

# コミュニティ支援に関する実践研究と Weblogを用いた関心动向分析

---

東京大学人工物工学研究センター  
価値創成イニシアティブ(住友商事)寄附研究部門  
客員教員  
福原知宏



# 本日の内容

---

## 1. コミュニティ支援に関する実践研究

- KDDI FTTHトライアル(2002～2003年)をはじめとするコミュニティ内知識共有・創造に関する実践研究について

## 2. Weblogを用いた社会の関心動向に関する分析研究

- リスクコミュニケーションのための基礎データとして

- 知的コミュニティ基盤研究センターにおける研究活動との接点を探りたい(途中でのご質問もどうぞ:-)
-

## Part I

# オンラインコミュニティにおける知識共有と 分析活動支援に関する研究

PUBLIC OPINION CHANNEL

# 研究背景

## ■ インターネット上の様々なコミュニティ

– ブログ, 掲示板, SNS, オンラインゲーム, 仮想環境 (Second Life)

- 知識の共有と創出に貢献
- 既存の組織に比べ, 効果的に知識を共有し発展させる可能性

– 知識創造コミュニティ

- 円滑に情報共有し, 互いに競争/協調する中で新たな価値を生み出す集団

## ■ 知識創造コミュニティとは何か?

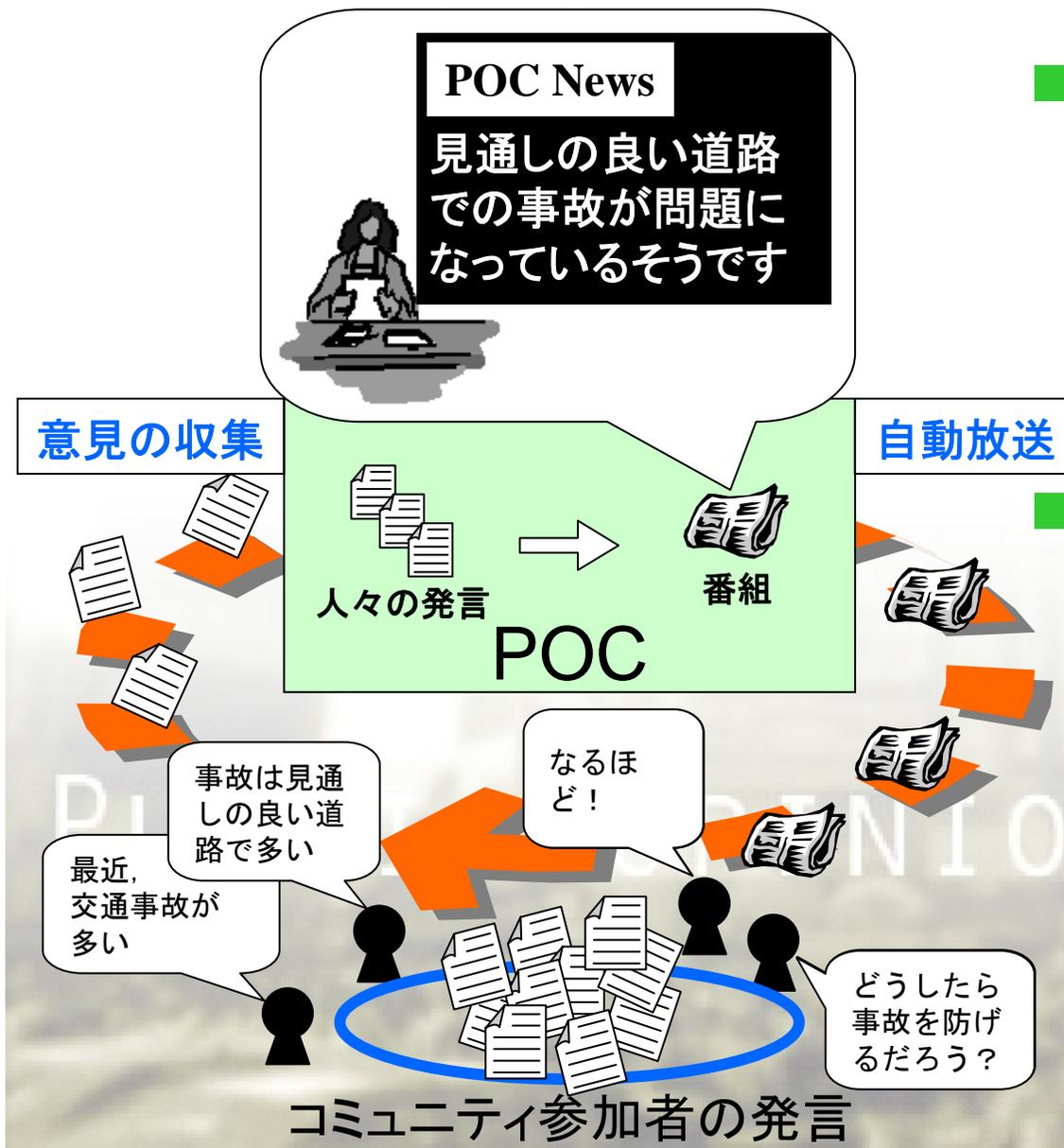
– 何がコミュニティを知的にするのか?

- 発言の量? 質? 雰囲気? 参加者数?

– どうすれば知識創造コミュニティを形成し, 維持できるのか?

コミュニティを知的にする要因を探り,  
得られた知見に基づいて知識創造コミュニティを実現したい

# 研究プラットフォーム： パブリック・オピニオン・チャンネル(POC)



## ■ 「街のつぶやき」を集める コミュニティ放送局

- コミュニティ参加者の些細な発言(知識)の収集
- 自動放送による知識の流通

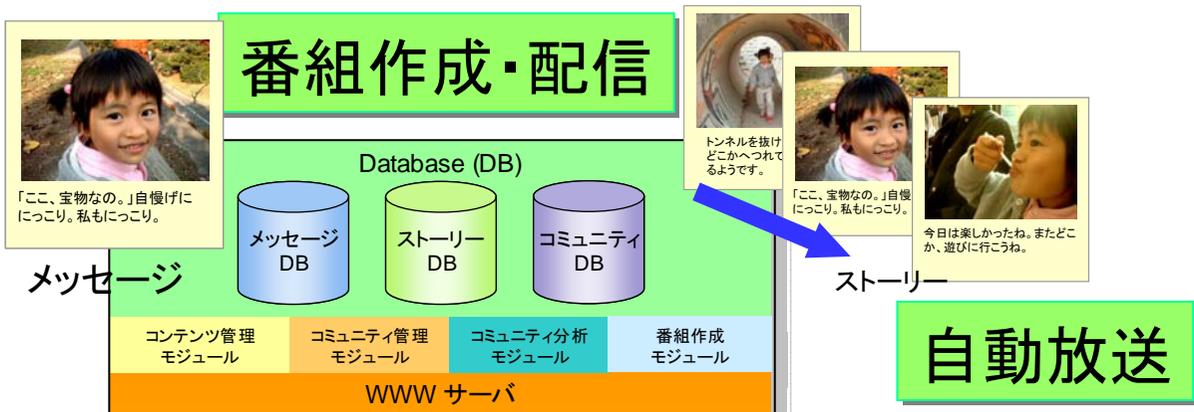
## ■ 意見の収集と放送による 知識循環

- コミュニティ内の人々の意見や考えが分かる
- コミュニティ内の情報共有, 意見交換の促進 (情報の風通しを良くする)

# POC実装システム

- メッセージ
  - 知識の断片
- ストーリー
  - 構造化された知識

情報発信



自動放送



POC Communicator



POCTV

コミュニティ

# POCで用いるデータ

## ■ メッセージ: 番組断片

- 100文字程度の文章と写真
  - タイトル, チャンネル, ファイル名, 投稿日時, IPアドレス, 投稿者のアカウント, 写真と本文
- 知識の断片
  - ストーリー(下記)の部品
  - 再利用を意識

## ■ ストーリー: 放送番組

- 順番を持つ複数のメッセージによるまとめ
- 手動および自動で生成

```
<?xml version="1.0" encoding="Shift_JIS" ?>
<opinion name="anonymous" img="1006746170.jpg">
  <title>My favorite park</title>
  <comment>This is my favorite park, she smiled.</comment>
  <url />
</opinion>
```

XML 表記

My favorite park



"This is my favorite park", she smiled.

