

平成 21 年 5 月 31 日現在

研究種目：若手研究（B）
 研究期間：2007～2008
 課題番号：19700099
 研究課題名（和文） 国際問題関連ドキュメント群を対象とした高度知識発見・情報分析システムの実現
 研究課題名（英文） Advanced Information Analysis and Knowledge Discovery System for International Relations Documents
 研究代表者
 佐々木 史織（SASAKI SHIORI）
 慶應義塾大学・大学院政策・メディア研究科・助教
 研究者番号：90383934

研究成果の概要：

本研究では、Web上に日々公開される国際問題に関するドキュメントデータから意味的なメタデータと時間的・空間的情報を自動的に抽出し、高度な知識発見を可能とする情報分析システムを設計、実現した。利用者は、本システムに各自の関心・視点に基づくキーワードを入力するだけで、専門的な文書内容の時系列的变化や地域別差異、問題領域別・情報源別の分析結果を多角的・定量的に獲得し、視覚的に把握することが可能となった。

交付額

（金額単位：円）

	直接経費	間接経費	合計
2007年度	2,300,000	0	2,300,000
2008年度	900,000	270,000	1,170,000
年度			
年度			
年度			
総計	3,200,000	270,000	3,470,000

研究分野：データベース，情報分析，国際関係論，教育工学

科研費の分科・細目：情報学・メディア情報学・データベース

キーワード：情報システム，可視化，ディレクトリ・情報検索，データベース，国際関係論，意味的連想検索，時空間的分析

1. 研究開始当初の背景

現在、広域ネットワーク上には、膨大な量の国際ニュース記事、各国政府や国際機関の公式発表、政府高官の談話、プレス・ブリーフィング、政策ステートメント、議会議事録、NGOの活動記録といった国際関係・国際政治に関連する重要なドキュメントデータが日々公開されており、これらの大量のデータから迅速かつ的確な情報を獲得することは、国際関係に携わる研究者・実務者のみならず、社会全体の緊急課題となっている。しかし、国際関係論や国際政治学の分野において行

われているドキュメント分析は、新聞や機関紙、演説、声明、交換文書、書簡などを対象とした政策決定者の認知、態度、対外イメージ、世論などの事後的・解釈学的な分析が主である（図1）。また、内容分析（Ole R. Holsti, "Content Analysis," 1968.）や認知構造図（Robert Axelrod ed., *The Structure of Decision: The Cognitive Maps of Political Elites*, 1976.）といったこの分野における従来の定量的なドキュメント分析手法は、主にドキュメントの静的な性質を対象とした知識発見の方法であり、分析者の視点や関心

に応じて意味内容を多角的・動的に分析することは困難であった。さらに、これらの手法に伴う概念や文章のコード化、信頼性のチェックなどは高度に専門的な知識を要するため、大量のデータを対象とした比較分析には不向きであった。

こうした従来の手法に対し、意味の数学モデルによる意味的連想検索方式（Yasushi Kiyoki, "Multimedia Data Management - using metadata to integrate and apply digital media -", 1998）は、分析者の視点や分析時の文脈に応じてダイナミックに言葉およびデータの意味解釈を実現する方式であり、その意味解釈を多次元直交ベクトル空間における相関量計算によって定量的に行うことを特徴としている。したがって、本方式を国際問題に関するドキュメントの検索・分析に応用することにより、対象ドキュメントのコード化を経ずに客観的な形で大量のデータの比較分析が可能になる。さらに、ドキュメントから抽出された時間的・空間的情報に基づいた時系列的変化の分析、意味的・動的クラスタリング分析、分析結果の問題領域別・情報源別・地域別の表示などの分析結果が可視化されることにより、分析結果の直感的な理解が容易になると考えられる。

また、学術的には、図1に示すように、国際関係分野と情報検索・データマイニング分野の研究が結びつくことにより、政策科学におけるエマージェントな国際問題に関する新しい研究領域が形成されると考えられる。加えて、本研究で実現されるシステムにより、国際問題に関する情報を自動的かつ定量的に獲得し、さらにドキュメントの意味解釈を伴った分析結果を獲得することが可能となることは、大きな社会的寄与になると考えられる。

