

「電気通信サービスにおける情報信憑性検証技術 に関する研究開発」

平成 2 1 年 1 2 月 1 日

総 務 省
総 合 通 信 基 盤 局
電 気 通 信 技 術 シ ス テ ム 課

「情報の巨大集積化と利活用基盤技術開発」連携施策群における位置づけ

目標

Web上及び非Web上にある大量かつ多様な情報を、個人が簡便、的確、かつ安心して収集、分析できる

次世代の情報検索・情報解析技術基盤を構築する。

2012年までに、情報爆発時代に対応した超巨大データの戦略的活用を可能とする。

非Web空間(実空間)

Web(サイバー空間)

[2009年度からの連携枠組み]

必要性

- Web情報の価値の拡大
 - Web情報の学術研究、経済活動への高度活用に対する期待
- ↓
- Web情報分析のための技術の発展が必要

情報の活用

ビジネス、生活、教育・研究、医療等

Web社会分析プロジェクト

- Web情報(テキスト、動画、画像、音声等)を効率よく収集してアーカイブを形成
- 蓄積したWeb情報の高度な分析

過去の情報を含めた高度な分析

連携を拡充

情報の信憑性を検証

[2008年度までの連携枠組み]

情報大航海プロジェクト

- 巨大集積化した情報を再利用できるように仕分け
- 整理整頓して器に格納

分析手段

情報の分析

情報大航海の一部成果も活用し、研究開発を効果的に推進

新たな分析手段

センサ情報の社会利用コンテンツ化プロジェクト

- プライバシ情報を管理
- センサ情報のコンテンツ化

超高性能データベースプロジェクト

処理の効率を格段に向上

情報信憑性検証技術プロジェクト

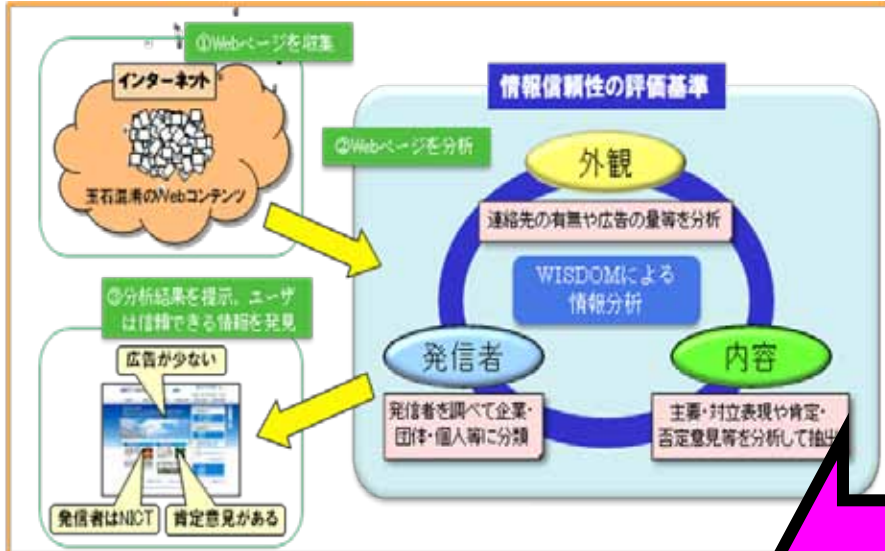
- 格納された情報の純度を上げるべく精錬
- 純度の保証付きでユーザーに提供

電気通信サービスにおける 情報信憑性検証技術に関する研究開発

電気通信サービスにおける情報信憑性検証技術に関する研究開発

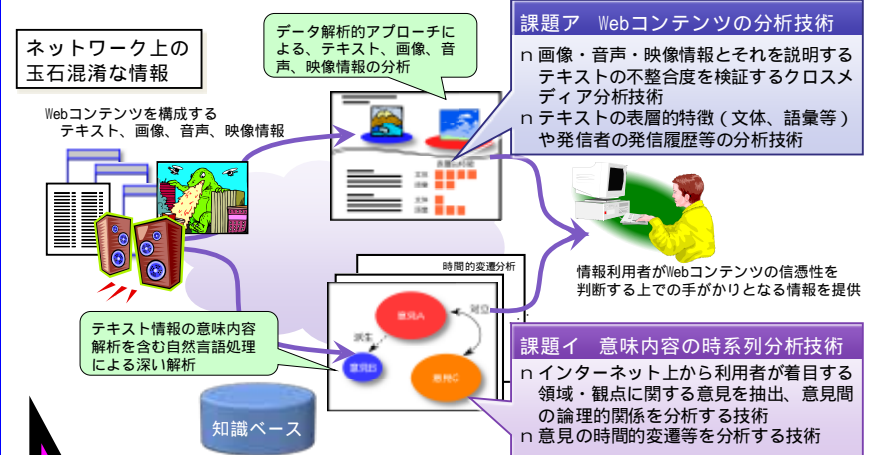
ネットワーク上の文字、音声、映像情報について、偽りの情報、信頼性の低い情報等进行分析する技術を確立し、信頼できる情報を提供することで、誰でも思いのまま、簡単に、信頼して、コンテンツを扱い、高度に利活用できる環境を実現

NICT研究



大量のWebコンテンツから自然言語処理技術に基づく4種類の情報分析手法で、ユーザが信頼性を判断する基準となる情報を提供

委託研究 研究期間:平成19年度～平成22年度(4年間)



相互補完

データ解析のアプローチによるWebコンテンツを構成するテキスト、画像、音声、映像情報の分析技術
自然言語処理による深い解析による情報相互の論理的関係性、時間的変遷の分析技術

