



*e-Business*での顧客の 行動パターンを発見しよう！！

SAS Institute Japan Ltd.
Solution Planning Center

東 一成

jpnkaa@jpn.sas.com

有山 茂芳

jpnsba@jpn.sas.com

The Power to Know.



The Power to Know.

*net*の発展

- ◆ 日本のインターネットの利用人口は3263.6万人(全世界では4億人)。
- ◆ ブロードバンド化が進めば、さらに加速することが予想されるが、PCだけではなく、携帯電話やPHSによる接続が増大し、PCを追い抜く勢い。
- ◆ オンラインショッピングは、書籍・旅行/宿泊/チケット・衣料・食品等を中心として増加している。
(インターネット白書2001より)
- ◆ インターネット広告も1999年→2000年で倍増し、携帯向け広告も本格的にスタートしている(2001年の推定市場規模は約1兆円)。
(電通調べ)
- ◆ ISPの加入者数は1289.9万人。
(財団法人ニューメディア開発協会調べ)
- ◆ 電子商取引の市場規模
 - 最終消費財 →約7.9兆円
 - 中間財(B2B) →約100兆円
 - モバイル関連 →約4.6兆円
(総務省情報通信白書調べ)

netの多様性

- ◆ e-Businessの重要なチャネルとしての認識が浸透
 - Web,e-mail,mobileなどのチャネル

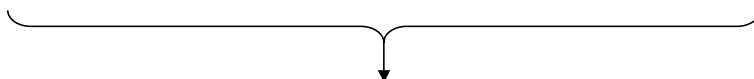
ただし、活かし方に関してはまだ模索状態

- ◆ Market Placeなどのプラットフォームビジネスの台頭
- ◆ 広告収入、通販、中間手数料収入、会員費収入など、ネットを基盤とした様々なビジネスモデルが転換期を迎えている

3

net活用の課題

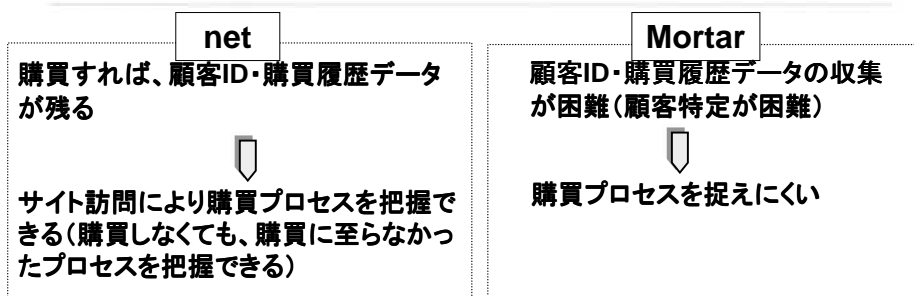
- ◆ netの活用の際し、多くの企業が以下のような課題を抱えている
 - データ収集:どのようにしてデータを得ればよいのか？
 - データ蓄積・管理:いかにして大量のデータを管理するのか？
 - データ分析・活用:新しいチャネルの活かし方は？



それでもnet上におけるデータの分析をするの？

4

netの優位性の源泉



Webログが強力な情報源

Webログの収集・蓄積・分析の仕組み作りが重要

5

Webログって何？

◆ Webログの概要説明

- Webサーバーに日々蓄積される、ユーザーがサイトにどのように訪れどのようなファイルを要求したかの記録。
- Webサーバーによりファイル形式は数種類ある
- ファイル形式によって残せるログの内容が変わる
- 膨大かつ冗長な構造で、必要な情報を得ることは困難

```
1999-11-07 00:11:46 161.69.245.198 - W3SVC1 WEB3 161.69.5.90 GET /Default.asp - 200 0 262 348 0 80
HTTP/1.1 Mozilla/4.0+(compatible;+MSIE+4.01;+Windows+NT;+NAI+Internal+Build+5) - -
1999-11-07 00:11:49 161.69.245.198 - W3SVC1 WEB3 161.69.5.90 GET /index.asp - 200 0 6660 357 0 80
HTTP/1.1 Mozilla/2.02+(X11;+I;+SCO_SV+3.2+i386) - -
1999-11-07 00:11:51 161.69.245.198 - W3SVC1 WEB3 161.69.5.90 GET /ssi/corner.jpg - 304 0 141 364 15
80 HTTP/1.1 SPRY_Mosaic/v7.36+(Windows+16-bit)+SPRY_package/v4.00
- http://internal.nai.com/index.asp
1999-11-08 00:51:07 161.69.245.57 north_america#dwallac W3SVC1 WEB6 208.228.228.222 GET
/ssi/bullets.gif - 200 0 5286 536 484 80 HTTP/1.1
Nokia-Communicator-WWW-Browser/3.0+(Geos+3.0+Nokia-9110) - http://internal2.nai.com/cgi-
bin/pricing?TEMPLATE=applicant/htdocs/person-show-pro.html
```