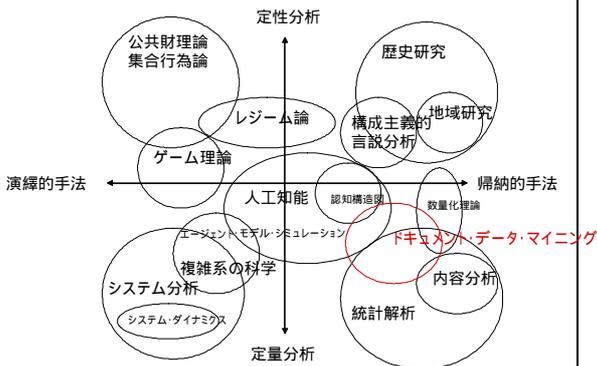


図1：政策科学における諸方法論と本研究の位置づけ

2. 研究の目的

本研究では、意味的連想検索方式および国際関係分野の意味検索空間を用いて、広域ネットワーク上に日々公開される大量の国際問題に関するドキュメントデータから意味的なメタデータと時間的・空間的情報を自動的に抽出し、高度な知識発見を可能とする情報分析

システムを実現する。本研究の目的は、外交・安全保障・経済・金融・環境・技術・エネルギー・民族・宗教と多岐に渡る国際問題に関するメディアデータ、特にドキュメントデータを対象とし、専門的知識を反映した高度な知識発見・分析を実現することにある。具体的には、利用者（分析者）の視点や設定する問題領域に応じた多角的・動的なドキュメント群の検索、ドキュメント群の意味内容・コンテキストの時系列的变化の分析、意味的・動的クラスタリング分析、分析結果の問題領域別・情報源別・地域別の表示などのヴィジュアリゼーションの機能を含めた情報検索・分析システムの実現である。

研究代表者は従来の研究成果として、意味的連想検索方式を国際関係分野のドキュメント検索に応用し、当該分野の意味的連想検索空間の生成方式を提案している。（Sasaki, S. and Kiyoki, Y., "Space Creation and Evaluation Method using Specialized and General Knowledge for Semantic Associative Search," 2005.）この意味的連想検索空間は国際関係の専門用語間および専門用語と一般的な言葉との関連性を記述した空間であり、これを用いることにより、国際関係の専門家が専門用語のキーワードを用いてニュース記事や演説などの一般語によるドキュメントを検索可能であると同時に、専門的知識を持たない利用者も一般的なキーワードを用いて専門用語を含むドキュメントを検索することが可能な環境を提供している。この国際関係意味検索空間に、本研究において開発される知識発見・情報分析システムが連結されることにより、より実用的かつ総合的な国際問題に関する知識発見・情報分析システムが実現されると考えられる（図2）。

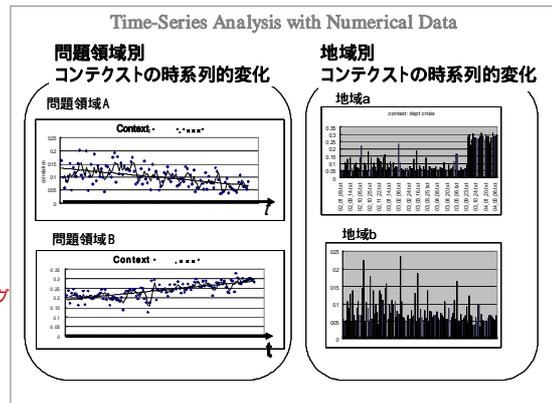


図2：システムの出力情報イメージ（一部）

本研究は、以上の背景および研究目的に基づき、これまでの研究において生成した国際関係意味検索空間を用いて、広域ネットワーク上に日々公開される大量の国際問題に関するドキュメントデータを自動的に検索し、高度な知識を発見する、総合的な情報検索・分析システムを実現する。具体的には、利用者（分析者）の視点や問題領域に応じてドキュメント群を多角的・動的に検索し、ドキュメント内容・コンテキストの時系列的变化や

