

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5072160号
(P5072160)

(45) 発行日 平成24年11月14日(2012.11.14)

(24) 登録日 平成24年8月31日(2012.8.31)

(51) Int. Cl.	F I	
G06F 13/00	(2006.01)	G06F 13/00 54OR
G06F 17/30	(2006.01)	G06F 17/30 11OF
G06Q 30/02	(2012.01)	G06F 17/30 13OA
G06Q 10/00	(2012.01)	G06F 17/60 17OA
G06Q 50/00	(2012.01)	G06F 17/60 326

請求項の数 18 (全 33 頁) 最終頁に続く

(21) 出願番号	特願2001-552567 (P2001-552567)	(73) 特許権者	503416113
(86) (22) 出願日	平成12年10月25日 (2000.10.25)		ネットレイティングス・インコーポレーテッド
(65) 公表番号	特表2004-504649 (P2004-504649A)		アメリカ合衆国ニューヨーク州ニューヨーク、120ウェスト45ストリート35階
(43) 公表日	平成16年2月12日 (2004.2.12)	(74) 代理人	100110412
(86) 国際出願番号	PCT/US2000/029352		弁理士 藤元 亮輔
(87) 国際公開番号	W02001/052462	(72) 発明者	グレゴリー・ジェイ・ロークハート
(87) 国際公開日	平成13年7月19日 (2001.7.19)		アメリカ合衆国ワシントン州シアトル、エヌイー、39アベニュー13200
審査請求日	平成19年9月26日 (2007.9.26)	(72) 発明者	クレイグ・ビー・ホーマン
(31) 優先権主張番号	60/175,665		アメリカ合衆国ワシントン州シアトル、ガーラーストリート201
(32) 優先日	平成12年1月12日 (2000.1.12)	(72) 発明者	クリスタ・エル・コロル
(33) 優先権主張国	米国 (US)		アメリカ合衆国ワシントン州シアトル
(31) 優先権主張番号	60/231,195		最終頁に続く
(32) 優先日	平成12年9月7日 (2000.9.7)		
(33) 優先権主張国	米国 (US)		

(54) 【発明の名称】 ワールドワイドウェブのデジタルコンテンツの普及を見積もるシステム及び方法

(57) 【特許請求の範囲】

【請求項1】

ネットワーク上のデジタルコンテンツの普及を見積もる方法であって、
コンテンツオブジェクトが所定の頻度で変化するウェブページがアクセスされた回数の見積もりを取得するステップと、

前記ウェブページを要求し、それに応じてコンテンツファイルを取得することを繰り返すステップと、

前記繰り返すステップによって得られた前記コンテンツファイルに第1のコンテンツオブジェクトが含まれている数を求めるステップと、

前記ウェブページが要求された合計回数を求めるステップと、

前記繰り返すステップによって得られた前記コンテンツファイルに前記第1のコンテンツオブジェクトが含まれている数と前記ウェブページが要求された合計回数と前記ウェブページがアクセスされた回数の見積もりとに基づいて、前記第1のコンテンツオブジェクトが前記ウェブページの訪問者に提供された回数を見積もるステップを有する方法。

【請求項2】

前記ウェブページがアクセスされた回数を見積もりの少なくとも一部は、プロキシから取得する請求項1に記載の方法。

【請求項3】

自動広告普及システムによって行われる請求項1に記載の方法。

【請求項4】

前記ウェブページがアクセスされた回数の見積もりの少なくとも一部は、パネリストコンピュータから取得する請求項 1 に記載の方法。

【請求項 5】

前記コンテンツオブジェクトは広告である請求項 1 に記載の方法。

【請求項 6】

前記第 1 のコンテンツオブジェクトが訪問者に提供された回数を見積もるステップは、
前記ウェブページが要求された合計回数を、前記繰り返すステップによって得られた前記コンテンツファイルに前記第 1 のコンテンツオブジェクトが含まれた回数で割ることで
前記コンテンツオブジェクトの回転率を求めるステップと、

前記ウェブページがアクセスされた回数の見積もりに前記回転率を掛けることで前記第 1 のコンテンツオブジェクトが訪問者に提供された回数を求めるステップとを含む請求項 1 に記載の方法。

10

【請求項 7】

コンテンツオブジェクトが所定の頻度で変化するウェブページがアクセスされた回数を見積もる見積もり手段と、

前記ウェブページを要求し、それに応じてコンテンツファイルを取得することを繰り返すプロパーと、

前記繰り返すステップによって得られた前記コンテンツファイルに第 1 のコンテンツオブジェクトが含まれている数を求め、前記ウェブページが要求された合計回数を求め、前記繰り返すステップによって得られた前記コンテンツファイルに前記第 1 のコンテンツオブジェクトが含まれている数と前記ウェブページが要求された合計回数と前記ウェブページがアクセスされた回数の見積もりとに基づいて、前記第 1 のコンテンツオブジェクトが前記ウェブページの訪問者に提供された回数を見積もる統計要約システムとを有するシステム。

20

【請求項 8】

前記ウェブページがアクセスされた回数の少なくとも一部は、プロキシから取得する請求項 7 に記載のシステム。

【請求項 9】

自動広告普及システムによって行われる請求項 7 に記載のシステム。

【請求項 10】

前記ウェブページがアクセスされた回数の見積もりの少なくとも一部は、パネリストコンピュータから取得する請求項 7 に記載のシステム。

30

【請求項 11】

前記コンテンツオブジェクトは広告である請求項 7 に記載のシステム。

【請求項 12】

前記統計要約システムは、

前記ウェブページが要求された合計回数を、前記繰り返すステップによって得られた前記コンテンツファイルに前記第 1 のコンテンツオブジェクトが含まれる数で割ることで前記コンテンツオブジェクトの回転率を求め、前記ウェブページがアクセスされた回数の見積もりに前記回転率を掛けることで、前記第 1 のコンテンツオブジェクトが訪問者に提供された回数を求めることによって、

40

前記第 1 のコンテンツオブジェクトが訪問者に提供された回数を見積もる請求項 7 に記載のシステム。

【請求項 13】

機械が読み取り可能な記録媒体であって、

実行されたときに少なくとも、

前記コンテンツオブジェクトが所定の頻度で変化するウェブページがアクセスされた回数の見積もりを取得するステップと、

前記ウェブページを要求し、それに応じてコンテンツファイルを取得することを繰り返すステップと、

50

前記繰り返すステップによって得られた前記コンテンツファイルに第1のコンテンツオブジェクトが含まれている数を求めるステップと、

前記ウェブページが要求された合計回数を求めるステップと、

前記繰り返すステップによって得られた前記コンテンツファイルに前記第1のコンテンツオブジェクトが含まれている回数と前記ウェブページが要求された合計回数と前記ウェブページがアクセスされた回数の見積もりとに基づいて、前記第1のコンテンツオブジェクトが前記ウェブページの訪問者に提供された回数を見積もるステップと、

を機械に行わせる指示を記録した機械が読み取り可能な記録媒体。

【請求項14】

前記ウェブページがアクセスされた回数を見積もりの少なくとも一部は、プロキシから取得する請求項13に記載の機械が読み取り可能な記録媒体。

10

【請求項15】

自動広告普及システムによって、前記機械が読み取り可能な記録媒体に記録された指示が実行される請求項13に記載の機械が読み取り可能な記録媒体。

【請求項16】

前記ウェブページがアクセスされた回数を見積もりの少なくとも一部は、少なくとも一つのパネリストコンピュータから取得する請求項13に記載の機械が読み取り可能な記録媒体。

【請求項17】

前記コンテンツオブジェクトは広告である請求項13に記載の機械が読み取り可能な記録媒体。

20

【請求項18】

前記機械が読み取り可能な記録媒体に記録された指示は、

前記ウェブページが要求された合計回数を、前記繰り返すステップによって得られた前記コンテンツファイルに前記第1のコンテンツオブジェクトが含まれる数で割ることで前記コンテンツオブジェクトの回転率を求め、

前記ウェブページがアクセスされた回数を見積もりに前記回転率を掛けることで前記第1のコンテンツオブジェクトが訪問者に提供された回数を求めることによって、

前記第1のコンテンツオブジェクトが訪問者に提供された回数を見積もる請求項13に記載の機械が読み取り可能な記録媒体。

30

【発明の詳細な説明】

【0001】

関連出願の相互参照

本出願は、参照して結合される、2000年1月12日に米国特許商標庁に提出された特許証用仮出願番号60/175,665号と、2000年9月7日に米国特許商標庁に提出された特許証用仮出願番号60/231,195号の優先権を主張する。

【0002】

発明の分野

本発明は、一般に、インターネット等のコンピュータネットワークで配布されるデジタルコンテンツを追跡及び測定するシステム、方法及びコンピュータプログラムプロダクトに関する。特に、本発明は、それらの広告の普及の測定を計算するために、オンライン広告データを収集し、データを解析し、データを使用するシステム、方法及びコンピュータプログラムプロダクトに関する。

40

【0003】

発明の背景

インターネットとワールドワイドウェブ(「ウェブ」)の人気の増加は、部分的には、ウェブページが採用可能なインタラクティブ技術に依存する。これらのインタラクティブ技術は、当該技術が、固定されたアイコンのスポンサー広告、変化するバナーとボタン、侵入型の広告(即ち、ユーザの作業を中断させてスクリーンディスプレイの相当の割合を占領するオンライン広告)など、新しい広告様式を導入したため、直接に宣伝媒体としての

50

ウェブに影響を与える。広告の作成が異なるとしても、見る者に対する影響は伝統的広告と同様である。例えば、ウェブページ上のバナー広告又はロゴアイコンは見る者に対して製品の印象を形成するが、それは商標名やスローガンを提供することによって、製品を宣伝する伝統的なビルボード広告に等価である。同様に、ウェブページ上のスポンサーのロゴは見る者に対してスポンサーの印象を形成するが、それは大学の野球の試合でスコアボードのスポンサーのロゴを見るのと等価である。

【 0 0 0 4 】

過去数年に亘るインターネットの急速で爆発的成長は、その拡張の大きさと速度を定量化する品質統計に対する高い需要を形成してきた。幾つかの伝統的な測定方法は、インターネットとそのユーザについて有用な統計を生み出しているが、インターネットの複雑性により、これらの方法の幾つかは多くの重要な問題に 대응することができないままになっている。

10

【 0 0 0 5 】

オンライン広告は、伝統的方法が十分に測定に力を貸さない一分野である。毎日、何千もの電子広告が数百万のウェブページから現れては消える。オンライン広告活動の一時的性質は、広告活動を正確に測定するための新規な方法を保証する。

【 0 0 0 6 】

現存の広告追跡評価システムは、ウェブページの収集を自動化しているが、オンライン広告の収集を自動化することはできない。オンライン広告の内容は経時的に変化又は変動するので、特定広告の頻度の正確な再現は、正しい割合で関連ウェブページの継続的なサンプリングをすることを必要とする。更に、ウェブの薄い寸法により、サンプリングアルゴリズムは、リソース（即ち、ネットワーク帯域、データベース記憶装置、プロセッサ時間など）の割り当てを最適化し、同時にインターネットのカバーする範囲を最大にするように正しく調節されなければならない。現存の広告追跡評価システムは、リソース割り当てに対して最適化されておらず、正しい割合で関連ウェブページを継続的にサンプリングしないので、これらの需要を満足することはできない。

20

【 0 0 0 7 】

上述した現存のシステムの欠点に関して、リソースをより賢く使用し、訪れたウェブサイトによりフレンドリーで、拡張性があり、正確な測定が可能な広告追跡評価システムの需要が存在する。ここで開示されている本発明は、この需要に着目している。なお、本発明に関連する先行技術として、特許文献 1 ~ 特許文献 1 0 に開示のものがある。

30

【 特許文献 1 】

米国特許第 5 , 7 3 2 , 2 1 8 号明細書

【 特許文献 2 】

米国特許第 5 , 7 2 7 , 1 2 9 号明細書

【 特許文献 3 】

米国特許第 5 , 8 3 5 , 7 5 8 号明細書

【 特許文献 4 】

米国特許第 5 , 8 7 8 , 2 2 3 号明細書

【 特許文献 5 】

米国特許第 5 , 9 9 9 , 9 6 8 号明細書

40

【 特許文献 6 】

米国特許第 6 , 0 1 2 , 0 8 3 号明細書

【 特許文献 7 】

米国特許第 6 , 0 4 9 , 8 2 1 号明細書

【 特許文献 8 】

米国特許第 6 , 1 1 2 , 2 1 2 号明細書

【 特許文献 9 】

米国特許第 6 , 6 0 1 , 1 0 0 号明細書

【 特許文献 1 0 】

50

米国特許出願公開公報第 2 0 0 2 / 0 0 0 2 5 7 1 A 1 号明細書

【 0 0 0 8 】

発明の要約

本発明は、インターネット等のコンピュータネットワークで配布されるデジタルコンテンツを追跡及び測定するシステム、方法及びコンピュータプログラムプロダクトである。システムは、オンライン広告データを収集し、データを解析し、それらの広告の普及の測定を計算するのにデータを使用する。

【 0 0 0 9 】

好ましい実施形態においては、無料の方法と様々なソースからのトラフィックデータは、インテリジェントエージェント（即ち、環境と対話、学習及び適応するソフトウェア）であるトラフィック解析システムを刺激する。トラフィック解析システムは、処理されたデータをデータベースに格納する前に、トラフィックデータを清浄及び要約することによって、生トラフィックデータを処理する。統計要約システムが、広告頻度、印象及び支出を計算する時、それは、トラフィック解析システムからの処理データに頼る。

10

【 0 0 1 0 】

広告サンプリングシステムは、「プロ - バー」又は「クラウドプローバー（Cloud prober）」としても知られており、プローブ（即ち、モニタ）する最も重要で影響力のあるウェブサイトを経続的に探索する強固な方法を使用する。更に、本発明によって実行されるサイト選択と定義は、完全に単一名実体がそのように報告されることを保証することを各ウェブサイトを有するウェブページに命令する。広告サンプリングシステムは、代表的なサンプルを取得するために様々な頻度でウェブページを検索するのにインテリジェントエージェント技術を使用する。これにより、クラウドプローバーは、各広告がトラフィックデータに現れる頻度を正確に評価することができる。クラウドプローバーがウェブページを取ってきた後で、広告サンプリングシステムはウェブページから広告を抽出する。好ましい実施形態においては、広告抽出装置は、「抽出装置」としても知られているが、ヒューリスティック抽出処理である自動広告検出（「AAD」）処理を喚起し、ウェブページから広告の全てを自動的に抽出する。

