

科学研究費助成事業 研究成果報告書

平成 26 年 5 月 7 日現在

機関番号：82624

研究種目：基盤研究(C)

研究期間：2010～2013

課題番号：22500875

研究課題名(和文) 萌芽的研究分野の探索手法に関する研究

研究課題名(英文) A Study on Searching Method for Emerging Research Field

研究代表者

古川 貴雄 (Furukawa, Takao)

文部科学省科学技術・学術政策研究所・科学技術動向研究センター・上席研究官

研究者番号：70262699

交付決定額(研究期間全体)：(直接経費) 3,100,000円、(間接経費) 930,000円

研究成果の概要(和文)：急速に発展する研究領域の革新的技術を検出、識別、分析するために、論文とカンファレンスセッションを分析して研究トピックの時間的な変化を示す手法を提案する。ここでは、2002年から2011年に開催されたWorld Wide Web (WWW) カンファレンスセッションの推移を示し時系列ネットワークから当該研究領域の発展を可視化した。Web関連技術の発展過程において、過去の研究トピックを総括する収束セッションノードと新たな研究トピックに分かれる分岐セッションノードにより、研究トピック間の相互作用が示され、当該研究の発展の過程で重要な役割を果たしたことが示唆された

研究成果の概要(英文)：In this report, we propose a method to analyse chronological changes in research topics as observed from papers and conference sessions, which enables us to detect, identify, and analyse the evolutionary process of emerging technologies in the many rapidly growing research fields. Chronological networks of conference sessions from World Wide Web conferences between 2002 and 2011 are used to visualize the scientific and technical streams formed by the session sequences. In the evolutionary process of emerging web-based technology, the convergent session nodes that recapitulate past research topics and the divergent session nodes in the networks play significant roles in promoting interactions among research topics.

研究分野：総合領域

科研費の分科・細目：科学教育・教育工学・科学教育

キーワード：革新的技術 発展過程 動向分析 学会 ネットワーク分析 計量書誌学 テキストマイニング

1. 研究開始当初の背景

革新的技術を形成する科学に基づくイノベーションは、新しい産業を創出し、既存産業を転換する可能性を秘めている。革新的技術は、既存技術の改善の積み重ねとは根本的に異なり、技術的な不連続な変化をもたらすとともに、従来の産業、市場、企業に対して破壊的なインパクトを与えている。その点で、革新的技術への戦略的な研究開発(R&D)投資は、産業競争力を効果的に高めることに寄与することから、企業の経営層をはじめとする民間セクターだけでなく、政策立案者や行政からも注目を集めている。

様々な科学技術のトピックとその応用について構造的な関係を記述する科学技術ロードマップは、公的・民間セクターの両方で戦略的研究開発の立案に用いられてきた。ロードマップを作成するロードマッピングの過程では、専門家の集団における新たな集合知を創出する手法と、コンピュータを用いた学術文献の自動分析によって抽出した知識を活用する手法が組み合わさることが一般化している。今後も、ロードマッピングにおいて専門家の知識は不可欠であることに変わりはないが、学術文献の爆発的な増大、計算機科学の進展やコンピュータハードウェアの性能向上を考慮すると、情報処理技術を利用した分析手法の重要性がより高まると考えられる。

2. 研究の目的

本研究では、情報処理技術を利用したロードマッピングの有効性と効率性の改善を目的とし、急速に発展する研究領域における革新的技術を検出、識別し、その発展を過程を分析する手法を提案する。革新的技術は何らかの破壊的なインパクトをもたらしていることに間違いはなく、それによって我々の生活は劇的に変化している。情報通信技術(ICT)の進歩は代表的な事例であり、その中でも特にインターネット、Web、モバイル関連技術は、パーソナルコミュニケーションだけでなく商業や製造業におけるビジネスコミュニケーションの形態も大きく変化させている。ここでは、影響が大きく、かつ、広範にわたるWeb関連技術に注目し、インパクトの大きな最先端研究の論文が発表されるトップランクの国際学会を取り上げ、研究トピックの変遷から当該領域における革新的技術を検出、識別し、その発展過程を分析する。