6

「e」で顧客を捉える 個人データとWebアクセスログデータの 結合方法の考え方 1

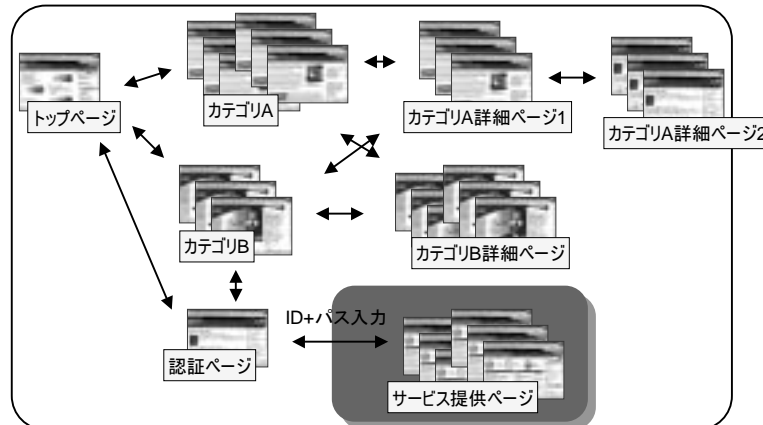
- ◆ サイト構成の一般例



パブリックなサイト



プライベートなサイト



7

「e」で顧客を捉える 個人データとWebアクセスログデータの 結合方法の考え方 2

- ◆ 顧客識別の仕組みとそれぞれの特徴

1. Cookiesを使用する

- Webサーバーで設定可能
- 精度の高いセッション管理
- CookieとIDが1対1で一致しない場合がある(顧客特定の精度)

例: 会社のマシンと家のマシン使い分け
同じマシンを家族(複数人)で共有
Cookiesファイルを削除される可能性
ブラウザによるCookies使用禁止設定
マシンを買い換えたら?
OSを再インストールしたら?

↓
全体の傾向を見る

パブリックサイト・無料サービス向け

2. IDとパスワードを入力させる

- Webサーバーだけでは構築不可能
- 精度の高いセッション管理
- 完全に顧客を特定できる
- ユーザーに負担を強いる

↓
特定のサイトにおける、
顧客を特定した動きを見る

プライベートサイト・有料サービス向け

8

「e」で顧客を捉える 個人データとWebアクセスログデータの 結合方法の考え方 3

◆ Cookiesとは？

- Webサーバーで発行するテキストファイル
 - Webサーバーでテキスト設定可能
(IDとして使用する)
 - ユーザーのブラウザに保存
(InternetExplorer、NetscapeNavigator、それぞれに保存)
 - ブラウザに保存しておくタイプ(SITESERVERID)と、
サイトのセッション終了時に削除するタイプ(ASPSESSIONID)がある
 - サイト単位で発行し、他のサイトでは使用できない
 - ブラウザに保存する(書き込む)のは許可が必要だが、
一度保存したファイルを見るのに許可はいらない
 - Cookiesで得た情報の二次利用には注意が必要

9

「e」で顧客を捉える 個人データとWebアクセスログデータの 結合方法の考え方 4

◆ 結合のためのデータ取得イメージ

- 既存ユーザーによるパブリックサイト(Myページ)へのアクセス

- ① ブラウザ → Webサーバー(アクセス)
- ② ブラウザ → Webサーバー(Cookie読込)
- ③ Webサーバー → アプリサーバー(Cookie情報)
- ④ アプリサーバー → DB(Cookie情報)
- ⑤ アプリサーバー ← DB(ユーザー情報)
- ⑥ Webサーバー ← アプリサーバー(コンテンツ情報)
- ⑦ ブラウザ ← Webサーバー(コンテンツ配信)