20

【 0 0 1 1 】

ウェブページから広告を抽出した後で、広告サンプリングシステムは、広告断片を解析するために分類エンジンを喚起する。分類装置は各断片を処理し、断片に対する分類を決定し、その後、断片と分類データをデータベースに格納する。広告サンプリングシステムによって実行された解析及び処理の結果は、クライアントが容易にクエリー可能な広告活動の豊富なカタログである。

30

【 0 0 1 2 】

本発明は、データベースのデータにアクセス及び更新するのにウェブフロントエンドとユーザインターフェースを使用する。ウェブフロントエンドは、本発明のクライアント又はユーザに、トラフィック解析によって移植されたデータベースへのクエリーインターフェース、広告サンプリング及び統計的要約システムを提供する。ユーザインターフェースは、システムアカウント管理、サイト管理、分類法管理、広告コンテンツ分類及びレートカード収集に対する別個のコンポーネントを含むグラフィカルユーザインターフェースである。ユーザインターフェースにより、アカウントマネージャー及びオペレータは、本発明を維持及び管理することができる。また、ユーザインターフェースは、本発明によって収集された莫大なデータ量の正確性と完全性を検証するために、メディアエディタがデータベース内のデータを検討することも可能にする。このデータの完全性処理は、システムを基準化してそれを常に変化する環境に適合するために、異常又は範囲外のデータ点を定期的に調査する。

40

【 0 0 1 3 】

添付図面は、本発明の詳細を、その構成と動作の両方において最も良く示している。これらの図において同様の参照符号と呼称同一部材を表す。

【 0 0 1 4 】

50

発明の詳細な説明

図1は、本発明の好ましい実施形態の環境を示し、インターネット100と、ウェブサイト110と、トラフィックサンプリングシステム120と、広告普及システム130と、クライアント140とを有する。本発明は、オンライン広告の属性、配置及び普及に関するデータを収集するのにインテリジェントエージェント技術を使用する。このデータは、広告統計の最新の見積もりをユーザに提供し、ユーザが競争的長所を獲得することを補助する。

【0015】

図1に示すように、インターネット100は、クライアント140とウェブサイト110と通信することをトラフィックサンプリングシステム120と広告普及システム130に可能にする公共通信ネットワークである。好ましい実施形態はインターネット100を使用しているが、本発明は、イントラネットやエクストラネットなど他の公的又は私的ネットワーク構成の使用も想定している。イントラネットは、インターネット100と同様の機能を有する私的通信ネットワークである。企業などの組織は、組織のネットワーク上のリソースにアクセスする安全な手段を組織構成員に提供するためにイントラネットを形成する。エクストラネットも、インターネット100と同様に機能する私的通信ネットワークである。イントラネットと対照的に、エクストラネットは、組織の非構成員に組織のネットワーク上のあるリソースにアクセスすることを許容する安全な手段を組織に提供する。また、本発明は、商標付きネットワークプロトコルと同様に、イーサネットやトークンリングなどのネットワークプロトコルを使用することも想定している。

【0016】

トラフィックサンプリングシステム120は、インターネット100上のウェブ活動を監視及び記録するプログラムである。トラフィックサンプリングシステム120は、インターネット100上の(図示しない)ウェブサーファとウェブサーバー112との間のトラフィックデータの間保管所である。図1に示すウェブサーバー112は、適当なオペレーティングシステム、ハードウェア、通信プロトコル(例えば、トランスミッションコントロールプロトコル/インターネットプロトコル)、ウェブページの収集を務めるウェブサーバーソフトウェアを含む従来のパーソナルコンピュータ又はコンピュータワークステーションである。(図示しない)ウェブサーファは、ウェブブラウザを典型的に使用して、ウェブサイト110に関連付けられたユニフォームリソースロケータ(URL)114、116、118を要求することによって、ウェブサーバー112と通信する。ウェブサーバー112に対して(図示しない)ウェブサーファによってなされたURLの要求を記録可能な、いかなるプログラム又は装置も、トラフィックサンプリングシステム120の、本発明が要求する機能を実行することができる。トラフィックサンプリングシステム120は、その後、広告普及システム130に使用される各ウェブサイト110に対してトラフィックデータを集める。

【0017】

本発明は、メディア・メトリクス(Media Matrix)視聴測定製品に類似した機能を提供する商業的に利用可能ないかなるトラフィックサンプリングシステムも利用することができる。トラフィックデータサンプルを取得するための他の可能な機構は以下のものを含む。

【0018】

1. 「プロキシキャッシュサンプリング(Proxy Cache Sampling)」は、ユーザクリックストリームデータ等のデータと、プロキシキャッシュサーバーの世界的な配布された階層からウェブページ要求を収集する。このデータは、ウェブオブジェクトに対してプリフェッチ及びキャッシュサービスを提供する中間機構を通過する。1999年5月の時点で、本発明によって計算されたトラフィック統計は、世界的には更なる1100と共に、合衆国で9つの第1層と約400の第2層キャッシュからの生データの蒸留を表す。

【0019】

2. 「クライアントサイドパネルコレクション (Client-Side Panel Collection)」は、クライアント側機構を介して各パネリストからサンプルデータを検索し、そのデータを収集保管所に移転する。クライアント側機構は、ブラウザロケーションバー、ユーズブラウザ、クライアント側プロキシ又はTCP/IPスタックフックを監視することができる。

【0020】

3. 「トランスコーダ」は、通常、広告歳入又はページヘッダー/フッターを作成する要素を追加する目的でHTMLを書き直すプロキシである。無料インターネットサービスプロバイダー (ISP) は典型的にこの技術を使用する。

【0021】

4. ネットワーク混雑を緩和するかユーザの経験を向上するような方法で、ウェブページ又はサイトを複製するいかなるコンテンツ配布機構

【0022】

5. URLに対する要求を評価し、かかる要求を許可又は不許可する行為をなすいかなるコンテンツフィルタリング機構

【0023】

6. インターネットサービスプロバイダー (ISP) 又は個人のウェブサイトが維持するサーバログから

【0024】

図2は、図1の広告普及システム130の詳細を拡張し、ネットワーク環境と広告普及システムを構成する構成要素との関係を示している。広告普及システム130は、格納のためにデータベース200にデータを通信する、トラフィック解析システム210と、広告サンプリングシステム220と、統計要約システム230とを有する。アカウントマネージャー260、オペレータ262と、メディアエディタ264は、管理機能を実行するために、ユーザインターフェース240を介してデータベース200にアクセスすることができる。クライアント140は、ウェブフロントエンド250を介してデータベース200にアクセスすることができる。

【0025】

トラフィック解析システム210は、トラフィックサンプリングシステム120から生トラフィックデータを受信する。トラフィック解析システム210は、トラフィックデータからインターネット100上で特定のユーザを識別するかもしれない情報を除去することによって生トラフィックデータを清浄し、その後、匿名データをデータベース200に格納する。トラフィック解析システム210は、インターネット100上のあらゆる重要なウェブサイトに対して世界的トラフィックを見積もる。本発明は、このデータを、そのページ上の変化頻度の見積もりがある場合に広告印象の数を計算するのみならず、プローブマッピングシステム320においても使用する。ある実施形態においては、トラフィック解析システム210は、インターネット100のキャッシュサイトからトラフィックデータを受信する。

【0026】

広告サンプリングシステム220は、どのURLがウェブサーバー112から検索されたサンプルに含まれるかを決定するのに匿名トラフィックデータを使用する。広告サンプリングシステム220は、URL114、116、118を検索し、広告を記述する付随特徴と共に広告を抽出するのにインターネット100を介してウェブサーバー112に連絡する。独創的なものの検索の成功レートは高い。解析は、本発明が使用された独創的なものの95%以上を捕獲することを示している。広告サンプリングシステム220は、これらの広告特徴をデータベース200に格納する。広告サンプリングシステム220、例えば、オンライン・メディア・ネットワーク・インテリジェント・コレクション (Online Media Network Intelligent Agent Collection: OMNIAC) 又は、クラウドプローバーは、繰り返し重要なウェブサイトをプローブし、プローブによって戻された各ウェブページから広告を抽出し、各ウェブページ

10

20

30

40

50

の広告を種類、技術、広告主別に分類する。

【0027】

また、トラフィック解析システム210及び広告サンプリングシステム220も、インターネット100から検索されたデータを統計要約システム230に定期的な処理のために提供する。統計要約システム230は、毎週サイト毎に、広告頻度、印象及び支出を計算する。

【0028】

本発明に対するグラフィカルユーザインターフェースは、ユーザインターフェース240とウェブフロントエンド250を含む。アカウントマネージャー260、オペレータ262及びメディアエディタ264は、クライアント140によるウェブフロント250へのアクセスを管理(例えば、アカウント及びパスワード管理)し、サイトとプローブ命令を定義し、広告分類法、コンテンツ分類及びレートカード収集を広告普及システム130のために管理するために、ユーザインターフェース240にアクセスする。ウェブフロントエンド250は、トラフィック解析システム210、広告サンプリングシステム220及び統計要約システム230によって生成されるように、クライアント140がデータベース200から広告測定結果を検索するのに使用するウェブブラウザインターフェースである。

【0029】

図3は、図2に示す広告普及システム130の構成要素を有する論理コンポーネントを示すために、広告普及システム130の詳細を更に拡張している。また、図3は、ネットワーク環境とそれらの論理コンポーネントとの関係も示している。

【0030】

トラフィック解析システム210は、匿名システム310とトラフィック要約312処理とを有する。

【0031】

匿名システム310は、インターネット上で特定のユーザを識別する情報を除去することによって、トラフィックサンプリングシステム120から受信したデータを清浄する。データは、全てのユーザ情報(例えば、送信元インターネットプロトコル(IP)番号又はクッキー)を暗号的に機密保持された一方向ハッシュ関数を通過させることによって匿名にされる。これは、結果データの価値を損なうことなしに、ウェブユーザに対して最大のプライバシーを確保する。匿名システム310は、清浄されたデータをトラフィック要約システム312に提供し、それは、次いで、収集されたURLコンテンツ情報をデータベース200に格納する。

【0032】

トラフィック要約処理312は、匿名システム310から清浄データを受信する。匿名トラフィックデータは、個々のURL、ドメイン及びウェブサイトに対して週又は月別にトラフィック合計を生み出すのために要約される。トラフィック要約処理312は、サンプルから全世界的トラフィックを推定するために係数を重み付けすることによってデータを測る。

【0033】

図3における広告サンプリングシステム220は、プローブマッピングシステム320と、ウェブページ検索システム322と、ウェブブラウザエミュレーション環境324と、広告抽出装置326と、構造分類装置328とを含む。

【0034】

プローブマッピングシステム320は、プローブマップ、即ち、広告サンプリングシステム200が訪れるURL114、116、118を生成する。このプローブマップは、広告サンプリングシステム220が個々のウェブサイト上の広告変化を測定するのを補助する。本発明の好ましい実施形態は、プローブマップの様々なウェブページを継続的にフェッチする。代替的实施形態においては、本発明は、プローブマップの各URLを略6分毎に訪れる。別の実施形態は、ウェブサイト全体と問題の個々のウェブページを訪れるトラ

10

20

30

40

50

フィック量、ウェブページ上で履歴上見られた広告数、及び、別のサンプリングされたページに対する履歴上観測された広告変化の類似性を含む幾つかの要因を考慮することによって、フェッチングレートを変化させることができる。

【0035】

ウェブページ検索システム322は、それがどのウェブページをサンプリングする必要があるか及びサンプリングの頻度を決定するのに、プローブマッピングシステム320によって生成されたプローブマップを使用する。プローブマッピングシステム320によって生成されたプローブマップにおける各URLに対して、ウェブページ検索システム322はウェブページをフェッチして、ウェブページから各広告を抽出し、広告の属性をデータベース200に格納する。プローブマップにおいて各URLから検索されたデータは、各広告が特定のウェブサイトで表示される頻度を計算するのに使用される。

10

【0036】

各ウェブページに対して、ウェブブラウザエミュレーション環境324は、ブラウザにおけるウェブページのディスプレイをシミュレートする。このシミュレーションは、本発明が、静的広告だけでなく、ウェブページに埋め込み可能な、ジャバスクリプト、パール(Perl)、ジャバ、C、C++又はHTMLなどの言語で記載されたソフトウェアプログラムによって生成された動的広告も検出するであろうことを保証する。

【0037】

広告抽出装置326は、ウェブブラウザエミュレーション環境324によって実行されたシミュレーションからオンライン広告を抽出する。広告抽出装置326は、特に興味のあるプローブマッピングシステム320によって返却されたウェブページから抽出された広告コンテンツの特徴(即ち、断片)を識別する。広告は抽出する最も興味深い動的特徴であるが、本発明の代替的实施形態は、プロモーション、調査及びニュース記事を含むデジタルコンテンツのいかなる種類を収集する抽出技術を使用することができる。広告抽出装置326は、規則基準抽出、ヒューリスティック抽出、比較抽出を含む様々な広告抽出方法を使用することができる。

20

【0038】

規則基準抽出は、規則を作成するのにユーザインターフェース240を使用するためにメディアエディタ264に依存する。ユーザインターフェース240は、データベース200に規則を格納し、広告抽出装置326は、ウェブページ検索システム322が検索する各ウェブページに規則を適用する。規則運営の効果は、ウェブページ(即ち、広告を含むページの一部)からHTML断片を識別及び抽出することである。広告抽出装置326は、まず、フェッチされたウェブページのHTML表現を整形式のXML表現に変換する。この変換に続いて、規則はウェブページのXML表現のパースツリーに適用される。

30

【0039】

ヒューリスティック抽出は、ウェブページを表示するというウェブブラウザエミュレーション環境324要求に回答してウェブページが生成される時に広告は広告サーバーによって典型的に挿入されるので、HTML又はXMLソースコードレベルで広告の類似性に依存する。ヒューリスティック抽出は、手がかりのためにソースコードを解析し(例えば、既知の広告サーバーの名前への参照)、それらの手がかりの周りの断片を抽出する。この方法の長所は、抽出が自動的で、メディアエディタがこの規則を作成する必要がないことである。

40

【0040】

比較抽出は、繰り返し同一のウェブページをフェッチする。この抽出方法は、コンテンツがバージョンからバージョンに変化するかどうかを決定するために、異なるバージョンのウェブページを比較する。ある程度の頻度で変化するウェブページの一部は、通常は、広告であり、抽出される。

