

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) 研究の目的

本研究の最終目標は、本研究課題のアに対応して、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、画像等マルチメディア情報の内容の特徴量分析、発信者分析等をもとに「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、および、開発技術の ISP などの実環境における実証実験を行うことである。

それぞれの課題において各種の評価実験を設定し、課題ア－1、課題ア－2については、それぞれ、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 65%以上を達成する。

さらに、国内外の学術会議において積極的に発表等を行ったり、特許を取得したりすることにより、本課題の有効性を世の中に広く発信する。

(2) 研究期間

平成 19 年度から平成 22 年度（4 年間）

(3) 委託先企業

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

(4) 研究予算（百万円）

平成 19 年度 119.6 百万円

平成 20 年度 125.0 百万円

平成 21 年度 105.6 百万円

平成 22 年度 99.3 百万円（予定）

(5) 研究開発課題と担当

課題ア－1 画像・音声・映像情報の分析技術

課題ア－1－1 (京都大学)

画像・音声・映像情報の信憑性検証のためのクロスメディア情報
分析・検索技術の開発

課題ア－1－2 (兵庫県立大学)

発信者・ユーザ意図検出と画像等の周辺コンテンツの
クロスメディア情報分析による信憑性検証

課題ア－2 テキスト情報の分析技術

課題ア－2－1 (京都大学)

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－2－2 (京都産業大学)

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

課題ア－2－3 (株式会社きざしカンパニー)

マニア指向ランキング機能を有するブログサーチエンジン

(6) 主な研究成果

特許出願： 1 件

外部発表： 62 件

具体的な成果

(1) Web 画像情報の信憑性

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判断するため、他の情報源から関連する画像を自動的に検索し比較分析して提示することで、サイトで掲載されている画像がどの程度実物と相違するかということや、そのサイトで掲載されている画像がどの程度標準的・典型的なものかを利用者にとって判断しやすいサーチエンジンおよびインタフェースを試作開発しました。また、ある人物や組織などに関する Wikipedia ページに含まれている画像が、どの人物や組織とどの程度関係が深いのかといった、画像の整合性を検証するエンジンおよびインタフェースを試作開発しました (京都大学)。

(2) 動画・音声コンテンツの信憑性

ウェブ上の動画・音声コンテンツの内容の信憑性を利用者が判断できるようにするために、動画や音声に対して他の多くの視聴者がつけた感想・評価コメント (ソーシャルアノテーション) を自動的に

取得・分析し、動画・音声コンテンツが信頼できる内容かどうか、見るに値する内容かどうかを判断するのに有用な情報を提示するシステムや、インターネットや TV のニュースなどの映像・音声コンテンツが伝えている内容が、他局や他のメディア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・音声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、字幕データや関連情報をもとに自動分析して見せてくれるシステムを開発しています（京都大学）。また、情報発信者が何を伝え、何を伝えていないかといった編集者の意図を自動分析し、比較しやすく提示する技術（兵庫県立大学）も開発しています。

(3) Web 上の電子地図情報の信憑性

Web 上の電子地図は広く普及していますが、電子地図は頻繁に更新されておらず、すでに存在しない地物が地図上に掲載されていることがあります。また、制作者の主観によって表示の制御がおこなわれるなど、電子地図の信憑性は重要な課題となっています。本研究では、電子地図データと関連 Web 情報アーカイブを用いて、このような地物の誤記載や未記載の量、表示基準の不統一性を電子地図の信憑性を量る尺度として計算する手法を開発しています（兵庫県立大学）。

(4) Web テキスト情報の信憑性

Web 検索エンジンで検索された Web ページや、質問回答サイトの回答文などのテキスト情報を対象とし、テキスト情報自身の内容の分析と、テキスト情報の社会的な支持の分析によって、信憑性を判定する技術を開発しています。テキスト情報自身の内容の分析としては、文章の体裁、検索クエリに関する代表的な話題をどの程度網羅しているか、ページ内容が検索クエリに関する専門的な内容をどの程度含んでいるか、ページの内容が検索結果の中でどの程度メジャーな内容かといった指標を扱います。テキスト情報の社会的な支持の分析としては、ソーシャルブックマークの分析やリンク元ページの地理的分布の分析を行います。（京都大学、兵庫県立大学）。

(5) Web コンテンツの発信者の信頼度

Web コンテンツの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は、発信情報の信憑性を判断するために重要な情報です。本研究では、各新聞社の過去の記事データを分析して感情表現語辞書を開発しました。評価実験を通じて、この感情表現誤辞書の精度について 10%程度の平均誤差を確認しています。この辞書

を用いて、新たな記事データのセンチメント値を算出し地域ごとに集計することで、同一の事象・事件を報道する各発信者のセンチメントの相違を提示できるシステムを開発しました。これにより、立場の違いによる情報発信の相違が分かるため、ユーザは話題および各記事に対する信憑性を判断できます。評価実験では開発したシステムが 68%の精度で信憑性の判断材料として有効であることが確認できました。

また、発信者が発信している情報の分野にどの程度熟知しているかを、過去の発信情報履歴を分析して推定する手法を開発し、ブログ情報の書き手の熟知度分析などに応用しています。ここでは単に熟知度の推定だけでなく、その時のセンチメント分析を行うことで、発信された情報の信憑性を情報閲覧者が推定可能となるような取り組みも行っています（京都産業大学）。

(6) 熟知度分析に基づくブログの書き手（ブロガー）の自動分類とブログランキング手法の開発と実証実験

京都産業大学による情報発信者の熟知度分析手法を用いて、発信者の自動分類技術およびサーチエンジンの拡張開発を行いました。分類カテゴリ（熟知領域）の大幅な拡張、エンジン的高速化、支援情報の提示などの精度・機能の拡張を実施しました。実証実験システムを株式会社きざしカンパニーの Web サイトにて公開し、評価実験を行い、公開を継続しています。