3. 研究の方法

(1) 分析データ

Web 関連技術は、この 20 年間に急速に発展した革新的技術として注目を集めている。本研究では、Web 関連技術に焦点を当て、当該研究領域においてトップランクに評価されている WWW カンファレンスを取り上げ、革新的技術の発展過程を分析する。WWW カンファレンスの場合、セッション名は Web 上で公開されているカンファレンスプログラムに記載されている。なお、WWW カンファレンスの情報を調査した段階では、カンファレンスのセッション名は IEEE、ACM

などの学術団体による文献データベース、Scopus や Web of Science のような商業学術出版社のデータベースのいずれにも収録されていなかった。2002 年から 2011 年の間に開催された WWW カンファレンスのプログラムや文献データベースを調べ、894 件のプロシーディングペーパーと 295 件のセッションに関する情報を得た。論文数は 75 件から 115 件の間で多少は変動しているが、セッション数はこの期間を通じて大きな変化は見られない(図 1)。セッション数は平均で約 30 件あり、1 つのセッションに平均で 3 件の論文が含まれることになる。カンファレンスには、ペーパーセッションやレギュラーセッションと呼ばれるセッションに加えて、ポスターセッション等も併設されているが、今回の分析では、セッション名との関係が明示されているペーパーセッションに限定して分析した。

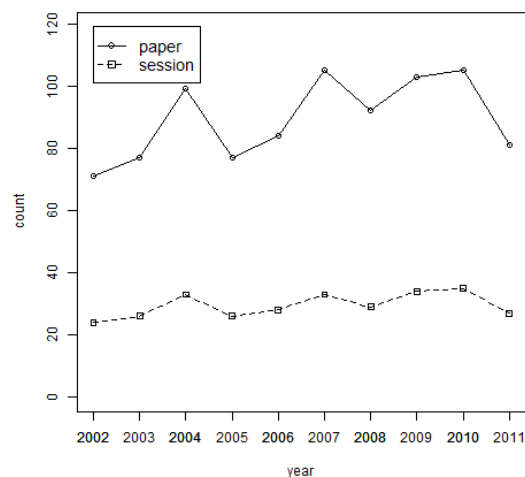


図 1 2002 年から 2011 年に開催された WWW カンファレンスにおける論文数とセッション数の関係

(2) 論文・セッションの類似度

論文は題目、著者、著者の所属、キーワード、引用文献などの属性をもつが、ここでは分析を簡略化するためアブストラクトのテキストデータのみを用いた。まず、論文の内容を要約したアブストラクトの文書データは、term frequency-inverse document frequency (*tf-idf*)の値を要素とするベクトルとして記述する。*tf-idf* は、簡単な単語の出現頻度よりも、特定の文書データに含まれる単語の重要性を強調した指標である。現在では、*tf-idf* 以外にも文書データをベクトルとして記述する手法が提案されている。例えば、LSA や LDA といった手法では、単語の集合から潜在的な意味情報を抽出することに焦点を当てている。これらの手法では、潜在的な意味情報を抽出するために、高次元空間の文書ベクトルを低次元空間に投影する操作が行われているが、投影される空間の次元など未知の定数を事前に決定しておく必要がある。このような未知の定数を決定する方法により結果が変化するという不安定性もあるため、ここでは、文書ベクトルとして *tf-idf* を用いることにした。

以下に論文間類似度の定義を示し、論文間類似度に基づくセッション間類似度の計算方法

を示す。

論文間類似度

tf-idf ベクトルによって記述された論文 i と論文 j をそれぞれ、ベクトル x_i と x_j と表記し、これらの論文間類似度を次のように定義する。

$$s_{ij} = x_i \cdot x_j / \|x_i\| \|x_j\| \quad (1)$$

tf-idf は非負の値をとることから論文間類似度 s_{ij} の範囲は 0 から 1 になる。論文間類似度 s_{ij} が 1 に等しい場合、論文間の *tf-idf* ベクトルの比率が一致し、論文間類似度 s_{ij} が 0 の場合、2 つの論文間には共通する単語が存在しないことになる。