ログが残るタイミング

- ◆ Webログとアプリログ・DBのデータを紐付けするキーとしてCookies
を使用する



10

192.168.0.1 www.sas.com/jp/products/base.html sasjpnba5317

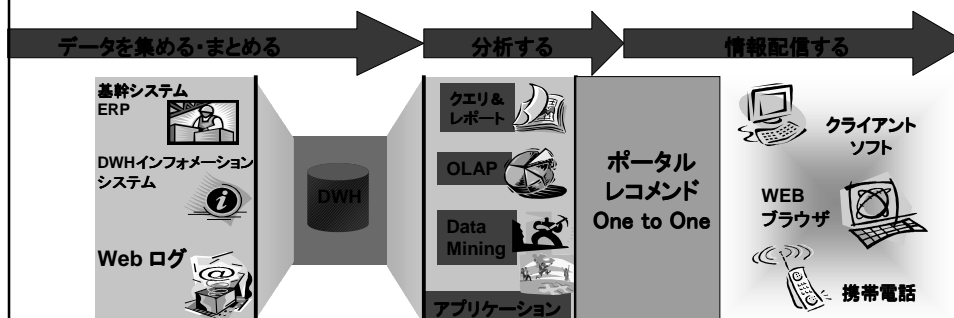
Webログと他のデータソースの統合分析環境

- それぞれ単独でもわかることは多い
 - ◆ Webログ・・・サイト全体の傾向を把握する(自社のことを知る)
 - ◆ アプリログ・DB・・・限られたサイトの中での顧客を特定した行動パターンを分析する
- ◆ SASのWebログに対する捉え方
 - ビジネスにおける様々な業務で発生する、多くのデータソースの一つ
 - ◆ Webログだけで分析できることは限られており、他のビジネスデータと連携することでより大きな効果を生む
 - ◆ 何を分析するかにより、どのようなデータを用意するかを判断する事が必要

11

SAS Institute Japanの e-Intelligenceソリューション

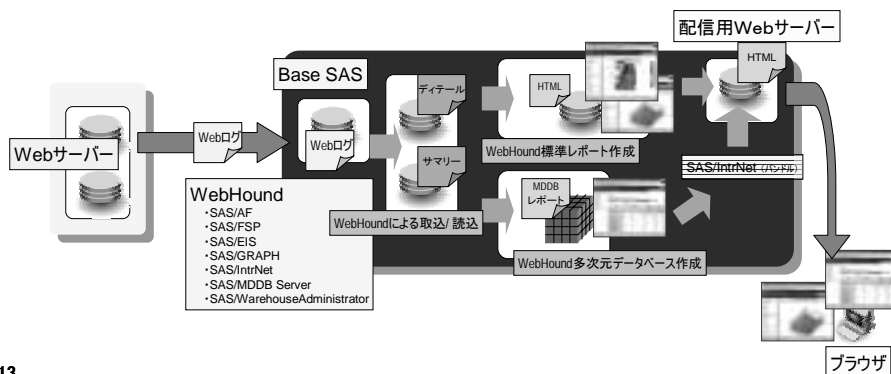
- ◆ e-Intelligenceソリューションの説明
 - Webログの分析
 - データの統合(ETL、DatawareHousing、メタデータ管理)
 - データマイニングによる深い分析



12

WebHoundとは？

- ◆ 日々Webサーバーに蓄積される膨大な量のWebログデータを高速に処理し、分析・レポートニングするツール



13

WebHoundによるWebログ取り込み

- － Webログを効率よく取り込む
 - ◆ 生のWebログを設定したディレクトリに保存しておくだけ
 - ◆ GUI (WebHoundAdministrator) による基本設定画面により、Webログをどのように取り込むかの設定を変更可能
 - ◆ Webログを取り込むためのマクロプログラム (SASのデータステップ) を書き換えることにより、前処理やロジック変更に対応可能

※前処理・ロジック変更例

- － 必要のない項目を取り込まない
- － Webサーバー間のタイムラグを吸収する
- － リクエストファイル中、「index.htm*」、「/」を統合する
- － Webログだけではなく他のアプリケーションのログも取り込む



14

WebHoundによるデータセットへのロード

- Webログをデータセットに加工する
 - ◆ ディテールデータセットとサマリーデータセットの2パターン
 - ◆ どのくらいの期間(回数)データを保持するか指定可能
 - ◆ 大容量データに対するスケーラビリティの高さ
 - ◆ 取り込みパフォーマンスの高さ
 - ◆ SASデータセットに変換し、蓄積するため、SASによる分析環境への移行が容易
 - ◆ 取り込んだデータをデータセットに変換するマクロプログラム(SASのデータステップ)を書き換えることにより、どのようなデータセットにするかを柔軟にカスタマイズ可能

15

WebHoundのレポーティング

- 静的レポート・・・定型的な集計レポートのテンプレートを約60種類用意
- 動的(MDDB)レポート・・・業務に合った切り口でレポーティング
- カスタマイズレポート・・・テンプレートをカスタマイズ可能

レポートのテンプレート例

- ◆ リクエストの多いページ
- ◆ 日/時間ごとのページリクエスト数
- ◆ エントリー/エグジットポイントTOP10
- ◆ 週ごとのセッション数
- ◆ セッション数と訪問者数の比較
- ◆ 404エラーコードのリクエスト名
- ◆ 参照サイトトップ25