(7) 情報信憑性検証に関する研究開発に関する国際的な情報発信

平成 20 年 10 月に米国で開催された ACM CIKM 国際会議において情報信憑性に関する国際ワークショップ WICOW2008 (<http://www.dl.kuis.kyoto-u.ac.jp/wicow2/>) を開催しました。また、平成 21 年 4 月にスペインで開催される World Wide Web 2009 (WWW2009) 国際会議において、情報信憑性に関する国際ワークショップ WICOW2009 (<http://www.dl.kuis.kyoto-u.ac.jp/wicow3/>) の開催を計画しました。

(7) 研究開発イメージ図

次ページより添付

「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コテンツから画像・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。

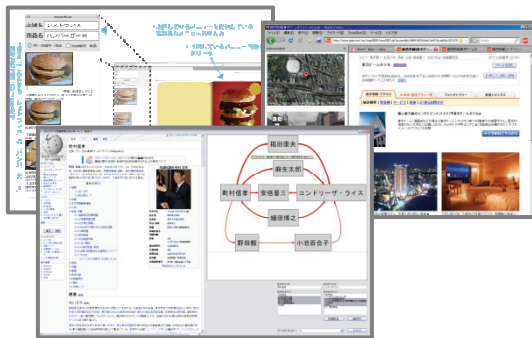
4. 研究開発の期間及び体制

- 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

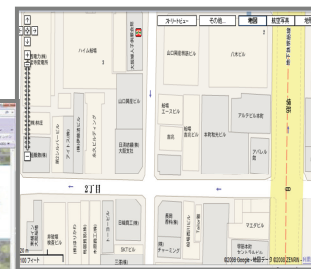
(周辺コンテンツや他情報源から収集したテキスト情報の分析・画像特徴量分析による信憑性検証)



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

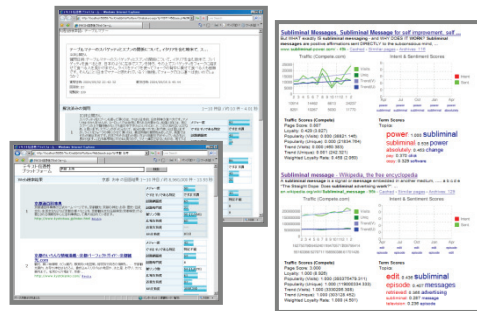


Webアーカイブを用いた
電子地図の誤記載度・
未記載度分析

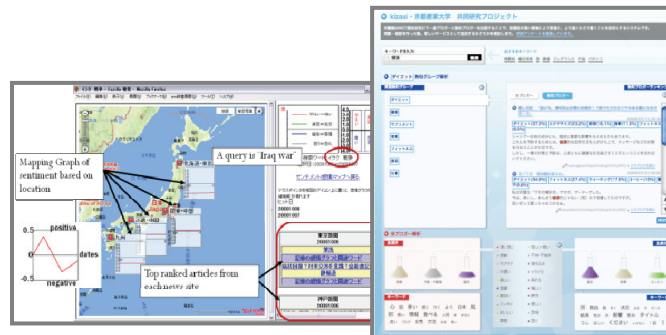
ヤフー株式会社の
ポータルサイトや検索エン
ジン連動での実証実験

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

(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

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



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

②テキスト情報の分析技術の主な成果

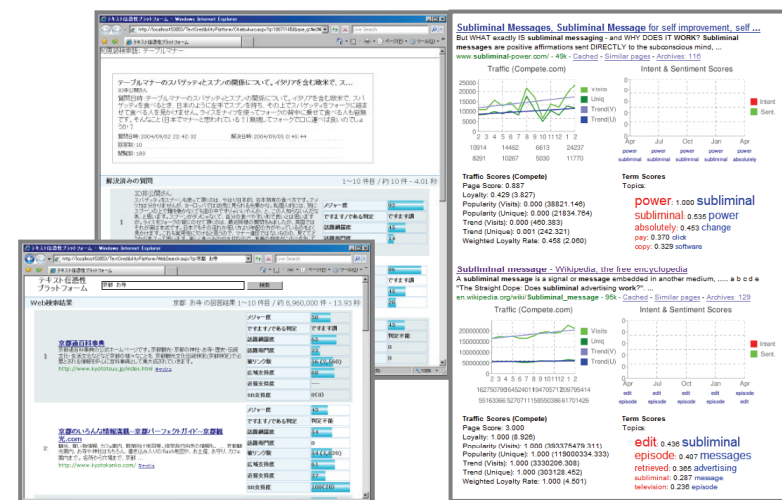
Webテキスト情報の信憑性検証技術の開発
(Webページのメジャー度・話題網羅度・社会的な支持分析、発信者分析に基づく信憑性検証)

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

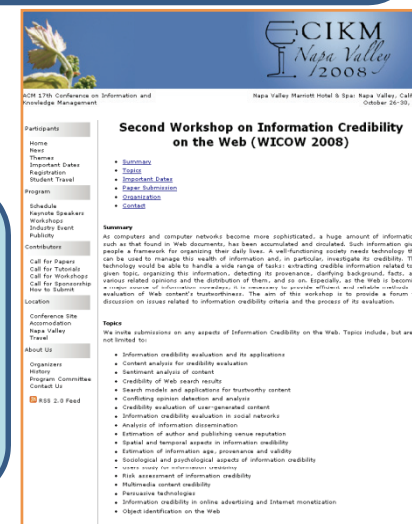
1. これまで得られた成果(特許出願や論文発表等)

	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。

