

データマイニング事例集



理学博士達で創るIT会社
株式会社 アイズファクトリー



i's FACTORY co., ltd.
Data mining opens the door to tomorrow.

どの事例も、お客様のご要望に合わせて、自由にカスタマイズ可能です！

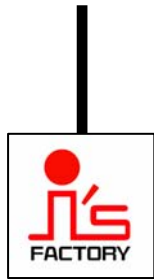
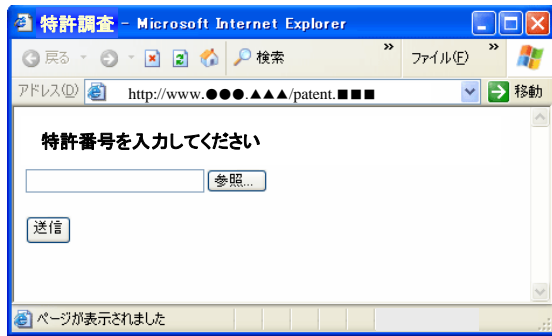
番号	事例	応用例	ページ
【1】	特許情報の類似検索、特徴語分析	インターネットからのマーケティングレポート作成	p.10
【2】	特許情報の時系列クラスター分析	コールセンターに寄せられる苦情の可視化 掲示板サイトの投稿傾向の可視化	p.11
【3】	携帯サイト投稿分析	問い合わせメールの自動分類 掲示板等のトピック把握や管理	p.12
【4】	POSデータを使った会員分類	ECサイト、小売店、銀行等の顧客分類	p.13
【5】	企業評価モデル	企業ごと成長性ランキング 株価予測	p.14
【6】	パーソナライズドWeb検索エンジン	携帯サイト、ポータルサイト、ブログ、サイトコールセンター等でのユニークかつユーザ満足度の高い検索	p.15
【7】	不適切コメントの自動チェックシステム	携帯サイト、ポータルサイト、ブログ、掲示板等に投稿される書込み、コメントの管理	p.16
【8】	名寄せ(定期チェックシステム)	金融機関等の多件顧客情報を持つ企業の顧客一元管理 M&A、部門譲渡などによって生じる顧客情報の統合	p.17

もし、お客様のイメージしているものが記載されていなくても、心配はご無用です。いつもお客様のご要望に沿えよう、柔軟に対応させていただきます。

■ 特許情報の類似検索、特徴語分析

分析目的 類似特許調査、技術評価を効率化する

イメージ図 応用対象

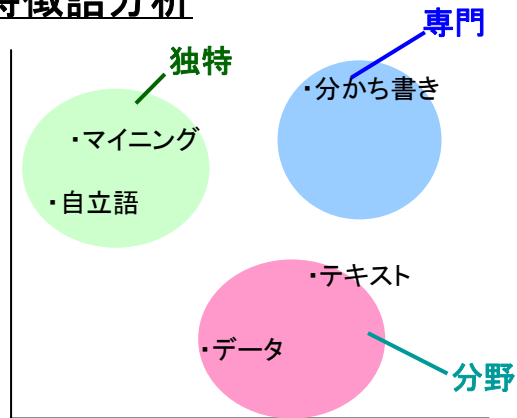


わずか**10分**
自動でPDF作成

類似特許抽出

- ・文書構造解析方法
自立語に分かち書きを行いテキストデータをマイニングする...
- ・マイニング装置
回答者から収集された大量のデータの中から、回答者の連想に...
- ・特徴語抽出システム
別の文書データ群に由来する概念辞書間で生じた単語に対する概念の違いや変化を検出...

特徴語分析



・インターネットからのマーケティングレポート作成

特徴

- ・Webインターフェースから分析が行える
- ・特許番号を入力するだけで良い
- ・PDFで自動レポート生成 (所要時間: 10分)

