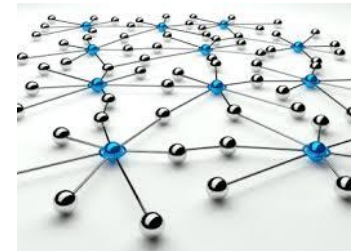


Webマイニングの手法

- Webコンテンツマイニング
 - Webの内容（contents）を分析
 - 主にテキストマイニングといった手法
- Web構造マイニング
 - Webサイトのリンク関係、生成・消滅、分裂・統合、といった構造を分析
 - 主にグラフマイニング（ネットワーク分析）
- Webログマイニング
 - Web上の行動ログ（アクセスログ、検索履歴）の統計情報にもとづいて、コミュニティや個人の思考を分析
 - ログ解析（アクセス解析）、統計分析、協調フィルタリング

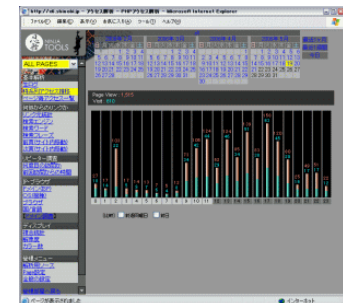
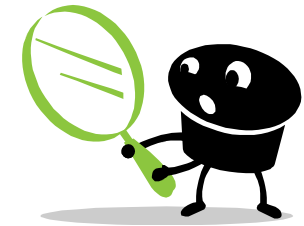


Webログマイニングの活用

- アクセス解析

- Webのアクセスログから有用な知識を発見する
マイニング技術

- 顧客の特性、人気のあるコンテンツ、インターネット広告の投資効果データ等

[illegible]

アクセス解析結果の応用



- 閲覧ページ、滞在時間 → 人気コンテンツ
- 時系列アクセス推移・ドメイン → ユーザ層
- 検索ワード・エンジン → 何処からのリンクか
- ...



- Webサイトの改善分析
- インターネット広告の投資最適化
- 利用者の嗜好分析
- 優良顧客分析
- ユーザのセグメンテーション

Webサイトの改善分析

- 顧客ニーズや売上の変化への対応
 - Webサイトの改善の必要性
- 閲覧状況をアクセスログ情報の利用
 - Yahoo!JAPAN、Amazon、楽天
 - 閲覧者の閲覧状況をアクセスログ情報から取得して、Webサイトの改善につとめている



(例) 検索エンジン・ワード

《 検索ワード Ranking》 Analyzable : 48,874 》				
順位	検索エンジン	検索ワード	シェア	検索した訪問者数
1	[Google]	javascript	9%	4,410
2	[Google]	php	5.5%	2,720
3	[Google]	PHP	3%	1,493
4	[Google]	JavaScript	2.6%	1,316
5	[Yahoo!JAPAN]	javascript	1.6%	822
6	[Yahoo!JAPAN]	php	1.3%	670
7	[Google]	配列	0.9%	477
8	[Google]	select	0.9%	443
9	[Yahoo!JAPAN]	PHP	0.8%	403
10	[Google]	Javascript	0.7%	374
... 略 ...				

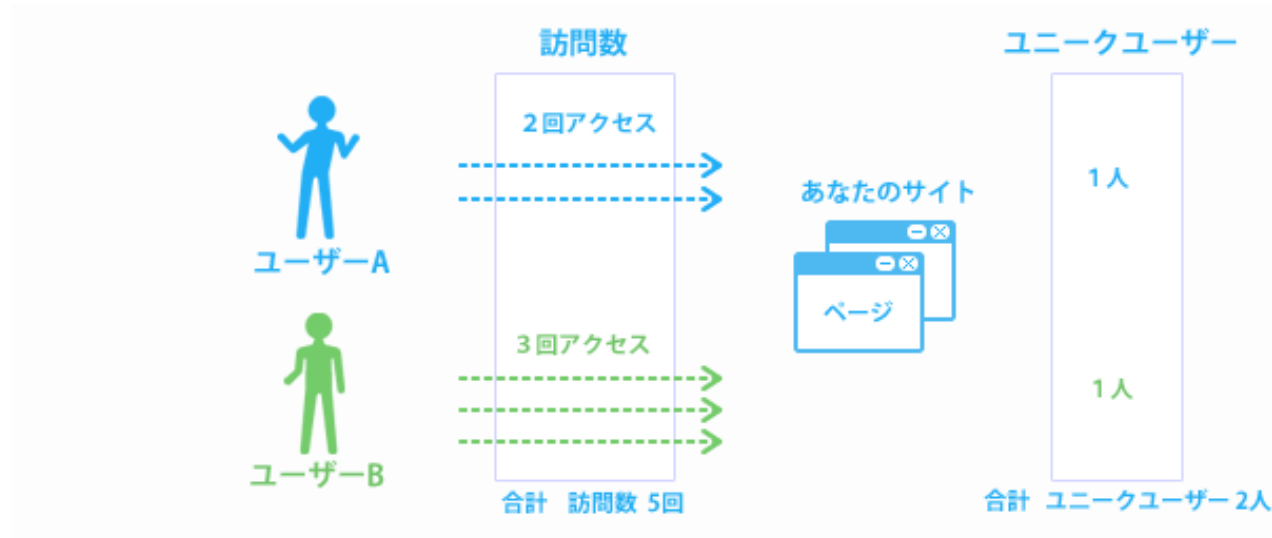
アクセスログ解析の指標 (1)

- 代表的な指標

- ページビュー (PV : Page View)
- ユニークユーザー数 (UU : Unique User)
- セッション数
- 滞在時間
- クリック率

アクセスログ解析の指標 (2)

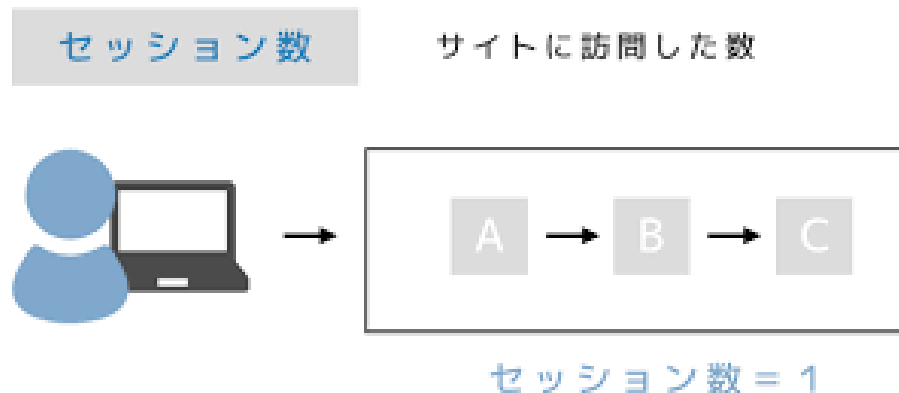
- ページビュー (PV)
 - Webブラウザからの1回のページリクエスト
- ユニークユーザー数 (UU)
 - 一定期間内にそのWebサイトにアクセスした訪問者数の単位
 - 最初のアクセスだけをカウントし、それ以降をカウントしない
 - Webブラウザに保存されるcookieを利用する



アクセスログ解析の指標 (3)

• セッション数

- 個人がそのWebサイトを1ページ以上アクセスした行動単位の総数
 - Webサイトを訪問したユーザにより開始された個々のセッションの数
 - セッションはある一定以上経過したらそれ以降は別の新しいセッションとみなされる
 - ある一定時間内に戻ってきた場合には同一セッションとなる



アクセスログ解析の指標 4

• 滞在時間

- 個人の1回の訪問（セッション）の間の経過時間
 - 1ページごとの訪問時間は閲覧時間と呼び異なる指標
 - 滞在時間が長いからといって重要であるということはいえない（別の作業を行っている可能性があるから）



• クリック率

- リンクや広告がクリックされる割合を示す指標
 - $(\text{リンクや広告がクリックされた総数}) / (\text{リンクや広告を含むページの総表示回数})$



同一性判定

- ユニークユーザの判定は難しい
 - 現実的には、IPアドレスと「ユーザエージェント」名の組合せが用いられる
- ユーザエージェント (User-Agent)
 - ユーザエージェント(UA): ログファイルに記述されているクライアントのブラウザやOSの情報



(例) ユーザエージェント



- Internet Explorer 11
 - Mozilla/5.0 (Windows NT 6.3; WOW64; Trident/7.0; Touch; rv:11.0) like Gecko
 - ...
- Safari
 - Mozilla/5.0 (Macintosh; Intel Mac OS X 10_8) AppleWebKit/536.25 (KHTML, like Gecko) Version/6.0 Safari/536.25
 - ...
- iPhone
 - Mozilla/5.0 (iPhone; U; CPU like Mac OS X; en) AppleWebKit/420+ (KHTML, like Gecko) Version/3.0 Mobile/1C28 Safari/419.3
 - Mozilla/5.0 (iPhone; CPU iPhone OS 6_0 like Mac OS X) AppleWebKit/536.26 (KHTML, like Gecko) Version/6.0 Mobile/10A403 Safari/8536.25
 - ...
- Android
 - Mozilla/5.0 (Linux; Android 4.1.1; Nexus 7 Build/JRO03S) AppleWebKit/535.19 (KHTML, like Gecko) Chrome/18.0.1025.166 Safari/535.19
 - ...
- Google Chrome
 - Mozilla/5.0 (X11; Linux x86_64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/28.0.1500.52 Safari/537.36
 - ...