地域別差異に関する分析を行い、その分析結果を問題領域別・情報源別・地域別に表示する機能を持つシステムの構築を行う。利用者はこのシステムを利用することにより、一般語のキーワードを用いて問題領域や関心領域を設定し、その領域に関連する国際問題に関する情報およびドキュメントの意味解釈と時間・空間的情報を伴った分析結果を、自動的かつ定量的に獲得することが可能となる。

3. 研究の方法

本研究では、高度知識発見・情報分析システムの実現のために、以下の計画に基づきシステムの設計および実現を行った。

(1) 利用者が検索条件を入力するに応じて広域ネットワーク上の複数情報源から該当ドキュメントを収集する機能を設計、実現する。この機能により、利用者は自動的に問題領域別・ネットワーク上の情報源別・期間別にドキュメントデータを収集することが可能になる。

(2) 収集したドキュメント群から自動的にメタデータを抽出する機能を設計、実現する。この機能により、検索空間に写像するための国際問題に特化した専門的なメタデータを自動的に獲得することが可能となる。

(3) メタデータ化されたドキュメントの情報をベクトル化し、国際関係意味的連想検索空間に写像する機能を設計、実現する。この機能により、空間に写像された情報と利用者が入力した視点や関心領域に関するキーワードと意味的な近さが意味的検索方式によって相関量として計量される。

(4) 意味検索空間の修正およびドキュメントデータベースの修正を実現する学習システムを設計、実現する。この学習システムを用いることにより、検索空間およびドキュメントデータベースに国際関係の専門家の知識を的確に反映することが可能となる。

(5) 国際問題に関する時間的・空間的情報データベースの設計、実現を行う。必要に応じて研究補助に依頼し、データベース入力を行う。

(6) (1)~(4)の情報検索システムと(5)の時間的・空間的情報データベースとの連結・統合を行うグローバル・アナライザーの設計、実現を行う。

(7) (6)の連結により得られた情報から新たな知識を発見する分析システムを設計、実現する。

(8) (7)から得られた分析結果を時系列およびイシュー別・地域別に表示するヴィジュアライゼーション機能を設計、実現する。これらの分析システムおよびヴィジュアライゼーション機能を用いることにより、利用者は(1)の機能における検索条件の入力ならびに(3)における問題領域に関するキーワードを入力するだけで、国際問題に関するドキュメント群を多角的・動的に検索し、ドキュメント内容・コンテキストの時系列的变化や地域別差異に関する分析、問題領域別・情報源別・

地域別に表示された分析結果を自動的かつ定量的に獲得することが可能となる。

4. 研究成果

本研究では、3.研究の方法に記した(1)~(8)の基礎技術・機能を設計・実現・連結することにより、国際問題関連ドキュメント群を対象とした高度知識発見・情報分析システムを構築した。

(1) 利用者の検索条件に応じたドキュメント収集機能：この機能により、利用者は自動的に問題領域別・ネットワーク上の情報源別・期間別にドキュメントデータを収集することが可能になった。(2) 意味的メタデータ自動抽出機能：この機能により、検索空間に写像するための国際問題に特化した専門的なメタデータを自動的に獲得することが可能となった。(3) 意味的連想検索空間マッピング機能：この機能により、空間に写像された情報と利用者が入力した視点や関心領域に関するキーワードと意味的な近さが意味的相関量として計量可能となった。(4) 時間的・空間的情報抽出機能およびデータベース：国際問題に関するドキュメントデータについて意味的な分析と共に時間的・空間的情報の分析が可能となった。(6) 意味的・時間的・空間的情報データベース連結グローバル・アナライザーを設計、実現を行った。(7) 意味的・時間的・空間的分析機能：この機能により、国際問題に関するドキュメントデータについて意味的な分析と共に時間的・空間的情報の分析が可能となった。(8) 時系列およびイシュー別・地域別ヴィジュアライゼーション機能：この機能により、ドキュメント群の意味的・時間的・空間的な分析結果を利用者に分かりやすく視覚的に表示することが可能となった。(9) 利用者が検索条件を入力するユーザ・インタフェースの設計、実現を行った。