16

WebHoundのレポートिंग 2

- MDDDBレポート・・・多次元データベース(バンドル)によるOLAP
縦軸横軸を変えたり、ドリルダウンしてより深く分析する

OLAPレポートの例

- リクエストの多いページ
↓
- ページごとのステータスコード
- エントリーページ
↓
- ページごとの参照ページ情報



17

WebHoundのレポートिंग 3

- カスタマイズレポート・・・マクロプログラムの修正、追加により、
業務に合わせたテンプレートを作成する

カスタマイズレポートの例

特定のページの参照ログを一定期間
集計し、毎日レポートする

特定のページからのクリックストリームを
ツリービュー(右図)で見る

検索キーワードの移り変わりを
ツリービューで見る

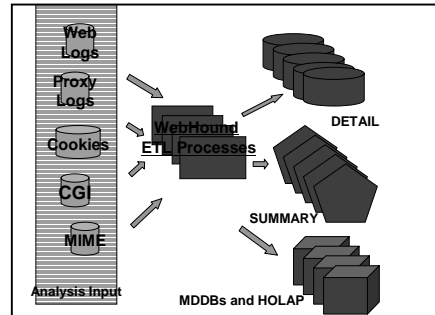


18

WebHoundによるETL Process

◆ ビジネスデータと連携

- Warehouse Administrator ソフトウェア (WA) を利用し、ビジネスデータとの連携
 - WebログデータをSASデータセットに変換し、WAで管理
 - MCIF、DWHに格納されているビジネスデータとの統合



一連のメタデータをWebHoundが生成し、WarehouseAdministratorが読み込み・管理

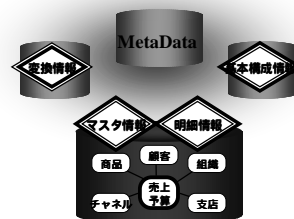
19

メタデータとは？

○メタデータとは？

- データに関わるデータ→
データの構成、内容、キー、索引等

メタデータは膨大なデータウェアハウスの中から、目的のデータの発見を容易にしてくれる



メタデータの役割

- データウェアハウス上にあるデータソースはどこから来たのか？
- 業務系システム→データウェアハウスに至るまでの変形は？
- データウェアハウスの内容の記述(権限、フィールド情報等)
- タイムスタンプやローディングスケジュールの為にバージョン付け

20

WarehouseAdministrator

メタデータ管理 & ETLツール



21

WarehouseAdministrator

メタデータ管理 & ETLツール

WebHoundによって書き出されたメタデータをWarehouseAdministratorに読み込む事で、SASデータセットに変換されたWebログの情報をすぐに見る事が可能となっている。

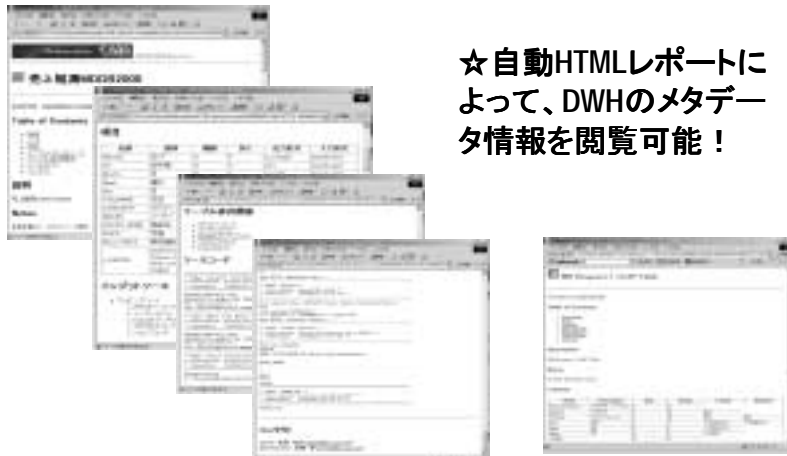
これにより、データの形式、保存場所、ETLのためのプロセス管理がメタデータにより行う事が可能となってくる。