メッセージ



トンネルを抜け  
どこかへつれて  
るようです。



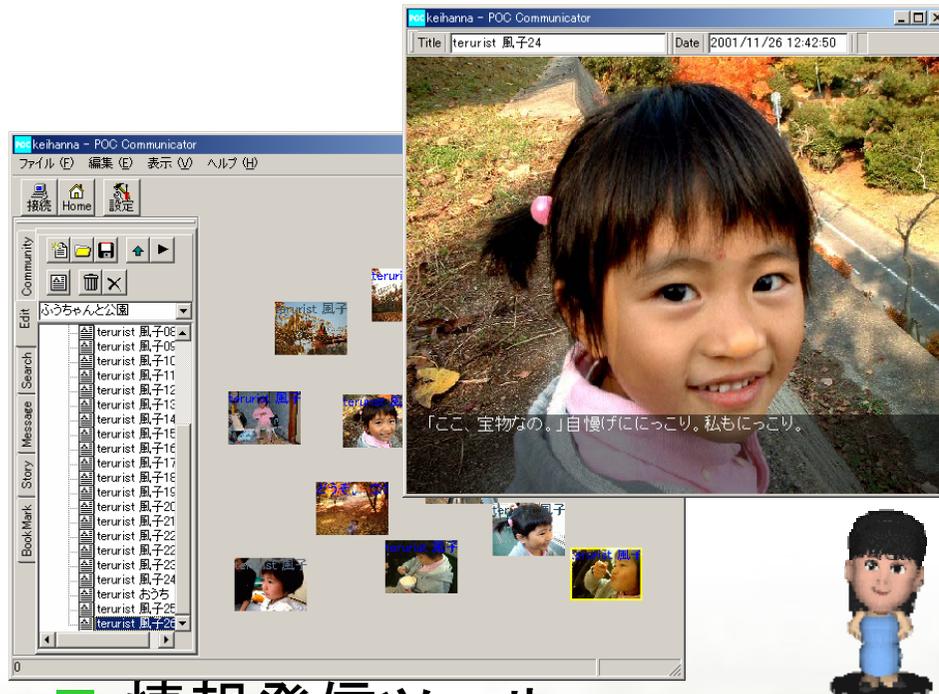
「ここ、宝物なの。」自慢  
にっこり。私もにっこり。



今日は楽しかったね。またどこ  
か、遊びに行こうね。

ストーリー

# 情報発信・閲覧ツール



## ■ 情報発信ツール: POC Communicator

- メッセージとストーリーの作成ツール
- **メッセージの循環提示**による過去発言からの新たな発言の引き出し



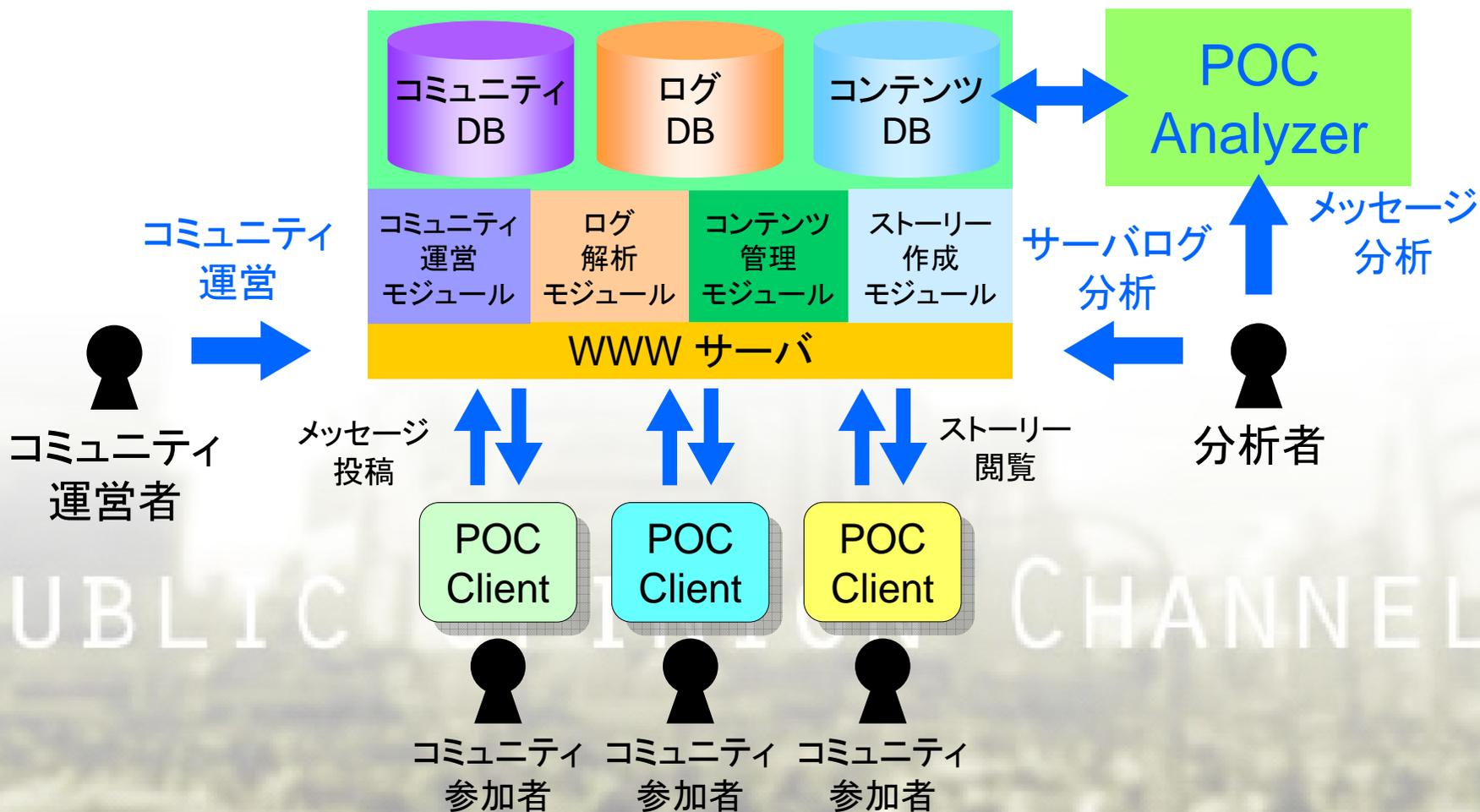
## ■ 情報閲覧ツール: POCTV\*

- テレビ番組(トークショー)形式による番組視聴ツール
- 書き言葉話し言葉変換による聞き手の理解の向上

\*久保田 秀和, 山下 耕二, 福原 知宏, 西田 豊明.  
POC caster: インターネットコミュニティのための  
会話表現を用いた情報提供エージェント,  
人工知能学会論文誌, Vol. 17, No. 3, pp.313-321(2002).

# POCにおける コミュニティ運営・分析支援機能

## POC サーバ



# コミュニティ支援システムの開発サイクル

システム  
開発計画

分析者

分析

- ・コミュニティの観察／調査
- ・実験データの分析

実験データの収集

コミュニティ支援  
システムの  
開発サイクル

システムへの設計指針／  
フィードバック

実験

- ・コミュニティ内での  
システム運用

コミュニティ  
運営者

設計

- ・システム設計案／  
修正案の作成

システム  
設計者

設計案／修正案の  
システムへの実装

PUBLIC OPINION CHANNEL

# 目的

## ■ 実験過程支援のための基本機能の提供

### 1. データ分析作業の支援

- サーバログの分析
- メッセージの分析

### 2. コミュニティ運営作業の支援

- コミュニティの設定
- ユーザの管理
- コンテンツの管理



# コミュニティ運営・コンテンツ管理モジュール

Public Opinion Channel: Message Manager - Microsoft Internet Explorer

ID	Author	Title	Date	Actions
252	khn07	bbs1025855211.xml この花は、何でしょう?	05 16:46:51	削除 同意
253	khn07	bbs1025855034.xml そう、トロピカルムード溢れる花	2002-07-05 16:43:54	メッセージの削除 同意 削除
254	khn07	bbs1025854968.xml この花は、何でしょう?	2002-07-05 16:42:48	メッセージの削除 同意 削除
255	khn07	bbs1025854276.xml kubo5舟下り	2002-07-05 18:31:16	メッセージの削除 同意 削除
256	khn06	bbs1025852848.xml 鯉鱒柄のエントランス	2002-07-05 16:07:28	メッセージの削除 同意 削除

メッセージ・番組一覧

Public Opinion Channel: Members Services - Microsoft Internet Explorer

Title	Body
terurist 風子24	「ここ、宝物なの。」自慢げにこっそり。私もこっそり。
terurist おうち	「おうちにかえろ。」
terurist 風子25	帰りに、アイスを買ってもらって上機嫌。ちょっと大きいかな?
terurist 風子26	...

ストーリー内容表示

Recent Postings - Microsoft Internet Explorer

No.	Type	Title	Author	Community	Date	Time
0	MESSAGE	大規模テキスト知識ベースに基づく自動質問応答 - ダイアログナビ		nklab-public	2002-06-27	21:33:09
1	MESSAGE	test		kc_test	2002-06-27	21:26:54
2	MESSAGE	2001年1月23日JIPPO ECワークショップ		nishida-internal	2002-06-27	14:12:05
3	MESSAGE	読書百篇意自ら通ず		nishida-internal	2002-06-27	13:59:53
4	MESSAGE	飲んで、食べて、笑って		nishida-internal	2002-06-27	13:59:03
5	MESSAGE	...		nishida-internal	2002-	0:58:23