以上の機能群の構築と統合により、国際問題に関する高度知識発見・情報分析システムの総合的な実験システムを実現し、これら研究成果を国内外の学会において発表した。利用者は、本システムに検索条件の入力および問題領域に関するキーワードを入力するだけで、国際問題に関するドキュメント群を多角的・動的に検索し、ドキュメント内容・コンテキストの時系列的变化や地域別差異に関する分析、問題領域別・情報源別・地域別に表示された分析結果を自動的かつ定量的に獲得することが可能となった。

さらに、以上の基礎技術および機能群を以下の4つの応用分野に適用し、複数のアプリケーション・プロトタイプ・システムを構築した。

(S-1) 意味的時空間的分析機構を有する4次元世界地図描画システム：本システムは、

歴史的・文化的なドキュメントデータ間の意味的・時空間的な関連性について、利用者の興味・関心・視点・テーマに応じて動的に計量・分析し、その結果を時系列で並べられた世界地図（4D World Map Set）や3次元空間、2次元グラフなどの多次元表現メディア上で可視化・表示するものである。本システムは、本研究において実現した基礎技術・機能の(1)~(8)全てを用いて構築されている。本システムの構成図を図3に、ユーザ・インタフェースを図4に、分析結果・出力例を図5、図6に示す。図5は、意味的コンテキストを固定し、関連するドキュメントの地理的・空間的な広がりや時系列変化を示した結果であり、図6は、時間軸を固定し、意味的コンテキストを変化させることによるドキュメントの地理的・空間的分布の相違を示したものである。本システムの特徴は、多面的な内容を含む歴史的・文化的文書データを対象として、利用者の興味・関心・視点と時空間情報の組み合わせに応じて、動的に、多種多様な歴史的・文化的ストーリーを生成可能とする点にある。

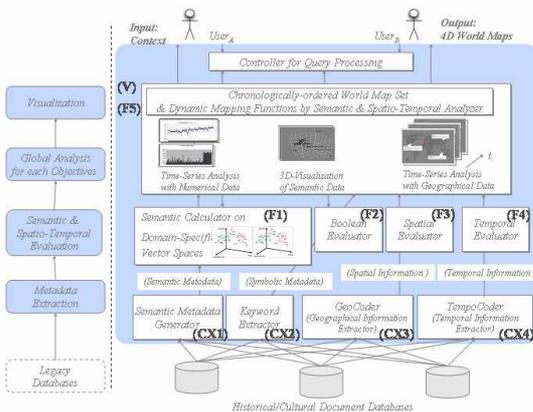


図3：4次元世界地図描画システム(4D World Map System)のシステム構成図

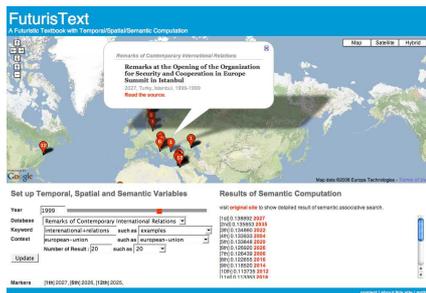


図4：4次元世界地図描画システム(4D World Map System)のユーザ・インタフェース

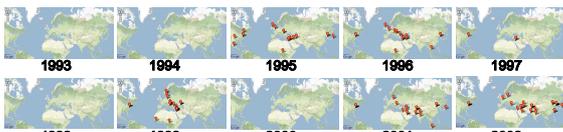


図5：4次元世界地図描画システム(4D World Map System)の出力例1



図6：4次元世界地図描画システム(4D World Map System)の出力例2

(S2) 歴史関連ドキュメントを対象とした多元的情報可視化システム：本システムは、歴史に関連するドキュメントデータを多元的情報表現を持つ媒体へと変換する歴史情報表現システムである。歴史関連情報を意味的な関連性計量を伴って時間的・空間的情報空間へ写像することにより、歴史的事象間の関連性の可視化を実現する点が特徴である。本研究における基礎技術・機能の(1)(4)(8)を用いて、歴史に関連する文書データから地理情報・年代情報を抽出し、文書データを時代別に分類すると共に、その分類情報を二次元マップ上に可視化している。システム構成図を図7に、出力例を図8に示す。

