

平成22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、高度に利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

①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と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。

②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。

①画像・音声・映像情報の分析技術

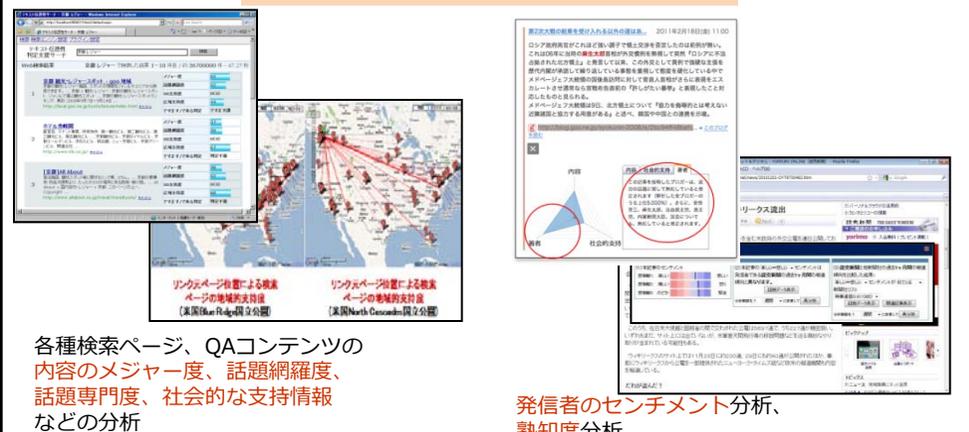


映像・音声の偏り分析や視聴者評価情報の分析

Web画像の典型度・過不足性・整合性の分析

Webアーカイブを用いた電子地図の誤記載・未記載度の分析

②テキスト情報の分析技術



各種検索ページ、QAコンテンツの内容のメジャー度、話題網羅度、話題専門度、社会的な支持情報などの分析

発信者のセンチメント分析、熟知度分析

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)

①画像・音声・映像情報の分析技術の主な成果

①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。



典型的な画像を提示

Wikipediaの画像とその記事およびキャプションに対するふさわしさを検証する技術を開発しました。記事オブジェクトとキャプション中の複数のオブジェクトとの関係(関連度)と、キャプションと画像との関係(典型度)に着目し、信憑性を判定しています。また、キャプションによりふさわしい代替画像の検索もできるシステムを開発しました。



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



分析対象の地図画像を選択

分析結果の視覚的提示

大きな地名ほど空間的に不正確

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。

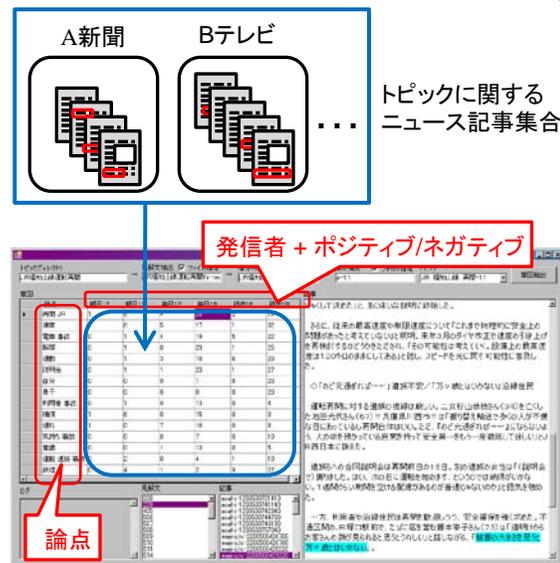


類似動画との比較結果を提示

映像ニュースが伝えている内容が、他ではどのように伝えているか、差異はどの程度あるのかなど、視聴している映像コンテンツの全体における偏りや説明と主張の不整合を、人物に関する表現や記述極性に着目して分析するエンジンを開発しました。また、開発したエンジンとWISDOMとの連携も行いました。



このシステムはある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、発信者ごとに各論点をどのくらい**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。
また、伝えている論点に偏りがいないかを数値として計算する技術も開発しました。



