

「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

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

- 画像・音声・映像情報がWebコンテンツの一部としてインターネット上に大量に流通しているが、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。さらには、オリジナルのコンテンツを、悪意を持って改ざんしたWebコンテンツも多く存在するようになってきている。画像や映像、音声情報を含んだWebコンテンツはテキストだけのものよりも多くの情報量があり、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内容から画像・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。

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)

研究開発の概要のイメージ

①画像・音声・映像情報の分析技術

(周辺コンテンツや他情報源から収集したテキスト情報の分析・画像特徴量分析による信憑性検証)



Web画像の典型度・過不足性・整合性の分析

映像・音声の偏り分析や視聴者評価情報の分析

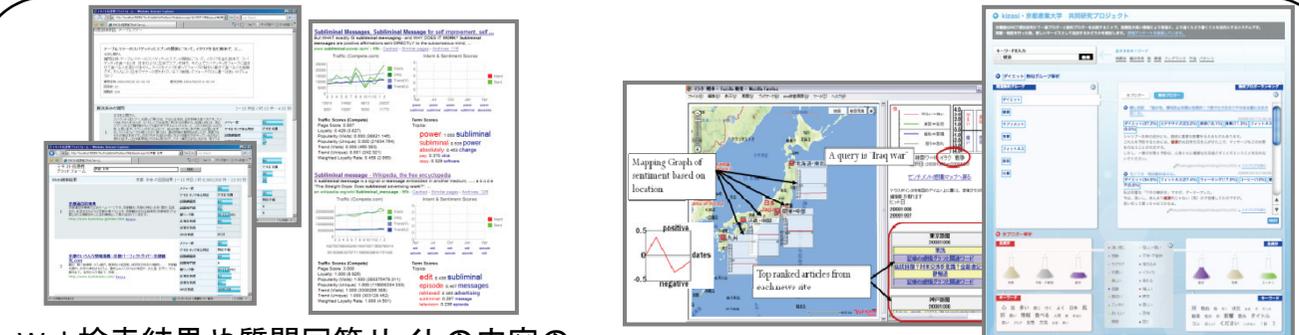
Webアーカイブを用いた電子地図の誤記載度・未記載度分析

ヤフー株式会社のポータルサイトや検索エンジン連動での実証実験

(株)きざしカンパニーのポータルサイトや検索エンジン連動での実証実験

②テキスト情報の分析技術

(Webページのメジャー度・話題網羅度・社会的な支持分析、発信者分析に基づく信憑性検証)



Web検索結果や質問回答サイトの内容のメジャー度、話題網羅度、社会的支持の分析

発信者のセンチメント分析／熟知度分析

情報信憑性検証基盤技術の確立
国際的な情報発信

①画像・音声・映像情報の分析技術の主な成果

Web上のマルチメディア情報の信憑性検証技術の開発

(周辺コンテンツや他情報源から収集したテキスト情報の分析や画像特徴量分析による信憑性検証)

Web画像情報の信憑性 (京大・兵庫県大)

- 画像が、どの程度実物と相違するか、どの程度標準的なものを他の情報源から関連画像を自動的に検索し比較分析する技術を開発
- 画像集合の過不足性に基づく信憑性判断支援技術を開発
- Wikipediaの画像とその記事に対する整合性を検証する技術を開発
- 電子地図データの地物の誤記載や未記載をWebアーカイブを用いて、分析発見する技術を開発



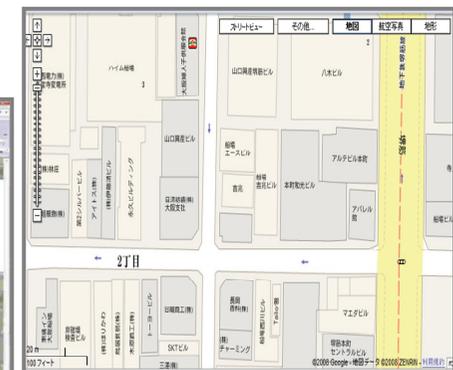
Web画像の典型度・過不足性・整合性の分析

Web動画・音声コンテンツの信憑性 (京大・兵庫県大)

- 動画に対し視聴者が付した感想・評価コメントを自動取得・分析し、動画の信頼性判定を行う技術を開発
- Web上の動画やTVニュースを、他局や他メディアではどのように伝えているかなど、対象コンテンツの全体における偏り(不整合度)を、字幕データや関連情報をもとに分析提示する技術を開発