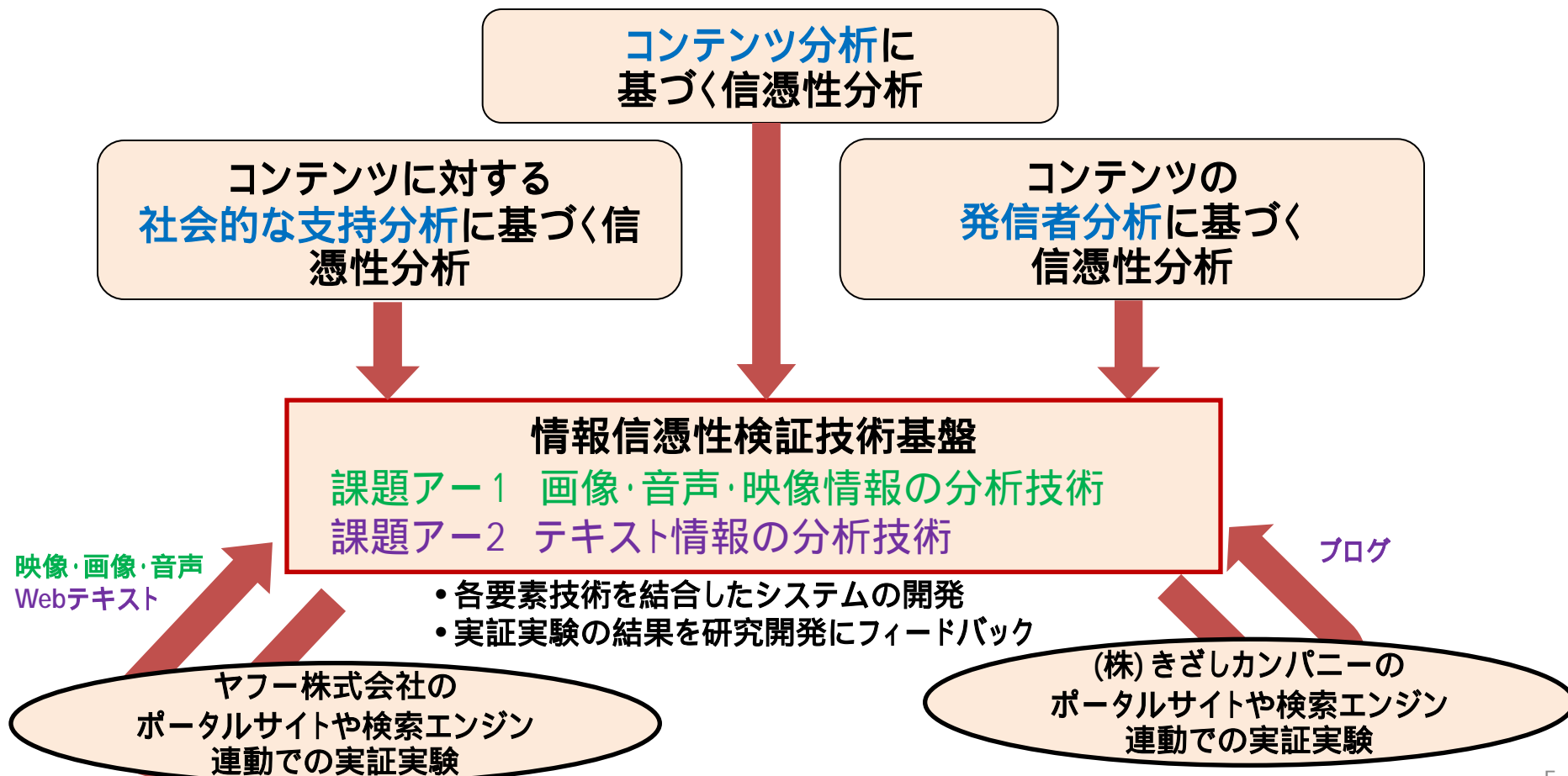
誰もが安心・安全にWebコンテンツを利活用できるための情報信憑性検証技術を実現

課題ア Webコンテンツ分析技術

研究代表者 京都大学大学院
兵庫県立大学
京都産業大学
(株)きざしカンパニー
(研究協力: ヤフー株式会社)

課題ア 研究開発の目的、方法

Web検索エンジン等によって得られる画像・音声・映像やテキストといったWebコンテンツの信憑性・信頼性を、情報利用者が適切に判断するための情報を、画像・音声・映像などの周辺コンテンツやテキストの表層的特徴分析をもとに、「現実的な処理時間」で収集・分析・提示できる機能の研究開発を実施。



課題ア 研究課題全体中間目標

平成22年3月末時点

研究目的

- ・Web検索エンジン等によって得られる画像・音声・映像やテキストといったWebコンテンツの信憑性・信頼性を情報利用者が判断するために資する情報を、画像・音声・映像などの周辺コンテンツやテキストの表層的特徴分析をもとに「現実的な処理時間」で収集・分析・提示できる機能の研究開発
- ・開発技術のISPなどの実環境における実証実験

評価実験の数値目標

【精度】

- ・10以上のトピックについて人間による評価と比較し、**精度75%以上**

【処理時間】

課題ア - 1 (画像・音声・映像情報の分析技術)

- ・**現実的な処理時間**

課題ア - 2 (テキスト情報の分析技術)

