

平成21年度「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」 課題A「Webコンテンツ分析技術」の開発成果について

1. 施策の目標

ネットワーク上の文字、音声、画像・映像情報について、情報の信頼性を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、誰でもが思いのまま、簡単に、信頼して、コンテンツを取り扱い、高度に活用できる環境を実現する。

2. 研究開発の背景

一般の人々が通常のPCによって容易に映像や音声情報を編集できるようになったこともあり、価値の低い情報や信頼性の低い情報が急増している。一旦誤った情報がインターネット上に流通すると、多くの情報利用者が誤った内容を信じてしまうことが危惧される。その大量のWebコンテンツを構成するテキスト、画像、音声、映像情報等の信頼性をデータ解析的な手法に基づき分析する技術を開発する必要がある。

3. 研究開発の概要と期待される効果

画像・音声・映像情報の分析技術：画像・音声・映像情報とそれを説明するテキスト情報や周辺のコンテンツから画像・音声・映像情報の内容を推定し、インターネット上から同等のテキスト情報とそれに関連づけられた画像・音声・映像情報を取得し、対象とする情報と取得された情報を分類するなどして、画像・音声・映像情報とそれを説明するテキスト情報の不整合度を検証する技術について研究開発を行う。Webコンテンツ上の画像・音声・映像を含む情報の信憑性を評価する手がかりを情報利用者に提示することによって、画像・音声・映像情報とそれを説明するテキスト情報の不整合度を検証することができるようになる。

テキスト情報の分析技術：Webコンテンツを構成するテキスト表現の表層的特徴（文体、表現、語彙等）やテキスト情報の発信者の専門性等、Webコンテンツの信憑性を評価する手がかりとなる側面におけるテキスト情報の分析技術について研究開発を行う。情報利用者がWebコンテンツを構成するテキスト情報の信憑性を評価する基準を与えることができるようになる。

画像・音声・映像情報の分析技術



テキスト情報の分析技術



4. 研究開発の期間及び体制

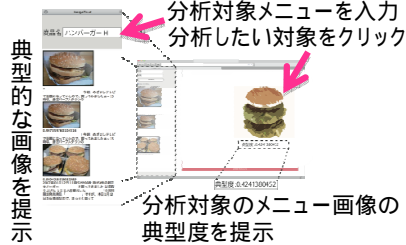
平成19年度～平成22年度(4年間)

NICT委託研究(京都大学、兵庫県立大学、京都産業大学、株式会社きざしカンパニー)

画像・音声・映像情報の分析技術の主な成果

画像・音声・映像情報の分析技術

・ウェブ上で公開されている撮影者の意図が強く反映された画像の信憑性を検証するため、関連画像を自動的に検索し、そのサイトで掲載されている画像がどの程度標準的・典型的なものを判断できるようなシステムを開発しました。



・Wikipediaの画像とその記事に対するふさわしさを検証する技術を開発しました。記事に対応するオブジェクトと、画像で表現される複数のオブジェクトとの関係に着目し、その関係の強さ、記事がその関係と整合しているか、画像がその関係を表すのに典型的かどうかを評価・図示し、総合的なふさわしさを算出しています。



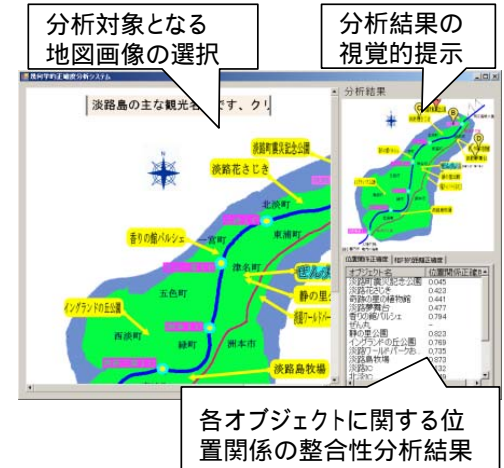
・映像コンテンツと、そのコンテンツに対して視聴者が付けているアノテーションの関係に着目し、同じ映像を使っているが視聴者の評価が肯定と否定に大きく割れているようなコンテンツは編集に問題があるとみなし、ユーザが指定した映像と最も対立しているような映像をユーザに提示するシステムを開発しました。これにより、映像を試聴する際に編集の違いを考慮して視聴できるようにします。