セッション間類似度

セッション間類似度は、セッションに含まれるすべての論文ペアについて求めた論文間類似度の平均値と定義する。次の式でセッション I とセッション J 間の類似度を定義する。

$$S_{I,J} = \frac{\sum \sum s_{ij}}{N_I N_J} \quad (2)$$

ここで、 N_I と N_J はセッション I とセッション J に含まれる論文数を示す。

(3) 時系列ネットワーク生成アルゴリズム

図 2 に示すアルゴリズムによりカンファレンスセッションの時系列ネットワークを生成する。各セッションは木構造ネットワークを構成するノードに対応するため、2 つのセッションノードを接続するエッジの挿入を繰り返すことでカンファレンスセッションの時系列ネットワークが生成される。

- (1) 基準年からルートノードとなるセッションを選択する。基準年以外の全セッションノードが接続されるセッションノードの候補となる(図 2 (a))。
- (2) 各セッション候補について、ルートセッションとの類似度を計算する(図 2(b))。
- (3) セッションペアの類似度が設定値よりも大きい場合、セッションノード間を接続するエッジを挿入する。接続されたセッションは候補セッションノードから除く(図 2(c))。
- (4) 新たに接続されたリーフノードを選択し、リーフノードが含まれる年のセッションを候補セッションノードから除く。
- (5) 各候補セッションノードについて、リーフセッションノードとの類似度を計算する(図 2 (d))。
- (6) ステップ(5)で計算したセッション間類似度が設定値よりも大きい場合には、これらのセッションノードを接続するエッジを挿入する(図 2 (e))。
- (7) 全セッションのペアについて接続が確認されるまでステップ(4)に戻って処理を続ける。

基準年が調査期間の途中であった場合、上記の処理は前述の後方処理と、時間を反転した前進処理を組み合わせ、セッション間を接続す

る時系列ネットワークを生成する。

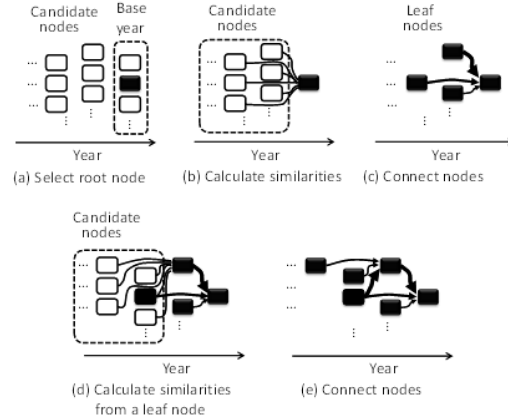


図 2 カンファレンスセッションの時系列ネットワークを生成するアルゴリズム

4. 研究成果

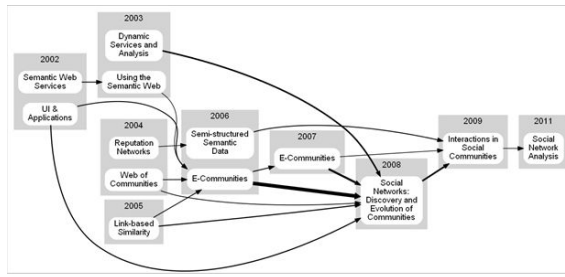
カンファレンスセッションの時系列ネットワークを生成するために、2002 年から 2011 年に開催された WWW カンファレンスのペーパーセッションで発表されたすべての論文からアブストラクトを抽出して分析した。本研究では、WWW カンファレンスを代表するような研究の発展過程を分析するために、次のセッションを時系列ネットワークのベースノードとして選択した。

(1) Social Network Analysis (WWW 2011)