22

WarehouseAdministrator

メタデータ管理 & ETL ツール

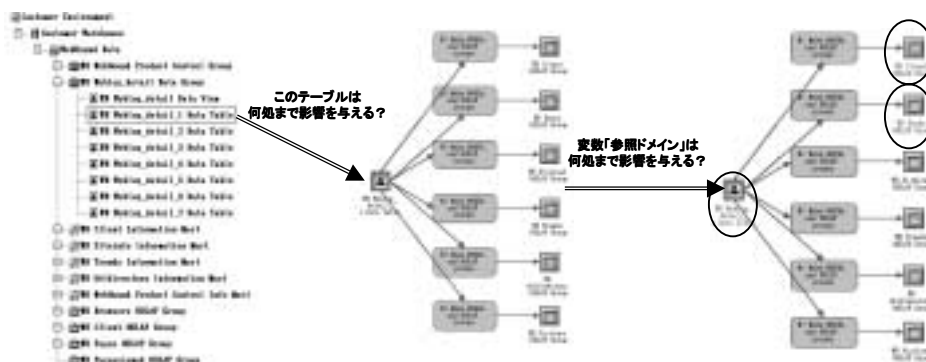


23

WarehouseAdministrator

メタデータ管理 & ETL ツール

多くのアドインツールが用意されており、「影響要因分析」などを行うImpactAnalysis等がある。



24

顧客データとの統合における WarehouseAdministratorの役割

- ◆ WarehouseAdministratorは情報系システムのためのDWH構築のために開発された。
- ◆ よって当然、Webログだけではなく、企業内に存在する様々なデータの統合が可能。
- ◆ WebHoundで紹介したような、Webログと個人を結び付けるロジックを用いる事により、本当の意味での「個」を認識したWebログ分析が可能となる。
- ◆ よりカスタマイズされ、よりニーズに合ったサービスを提供するための有効なインフラを構築する。

25

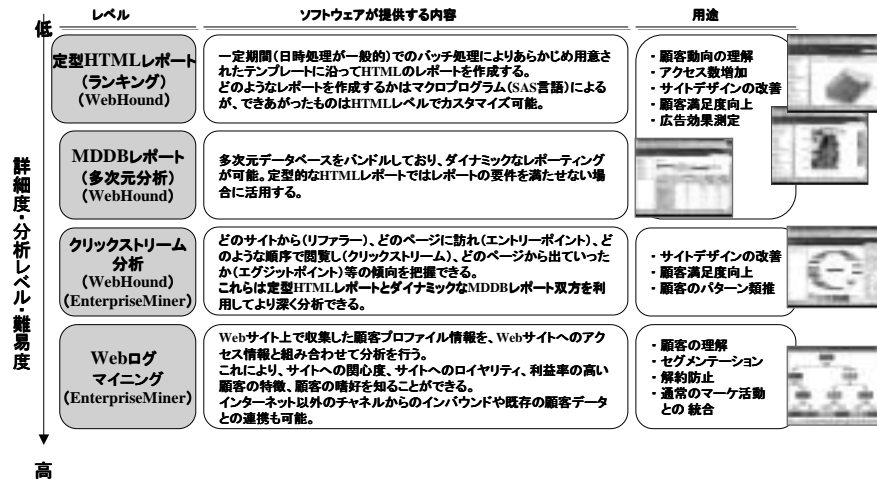
EnterpriseMinerによるデータ分析

- ◆ eビジネスを成功させるには、データ分析は不可欠なもの。
 - Webログデータとデータマイニングの技術を利用して、より深い分析を行っていく。
- ◆ Webログを使うことにより、「店の中のどのフロアーをどの順序で見て回ったか？」というような分析が可能となる。
- ◆ Webログ分析の成功のカギは、やはり「データ」の質。その中でも、「個人が特定できるかできないか？」によって、分析内容に大きな差が出てくる。