【0041】

構造分類装置328は、各広告をパースし、構造コンポーネントをデータベース200に格納し、それらのコンポーネントを統計要約システム230に渡す。広告抽出装置326

50

によって抽出された各広告断片は、構造分類装置 3 2 8 によって解析される。構造分類装置 3 2 8 によって実行される処理は、重複断片排除、構成断片解析、重複広告検出を含む。

【 0 0 4 2 】

構造分類装置 3 2 8 は、現在の広告断片をデータベース 2 0 0 内の他の断片と比較にすることによって重複断片排除を実行する。2つの広告断片は、断片が同一であれば、重複である（例えば、各断片は全く同一のHTMLコンテンツを有する）。構造分類装置 3 2 8 は、現在の断片がデータベース内の断片の重複であると判断すると、広告サンプリングシステム 2 2 0 は、断片の別の観測をログし、処理断片を継続する。

【 0 0 4 3 】

構造分類装置 3 2 8 は、断片の「物理的種類」（即ち、広告を構成するのに使用されるHTMLソースコード）を決定することによってウェブページのXML表現に関して構成断片解析を実行する。本発明が認識する物理的種類は、バナー、形式、単一リンク及び埋め込まれたコンテンツを含む。バナー広告断片は、一又は二の囲まれた画像を有してFORM又はIFRAMEタグを有しない単一のHTMLリンクを含む。形式広告断片は、IFRAMEタグを有しない単一のHTML形式を含む。単一リンク広告断片は、テキストであるがIMG、FORM又はIFRAMEタグのないリンクを含む。埋め込まれたコンテンツ広告断片は、IFRAMEタグを使用して外部実体を参照する。この解析実行後に、構造分類装置 3 2 8 は、データベースにおいて広告断片を更新する。バナー広告断片に対して、構造分類装置 3 2 8 はリンクと画像URLをデータベース 2 0 0 に格納する。形式広告断片は、各HTML制御をそのデフォルト値に設定するユーザ提出をシミュレーションすることによってURLの作成を要求とする。構造分類装置 3 2 8 は、このURLと「形式署名」（即ち、この形式内の全制御のコンテンツを独自に記述するストリング）をデータベース 2 0 0 に格納する。単一のテキスト広告断片に対して、構造分類装置 3 2 8 は、リンクに対してURLとリンク内に含まれる全テキストをデータベース 2 0 0 に格納する。埋め込まれたコンテンツ広告断片に対して、構造分類装置 3 2 8 は、外部参照に関連付けられたURLをデータベース 2 0 0 に格納する。このURLは、システムによってロードされ、参照された文書はロードされる。一旦ロードされた文書が構造解析されると、原断片は、新しい断片の解析に由来するいかなる属性も受け継ぐ。

【 0 0 4 4 】

構造分類装置 3 2 8 は、既知の物理的種類を有する各広告断片に対して、これらの断片が広告を表すので、重複広告検出を実行する。各独自性のある広告は、データベース 2 0 0 に格納された、どのサイト定義が断片と関連付けられているかを含む情報を有する。構造分類装置 3 2 8 による独自性の決定は、断片の各種類に対して異なる基準に依存する。あらゆる種類の定義に対する最初のステップは、記録に関連付けられた全URLを区別することである。画像を参照するURLがロードされ、重複画像がメモされる。HTMLリンクURLは、「クリックURL」としても知られており、新しい広告が作成される毎に生じる。全てのHTTPが向け直し後に、クリックURLの最終目的地がメモされる。また、これはシミュレーションされたリンク提出に対してもなされる。一旦全てのURLが区別されると、構造分類装置 3 2 8 は、広告が独自であるかどうかを決定する。バナー広告断片は、それらが同一の画像数を有し、画像が同一で、かつ、目的地URLが同一であれば、独自性があるとみなされる。形式広告断片は、同一の署名と同一の目的地URLを有すれば、独自性があるとみなされる。単一リンク広告断片は、それらが同一のテキストコンテンツと同一の目的地URLを有すれば、独自性があるとみなされる。

【 0 0 4 5 】

統計要約システム 2 3 0 は、データベース 2 0 0 の独自の広告毎に広告統計を計算する。本発明は、ウェブサイト毎に、広告印象（即ち、人間が広告を見る回数）を計算する。本発明は広告印象 I を、 T をサイトに行くトラフィック、 R をそのサイトの広告変化として、式 $I = T \times R$ を利用して計算する。また、本発明は、 I をウェブサイトに対する広告印象、 RC をウェブサイトに対するレートコードとして、式 $S = I \times RC$ を利用して支出 S

10

20

30

40

50

も計算する。殆どの広告購入は、量購入割引の複雑な取引であるので、我々の数は購入合計の現実の費用を必ずしも表さない。

【 0 0 4 6 】

ウェブフロントエンド 2 5 0 は、クライアント 1 4 0 に、トラフィック解析システム 2 1 0、広告サンプリングシステム 2 2 0 及び統計要約システム 2 3 0 によって移植された、データベース 2 0 0 へのクエリーインターフェースを与えるグラフィカルユーザインターフェースである。クライアント 1 4 0 は、クライアント 1 4 0 が選択するサービスのレベルに依存する一以上の産業部門に対してグラフィカルで表にまとめられた報告を作成、格納、編集及びダウンロードするのにウェブフロントエンド 2 5 0 を使用することができる。

10

【 0 0 4 7 】

図 3 のユーザインターフェース 2 4 0 は、システムアカウント管理 3 4 0、サイト管理 3 4 2、分類法管理 3 4 4、広告コンテンツ分類 3 4 6 及びレートカード収集 3 4 8 に対する別個のコンポーネントを含む。

【 0 0 4 8 】

アカウントマネージャー 2 6 0 は、ウェブフロントエンド 2 5 0 の管理を単純化するのに、ユーザインターフェース 2 4 0 のシステムアカウント管理 3 4 0 モジュールを使用する。アカウントマネージャー 2 6 0 は、ユーザアカウントを作成及び削除し、ユーザアカウントパスワードを管理し、ウェブフロントエンド 2 5 0 の健康全体をチェックするのに、システムアカウント管理 3 4 0 モジュールを使用する。

20

【 0 0 4 9 】

オペレータ 2 6 2 は、サイト定義の管理を単純化するのに、ユーザインターフェース 2 4 0 のサイト管理 3 4 2 モジュールを使用する。インターネット広告事務局 (Internet Advertising Bureau) からの解析者は、ウェブ広告全体ドルの 9 0 % 以上が上位 5 0 のウェブサイトで支出されていると見積もっている。サイト選択は、メディア・メトリックス、ニールセン/ネットレーティング (Nielsen / Net Ratings)、及び、データベース 2 0 0 のプロキシトラフィックデータからのデータを考慮することによって上位 1 0 0 の広告を選択することによって開始する。これらのリストは、低トラフィックレベルのウェブサイトを降格し、高トラフィックレベルの新しいサイトを促進することによって定期的に更新される。また、本発明は、基幹産業において相当のコンテンツを与えるウェブサイトも含む。サイト定義に含まれるように選択されたサイトは、広告を提供しない、外国から派生する、フレームセットの一部であるセクションを除去するためにサイトの構造が解析させなければならない。yahoo.co.jp など外国から派生するサイトは、ホスト国で広告を販売するので、本発明によって計算された測定には適用できない。HTML フレームセットを使用するウェブサイトは、変化率を広告を含むフレームセットのセクションからのトラフィックに適用するためにのみ大変注意深く取り扱われる。これらの組み合わせられた除外は、広告印象の正確な見積もりをなすための鍵である。また、本発明は、登録要求のために (例えば、メールページなど) 直接は測定できないセクションにタグも付す。ウェブサイトは頻繁に変化するので、この構造解析は定期的に繰り返される。最終的に解析段階は、より時期を得た更新を許容するために、変更されたサイトに自動的にフラグを立てる。

30

40

【 0 0 5 0 】

メディアエディタ 2 6 4 は、ユーザインターフェース 2 4 0 の分類法管理 3 4 4、広告コンテンツ分類 3 4 6 及びレートカード収集 3 4 8 モジュールを使用する。分類管理 3 4 4 モジュールは、広告の業界、企業及び製品を含むコンテンツ分類中に、広告に割り当てられた属性の作成と維持を単純化する。分類法は、各属性に名前を付け、その種類、由来及びセグメント会員を特定する。例えば、企業ホングは、自動車業界で起源をなし、業界セグメント自動車メーカーに属する。広告コンテンツ分類 3 4 6 コンポーネントは、メディアエディタ 2 6 4 がコンテンツ分類を実行するのを補助する。

【 0 0 5 1 】

50

構造分類装置 328 は、広告が宣伝しているものを決定するために、自動化された広告可能な割り当てを実行する。この処理は、「広告可能性」（即ち、広告が宣伝する各「物」を記述する属性）を各広告断片に割り当てることを含む。本発明の別の実施形態においては、広告サンプリングシステム 220 は、広告可能性を各広告に割り当てるのにヒューリスティックな伸張可能なセットを使用する。しかし、好ましい実施形態においては、使用される自動化された方法のみがロケーション分類である。ロケーション分類は、一連の広告可能性を広告に割り当てるために、目的地 URL に依存する。メディアエディタ 264 は、一連の分類されたロケーションを維持するのにユーザインターフェース 240 を使用する。例えば、メディアエディタが初めて、クリックスルー URL が `www.honda.com` である広告を観察する時に、それはこの URL を広告主「ホンダモーターズ」に 10
 関連付けるように入力する。また、同一のクリックスルー URL を含むその後のいかなる広告もホンダの広告として認識されるであろう。分類されたロケーションは、ホスト、URL パスプリフィックス及び広告可能性を有する。ロケーション分類は、目的地 URL のホストが分類されたロケーションのホストと分類されたロケーションのパスプリフィックスが目的地 URL のパスの始めに一致すれば、分類されたロケーション広告可能性を広告に割り当てる。

【0052】

構造分類装置 328 は、広告可能データの品質チェックとして人的広告可能性割り当て及び検証を実行する。このフェーズは最も人的集中がなされる。メディアエディタ 264 は、各広告を表示し、自動広告可能割り当てを検証し、広告調査後に適当になる他のいかなる 20
 適当な広告可能性及び広告の目的地をも割り当てるのにユーザインターフェース 240 のグラフィカルユーザインターフェースモジュールを使用する。また、ロケーション分類データベースも、典型的にこの時維持される。

【0053】

メディアエディタ 264 は、指定された広告主と同様に、トラフィック解析システム 210 によって識別されたウェブサイトに対する連絡及びレートカード情報に入るのに、レートカード収集 348 モジュールを使用する。レートカード入力は、適用可能なクォーター（例えば、Q4 2000）、ピクセルの広告寸法、費用構造（例えば、CPM、フラットレート又はクリック毎）、様々な量と期間の購入に対する費用スケジュールを含む。また、メディアエディタ 264 は、オンラインメディアキットの URL アドレスと、レート 30
 がここに公開されているかどうかを記録する。ウェブサイト又は広告主に対する連絡情報は、ホームページ、名前、電話及びファックス番号、電子メールアドレス、住所を含む。

【0054】

図 4A 乃至 4C は、広告普及統計を計算するための好ましい方法を示している。広告普及統計の計算は、重み付けされた値と広告普及統計を計算するのに、トラフィック解析システム 210 に由来する期待値と広告普及システム 220 に由来する観測値を使用する反復処理である。図 4A 乃至図 4C は、ウェブサーバ P410 とウェブサーバ Q420 によって提供される 2 つのウェブサイトを含むインターネット 100 上のネットワークを各々示している。図 4A は、ネットワークに対する例示的なトラフィック期待値を示している。図 4B は、ネットワークに対する例示的なトラフィック観測値を示している。図 4C 40
 は、ネットワークに対する例示的なトラフィック加重値を示している。

【0055】

処理の最初のステップは、トラフィック分析システム 210 からの結果を正規化することである。トラフィック分析システム 210 は、トラフィックデータサンプルに各ウェブページによって受信されたトラフィックを与える。図 4A は、「トラフィック = 」とラベルされたインターネット 100 の各ウェブページ 411 - 416、421 - 424 で受信された例示的なトラフィックを示す。プローブマッピングシステム 320 によって生成されたプローブマップは、ウェブページ 411 - 416、421 - 424 の各々に対する入力を含む。また、プローブマップは、各ウェブページ 411 - 416、421 - 424 がプ 50
 ローブマップで消費する「領域」も含む。図 4A は、各ウェブページ 411 - 416、4

21-424が、「領域=」でラベルされたプローブマップで消費する例示的な領域を示す。正規化された結果は、ウェブページがプローブマップで消費する領域を、トラフィックサンプルにおける各ウェブページに対する領域の合計で割ることによって計算される。図4Aにおいて、ウェブページP1-411に対する正規化された値又は可能性は、ウェブページP1、P2、P3、P4、P5、P6、Q1、Q2、Q3及びQ4に対する領域の合計（即ち、120）で分割された、ウェブページP1に対する領域（即ち、15）である。従って、正規化された値は、0.125又は12.5%である。また、正規化された値に加えて、システムは、ウェブページに対する領域によってウェブページに対するトラフィックを割ることによってスケールを決定もする。図4Aにおいて、ウェブページP1-411に対するスケールは、ウェブページP1に対する領域（即ち、15）によって分割されたウェブページP1に対するトラフィック（即ち、150）であるので、ウェブページ1に対するスケールは10である。表1は、図4Aの残っているウェブページに対するスケールと可能性の値をまとめたものである。

10

【0056】

図4Bは、「フェッチ=」とラベルされたインターネット100の各ウェブページ411-416、421-424における例示的なウェブページフェッチを示している。また、図4Bは、例えば、広告A1のビューの数を示すための「A1ビュー=」、広告A2のビューの数を示すための「A2ビュー=」などのラベルを有するウェブページ411-416、421-424に現れる各広告の例示的なビューの数も示している。

20

【0057】

図4Cは、「フェッチ=」とラベルされたインターネット100の各ウェブページ411-416、421-424における例示的なウェブページ加重フェッチを示す。また、図4Cは、例えば、広告A1のビューの数を示すための「A1ビュー=」、広告A2のビューの数を示すための「A2ビュー=」などのラベルを有するウェブページ411-416、421-424に現れる各広告の例示的なビューの数も示す。計算処理の次のステップは、ウェブサイトの各ウェブページ4411-416、421-424に対して、図4Bからの観測されたフェッチと図4Aからの割合との積を足し合わせるることによって、各ウェブサイト410、420に対するスケールドフェッチを計算することである。次に、ウェブサイトの各ウェブページ4411-416、421-424に対して、図4Aからのトラフィックを足すことによって、各ウェブサイト410、420に対するトラフィックを計算する。レートカード又はCPMは、各ウェブサイト410、420に対するメディアエディタ264によって割り当てられた値である。表2は、図4A乃至4Cに対するスケールドフェッチ、トラフィック及びCPMをまとめたものである。