人物とその関係に関する記述に着目して、視聴している映像ニュースの報道全体における偏りや説明と主張の不整合を、字幕データや関連映像をもとに分析して提示するシステムを開発

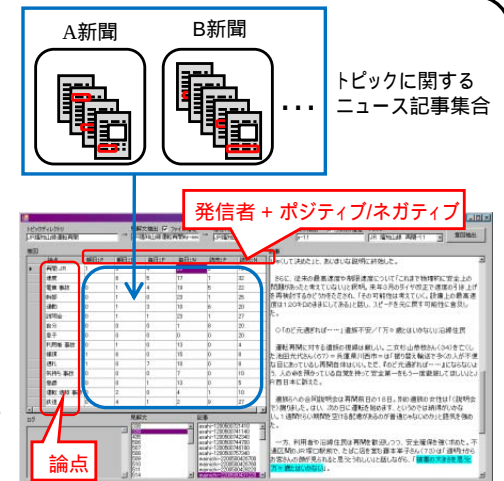
・このシステムは、Web上の略地図画像に対して、地理オブジェクト情報を抽出し、その地理オブジェクトの略地図上での位置関係と、オンライン地図から抽出した実空間上での位置関係を比較し、オブジェクトの相対的な配置の整合性を分析するものです。

・これにより、略地図画像から正しい地理情報を読み取ることが可能となります。



・このシステムは特定のトピックに関するニュース記事集合から**主要な論点**を抽出し、発信者(報道機関など)毎に各論点をどのように(どのくらい**ポジティブ/ネガティブ**)に伝えているかを編集意図とみなし、表形式で提示するものです。

・これにより、発信者の編集意図を比較しながらのニュース記事の視聴・閲覧が可能になり、ニュースの偏った理解を避けることができます。



テキスト情報の分析技術の主な成果

テキスト情報の分析技術

・このシステムはWeb検索結果、QAコンテンツ、レストラン検索結果などと共に下記のような指標を提示することでユーザの信憑性判定を手助けします。

コンテンツの公正さを表す指標

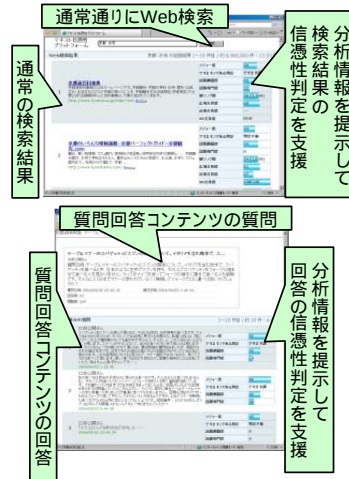
コンテンツそのものの特徴の分析

- ・ **メジャー度**
- ・ **話題網羅度**
- ・ **話題専門度**
- ・ **体裁(ですます / である判定)**

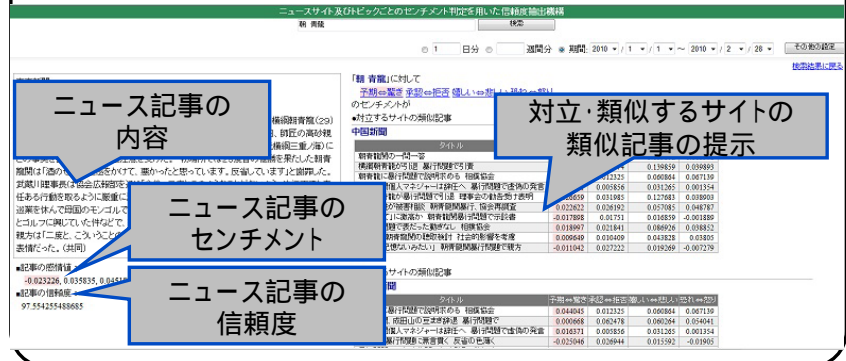
コンテンツの社会的支持度を表す指標

コンテンツの外部からの評価の分析

- ・ **ソーシャルブックマーク件数**
- ・ **被リンク数**
- ・ **広域支持度 / 近接支持度**



・このシステムは、任意の話題に対してセンチメント排反度・一致度が高いニュースサイトを発見でき、センチメントの排反・一致関係を分析することでニュース記事の信頼度を算出します。