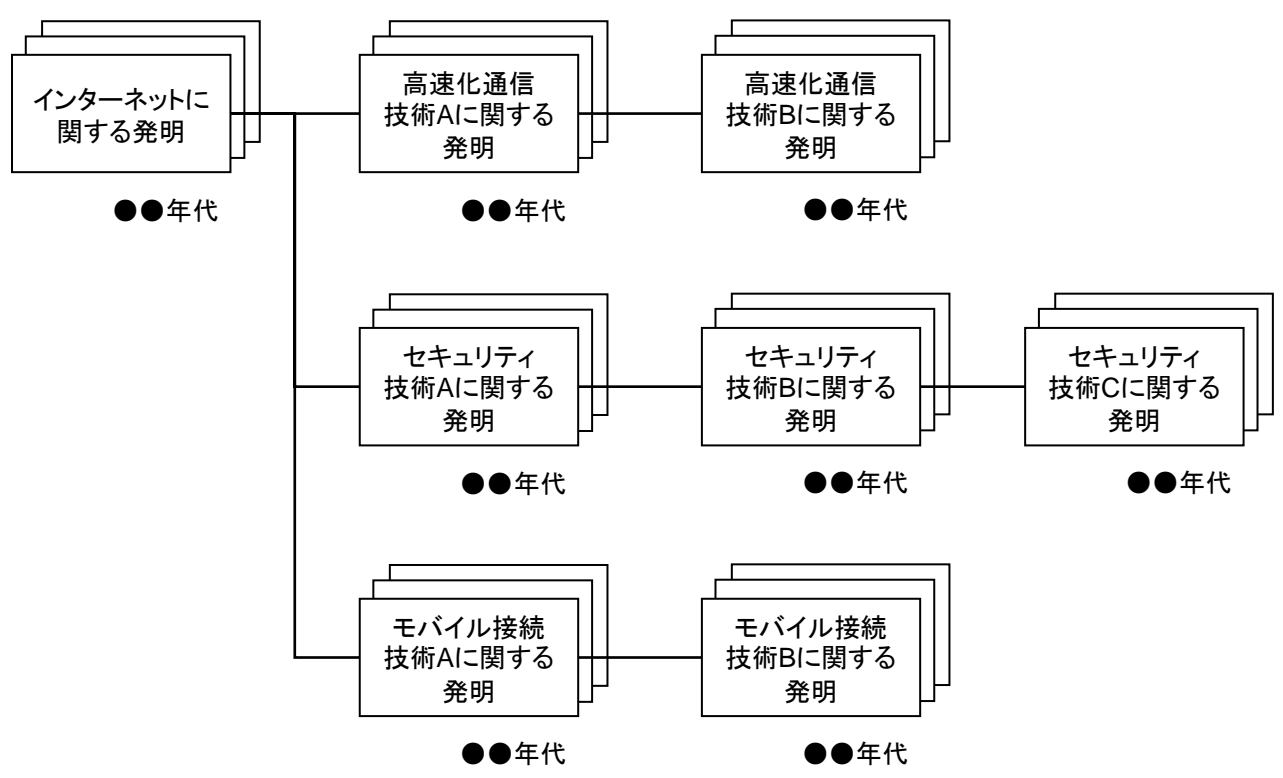
概略

- (1) 特許番号の入力
- (2) 特許広報700万件のキーワード重み付け、文書ベクトル化
- (3) レポート生成

■ 特許情報の時系列クラスター分析

分析目的 特許情報を可視化し、技術発展図を自動作成する

イメージ図 **応用対象**



応用対象

- ・コールセンターに寄せられる苦情の可視化
- ・掲示板サイトの投稿傾向の可視化

特徴

- ・自動で図を作成できる
- ・時系列クラスター分析ができる
- ・クラスターへの自動タイトル付与

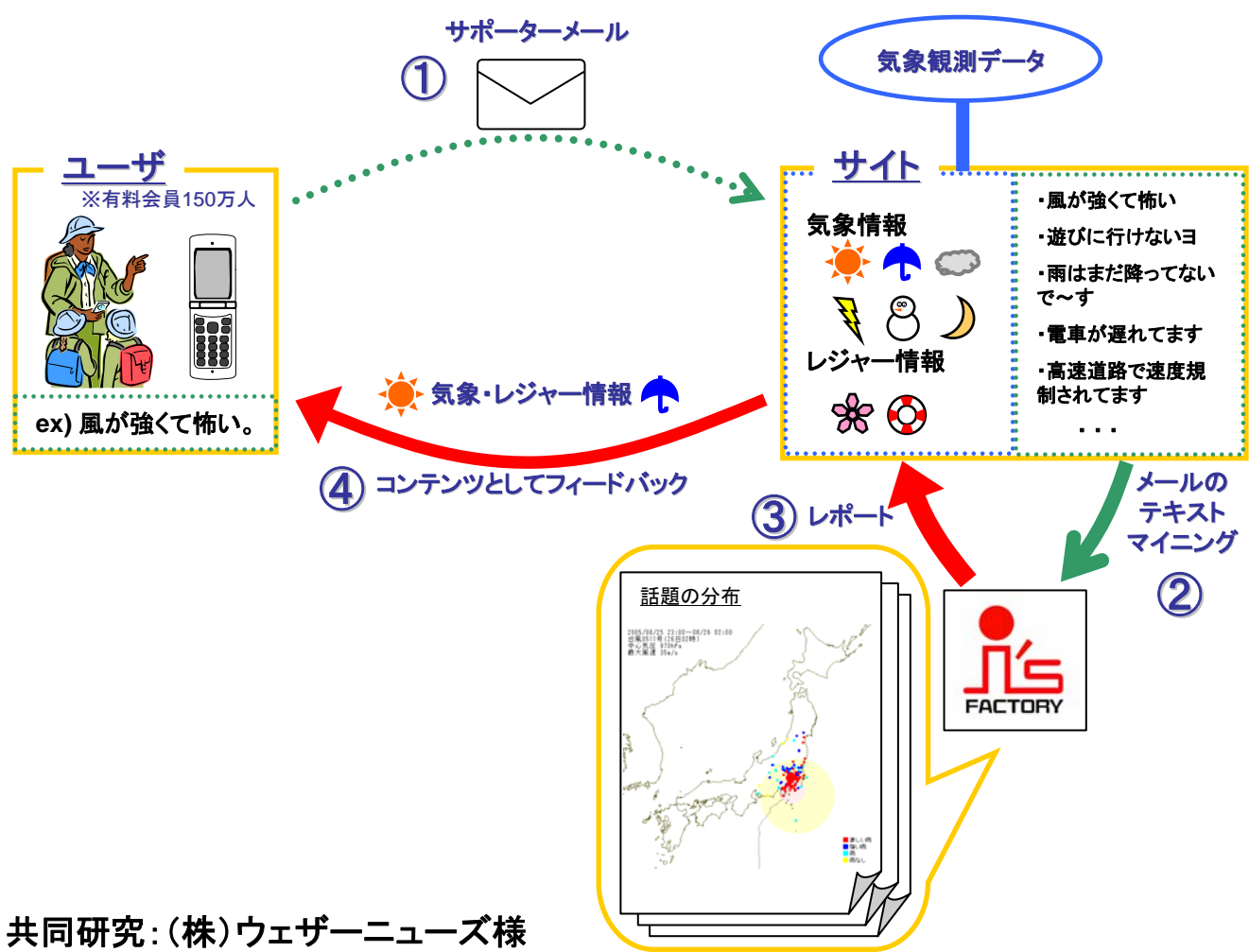
概略

- (1) 特許文書の分類
- (2) 文書群内容の自動判定
- (3) 技術年代の自動判定

■ 携帯サイト投稿分析

分析目的 CGMにより収集した身近な気象情報から、コンテンツを作成する

イメージ図



応用対象

- ・問い合わせメールの自動分類
- ・掲示板等のトピック把握や管理

特徴

- ・短い文章に適用可能
- ・投稿メールの自動的な高精度カテゴリ分類
- ・特徴語抽出によるトピック把握

概略

①サイトユーザ150万人からのメール投稿テキストを②特徴語解析して③レポートニングし④コンテンツとしてユーザにフィードバックすることで、Grass Roots (草の根)的に身の回りの情報を取得することができます。

