

アクセス解析の集計と用語定義ガイドライン

第1版

アクセス解析イニシアチブ
2009年10月27日

著作権について

著作権は「アクセス解析イニシアチブ LLC」にあります。クリエイティブ・コモンズで規定している、表示（著作権表示を求めます）非営利（非営利目的に限って使用を認めます）改変禁止（そのままの形でのみ利用を認める）の範囲でご利用下さい。つまり配布にあたってコピー代等を含めた金銭の授受を伴わないかたちで、出典を明記して頂ければ、全文掲載の形での利用を許諾します。また、雑誌等商用媒体への全文掲載にあたっては「アクセス解析イニシアチブ LLC」までお問い合わせください。

1. データ収集方法と集計方法、定義の重要性

アクセス解析に用いられるウェブサイトの利用データは、メディア・リサーチ・データの一つである。

リサーチ・データは、出てきた結果を正しく読むために、通常調査対象を選ぶサンプリング手法や質問文と回答の選択肢などの質問内容、サンプル数と回答者数や回収率、そして集計方法を明らかにする。程度の差はあれ、まともな調査であれば、調査結果の数字を公表する際に、最低限の調査手法などの情報を付けることが一般的だ。

アクセス解析も上記と同様のことが必要である。それではアクセス解析のケースでは、何を最低限明らかにしなければならないのだろうか。まず次のことを明確にしないとデータを正しく利用できない。

A. データ収集方法の明示

これは何をどう収集しているのかということである。どういうデータをどのような仕組みによって取得しているかという方法を明らかにする必要がある。なぜならデータ収集方法が違えば、それだけで同じサイトでも違う数字が出てくるからだ。

B. 集計方法の明示

同じ指標でもツールによって集計方法が違うこともありうる。どれが正しいということではなく、どのようなバリエーションがあるのか、それはどういう計算式によって算出されているのかを明示する必要がある。

C. 使う言葉の意味の明示

同じ言葉でも違う意味で使っているケースもあるし、同じ意味なのに違う言葉を使っている場合もある。どれが正しいということではなく、どのようなバリエーションがあるのか、それはどういう意味なのかといったこと明示する必要がある。

この三つを明示することにより、サービス提供側の最低品質保証のレベルの確保、そして利用者側の最低限の理解の確保が可能になる。

2. データ収集方法の明示

ウェブサイトの利用状況データは、様々な手法によって取得することができる。以下は代表的なデータ収集手法だが、どのような手法によってデータを収集したかを明らかにすることで、収集方法の特性に応じた正しい解釈を促進することになる。利用したツール名まで明記する必要はないが、最低限下記のような分類レベルで明示するのが望ましい。

A. サーバーログを利用する方法

内容：ウェブサーバーの機能を使って収集されるサーバーログを利用するもの

特徴：ウェブサーバーへのリクエストが記録されるので、検索エンジンのクローラーなど、人のウェブサイトの閲覧行動以外にも含めた全ての要求が記録される。そのため人の閲覧行動だけを集計したい場合は、それらのデータを除いて集計することが必要である。また戻るボタンによる閲覧など、キャッシュから読み込まれ、ウェブサーバーにリクエストがいかないクライアント側の閲覧行動は取得できない

注意：キャッシュからの読み込みデータは取得できないので、閲覧経路の正確な把握などで精度が低くなる可能性がある。また、ブラウザのユーザーエージェントを詐称するクローラーのデータが混在する可能性がある。

B. JavaScript を利用して、ブラウザ側の行動データを取得する方法

内容：JavaScript の計測タグを使って、ブラウザのデータを収集するもの

特徴：検索エンジンのクローラーなどは基本的に JavaScript を無視するため、人のブラウザによるウェブサイトの閲覧行動だけを取得することになる

注意：検索エンジンのクローラーなどの挙動を知りたい、といったニーズには応えられない。また、セキュリティその他の理由からブラウザで JavaScript の動作をオフにしているごく一部の閲覧者の行動は取得されない。

C. ウェブサイトの直前を流れるパケットデータの一部を取得する方法

内容：ウェブサーバーを出入りするパケットデータの一部を利用するもの

特徴：サーバーログと同様、サーバー側でのデータ取得に特有の特徴に加え、サーバーログでは記録されない、post メソッドによるリクエスト情報の収集も可能

注意：キャッシュからの読み込みデータは取得できないので、閲覧経路の正確な把握などで精度が低くなる可能性がある

D. 外部サービスのデータを使う場合

上記AからCまでとは区分のレベルが違うが、レンタルサーバーやプログサービスなどでサービス提供側から提供される数値が、上記などのどの手法によって取得されたデータなのか、自分のアクセスを含むものなのか排除されるものなのかなどについて明示するのが望ましい。