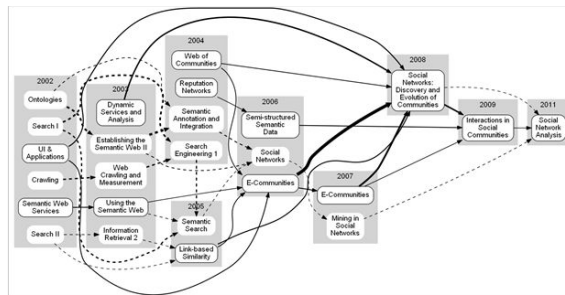
2011 年に開催された WWW カンファレンスの Social Network Analysis というセッションに注目し、2002 年から 2011 年の 10 年間に開催された WWW カンファレンスセッションとの関係を時系列ネットワークに示した結果を図 3 に示す。図 3(a)と図 3 (b)は、セッション間の接続を判定するためのセッション間類似度のしきい値を高く設定した結果と低く設定した結果を示している。図 3(b)の実線で囲まれたセッションノードと実線の矢印で表示された部分は図 3 (a)の時系列ネットワークに一致するため、図 3(a)は図 3(b)の部分集合となっている。一方、図 3(b)の実線で囲まれていないセッションノードと点線で示された矢印は、図 3(a)に対する付加的部分ネットワークを示している。また、セッション間類似度の大きさを矢印の太さに反映しているため、太い矢印で接続された 2 つのセッションは類似度が大きいことを示している。

図 3 の右側に表示された 2011 年の Social Network Analysis というセッションから左側に遡って関連するセッションを辿ると、2009 年の Interactions in Social Communications を経由して、2008 年の Social Networks: Discovery and Evolution of Communities (以下、DEC と省略して表記する。)に辿りつく。図 3(a), (b)のいずれも、2008 年の Social Networks: DEC というセッションには過去の 6 つのセッションから接続されている。さらに、2008 年の Social Networks: DEC から遡ると、2003 年の Dynamic Services and Analysis、2006 年、2007 年の E-Communities に到達する。この時系列ネットワークの構造を見ると、2008 年

の Social Networks: DEC というセッションは収束セッションノードとみなすことができ、このセッションにより過去には分かれていた研究トピックが統合されたことを示唆している。同様に、2009 年の Interactions in Social Communications や 2006 年の E-Communities も接続されるセッションが比較的多いことから、これらも収束セッションノードとみなすことができる。これらのセッションは、過去の研究トピックを統合し、その後の Social Networks 研究の発展に寄与した重要なセッションであると考えられる。



(a) High similarity threshold



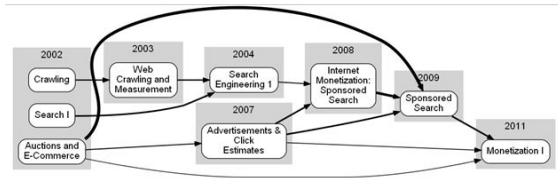
(b) Low similarity threshold

図 3 2011 年の Social Network Analysis セッションに至る時系列ネットワーク。セッション間類似度の大きさを矢印の太さに反映させている。実線で囲まれたセッションノードと実線で示された矢印は、上段の時系列ネットワークに一致する。

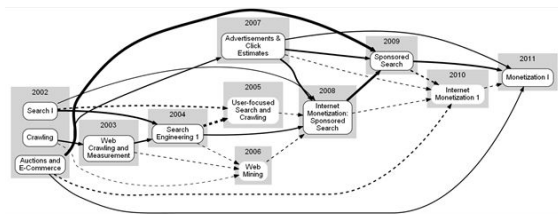
図 3 において、2006 年の E-Communities と 2008 年の Social Networks: DEC を接続する太い矢印は、このネットワークの中で、これらのセッション間の類似度が高いことを示している。また、2006 年の E-Communities から 2007 年の E-Communities を経由した 2008 年の Social Networks: DEC に至る間接的な経路も存在しているが、これらのセッション間を接続する矢印は太くないことからセッション間類似度もそれ程高くはないことがわかる。2006 年と 2007 年の E-Communities というセッションは同一の名称でありながら、セッション間を接続する矢印が太くはないため、セッション間類似度も高くない。この結果は、2008 年の Social Networks: DEC セッションに含まれる研究は、主に 2006 年の E-Communities セッションに含まれた研究から派生した内容であり、2007 年の E-Communities というセッションに含まれた研究との関連性はあまり高くはないことを示唆している。さらに、2004 年の Web of Communities を起点とする経路は、Social Networks 研究が community の研究から発展したことを示唆している。