共同研究: (株) ウェザーニューズ様

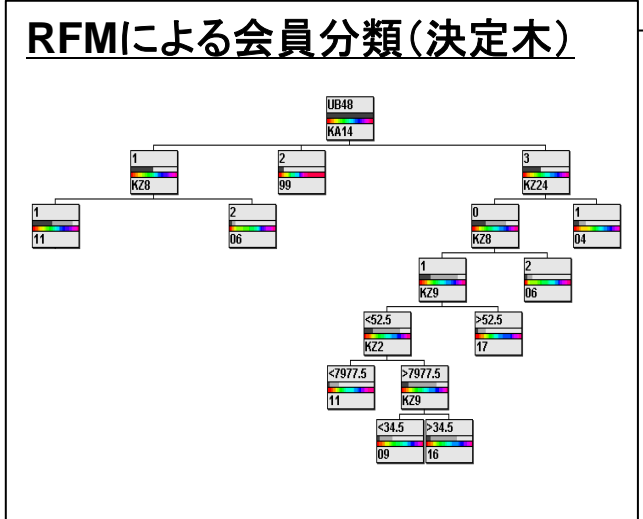
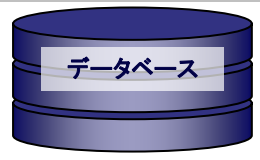
■ POSデータを使った会員分類

分析目的

ダイレクトメール配信を効率化する

イメージ図

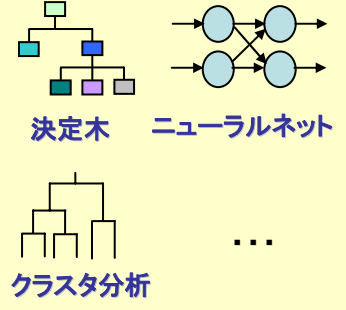
POS、会員属性、カード利用履歴



最適なセグメント化

<p>グループA</p>	<p>グループB</p>	<p>グループC</p>	<p>...</p>
<ul style="list-style-type: none"> ・男性、60代 ・回数：大 ・酒類 	<ul style="list-style-type: none"> ・女性、40-50代 ・金額：大 ・バッグ 	<ul style="list-style-type: none"> ・小中学生 ・回数：大、金額：小 ・食品 	

目的に応じた手法



応用対象

- ・ECサイト、小売店、銀行等の顧客分類

特徴

- ・隠れた特性を発見できることがある
 - ・目的に応じて様々な手法を用いることができる
- (1) 決定木
 - (2) ニューラルネット
 - (3) クラスタ分析 ほか

概略

従来のクレジットカードの会員分類は、非常にマーケットがセグメント化してきているため、非効率である。
 様々なデータマイニング手法を用いて**最適なセグメント化**を行うことができる。