26

Webログ分析の分類と活用

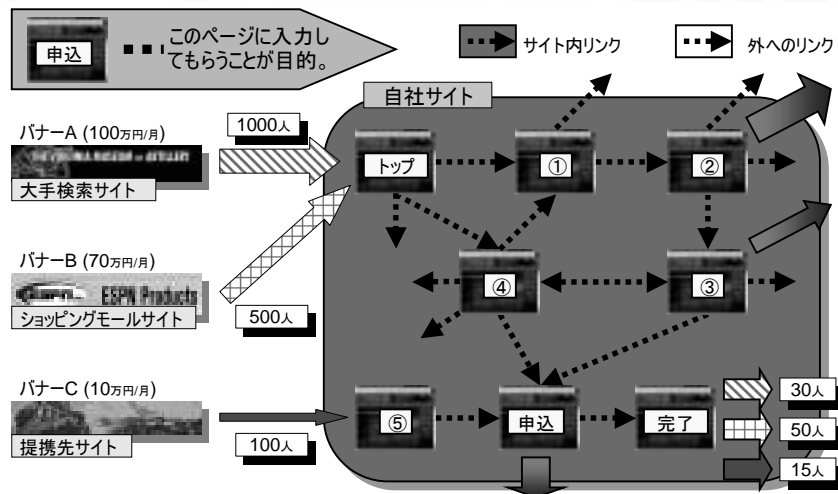
SAS Institute Japanが提供するソリューション



27

Webログ分析のトレンド

クリックストリーム分析とは？

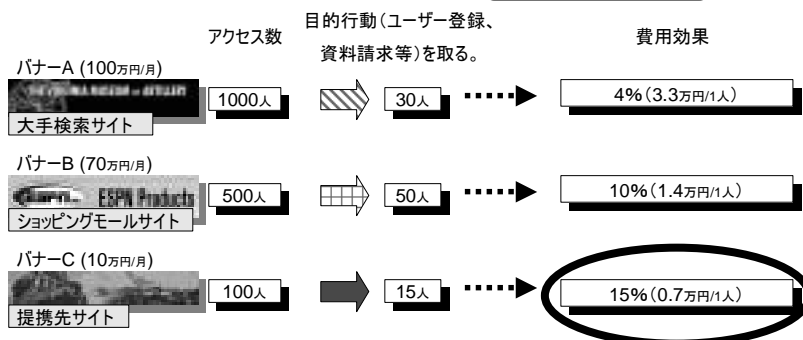


28

どのバナーが効果が高いのか？

◆ バナー広告別費用効果の算定

「申込」ページで入力し、「完了」ページに至った人のバナー別獲得数を見る。



29

アソシエーション・リンク分析による クリックストリーム分析

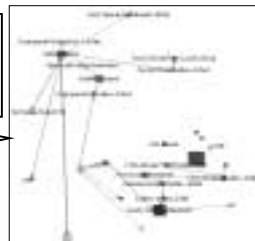
- EnterpriseMiner4.1では以下のノードが利用できる。
 - 「アソシエーション」ノード
 - 「リンク分析」ノード(評価版)
- 両方とも多くの履歴データの中から、発生するパターンやルールを見つけ出す手法になっている。
- 訪問時期やドメイン別、その他の条件の違いにより、クリックストリームは変化しているのか? 「リンク分析」ノードでは視覚的に発見する事ができる。
- ただの組み合わせだけではなく、Webログは時刻も入っているので、閲覧の順番などの分析することも可能となっている。



組織がacの
ユーザからの
クリックスト
リーム結果

組織がcoの
ユーザからの
クリックスト
リーム結果

訪問者のドメインがCOと
ACでパターンが違う。これ
はニーズの違いとも考えら
れる。



30

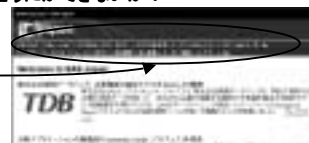
顧客の傾向をURL単位ではなく 大まかに傾向を掴めないか？

SASのURLは以下の通り。

- /offices/asiapacific/japan/news/press/199909/17.html
- /offices/asiapacific/japan/event/seminar/bsc.html
- /offices/asiapacific/japan/software/stratv.html

- 最後のHTMLレベルまでを考えて分析すると、非常に複雑になるので、どうにかできないか？

ファイル単位のレベルではなくて、ディレクトリのレベルで考えてみる。
上記のURLだと「NEWS」「EVENT」「SOFTWARE」と言った単位。



また、訪問者のセッションID (102.183.221.202.ts.2ijj.netMozilla4.0compatibleMSIE5.01WindowsNT5.0 2001) では複雑すぎるので、訪問者の組織で考えてみる。
例えば「ac」「co」「ne」「com」などでまとめてみる。

組織とディレクトリの関係を見てみる

URLのディレクトリレベル

	ディレクトリ	アクセス数	ディレクトリ	アクセス数	ディレクトリ	アクセス数	ディレクトリ	アクセス数	ディレクトリ	アクセス数	ディレクトリ	アクセス数
1	ac	112	co	282	ne	200	com	121	net	94	org	11
2	ac	1029	co	2612	ne	2076	com	2184	net	2786	org	198
3	ac	898	co	2652	ne	960	com	1836	net	300	org	8
4	ac	11	co	118	ne	47	com	27	net	52	org	5
5	ac	932	co	2829	ne	1182	com	4266	net	964	org	14
6	ac	170	co	988	ne	276	com	466	net	189	org	27
7	ac	276	co	1166	ne	360	com	1117	net	282	org	40
8	ac	920	co	2829	ne	1182	com	4266	net	964	org	14