30

【0058】

計算処理の次は、ウェブサイト410、420の各ウェブページ411-416、421-424に対して、図4Bからの広告ビューと図4Aからのスケールの積を足し合わせるることによって、各ウェブサイト410、420の各広告に対して、スケールド観測を計算することである。計算の最終ステップは、各ウェブサイト410、420の各広告に対して広告普及統計（即ち、周期、印象及び支出）を計算することである。頻度は、各ウェブサイト410、420の各広告に対して、スケールド観測をスケールドフェッチによって割ることによって計算される。印象は、各ウェブサイト410、420の各広告に対して、頻度を上記表2からのトラフィックと掛けることによって計算される。支出は、各ウェブサイト410、420の各広告に対して、印象と上記表2からのCPMを掛けることによって計算される。表3は、図4A乃至図4Cのデータを使用して、ウェブサイトP410に対するスケールド観測、頻度、印象及び支出をまとめたものである。表4は、図4A乃至図4Cのデータを使用して、ウェブサイトQ420に対するスケールド観測、頻度、印象及び支出をまとめたものである。

40

【0059】

【表3】

	スケールド観測	頻度	印象	支出
A1	55.0	0.28	52.58	\$1.84
A2	85.0	0.44	81.27	\$2.84
A3	6.0	0.03	5.74	\$0.20
A4	3.5	0.02	3.35	\$0.12
A5				

【表4】

	スケールド観測	頻度	印象	支出
A1	29.5	0.69	34.99	1.75
A2	12.0	0.28	14.23	0.71
A3	12.0	0.28	14.23	0.71
A4	12.0	0.28	14.23	0.71
A5	1.5	0.03	1.78	0.09

図4Dは、広告普及統計を計算する代替的实施形態を示す。本実施形態においては、プロバーは変化測定精度を最適化するのに調節される。フィールドにおける正確性の統計見積もりは、広告サーバーの非静止な性質により実行が困難である。6分毎にプローブすると、1週間の測定期間に亘って変化頻度に0.06%分解能を有する。

【0060】

また、図4Dの代替的实施形態においても、プローブは、各サイトの広告変化を正確に測定するために、サイト間で配布される。サイトに割り当てられたプローブURLの数は、3つの変数から決定される。第1は、全てのサイトに一定である。ある数のプローブURLが最小サイトでさえ変化を正確に測定するのに必要である。プローブの半分はこの変数に割り当てられる。第2の変数は、40%で重み付けされ、サイトに行くトラフィックの量であり、各プローブURLは、インターネットトラフィック合計の割合を表す。20の最大サイトは、これらのプローブの75%以上を受信する。最後に、サイトの複雑性は、我々のプロキシトラフィックデータで発見される独自のURLの総計によって測定され、余分なプローブURLを受信するより複雑なサイトと共に、考慮される。これは、プローブ配布の残り10%を明らかにする。プローブURLは、サイトを領域（即ち、広告変化特徴が類似する可能性があるページのセット）に分割するのにサイト・シュレッター・アルゴリズムを使用して選択可能である。領域の配布は、サイトが覆う範囲と従って広告変化精度が最大になるように数学的に設計される。単一のURLが、各領域からの広告変化を表すように選択される。このURLは、その領域に広告を含む最もトラフィックが重いページとして選択される。アルゴリズムは、日付特定ページ又は、1999年8月月食などの時限事象を参照するページを回避する。

【0061】

図4Dの代替的实施形態は、各ウェブサイト430に対して変化及びトラフィックの見積もりを組み合わせることによって、広告印象を計算する。これを行うために、システムは、サイト・シュレッター・アルゴリズムを使用してその構成ステムにサイトを分割する。各広告スロットにおける広告変化は、その関連付けられたステムに関する広告印象を見積もるために、計算及び適用される。プローブなしのステムに関する広告変化は、トラフィックによって重み付けられた、類似レベルでのプローブの広告変化の平均から見積もられる。

【0062】

例えば、図4Dにおいて、サンプルサイトツリーは、メインページと14の二次ブランチから離れた5つの主要ブランチに配置された5つのプローブURL 431 - 435、P₁ - P₅を有する。各ページの数、そのページに行くサンプルトラフィックである。ホームページ「www.testsite.com」のプローブP₁は、88のページビューの

10

20

30

40

50

トラフィックと共に、そのメインページに行くトラフィックに適用される変化Rを測定する。ブランチAは、プローブURL「www.testsite.com/A/」を有し、そのブランチのトップレベルページに配置される単一プローブP₂を有する。この単一プローブURLの変化はR_Aとして見積もられ、21のページビューの合計であるそのステム全体に対するトラフィックに適用される。ブランチCは、プローブURL「www.testsite.com/C/third.html」を有し、トラフィックの重い二次ブランチページ上のプローブP₃を有する。このページの変化R_Cは、25のページビューの合計に亘って、そのステム上でツリーのレベル高い二次ブランチページに適用される。ブランチEは、合計61のページビューであるサイトに対するトラフィックの大部分を受信するため、2つのプローブP₄及びP₅に割り当てられる。これらは、2つの二次ブランチページ「www.testsite.com/E/first.html」及び「www.testsite.com/E/third.html」である。各々の変化は、それらの個々のページに対してトラフィックに適用される。そのブランチ上の残りの18のページビュー（2つの二次ページとそのブランチのトップレベルページの8からの10のページビュー）に対して、重み付けされた変化は、 $R_E = ((13 \times R_{E1}) + (30 \times R_{E3})) / (13 + 30)$ で計算される。ステム変化の解析は、サイトの96%以上に対する広告印象をもたらす。最後の2つのブランチB及びDに対する印象は、トラフィックによって重み付けられた、隣り合うブランチからの平均変化によって、 $R_B = R_D = ((21 \times R_A) + (25 \times R_C) + (61 \times R_E)) \div (21 + 25 + 61)$ で計算される。この解析は、各独自の広告に対してサイトに亘る印象全体をもたらす。図4Dの代替的实施形態によって実行される最終計算は、印象とレートカードの積としての支出である。

【0063】

図5は、広告普及システム130がトラフィックサンプリングシステム120とウェブページ検索システム322によって検索された情報を格納するのに使用可能なデータベース構造を示している。好ましい実施形態は、データベース200をパーティションに分割する。各パーティションは、データベース200などの独立したデータベースと同様の機能を実施することができる。また、パーティションに分割されたデータベースはパーティション内のデータの管理を単純化する。好ましい実施形態がデータベースパーティションを使用している場合、本発明は、各パーティションを独立したデータベースにし、各データベースを別個の汎用コンピュータワークステーション又はサーバーに配布すると同様に、これらのパーティションを一のデータベースに統合することを想定している。本発明のデータベース200に対するパーティションは、サンプリング記録510、プローブ定義520、広告サポートデータ530及び広告要約540を含む。本発明の好ましい実施形態は、データベースとパーティションを作成及び管理するのに、オラクルコーポレーション(Oracle Corporation)によって製造されるオラクル8i製品などのリレーショナルデータベース管理システムを使用する。好ましい実施形態がリレーショナルデータベースを使用している場合、本発明はオブジェクト指向データベース管理システムなど、他のデータベースアーキテクチャの使用を想定している。

【0064】

データベース200のサンプリング記録510パーティションは、トラフィックデータ512、広告ビューロギング514及び広告構造516領域に論理的に分割されたデータベーステーブルを有する。

【0065】

トラフィックデータ512領域は、トラフィックサンプリングシステム120、匿名システム310及び統計要約システム230によって処理されたデータを含む。このスキーマに格納されたデータは、「書き換えられた(munged)」URLと、各URLがトラフィックソース毎に所定期間に亘って受信するトラフィックのカウントとを含む。「書き換えられた」URLは、プロトコルフィールドが除去され、ホスト名におけるドットコンポーネントの順番が逆転された通常のURLである。例えば、本発明は、http://

10

20

30

40

50

www.somesite.com/foodなどの通常のURLを、プロトコルフィールド（即ち、「http://」）を除去し、ホスト名（即ち、「www.somesite.com」）におけるドットコンポーネントの順番を逆転させることによって、書き換えられたURLに変換する。本例で結果としての書き換えられたURLは、「com.somesite.www/food」である。本発明は、トラフィックデータ解析処理を非常に強めるのに、この商標付きURL形式を使用する。トラフィックサンプリングシステム120は、トラフィックデータ512領域をデータベース200に移植する。プロンプマッピングシステム320は、ウェブページ検索システム322及び統計要約システム230が広告印象及び支出統計の計算をするのを補助するために、トラフィックデータ512領域内のデータにアクセスする。

10

【0066】

広告ビューロギング514領域は、インターネット100上で遭遇された各広告に対して、時間、URL及び広告識別子をログする。また、この領域も、システムが広告を以前は含んでいたウェブページ内に広告を検出しないたびにログする。更に、システムは、システムが潜在的な広告を検出するが、構造分類中に広告を認識することができないたびにログする。広告サンプリングシステム220の構造分類装置328及びウェブページ検索システム322は、データベース200に広告ビューロギング514領域を移植する。統計要約システム230は、広告ビューロギング514内のデータにアクセスし、各広告が各サイトに現れる頻度を決定する。

【0067】

広告構造516領域は、システムによって突き止められた各独自性のある広告を特徴づけるデータを含む。このデータは、広告のコンテンツ、広告の種類（例えば、画像、HTML形式、フラッシュなど）、広告にリンクされた目的地URL、及び、コンテンツ分類及び診断中に使用され、広告が最初に見られた場所、どの広告の定義が最初に広告を作成したかを含む幾つかの項目を含む。広告サンプリングシステム220の構造分類装置328コンポーネントは、広告構造516領域をデータベース200に移植する。ユーザインターフェース240は、広告構造516領域内のデータにアクセスし、分類編集集中にメディアエディタ264に各広告を表示する。また、ウェブフロントエンド250は、クライアント140に広告を表示するために、広告構造516領域中のデータにもアクセスする。

20

【0068】

データベース200のプローピング定義520パーティションは、サイト定義522、プロンプマップ524及び広告抽出規則定義526領域に論理的に分割されたデータベーステーブルを有する。

30

【0069】

サイト定義522領域は、システムが領域にプローブするインターネット100の一部を分割する。一次領域定義は、「サイト」であり、システムが解析、サンプリング及び要約する必要がある結合した実体である。システムは、包括的及び排他的な書き換えられたURLプレフィックスで各サイトを定義する。「書き換えられたURLプレフィックス」は、それがプレフィックスである全ての書き換えられえたURLの領域を表す、書き換えられたURLである。「包括的な書き換えられたURLプレフィックス」とは、URLがある実体の一部であることを特定する。「排他的な書き換えられたURLプレフィックス」とは、URLがある実体の一部ではなく、包括的なプレフィックスによって含まれる実体の一部をただ乗りしていることを特定する。例示的に、以下のものは、トラフィックサンプルにおいて一連のURLの処理に起因し得る、書き換えられたURLのリストである。

40

1. com.somesite/
2. com.somesite/fo
3. com.somesite/foob
4. com.somesite/fooblah
5. com.someothersite/

【0070】

50

「somesite」に対するサイト定義が包括的なURLプリフィクス「com.somesite/」と排他的なURLプリフィクス「com.somesite/food/bar」を含む場合、上記にリストされたサンプルURLへのこのサイト定義の適用は、URL 1、2及び4を含むシステムを生み出す。URL 3は、「com.somesite/food/bar」の明示的除外のために、サイト定義の一部ではない。URL 5は、包括的なURLプリフィクス「com.somesite/」に含まれることは決してないために、サイト定義の一部ではない。ユーザインターフェース240は、サイト定義522領域をデータベース200に移植する。プローブマッピングシステム320は、どのURLがプローブされるべきかを決定するために、サイト定義522領域内のデータにアクセスする。統計要約システム230は、サイトに含まれるURLへのトラフィックを足し合わせることによってサイトへのトラフィックレベルを決定するために、サイト定義522領域にアクセスする。

10

【0071】

プローブマップ524は、システムが測定している各サイトにおける各URLに対する重みを含む。この重みは、システムが各プローブに対してURLを選択する可能性を決定する。システムは、データベース200内のトラフィックデータとプローブ記録に対して複雑な反復アルゴリズムを起動することによって重みを生成する。トラフィックデータの解析は、どのURLが訪問されたか、ユーザがどれくらい頻繁にそれらのURLを訪問したかを見分けることができる。解析の結果は、任意の単一URLへ割り当てるための最大数のプローブなどのある制限があれば、システムがこれらのURLの広告サンプリングを実行することを保証する。データベース200のサンプリング記録510パーティションのデータは、過去の行為のためにどのURLが特別処理を必要としているかを決定するのに有用である（例えば、システムがURL内の広告を検出しなければURLが余り頻繁にはサンプリングされない）。プローブマッピングシステム320は、データベース200にプローブマップ524領域を移植する。プローブマッピングシステム320は、プローブを割り当てるためにプローブマップ524内のデータにアクセスする。統計要約システム230は、どのURLが、プローブマップ制限実施の効果に反対するために、スケールされたそれらの変化を有するべきかを決定するために、プローブマップ524領域内のデータにアクセスする。

20

【0072】

広告抽出規則定義526領域は、典型的に正規化されたHTML文書を表し、システムが広告であるとみなすコンテンツの部分を示すエクステンシブル・マークアップ言語（XML）タグを記述する。システムは、「XML構造」と「XML特徴」で抽出規則を定義する。「XML構造」は、他のXMLノードに対する様々なXMLノードの位置を参照する。例えば、画像（IMG）ノードを含むアンカー（A）ノードは、広告になる見込みがある。広告コンテンツを一致するこの構造検出処理を使用した後で、システムは、コンテンツが広告であるかどうかを決定するために、コンテンツの特徴を調べる。前例を繰り返すために、画像ノードがサブストリング「広告サーバー（adserver）」を含むリンク（href）特徴を含んでいれば、広告になる見込みが大変高い。特徴は、例にあるように、単純なサブストリング又はより複雑な通常表現に基づいて一致する。抽出規則の別の形式は、「Xポインタ（Xpointer）」などのある形式のXMLパス仕様を使用してXML構造の特定のノードへの指示であってもよい。メディアエディタ264は、広告抽出規則定義526領域をデータベース200に移植する。広告サンプリングシステム220の広告抽出装置526は、各プローブされたページのどの部分が広告を表すかを決定するために、広告抽出規則定義326領域のデータにアクセスする。