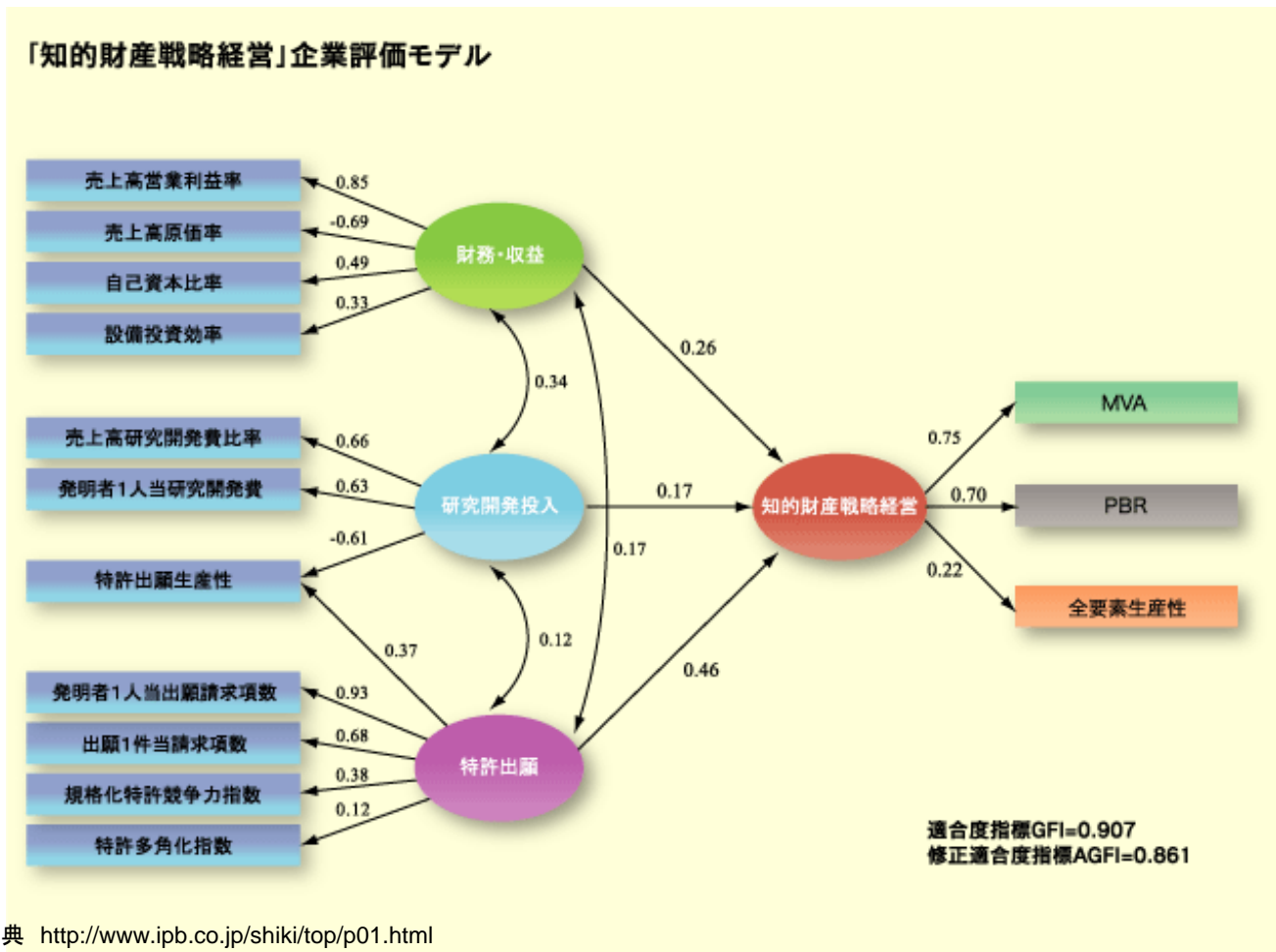
■ 企業評価モデル

分析目的

隠れた相関性・因子の発見

イメージ図

応用対象



・企業ごと成長性ランキング
・株価予測

特徴

・原因と結果の相関を数値モデル化する

概略

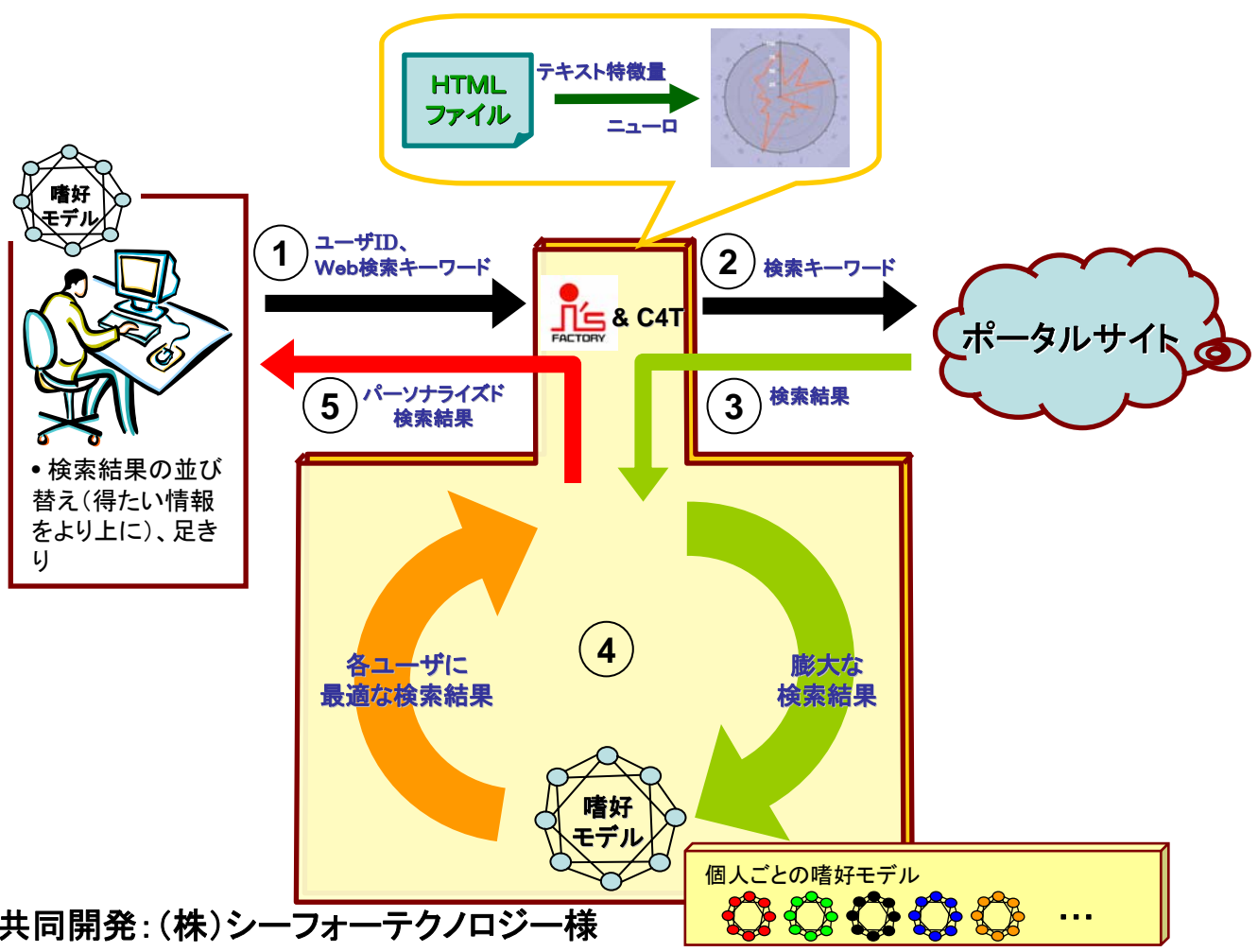
財務指標、経済指標等から
 (1) 企業の収益にどの因子が影響しているか？
 (2) 各因子同士の相関はどれくらい強いかなどを探索する

