし・対象外	7339	3.82%

表 5 影響度大の非互換要因(上位5件)

順位	ID	タイトル	検出数	全サイトに占める割合
1	80	z-index のマイナス値	1382	0.72%
2	98	VBScript	659	0.34%
3	88	タグや<a>タグの URL	414	0.22%
4	34	<object>タグの type 属性	350	0.18%
5	42	language 属性	250	0.13%

示すとともに、以下、詳細を説明する。

最も検出数の多かった「z-index のマイナス値」(ID 80)に関しては、Generator 記述とのクロス分析から IBM Homepage Builder によって生成されたコンテンツがほとんどを占めることがわかった。z-index 値にマイナスの値を設定する場合には、基準(ゼロ)となる値を設定しなくてはならないが、Homepage Builderはこの値を設定しない。このようなコンテンツをFirefoxで表示すると、z-index が指定された要素すべてが非表示となってしまふ。特定オーサリングツールの利用が、影響度大の非互換要因を、100コンテンツに1件弱の割合で発生させているというのは、注目に値する。

2位の「VBScript」(ID 98)は、IEのみが解釈できるVBScriptを用いたサイトである。Firefoxでも正常に閲覧するためにはスクリプトをJavaScriptで書き直す必要があり、大きな労力が必要となる。5位の「language属性」(ID 42)もスクリプトと関係しているが、この要因は必ずしも非互換をひきおこすとは限らない。language属性で指定された言語がJavaScriptであればFirefoxでも実行できるからである(正確には<meta>タグでの指定にも依存する)。

3位の「タグや<a>タグの URL」(ID 88)は、URLのディレクトリの区切り文字が Windows で使われる”¥”になっているために生じる問題である。Firefox では、タグであれば画像が表示されず、<a>タグであればクリックしてもリンク先へ移動することができない。

4位の「<object>タグの type 属性」(ID 34)は<object>タグに type 属性が設定されていないため、指定されたコンテンツが再生できないという問題が生じる。type 属性を設定する代わりに<object>タグの内部に<embed>タグを記述すれば、Firefox でも再生可能であるが、検出された 250 のサイトでは、このような処置も行われていない。

6.3 <meta>タグ上の Generator 記述含有率

調査では、非互換要因の混入に特定のオーサリングツールが関わっているかについて着目し、<meta>タグのうち Generator 記述について抽出を行った。その結果、約 3 分の 1 のサイト (33.4%) で Generator 記述が行われており、コンテンツ作成にオーサリングツールが活用されている様子が窺える。

記述されているオーサリングツールの製品、バージョンは実に多岐にわたっているが、ベンダー単位で見ただけの場合、IBM Homepage Builder が 44,707 件で全体の 68%以上を占めており、他を圧倒して利用されていることがわかる (表 6)。製品自体がそれだけのシェアを

表 6 Generator 記述数比較

Adobe 関連	6,151
IBM Homepage Builder 関連	44,707
Microsoft 社各種オーサリングツール関連	7,495
MSHTML 関連	1,143
その他	6,179
計	65,675

表 7 JPドメイン別検出率

ドメイン種別	非互換要因検出率	総検出数
AC.JP	0.397	3013
AD.JP	0.590	141
CO.JP	0.502	30379
ED.JP	0.460	866
GO.JP	0.304	1090
GR.JP	0.489	863
LG.JP	0.340	82
NE.JP	0.670	22011
OR.JP	0.514	8013
その他の JPドメイン	0.640	27522

持つと判断できるが、オーサリングツールごとの非互換要因に対する特性を知るには、この比率を考慮する必要がある。

6.4 Generator 記述サイトの非互換要因傾向性

前述のように、Generator 記述サイトにおける非互換要因の含有率を分析した結果、Generator 検出サイトの 33.8%に非互換要因が含まれていた。これは全サイトにおける非互換要因含有比率 (32.8%) とほぼ等しい。一方、1 サイトあたりの検出数は 1.19 件であり、全サイトの平均の 0.57 件と比較すると約 2 倍である。Generator 記述サイトでは 1 サイトに複数の非互換要因が含まれることが多いことがわかる。