アクセス解析方法

アクセス解析方法下記の3種類がある。

1. サーバログ解析型アクセス解析

- Webサーバのログを解析し、ユーザのアクセス状況を解析する。

2. パケットキャプチャ型アクセス解析

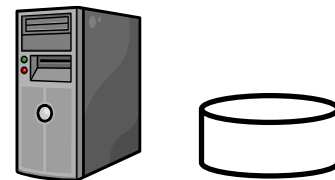
- Webサーバのアクセスログではなく、スイッチング・ハブを流れるパケットの情報を利用して解析を行う。

3. タグ挿入型アクセス解析

- Webページにタグを挿入し、ユーザのアクセス状況を解析する。

1. サーバログ解析型アクセス解析 (1)

- Webサーバのログを解析し、ユーザのアクセス状況を解析する。



127.0.0.1 - frank [10/Oct/2000:13:55:36 -0700] "GET /apache_pb.gif HTTP/1.0" 200 2326

アクセス
(サーバ
にリクエ
スト) し
たクライ
アントの
IPアドレ
ス

HTTP認証
によって
アクセス
した人の
ユーザID

アクセス
した
(サーバ
がリクエ
ストを受
け取っ
た) 時刻

クライア
ントが
使ったメ
ソッド

クライア
ントがア
クセスし
たリソー
ス (パス/
クエリ文
字列)

クライア
ントがア
クセスで
使ったプ
ロトコル

サーバ
がクライ
アントに返
すステ
ータス
コード
(200
は成
功)

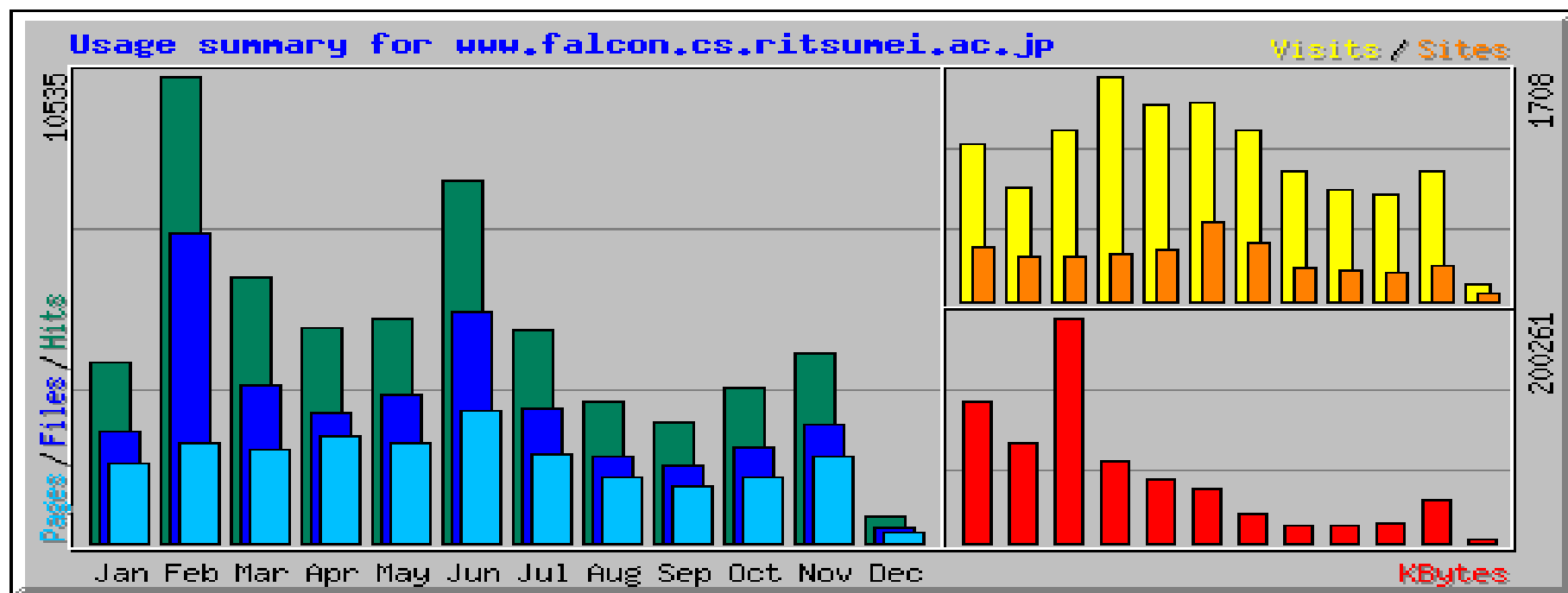
サーバがク
ライアント
に送信した
サイズ (応
答ヘッダを
除く)

アクセス解析例

ページ	訪問者数		滞在時間		平均	
	今週	累計	今週	累計	今週	累計
合計	1141	39829	5009	341833	4.39	8.58
トップページ	197	7494	2901	112104	14.73	14.96
物件概要	32	1162	95	23693	2.97	20.39
会員登録フォーム	0	59	0	0	0	0
ロケーション（簡易）	41	2474	30	11803	0.73	4.77
アクセス（簡易）	50	2680	0	11595	0	4.33
計画概要（簡易）	37	2134	0	9134	0	4.28
プラン（簡易）	57	2307	0	11678	0	5.06
外観（簡易）	33	1751	0	3473	0	1.98
構造・工法（簡易）	19	1147	0	1783	0	1.55
モデルルーム（簡易）	45	1990	339	7382	7.53	3.71
設備・仕様（簡易）	28	1377	0	3387	0	2.46
資料請求フォーム	0	17	0	0	0	0
トップページ（詳細）	1	157	0	2684	0	17.1
ロケーション（詳細）	0	100	0	659	0	6.59
東難波の歴史	0	38	0	691	0	18.18
良好な住環境	0	47	0	599	0	12.74
現地周辺地図	0	56	0	0	0	0
アクセス（詳細）	0	105	0	8770	0	83.52
各主要駅のアクセス	0	36	0	31	0	0.86
新線開業アクセス	0	35	0	183	0	5.23

(例) Webalizerによるアクセスログ解析

Webalizerによるアクセスログ解析例



2. パケットキャプチャ型アクセス解析

- パケットキャプチャ型アクセス解析
 - スイッチング・ハブを流れるパケットの情報を利用して行う解析
 - パケット情報(IPV4)
 - 識別子・フラグ・断片位置
 - 送信元アドレス
 - 宛先アドレス
 - プロトコルID
 - RTMetricsやVissionalistのようなツールが使用される
 - リアルタイム解析が可能



3. タグ挿入型アクセス解析



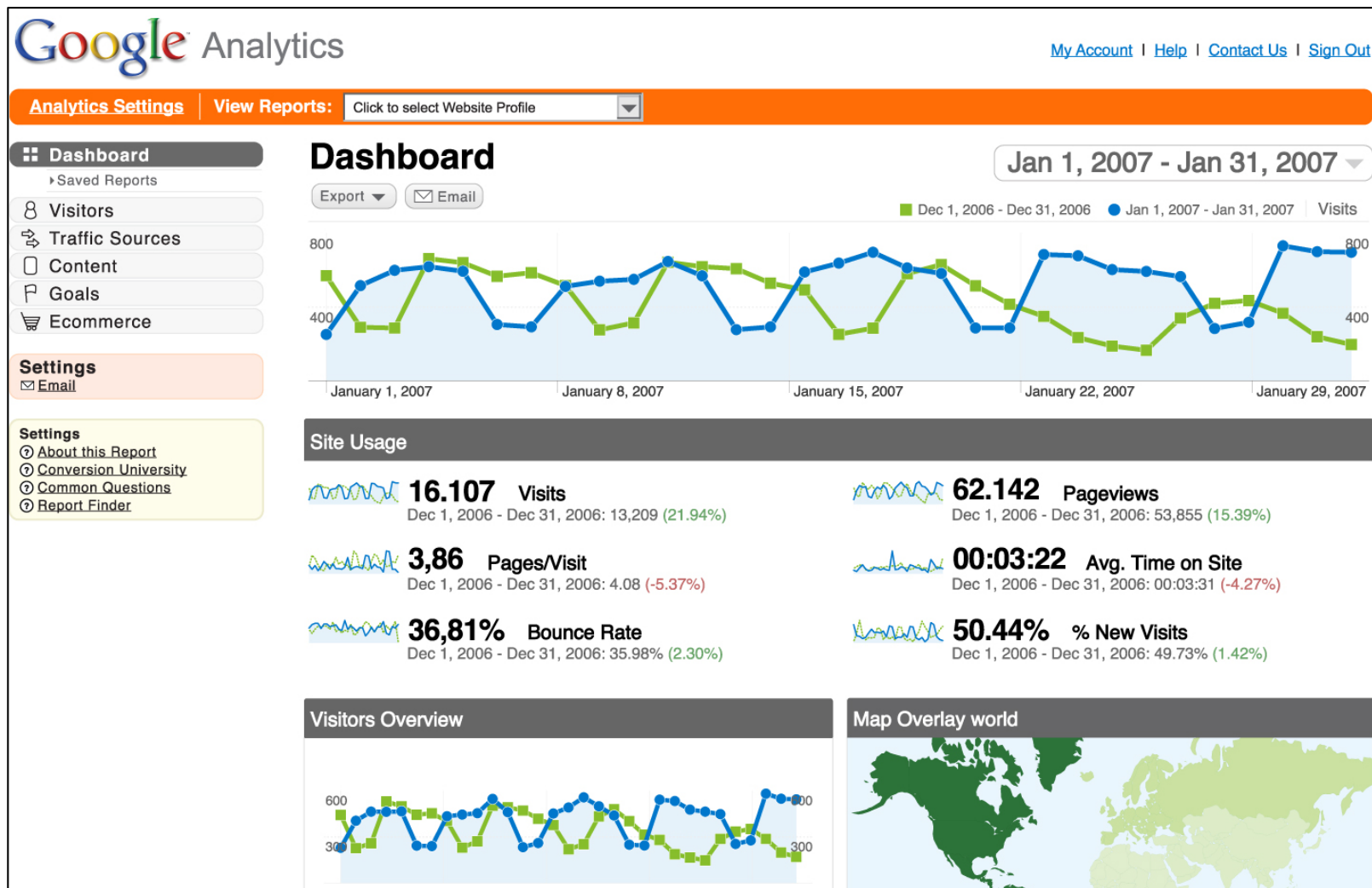
- タグ挿入型アクセス解析
 - 直接、Webページのhtmlにタグを挿入
 - ユーザがこのhtmlソースにアクセスすると、このjavascriptが実行され、その内容が解析用サーバに送られる
- (例)
 - HTMLソース内に下記のようなJavascriptを埋め込んでおく

```
<script src=http://www.....js"
type="text/javascript"></script>
```

```
<script type="text/javascript"> _ua="UA
xxxxxx"; UATracker(); </script>
```

- ユーザがこのhtmlソースにアクセスすると、このjavascriptが実行され、その内容が解析用サーバに送られる

(例) Google Analytics



アクセス解析の難しさ



- クローラーによる影響
 - サーバログ解析型やパケットキャプチャ型では影響大
- ブラウザの戻りボタン
 - サーバログ解析型やパケットキャプチャ型では取得できない
- pdf等の非htmlファイルへのアクセス
 - タグ導入型では取得できない
- その他
 - タグ導入型では埋めこみミスによる取得できない可能性が発生する

インターネット広告

アクセス解析結果の応用



- 閲覧ページ、滞在時間 → 人気コンテンツ
- 時系列アクセス推移・ドメイン → ユーザ層
- 検索ワード・エンジン → 何処からのリンクか
- ...