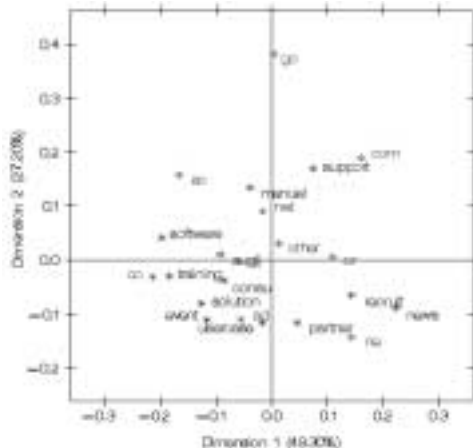
組織

アクセスの度数

- 上記のデータから「どのような組織に属している人は、どのようなものに興味があるのか？」といった傾向などを導き出す事は出来ないだろうか？
- 「組織単位」を企業単位にしたり顧客単位に発展させられるのではないか？

組織とディレクトリなど訪問者の関連性

- 対応分析と、その結果のグラフ化を行ってみる。



グループ別に変数の度数を集計



対応分析プログラム

```
proc corresp data = Dataset outc = out1 cellchi2;
  var event -- recruit;
  id group;
run;
%plotit(data=out1,datatype=corresp,vref=0,href=0,
        color=black plotvars=Dim1 Dim2);
```

このような軸の解釈はできないであろうか？

第1成分(Dimension 1)

正：会社動向 負：ソフトウェア(スキルアップ)

第2成分(Dimension 2)

正：ソフトウェア(サポート) 負：会社情報

33

組織とディレクトリなど訪問者の関連性

```
proc corresp data = Dataset outc = out1 cellchi2;
```

前のページにあるプログラムの、「proc corresp」ステートメントの「cellchi2」オプションを利用することにより、項目間の関連の強さを数値的に見ることができる。

	event	consult	support	software	training
ac	28.77	3.81	1.58	196.00	22.11
ad	3.00	0.81	28.27	6.37	3.81
ca	85.11	2.22	77.44	96.54	187.30
com	32.09	9.22	388.00	37.88	28.30
an	6.74	0.37	15.25	2.08	3.79
ae	0.05	0.06	183.41	76.64	47.76
af	7.33	0.06	31.23	3.32	0.33
ag	8.00	2.22	2.64	0.34	17.13
other	0.21	0.74	26.68	8.81	3.17

SASのアウトプットからは

「ac」→「software」

「co」→「training」

「com」→「support」

といった関連の強さなどが考えられる。

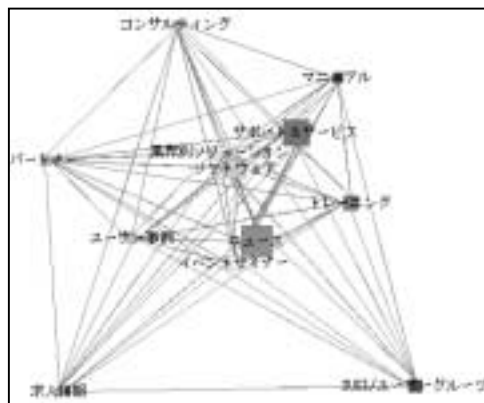
34

セッションIDとディレクトリでの アソシエーションとリンク分析

アソシエーション分析

[illegible]