個別の非互換要因ごとの傾向を知る上で、Generator 記述サイトは全体の 3 分の 1 であるので、その数字より検出比率が大きく変わるものは、オーサリングツール利用によって非互換要因 ID に対応する要因が埋め込まれやすい (または埋め込まれにくい) ことを示していると考えられる。すなわち、これらの非互換要因 ID に関しては、その非互換を生じる原因がオーサリングツールの利用に起因すると考えられるものであることを示している。

この傾向を数字の上で確認するために、母数となる全体の検出数との比率を求めた。ただし、サンプル数が少ないことによる誤差の影響を抑えるために、母数となる検出数が 1,000 以上の非互換要因 ID について、Generator 記述含有比率である 33.8%からの解離差を求め、20%以上の差があるものを図 11 としてまとめた。このような非互換要因 ID は 11 個あり、これらはオーサリングツールの利用の有無によって出現が大きく変動する可能性が考えられる。

6.5 JPドメイン別の非互換要因検出率

JPドメイン別の非互換要因検出率を表 7 に示す。教育機関 (AC.JP)、政府関連 (GO.JP)、地方自治体関連 (LG.JP) のサイトは、他のドメインと比較すると非互換要因検出率が低いことがわかる。政府や地方自治体関連のサイトで非互換要因が少ない理由として、アクセシビリティへの配慮が求められていることも挙げられるだろう。IE からの閲覧を前提とせず、Firefox や Mac の Safari ブラウザ、あるいは視覚障がい者のための音声ブラウザからのアクセスを考慮した結果であると推測される。

6.6 プロバイダ別の非互換要因検出率

今回の調査対象となったサイトには、URL から同一 ISP (Internet Service Provider) やポータルに所属すると判断できるサイトが含まれている。その多くは、ISP やポータルが提供するホームページサービスやブログサービスを利用している個人のページと考えられる。このようなサイトは、サービスプロバイダが提供するテンプレートなどを利用している可能性があるため、プロバ

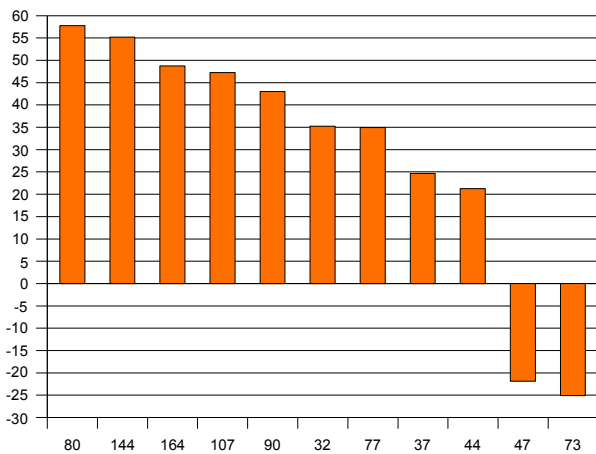


図 11 検出率解離率上位 ID

イダごとの非互換要因項目を集計することにした。これによって、各プロバイダが非互換要因を持つサイトをどれだけ含んでいるかがわかる。

まず、同一ドメイン中に 500 件以上のサイトを持つドメインを集計し、それを調査対象とした(表 8)。主力プロバイダ全体での非互換要因検出数は 21,753 件であり、非互換要因検出総数の 19.9%に相当する。

なお、Yahoo! Japan については、ホームページエリア提供事業の Yahoo!ジオシティーズを、ポータル事業である Yahoo! Japan とは別に集計した。ユーザが自由にホームページを作成できる Yahoo!ジオシティーズの非互換要因含有比率は 0.718 件と他プロバイダ並みであるが、ブログサービスなどあらかじめユーザカスタマイズ可能なテンプレートを用意しているポータル事業の方は含有比率が 0.191 件となり、ポータル事業に属するサイトについては非互換要因があまり含まれていないという結果となった。その他のプロバイダについては 0.5 件以上の含有率であることを考えると、サイトの提供テンプレートなどで Web 互換性を意識した対応がなされていると考えられる。このことから、提供するサービス内容によっても非互換要因の含有比率に差が出るのが推測される。