30

40

【0073】

データベース200の広告サポートデータ530パーティションは、広告可能分類法532、広告情報534、レートカード536及び拡張された広告可能情報538領域に論理的にセグメント分割されたデータベーステーブルを有する。

【0074】

50

広告可能分類法 5 3 2 領域は、広告が宣伝しているものを記述する属性である広告可能性の階層的な分類法を含む。この分類法は、業界、企業、製品、ウェブサイト、ウェブサブサイト、メッセージなどを含む。階層の各ノードは、それが表す実体の種類と親ノードを特定する種類を有する。例えば、階層は、製品が企業内に存在し、それは次いで業界内に存在することを特定する。メディアエディタ 2 6 4 はデータベース 2 0 0 に広告可能分類法 5 3 2 領域を移植する。ユーザインターフェース 2 4 0 は、統計データを生成して企業、業界などが広告しようとする場所を記録するために、広告可能な分類 5 3 2 領域内のデータにアクセスする。また、ウェブフロントエンド 2 5 0 は、この情報をクライアント 1 4 0 に示すために、広告可能分類法 5 3 2 領域内のデータにアクセスする。

【 0 0 7 5 】

広告情報 5 3 4 は、システムによって記録された各独自性のある広告が宣伝するものを記述するデータを含む。広告情報 5 3 4 領域のこのテーブルは、広告可能性を広告に関連付けている。例えば、システムは、広告が企業を宣伝していることを示すために、広告可能性の企業種類を特定の広告に関連付けることができる。システムは、広告可能性を広告に関連付けるのに以下の方法を使用する。

【 0 0 7 6 】

1 . 「直接分類」は広告可能性を直接に広告に割り当てる。例えば、メディアエディタ 2 6 4 は、特定の広告が「ホンダ」の広告可能性を宣伝することを特定することによって直接分類を作成する。

【 0 0 7 7 】

2 . 「ロケーション分類」は広告可能性を、システムが広告の目的地を適合するのに使用するロケーションプレフィックスに割り当てる。例えば、メディアエディタ 2 6 4 は、ロケーション「com.honda」がホンダに対する広告を示すことを特定することによって、ロケーション分類を作成する。従って、「com.honda.www/cars」を指示する広告は、広告をホンダに関連付ける。

【 0 0 7 8 】

3 . 「先祖分類」は広告可能性の先祖を広告に割り当てる。例えば、直接分類がホンダを広告に割り当てれば、「自動車」業界の広告可能性はホンダの前任である。先祖分類は、自動車を広告に関連付けるのにこの関係を使用する。

【 0 0 7 9 】

メディアエディタ 2 6 4 は、データベース 2 0 0 に広告情報 5 3 4 を移植する。ユーザインターフェース 2 4 0 は、統計データを生成するために、広告情報 5 3 4 領域のデータにアクセスする。

【 0 0 8 0 】

レートカード 5 3 6 領域は、ウェブサイト上の広告費用を記述するデータを含む。これらのコストは、インターネット 1 0 0 の広告主が広告購入費用を決定するのに使用する特定の形状、大きさ又はランレングス毎に金銭的価値を含む。システムは、システムがプロンプする各ウェブサイトに対して、レートカードデータを格納する。メディアエディタ 2 6 4 は、レートカード 5 3 6 領域をデータベース 2 0 0 に移植する。ユーザインターフェース 2 4 0 は、統計データを生成するために、レートカード 5 3 6 領域内のデータにアクセスする。

【 0 0 8 1 】

拡張された広告可能情報 5 3 8 領域は、分類法階層で直ちに捉えられない特定種類の広告可能性の追加情報を含んでいる。特に、これは、企業連絡情報、ウェブサイト及びメディアキット URL などウェブサイト及び企業に関連する追加情報を含んでいる。この情報は、プロンプされた実体について追加情報をクライアント 1 4 0 に与えることによって、本システムの有用性を拡張する。例えば、クライアント 1 4 0 は、システム報告から直接に企業連絡情報へのハイパーリンクに従うことができる。メディアエディタ 2 6 4 は、データベース 2 0 0 に拡張された広告可能情報 5 3 8 領域を移植する。ウェブフロントエンド 2 5 0 は、価値が付加された情報をクライアント 1 4 0 に届けるために、拡張された広告

10

20

30

40

50

可能情報 5 3 8 領域のデータにアクセスする。

【 0 0 8 2 】

データベース 2 0 0 の広告要約 5 4 0 パーティションは、広告統計 5 4 2、データの完全性 5 4 4 及び広告情報要約 5 4 6 領域に論理的にセグメント分割されたデータベーステーブルを有する。

【 0 0 8 3 】

広告統計 5 4 2 領域は、広告が各ウェブサイトでどの程度頻繁に出現したかを記述する。システムは以下の統計を計算してこの領域に格納する。

【 0 0 8 4 】

1 . 広告を表示するページビューのページビュー総計に対する割合。システムは、プロ
ープ記録を解析することによってこの統計を決定する。 10

【 0 0 8 5 】

2 . 広告が受けた印象の数。システムは、この統計を、サイト定義とトラフィックデータを使用してウェブサイトに対するトラフィックレベルを測定し、その測定に上述の計算されたページビューの割合を掛けることによって決定する。

【 0 0 8 6 】

3 . 広告が受信した支出の量。システムは、レートカード情報を上述の計算された広告が受信した印象の数に適用することによって、この統計を決定する。

【 0 0 8 7 】

統計要約システム 2 3 0 は、データベース 2 0 0 に広告統計領域 5 4 2 を移植する。ウェブ
フロントエンド 2 5 0 は、支出、印象及び広告変化をクライアント 1 4 0 に報告するため
に、広告統計 5 4 2 領域内のデータにアクセスする。 20

【 0 0 8 8 】

データ完全性 5 4 4 領域は、流行及びタイムスライス解析に起因する統計異常値及び他の潜在的な例外についての詳細な情報を含む。この自動化された監視と解析は、システムが正確な解析データを含むであろうことを保証する。また、システムは、分析データの精度を検証するために、システムへの入力として実世界広告情報を使用する。統計要約システム 2 3 0 によって実行されるデータ完全性解析システムは、データ完全性 5 4 4 領域をデータベース 2 0 0 に移植する。オペレータ 2 6 2 は、潜在的誤差を検出して一般的なシステムの健全性を監視するために、データ完全性 5 4 4 領域にアクセスする。 30

【 0 0 8 9 】

広告情報要約 5 4 6 領域は、コンパクトで配布しやすい形式で広告情報をまとめる。システムは、広告サポートデータ 5 3 0 パーティションからこの領域のデータを抽出する。データが広告サポートデータ 5 3 0 パーティションのデータのように記述的でなければ、それは正確なクエリーを高速で実行する能力を提供する。広告サポートデータ 5 3 0 パーティションは、各広告と企業、製品又は業界とを関連付ける。システムが、同一種の多数の広告可能性を広告と関連付けていれば、以下のような割り当て優先システムを使用してそれらの関連物を要約するために、単一の広告可能性が選択される。

【 0 0 9 0 】

1 . 直接分類を使用して広告に関連付けられた広告可能性は、可能な最高の優先権 M を
受ける。 40

【 0 0 9 1 】

2 . ロケーション分類を使用して広告に関連付けられた広告可能性は、それらが割り当てられたロケーションプリフィックスのストリング長と等しい優先権を受けるので、長いロケーションプリフィックスストリングは、短いロケーションプリフィックスストリングよりも高い優先権を受ける。

【 0 0 9 2 】

3 . 先祖分類を使用して広告に関連付けられた広告可能性は、割り当てられた先祖の優先権を受ける。

【 0 0 9 3 】

4. 広告は、各種類で広告可能な最高優先権を受ける。

【0094】

5. 同一種類と優先権を有する2つの先祖が広告に割り当てられた場合には、コンフリクトが発生し、メディアエディタ264によって訂正されなければならない。

【0095】

統計要約システム230は、広告情報要約546領域をデータベース200に移植する。ウェブフロントエンド250は、クライアント140用の報告を作成するために、広告情報要約546領域にアクセスする。

【0096】

以下の記述は、図5に示すデータベース構造の一実施形態を説明している。このデータモデルは、オラクルデータベースにおいて符号化されている。テーブル構造は、3つの環境、コアスキーマ、解析スキーマ及びフロントエンドを有する。コアスキーマはバックエンド環境を説明し、それにより、クラウドプロバーが提供する情報を指示、監視及び増大するために、クラウドプロバーは、ウェブを継続的に調査し、本発明のために、広告活動及びオペレータ及びメディアエディタに留意しながら、当面の匿名処理を指示することができる。解析スキーマは、OMNIA Cとしても知られる広告サンプリングシステムが、ウェブから集められた情報に厳格なデータ解析処理を適用することを可能にするバックエンド環境である。フロントエンドスキーマは、本発明のクライアントが、データにアクセスし、データベースクエリースtringを形成し、報告を作成することを補助する。

【0097】

「コアスキーマ」を有するデータベースオブジェクトは、OMNIA Cシステムの様々なコンポーネントによって最も頻繁に使用される。このスキーマに依存するコードベースは、ウェブから広告を引き出すバックエンドプロセッサの実行を含む。また、OMNIA Cに関連付けられている他のコンポーネントによって利用されるデータベーススキーマは、コアスキーマのテーブルの一部又は全部から構成される。コアスキーマは概念的に、宣伝、広告、プローブ及びサイトを含む4つのサブスキーマから構成される。宣伝サブスキーマは、各広告が宣伝する「広告可能な」実体についての情報を保持する。広告サブスキーマは、システムが突き止めて解析する広告を記述する。プローブサブスキーマは、プローブ処理に対する「いつ」、「どこで」、「どのように」を定義する。サイトサブスキーマは、構造サイト定義及びレートカード情報を含むウェブサイトを記述する。

【0098】

4つのサブスキーマの中で、宣伝は最も汎用に機能し、従って最も頻繁に参照される。このサブスキーマにおける主要テーブルは、ADVERTISABLEであり、それは広告可能性を定義する。OMNIA Cの世界における概念的実体の殆どは広告可能性であり、業界、企業、製品、サービス及びウェブサイトは全てここで定義される。種類フィールドは、ADVERTISABLE__TYPEテーブルを参照しながら、異なる種類の広告可能性を区別し、親フィールドは階層的に記録を整理し、「業界が企業を含む」や「企業が製品を作る」などの関係を設定する。

【0099】

ADVERTISABLEに定義された親子関係によって仄めかされた固有のグループ分けに加えて、ADVERTISABLE__GROUP__MEMBERは広告可能性を更にグループ分けするのに使用される。このようにして定義されたグループの例は、自動車クラス、旅行業界セグメント、コンピュータハードウェアの種類を含む。

【0100】

宣伝サブスキーマの他のテーブルは、各広告によって宣伝されるものを定義するように機能する。ADVERTISEESは広告可能性を直接広告に関連付ける。また、LOCATION__ADVERTISEES、CLASSIFIED__LOCATION及びLOCATION__MATCHESは、広告の目的地口ケーションを介して間接的に広告可能性を広告に関連付ける。

【0101】

10

20

30

40

50

上述の「広告」は、広告サブスキーマの主要テーブルであるADの記録への参照である。広告サブスキーマは、OMNIACの世界の各広告を定義する機能を有する。あらゆる独自性のある広告が、AD_DEFINITIONに一以上の記録と、ADに記録を有する。広告定義は、OMNIACがウェブから受信した独自のXML断片である。広告は、自動分類中に等価であると判断された広告定義のセットによって定義された独自の広告である。

【0102】

広告の他のテーブルは、AD及びAD_DEFINITIONによって参照される広告属性を含む。AD_TECHNOLOGYは、広告をなすのに使用される既知のウェブ技術を記述し、TEXTはある広告に対してテキストコンテンツを記述する。FUZZY_WEB_LOCATIONは、広告に発見されたファジーロケーションを含む。ファジーロケーションは、アンカーや画像など、システムによって処理される必要があるURLである。一旦OMNIACがファジーロケーションをロードすると、URLが画像を参照するとMIME_CONTENTへの参照がなされ、URLが別のHTMLページを参照すればDEST_WEB_LOCATIONがなされる。

10

【0103】

続いて、プローブサブスキーマは、OMNIACのプローブ及び広告抽出コンポーネントの行為を制御する。このスキーマの主目的は対象セットを定義することである。対象セットは、ある間隔で一セットのページをフェッチすることをOMNIACに命令する概念的構造であり、抽出規則と呼ばれる一連の規則を使用して広告を抽出する。各対象セットは、TARGET_SETの列に定義される。

20

【0104】

各対象セットを作成する、頻度、位置及び抽出規則は、STROBE、AD_WEB_LOCATION及びEXTRACTION_RULEにそれぞれ定義される。これらのテーブルにおける列の多対多の関係は、TS_RUNS_AT、TS_PROBES及びTS_APPLIESに規定される。

【0105】

第4及び第5のサブスキーマは、ウェブサイトについての情報を単に記録するサイトである。広告可能階層において定義された各サイト又はサブサイトは、SITE_DOMAIN及びSITE_MONTHLY_DATAの幾つかの列と共に、SITE_INFOに対応する記録を有する。SITE_DOMAINは、包括的又は排他的なURLシステムに関してサイトの物理的構造を記述する。SITE_MONTHLY_DATAは、月毎に各サイトに対して広告レートカード、第三者トラフィック見積もり、及び、キャッシュ統計を記録する。

30

【0106】

解析スキーマは、OMNIACの解析モジュールによって移植された幾つかの追加テーブルを含むコアスキーマへの拡張である。解析モジュールは、コアスキーマに保持された情報処理を担当のユニットであり、宣伝活動を正確に記述する細身のデータセットを作成する。

【0107】

コアスキーマと同様に、解析スキーマは、共通の機能を実現するテーブルからなる4つの概念サブスキーマから構成される。これらのサブスキーマは、宣伝分解、広告ビュー要約、スロット統計及びサイト統計を含む。宣伝分解サブスキーマは、属性と広告が宣伝しているものを含む、システム内の各広告についての情報を保持する。広告ビュー要約サブスキーマは、広告ビューを要約し、一日のうちで各広告が各スロットで見られた回数を記録する。スロット統計サブスキーマは、各期間中の各スロットに対する広告変化を記述する。サイト統計サブスキーマは、各期間中の広告変化を含むサイト情報を記述する。