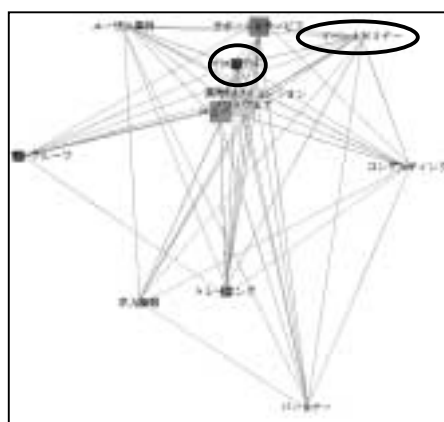
リンク分析



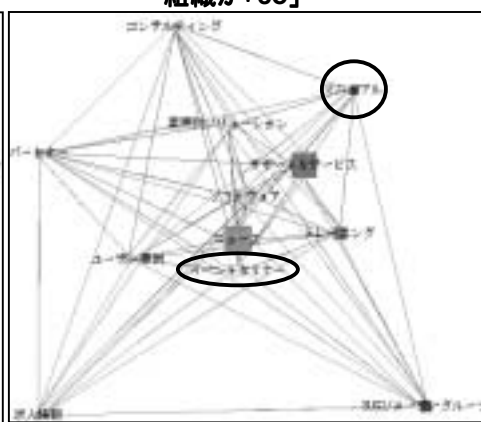
35

組織別に違いはあるか？

組織が「ac」



組織が「co」



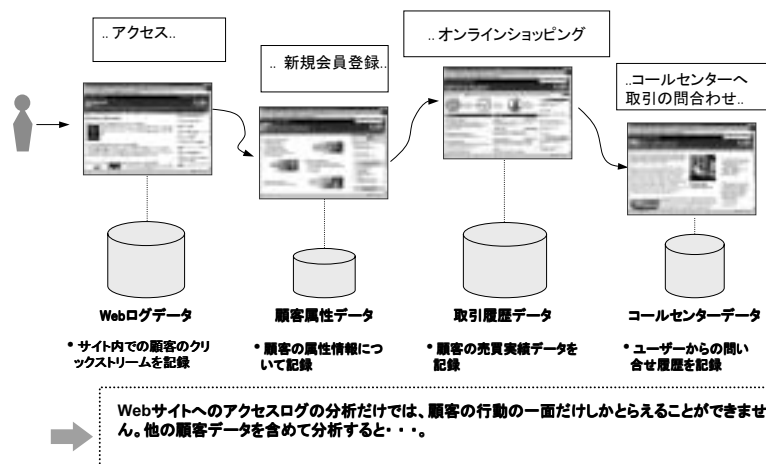
36

より高度な分析に向けて

37

Webログと顧客データによるより高度な顧客分析の実現

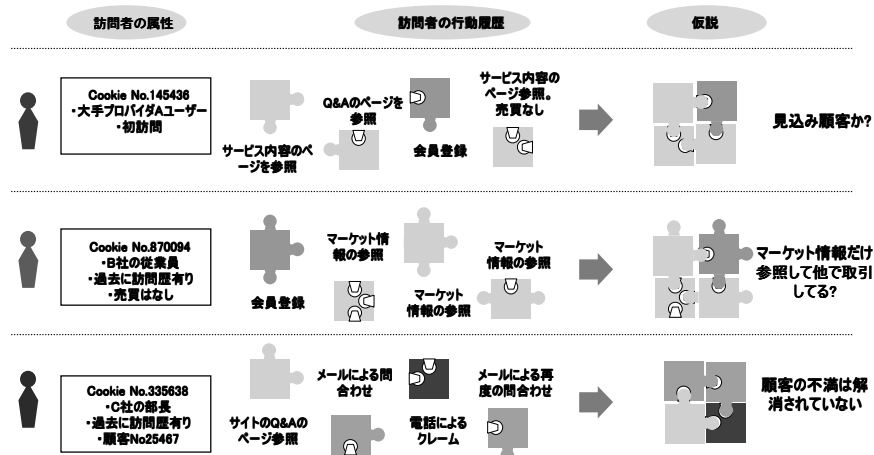
顧客の本当の姿を知るためには、Webサイトへのアクセスログ以外のデータソースを分析の対象データとすることが重要となります。



38

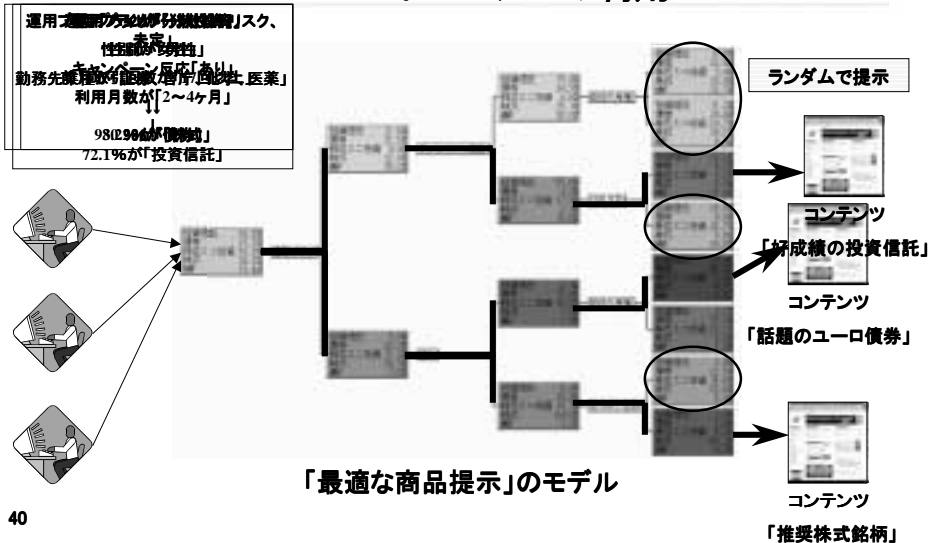
Webログと顧客データによるより高度な顧客分析の実現

Webサイト訪問者が残した様々な足跡を分析することによって、顧客像についての仮説を立案することが可能となり、One to Oneのサービス、対応を行うことが可能になります。



39

ツリー分析と レコメンドエンジンの利用



40



The Power to Know™