出典 <http://www.ipb.co.jp/shiki/top/p01.html>

■ パーソナライズドWeb検索エンジン

分析目的 個人の趣味嗜好に応じて結果が変化する検索エンジンを提供する

イメージ図 **応用対象**



応用対象

- ・携帯サイト、ポータルサイト、ブログ、サイトコールセンター等でのユニークかつユーザ満足度の高い検索

特徴

- ・各ユーザにとって、最適な検索結果を簡単に提供することができる
- ・各ユーザごとに嗜好モデルを構築
- ・テキストデータの数値化(特徴量)により、ニューラルネットワークが応用可
- ・ニューラルネットワーク技術により、自動学習が高速・高精度になり、最適な推測・予測が可能(C4T共同開発)

概略

気に入ったサイトをお気に入りに登録するのと同様に、Web検索エンジンへ、自分の趣味嗜好を学習させる。

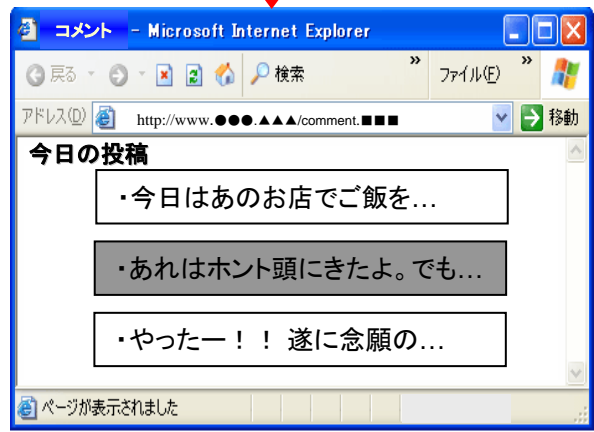
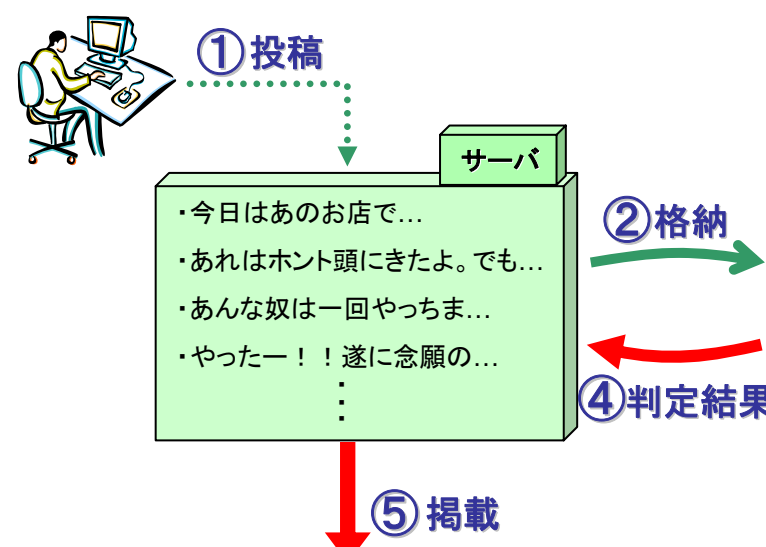
すると誰でも簡単に、自分が本当に探しているサイトへ、辿り着けるようになる。