40

【0108】

広告分解サブスキーマにおける主要テーブルは、広告属性を記述する非正規化された記録を含むAD_INFOである。AD_INFO記録は、ADテーブルのIDから入力され

50

る。AD__INFO記録は、完全に分類されて有効な広告を表す各AD記録に対して存在する。AD__INFOは、コアスキーマテーブルADVERTISES及びLOCATION__ADVERTISESに記載された宣伝関係から解析モジュールによって移植される。

【0109】

広告によって宣伝されているものを特定するAD__INFOのフィールドは、CATEGORY（業界）、ORGANIZATION（企業）、ORGANIZATION__GROUP（業界セグメント）ORGANIZATION__OVERGROUP、COMMODITY（製品/サービス）、COMMODITY__GROUP（製品/サービスセグメント）、COMMODITY__OVERGROUP及びMESSAGEである。

10

【0110】

また、AD__INFOは、幾つかの非宣伝属性を記述するフィールドも含む。AD__SLOT__TYPE.IDを参照するFORMATは広告の形式因子を特定する。AD__TECHNOLOGY2.IDを参照するTECHNOLOGYは、広告を実現するのに使用される技術を特定する。DEFINITION、IMAGE及びDESTINATIONは、広告に関連付けられたAD__DEFINITION、IMAGE及びDEST__WEB__LOCATION記録を特定する。これらの3つのフィールドは、ADテーブルのミラーフィールドである。

【0111】

宣伝分解スキーマは、AD__INFOに加えて幾つかのテーブルを含んでいる。ADV__IMPLICATIONは、ADVERTISABLEの階層に由来する広告可能関係のキャッシュである。これは、解析モジュールの動作を促進するのに使用される。AD__INFO__FLATTENEDは、ADVERTISABLEを参照するAD__INFOのフィールドの各々に対して広告/広告可能性のペアを含むAD__INFOのより容易にクエリーされるバージョンである。最後に、AD__TECHNOLOGY2は、フロントエンドにおいてユーザに提供可能で、解析モジュールによって理解された広告技術を記述する。

20

【0112】

広告ビュー要約サブスキーマは、単一のテーブルPLACEMENT__SUMMARYをカバーする。PLACEMENT__SUMMARYは、日、広告及びスロットから入力され、CNTフィールドに、広告が特定の日にスロットで見られた回数を含む。

30

【0113】

解析モジュールは、nが問題の日のIDである場合に、その一つが日毎に存在し、APD__nテーブルに記録されたヒットを集めることによってPLACEMENT__SUMMARYを移植する。これらのテーブルは、広告ヒットがシステムに流れるとバックエンドによって作成及び移植される。

【0114】

解析スキームにおける第3のサブスキームは、スロット統計である。このサブスキームは、スロットそれ自体の情報に加えて、スロットに関する広告行為を記述する。スロットは、広告が変化し、ロケーションID（AD__WEB__LOCATED.IDへの参照）と抽出規則ID（EXTRACTION__RULE.IDへの参照）によって現在定義されているウェブ上のロケーションである。

40

【0115】

スロット統計の主要テーブルは、ビュー合計と各期間中の各スロットにおける各広告に対する相対頻度SLOT__AD__VIEWSである。このテーブルの主要な鍵は、PERIOD__TYPE、PERIOD、LOCATION__ID、RULE__ID及びAD__IDのフィールドから構成されている。2つのフィールドは、主要な鍵の外に存在する。CNTは広告ビューの総数を保持し、FREQUENCYは相対頻度を保持する。

【0116】

また、このサブスキーマには、個々の広告の範囲外で一般スロット情報を記録するSLO

50

T_SUMMARYも存在する。そこで、このテーブルは、PERIOD_TYPE、PERIOD、LOCATION_ID及びRULE_IDフィールドから入力される。CNTフィールドは、スロットの広告ビュー総計を記録する。このフィールドは、相対頻度を決定するために、SLOT_AD_VIEWS.CNTに分割される。また、SLOT_SUMMARYにあるのは、スロットで最も頻繁に見られる広告の種類を特定するSLOT_TYPEフィールドと、スロットが置かれているのはどのサイトであるかを特定するSITE_IDである。

【0117】

スロット統計サブスキームの最終テーブルは、SLOT_TYPE_COUNTである。このテーブルは、SLOT_SUMMARY.SLOT_TYPEに使用されるべき値を決定するのに使用される。各広告様式が見られた回数は記録され、最も多くのビューを受信するスロットの種類がSLOT_SUMMARY.SLOT_TYPEに差し込まれる。

10

【0118】

図6は、広告普及システム130の機能ブロック図である。広告普及システム130のメモリ610は、本発明に従って、インターネット100上のトラフィックデータを解析し、そのトラフィックデータから広告データをサンプリングし、広告データを特徴づける要約データを生成するソフトウェアコンポーネントを格納する。システムバス612は、広告普及システム130のメモリ610をトランスミッションコントロールプロトコル/インターネットプロトコル(TCP/IP)ネットワークアダプタ614、データベース200及び中央プロセッサ616に接続する。TCP/IPネットワークアダプタ614は、広告普及システム130とインターネット100の間のネットワークトラフィックの通過を容易にする。中央プロセッサ616は、メモリ610に格納された、プログラムされた命令を実行する。

20

【0119】

図6は、オブジェクトモデルとして構成された広告普及システム130の機能モジュールを示す。オブジェクトモデルは、オブジェクト指向ソフトウェアプログラムを、広告普及システム130の主要な機能とアプリケーションを実行するコンポーネントにグループ化する。図6のオブジェクト指向ソフトウェアプログラムコンポーネントを適当に実行するために、エンタープライズ・ジャバビーンズ仕様を使用することができる。「ビルディング・ジャバ・エンタープライズ・システム・ウィズ・J2EE(Building Java Enterprise Systems with J2EE)」(サムズパブリッシング(Sams Publishing)、2000年6月)と題されたポール・ジェイ・ペロン(Paul J. Perrone)等による本は、エンタープライズ・ジャバビーンズ仕様を使用して開発されたジャバエンタープライズアプリケーションの説明を与える。「ビギニング・Eコマース(Beginning E-Commerce)」(ロックスプレス社(Wrox Press Inc.)、2000年)と題されたマチュー・レイノルド(Matthew Reynolds)による本は、電子商取引の用途に対するウェブサーバーの設計におけるオブジェクトモデルの使用の説明を与える。

30

【0120】

広告普及システム130のメモリ610に対するオブジェクトモデルは、プレゼンテーション層620、インフラストラクチャオブジェクトパーティション630、及び、ビジネス論理層640を含む三層アーキテクチャを採用している。オブジェクトモデルは、ビジネス論理層640を、アプリケーションサービスオブジェクトパーティション650とデータオブジェクトパーティション660という2つのパーティションに更に分割する。

40

【0121】

プレゼンテーション層620は、広告普及システム130と、クライアント140、アカウントマネージャー260、オペレータ262及びメディアエディタ264との間のインターフェースを管理するプログラムを保持する。図6において、プレゼンテーション層620は、TCP/IPインターフェース622、ウェブフロントエンド624及びユーザ

50

インターフェース 626 を含む。プレゼンテーション層 620 を適当に実現するために、ハイパーテキストトランスファープロトコル (HTTP) を介して本発明のクライアント 140、アカウントマネージャー 260、オペレータ 262 及びメディアエディタ 264 と対話するのにジャバサーブレットを使用することができる。ジャバサーブレットは、クライアント 140、アカウントマネージャー 260、オペレータ 262 及びメディアエディタ 264 からの要求メッセージを取り扱い、クライアント 140、アカウントマネージャー 260、オペレータ 262 及びメディアエディタ 264 に応答メッセージを返却する要求/応答サーバー内で起動する。ジャバサーブレットは、ウェブサーバー環境内で起動するジャバプログラムである。ジャバサーブレットは、入力としての要求を受け取り、データをパースし、論理演算を実行し、クライアント 140、アカウントマネージャー 260、オペレータ 262 及びメディアエディタ 264 に戻す応答を発行する。ジャバランタイムプラットフォームは、多くの要求に同時にサービスを与えるためにジャバサーブレットをプールする。ジャバサーブレットを使用する TCP/IP インターフェース 622 は、HTTP プロトコルを利用してクライアント 140、アカウントマネージャー 260、オペレータ 262 及びメディアエディタ 264 と通信するウェブサーバーとして機能する。TCP/IP インターフェース 622 は、HTTP 要求をクライアント 140、アカウントマネージャー 260、オペレータ 262 及びメディアエディタ 264 から受領し、その要求の情報をビジネス論理層 640 内のビジットオブジェクト 642 に渡す。ビジットオブジェクト 642 は、ビジネス論理層 640 から戻される結果情報を TCP/IP インターフェース 622 に渡す。TCP/IP インターフェース 622 は、これらの結果を、クライアント 140、アカウントマネージャー 260、オペレータ 262 及びメディアエディタ 264 に HTTP 応答において返信する。TCP/IP インターフェース 622 は、TCP/IP ネットワークアダプタ 614 を介してインターネット 100 とデータを交換する。

10

20

【0122】

インフラストラクチャオブジェクトパーティション 630 は、ビジネス論理層 640 の代わりに、管理及びシステム機能を実行するプログラムを保持する。インフラストラクチャオブジェクトパーティション 630 は、オペレーティングシステム 636 とデータベース管理システム (DBMS) インターフェース 632、管理者インターフェース 634、ジャバランタイムプラットフォーム 638 に対するオブジェクト指向ソフトウェアプログラムコンポーネントを含む。

30

【0123】

ビジネス論理層 640 は、本発明の実体を実行するプログラムを保持する。図 6 のビジネス論理層 640 は、ビジットオブジェクト 642 の多数の例を含む。ビジットオブジェクト 642 の別個の例は、TCP/IP インターフェース 622 を介してウェブフロント 624 又はユーザインターフェース 626 のいずれかによって開始される各クライアントセッションに対して存在する。各ビジットオブジェクト 642 は、単なる単一のインタラクティブセッション又はメソッドコール中ではない開始からクライアントセッションの終了を通じてパーシステント記憶領域を含むステートフルセッションである。パーシステント記憶領域は、URL 114、116、118 又はクライアント 140、アカウントマネージャー 260、オペレータ 262 及びメディアエディタ 264 のいずれかに関連付けられた情報を保持する。また、パーシステント記憶領域は、データベース 200 クエリーからのクエリー結果セットなどの TCP/IP インターフェース 622 を介して、広告普及システム 130 とトラフィックサンプリングシステム 120 との間で交換されるデータを保持する。

40

【0124】

トラフィックサンプリングシステム 120 が URL 114、116、118 についての情報収集を終了すると、それはメッセージを TCP/IP インターフェース 622 に送信し、ビジットオブジェクト 642 を作成する方法を喚起してビジットオブジェクト 642 状態の接続に関する情報を格納する。次いで、ビジットオブジェクト 642 は、トラフィックサンプリングシステム 120 によって検索された情報を処理するために、トラフィック

50

解析アプリケーション 6 5 2 の方法を喚起する。トラフィック分析アプリケーション 6 5 2 は、匿名システム 3 1 0 及びプローブマッピングシステム 3 2 0 からの処理されたデータを、トラフィック解析データ 6 2 2 状態及びデータベース 2 0 0 に格納する。図 7 A 及び 7 B は、より詳細に、トラフィックサンプリングシステム 1 2 0 から取得した各 URL 1 1 4、1 1 6、1 1 8 に対してトラフィック解析アプリケーション 6 5 2 が従う処理を説明している。図 6 は中央プロセッサ 6 1 6 がトラフィック解析アプリケーション 6 5 2 を制御するものとして示しているが、トラフィック解析アプリケーション 6 5 2 によって実行される機能が広告普及システム 1 3 0 に類似して構成された別個のシステムに配布可能であることが理解されるであろう。

【 0 1 2 5 】

トラフィック解析アプリケーション 6 5 2 がトラフィックサンプリングシステム 1 2 0 によって識別された URL 1 1 4、1 1 6、1 1 8 を処理した後で、ビットオブジェクト 6 4 2 は、ウェブサイト 1 1 0 から URL 1 1 4、1 1 6、1 1 8 を検索するために広告サンプリングアプリケーション 6 5 4 の方法を喚起する。広告サンプリングアプリケーション 6 5 4 は、埋め込まれた広告を抽出してそれらの広告を分類することによって、検索されたウェブページを処理する。広告サンプリングアプリケーション 6 5 4 は、ウェブページ検索システム 3 2 2 によって検索され、ウェブブラウザエミュレーション環境 3 2 4、広告抽出装置 3 2 6 及び構造分類装置 3 2 8 によって処理されたデータを広告サンプリングデータ 6 6 4 状態及びデータベース 2 0 0 に格納する。図 7 A、7 C、7 D 及び 7 E は、より詳細に、トラフィックサンプリングシステム 1 2 0 によって識別された各 URL 1 1 4、1 1 6、1 1 8 に対して広告サンプリングアプリケーションが従う処理を説明している。図 6 は、中央プロセッサ 6 1 6 が広告サンプリングアプリケーション 6 5 4 を制御するものとして示しているが、当業者は、広告サンプリングアプリケーション 6 5 4 によって実行される処理が広告普及システム 1 3 0 に類似の構成を有する別個のシステムに分配可能であることを理解するであろう。

【 0 1 2 6 】

トラフィック解析アプリケーション 6 5 2 と広告サンプリングアプリケーション 6 5 4 がトラフィックサンプリングシステム 1 2 0 によって識別された URL 1 1 4、1 1 6、1 1 8 を処理した後で、ビットオブジェクト 6 4 2 は、データに対する要約統計を計算するための、統計要約アプリケーション 6 5 6 内の方法を喚起する。統計要約アプリケーション 6 5 6 は、URL 1 1 4、1 1 6、1 1 8 に埋め込まれた各広告に対して広告印象、支出及び査定統計を計算する。統計要約アプリケーション 6 5 6 は、統計データを統計要約データ 6 6 6 状態及びデータベース 2 0 0 に格納する。図 7 E は、トラフィックサンプリングシステム 1 2 0 によって識別された各 URL 1 1 4、1 1 6、1 1 8 に対して、統計要約アプリケーション 6 5 6 が従う処理をより詳細に説明している。図 6 は中央プロセッサ 6 1 6 が統計要約アプリケーション 6 5 6 を制御するものとして示しているが、当業者は、統計要約アプリケーション 6 5 6 によって実行される機能が広告普及システム 1 3 0 に類似の構成を有する別個のシステムに分配可能であることを理解する。