(2) Monetization I (WWW 2011)

図 4 に、2011 年の Monetization I セッションから遡って生成したカンファレンスセッションの時系列ネットワークを示す。図 4(a)、(b)はセッション間の接続を判定するセッション間類似度のしきい値を変化させて生成した結果である。学術研究の文脈で考えると、Monetization というトピックは一般的な研究とは考えにくい。そのため、当該領域の研究者でなければ、WWW カンファレンスで Monetization 研究が扱われるようになった背景を理解するのは容易ではない。2002 年から 2011 年の全セッション名を確認すると、monetization という単語は 2008 年の Internet Monetization: Sponsored Search で初めて現われる。このセッションノードは、2002 年の Auctions and E-Commerce、2007 年の Advertisements & Click Estimates、2009 年の Sponsored Search に接続されている。これらのセッション間の関係から、インターネットオークションや広告技術を含む電子商取引等に関する研究が、Monetization 研究を形成したことが読み取れる。



(a) High similarity threshold



(b) Low similarity threshold

図 4 2011 年の Monetization I セッションに至る時系列ネットワーク。セッション間類似度の大きさを矢印の太さに反映させている。実線で囲まれたセッションノードと実線で示された矢印は、上段の時系列ネットワークに一致する。

図 4(a) では、2002 年の Auctions and E-Commerce と 2007 年の Advertisements & Click Estimates から出た 3 本の矢印が他のセッションノードに接続され、図 4(b)では 4 本の矢印が他のセッションノードに接続されている。これらの分岐セッションノードは、その後の Monetization 研究に大きな影響を与えたセッションとみなすことができる。時系列ネットワークの詳細な構造を示す図 4(b)を見ると、2002 年の Search I と 2004 年の Search Engineering 1 から、それぞれ 3 本の矢印が出て、他のセッションノードに接続されている。これらのセッションも 2002 年の Auctions and E-Commerce と同様に、その後の研究に大きな影響を与えたセッションとみなすことができる。図 4 (b)を見ると、2002 年の Auctions and E-Commerce は、2007 年から 2011 年までのセッションノードに接続されている。これ

らのセッション間の関係をみると 2002 年からは最短でも 5 年以上の間隔がある。特に、2002 年の Auctions and E-Commerce と 2009 年の Sponsored search の間には 7 年の間隔があるにもかかわらず、太い矢印で接続されたセッション間類似度は高い関係であることがわかる。このような結果は、2002 年の Auctions and E-Commerce に関連する研究が、数年の間隔を置いて再び盛んになったことを示唆している。

図 4(a)を見ると、2009 年の Sponsored Search と 2011 年の Monetization I は、それぞれ 3 つの過去のセッションノードから接続されていることから、収束セッションノードとみなすことができる。図 4(b) に示された 2008 年の Internet Monetization: Sponsored Search と 2010 年の Internet Monetization 1 も複数の過去のセッションノードに接続されていることから、過去の研究を統合した収束セッションノードと考えることができる。

(3) Semantic Analysis (WWW 2011)

図 5 に、2011 年の Semantic Analysis セッションから遡って生成したカンファレンスセッションの時系列ネットワークを示す。図 5 (a)、(b) はセッション間の接続を判定するセッション間類似度のしきい値を変化させて生成した結果である。Semantic Analysis は、前述した Social Networks や Monetization 研究と比較して、この研究領域では比較的歴史の長い伝統的な研究テーマと言える。