共同開発: (株)シーフォーテクノロジー様

■ 不適切コメントの自動チェックシステム

分析目的 Webサイト投稿コメントの管理を効率化する

イメージ図 応用対象



応用対象

- ・携帯サイト、ポータルサイト、ブログ、掲示板等に投稿される書込み、コメントの管理

特徴

- ・人的作業の**大幅減**
- ・**学習**によるフィルタ最適化
- ・自動/手動の**役割分担**による最適学習
- ・サイト内容に応じて**カスタマイズ可**
- ・不適切コメントの3段階評価
- ・grey、black判定コメントに関して、**アラートメール**を管理者へ送付
- ・blackコメントの**自動排除**

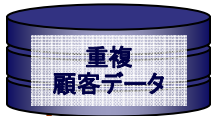
概略

不適切判定フィルタによって投稿コメントを3段階(white, grey, black)に評価する。blackは自動的に排除し、greyはメールでアラートされる。人がgreyコメントを最適判定し、フィルタに学習させる。フィルタは、**使えば使うほど最適化**されていく。

■ 名寄せ(定期チェックシステム)

分析目的 顧客情報の一元化を行い、最適状態を維持する

イメージ図 **応用対象**



鈴木太郎 東京都千代田区神田錦町2-9-1
 鈴木 太郎 神田錦町二丁目九番一号 ○○ビル
 スズキ タロウ 千代田区神田錦町二丁目九番一号
 鈴木太郎 千代田区神田錦町2-9-1○○ビル4F
 佐藤 次郎 東京都千代田区須田町3-3-2 2F
 サトウ ジロウ 千代田須田町3丁目3の2 △△ビル
 高橋 三郎 東京都千代田区内神田4丁目1番地の5
 高橋三朗 千代田区内神田4-1-5コーポ××302
 …



A 鈴木 太郎
 東京都千代田区神田錦町2-9-1 ○○ビル 4F
 B 佐藤 次郎
 東京都千代田区須田町3-3-2 △△ビル 2F
 C 高橋 三朗
 東京都千代田区内神田4-1-5 コーポ×× 302号
 …

定期チェックシステム

通常、新たに追加される顧客データによって、再び重複が生じてしまう。
 この重複した顧客データを、定期的に名寄せすることのできるシステム(データの種類によってカスタマイズ)を構築する。
 名寄せ処理は、簡単な操作で行うことができ、人手で行うのに比べ、長期的に高い投資対効果が期待できる。

応用対象

- ・金融機関や、多件数の顧客情報を有する企業の顧客一元管理
- ・M&A、部門譲渡などによって生じる顧客情報の統合

特徴

- ・表記揺れに対応
- ・重複判定の段階評価
- ・一度きりでない、定期チェックシステム
- ・名前、住所、電話番号など複数の情報を取り扱い可
- ・90万件の重複データ処理の実績あり

概略

顧客データを正規化(全角/半角などの表記統一)し、テキストマイニング業務で培った弊社独自の技術で重複判定を行い、一元化を行う。
 チェックシステムを構築し、定期的な名寄せを可能とする。

- ### 名寄せ処理
1. 顧客データの正規化 (全角/半角などの表記統一)
 2. 重複判定(表記揺れに対応)
 3. バッチ処理