- ・**数秒 ~ 20秒**程度以内

本課題の有効性に関する情報発信

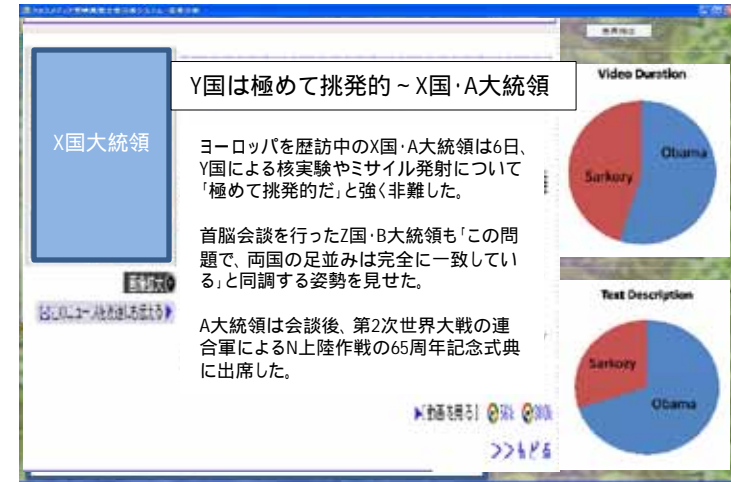
- ・国内外の学術会議における積極的な発表
- ・特許の取得

クロスメディア型映像整合性分析 [課題ア - 1 - 1]

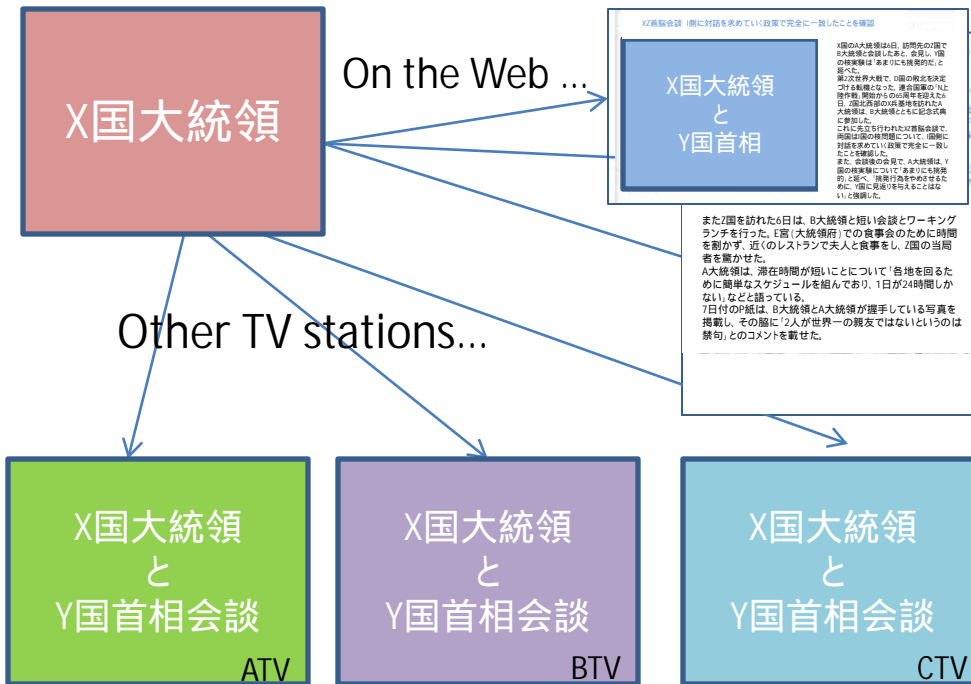
「映像による実世界の部分クリッピング」の発見

映像コンテンツの字幕と周辺テキストを用いて、関連情報と比較分析し、映像コンテンツの説明*の違い(不整合)を発見する。実世界の特定箇所・側面しか説明できていない場合はそれを明らかにし、その他の箇所・側面およびその差異を提示してユーザによる映像の信憑性判断を支援。

* : 本研究では、映像コンテンツの説明 = 周辺テキストとする。



不整合分析



関連情報収集

N局はX国大統領とY国首相が会談で完全一致したことを報道しているが、ほかの局は二人の関係の悪さに着目している。

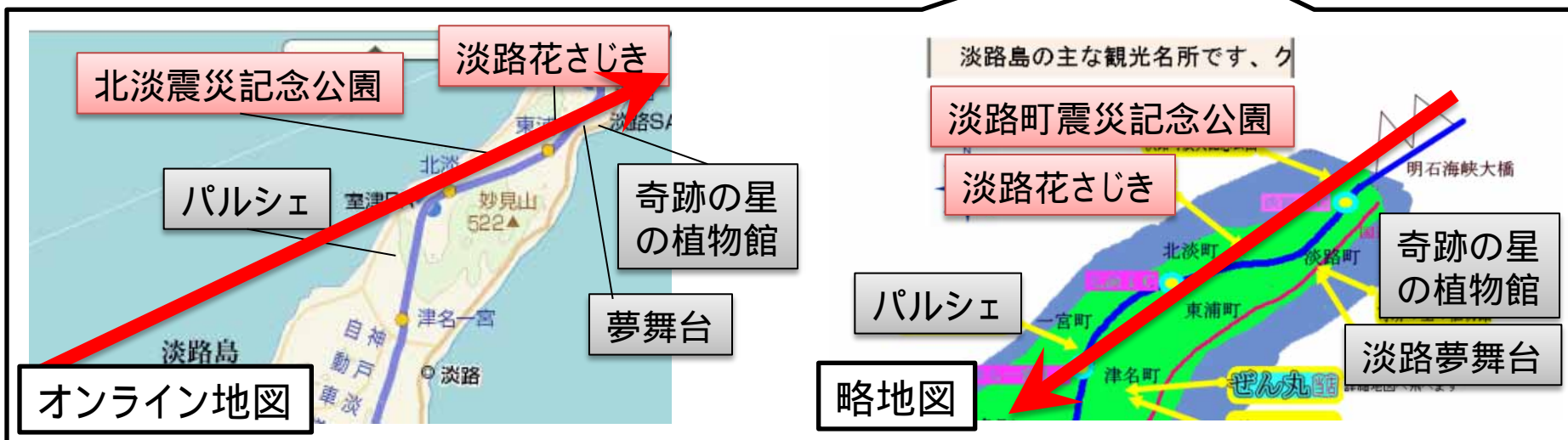


実空間への要求意図に基づく地理的整合性分析エンジンの開発

[課題ア - 1 - 2]