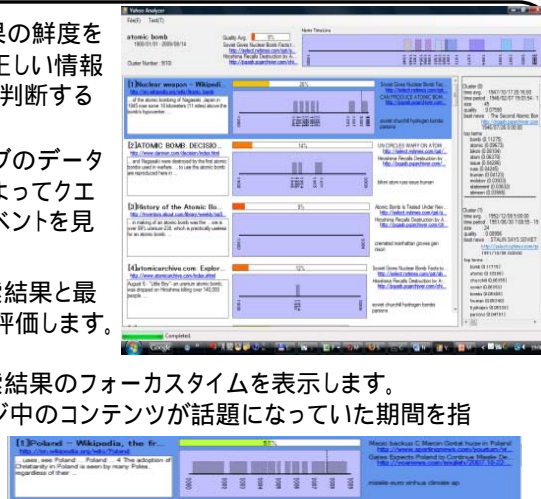


・このシステムは検索結果の鮮度を評価する事で、現時点で正しい情報を提供しているかどうかを判断する支援を行います。

・まず、ニュースアーカイブのデータを使い、クラスタリングによってクエリに関連した最新のイベントを見つけます。

・さらに、それぞれの検索結果と最新のイベントの関連性を評価します。

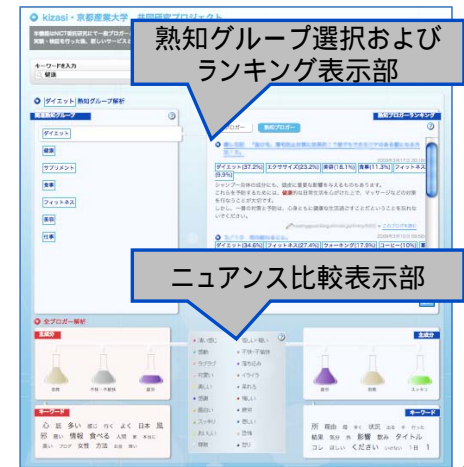
・最後にそれぞれの検索結果のフォーカスタイムを表示します。フォーカスタイムはページ中のコンテンツが話題になっていた期間を指し示します。



・このシステムはブロガーの体験熟知度を算出し、プロゲントリのランキングを提供します。

・また、視点の異なる複数のランキングを提示するとともに、ブロガーの特性に関する補助情報を提示します。

・これにより、ユーザは閲覧するプロゲントリの信頼性を自分なりに判断することが可能となります。



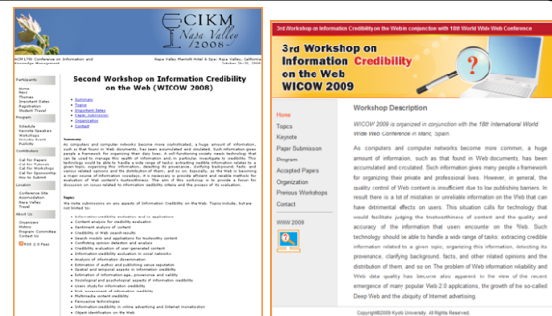
1. これまで得られた成果(特許出願や論文発表等)

	特許出願	論文	研究発表	報道発表	標準化提案	学会での受賞
「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」 課題ア Webコンテンツ分析技術	2	39	83	1	0	7

2. 研究成果発表会等の開催について

(1) 国際ワークショップを2年連続で開催

Web上のコンテンツの信憑性に関する国際ワークショップ WICOWを2年連続で開催した。平成20年10月には米国で開催されたCIKM2008において、平成21年4月にはスペインで開催されたWWW2009において、本ワークショップを開催した。これらの会議は情報検索やWebの分野における最も権威のある国際会議であり、情報信憑性についての高度な議論を推進することができた。



(2) 国際ワークショップを開催予定、国際会議におけるミニトラックを主催予定

平成22年度4月に米国で開催されるWWW2010において、WICOW2010ワークショップを3年連続で開催する予定。WWW2010はWebの分野における最も権威のある国際会議であり、情報信憑性についての深い議論が期待される。平成23年1月には、米国で開催されるHICSS2011国際会議におけるミニトラック「Web Information Credibility Analysis」を主催し、さらに高度な議論を推進する予定である。