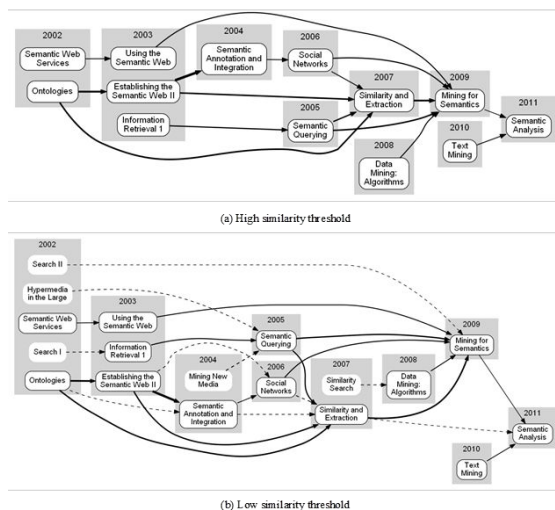


図 5 2011 年の Semantic Analysis セッションに至る時系列ネットワーク。セッション間類似度の大きさを矢印の太さに反映させている。実線で囲まれたセッションノードと実線で示された矢印は、上段の時系列ネットワークに一致する。

図 5 (a)を見ると、2000 年代初めの 2002 年に頭に始めに Semantic Web や Ontologies を含むセッションが登場している。2007 年の Similarity and Extraction や 2009 年の Mining for Semantics は、過去に実施された複数のセッションに接続されていることから、収束セッションノードとみなすことができる。そのため、2007 年の

Similarity and Extraction において過去の Semantic Web 技術が統合され、2009 年の Mining for Semantics で Semantic Web と Data Mining 技術が統合されたと推察される。図 5(b) の点線で示した矢印は、図 5(a) で示されなかった潜在的なセッション間の関係を示している。結果として、これらの時系列ネットワークは、Semantic Analysis に関する研究は Data Mining 技術を取り込みながら発展したことを示唆している。

5. むすび

公的、あるいは、民間の科学技術投資の方向性を議論する際に、当該領域の専門家だけでなく非専門家を含めたステークホルダーの間で、先端研究の動向に関する共通認識をもつことは重要である。しかし、急速に発展している研究領域の動向については、非専門家が研究の動向を正確に把握することは極めて困難である。本研究では、ロードマッピングへの応用を想定し、革新的技術を検出、識別し、その発展過程を分析するための手法を提案した。具体的には、カンファレンスのセッションとプロシーディングペーパーを用いて時系列ネットワークを生成する手法について述べ、2002 年から 2011 年に開催された WWW カンファレンスについてセッションをノードとする時系列ネットワークを生成し、研究トピックの変遷から最先端研究の動向を可視化した。

分析対象の範囲、データ形式、データ間の接続関係、分析結果の安定性の観点から、計量書誌学的手法とテキスト分析手法を比較した結果についても述べた。このような比較から、学術文献を用いて専門家にも認識されていないような知識の抽出を目的とした場合、不明瞭なデータを扱うことのできるテキスト分析が適することを示した。さらに、カンファレンスセッションの時系列ネットワーク分析から得られた特徴的なセッションから、研究トピック間の相互作用が研究の発展過程において重要な役割を果たすことが示唆された。例えば、以下のセッションによって研究の発展過程が特徴づけられた。

- 過去のセッションとの接続が多い収束セッションは、過去の研究トピックを統合したと考えられる。
- その後のセッションとの接続が多い分岐セッションは、他の研究に大きな影響を与えたセッションと考えられる。

提案手法の特徴は、研究者コミュニティにおける将来展望、新たな研究領域を開拓しようとする意思が反映されたカンファレンスセッションに注目して分析した点にある。基本的に、カンファレンスセッション名は、特定の研究を代表する概念に相当するため、共引用関係を用いた学術文献のクラスタリングなどのボトムアップ手法で必要とされるクラスタの解釈を行う必要がない点も提案手法の利点と言える。また、医学、物理学、化学領域の代表的な学会の年次大会について調査したところ、名称の付与されたセッションが行われていることから提案手法が他の研究領域