これ以外のデータ収集方法もある。その場合も上記のように、それはどのような方法なのかということ、それに伴ってデータを読む場合に気をつけなければいけない特徴や注意ということも明らかにすることが望ましい。

ウェブサイトのアクセスデータは、同じサイトのデータでも別の手法を使えば違う数字が出てくる。つまり大事なことは、どの数字が正しいということではなく、それぞれの特性が違うので、それを理解して使うということ。また取得可能なデータの精度も 100% 確実なものはない。しかしだからこそ、どういった方法で取得したデータなのかということを知り、そのデータの特性を理解して、その範囲の中で活用することが必要になる。例えば瞬間の絶対値を見るのではなくトレンドで数字を追うといった視点である。

3. 主要指標の定義

A. 「ページビュー」と「ページビュー数」(単位: ページ)

ページビュー数をカウントするためには、何を「ページビュー」と定義するかが明示されていなければならない。一般的にページビューという場合の「ページ」は、一つの html ファイルを示す。つまり一つのユニークな URL で判別できるページを表す。この一つ一つの URL がどれだけ見られたかをカウントした「ページビュー数」というのを基本的な定義とする。また人によって能動的にリクエストされたものをカウントするのが望ましい。

もちろんこの原則に当てはまらないページが多くある。下記のケースを参考にしながら、それぞれの計測における「ページ」の定義を明確にすることを推奨する。

1. フレームページの構成の場合

親フレームと子フレームの全てをページとしてカウントすると、ユーザーが実質的に 1 ページ見ているのにも関わらず、これらを構成する全ての html ファイル数をカウントすることになるので注意が必要である。

2. Ajax や Flashなどで作成されたページの場合

ユーザー側の行動としては、クリックをして次のページを見ているような意識にも関わらず、リクエストがサーバーに飛ばないといったページの作りをしているものが多数ある。普通この場合は、その 1「クリック」イベントに対して、ページビューは発生しないが、それらに 1 ページビューを割り当てるという考え方は合理的である。JavaScript 型の計測をしている場合は、同様の計測コードを実装することによって、計測したい「イベント」を 1 ページ相当としてカウントしてもよいだろう。

3. リダイレクトページの場合

リダイレクトは広告効果の測定のため、あるいはサイトの引越のため、外部リンクのクリック計測のためなど、様々な利用方法がある。閲覧する中身のないリダイレクトに関しては、カウントから外すことを推奨する。人によって能動的にリクエストされたものでないからだ。

その他にも URL が異なるのにコンテンツが全く同じ(セッション ID や計測用のダミーパラメータが付与されているなど) ケースや、逆に同じ URL なのにコンテンツが違う (post メソッドを利用している場合や、パーソナライズなどのカスタマイズがされているような場合など) ケースがある。この場合に URL 別のページビュー数をどうカウントするのかについて、明記しておくのが望ましい。また上記とはレベルが異なるが、RSS フィードを読むといったコンテンツの閲覧方法もあるが、ここでは扱わない。

B. 「訪問」と「のべ訪問者数」(単位：人もしくは回)

まず「訪問」の定義をする。「訪問」とはウェブサイトの閲覧において、一連のまとまった閲覧行動を意味する。ウェブサイトの閲覧行動において、30分を超える間隔があいたら別の利用行動、別の「訪問」と見なそうということになっている。「訪問」という言葉の他に、「ビジット」「セッション」といった名前も同じ意味合いを持つ。

30分というのが事実上の世界標準となっているが、ツールによっては、30分の代わりに15分などに変更して集計することも可能である。そのサイトで「訪問」を時間間隔の視点からどう定義するかは、サイトの特性にもよるので、ツールによっては可変としているものもある。しかしその都度恣意的に変化させるのはおかしいし、比較もしにくくなるので、こうした標準からあまり動かさないのがよいだろう。

「のべ訪問者数」とは、特定期間におけるサイトへの総「訪問」回数を意味する。Cで解説する特定期間における同一人物の重複を省く「ユニーク訪問者数」と区別をすることが必要なので、指標名としては、「のべ訪問者数」と「ユニーク訪問者数」を第1候補として記載した。「訪問者」(あるいは「ユーザー」「ビジター」など)というのが紛らわしい場合は、「のべ訪問回数」と表現するとよいだろう。いづれにしても、「訪問」の言葉のバリエーションに伴い、特定期間におけるサイトへの総訪問回数も様々な表現方法があるので、どの言葉で表現するのかを明示しよう。