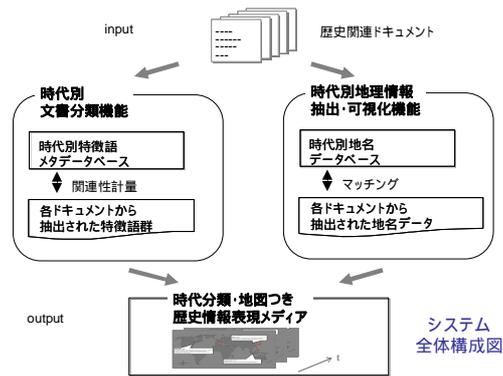


図7：歴史関連ドキュメントを対象とした多元的情報可視化システムの構成図

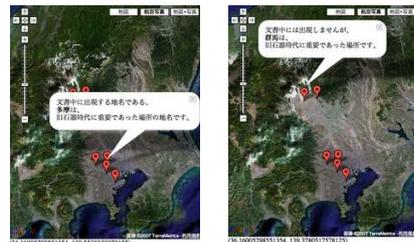


図8：歴史関連ドキュメントを対象とした多元的情報可視化システムの出力例

(S3) 化学プラント事象を対象としたドキュメントデータ因果関係計量システム：本システムは、化学プラントにおいて突発的に発生するトラブルに対して解決策に関連のあるドキュメントデータを提示するための、意味的連想検索を用いた因果関係計量システムである。この空間を用いることにより、利用者は、発生するトラブルと過去の原因・

結果関係の履歴との関係性を計量し、原因箇所を特定し、その解決策に関連のある再利用可能なドキュメント群を自動的に検索することが可能となる。本研究における基礎技術・機能の(2)(3)を用いることにより、ドキュメントデータから原因事象・結果事象に関する特徴語をメタデータとして自動的に抽出し、事象とドキュメント間の直接的・間接的な原因・結果の関連性計量を行っている。図9にシステム構成図、図10に実験結果例を示す。

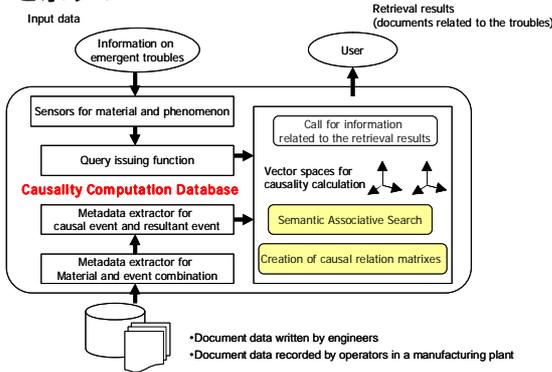


図9: 化学プラント事象を対象としたドキュメントデータ因果関係計量システムの構成図

Retrieval result of "causal event" of "c6"				Retrieval result of "causal event" of "c6" with query expansion			
Ranking	event ID	correlation	judgment	Ranking	event ID	correlation	judgment
1	c3	0.741724	t1	1	c10	0.691142	t2
2	c8	0.492359	t1	2	c7	0.495663	t2
3	c7	0.407728	t2	3	c1	0.318513	t2
4	c6	0.346576	t0	4	c24	0.243730	f
5	c12	0.285119	f	5	c3	0.238896	t1
6	c19	0.254898	f	6	c13	0.207097	f
7	c4	0.219637	f	7	c8	0.201136	t1
8	c1	0.218680	t2	8	c18	0.185162	f

Compared to the result without query expansion, indirect causal events c7 and c1 jumped up in the ranking, and a direct causal event c3 and c8 was ranked down with query expansion.