- Webサイトの改善分析
- インターネット広告の投資最適化
- 利用者の嗜好分析
- 優良顧客分析
- ユーザのセグメンテーション

インターネット広告の投資最適化

- インターネット広告は投資の費用対効果 (ROI: Return on investment) が優れている
 - TVのCM、新聞広告、雑誌広告、ラジオ広告等と比べて
- 投資の最適化
 - コンバージョン（商品の購入、資料請求、契約）が多いキーワード、リンクを選別



インターネット広告

- インターネットのWebサイトや電子メール等を用いて、企業が製品やサービスの宣伝活動を行うこと
- Webページ向け、電子メール向け、検索エンジン向け、携帯電話向けなど、様々な形態の広告が使用されている
- 新しい形態の広告が盛んに提案されて実施されている
- 広告料金形態も多様
 - 露出回数見合い
 - クリック回数見合い
 - 成果数見合い

拡大するインターネット広告市場

日本の広告市場 6兆1,522億円

動画広告活用
の本格化

スマートフォン市場
の成長

インターネット広告市場
1兆519億円
前年比+112.1%

運用型広告の拡大

新しいマーケティング
手法の確立

出典：電通「日本の広告費」

中国のWeb広告市場

デジタル広告比率は50%以上

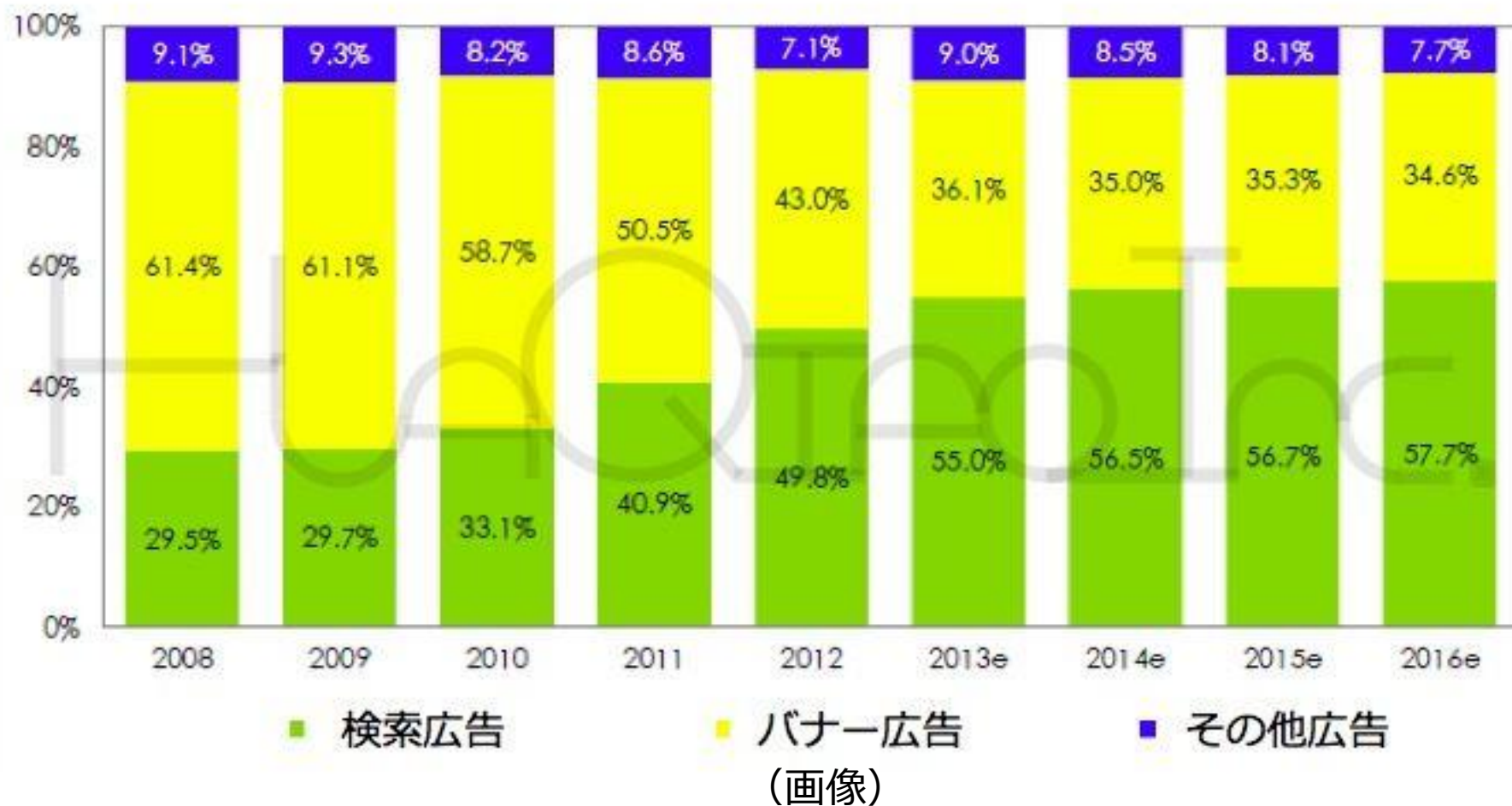
Total Media Ad Spending in China, by Media, 2014-2020

billions

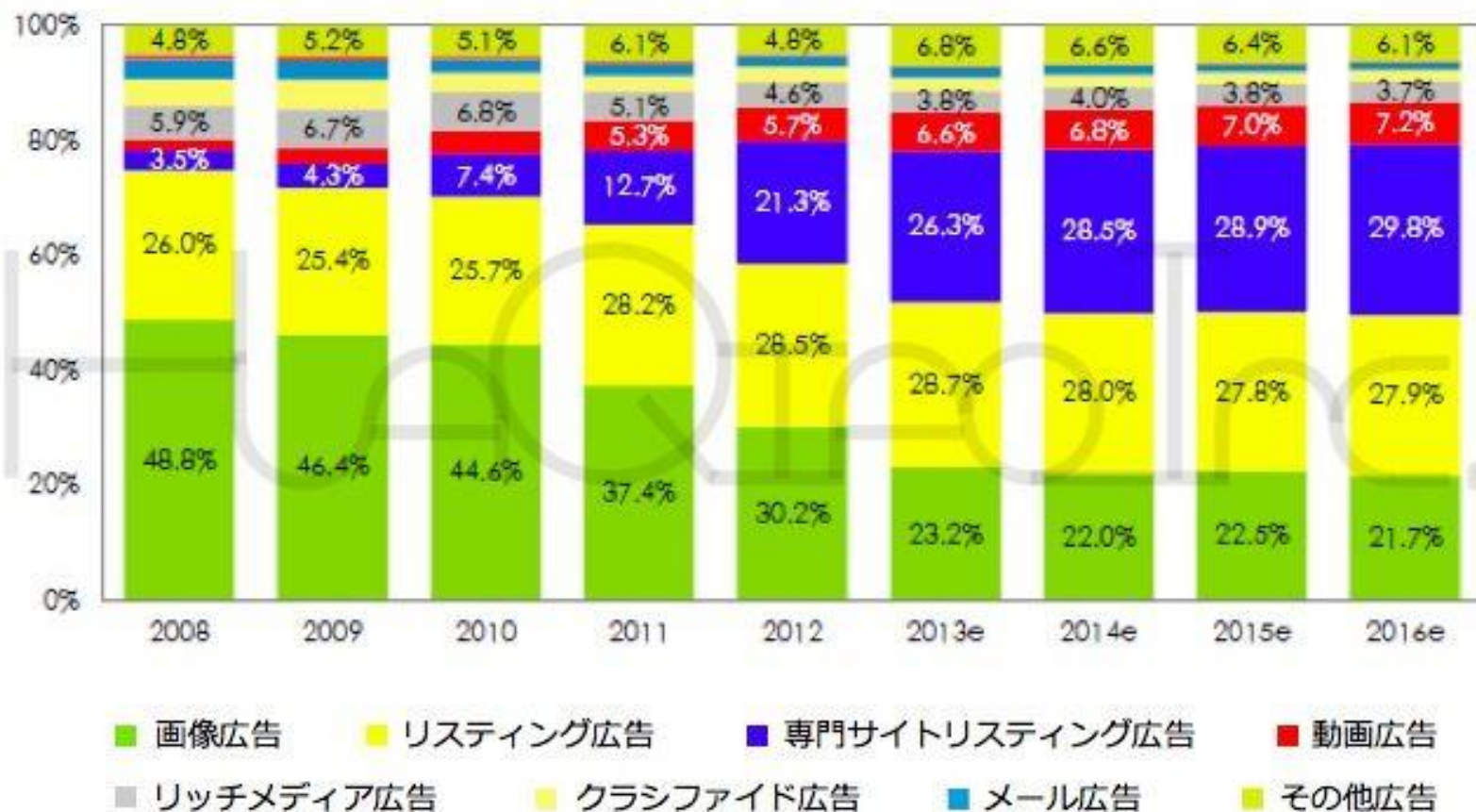
	2014	2015	2016	2017	2018	2019	2020
Digital	\$23.55	\$31.09	\$40.42	\$50.52	\$62.14	\$73.32	\$83.59
—Mobile	\$8.21	\$17.07	\$27.31	\$37.63	\$49.18	\$60.25	\$70.43
TV*	\$20.01	\$19.61	\$18.92	\$18.64	\$18.73	\$18.83	\$18.92
Out-of-home	\$7.60	\$8.36	\$9.02	\$9.48	\$9.86	\$10.15	\$10.35
Print	\$6.89	\$5.96	\$5.50	\$5.21	\$5.01	\$4.92	\$4.96
—Newspapers*	\$5.97	\$5.14	\$4.72	\$4.46	\$4.27	\$4.18	\$4.22
—Magazines*	\$0.92	\$0.82	\$0.78	\$0.76	\$0.74	\$0.73	\$0.73
Radio*	\$3.99	\$4.31	\$4.49	\$4.62	\$4.71	\$4.76	\$4.81
Total	\$62.04	\$69.33	\$78.35	\$88.47	\$100.45	\$111.98	\$122.63

Note: converted at the exchange rate of US\$1=RMB6.23; excludes Hong Kong; numbers may not add up to total due to rounding; *excludes digital
Source: eMarketer, March 2016