映像・音声の偏り分析や
視聴者評価情報の分析



Webアーカイブを用いた
電子地図の誤記載度・
未記載度分析

②テキスト情報の分析技術の主な成果

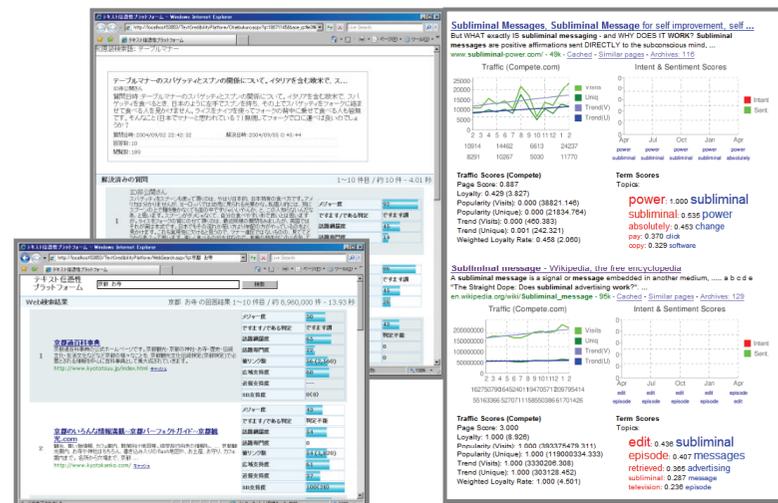
Webテキスト情報の信憑性検証技術の開発
 (Webページのメジャー度・話題網羅度・社会的な支持分析、発信者分析に基づく信憑性検証)

Webテキスト情報の信憑性(京大)

- サーチエンジンで検索されたWebページの、文章の体裁、検索語に関する典型的な話題の網羅性、話題の専門性、検索結果内での内容のメジャー度、社会的・地理的な支持度という指標を示して信憑性判断する技術の開発
- 上記の様々な分析手法について時間的側面からの分析技術の開発

Webコンテンツの発信者の信頼度分析(京産大) 熟知度分析によるブロガーの自動分類とブログランキング手法の開発・実証(きざしカンパニー)

- 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か、発信している分野にどの程度熟知しているかを分析するため、感情表現語辞書、および、過去の発信情報履歴を分析してブログ情報の書き手の熟知度を分析する技術の開発
- 発信者の熟知度分析手法を用いて、ある話題に関するブログを検索した場合に、その話題に関連する分野に熟知しているブロガーのブログを上位にランキングする手法を開発し、(株)きざしカンパニーのWebサイトでの実証実験を開始(平成20年10月1日)



検索ページの内容のメジャー度、
 話題網羅度、社会的な支持情報分析



発信者のセンチメント分析／熟知度分析

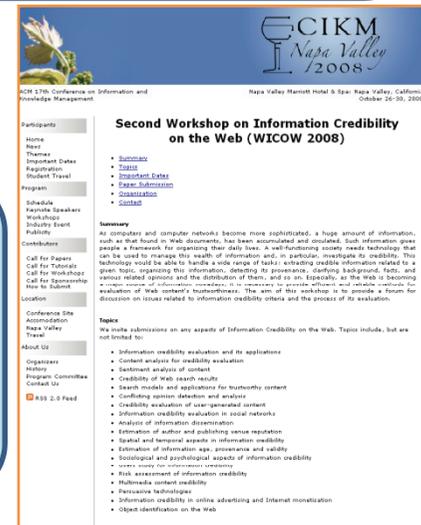
1. これまで得られた成果(特許出願や論文発表等)

	特許出願	論文	研究発表	報道発表	標準化提案	学会における受賞
「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	1	16	43	1	0	6

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WICOW2008を開催

平成20年10月に米国で開催されたACM CIKM 2008国際会議において、世界でもほぼ唯一のWeb上のコンテンツの信憑性についての国際ワークショップWICOW2008を開催した。CIKM国際会議は情報検索、データベース、ナレッジマネジメントの分野において最も権威のある会議の1つであり、情報信憑性についての高度な議論を推進した。



② 国際ワークショップWICOW2009の開催を計画

WICOW2008に引き続き、平成21年4月にスペインで開催されるWWW 2009国際会議において、Web上のコンテンツの信憑性についての国際ワークショップWICOW2009を計画し、開催が決定している。WWW国際会議はWeb研究の最も権威のある会議であり、Webコンテンツの情報信憑性についての研究の重要性について深い議論が期待される。