【 0 1 2 7 】

図 7 A は、全ての主要な業界、チャネル、宣伝形式及び種類に亘ってオンライン宣伝活動を追跡及び比較することによってオンライン広告の価値を測定する広告普及システム 1 3 0 の処理のフローチャートである。処理 7 0 0 は、インターネット 1 0 0 からのトラフィックデータをサンプリングすることによってステップ 7 1 0 で開始する。図 7 B は、ステップ 7 1 0 をより詳細に説明するものである。ステップ 7 2 0 は、サイト選択を実行し、広告普及システム 1 3 0 のためのサイト定義を定義及び精緻するのにサンプリングされたトラフィックデータを使用する。ステップ 7 3 0 は、サンプリングされたトラフィックデータに基づいてプローブマップを生成するのにサイト選択と定義処理の結果を使用する。図 7 C は、より詳細にステップ 7 3 0 を説明している。ステップ 7 4 0 は、ステップ 7 3 0 で識別されたプローブサイトからサンプルデータを収集するため、インターネット 1 0 0 を訪れるのにステップ 7 3 0 からのプローブマップを使用する。図 7 D は、ステップ 7

10

20

30

40

50

40を詳細に説明している。ステップ740で検索された各URLに対して、ステップ750はURLから広告を抽出し、ステップ760は各広告を分類し、ステップ770は各広告からの統計を計算する。図7E及び7Fは、それぞれ、ステップ760及び770をより詳細に説明している。最後に、処理700は、システム内のデータの完全性と解析結果を検証するために、データの完全性のチェックをステップ780において実行する。

【0128】

図7Bは、図7Aのステップ710からのトラフィックデータをサンプリングする処理をより詳細に説明するフローチャートである。処理710は、トラフィックサンプリングシステム120などのウェブトラフィック監視からデータを収集することによってステップ711で開始する。処理710は、データを清浄してサンプルの匿名性を保証するために、ステップ712においてウェブトラフィック監視によって検索されたデータからユーザ情報を除去する。清浄されたサンプルにおける各URLに対して、ステップ713は、トラフィックデータにおいて観察されたウェブページのビューの数を測定する。ステップ714は、インターネット100の全世界までサンプルのウェブページビューの測定された数を統計的に推定することによって、処理710を想定している。

10

【0129】

図7Cは、図7Aのステップ730からのサンプリングされたトラフィックデータに基づいてプローブマップを生成する処理をより詳細に説明するフローチャートである。処理730は、適当なサイト定義に該当するサンプルトラフィックデータの部分集合を解析することによって、ステップ731で開始する。ステップ731の解析に続いて、ステップ732はサンプリングされたトラフィックデータに基づいて初期プローブマップを構築する。ステップ733は、初期プローブマップにおけるURLに対してデータベース200内の過去の広告測定結果を解析する。ステップ734は、サンプリング手順を最適化するシステムパラメータと共に、これらの履歴結果を使用する。ステップ735は、サンプル結果を監視してシステムを必要に応じて調節することによって、処理730を完了する。

20

【0130】

図7Dは、図7Aのステップ740からのサンプルデータを収集するためにインターネット100をプローブする処理をより詳細に説明するフローチャートである。処理740は、インターネット100からのウェブページのフェッチによってステップ741で開始する。ステップ741からのウェブページは、ブラウザ内でのウェブページの表示をシミュレートするために、ステップ742でウェブページブラウザエミュレーション環境に渡される。このシミュレーションにより、広告普及システム130は、ウェブページに埋め込まれた広告を検出することができる。これらの広告は、ジャバスクリプトコード、ジャバアプレット又はサーブレットコード、若しくは、パールスクリプトなどの共通ゲートウェイインターフェースコードで埋め込み可能である。また、ステップ742のシミュレーションにより、広告普及システム130は、ウェブページ内で動的及びインタラクティブな広告を検出することができる。ステップ742のシミュレーション後に、ステップ743は広告データをウェブページから抽出し、ステップ744は広告データをデータベース200に格納する。ステップ745は、プロセス740がより多くのサンプルデータを収集するために別のウェブページをフェッチする必要があるかどうかを判断する。好ましい実施形態においては、処理740は、継続的にウェブページをインターネット100からサンプリングする。当業者は、ステップ745によって実行される機能は、サンプル宣伝データを収集するためにインターネット100のプローブをスケジューリングするスケジューリングシステムに関連付けられることができることを理解する。

30

40

【0131】

図7Eは、図7Aのステップ760からの宣伝データを分類する処理をより詳細に説明するフローチャートである。処理760は、断片が重複するかどうかを判断することによって、ステップ761の広告断片の解析を開始する。広告普及システム130が最初に広告断片に遭遇すると、ステップ762は、断片の内部構造を解析する。ステップ762に続いて、又は、ステップ761が広告断片が重複であると判断すると、ステップ763はウ

50

ウェブページから広告の外部コンテンツを検索する。ステップ764は、その後、外部コンテンツを以前に観測された広告と比較する。ステップ765は、広告が重複するかどうかを判断するために、ステップ764の比較の結果を解析する。広告普及システム130が広告に最初に遭遇すると、ステップ766は、データベース200に新しい広告の構造を記録することによって新しい広告を処理することを開始する。ステップ767は、その後、自動広告分類を実行し、分類の種類をデータベース200に格納する。ステップ768は、広告分類の人的検証を実行することによって、新しい広告の処理を完了する。ステップ768に続いて、又は、ステップ765が広告が重複すると判断すると、ステップ769は、広告の観測を示すために、データベース200の広告ビューログを更新する。

【0132】

図7Fは、図7Aのステップ770からの広告統計を計算する処理をより詳細に説明するフローチャートである。処理770は、広告測定結果を要約することによって、ステップ771において広告統計の計算を開始する。ステップ772において、処理770は、広告測定結果を重み付けするのにステップ730で生成されたプローブマップを使用する。広告頻度は、各ウェブページ要求に対してステップ773で計算される。ステップ774は、各広告に対して広告印象を計算するのに、ステップ710からのサンプルトラフィックデータ及びステップ773からの広告頻度を使用する。ステップ775は、ステップ774からの広告印象とメディアエディタ264によって入力されたレートカードデータとをユーザインターフェース240のレートカード収集348モジュールに組み合わせて広告支出を計算することによって、処理770を完了する。

【0133】

本発明において開示された実施形態は十分に機能するシステムを説明しているが、ここで開示された実施形態に均等である他の実施形態が存在することが理解されるであろう。本出願を検討した者には無数の変形及び変更が起こるのであるから、本発明はここで図示及び説明された構成及び動作そのままに限定されるものではない。従って、全ての適当な変形及び訴え得る均等物が特許請求の範囲に該当することが企図されている。

【図面の簡単な説明】

【図1】 本発明による広告普及システムの環境を示すネットワーク構成図である。

【図2】 図1のネットワーク構成図をより詳細に示し、広告普及システムを有するネットワーク環境と構成要素との間の関係を示している。

【図3】 図2のネットワーク構成図をより詳細に示し、広告普及システムを有する構成要素と準構成要素との間の関係とネットワーク環境への接続を示している。

【図4A】 広告普及統計の計算に使用される期待値を示す例示的なウェブサイトである。

【図4B】 広告普及統計の計算に使用される観測値を示す例示的なウェブサイトである。

【図4C】 広告普及統計の計算に使用される加重値を示す例示的なウェブサイトである。

【図4D】 広告普及統計の代替的方法を示す例示的なウェブサイトである。

【図5】 広告普及システムが使用可能なデータベース構造の一例を示す。

【図6】 ハードウェアとソフトウェアコンポーネントの構成を示す広告普及システムの機能ブロック図である。

【図7A】 オンライン広告によって生成されるオンライン広告と活動の品質を測定する広告普及システムの処理のフローチャートである。

【図7B】 図7Aからのトラフィックデータをサンプリングする処理をより詳細に説明するフローチャートである。

【図7C】 図7Aからのサンプリングされたトラフィックデータに基づいてプローブマップを生成する処理をより詳細に説明するフローチャートである。

【図7D】 図7Aからのサンプルデータを収集するためにインターネットをプローブする処理をより詳細に説明するフローチャートである。

10

20

30

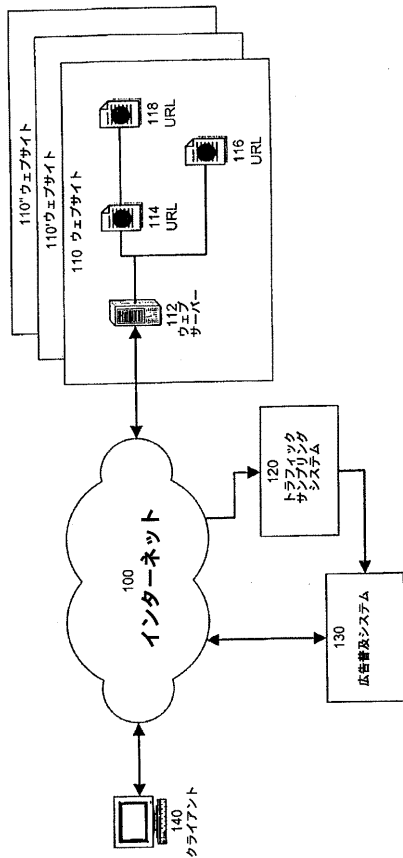
40

50

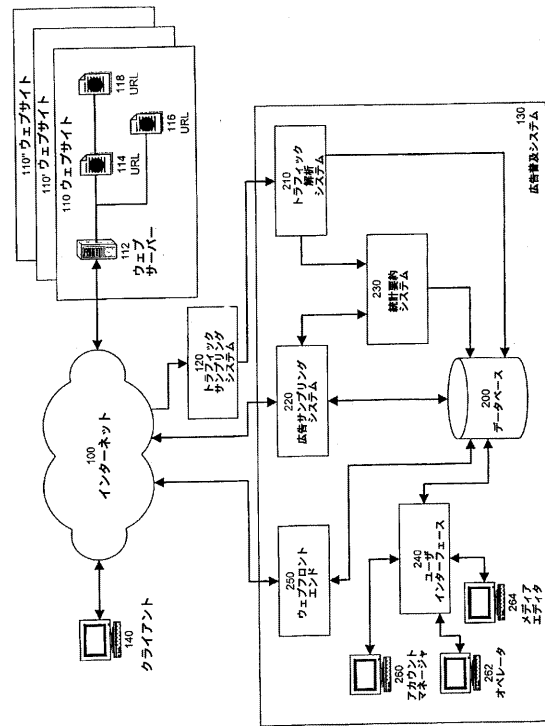
【図7E】 図7Aからの広告データを分類する処理をより詳細に説明するフローチャートである。