2008～2016年 中国ネット広告市場シェア【広告形式別】

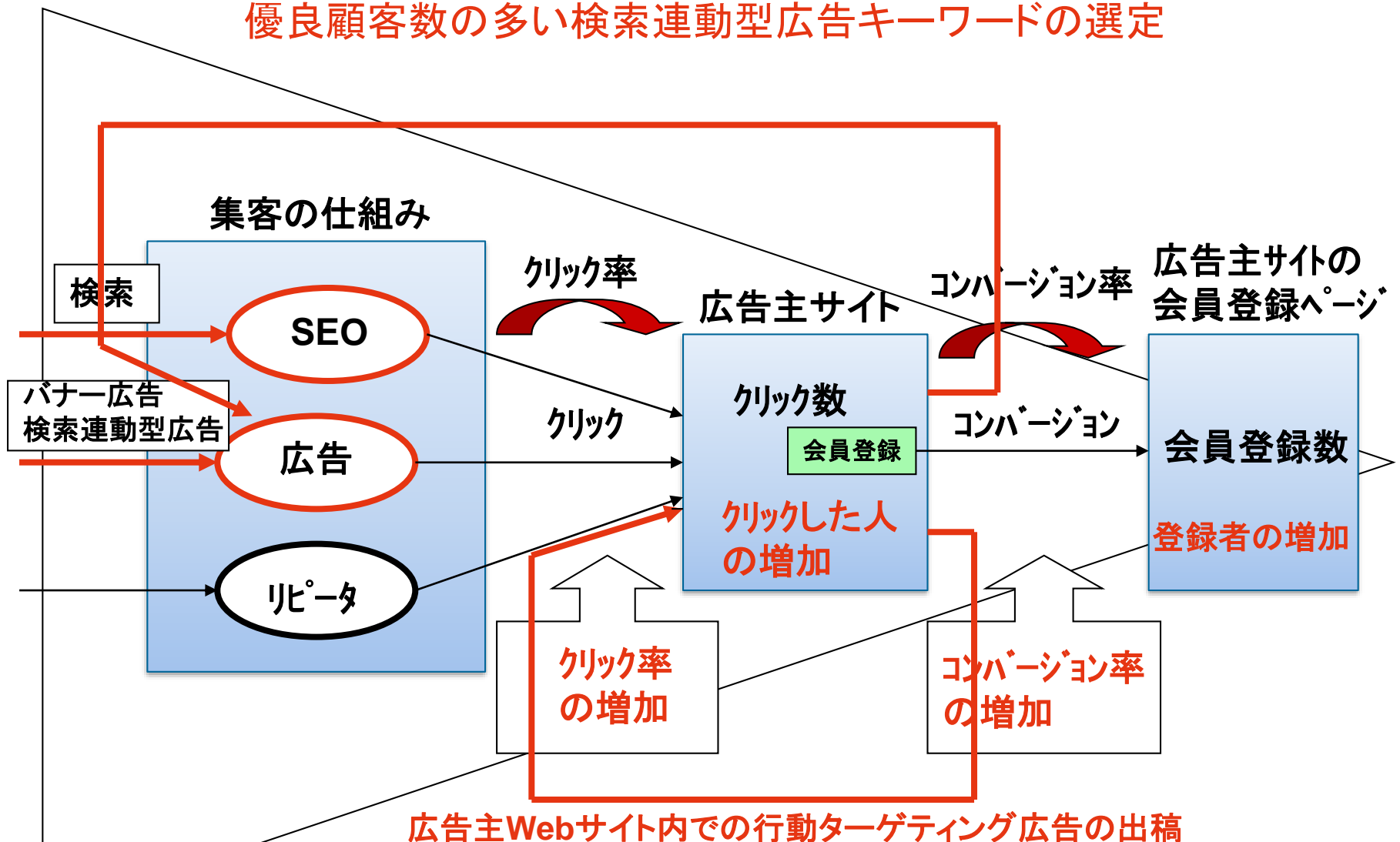


2008～2016年 中国ネット広告市場シェア【クリエイティブ形式別】



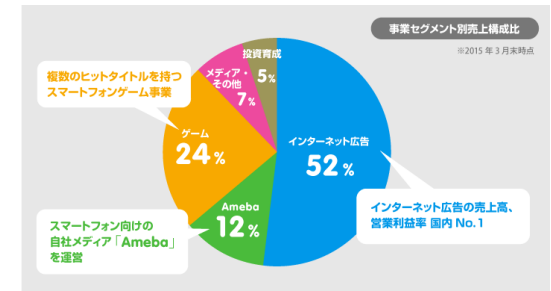
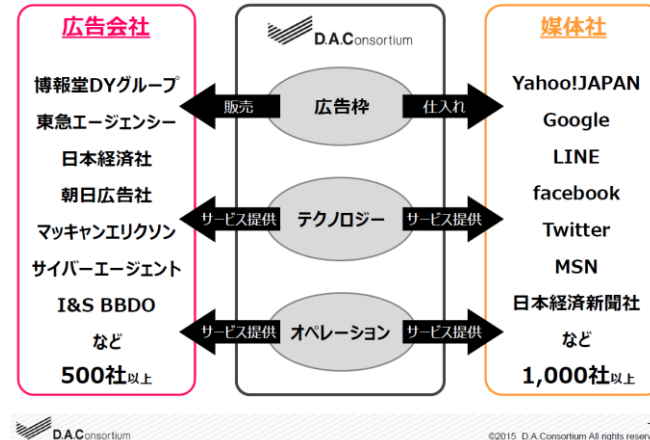
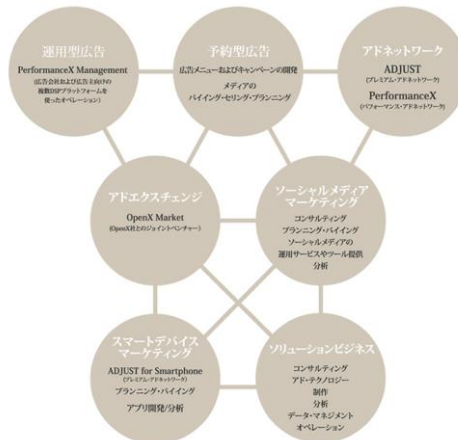
検索広告

優良顧客数の多いバナー広告ポータルサイトの選定
優良顧客数の多い検索連動型広告キーワードの選定



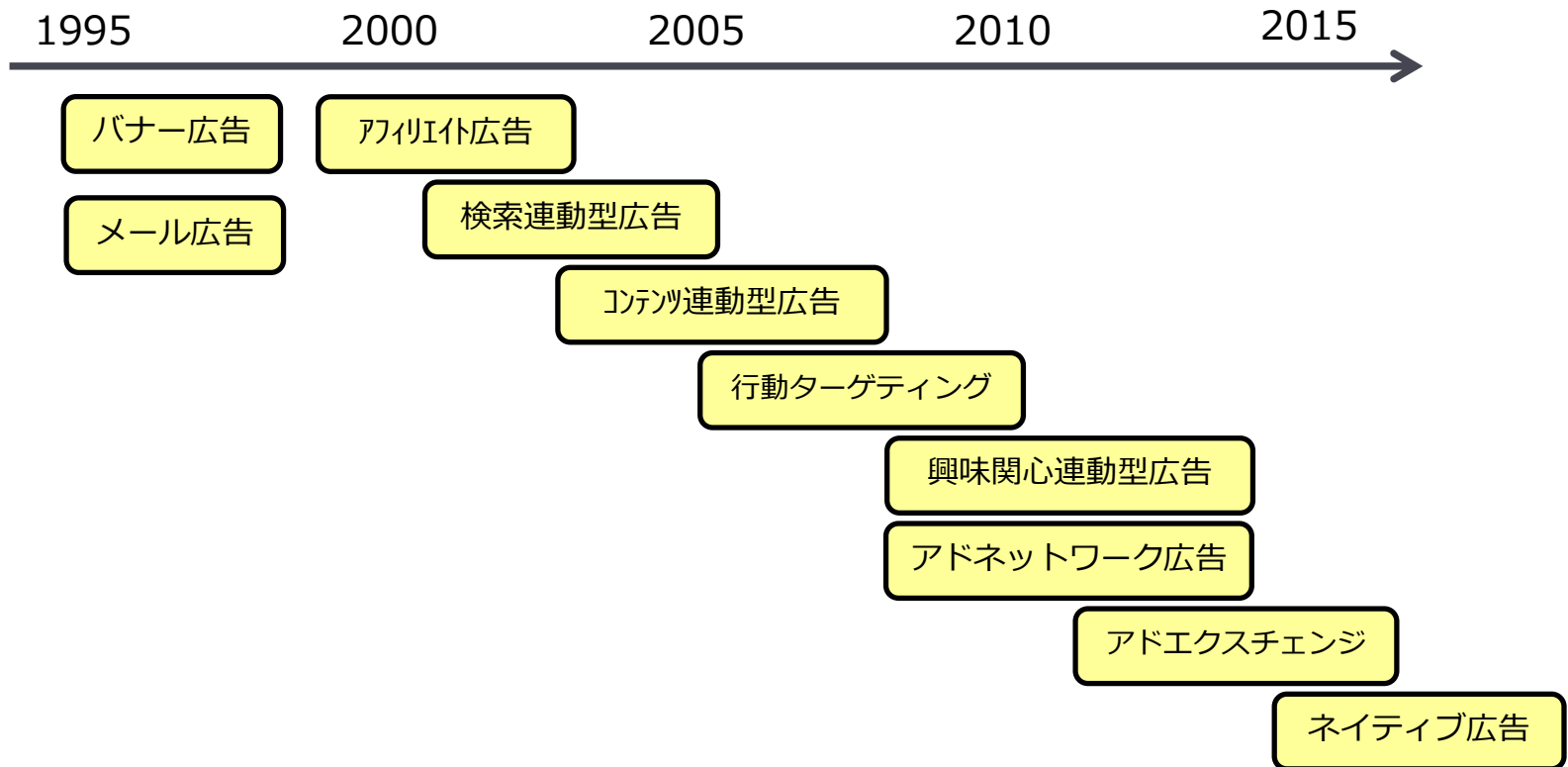
インターネット広告

- インターネット広告は広告配信会社によって実施されている
 - サイバーコミュニケーション（電通系）
 - デジタルアドバタイジングコンソーシアム（博報堂系）
 - サイバーエージェント（独立系）など



サイバーコミュニケーション デジタルアドバタイジングコンソーシアム サイバーエージェント

インターネット広告の歴史



インターネット広告 歴史(1)

1995 2000 2005 2010 2015 →

バナー広告

アフィリエイト広告

メール広告

検索連動型広告

バナー広告:

Webページ内に横長の長方形の枠（バナー）を設け、その枠内に広告を表示する広告手法

- Webページの上部あるいは下部に表示されることが多い
- 広告主のサイトにリンクが張られており、閲覧者が枠内をクリックすると広告主サイトが閲覧できる



60×468ピ
クセル



90×728ピ
クセル

インターネット広告 歴史(2)



バナナ広告

アフィリエイト広告

メール広告

検索連動型広告

メール広告:

メールに掲載される広告

- 数行程度のテキストメールやメールマガのヘッダやフッターに挿入された広告文がある
- テキストメール広告だけでなく画像等が利用できるHTML形式のHTMLメール広告がある。

サーバがうるさくて仕事にならない

サーバと同じ場所で仕事って…ウルサイ&空調費をつけなくちゃ大変！
 静音サーバラックなら、エアコンの音よりも静かになって、作業効率が格段にアップします！ 詳しくは⇒ <http://www.key-rp.com/dbseries/>