Web地図画像の整合性分析

- 方角, 経路, 距離の整合性分析技術の開発
- 略地図データとオンライン地図データの対応付けにより, オブジェクトの位置関係に関する信頼性を提示



サーチエンジンの索引データを用いたテキストの表層的特徴分析技術

[課題ア - 2 - 1]

レビューや広告テキストの信憑性判定支援

- 「ブランド」としての信憑性について判定
- 「ブランド」であれば言及されるべき意味のある「属性」を取得
 1. 確実に「ブランド」である語を取得(商標DBなどを利用)
 2. Web検索エンジンを利用して「ブランド-属性」の二部グラフを作成
 3. 得られた「属性」について言及されていれば「ブランド」である可能性が高い



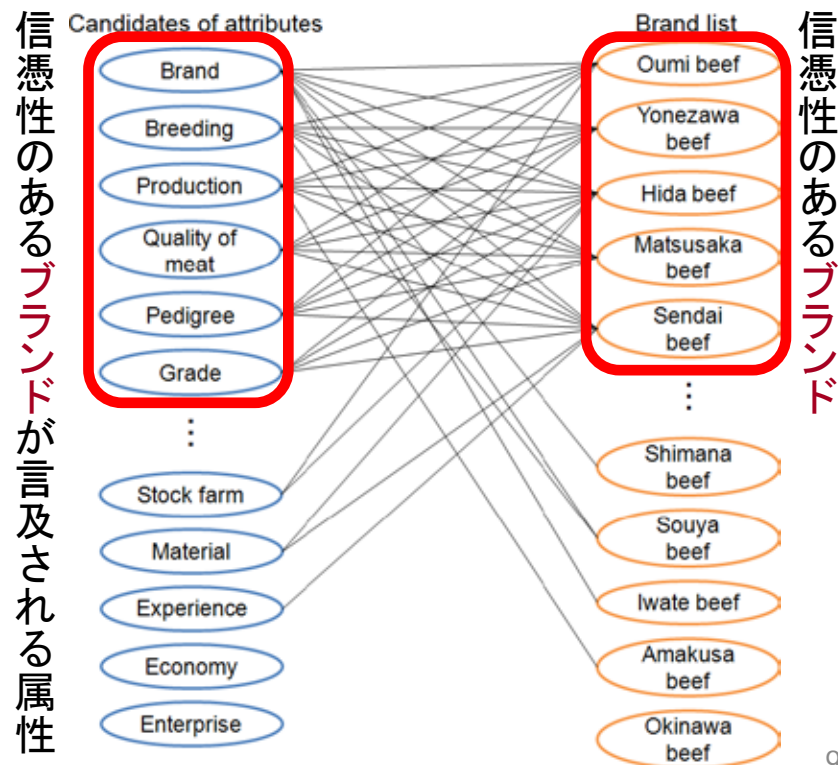
Kobe beef grilled over a Bincho charcoal flame. Come taste the finest A5 and A4 quality Kobe beef and Tajima beef steak in our restaurant located in Sannomiya, Kobe.

Sushi Wushi is a unique traditional Japanese restaurant in leafy Eltham, North Sydney. Enjoy Kobe beef in Kobe's Sannomiya. The highest grades of Kobe beef available, A5 and A4, are grilled broiled over a Bincho charcoal flame. This is the best way to experience Kobe beef. While we naturally pay close attention to our meats, we also offer We pride ourselves on homestyle sauces, dressings, and desserts. Our Kobe pasta is prepared with sauce made fresh ingredients and prepared from Kobe beef fat as well. Finish up with some of our delicious ice cream, made without customers with great additives.

Our menu has a selection of excellent traditional fresh Japanese Sushi and Sashimi or you can indulge in one of our favorites from the grill like our tasty Osaka Beef!... [More details](#)



[More photos](#) [Visit Website](#)