7 イン트라ネット調査

イントラネットを対象とした Web サイト実態調査に関しては、情報サービス企業 4 社と IPA のあわせて 5 組織を対象とした。

なお、これらのイントラネットに対する調査結果は、機微情報の取扱いの関係上、公開する成果物は具体的なシステム構成や画面例等を省略し、抽象度を高めたものとした。イントラネットの調査は人手による調査と、当事者である情報システム部に対するインタビュー調査により状況を分析した。

イントラネットを対象とした調査では、特定ブラウザに依存した Web システムの構築が行われている実態が明らかになった。いくつかの問題点に関しては原因と対策も明らかにでき、その中には明らかなバグである

にもかかわらず、IE では問題が顕在化しなかったためにバグとして認識されていなかった例も存在した。

システム管理者に対するインタビュー調査では、イントラシステム作成に関する実情やコスト意識、また IE のバージョンが 6.0 から 7.0 に移行するタイミングにあたり、慎重にならざるを得ない現状が明らかになった。同時に、このコンテンツの依存性を生み出す要因が、オーサリングツールだけでなく、イントラシステムを構成するパッケージにもあること、また、ベンダーの保証する環境が特定の Web ブラウザに限定されるために、運用環境でも対象を絞らざるを得ないこと等が浮き彫りとなった。

8 まとめ

今回の網羅的な調査によって、非互換問題を取り巻くさまざまな実態が浮き彫りとなった。たとえば、非互換発生状況と Generator 情報をクロス集計して分析した結果、特定のツールが非互換要因を生成しているケースが多いという現状が明らかになった。特に、表示されなかったり、対話的要素が動作しなかったりというような、影響度「大」のカテゴリに属する非互換要因に目を向けると、1,382 件となっている。このうち、最多の非互換要因として検出された z-Index に関する問題は、特定のオーサリングツールによって生成される傾向が非常に高いことが判明した。非互換要因の発生率が高くないとはいえ、特定のオーサリングツールがこのような状況を生み出している事実は無視できない。

一方で、ISP 別のクロス集計結果から、特定のテンプレ

表 8 プロバイダ別検出率

プロバイダ	サイトあたりの非互換要因検出率
Yahoo! ジオシティーズ	0.718
OCN	0.625
BIGLOBE	0.760
Infoseek	0.804
Yahoo! Japan	0.191
DION	0.646
So-net	0.606
ぷらら	0.741
ASAHI ネット	0.525
DTI	0.590
hi-ho	0.754
ODN	0.648
livedoor	0.638
J:COM	0.821
楽天	0.500
リムネット	0.581

レートを提供することが、非互換要因の発生を抑制することも示唆されている。最近では、良質なテンプレートやライブラリが充実し、スクラッチからコンテンツを作成する機会も減る傾向にある。したがって、これらのテンプレートが、ツールの不具合をカバーしていくことを期待したい。また近年の動向として、一般企業や個人サイトにおいても、コンテンツ管理システム(CMS)やブログツールといった、プロバイダ側で用意するフォーマットに従った情報提供形態が増えてきている。これらが非互換要因を含まないコンテンツを提供するという前提に立てば、これらの普及は互換性の向上に非常に効果的であろう。

政府系機関や地方自治体のサイトでは、アクセシビリティへの配慮が非互換要因の発生を抑えている可能性が高いという一方で、一般企業や個人サイトは、非互換要因を含む確率が比較的高いという結果が得られている。これは、これらのサイトでは問題を含むツールに頼ることが多いと考えられるほか、コストに対してのみ意識が集中する傾向にあることも、システム管理者やクリエイターに対するインタビュー結果から導かれている。コンテンツ発注者に対しては、中途半端なコスト意識で特定のブラウザのみに対応するという姿勢は、長期的にはコスト増を招く可能性がある、という啓蒙を行う必要がある。また、そもそも特定のブラウザに対するシェアの問題から、自分が利用しているブラウザ以外を知らない、ひいてはコンテンツの非互換問題が世の中に存在することを認識せずにコンテンツを制作する、あるいはコンテンツクリエイターに制作を発注するユーザもいる。これらのユーザに対しても、啓蒙の機会を与えることが望ましい。