■グループ内メール共有システム GroMa

メールもやる気も ひとつにします

- ・1つのアドレスを複数ユーザで共有
- ・受信箱と送信箱の一体化で漏れなし
- ・グループ内のメール状況が一目瞭然

■他社を出し抜く処理能力アップならお任せ！>>> <http://www.key-p.com>

<https://www.webfile.jp/ad.html>

インターネット広告 歴史(3)

1995 2000 2005 2010 2015 →

バナー広告

アフィリエイト広告

メール広告

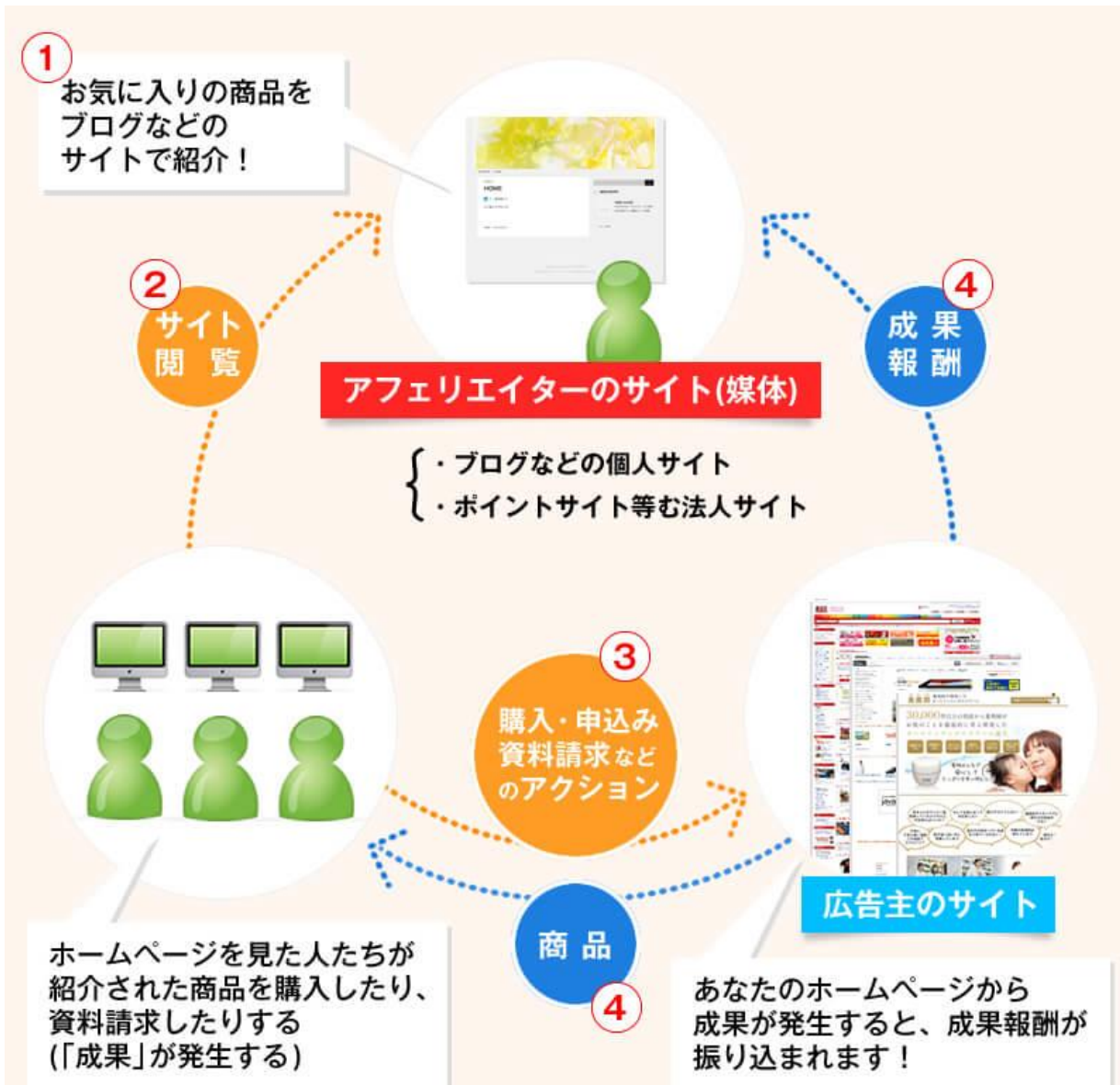
検索連動型広告

アフィリエイト(Affiliate)広告:
成果報酬型インターネット広告

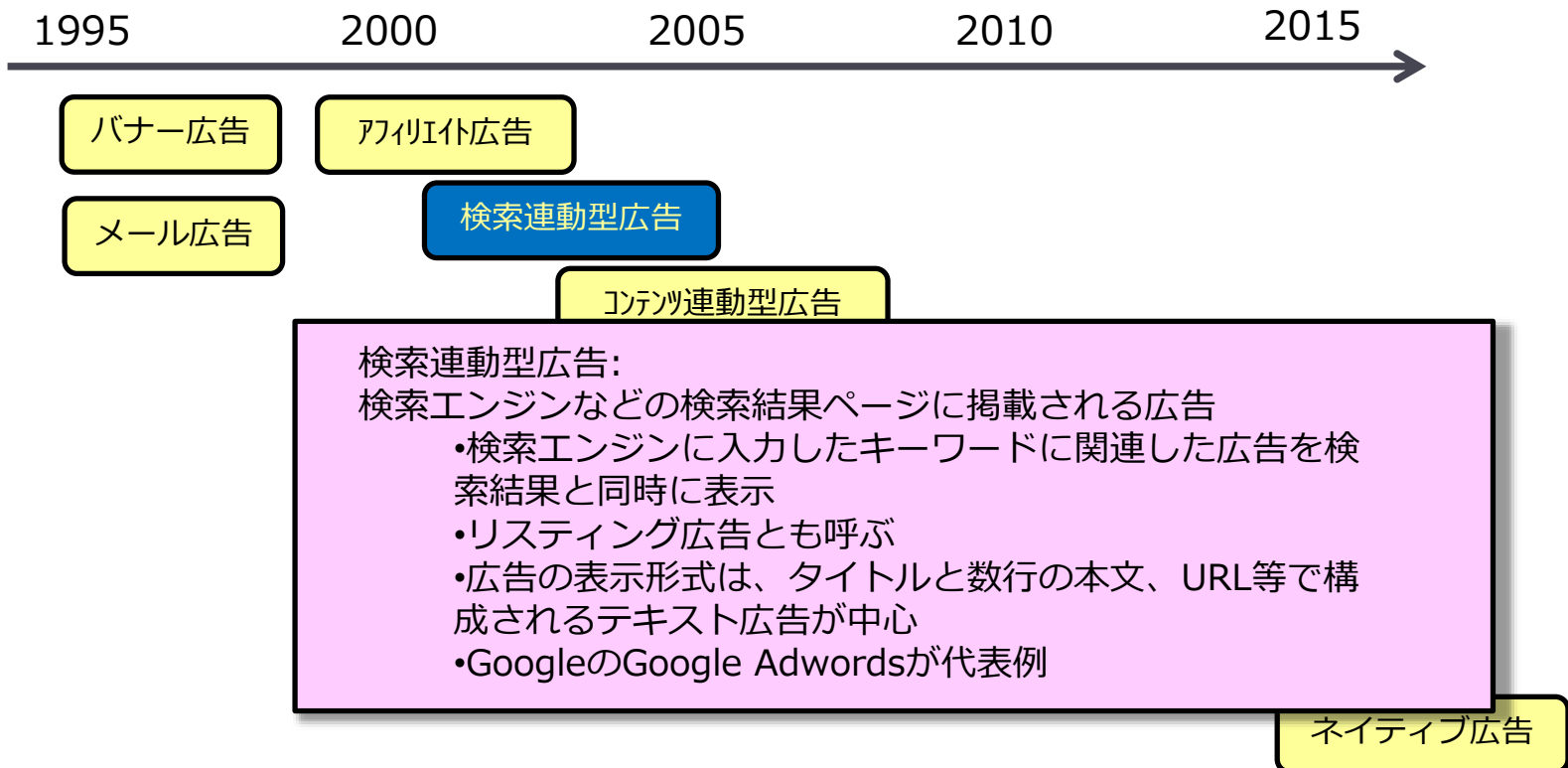
- あるWebサイトに表示された広告を閲覧者がクリックしたときや、クリックの後、その閲覧者が商品やサービスの購入をしたときに、そのWebサイト提供者に成功報酬を与える広告形態
- クリックしたときの成功報酬をクリック報酬、商品購入したときの成功報酬をアクション報酬と呼ぶ
- 1回につき一定の金額を支払う定額型報酬と、購入した金額に比例した金額を支払う定率型報酬に分けられる
- なお、英語では、Affiliate Marketingと呼ぶことが多い

告

アフィリエイト広告 (Affiliate Advertising)



インターネット広告 歴史(4)



検索連動型広告(Listing Advertising)

The screenshot shows a Google search results page for the query "ホームページ 集客" (Homepage Lead Generation). The search bar at the top shows the query and a search button. Below the search bar, the text "検索連動型広告" (Listing Advertising) is overlaid in large red characters. A red line points from this text to a red box on the right side of the page, which contains a list of search results. Another red line points from the same text to a red box on the left side of the page, which also contains a list of search results. The search results are listed in a standard Google format, with the title, URL, and a brief description. The results are as follows:

他のキーワード: 集客方法 ネットショップ 集客 ホームページ アクセスアップ
ホームページ 集客 無料

売上げ獲得型ホームページ制作 - hp-syuukyaku.info
www.hp-syuukyaku.info/page/ ▼
SEO対策付きホームページで集客 初期費用0円で集客する

集客に強いホームページ制作会社 - webkikaku.co.jp
www.webkikaku.co.jp/ ▼ 03-5456-0610
ライブルを出し損きたい社長さん大歓迎! フランク上のご提案で集客をサポート
ランディングページ制作・女性をターゲットにしたい・集客がしたい
〒東京都渋谷区 富士ビル40 7F 桜丘町15-14

ホームページ 集客200社一括比較 - homepage-seisaku.jp
www.homepage-seisaku.jp/ ▼
【無料&簡単に】ホームページ 集客を一括比較 おかげでご紹介数10万件突破!
対応満足度95%以上の紹介・ご相談無料・全国各地の企業多数ご紹介

Googleでホームページ作成
www.google.com/Apps/Sites ▼
文書を作成する感覚でサイトを作る。
30日間無料試用。

ネット集客どうすれば良いの?
www.quartet-communications.com/ ▼
検索広告で効率集客。まずは2015年版
最新実績資料50社分を無料ダウンロード

集客できるホームページ
www.ogunobu.com/ ▼
リスティングでお金を使っているのに
反応が少ないと困っていませんか?