【図7F】 図7Aからの広告統計を計算する処理をより詳細に説明するフローチャートである。

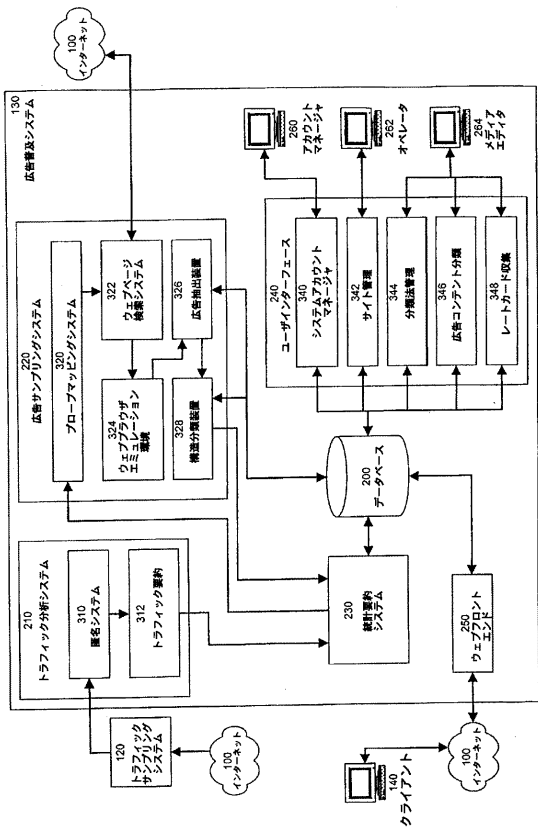
【図1】



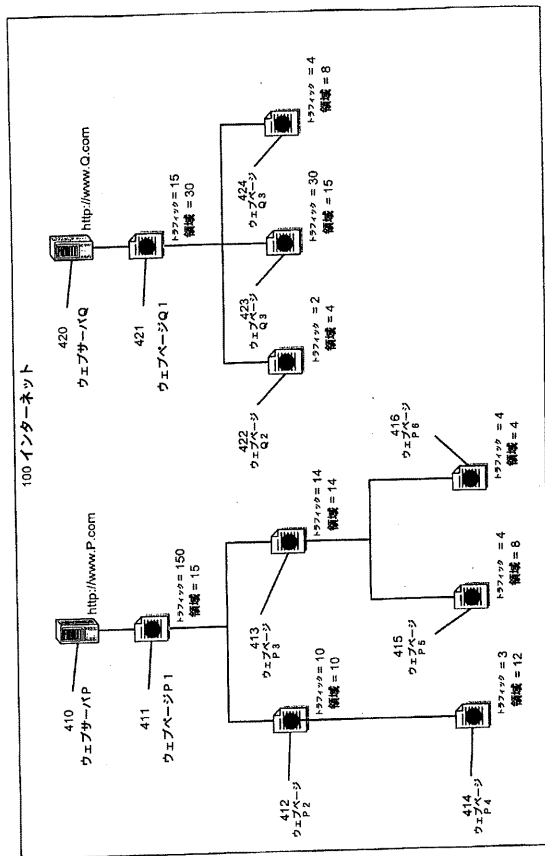
【図2】



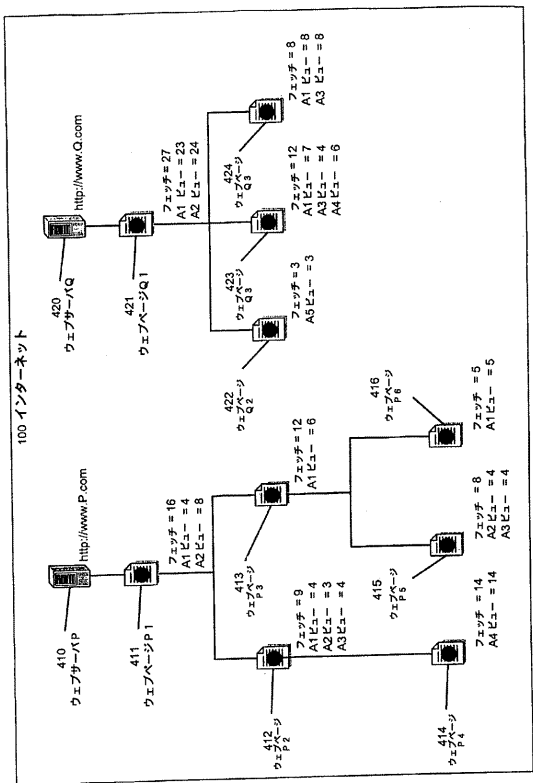
【図 3】



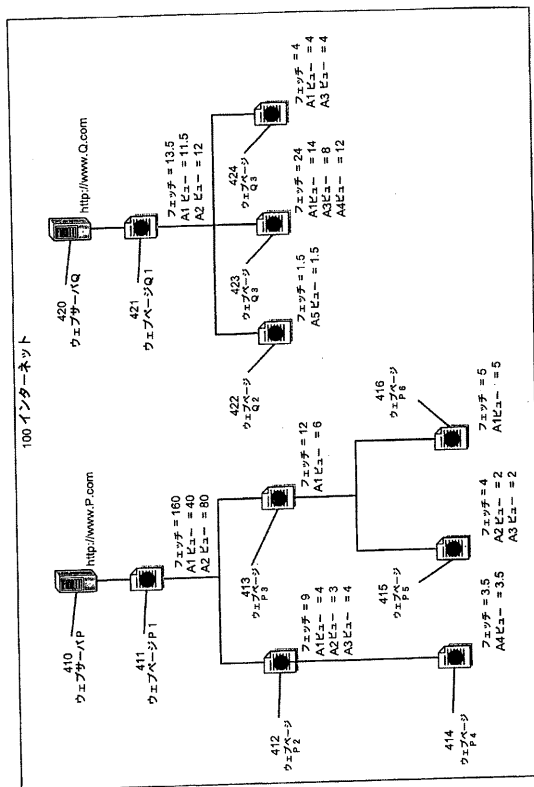
【図 4 A】



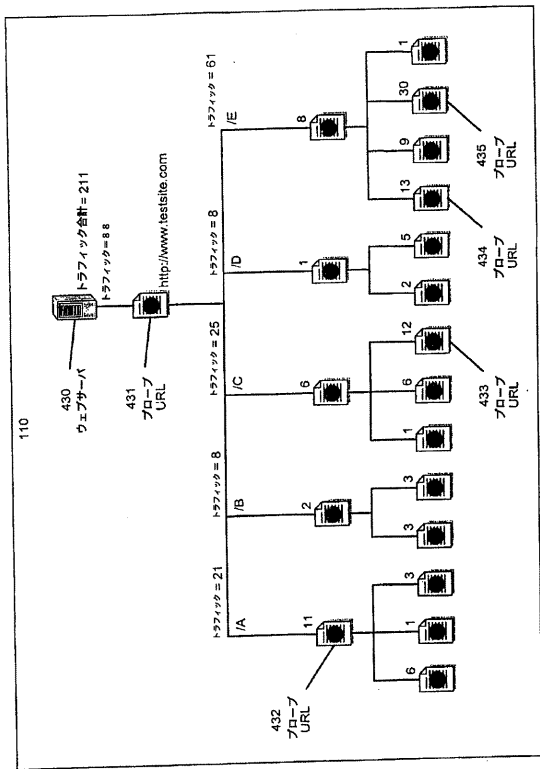
【図 4 B】



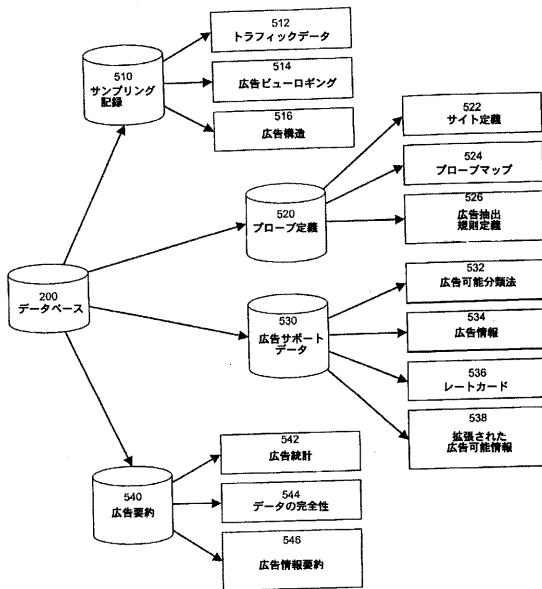
【図 4 C】



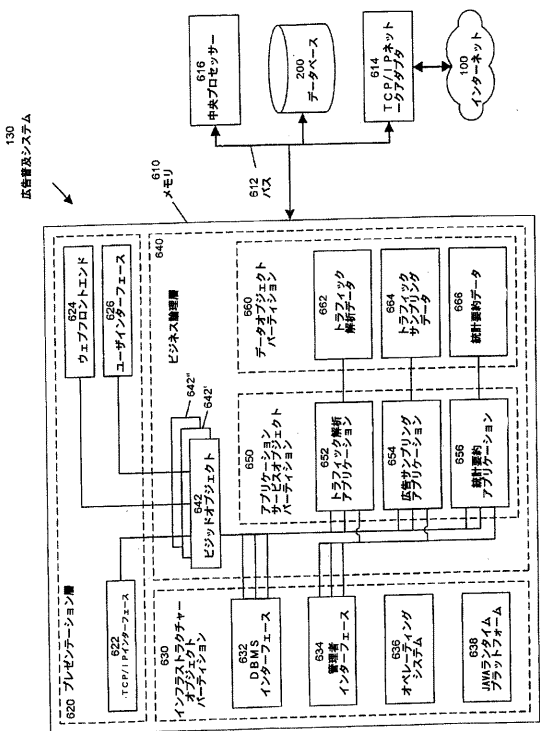
【図4D】



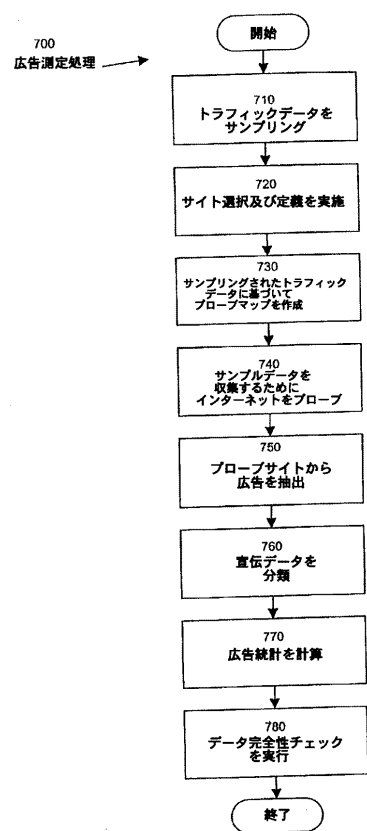
【図5】



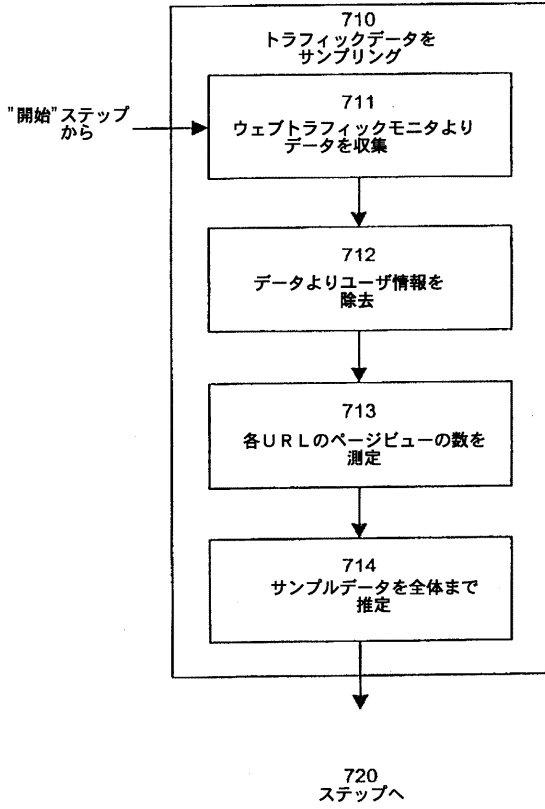
【図6】



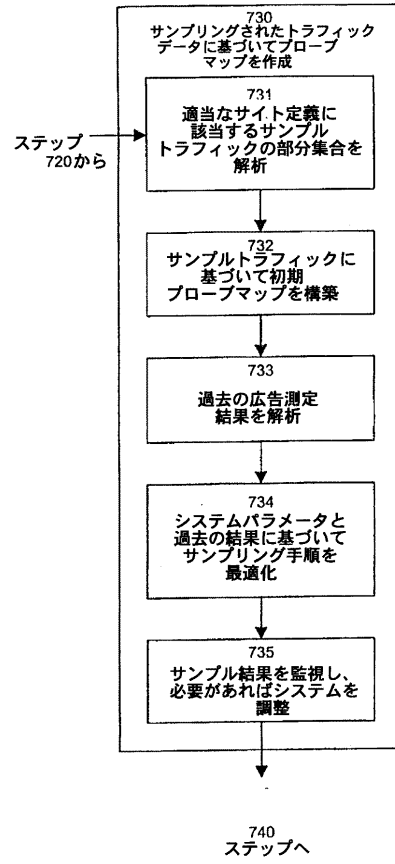
【図7A】



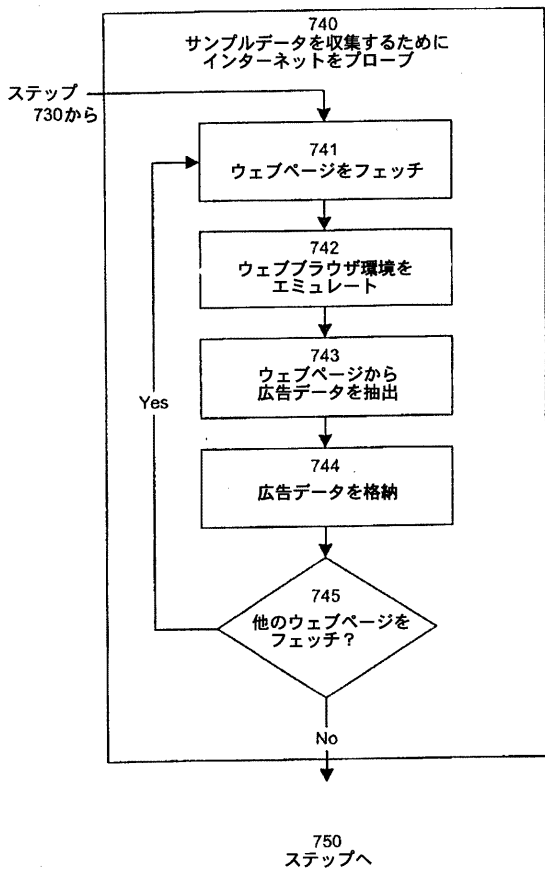
【図7B】



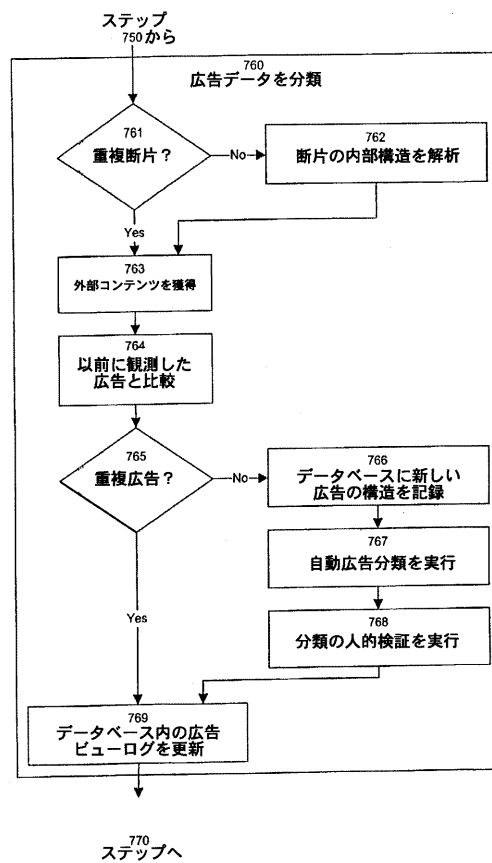
【図7C】



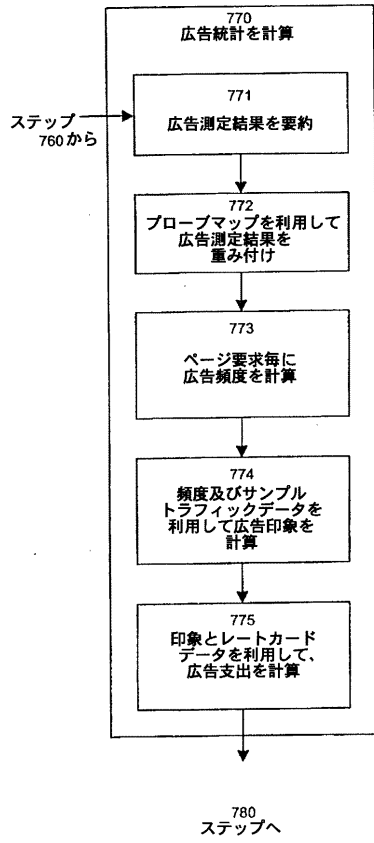
【図7D】



【図7E】



【図7F】



フロントページの続き

(51)Int.Cl. F I
G 0 6 F 19/00 (2011.01) G 0 6 F 17/60 5 0 4
G 0 6 F 17/60 Z E C
G 0 6 F 19/00 1 1 0

(72)発明者 ジェイムス・ティー・パートット
アメリカ合衆国ワシントン州シアトル、ガーラーストリート201

審査官 田上 隆一

(56)参考文献 特開平09-091308(JP,A)
特開平10-124428(JP,A)
特開平11-039258(JP,A)
特開平11-328226(JP,A)
小池 良次, ポストバナー時代到来!?インターネット広告戦線に異状あり, INTERNET
magazine, 日本, 株式会社インプレス, 1999年 9月 1日, No.56, p.
232~235
田中 慎二, マルチメディアビジネスの展開(第23回), ビジネスコミュニケーション, 日本
株式会社ビジネスコミュニケーション, 1998年 2月 1日, 第35巻 第2号, 頁11
6~123

(58)調査した分野(Int.Cl., DB名)
G06F 13/00