また、クリエイターに対するインタビューの結果から、特にクリエイターの側から Web コンテンツの特定ブラウザ依存を考えた場合、現在の日本では欧米製商用オーサリングツール(特定ベンダー)への依存性が極めて高く、これによって生成されたコンテンツに含まれる非互換性要因も多い。このことから、OSSによるオーサリングツールの開発を推進することも、依存性を解決するための方策のひとつとして検討されるべきであろう。

これらの調査の結果に基づき、我々は、特定プラットフォームに依存しないコンテンツの制作を推進するため、コンテンツクリエイターやコンテンツオーサリングツールベンダーに対する指針として、提言書をまとめた。その中で、以下の5項目の提言を行っている。

- ① 古いバージョンのオーサリングツールに対するフォローをすべし
- ② クロスブラウザ対応の利点を意識すべし
- ③ 最低限のチェック環境を用意すべし、また最低限のチェックを行うべし
- ④ 適材適所のコンテンツ表現手段を選択すべし
- ⑤ Web ブラウザはすべからず標準に従った実装

を目指すべし

今回の調査の成果が有効に活用され、OSS デスクトップ普及の一助となることを期待している。

謝辞

本研究成果は、独立行政法人情報処理推進機構オープンソースソフトウェア活用基盤整備事業の成果として纏められたものである。本プロジェクト遂行を支援する全ての関係者に感謝します。

参考文献

- [1] Netcraft, November 2006 Web Server Survey, http://news.netcraft.com/archives/2006/11/01/november_2006_web_server_survey.html
- [2] The Web Standards Project, <http://www.webstandards.org/>
- [3] もじら組, Web 標準化レポート, <http://bugzilla.mozilla.gr.jp/>
- [4] もじら組, 日本トップウェブサイトテストプロジェクト報告書, <http://www.mozilla.gr.jp/webtest/finalreport.html>
- [5] Mozilla, Firefox 「問題のあるウェブサイトを報告」DB, <http://reporter.firefox.org/>
- [6] TouchUpWeb, 「ウェブコンテンツの相互運用性を改善するフィルタリングサーバの開発」, <http://www.touchupweb.org/>
- [7] 大藤, 詳解 HTML&XHTML&CSS 辞典, 秋和システム, 2002.
- [8] 佐藤, 必携 HTML/CSS/JavaScript Web ブラウザー互換性辞典, インプレス, 2003.
- [9] ユニゾン, 超図解 HTML&CSS 辞典, エクスペディア, 2006.
- [10] De Antonellis, V., Melchiori, M., and Plebani, P., An approach to Web Service compatibility in cooperative processes, Proc. of SAINT 2003 Workshops, pp.95-100, Jan, 2003.
- [11] Xu, L., Xu, B., Nie, C., Chen, H., and Yang, H., A Browser Compatibility Testing Method Based on Combinatorial Testing, Lecture Notes in Computer Science, Springer Berlin / Heidelberg, Vol.2722/2003. pp.310-313, Feb. 2004.
- [12] Peter, K., Government web standards usage: USA-Standards Schmandards, <http://www.standards-schmandards.com/2005/government-web-standards-usage-usa/>
- [13] Andrew, R., Web Design Principles and Practices, <http://www.rickmann-design.co.uk/collections/principles>
- [14] IPA, 「OSS デスクトップの普及に資する Web コンテンツ互換性向上に関する調査」調査報告書, http://www.ipa.go.jp/software/open/osscc/download/Web_Research_Ja.pdf, 2007.
- [15] IPA, 「OSS デスクトップの普及に資する Web コンテンツ互換性向上に関する調査」非互換問題改善提言書, http://www.ipa.go.jp/software/open/osscc/download/Web_Recommendations_Ja.pdf, 2007