Web戦略コンサルティング
www.web-consultants.jp/ ▼
経験・知識が力。(Web)担当専任必置!

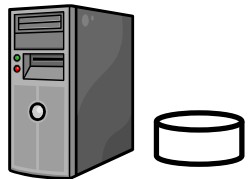
検索連動型広告の仕組み

ステップ1：人気の高いキーワードの選定

【1】人気の高いキーワードの選定

1ヶ月前の検索アクセスキーワードが分かる。

キーワード	アクセス数
①大阪 マンション	21, 125
②関西 マンション	12, 115
③大阪 不動産	8, 295
④大阪 購入マンション	1, 496
⑤関西 マンション物件	964



＜アクセス解析＞

《 検索ワード Ranking 》 Analyzable : 48,874 《				
順位	検索エンジン	検索ワード	シェア	検索した訪問者数
1	[Google]	javascript	9%	4,410
2	[Google]	php	5.5%	2,720
3	[Google]	PHP	3%	1,493
4	[Google]	JavaScript	2.6%	1,316
5	[Yahoo!JAPAN]	javascript	1.6%	822
6	[Yahoo!JAPAN]	php	1.3%	670
7	[Google]	配列	0.9%	477
8	[Google]	select	0.9%	443
9	[Yahoo!JAPAN]	PHP	0.8%	403
10	[Google]	Javascript	0.7%	374
... 略 ...				

検索連動型広告の仕組み

ステップ2：キーワードのオークション

希望する検索連動型広告のキーワードをオークションで入札する。

キーワード 大阪 マンション

1位	500円	A社
2位	450円	B社
3位	300円	C社
4位	200円	D社
5位	100円	E社



検索連動型広告の仕組み

ステップ3：検索連動型広告の表示

検索サイトでのスポンサー表示

検索結果

検索結果1

検索結果2

検索結果3

検索結果4

検索結果5

スポンサー

1 A社

2 B社

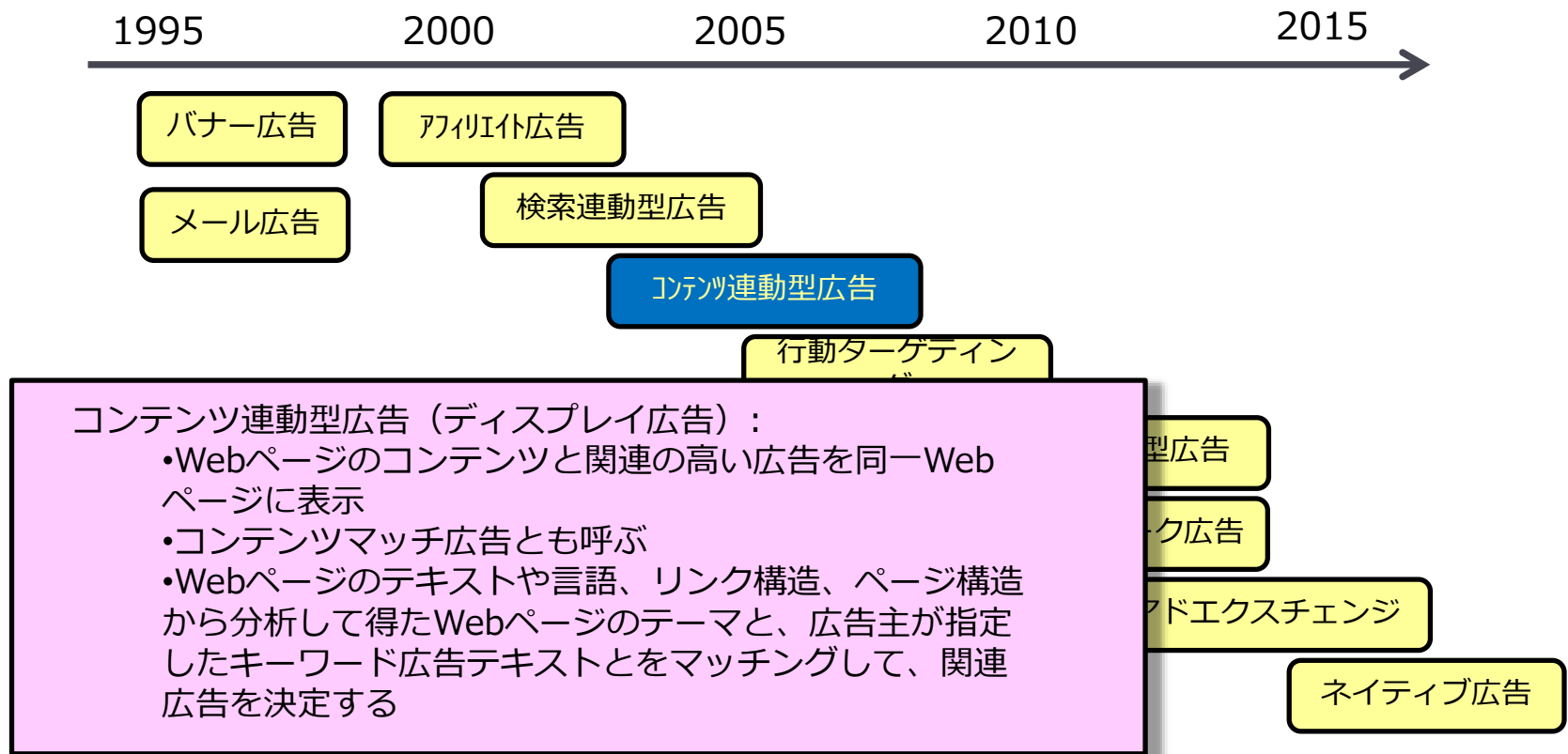
3 C社

4 D社

5 E社



インターネット広告 歴史(5)



コンテンツ連動型広告（ディスプレイ広告）



毎日の料理を楽しむに
cookpad

284万
レシピ

×

MYフォルダ

ひな祭り ちらし寿司 ほうれん草 鶏もも肉 半羽元

水菜のさっと煮



定番ですが…さっと簡単に出来ます

patisier

材料 (4人分)

水菜	1袋
豚肉 (コマ切れ)	150~200g
油揚げ	2~3枚
水	250cc
* 鰹だし(顆粒)	
* 薄口醤油	大さじ2~
* 砂糖	小さじ1/2
* 酒	大さじ1

1 水菜は、洗って根っこを落とし、4~5cmくらいに切る。油揚げは、油

2 鍋にお湯を沸かし、*印の調味料を入れ、豚肉に火が通るまで煮る。

3 水菜と油揚げを加え、さっと煮る。

4



南魚沼産コシヒカリ

JAAが誇る最上の100%南魚沼産コシヒカリを産地直送で販売。JA 魚沼みなみ 特産センター

毎週更新！おすすめ特集

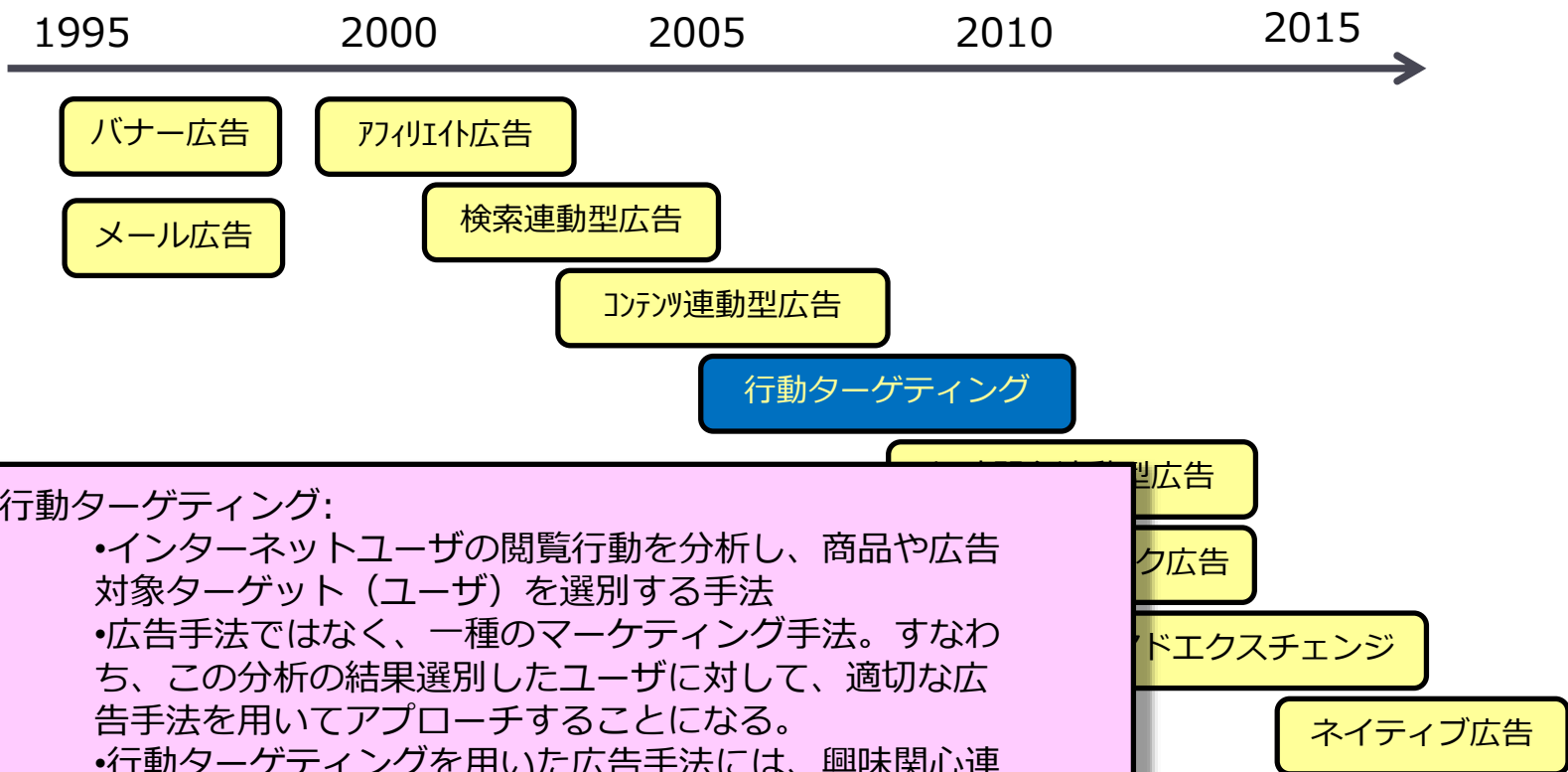
- 簡単！大絶賛！絶品おつまみ
- 失敗しない！簡単ちらし寿司
- みそバターが絶品！簡単一品
- ビールに合う！絶品旬レシピ
- 残りごはん♪絶品ランチ