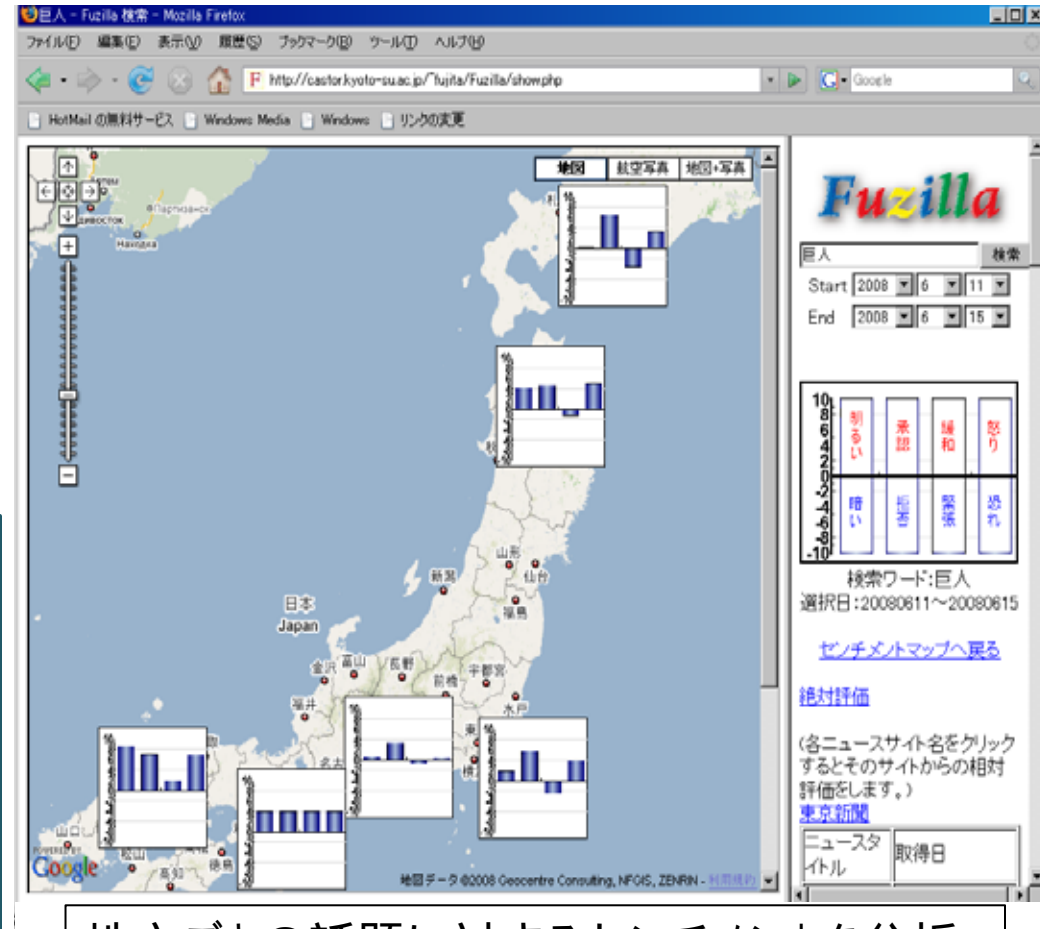


Webテキスト信憑性分析: 発信者のセンチメント(感情)分析 [課題ア - 2 - 2]

「発信履歴情報からの発信者熟知度・センチメント分析に基づく発信者分析エンジンの開発」

U 発信者がどのような感情(センチメント)で情報を発信しているのか? 発信している分野にどの程度熟知しているか? それら进行分析することで、話題に対する新聞社サイトや各国、さらにブロガーなど発信者の信憑性を計る技術の開発

1. 感情表現語辞書の構築・改良
2. 各記事のセンチメント(感情)を抽出
3. 話題ごと、発信者ごと、組織ごとのセンチメントの検出
4. 話題・発信者(個人や組織・団体)・記事などのセンチメントの相違を分析
5. 相違の提示による信憑性判断支援を実現



地方ごとの話題に対するセンチメントを分析
例) 話題: 「団体A」

「体験熟知度とセンチメントに基づく情報発信者分析および信憑性評価手法の開発」

U 体験熟知度判定技術の精度向上に取り組みると共に、センチメント分析技術の改良を行うことで、情報発信者分類機能の精度向上を実現する。また、情報発信者間の排反・分散関係の分析とこれを利用した信憑性評価機能の開発を行う。

1. 熟知領域リストおよび領域毎の共起語辞書の作成と改良
2. 情報発信者が過去に投稿したエントリの分析に基づく熟知度スコアの算出
3. 熟知度スコアに基づくエントリのランキングの算出
4. 各エントリのセンチメント分析と、これを用いた信憑性評価機能の開発



公開中の実証実験システム

課題ア 研究課題全体最終目標

研究目的

- ・Web検索エンジン等によって得られる画像・音声・映像やテキストといったWebコンテンツの信憑性・信頼性を情報利用者が判断するために資する情報を、画像・音声・映像などの周辺コンテンツやテキストの表層的特徴分析をもとに「現実的な処理時間」で収集・分析・提示できる機能の研究開発
- ・開発技術のISPなどの実環境における実証実験

評価実験の数値目標

【精度】

- ・10以上のトピックについて人間による評価と比較し、精度80%以上

【処理時間】

課題ア - 1 (画像・音声・映像情報の分析技術)

- ・現実的な処理時間

課題ア - 2 (テキスト情報の分析技術)

- ・数秒～10秒程度以内

本課題の有効性に関する情報発信

- ・国内外の学術会議における積極的な発表
- ・特許の取得

課題Ⅰ 意味内容の時系列分析技術


研究代表者 日本電気株式会社
国立大学法人 奈良先端科学技術大学院大学
国立大学法人 横浜国立大学

情報信頼性分析に対する課題イのアプローチ

[目的]
広い視野から偏りのない
判断を支援

分析結果を統合的に出力



話題のダイエット特集 

朝バナナを食べるだけで楽やせ！
超簡単！ その驚きの効果...
.....

着目言論

「朝バナナで痩せられる！」
その情報、信じてよいか？

[手段]
気になる情報の

言論空間での位置づけを
分析し、俯瞰的要約を
ユーザにレポート

言論空間を再構成する
偏りのない情報を分かりやすく

どこに注目すべきか？

論理空間に位置づける
他のページはどう言っているか？

反対意見は？
根拠は何か？

時間軸に位置づける
どんな経緯で語られたのか？

何が切っ掛けか？
今でも有効か？

課題Ⅰ 研究課題全体中間目標

平成22年3月末時点

イ-1-1 着目言論を論理空間に位置づける (奈良先端大)

- ・ **言論マップ生成**: 評価用10トピックにおいて言論間論理的関係の認識精度80%
- ・ **大規模知識ベース**: 事象(コト)間関係知識約50万関係を整備。実体(モノ)間知識約200万関係を精度90%で獲得

イ-1-2 着目言論周辺の言論空間を再構成する (横浜国大)

- ・ **重要言論・重要発信者抽出**: 評価用10トピックにおいて各抽出タスクを3分程度、R-精度80%で実現
- ・ **要約文書生成**: 評価用10トピックにおいて人手要約との比較で精度80%、3分程度で分析結果を提示開始

イ-2 着目言論を時間軸に位置づける (NEC)

- ・ **変化要因抽出**: 評価用10トピックにおいて適切度80%の抽出結果を3分程度で出力
- ・ **有効期限切れ判定**: 評価用10トピックにおいて精度80%の言論間関係、精度70%の有効期限ラベルを3分程度で出力

統合システム、実証実験 (NEC)

- ・ 統合システムを用いて実証実験を行い、信頼性判断支援としての有効性を確認

課題イ - 1 - 1:最終出カイメージ (言論マップ)

着目言明: バナナダイエットでダイエットできる?