ではこの「のべ訪問者数」は通常どのように集計するかというと、まず同一人物(どう判定するかは項目Cを参照)のデータごとに利用開始時間のデータを並べる。これがその人のサイトの時系列の閲覧行動と考える。次に各ページの閲覧開始時間の差が30分を超えているかいないかをチェックし、超えていたら別の「訪問」としてカウントしていく。

例：下記のケースは、計測対象サイトにおけるある人のページ閲覧記録である。各ページの閲覧開始時間が次のようになっていたとする。ページBとCの時間間隔が30分を超えているので、ここを境として別の訪問だとみなす。つまりこの例では「のべ訪問者数」が2とカウントする。

閲覧ページ	閲覧開始時間	次ページとの閲覧間隔
計測対象サイト内のページA	12:00	20分
計測対象サイト内のページB	12:20	40分 --30分を超えるため、ここで区切る
計測対象サイト内のページC	13:00	10分
計測対象サイト内のページD	13:10	不明

但しツールによっては、一旦サイトの外部へ出たことがわかっている場合は、計測対象サイト内の閲覧行動が30分以内であっても、別の「訪問」とカウントするケースもある。下記のケースでページBとCの閲覧間隔は10分だが、ページCの参照元に外部サイトが記録されていることから一旦外部サイトへ出て行ったと想定されるので、このサイト内では別の一連の閲覧行動として区別しようという考え方だ。これも一つの考え方として間違っていない。

閲覧ページ	閲覧開始時間	次ページとの閲覧間隔
計測対象サイト内のページA	12:00	20分
計測対象サイト内のページB	12:20	10分
外部サイトのページを閲覧		
計測対象サイト内のページC	12:30	10分
計測対象サイト内のページD	12:40	不明

また集計には「日」「週」「月」といった単位があるので、その境でセッションを集計し一旦切ってしまうのか、継続性を保つのかなどに関しては、ツールによって異なる。日を跨ぐ「訪問」を日別の集計ではどのように行っているのか、上記例のように一旦外部に出た時に「訪問」を繋げるのか切ることかといった繋ぎ方など、細かい点に関しては、各ツールベンダーに確認すること。

C. 「ユニーク訪問者」と「ユニーク訪問者数」(単位:人)

「ユニーク訪問者」の他に「ユニークユーザー」「ユニークビジター」「ユニークブラウザ」などとも表現する。ユニークという名称がついている通り、同じ人が何回訪問しても一人とカウントする指標である。レポート対象期間の「日」「週」「月」といった集計の長さにおいて、ユニークなサイト利用者が何人いたかということを表示する。

下記で「のべ訪問者数」と「ユニーク訪問者数」がどうカウントされるかを見ていく。 その日における1「訪問」を表示しているものとする。この他に6月、7月での訪問はなかったとする。

	6/30	7/1	7/2
Aさんの訪問	1	1	1
Bさんの訪問	1	1	1
Cさんの訪問	1	1	1
Dさんの訪問	1	1	1

6月月次の「のべ訪問者数」はこの期間の の数に相当するので3。6月の「ユニーク訪問者数」はAさんとCさんの二人なので2となる。7月月次の「のべ訪問者数」はこの期間の の数に相当するので6。7月の「ユニーク訪問者数」は4となる。7月1日の日別データでは「のべ訪問者数」が4で「ユニーク訪問者数」が3。7月2日の日別データでは「のべ訪問者数」が2で「ユニーク訪問者数」が2。「のべ訪問者数」は日別データを足し算すれば月次データと一致するが、「ユニーク訪問者数」は日別の足し上げは月次データとは一致しない。

ではこの同一人物の利用行動であるという判別はどのように行うかという点だが、各種方法があるので、詳細はツールベンダーに確認すること。いくつかの例を挙げておく。

1.クッキーによる判定

ウェブサイトの利用時にクッキーを付与し、それをアクセスデータと一緒に取得することで、同じブラウザからの閲覧であるとみなす方法である。

注意点としては、クッキーを拒否する設定のブラウザからのデータはカウントできない。また定期的にクッキーを削除するようなユーザーの場合は、その都度新規の別の訪問者とみなされてしまう。また同じ人が幾つものパソコンやブラウザを利用していれば、別々の訪問者とカウントされる。