もっと見る

クックパッドへのご意見をお聞かせください

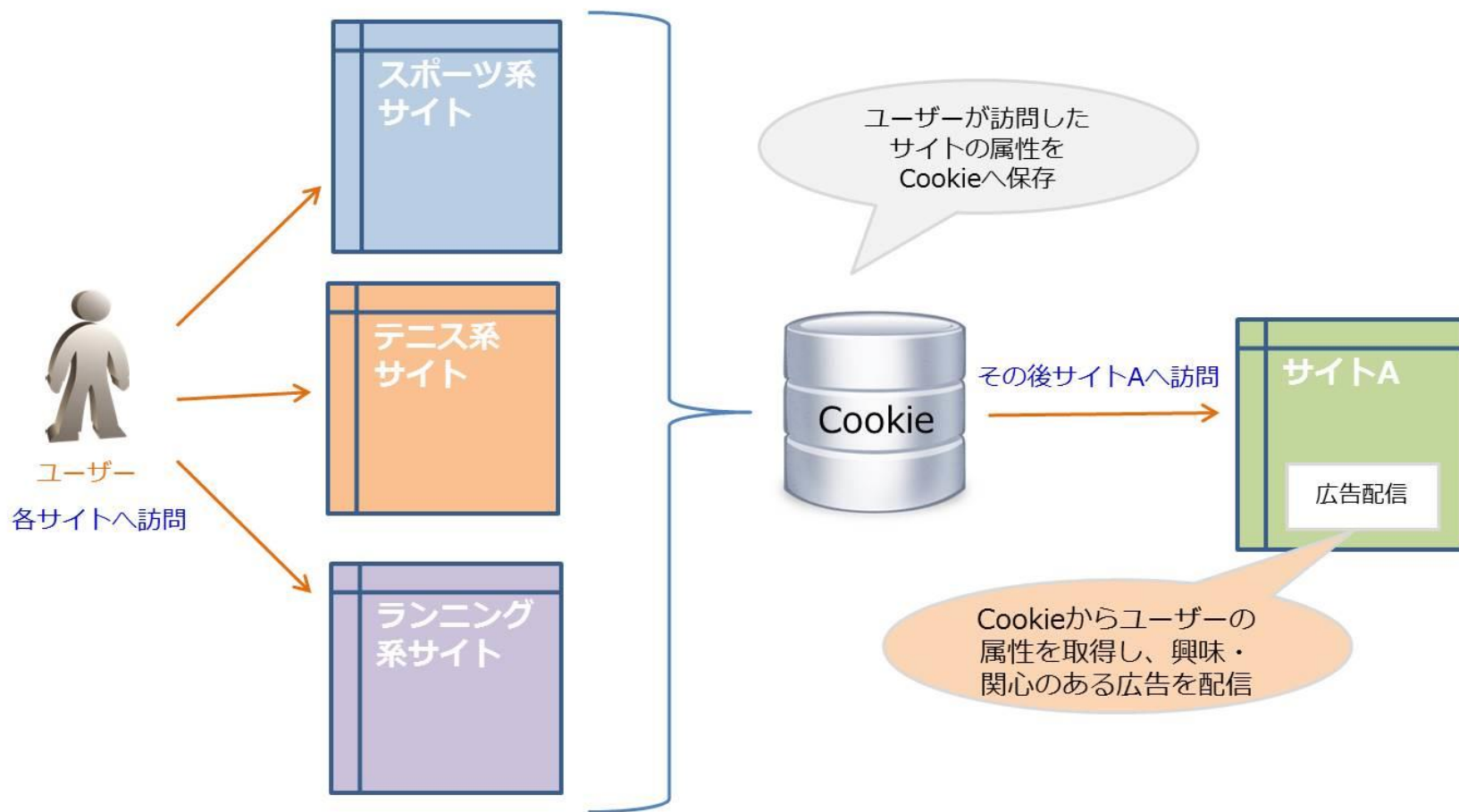
- サービスへのご意見・ご要望
- 機能の不具合

インターネット広告 歴史(6)

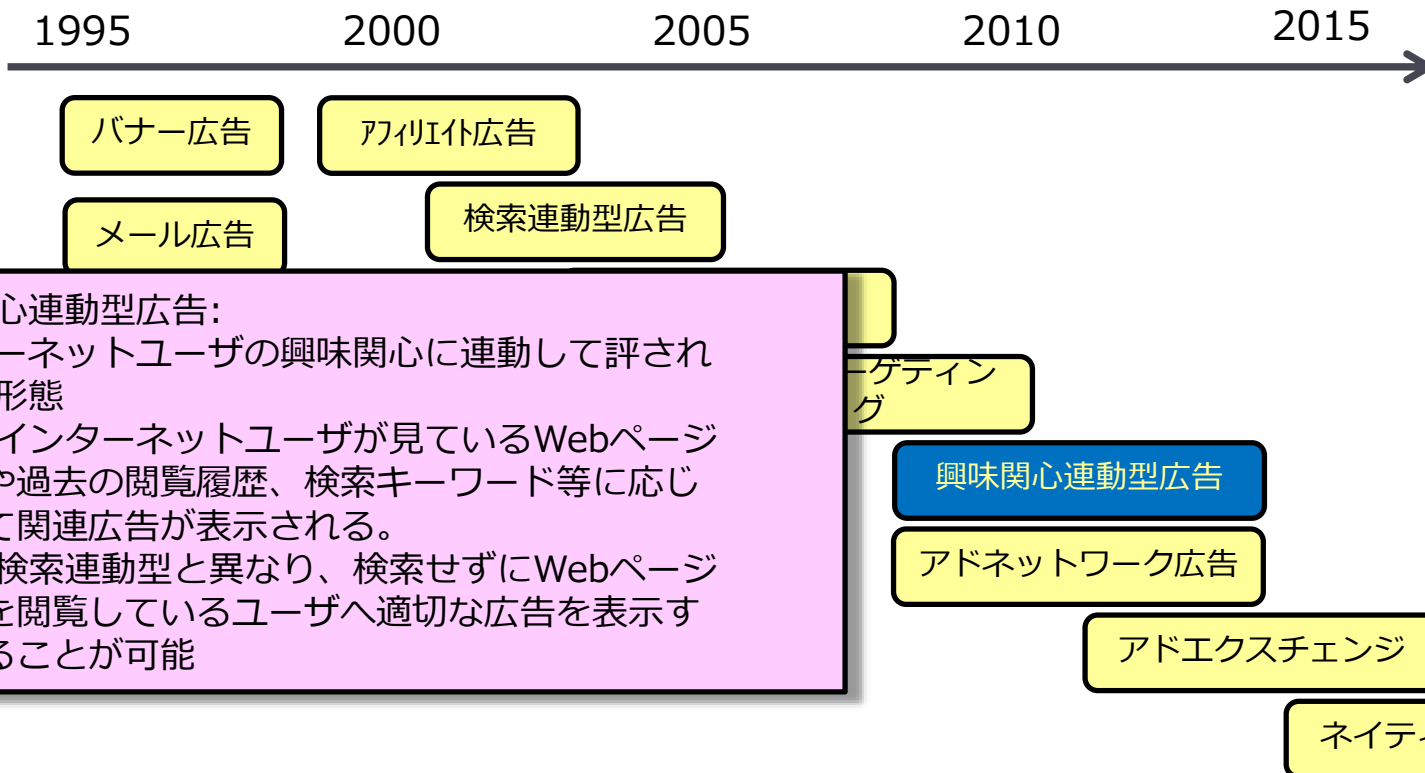


行動ターゲティング

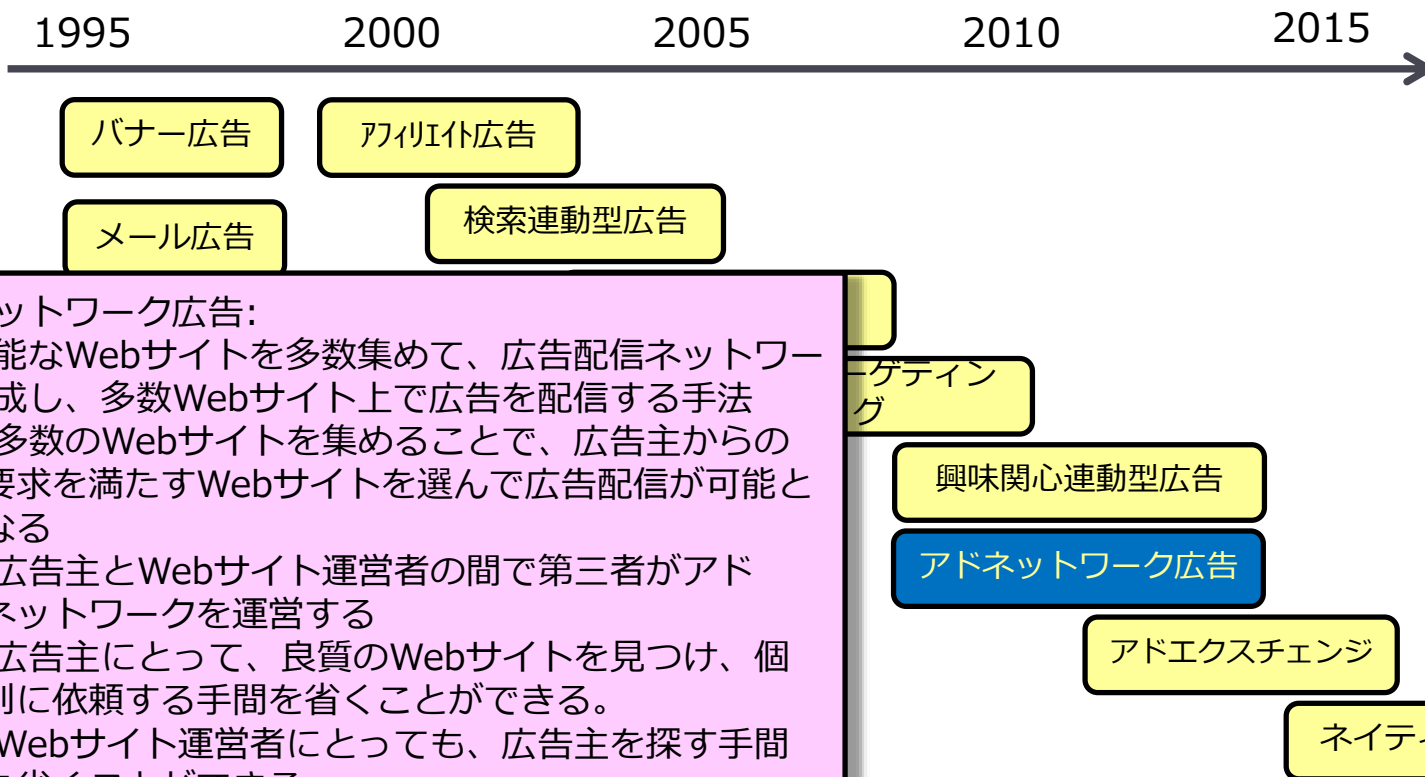
行動ターゲティング



インターネット広告 歴史(7)