着目 [朝、バナナを食べるだけでダイエットできる]

- ・朝、バナナを食べるだけで、気軽に痩せられる...
- ・朝食にバナナを食べるだけで痩せられる「朝バナナダイエット」...

根拠 [バナナは食物繊維で便秘解消してダイエットできる]

- ・食物繊維も豊富だからお通じもよくなりダイエットにも...
- ・繊維も多く「便秘」にも効果があり...

根拠 [バナナの持つ酵素の働きはダイエット効果がある]

- ・バナナに含まれる酵素を使ってダイエットできる
- ・バナナの持つ酵素の働きにより、ダイエット効果がある
- ・バナナは糖質分解酵素が脂肪燃焼を活発にする...

根拠 [バナナは糖質分解酵素が豊富である]

- ・バナナは消化を促進する酵素や食物繊維が含まれている
- ・バナナには脂肪を分解するための酵素が多く含まれている...

根拠 [私はバナナダイエットで1カ月に3キロ痩せました]

- ・実際、3ヶ月で5kgくらいはバナナダイエットで痩せました。
- ・私はバナナダイエットをして1カ月に3キロ痩せました。...

対立 [バナナを食べれば痩せるなんてことはない]

- ・バナナ食べて痩せるなんて、有り得ない訳です。
- ・朝バナナと水を飲んだだけで、痩せるわけがありません！...

根拠 [バナナの酵素が代謝を高めることはない]

- ・果物の中に含まれている...酵素が、体の中で活性を持つ、なんて話はない。
- ・バナナでは、含まれる酵素が代謝を高めるなんてことは絶対にありえない。...

根拠 [1ヶ月のバナナダイエットで効果なしでした]

- ・まず1ヶ月だけ試しましたが、やせませんでした。
- ・私は1ヶ月バナナダイエットしましたが効果なしでした。...

対立 [バナナダイエットは発がん性向上の恐れ]

- ・専門家が発癌性向上の恐れをバナナダイエットに警鐘...

限定 [快眠を得なければ効果は期待できない]

- ・快眠を得なければ、朝バナナの効果はあまり期待できません。...

限定 [バナナを食べて動かずにいたら、太る]

- ・バナナ食べて、動かずだと、同じカロリーのほかの食材よりも太る。...

関連キーワード

着目言論のサブピックや関連トピック、用語説明 ex. 酵素 アミラーゼ カリウム 食物繊維 森久美子 ...

背景

関連イベント(時系列分析結果) ex. 最初のイベント=事の起こり

各言論が主張する根拠等

「朝、バナナを食べるだけでダイエットできる」
・バナナの持つ酵素の働きはダイエット効果がある
・バナナ食物繊維で便秘解消してダイエットできる

「バナナを食べれば痩せるなんてことはない」
・バナナの酵素が代謝を高めることはない

「バナナダイエットは発がん性向上の恐れ」

対立言論の整理と疑似対立の判断

対立点1 「バナナの酵素は代謝を高める」

バナナの持つ酵素の働きはダイエット効果がある

バナナの酵素が代謝を高めることはない

対立点2 「バナナのカロリーは低い」

バナナは低カロリーで満腹感があります

両立できる可能性 高い

バナナは果物の中では水分が少ないためカロリーは高めです

対立言論の真偽判断または調停判断

対立点1 「バナナの酵素は代謝を高める」という点で対立しています。両者の根拠をよく読んで、どちらが正しいか判断してください。

バナナには、でんぷん(炭水化物)分解酵素である「アミラーゼ」が豊富に含まれているからです。バナナを食べると、普通なら食べ物を消化・吸収する為に使われる私たち自身の消化酵素を使わずに、**バナナ自体に含まれている酵素を使って、消化・吸収することが出来ます。**消化酵素を使わずに済めば、その分の潜在酵素は、代謝酵素として使うことが出来ますよね。代謝酵素がUPすれば、新陳代謝もUPし、ダイエットに繋がる、というわけです。
アイザック・エアラインスクール講師

酵素はタンパク質なのだから、食べても消化されて終わりだ。果物の中に含まれているごく微量の酵素が、体の中で活性を持つ、なんて話はない。家畜の飼料に酵素を添加して消化を助け増体を図るやり方があるけれど、かなりの量、時には飼料の1%近くになるほど大量の酵素を与えて消化を助けるような方法らしい。バナナでは、**含まれる酵素が代謝を高めるなんてことは絶対にありえない。**

松永和紀(フリーの科学ライター)

対立点2 「バナナのカロリーは低い」という点で一見対立していますが、以下の説明で対立が解消できるか調べましょう。

それと、フルーツが**お菓子やご飯より低カロリー**なのは当たり前。カロリーが低いかどうかは何と比べて低いのか？ってことがない意味がない話。実際はバナナは**フルーツの中でカロリーは高い**ほうです。オレンジ、みかん、グレープフルーツ、さくらんぼ、キウイ、ナシ、メロンなどのほうがカロリーは低いです。なぜバナナで朝食を代用できるのか、それは裏を返せば「十分なカロリーを摂取できるから」です。カロリーが低かったら朝バナナダイエットは成立しません。ハニースポットが出ているバナナはさらにカロリーが高いです。
教えて!goo