トピックに関するニュース記事集合

発信者 + ポジティブ/ネガティブ

論点

②テキスト情報の分析技術の主な成果

②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

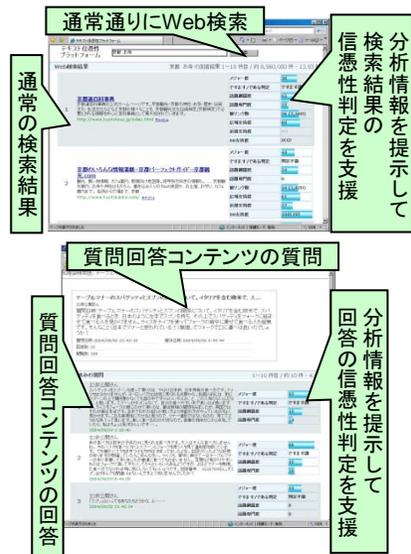
コンテンツそのものの特徴の分析

- ・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。

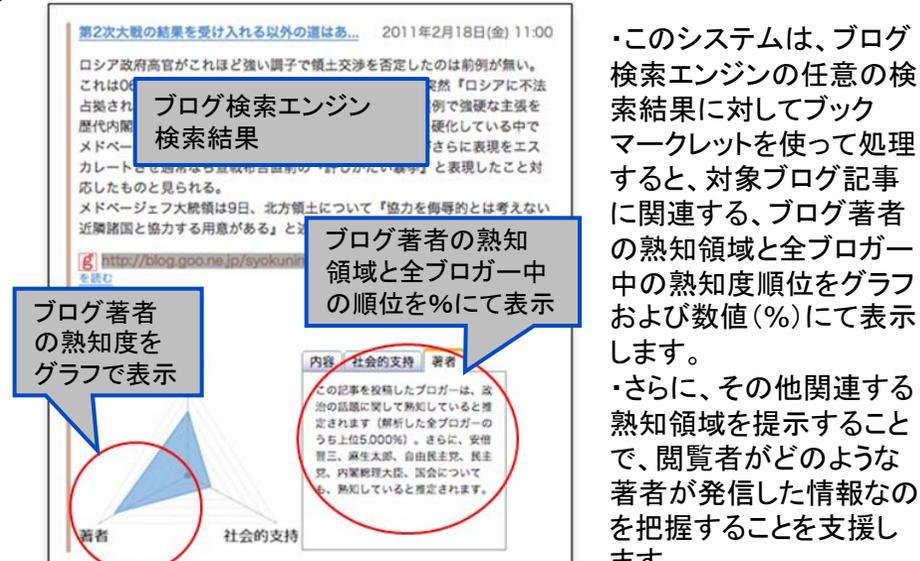


・検索結果の鮮度を評価する事で、現時点で正しい情報を提供しているかどうかの判断を支援します。

・例えば、「最新」といった語が書かれているWebページ中のコンテンツが、本当に新しいトピックが含んでいるかどうかをWebマイニング技術を用いて判定します。

・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ーに関連した最新のイベントを見つけ、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。

・それぞれの検索結果にはフォーカスタイム(ページ中のコンテンツが話題になっていた期間)が提示されます。



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	展示回	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	3 (1)	74 (31)	87 (21)	1 (0)	0 (0)	21 (1)	16 (1)

(1) 表彰・受賞

10th International Symposium on Web and Wireless Geographical Information Systems (W2GIS 2011)において、発表した論文「Extraction and Geographical Navigation of Important Historical Events in the Web」がBest Paper Awardを受賞

2. 国際会議のミニトラック主催、Web情報の信憑性に関する国際ワークショップを継続して開催

Web情報の信憑性に関する国際ワークショップWorkshop on Information Credibility on the Web (WICOW)を4年連続で開催した。今年度は、平成22年4月に、WWW2010(米国)においてWICOWワークショップを開催した。平成23年3月にはWWW2011(インド)において、WICOWとAIRWebが連携し、ウェブ情報の品質に関するワークショップJoint WICOW/AIRWeb Workshop on Web Quality (WebQuality 2011)を開催した。

さらに、平成23年1月に、HICSS44国際会議(米国)におけるミニトラック「Web Information Credibility Analysis」を主催した。より高度な議論を行うことができた。

我々の手によるこれらの活動によって、Web情報の信憑性に関する国際的で継続的な情報発信の場が形成された。