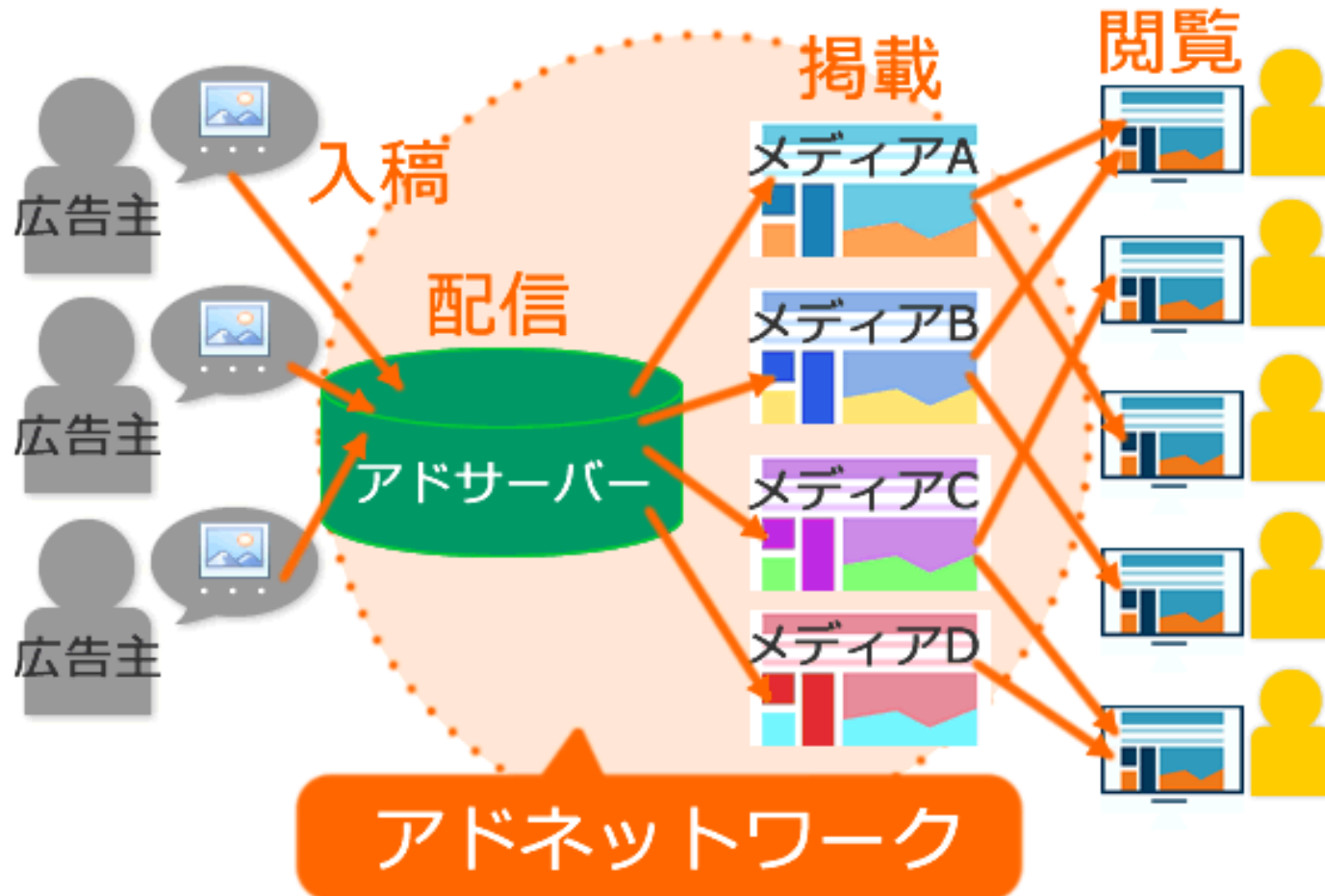


インターネット広告 歴史(8)

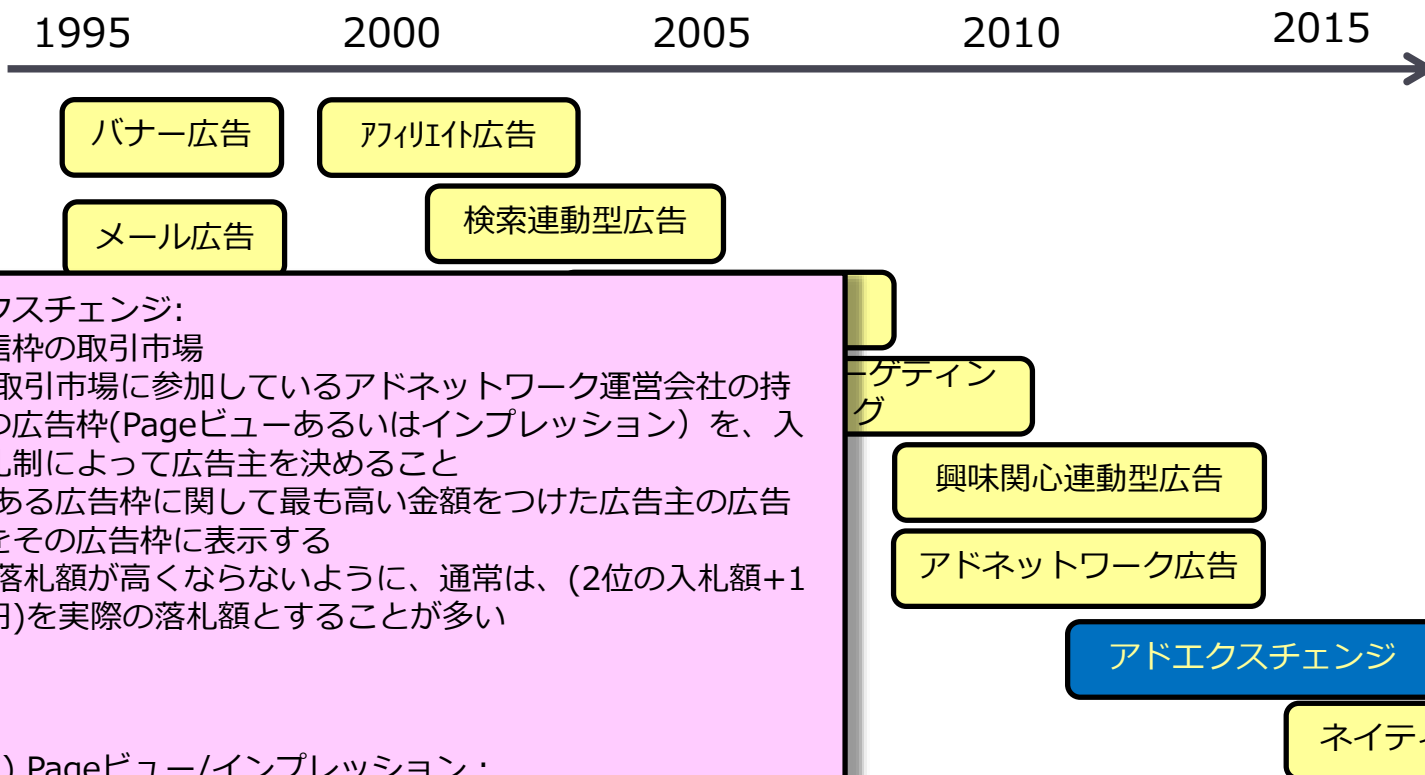


アドネットワーク広告(Ad Network Advertising)

アドネットワークの仕組み



インターネット広告 歴史(9)



アドエクスチェンジ: 広告配信枠の取引市場

- 取引市場に参加しているアドネットワーク運営会社の持つ広告枠(Pageビューあるいはインプレッション)を、入札制によって広告主を決めること
- ある広告枠に関して最も高い金額をつけた広告主の広告をその広告枠に表示する
- 落札額が高くなならないように、通常は、(2位の入札額+1円)を実際の落札額とすることが多い

注) Pageビュー/インプレッション:
Webページが読みこまれた回数, 閲覧回数

インターネット広告 歴史(10)

ネイティブ広告:

Webページのコンテンツの内容やデザインに広告を自然に溶け込ませ、ユーザにコンテンツの一部として広告を見てもらうことを目的とした広告概念

•次の条件を見たすもの

デザイン(本体と同一)、機能(本体と同様)、表示方法(本体と同様に表示)、配信方法(全ページではなく部分的に表示)、効果測定(ユーザと広告媒体との結びつきの度合いや絆の度合い、Engagement)、PR表記(広告であると明記)

•広告形式は、次の6種類に分類できる

インフィード型(Webページ内の記事間の広告)、検索連動型(検索連動型広告で上記条件を満たすもの)、レコメンドウィジェット型(レコメンド配信サービス利用のお勧め記事や関連記事)、プロモートリスティング型(検索上位にPRで表示される広告)、ネイティブな要素を持つインアド型(広告枠内に表示される本体コンテンツとの親和性の高い広告、カスタム型(その他)

10

<https://ferret-plus.com/images/5576529169702d0592ce0000>

Vivaldiの日本人COO様
なぜ今さらブラウザ業界に
参入ですか!?

ネットの夜明けから現代まで
「インターネットセキュ
リティ」の歴史を分かり
やすく解説するとうなる

ASCII.jp

GIGAZINE (ギガジン)

【アメックス】抽選で
海外往復航空券をプレゼント。

広告 アメックス

カメラ

二人

関心

ストリーフ広告

アドエクスチェンジ

ネイティブ広告

ネイティブ広告(Native Advertising)



インターネット広告の特徴

- インタラクティブ（あるいは能動性）
 - TVや新聞等の広告と異なり、インターネット広告は、ユーザの興味や閲覧目的等に適した広告を能動的に行うことが可能
- ターゲティング
 - 広告主が要求する対象ユーザに対して広告を実施することが容易
- 効果測定
 - 効果（広告と商品購入行動との関係等）を測定することが容易

インターネット広告の課題

- 広告の信頼性
 - うそ、誤認、誇大、誇張、不当表示等
 - 例： 広告を本体コンテンツと誤解
- ドメイン名紛争
 - わかりやすいドメイン、短いドメインが人気
 - ドメイン売買
- 利用者の機器や環境の違いへの対応(技術面)
- 著作権、プライバシー（個人情報収集）