「高い、低い」は相対的な表現です。「お菓子やご飯より」と「フルーツの中で」で比較対象が異なっていませんか？

論点の網羅 各論点の主張

注目すべき対立点 根拠の提示

調停要約

俯瞰的要約

ユーザが読み解くための要約

偏りのない情報を

分かりやすく

課題イ - 2 : 変化要因抽出の結果イメージ

バナナダイエットでダイエットできる？

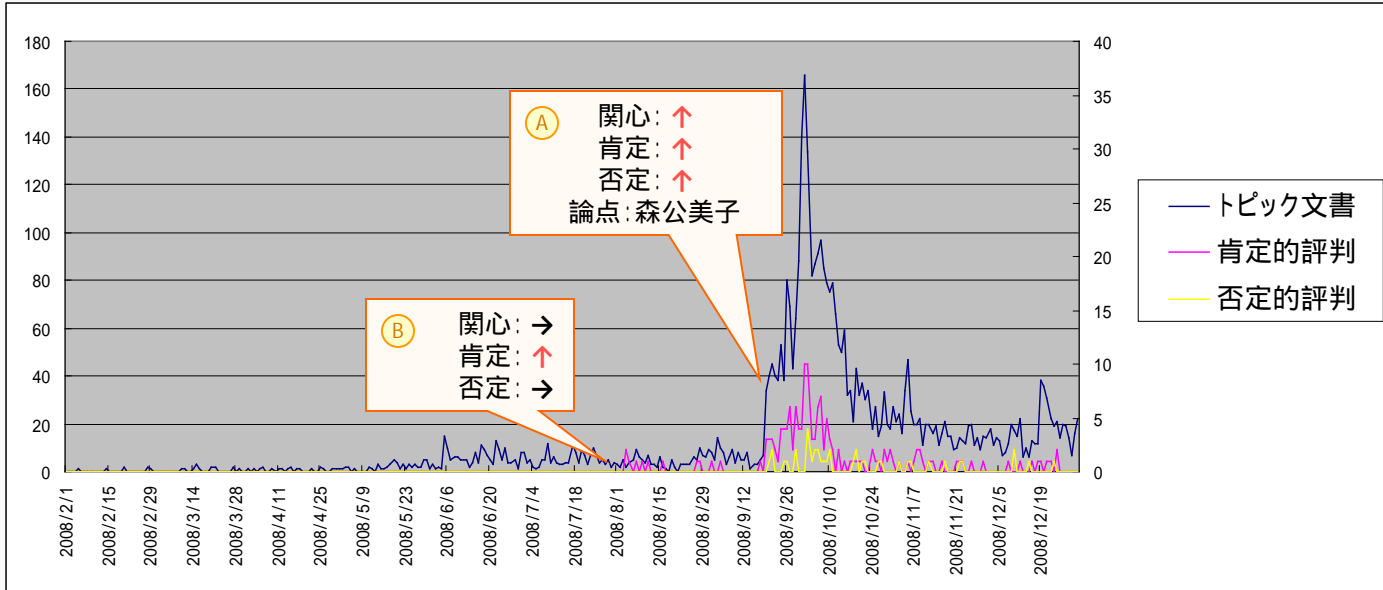
言論マップ

時系列分析

検索

要約

「バナナダイエットでダイエットできる？」の分析結果



- A 2008/9/21
朝バナナダイエットに森公美子さんが挑戦し、1ヵ月半で7キロの減量に成功
- B 2007/3/31
みのもんたさん司会の日本テレビ系「おもしろイイ!! テレビ」が6月5日放送から計3回も特集した。

「バナナダイエットでダイエットできる？」に関する評判に影響を与えた出来事 重要度順 時系列順

○ A 2008/9/21 朝バナナダイエットに森公美子さんが挑戦し、1ヵ月半で7キロの減量に成功

● どんな出来事？

朝バナナダイエットに森公美子さんが挑戦し、1ヵ月半で7キロの減量に成功！森公美子さんのせいで、スーパーなどではバナナの品切れが続出です。森公美子さん！朝バナナダイエットで超スリムになったら許してあげるけど、失敗したら許してあげないからね！
[朝バナナダイエット森公美子](#)