にも適用できる可能性が示された。提案手法には、いくつかの課題が残されているものの、革新的技術の検出、識別とその発展過程の分析手法として有用であると考えられる。

6. 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 4 件)

- [1] Takao Furukawa, Kaoru Mori, Kazuma Arino, Kazuhiro Hayashi, Nobuyuki Shirakawa, Identifying the evolutionary process of emerging technologies: A chronological network analysis of World Wide Web conference sessions, Technol. Forecast. & Soc. Chang., 査読有 (In Press)
<http://dx.doi.org/10.1016/j.techfore.2014.03.013>
- [2] Takao Furukawa, Nobuyuki Shirakawa, Kumi Okuwada, An empirical study of graduate student mobility underpinning research universities, Higher Education, 査読有, Vol. 66, No. 1, 2013, 17-37.
- [3] Takao Furukawa, Nobuyuki Shirakawa, Kumi Okuwada, Kazuya Sasaki, International mobility of researchers in robotics, computer vision and electron devices: A quantitative and comparative analysis, Scientometrics, 査読有, Vol. 91, No. 1, 2012, 185-202.
- [4] Takao Furukawa, Nobuyuki Shirakawa, Kumi Okuwada, Quantitative analysis of collaborative and mobility networks, Scientometrics, 査読有, Vol. 87, No. 3, 2011, 451-466.

(学会発表) (計 4 件)

- [1] 森 薫, 有野 和真, 古川 貴雄, 情報科学の国際学会セッションの時系列ネットワーク分析, 研究技術・計画学会 第 27 回年次学術大会, Oct. 27, 2012, 一橋大学 (東京都)
- [2] Kazuma Arino, Takao Furukawa, Nobuyuki Shirakawa, Kumi Okuwada, Temporal Network Analysis of Emerging Technologies: Topic Transition in World Wide Web (WWW) Conference, Proceedings of IEEM 2012 International Conference on Industrial Engineering and Engineering Management, Dec. 12, 2012, Hong Kong (China)
- [3] Nobuyuki Shirakawa, Takao Furukawa, Masatoshi Tamamura, Global Convergence

in Electric, Electronic and Information Communication Technology Research, The 5th ISPIM Innovation Symposium, Dec. 12, 2012, Seoul (South Korea)

- [4] Takao Furukawa, Nobuyuki Shirakawa, Kumi Okuwada, Kazuya Sasaki, Quantitative Analysis of International Mobility of Robotics Researchers and Characteristics of Domestic Robotics Research, 2011 IEEE Int. Conf. on Industrial Engineering and Engineering Management, Dec. 8, 2011, Singapore (Singapore)

(図書) (計 1 件)

- [1] Nobuyuki Shirakawa, Takao Furukawa, Minoru Nomura, Kumi Okuwada, Global Diffusion of Electrical and Electronics Engineering Research - An Extensive Quantitative Analysis of IEEE Publication (1980-2008) in Paul Wesling et al. (ed.) Advances in Technology and Innovation Management Volume I - 2012, 169-188, IEEE, 2012.

(産業財産権)

○出願状況 (計 0 件)

○取得状況 (計 0 件)

(その他)

6. 研究組織

(1) 研究代表者

古川 貴雄 (FURUKAWA, Takao)
文部科学省科学技術・学術政策研究所・科学技術動向研究センター・上席研究官
研究者番号: 70262699

(2) 研究分担者

白川 展之 (SHIRAKAWA, Nobuyuki)
文部科学省科学技術・学術政策研究所・科学技術動向研究センター・上席研究官
研究者番号: 20556071

佐々木 和也 (SASAKI, Kazuya)
宇都宮大学・教育学部・准教授
研究者番号: 60292570