最近の投稿傾向

movie コミュニティの設定

コミュニティ情報

コミュニティ名: movie

Owner: xxxxxx

Description: コミュニティの説明を変更する場合は下の欄を書き換えてください。映画の話題はこちらへ

現在の参加者一覧

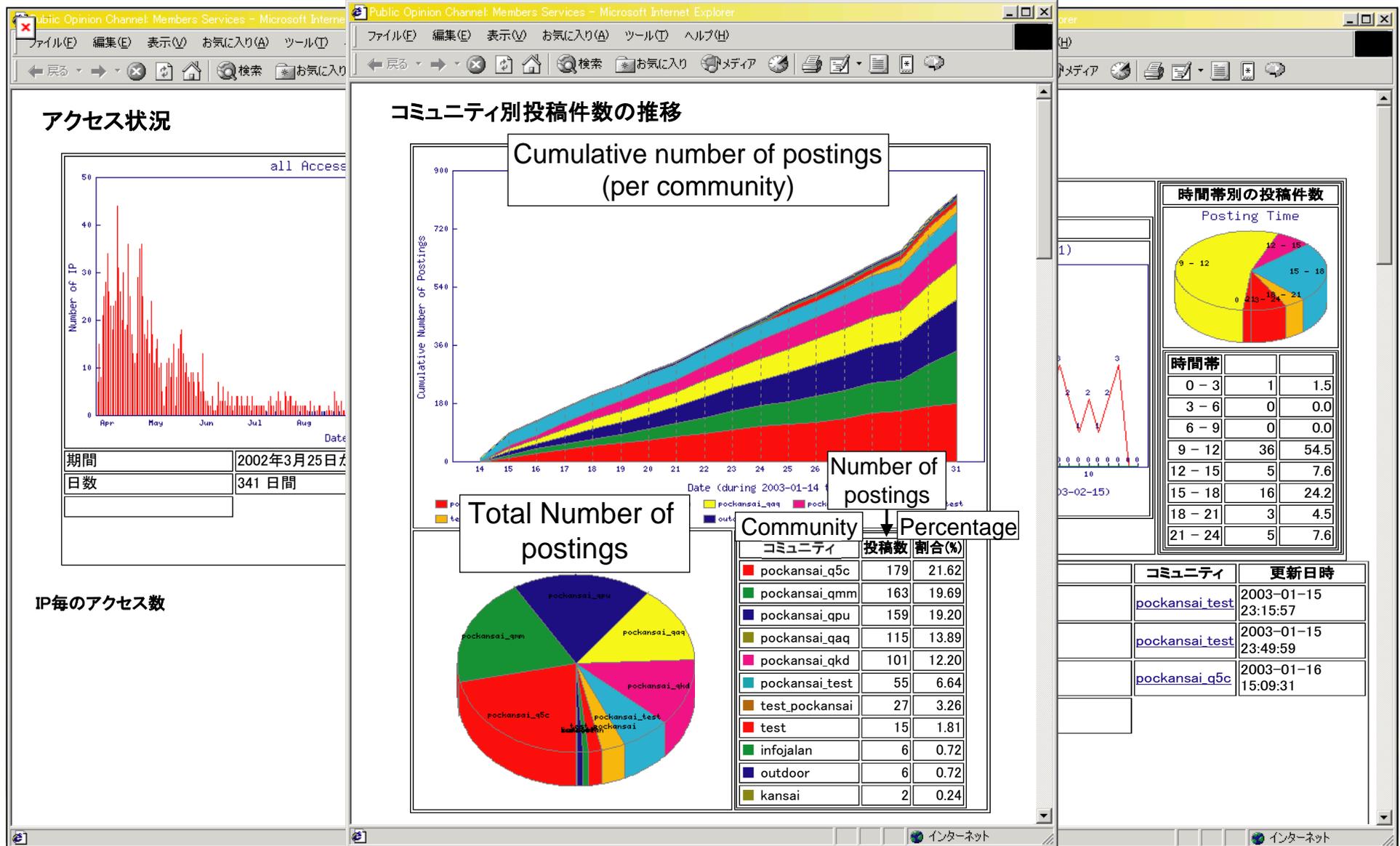
コミュニティの作成・設定

kc_talk	設定	東大「知識コミュニケーション/コミュニティウェア・グループウェア・デザイン」講義受講者の意見交換のためのチャンネル(投稿, 閲覧: 講義受講者)	2 件	84 件 (4 件増)
nklab-internal	設定	東大・西田・黒橋研究室の内部チャンネル(投稿, 閲覧: 西田・黒橋研究室)	10 件	10 件 (3 件増)
nklab-public	設定	東大・西田・黒橋研究室のパブリックチャンネル(投稿: 西田・黒橋研究室, 閲覧: 限定無し)	3 件 (1 件増)	3 件

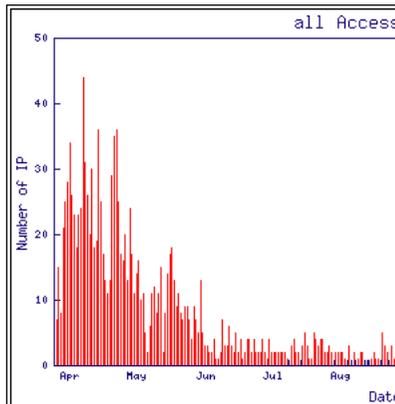
コミュニティ一覧

- コミュニティ運営者用インターフェース
- 全てブラウザ上でコミュニティ設定/コンテンツ確認可能

# ログ解析モジュール: レポート



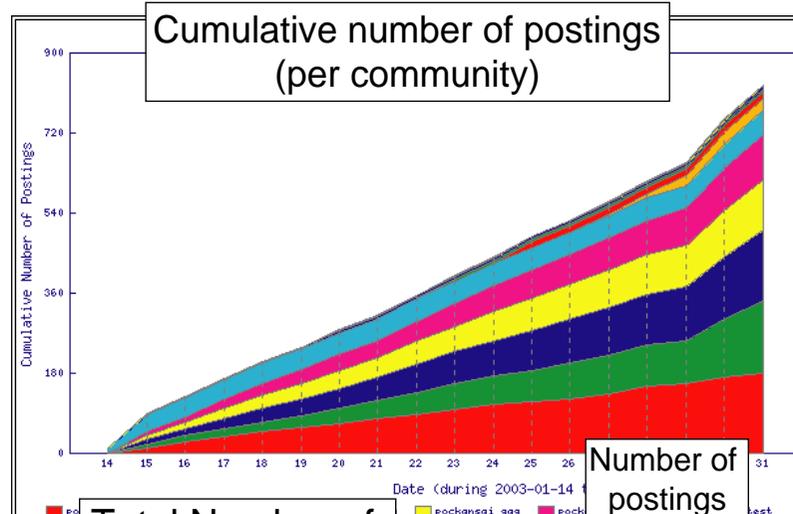
## アクセス状況



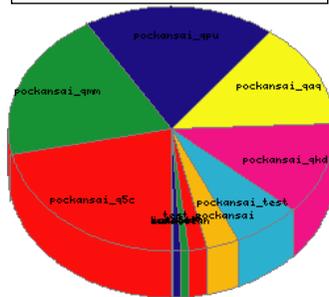
期間	2002年3月25日
日数	341 日間

## IP毎のアクセス数

## コミュニティ別投稿件数の推移



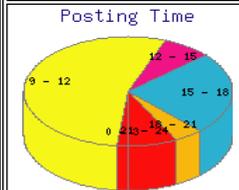
## Total Number of postings



## Community Percentage

コミュニティ	投稿数	割合(%)
pockansai_q5c	179	21.62
pockansai_qmm	163	19.69
pockansai_qpu	159	19.20
pockansai_qaq	115	13.89
pockansai_qkd	101	12.20
pockansai_test	55	6.64
test_pockansai	27	3.26
test	15	1.81
infojalan	6	0.72
outdoor	6	0.72
kansai	2	0.24

## 時間帯別の投稿件数



時間帯	投稿数	割合(%)
0 - 3	1	1.5
3 - 6	0	0.0
6 - 9	0	0.0
9 - 12	36	54.5
12 - 15	5	7.6
15 - 18	16	24.2
18 - 21	3	4.5
21 - 24	5	7.6

コミュニティ	更新日時
pockansai_test	2003-01-15 23:15:57
pockansai_test	2003-01-15 23:49:59
pockansai_q5c	2003-01-16 15:09:31

# POC Analyzer: メッセージ解析ツール

The screenshot displays the POC Analyzer interface. On the left, a network graph shows nodes representing messages (e.g., bbs161.xml, bbs162.xml) connected by arrows. A yellow box highlights a specific node with its details: 発言者: STM10j, タイトル: Re:[174] [172] 梅便り, 発言時刻: 2002/03/02 00:25. On the right, a text box contains the message content: いやはや失礼しました。花といえば、近世のお方は桜と思いきや、正統”梅”派もおられましたな。拙宅付近でも練習の甲斐あって上達した鶯の音が聞けますぞ。まずは春を愛でに中山さんに、爛漫か散りの春感を堪能に高台寺にと参りますか。 Below the graph, a table shows network statistics: 総ノード数: 12, 総リンク数: 15, 密度: 0.227273. At the bottom, a table lists analysis results for various nodes.

メール, NetNews, Web掲示板ログ等に適用可能

グラフの状態を示すネットワーク分析指標 (e.g., 密度の高い状態は話題の同質性を示す)

メッセージ間の明示的参照関係や類似度に基づく暗黙的关系を元にグラフ作成

キーワード及び類似度指定でグラフ作成

ID	ノード数	最大入次数	最大出次数	最大平均パス長	最大中心性	密度
6	15	3	4	2.25	0.25	0.17
7	4	3	3	1.00	0.00	1.00
8	5	3	4	1.00	0.00	0.80
9	12	3	4	2.73	0.46	0.23
10	3	2	2	1.00	0.00	1.00
11	4	3	2	1.00	0.00	0.67
12	2	1	1	1.00	0.00	1.00

解析結果を一覧表示

- メッセージ間の関係を数値化しグラフ表示
- 参照/被参照関係を元にグラフ生成
- メッセージ間類似度を計算して暗黙的关系も特定

# KDDI FTTHトライアルにおける 実証実験(2002年～2003年)

- 2002年3月25日～2003年2月28日  
(約1年間)
- 東京都新宿区・文京区の443世帯
- 光回線による100MbpsのLAN提供,  
セットトップボックス(STB)の配布
  - ビデオオンデマンド, インターネットカラ  
オケなど
  - POCTVによるコンテンツ配信
- サービス
  1. 我が街ストリーム
    - 地域情報やニュース, 名所や催し物情  
報など, 最新情報を含むコンテンツを  
定期的に配信
    - 3,000件の初期コンテンツ
  2. おしゃべり掲示板
    - システムに関する問い合わせなど



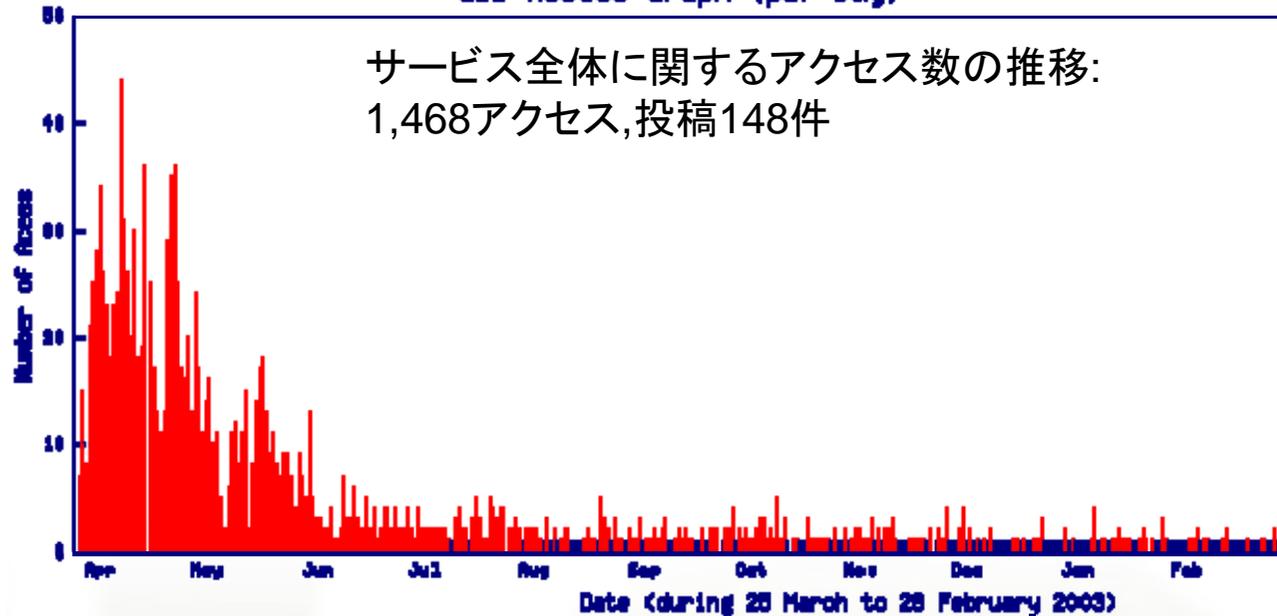
セットトップボックス(STB)と画面



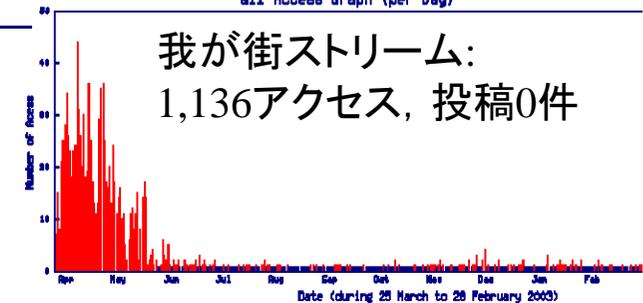
我が街ストリームで使われた写真

# FTTHトライアルにおける実証実験

all Access Graph (per Day)



all Access Graph (per Day)



all Access Graph (per Day)