● この出来事による影響

関心：↑ 評判：肯定↑ 否定↑ 新しい論点：森公美子

「この出来事で、『バナナダイエット』に多くの人々が関心を持つようになり、良い評判も増えました。

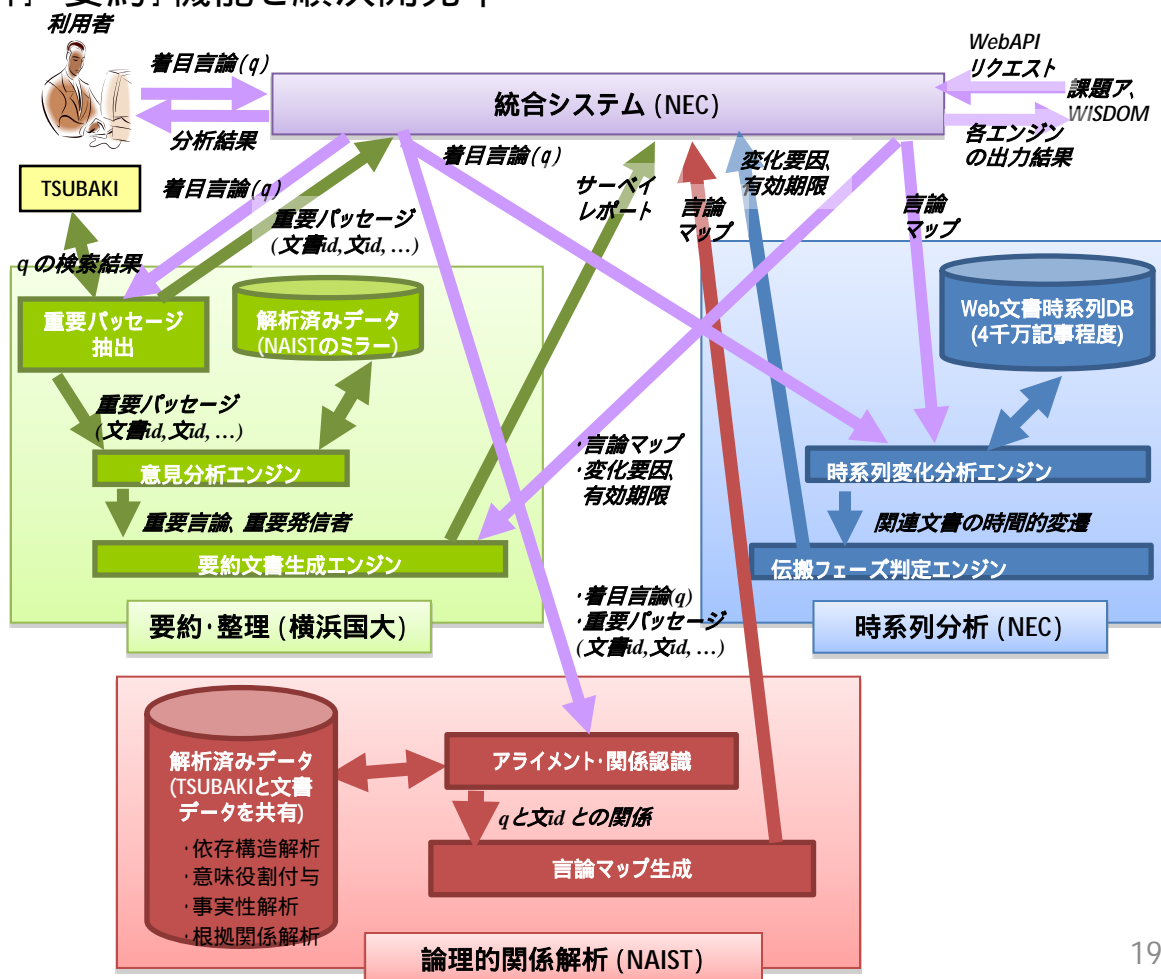
統合システム、実証実験に向けての活動

統合システムモックアップ第2版を設計・開発中

- 出力情報の信頼性判断支援としての意味・効果を整理し、それに合わせて、画面イメージやUI機能を再設計
- クライアント側での素早い応答が可能な java script + Flex ベースでの実装を採用
- 現在、「言論マップ」「時系列分析」「要約」機能を順次開発中

課題イの統合システムでは各エンジンを分散配置

- 重要パッセージ抽出にNICTと連携し、TSUBAKIの利用を想定
- 各エンジンで使用するDBをローカルに配置
 - ・Web文書はTSUBAKIの文書id, 文idで共通管理
- 9月頃までに各エンジンのWebAPIを策定
- 平成21年度末に、各エンジンと結合した統合システム第1版を開発



課題イ これまでの主な研究成果と達成度

イ-1-1 着目言論を論理空間に位置づける (奈良先端大)

- **言論マップ生成**: 論理的関係認識基本システムを開発。言論間の論理的関係の体系化とコーパスを構築(約2,000エントリ、一般公開済) 達成度60%
- **大規模知識ベース**: 事象間関係知識を精錬(事象間二項関係DB約5万3千件、述語シソーラス約2万3千語義)。WikipediaおよびHTML文書から上位下位関係知識獲得(約250万件、精度90%)、Wikipediaリダイレクトから同義関係知識を獲得(約20万件) 達成度90%

イ-1-2 着目言論周辺の言論空間を再構成する (横浜国大)

- **重要言論・重要発信者抽出**: 重要言論抽出のR-精度60%、発信者抽出の精度50% 達成度60%
- **要約文書生成**: 重要パッセージ抽出の精度50%。信頼性判断支援のための要約の情報構造を分析し、俯瞰的要約に加えて、ユーザが読み解くための調停要約を検討 達成度60%

イ-2 着目言論を時間軸に位置づける (NEC)

- **変化要因抽出**: 重要変化区間を選別し、そこから変化要因となる重要イベント抽出する基本手法を開発(キーワードレベル精度約70%)。評価評判表現を利用した高精度化手法を考案 達成度60%
- **有効期限切れ判定**: 物・サービス等に関して、変化点候補検証に基づく有効期限切れ判定の基本的手法を開発、その有効性を確認 達成度60%
- **時系列化**: Webページの記事分割・発信日付付与技術を開発(定型48サイト 精度99%、単一記事ページ 精度80%)、NICTコーパス他から時系列データベースを構築(1713万記事) 達成度80%

統合システム、実証実験 (NEC)

- 統合システムのUI機能を設計し、「言論マップ」、「時系列分析」機能のUIモジュールを試作。NICTを含めた連携・統合に向けて、エンジン間連携方針や、文書データの共有・管理方針を決定 達成度50%

関連技術の施策

- 関連技術の施策(インターネット上の違法・有害情報検出技術の研究開発) -

インターネット上の違法・有害情報検出技術の研究開発

時々刻々新たに流通する違法・有害情報を検出するには、効果的なコンテンツ・チェック方式の高度化が不可欠。現在インターネット接続事業者(ISP)等で用いられているのは、単語レベルで一致したものを検出する技術が一般的であるが、今後は、文脈の意味を解析して、文章単位で違法・有害な情報を検出することのできる技術を開発することにより、ISP等における違法・有害情報の検出の迅速化を図るとともに、検出の負担を軽減し、もってインターネット上の情報の適正化を推進することが重要な課題。

本研究開発は、独立行政法人情報通信研究機構(NICT)から株KDDI研究所に委託され、本年9月から研究が開始されている。実施期間は、平成21年度から平成23年度を予定。