またブラウザによっては、新しいタブで表示した場合や新しいブラウザで表示させた場合などで、クッキーの受け渡し方が異なる場合もある。

2.IPアドレスとユーザーエージェントの組み合わせによる判定

閲覧しているユーザーのIPアドレスと、ユーザーエージェント(OSとブラウザの組み合わせ)の組み合わせでユニークを区別しようという方法だ。注意点としては、企業などで同じIPアドレスから同じユーザーエージェント(企業内であればOSとブラウザのバージョンなども全く同じことが多いと考えられる)の閲覧行動が全て同一ユーザーとみなされる可能性が高くなる。またこちらも同じ人が幾つものパソコンやブラウザを利用していれば、別々の訪問者とカウントされる。

厳密には同一人物でも、ユニークデバイス(パソコン、携帯、IPテレビなど同じ人でもサイトに接続するデバイスは各種あるということ)、ユニークPC(一人で何台もパソコンは使っているということ)、ユニークブラウザ(1台のパソコンでも複数のブラウザを使っているということ)、ユニーククッキー(一つのブラウザでも複数のクッキーが発行されているということ)というような階層がある。そこまで厳密な議論はここではしないが、計測方法や利用実態によっては、このように現実との乖離があるということを知っておくこと。会員サイトでユニークIDによって同一人物の特定がより正確にできる場合でも、IDを共有しているかもしれないということを理解しておこう。

D. 「ページ滞在時間」と「訪問滞在時間」(単位：時分秒)

滞在時間の指標として、二つ挙げておく。「ページ滞在時間」とは各ページ別の滞在時間を表すもので、「訪問滞在時間」は1訪問全体の滞在時間を表すもの。

通常各ページの滞在時間は、該当ページと次のページの閲覧開始時間の差をもって、該当ページの閲覧時間とする。この方法だと、各「訪問」の最後の閲覧ページは次のページがないので、閲覧時間は算出できない。もちろんもっと精度の高い時間計測が可能であれば、それがどのような手法で取得しているものであるかを明示すればよいだろう。最後の閲覧ページの滞在時間の取り扱いの例としては、つぎのような場合が考えられる。

訪問滞在時間では、この不明の最後のページの閲覧時間を0秒や1分などと決めて計算をすることになる。

ページ滞在時間では、該当ページの滞在時間の平均値を算出するときに、この不明の最後のページの閲覧時間を0秒や1分などとして集計に含める場合(分母のページビュー数に1カウントし、分子の滞在時間で0秒や1分を加える)もあれば、母数に含めない場合(分母のページビュー数に加えず、分子にも加えない)もある。直帰率が非常に高いページの場合、0秒として集計に含めると実態とかけ離れることになるので、最後のページに関してはページ滞在時間の集計対象から外すのが望ましい。

例：下記のケースは、あるサイトにおけるある人のページ閲覧記録だ。各ページの閲覧開始時間が次のようになっていたとする。各ページの「ページ滞在時間」と「訪問滞在時間」を下記に表す。

閲覧ページ	閲覧開始時間	次ページとの閲覧間隔
計測対象サイト内のページA	12:00	20分
計測対象サイト内のページB	12:20	10分
計測対象サイト内のページC	12:30	15分
計測対象サイト内のページD	12:45	不明

ページ滞在時間：

ページAの滞在時間 20分

ページBの滞在時間 10分

ページCの滞在時間 15分

ページDの滞在時間 0秒あるいは1分など当初決めたルール、あるいはデータなし(ページビュー数の指標では1とカウントするが、滞在時間データの算出の分母のページビュー数に加えず、分子にも加えないということ)

訪問滞在時間：45分(20分+10分+15分)あるいは最後のページを1分と決めている場合は、それも加えて46分。

滞在時間は、利用シーンと一緒に考慮すべき指標である。例えばナビゲーション目的でしか作られていないトップページの場合は、早く次のページへ誘導することが望ましく、ページ滞在時間は短いほどよいといった考え方をする。入力フォームであれば、ページ滞在時間が長いとユーザーが迷っていると考えるよいだろう。

しかし読ませるコンテンツとナビゲーションが混在しているページなどの場合は、ページ滞在時間が長い方がよいのか短い方がよいのか何とも評価できない。滞在時間の長短が評価軸として比較的明確な場合に使うのが有効だろう。

以上