図10: 化学プラント事象を対象としたドキュメントデータ因果関係計量システムの実験結果例

(S4) 地域情報関連テキストを対象とした地域状況表示地図の動的生成方式: 本システムは、住所情報を含む地域情報関連テキストを対象として、ユーザの与えるコンテキストに応じた「地域状況表示地図」を動的に生成するシステムである。本システムは、コミュニティベースで蓄積された住所情報を含む文書の集合を、ユーザの与える任意のコンテキスト(任意のキーワードセット)によって評価し、その結果を画像として地図上に重ね合わせることにより、ある場所とその周囲環境の空間的配置に関してコミュニティが持つイメージ・印象・雰囲気、地域状況表示地図として可視化する。図11にシステム構成図、図12、図13に出力例を示す。

さらに、本システムの適用範囲をドキュメントデータから画像・動画などのマルチメディアデータに拡張するための基盤技術・機能の開発を行った。(1)画像データから文化別・

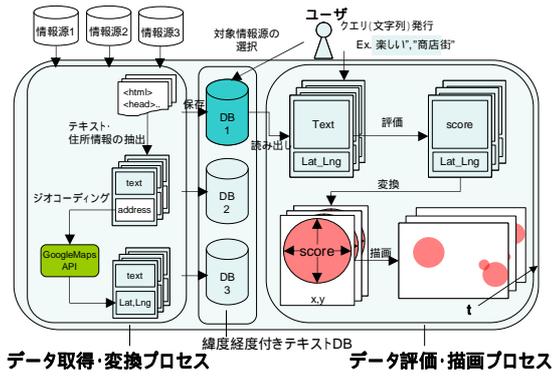


図11: 地域情報関連テキストを対象とした地域状況表示地図の動的生成システムの構成図



図12: 地域情報関連テキストを対象とした地域状況表示地図の動的生成システムの出力例1

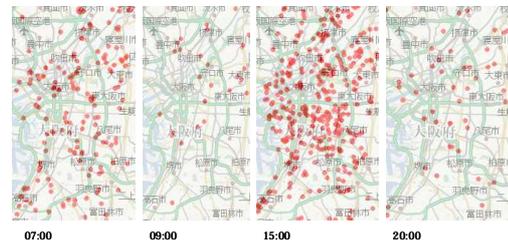


図13: 地域情報関連テキストを対象とした地域状況表示地図の動的生成システムの出力例2

時代別・作者別の特徴を抽出する機能、(2)音楽データから文化別・時代別・作者別の特徴を抽出する機能、(3)メディアデータを用いてユーザの意図を表現するクエリの生成機能、(4)メディアデータをクロス・カルチャルな意味空間へマッピングする機能。

今後は、これら機能をシームレスに統合し、歴史・文化に関する人文社会科学分野のマルチメディアデータ群を対象として、データ間の意味的・時空間的関連性を動的に計量・分析し、その分析結果を時間軸を伴った三次元地図(4D世界地図)上に可視化する情報分析システムを実現する。特に、歴史学、政治学、社会学、比較文化学といった学問領域においては、データ集合の共時的分析(データ内容の空間的な広がりに関する分析)や通時

的分析(時間経緯に伴うデータ内容の変化の分析)が重要であり,データ間の意味的な関連性や時空間的関連性を分析・可視化するシステムは,これらの分析や分析結果の直観的理解を支援するものと考えられる.

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計5件)

Sasaki, S., Takahashi, Y. and Kiyoki, Y., "The 4D World Map System with Semantic and Spatio-temporal Analyzers," Information Modelling and Knowledge Bases, Vol. XXI, (to appear), May 2010, 査読有

Shiori Sasaki, Yoshiko Itabashi, Yasushi Kiyoki and Xing Chen, "An Image-Query Creation Method for Representing Impression by Color-based Combination of Multiple Images," Information Modelling and Knowledge Bases, Vol. XX, pp.105-112, 2009, 査読有

佐々木 史織, テキストマイニングによる国際政策分析・情報可視化システムの構築, 法学研究, 83巻3号, 20 pages, 2009年, 査読無(招待論文)

Shiori Sasaki, Masato Yokoyama and Yasushi Kiyoki, "Causality Computation Database Creation with a Semantic Associative Search Method Applied to Emergent Troubles in a Chemical Plant," Proc. of The 3rd Information and Communication Technology Seminar (ICTS) 2007, pp.241-248, 2007, 査読有

Shiori Sasaki, Yasushi Kiyoki and Hiroyasu Akutsu, "An application of Semantic Information Retrieval System for International Relations", Information Modelling and Knowledge Bases, Vol. XIX, pp.62-79, 2007, 査読有