- フレーミングや無責任な発言は見られなかった
  - 利用規約の制限もあり発言数自体少なかった
- 我が街ストリームにおける知識提供
  - 繰り返しやってくる視聴者(リピータ)の存在
  - 1回数時間に渡る視聴

# 教育場面へのPOC応用:

## 「インタラクティブ・メディアを利用した知識創造過程の研究」\*

- (財)大阪市青少年活動協会の協力の下, POCを使った『伝えるプロセス』の支援に関する実験を実施
  - 2005年12月18日から2006年3月5日まで計6回
  - 大阪市内の小学生6名を対象
  - 大阪市内の博物館を見学, レポートを作成

実験用ポータルサイト  
<http://mc3.race.u-tokyo.ac.jp/>

PUBLIC OPINION CHANNEL

\*平成17-18年度科学研究費補助金若手研究(B)  
課題番号: 17700629  
研究代表者 山下耕二

# まとめ

- POCの概念を実装し、実践を通じて効果を検証した。
  - FTTHトライアルでは地域に関する知識が発掘され、POCTVによる知識提供がなされた
  - 大学講義での使用ではストーリーの共有による相互触発が見られた
- コミュニティ支援システムにおける  
コミュニティ運営・分析支援機能を提案・評価した
  - 実験状況把握への効果
  - メッセージ分析作業の効率化
  - コミュニティ運営作業の改善

Part II  
Weblogを用いた  
関心動向分析に関する研究

# はじめに

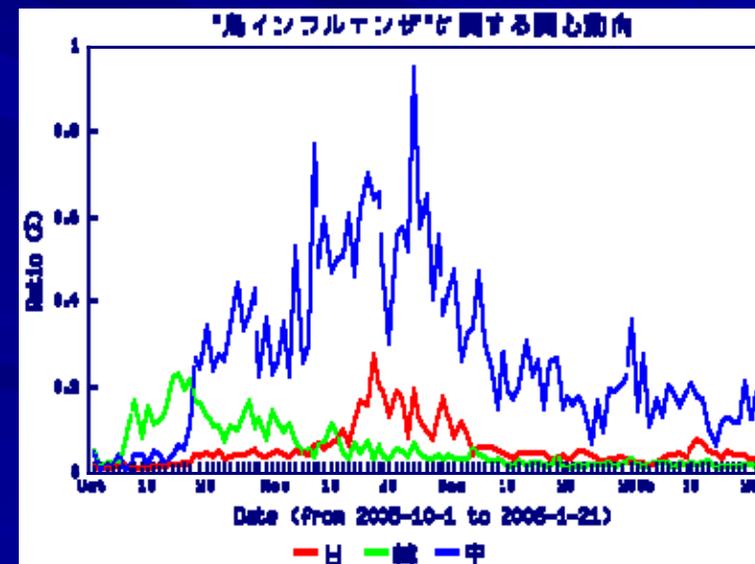
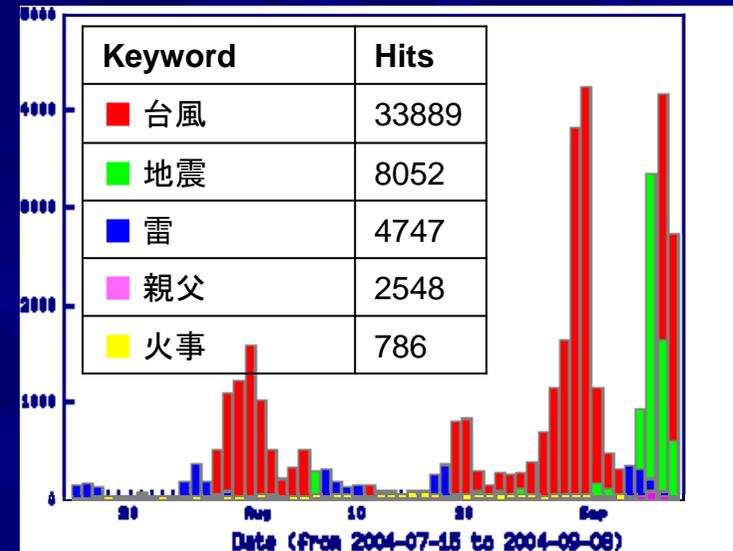
- Weblogや新聞記事等の時系列テキスト集合を用いた関心动向の分析について述べる

— 時系列テキスト集合:

- Weblog, 新聞記事, e-mailなど日付情報を含むテキスト集合

- 分析の視点

1. 社会的関心のパタン
2. 共起語を使った分析
3. 言語横断型関心分析
4. 感情表現を使った分析
5. 実世界データとの関連



# 社会問題と関心

## ■ 社会的関心と社会問題

### － 様々な社会問題と安全・安心

- 地震, 台風, 津波, テロ, 地球温暖化, 個人情報, 狂牛病, SARS, 鳥インフルエンザ...
- 社会運営にも大きな影響を与える.

### － 問題の解決に向けて

- 問題に対する社会の関心を把握することが重要!
  - － 人々がどのような問題に関心を寄せているかを知る.
  - － それによって必要な対策を講じる.

地震  
(阪神淡路大震災 (1995))

SARS

狂牛病

# Weblog (ブログ)

## ■ Weblog (Blog)

- 個人や組織が日々の出来事や考え、ある話題についての意見などを時系列的に記載する情報発信スタイル
- RSS/Atom による新着記事配信
- 多くの人々がブログサービスを利用。日常の出来事や社会の出来事についても記述。

- ブログ記事の収集と分析によって人々の関心動向を迅速に把握できるのではないか

Blogの例

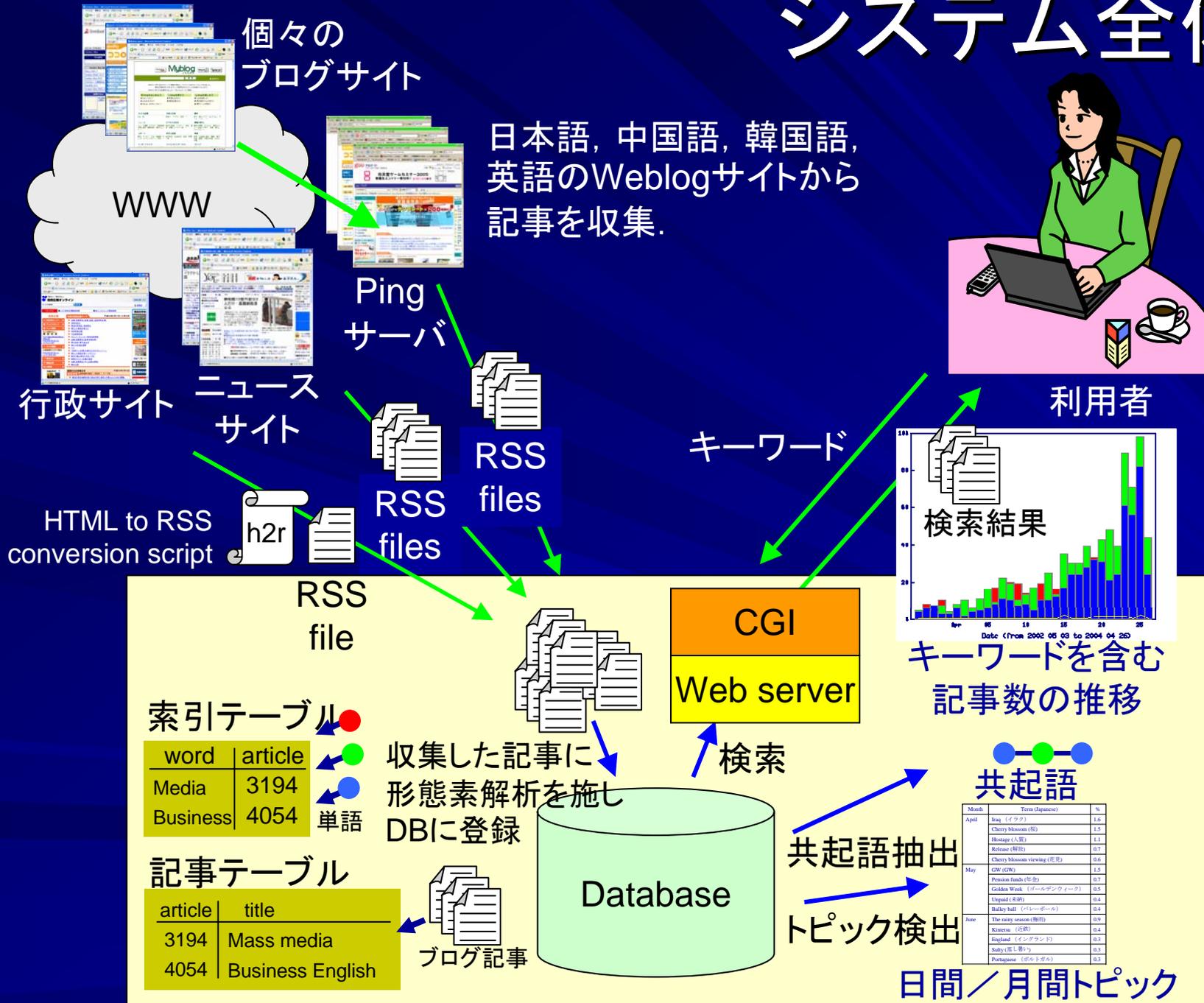
(<http://guonb.cocolog-nifty.com/blog/>)

# Weblog記事を用いた 関心解析システム: KANSHIN

- Weblog記事を収集し  
社会的関心を解析する  
システム
- 特徴
  - ブログ記事の収集と解析
  - 共起語検索
  - 特定期間の話題を抽出
    - 月間トピック, 日間トピック
  - 言語横断型分析
    - 日中韓英ブログ記事を対象



# システム全体像

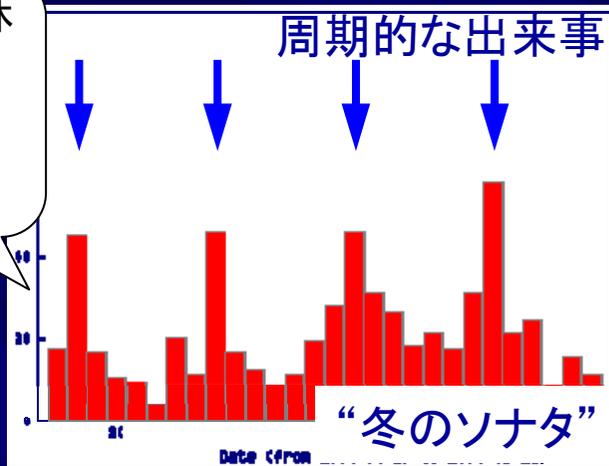


# 分析結果

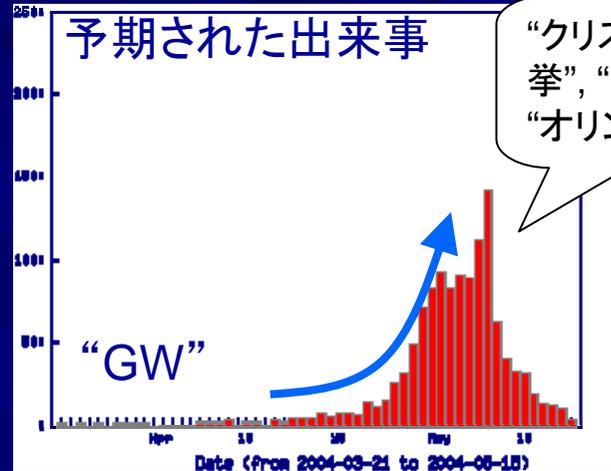
1. 社会的関心の分類
2. 共起語を用いた分析
3. 言語横断型関心分析
4. 感情表現を用いた分析

# Weblogに見られる関心パターン

“給料日”, “休日”, “BBQ”, “家族連れ”, “結婚式”, “同窓会”

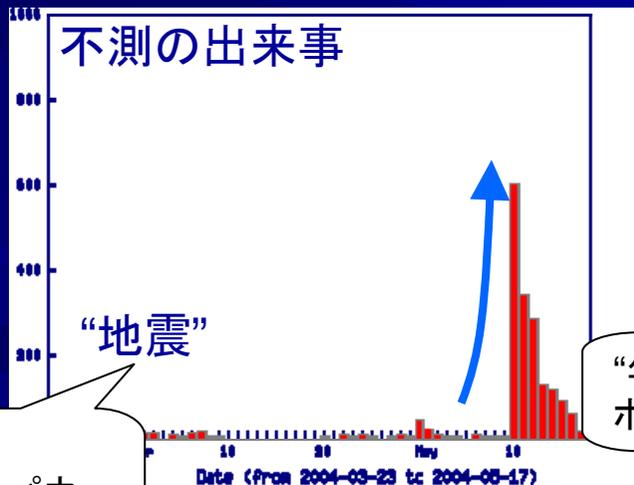


I. 周期型

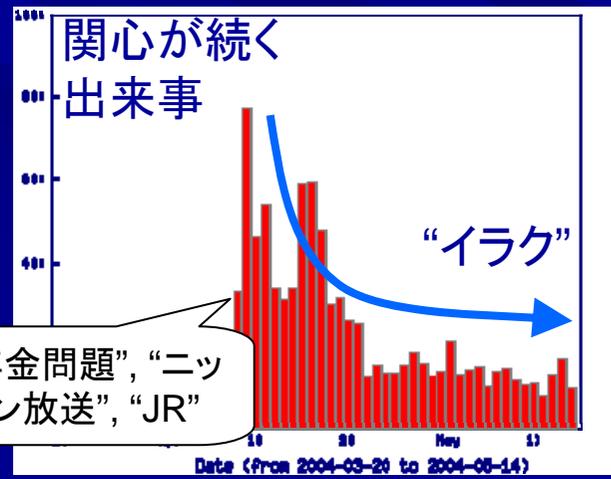


II. 漸次増加型

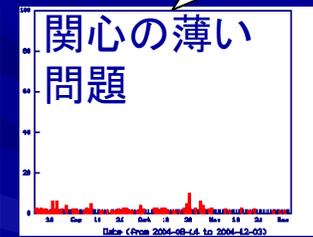
“酸性雨”, “大気汚染”, “水不足”



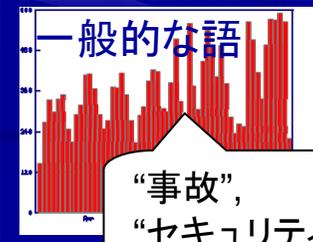
III. 突発型



IV. 関心持続型



V. その他



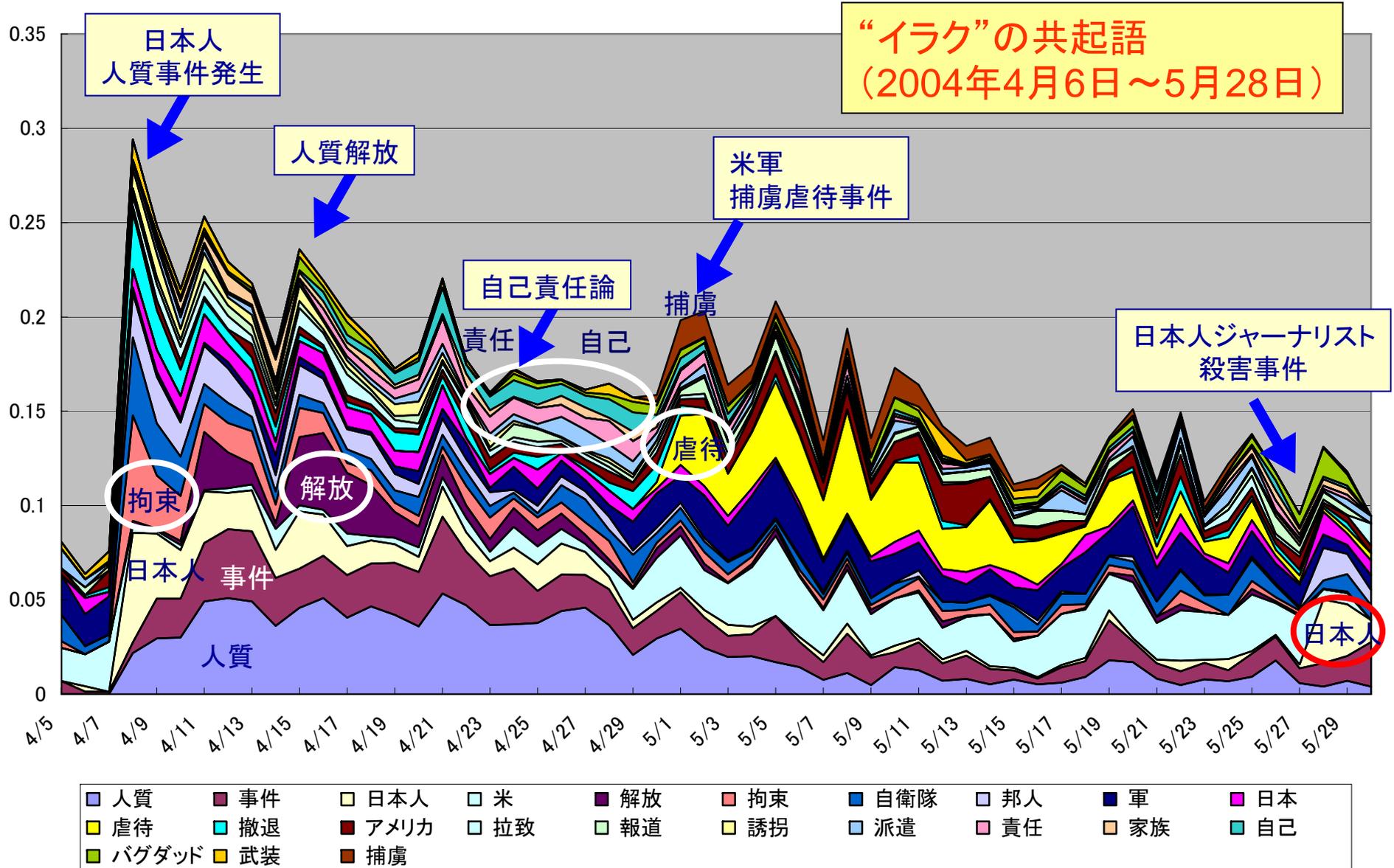
“事故”, “セキュリティ”

# 共起語を用いた分析

どんな言葉と共起していたか？

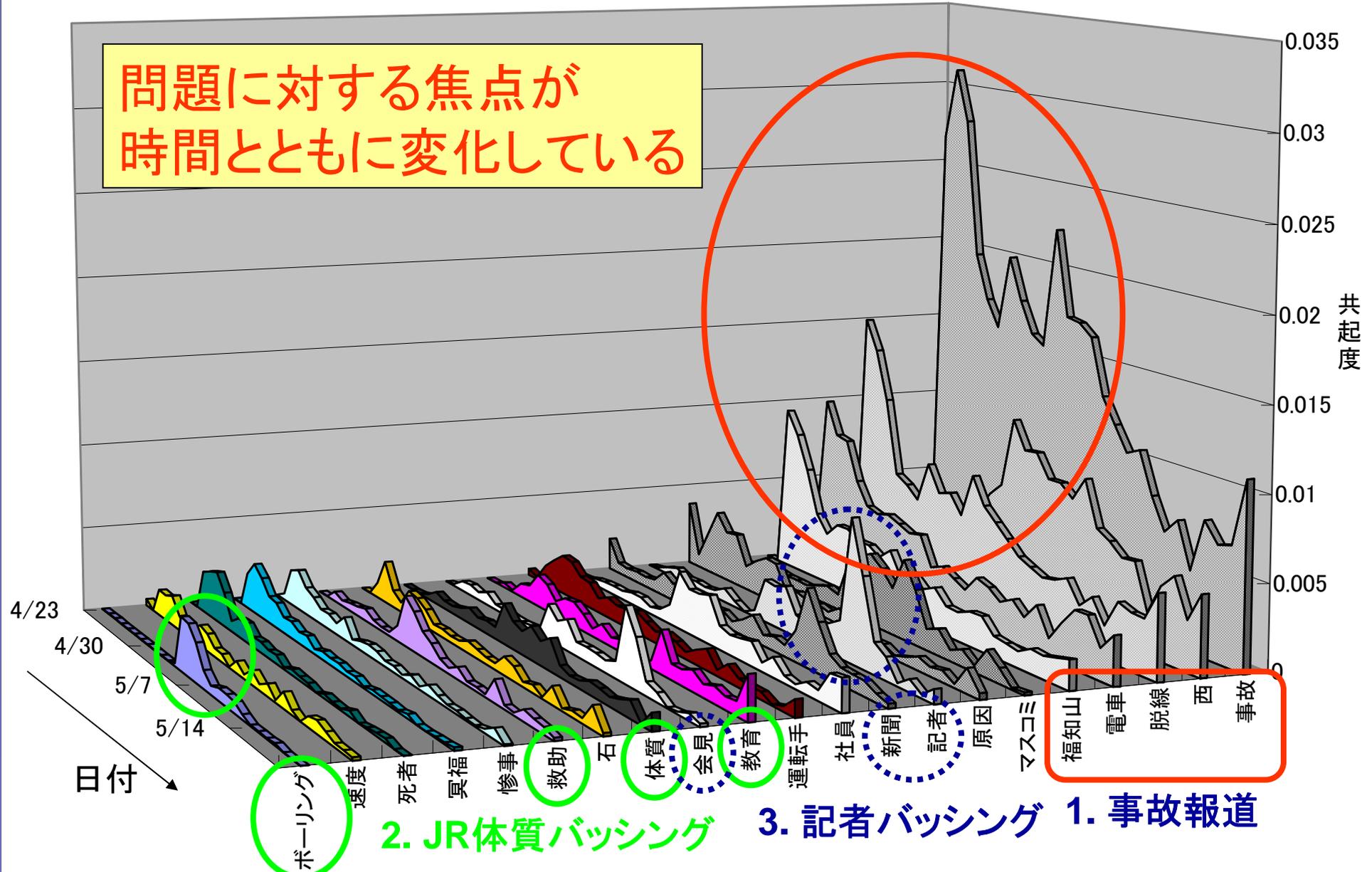
問題の焦点は何か？

# 共起に基づく分析 2004年4月イラク人質事件



# 共起に基づく分析

## 2005年4月JR西日本福知山線脱線事故

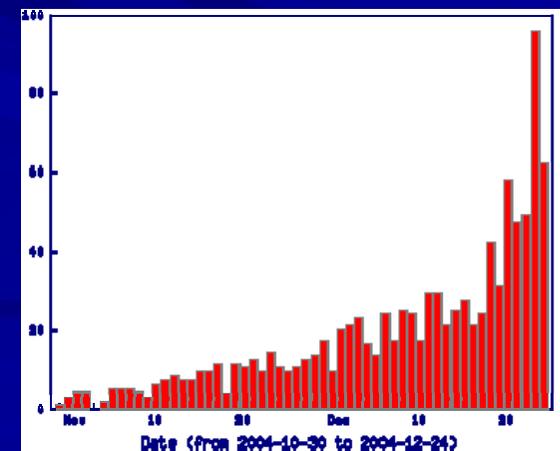


# 言語横断型関心分析

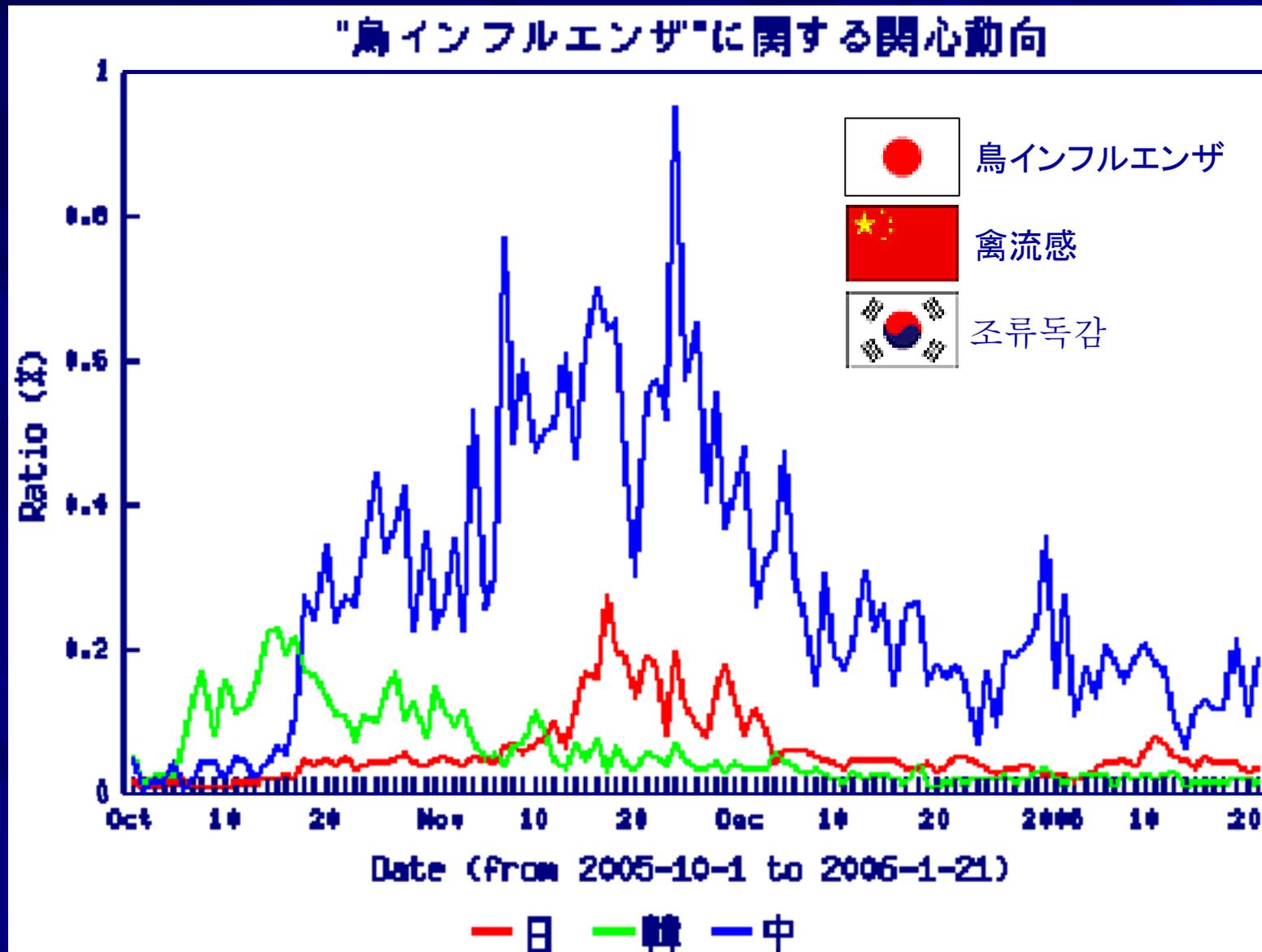
国内と海外の関心を比較する

# 言語横断型関心分析

- 各国の関心を知りたい
  - ある出来事に関する各国の関心を知りたい
- あるキーワードについて他の言語コミュニティの関心と比較する
  - 入力
    - 日本語キーワード
  - 出力
    - それぞれの言語での検索結果



# 鳥インフルエンザに関する関心比較



# 検索語の翻訳: Wikipediaを用いた対訳検索

入力(翻訳元言語)

 FIFAワールドカップ



Wikipedia(日本語)

他の言語

- Alemannisch
- العربية
- Български
- Bosanski
- Català
- Český
- Cymraeg
- Dansk
- Deutsch
- Ελληνικά
- English

他の言語の  
同一エントリーへのリンク

直接リンクが無い場合,  
英語版Wikipediaで確認

ターゲット言語へのリンク

出力(目標言語)



 축구 월드컵



 世界盃足球賽

## ■ 処理概要

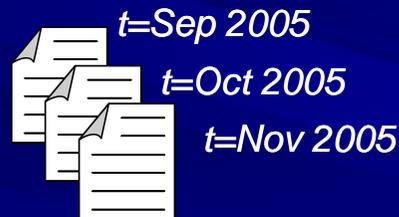
1. キーワードを翻訳元言語のWikipediaで検索.
2. (もしエントリーがあれば)ターゲット言語へのリンクをたどり,  
対訳表現を得る.
  - (ターゲット言語のエントリーが無ければ)英語エントリーから  
ターゲット言語へのリンクの有無を確認.
3. 対訳表現の獲得(あるいは対訳検索の失敗).

# 感情表現を用いた分析

人々は社会の出来事に対して  
どのような感情を抱いているか？

# 感情と話題推移の把握： 感情表現を用いた話題検出手法

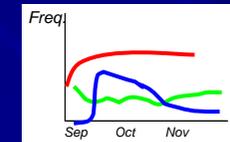
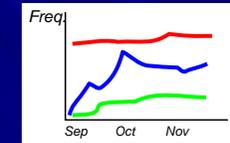
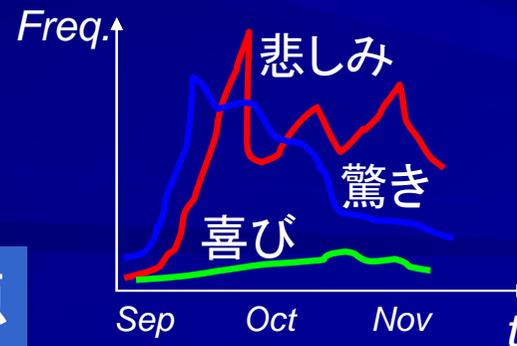
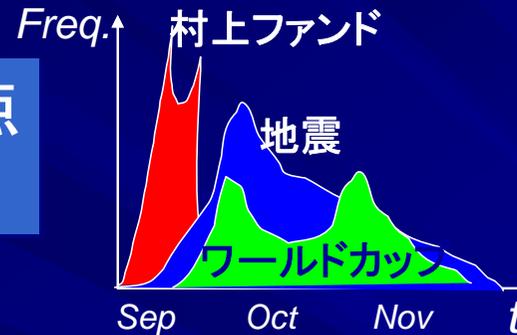
## 1. 話題推移の観点 (トピックグラフ)



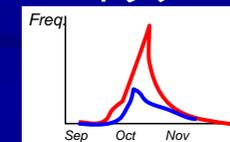
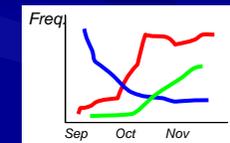
時系列

テキスト集合：  
(タイムスタンプ付き  
テキスト集合)

## 2. 感情推移の観点 (感情グラフ)



感情を指定

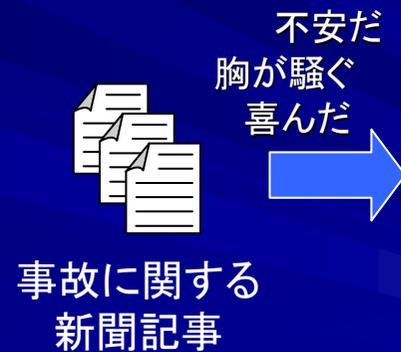


話題を指定

# 感情表現の例

- 不安
  - 不安, 心配, 怯えた, 気掛かり, 胸が騒ぐ, 安心できない, 青ざめた
- 悲しみ
  - 悲しい, 涙が, 目頭を押さえた, すすり泣く, 嗚咽を
- 怒り
  - 怒り, 苛立ち, 語気を強めた, 許せない, 憤った, 声を荒げた, 厳しく非難
- 安心・喜び
  - 安心, 喜んだ, ホットした, 躍り上がった, 笑顔で, 胸をなでおろした
- 苦悩
  - 苦悩, 苦渋, 頭を抱えて, やりきれない, 憂うつ,
- 疲労
  - 疲労, 疲れ, ぐったり, うんざり, 途方に暮れた, 疲れ果て, 疲労困ぱい, 残念,

- 不満
  - 不平, 不満, 納得いかない, ぶ然と, 難色を示した, 不公平, 不服,
- 衝撃
  - ショック, 驚き, 衝撃, 啞然と, 慌てて, 興奮して, 動転して



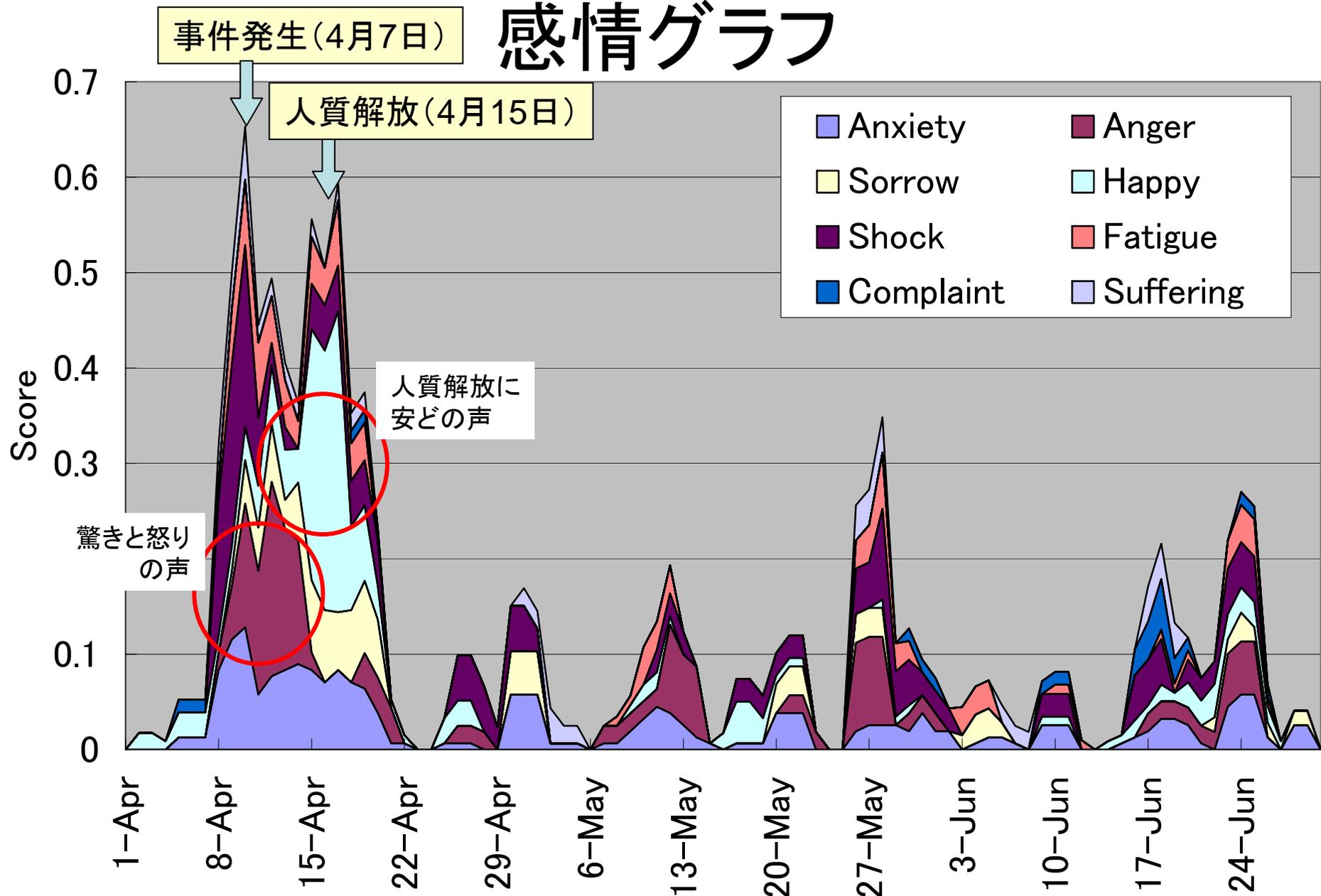
不安	苦悩
悲しみ	疲労
怒り	不満
安心・喜び	衝撃

感情表現  
(8カテゴリ, 383表現)

※2003年に発行された全国紙4紙, 地方紙3紙のうち化学工場事故に関する記事から383個の感情表現を手動で抽出

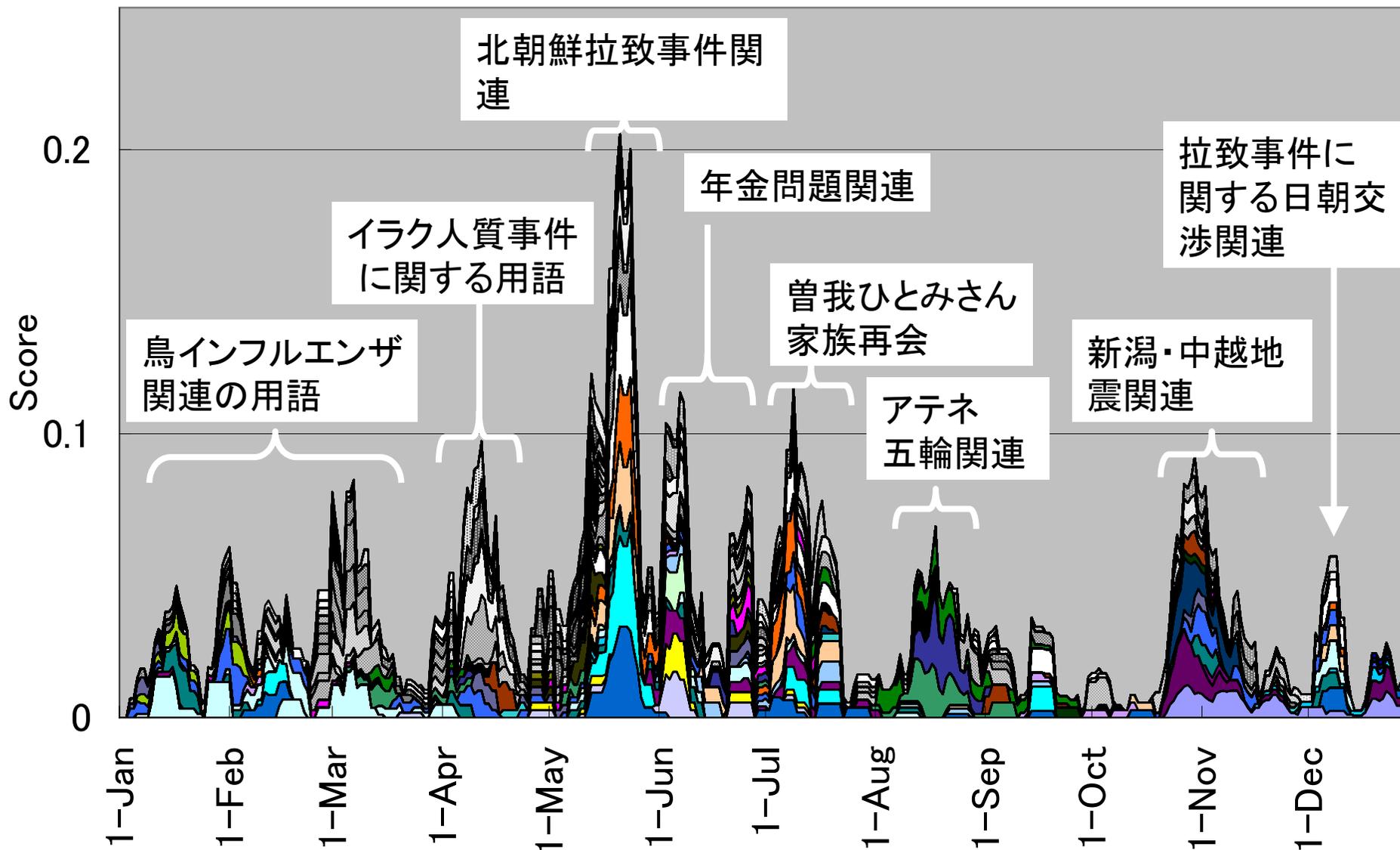
# 2004年4月イラク人質事件に関する

## 感情グラフ





# “不安”に対するトピックグラフ (朝日新聞2004年版から作成)



# 用語のクラスタリングによる話題検出

## ■ 問題

- いつ、どういう出来事があったか思い出せない。
- 例えば、2004年にはどんな出来事があったか？

## ■ 目的

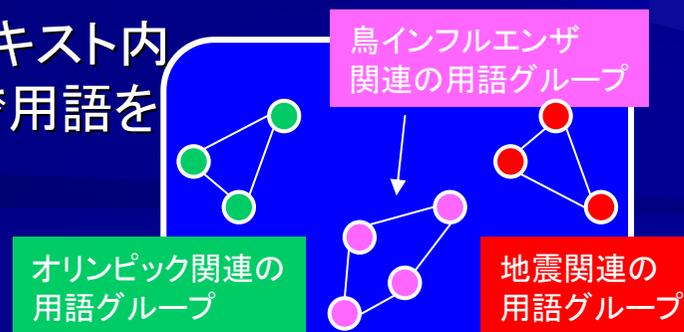
- 膨大なテキスト集合から重要な話題を確認する。

## ■ アプローチ

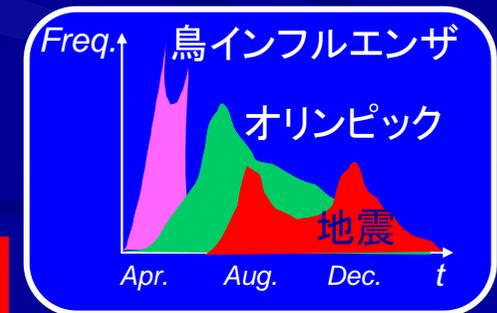
- 時間軸上の相関とテキスト内での共起を見ることで用語をクラスタリングする。



1) 時間軸上の相関と  
2) 記事中の共起でクラスタリング



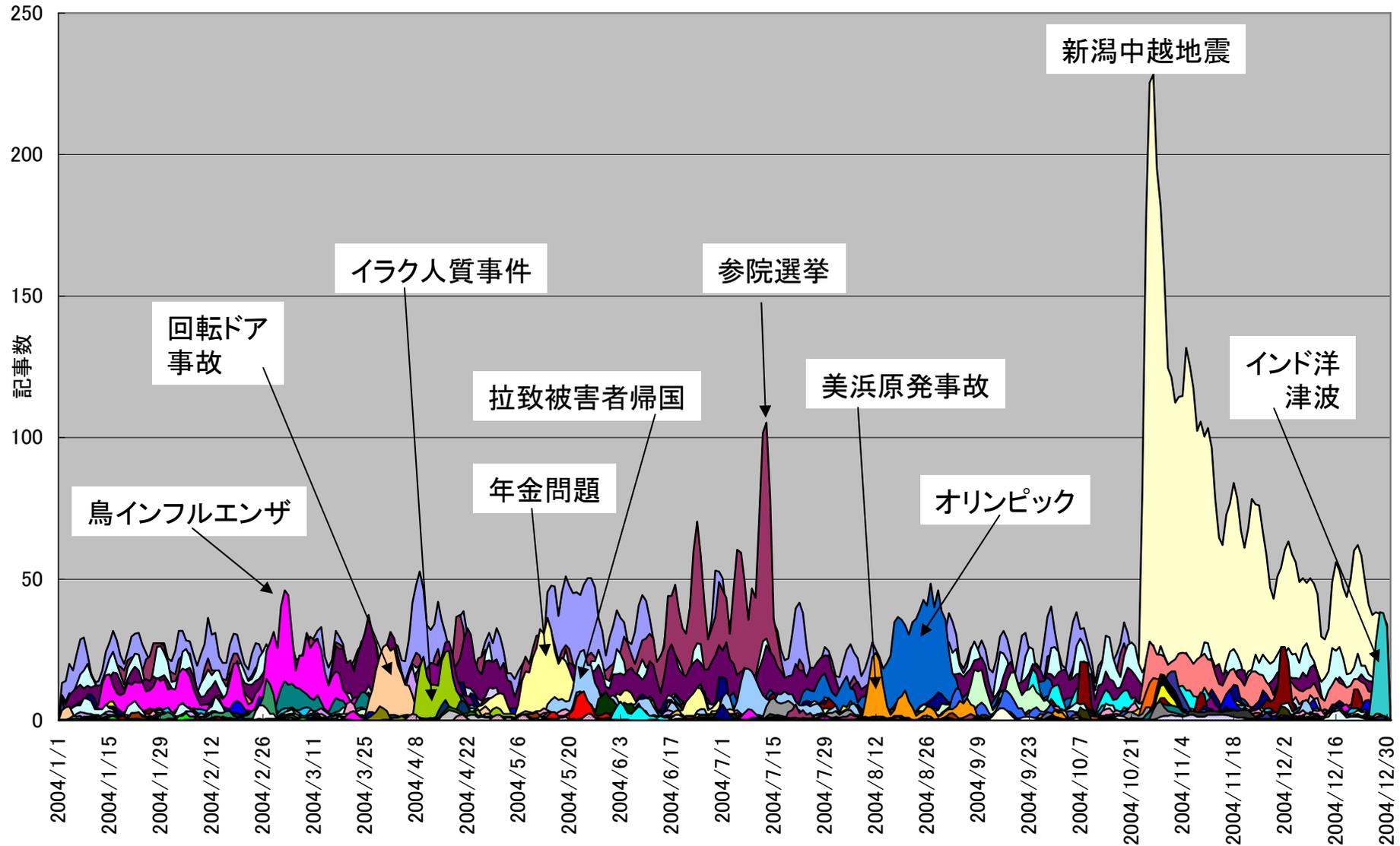
グラフ1: 話題グラフ



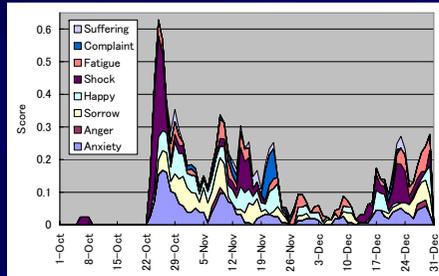
グラフ2: 時間軸  
話題分布グラフ



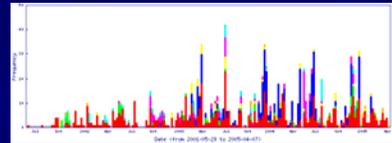
# 時間軸上での話題の分布 (2004年朝日新聞より作成)



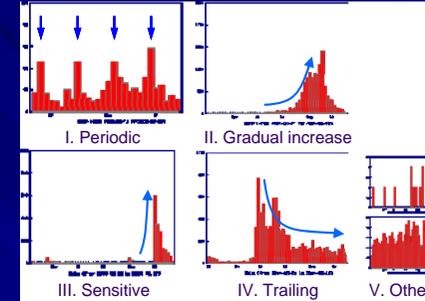
# 関心に対する視点



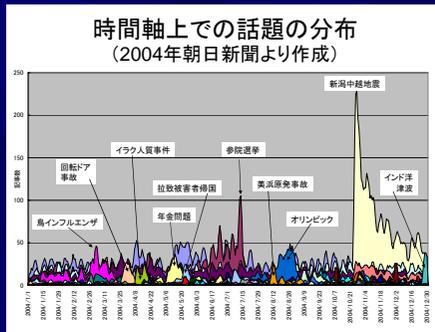
感情に注目した分析



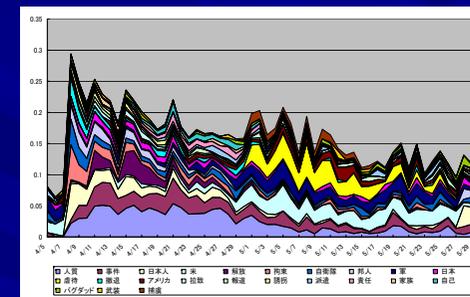
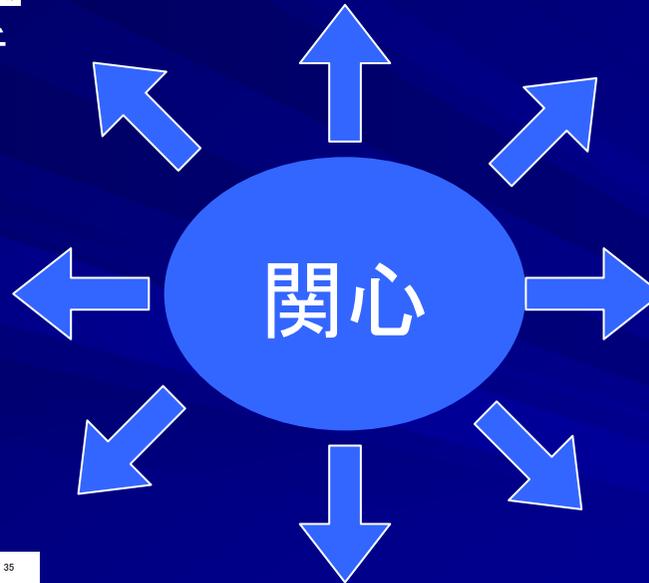
個人(キーパーソン)の関心



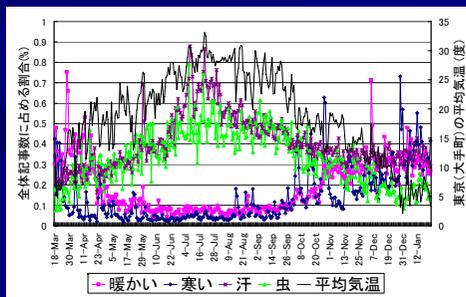
社会の関心



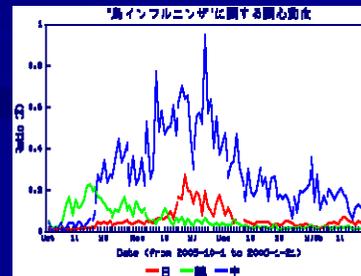
話題の自動検出



共起語を用いた  
話題の焦点に関する分析



実世界データとの関連



海外の関心との比較

## TODO

- 新聞, テレビ, 雑誌との比較
- 個人属性との関係
- (年代, 性別, 住所...)
- コミュニティ単位での分析

# まとめ

---

- コミュニティ支援に関する実践研究と、Weblogを用いた関心動向分析に関する研究について述べた。
    - POCを用いた学際的実践研究
    - Weblogや新聞記事等の時系列テキストを用いた関心動向分析
  - 今後、知的コミュニティ基盤研究センターとの連携・交流等を通じて接点を探ることができれば幸いです。
-