[学会発表](計11件)

Sasaki, S., Takahashi, Y. and Kiyoki, Y., "The 4D World Map System with Semantic and Spatio-temporal Analyzers," the 19th European-Japanese Conference on Information Modelling and Knowledge Bases, pp.11-24, June 1-5th, Maribor, Slovenia, 2009, 査読有

Heimbu"rger, A., Jaakkola, H., Sasaki, S., Yoshida, N. and Kiyoki Y., "Context-Based Knowledge Creation and Sharing in Cross-Cultural Collaborative Communities," the 19th European-Japanese Conference on Information Modelling and Knowledge Bases, Panel Session, June 1-5th, Maribor, Slovenia, 2009, 査読無

桐本篤, 佐々木史織, 清木康: 音楽と感性語の相関量計量による環境状況コンテキスト対応型音楽推薦システムの実現, 第1回データ工学と情報マネジメントに関するフォーラム (DEIM2009) 論文集, (accepted 8 pages), Mar. 8th-10th, 掛川, 2009, 査読無

守屋敬太, 佐々木史織, 清木康: 地域情報関連テキストを対象とした地域状況表示地図の動的生成方式, 第1回データ工学と情報マネジメントに関するフォーラム (DEIM2009) 論文集, (accepted 8 pages), Mar. 8th-10th, 掛川, 2009, 査読無

Sasaki S., and Kiyoki Y., "A Media-Query Creation Method for Representing Distinctive Impressions by the Combinations of Multiple Media Data Sets," the Institute of Electronics Information and Communication Engineers (IEICE) Technical Report, DE2008-32 (2008-09), pp.5-6, Sep. 21th-23th, 福島, 2008, 査読無

桐本 篤, 佐々木 史織, 清木 康: 風景画像とサンプル楽曲を用いた環境状況コンテキスト対応型音楽推薦システムの実現, 情報処理学会研究報告, 2008-DBS-146, pp. 157-158, Sep. 21th-23th, 福島, 2008, 査読無

高橋雄介, 佐々木 史織, 清木 康: 意味的時空間的分析機構を有する4次元世界地図描画システム, 信学技報, pp. 27-28, Sep. 21th-23th, 福島, 2008, 査読無

Shiori Sasaki, Yoshiko Itabashi, Yasushi Kiyoki and Xing Chen, "An Image-Query Creation Method for Representing Impression by Color-based Combination of Multiple Images," Proceedings of the 18th European-Japanese Conference on Information Modelling and Knowledge Bases (EJC2008), pp. 105-112, Tsukuba, Japan, June 2nd-6th, 2008, 査読有

Sasaki, S., Takahashi, Y. and Kiyoki, Y., "FUTURISTEXT: THE 4D WORLD MAP SYSTEM WITH SEMANTIC, TEMPORAL AND SPATIAL ANALYZERS," Proceedings of IADIS e-Society 2008 Conference, Algarve, Portugal, April 12-16th, 2008, 査読有

桐本 篤, 佐々木 史織, 清木 康: 風景画像を用いた環境状況コンテキスト対応型音楽推薦システムの実現, 第19回データ工学ワークショップ (DEWS2008) 論文集, (accepted 8 pages), Mar. 9th-11th, 宮崎, 2008, 査読無

井上勝哉, 佐々木史織, 清木 康: 歴史関連ドキュメントを対象とした多面的情報可視化システムの実現, 情報処理学会研究報告 Vol.2007, No.65, (2007-DBS-143), pp. 321-326, 夏のデータベースワークショップ 2007 (DBWS 2007), July 2nd-4th, 仙台, 2007, 査読無

[その他]

ホームページ等

<http://www.mdbl.sfc.keio.ac.jp/~sashiori/sashiori.html>

6. 研究組織

(1) 研究代表者

佐々木 史織 (SASAKI SHIORI)

慶應義塾大学・大学院政策・メディア研究科・助教

研究者番号: 90383934

(2) 研究分担者

なし

研究者番号:

(3) 連携研究者

なし

研究者番号: