慶應義塾大学学術情報リポジトリ
Keio Associated Repository of Academic resouces

Title	政策ドキュメントを対象とした分析・知識発見・可視化システムの実現
Sub Title	Advanced information analysis and knowledge discovery system for IR-related documents
Author	佐々木, 史織(Sasaki, Shiori)
Publisher	慶應義塾大学法学研究会
Publication year	2010
Jtitle	法學研究 : 法律・政治・社会 (Journal of law, politics, and sociology). Vol.83, No.3 (2010. 3) ,p.243- 279
JaLC DOI	
Abstract	
Notes	
Genre	Journal Article
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=AN00224504-20100328-0243

慶應義塾大学学術情報リポジトリ(KOARA)に掲載されているコンテンツの著作権は、それぞれの著作者、学会または出版社/発行者に帰属し、その権利は著作権法によって保護されています。引用にあたっては、著作権法を遵守してご利用ください。

The copyrights of content available on the KeiO Associated Repository of Academic resources (KOARA) belong to the respective authors, academic societies, or publishers/issuers, and these rights are protected by the Japanese Copyright Act. When quoting the content, please follow the Japanese copyright act.

政策ドキュメントを対象とした分析・知識発見・可視化システムの実現

佐々木史織

一　はじめに
二　国際政治学における政策分析の方法論
三　政策ドキュメント分析の方法論——情報科学の視点から——
四　国際関係ドキュメントを対象とした意味的・時空間的分析・可視化システムの実現
五　実ドキュメントデータを対象とした適用実験
六　まとめ

一　はじめに

本稿では、国際関係の政策関連ドキュメント群を対象として、ドキュメント間の意味的な関連性や時空間的分析を行い、その結果発見された知識を二次元グラフや時間軸を伴った地図上に可視化する情報分析・知識発見シ

ステムの基本設計と実現方式について述べる。本システムの特徴は、日々刻々と更新されるドキュメントデータを対象として、データ間の意味的・時空間的関連性を動的に計量・分析・可視化し、その大局的な概観をユーザ（分析者）に提示する点にある。ここで扱うドキュメントデータとは、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、特にテキストデータであり、日々広域ネットワーク上に公開される一般の国際ニュース記事、政府や国際機関の公式発表、プレス・ブリーフィング、政策ステートメント、政府高官の談話、議会議事録、NGOの活動記録などを指す。

薬師寺泰蔵はT・クーンとC・ポパーのパラダイム論争、および、G・アーモンドの「雲とからくり時計の論争」[5] を引用し、政策科学の対象は「とりとめもないぼんやりしたもの」[2] と表現している。政治現象は人間の行為・行動から成り立っており、その人間の行為・行動は決定論では予測できない不確実で不明瞭なものであること、さらに、人間の行為・行動は、具体的・物理的な強制力だけでなく、思想・主義・方針・意図・計画といった抽象的・非物理的な力（すなわち言葉の力）によって動かされることも多いことを指摘している。さらに、政策分析の対象は「雲が雲をコントロールするプラスチック・コントロールの世界」[2] であり、政策分析の目的は「雲のようなぼんやりとした政策が」「雲のような人間の心を動かし、最終的に物理的な行為を生み出す」過程の分析である [2] と述べている。

これらの指摘と前提に則ると、政策科学、特に国際関係や国際政治に関する政策分析を行うためには、「動き」と「言葉」は必要条件の視点であるといえる。動きとは、変化、変動、変遷、遷移、動態のことであり、分析対象はモノ・金・人・情報の流れなどの物理的移動から、人間の認知や認識・社会的価値・社会構造・制度など目に見えないものまで広範囲である。一方、言葉とは、語、概念、発話、記述、文章、記号、表象、シンボル、イメージ（色・形状・構図・テキスチャ）、非言語言語（動作・ジェスチャ・視線・音声）などを含む。人間の認知・

社会的価値・社会構造・制度など目に見えないものを分析するためには言葉の分析が不可欠である。

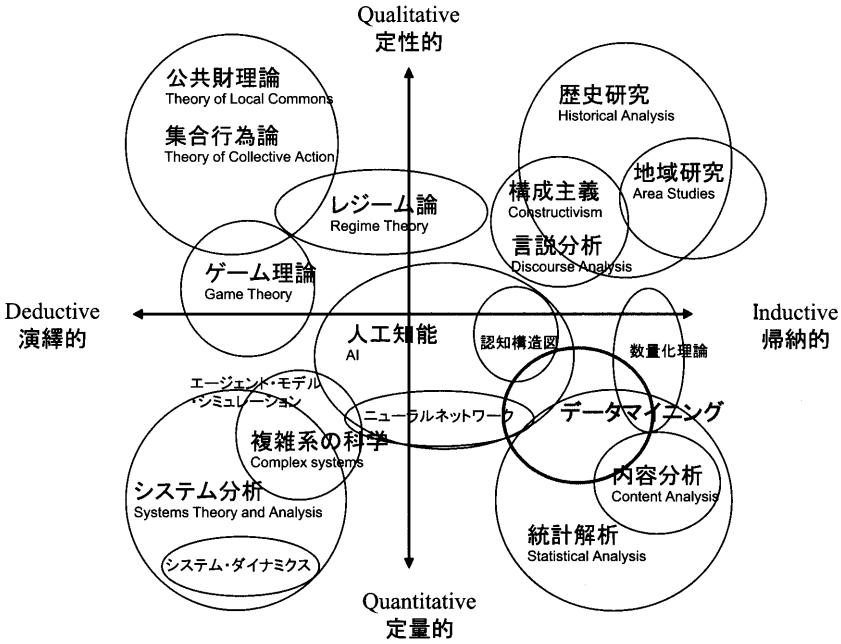
しかしながら、「目に見えないもの」について「動き」を観察・分析することは難しい。

現在、インターネットおよび広域ネットワーク上には膨大なメディアデータが存在し、その中には、国際関係に関するドキュメントデータも多い。これらのドキュメントは、WWW上で日々刻々と更新・配信され、デジタルライブラリやデジタルアーカイヴに大量に保存・蓄積されている。ポスト冷戦の相互依存の深化によってもたらされたグローバル化社会において、また、二〇〇〇年以降の情報技術の急速な発展によるグローバル・ネットワーク社会においては、国際関係論・国際政治学の研究者、政策形成過程や政策評価の研究者のみならず、政策策定実務者、さらには一般の生活者にとっても、これらのドキュメントデータからの迅速かつ適切な情報獲得は緊急かつ重要な課題の一つであるといってよい。

しかしながら、利用可能な情報やデータ量が爆発的に増加するにつれ、大量かつ多種多様なデータの中から有益な情報を的確に収集・獲得・管理し、適切な分析結果と知見を導き出すことは益々困難になっている。

本稿では、これらの問題意識に基づき、言葉の集合から「雲」の広がりと動きを描く方法論の一つとして、政策ドキュメントを対象とした意味的・時空間的分析機能を有する知識発見・可視化システムの実現方式について述べる。本システムの目的は、膨大かつ多様なドキュメントデータの中から何らかのパターンや知識を抽出し、空間的な分布と時系列的遷移に関する分析を行い、直観的理解が可能な形式で可視化することで、政策ドキュメント群全体の大局的把握を支援することにある。すなわち、「境界のはっきりしない」「ぼんやりとした」「刻々とその姿を変える」雲について、分析者がその起点の推測と結果の予測ができるよう、「おおまかな」流れと傾向と概観を描く試みの一つである。

図1　国際政治学における政策分析の方法論



二　国際政治学における政策分析の方法論

（一）定量的手法の位置づけ

国際政治学における政策分析の主な方法論について、概観を図1に示す。

図1の縦軸は、定量的手法であるか、定性的手法であるかの軸である。定量的分析とは、主に数値データおよび数値化された変数を用いた分析を指す。一方、定性的分析とは、歴史文書、外交文書、演説、書簡といった文字データや記述・テクストを用いた解釈学的分析を示している。なお、G・キング、R・コヘイン、S・ヴァーバは、定量的分析を「数字と統計手法を用いて一般論を抽出、または、因果に関する仮説を検証する手法」と位置づけ、定性的分析を「少数のケースに焦点を当て、インタビューや歴史的な洞察を行う手法」と位置づけている[6]。

図1の横軸は、帰納的方法であるか、演繹的方法であるかを示す軸である。定量的な帰納的分析とは、数値ないし言語データを統計手法によって整理・分析し、並べ替え、類型化、グラフ化によってある規則性・パターンを抽出する手法である。ここでは、分析者は準備仮説を立て、その仮説を裏付けるためのデータを収集・加工し、仮説の検証を行いながら分析結果と結論を導きだす。一方、演繹的分析とは、分析者があらかじめ既存の研究や理論、直観より得られたモデルに基づき、論理的に導かれた仮説の証明をモデルの変数操作や模擬実験（シミュレーション）等により行う手法である。なお、薬師寺泰蔵は一九七九年に、政策分析における統計手法を使用したモデルを「相関モデル」「傍観者モデル」「実証モデル」と分類し、シミュレーションを用いたモデルを「因果モデル」「実行者モデル」「機械論モデル」と分類している［7］。

（二）ドキュメント分析と定量的分析手法

国際政治学において帰納的定量分析として位置づけられる統計的手法は、一九五〇～一九六〇年代にかけて盛んに行われた。代表的研究としてP・スモーカーの軍事的相互依存に関するクラスター分析［8］、B・ラセットの国際システム内の地域特性についての因子分析［9］、D・シンガーらの戦争とパワー分布に関する回帰分析［10］、山本吉宣による戦争の確率統計モデル分析［11］、薬師寺泰蔵の逐次的回帰分析によるリチャードソン・モデルの再検討［12］などがある。

これら統計的手法のうち、新聞記事・演説・交換文書などの文字・音声データ（以下、「ドキュメント」という）を定量的に捉え、国際コミュニケーション・対外イメージ・メッセージ・シンボル・世論などを単語の出現頻度や言葉・概念の間の関係・構造について定量的に分析・理解する手法として「内容分析（Content Analysis）」「認知構造図（Cognitive Map）」が挙げられる。内容分析および認知構造図は共に、六〇年代～七〇年代にかけ

て国際関係論・国際政治学の分野に導入され、公表された文書の事後的分析を通じて政策決定者の認知、態度、あるいは対外イメージなどの分析に応用されてきた。特に、世論や政党の分析には、新聞や機関紙が対象ドキュメントとして用いられ、意見調査を行うことが困難な各国首脳の認知分析には、演説や声明、交換文書、書簡などが対象ドキュメントとして用いられてきた。

内容分析は、ドキュメント群中の単語またはコード化された文の出現頻度を計測し、各単語・コードと各ドキュメントとの相関度を計算し、多変量解析を行うものである。第二次世界大戦期のプロパガンダ分析によって大規模に実用化されたといわれ、戦後はH・ラズウェルによるシンボル分析［13］や、O・R・ホルスティらによるコミュニケーション分析［15］［16］によって国際政治学に応用されてきた。日本では、猪口孝による政策決定者間コミュニケーション分析［17］、武者小路公秀による歴代総理の帝国議会演説の価値内容分析［18］、高木誠一郎による人民日報の頻度分析［19］などの研究がある。

一方、認知構造図は、あるドキュメントの論理構造を概念（concept）のネットワークとみなし、各コンセプト間の因果関係を＋,－,0で表し論理演算を行うことでドキュメントの著者または発言者（政策決定者）の認知における論理回路を分析する手法である。一九五〇年代にE・トールマンによって提起された認知科学の一手法であり、人工知能や計算機による言語処理技術の発達に伴い、一九七〇年代には国際政治学の対外政策決定分析に応用された［20］。代表的な研究として、R・アクセルロッドの政策決定者認知分析［21］、山本吉宣の外交政策実務者認知分析［22］、C・ジョンソンらの冷戦下の外交交渉分析［23］がある。

これらのドキュメントを対象とした定量的分析手法は、近年のコンピュータによる情報処理技術や自然言語処理技術により、大量の非数値データの量的な分析と把握を可能とするものとなり、再評価の対象となりうるものである。従来の内容分析および認知科学的アプローチにおいては、手動（マニュアル）による分析が主流であっ

たため、データ処理の過程における恣意性（データ収集、コンセプトの抽出、ラベリング）の問題や、大量のドキュメントデータの収集やコード化が煩雑であること、コード化には特定問題領域に関する極めて専門的な知識が必要となること、信頼性を確保するために複数の人間によるチェックが必要となること等がボトルネックとなっていた。しかし、一九八〇年代後半からの広域ネットワーク（WWW）環境と電子情報処理技術の急速な発展により、近年では、これらの大量のドキュメントデータを対象とした自動的な定量的分析が可能となっている。

一方で、薬師寺泰蔵は、定性的・定量的、帰納的・演繹的といった従来型のオーソドックスな分類とは異なる国際政治学における新しい方法論として、「論理ストーリー主義」を提唱している［3］。この方法論は、自然科学分野から社会科学分野へ応用された従来の「論理実証主義」に対して、C・A・バースが演繹法、帰納法に次ぐ第三の方法として提唱したアブダクション（仮説的推論）に基づいている。薬師寺の主張によると、国際政治学または政治学は、「不可到達」かつ「不可観測」な世界を対象とするために、厳密な実験・観測が不可能であり、仮に、政治や社会をシステムとして捉えて実証分析を行う場合でも、システム各部分間の強い相互依存関係、分割可能性、可制御性について十分に注意を払わなくてはならない［2］［4］。ゆえに、国際政治や公共政策の分析には、統計量や「信用に値するストーリー」によって仮説を検定する論理ストーリー主義アプローチが有効である、という主張である［3］。

以下の章では、この論理ストーリー主義アプローチの一つの実践として、情報科学の観点からみた政策ドキュメント分析手法の可能性について述べる。

三　政策ドキュメント分析の方法論――情報科学の視点から――

情報科学（Information Science）とは、文章や数値など多様な形態で存在する多種多様な情報源から「情報」「知識」を選択し、これらを効率よく利用するためにコンピュータを利用して処理を行い、人間の情報処理能力を強化・支援する試みに関する学際的な科学である［27］。情報がコンピュータに記憶・計算可能な内部表現（記号列）に変換・形式化・蓄積されたものをデータと呼ぶ。データからどのような情報を抽出するか、また、情報からどのような知識を抽出するかは、利用者の意図・判断によって取捨選択される。したがって、事象や現象を客観的に記述した静的なものである「データ」から、ユーザの意図によって動的に意味が変化する「情報」を抽出し、さらに人間の知識構造を変化させるような「知識」を発見することが情報科学の目的であるといえる。

前章で述べたように、一九六〇年代より、社会学的観点から情報の生産・配布・収集・利用の過程を調査・統計分析により解明する研究が盛んになり、一九七〇年代には、プロトコル解析やアンケート調査などの行動科学の手法が研究されるようになった［27］。一九八〇年代には、これらの研究から得られた知見を基に数理的なモデルを設定し、目的とする知識形態に従った有用な知識を抽出する技術が、人工知能（ＡＩ）や認知科学の分野で発展する。これらのアプローチは「知識工学アプローチ」「知識発見（Knowledge Discovery）」「知識ベース（Knowledge Base）」と呼ばれ、ＡＩ、データ工学、計算理論、統計学、心理学、医学、経営学、社会分析など分野横断的・学際的な研究分野となっている［28］。

一方、データからの情報抽出、特に、人間による解釈・抽象化を介さずに可能な限り元の形でデータを扱い、そこから自動的に情報を抽出しようとするアプローチは、「情報検索的アプローチ」または「データマイニング」と呼ばれる［27］［28］［29］［32］。このアプローチでは、抽出・検索された情報の解釈はユーザに委ねられるた

め、ユーザに必要な情報を効率よく検索・提示できる方式、大規模なデータ群から「思いがけないパターン」[28] を発見するためのアルゴリズム、ユーザの役に立つ検索システムの設計・構築に主な関心が寄せられている。情報科学、とりわけデータ工学（Data Engineering）の分野において、情報検索とは、広義には「ユーザの持つ問題を解決できる情報をみつけだすこと」、狭義には「ユーザの検索質問（query）に適合する文書（document）を文書集合（document collection）の中からみつけだすこと」と定義される [27]。文書はテキストだけでなく、音声・画像・動画などのマルチメディア情報を含むこともある。

これら二つのアプローチは決して相反するものではなく、むしろ人文社会科学のような暗黙知的な専門知識による分析が重要な研究分野においては特に、実利用によって有用な情報・知識を獲得するためには、両者を不可分な関係で同時に利用する必要がある。知識工学アプローチにおいては、情報をコンピュータによる推論に直接利用できる程度まで形式化する必要があり、人間によって情報を体系化し、知識ベースを構築するにはコストが高いという問題がある。一方で、情報検索アプローチにおいても、適切な仮説を事前に持たずに適当にパターンを探索するのみでは、人間の状況判断や意思決定に有用な知識を発見・獲得するには至らない。

たとえば、国際政治分野で政策分析に利用される認知構造図手法において、自然言語処理によりドキュメントからコンセプトを自動抽出することは可能だが、正確なコンセプト間の「意味的な」関連性を記述する図を作成するためには、必ず人間によるチェックが必要となる。また、ドキュメント内の単語の出現頻度を計測し、統計処理によってパターンを抽出する内容分析においては、大量のデータを自動的に扱うために、機械的なパターンマッチングによる計測が行われるが、各単語の文脈に応じた多義性、すなわち、その単語がいかなる文脈において使用されているのかは反映されない。（たとえば、「engagement」という単語がドキュメント内で経済的な「債務・約束」という意味で使用しているのか、あるいは軍事的な「関与」を意味しているのか、「development」という単

語が経済的な「発展・開発」を指しているのか、軍事兵器の「開発」を表しているのかは計測結果に表れない。）

本研究では、国際関係論の専門知識を情報技術による意味の多次元ベクトル空間に反映させ、ドキュメント間の意味的な関連性について計量・分析する手法について述べる。また、知識ベースの研究成果をデータマイニング手法や情報可視化技術と組み合わせることにより、従来の情報検索アプローチと知識工学アプローチのブリッジとなるような高度な情報分析・知識発見システムの実現方式についても述べる。この方式を国際関係の政策ドキュメント分析に適用することにより、大量の政策ドキュメント群について、コード化を経ずに可能な限り原データに近い形で意味的な比較分析が可能になる［40］［41］［42］［43］。次章では、この高度情報分析・知識発見システムの基本構成について概要を紹介する。

四　国際関係ドキュメントを対象とした意味的・時空間的分析・可視化システムの実現

本章では、国際関係の政策関連ドキュメントデータ（および、マルチメディアデータ）を対象とした「意味的・時空間的分析・可視化機構を有する４Ｄ世界地図システム」の基本設計、諸機能の実現方式、およびプロトタイプシステムによる実験結果について述べる。本システムは、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「４Ｄ世界地図」と呼ばれる時間軸を伴った三次元地図上に可視化する情報分析・知識発見システムである［44］［45］。ここで扱うドキュメントデータとは、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、特にテキストデータであり、日々広域ネットワーク上に公開される一般の国際ニュース記事、政府や国際機関の公式発表、プレス・ブリーフィン

グ、政策ステートメント、政府高官の談話、議会議事録、ＮＧＯの活動記録などを指す。

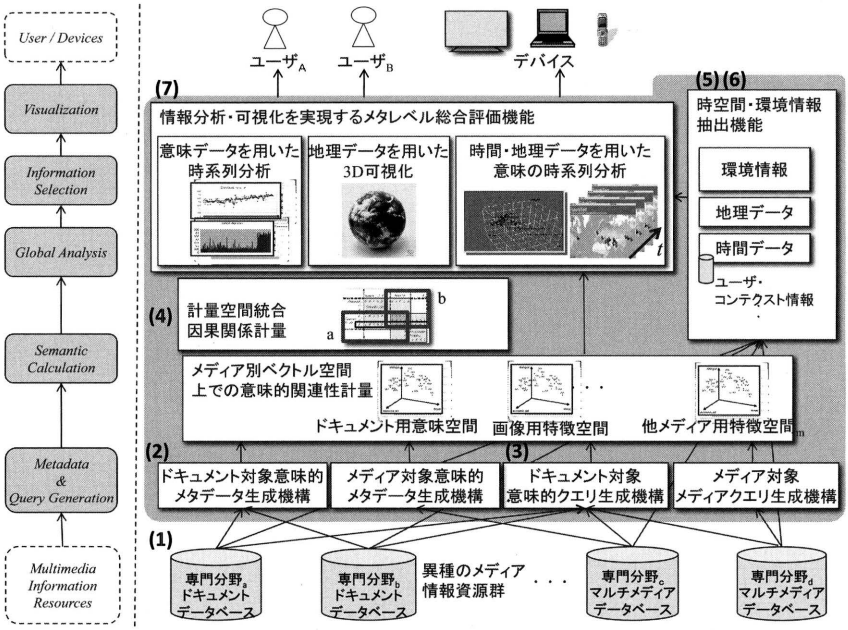
本システムの特色は、国際関係論の専門知識と情報工学におけるデータベース技術、データマイニング、情報可視化技術を統合する事により、多面的な内容を含む国際問題に関するマルチメディアデータ群を対象として、利用者の興味・関心・視点・問題領域と時空間情報の組み合わせに応じた動的な分析と可視化を行う点にある。本システムは、利用者・分析者の設定した問題領域毎の多種多様なストーリー（解釈）の生成を可能とするものである。

システム構成図を**図2**に示す。

本システムは、図2に示すように、主に二種類の機能群により構成される。第一に、時間情報・地理情報を伴ったメディアデータ群を、対象メディア別の多次元ベクトル意味空間に写像する機能群であり、第二に、意味空間において計量されたデータ間の関連性および時空間的分析結果を、４Ｄ世界地図上に写像するための機能群である。

(1)**データ収集機能**：利用者が入力した検索条件に応じて広域ネットワーク上の複数情報源から該当するマルチメディアデータをクローリングにより収集する機能である。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・期間別に関連するマルチメディアデータの収集が可能になる。(2)**メタデータ抽出機能**：収集したマルチメディア群からメディアデータ別に意味的メタデータおよび特徴量を抽出する機能である。この機能により、検索空間に写像するための専門的かつメディアに特化したメタデータを自動的に獲得することが可能となる。(3)**クエリ生成機能**：マルチメディア群から抽出された意味的メタデータおよび特徴量を対象として、複数の集合演算子を適用し、ユーザの意図に応じた意味的クエリおよびメディア・クエリを自動生成する機能を設計、実現する。(4)**意味的関連性計量機能**：メタデータ化されたメディアデータの情報をベクトル化し、専門分

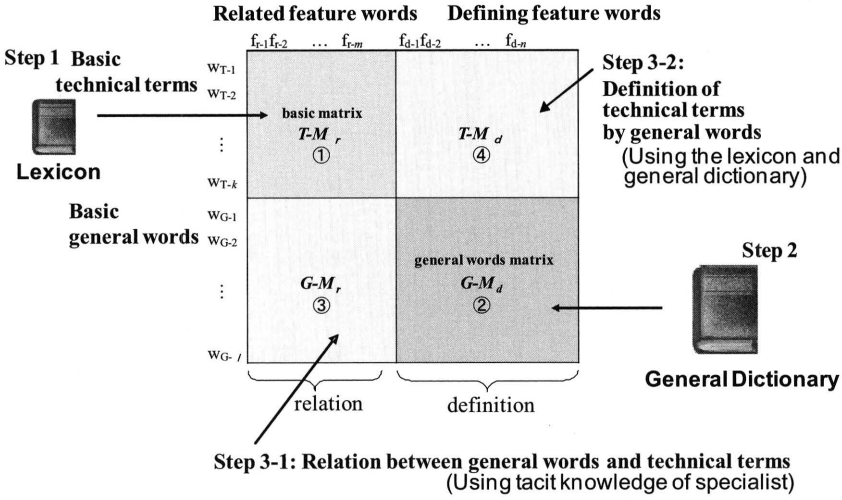
図２ 意味的・時空間的分析・可視化機構を有する４D世界地図システムの構成図



野別・メディアデータ別の検索空間に写像する機能である。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キーワードから自動生成されたメディア・クエリとの意味的な近さが相関量として計量される。(5)**時空間メタデータ抽出機能**：分析対象メディアデータおよびクエリ生成用メディアデータ群から時間的・空間的情報をメタデータとして抽出する機能である。また、ユーザのデバイスよりユーザコンテクストとしてユーザの位置・時刻を抽出する。(6)**環境情報抽出機能**：分析対象メディアデータおよびクエリ生成用メディアデータ群から環境情報をメタデータとして抽出する機能である。また、ユーザのデバイスよりユーザコンテクストとしてユーザのデバイス情報を抽出する。(7)**情報分析・可視化機能**：(1)～(5)の情報検索システムと(6)の時間的・空間的情報データベースとの連結・統合を行う総合評価機能である。分析結果を時系列およびイシュー別・地域別に表示可能とする。

(4)の意味的関連性計量には、清木康、北川高嗣によって提案された「意味の数学モデル」[31][32][34]と「意味的連想検索方式」[33][35][36][37]を適用する。この本式は、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行う。「部分空間（コンテクスト）選択」というオペレーションにより、対象ドキュメント群の意味的相関関係の動的計量を可能とする点が特徴であり、言葉の意味は文脈＝コンテクスト＝部分空間によって決まるという認識論から設計されたモデルである。また、(4)の意味的計量を行う多次元ベクトル空間の生成には、国際関係分野のドキュメント検索に応用するための意味的連想検索空間生成方式[40][41][43]を適用する。この空間は、専門分野の検索空間と一般知識の検索空間とを統合する方式[42]を適用している。この統合的意味的連想検索空間は、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、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

図3 国際関係ドキュメントを対象とした意味的連想空間生成のためのデータ行列



ユ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。

さらに、(2)(3)ではマルチメディアを対象とした検索用クエリの動的生成方式［61］を適用する。このマルチメディア・クエリ生成方式は、マルチメディア（文書・画像・音声・動画）群の集合から特徴量を抽出し、その特徴を用いて、利用者の意図を反映した検索用クエリを動的に生成する方式である。

また、(7)では既存のマルチデータベース技術［47］［48］［49］［50］［51］［52］、および、国際関係意味検索空間と情報分析・知識発見システムを連結し、国際問題に関するドキュメント群を対象とした意味的・時空間的分析機構を有する情報分析・知識発見システムの実現方式［43］［44］を適用している。本意味的・時空間的分析機構を有する情報分析・知識発見システムは、利用者の設定した視点・問題領域に意味的に関連するドキュメントが、地理的にどのように分布し、その分布が時系列的にどのように変化しているかを可視化するものである。

以下の節では、(4)の意味空間、および、(7)の時空間的分析・可視化機能について概要を述べる。

(一) 国際関係分野のドキュメントを対象とした意味のベクトル空間

本節では、国際関係分野（International Relations：以下ＩＲ）のドキュメントを対象とした検索・分析環境を実現するために意味的連想検索方式［33］［34］［35］［36］［37］を適用した多次元ベクトル空間の生成方式［40］［42］［43］の概要を示す。

本方式の特徴は、専門用語に関する情報源（用語集）を対象として生成する意味的計量空間と、一般語に関する情報源（一般辞書）を対象として生成する意味的計量空間を統合し、専門用語と一般語の両者の意味的関係の計量を行うことができる意味的連想検索空間を実現する点にある。

本方式は、ＩＲ分野に関するドキュメントデータ群を対象とした、専門用語を含有する意味的検索空間を生成することを目的としている。まず、専門用語集を用いて基本データ行列を生成する。次にこれを、一般辞書を用いて構築された既存のベクトル空間のマトリクスに合成する。生成および合成に際して、意味空間における「関連性（reference）」と「定義（definition）」というコンセプトを基本とし、専門用語を一般語で定義づけるプロセスと、一般語を専門語との関連によって特徴づけるプロセスを行う。

「特徴語（feature）」とは、意味的検索空間生成のための行列において横軸にあたる単語および用語を指し、「基本語」および「基本用語」は行列において縦軸にあたる基本データを示す。ＩＲ分野の基本用語は w_T と表し、一般の基本語を w_G と表すこととする。また特徴語のうち、基本語との「関連性」を表す特徴語は「関連特徴語（f_r）」として定義し、基本語の定義を表す特徴語を「定義特徴語（f_d）」と定義する。

図3は、空間生成のためのデータ構造（行列の構成）を示している。T-M_r は k 個のＩＲ基本用語（w_{T-1},

$w_{T-2}, \cdots, w_{T-k}$）について m 個のＩＲ分野の関連特徴語（$f_{r-1}, f_{r-2}, \cdots, f_{r-m}$）で特徴づけた、ＩＲ用語間の関連性を示したメタデータ行列であり、$G\text{-}M_d$ は l 個の一般基本語（$w_{G-1}, w_{G-2}, \cdots, w_{G-l}$）について一般の定義特徴語（$f_{d-1}, f_{d-2}, \cdots, f_{d-n}$）で定義を示したメタデータ行列である。ここに、ＩＲ基本用語について定義特徴語で特徴づけを行う部分 $T\text{-}M_d'$ および、一般語についてＩＲ関連特徴語で関連づけを行う部分 $G\text{-}M_r$ を加えることにより、統合メタデータ行列 $T/G\text{-}M_d$ が生成される。

（a－1）国際関係分野の基本データ行列の生成

行列 $T\text{-}M_r$ を生成する。第一に、ＩＲ分野を表現するために必要な特徴語群を準備する。専門用語集を用いて、各項目の説明文の中から関連する他の項目を抽出し、この集合を関連特徴語群とする。第二に、同用語集を用いて各項目を抽出し、この集合を基本用語群とする。第三に、関連特徴語群を用いて各基本用語の特徴づけを行う。同用語集を用いて、各基本用語の説明文に現れる関連特徴語には1を、現れない関連特徴語には0を、否定的な意味で現れる関連特徴語には-1を設定する。以上のプロセスにより、ＩＲ分野における基本用語と関連特徴語の関係を示す基本データ行列が生成される。

（a－2）一般語辞書によるデータ行列との合成

行列 $T\text{-}M_r$ と行列 $G\text{-}M_d$ を合成するため、部分 $G\text{-}M_r$ と部分 $T\text{-}M_d$ を生成する。

・専門語による一般基本語の関連づけ

部分 $G\text{-}M_r$ を生成する。すなわち、l 個の一般基本語（$w_{G-1}, w_{G-2}, \cdots, w_{G-l}$）について、$m$ 個のＩＲ分野の関連特徴語（$f_{r-1}, f_{r-2}, \cdots, f_{r-m}$）で特徴づける。

・一般基本語による専門語の定義づけ

部分 $T\text{-}M_d$ を生成する。すなわち、k 個のＩＲ基本用語（$w_{T-1}, w_{T-2}, \cdots, w_{T-k}$）について、一般の定義特徴語

$(f_{d-1}, f_{d-2}, \cdots, f_{d-n})$ で特徴づける。

・その他の語の追加

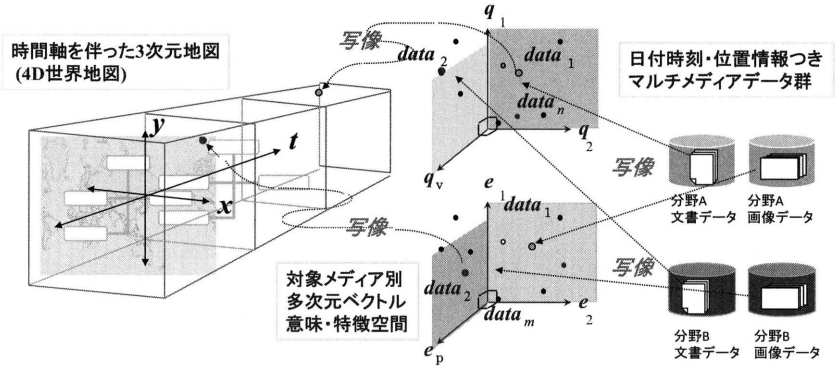
行列 $T\text{-}M_r$ と行列 $G\text{-}M_d$ のいずれにも存在しないが検索対象ドキュメント群に頻出する語を、基本データとして縦の列に追加し、IRの関連特徴語と一般語の定義特徴語で特徴づける。

以上の生成・合成プロセスにより、統合行列 $T/G\text{-}M_d$ が生成される。

(b) 実現方式

本方式による実現例として、汎用されているIR用語集の"Dictionary of International Relations"[38]を用い、(a-1)(a-2)で示した方法で行列を生成した。この用語集では、七一六の専門用語(項目)について、その定義、出典、歴史および他の用語(項目)との関連性を説明している。そのうち、七一六の各項目の説明文から関連項目のみを関連特徴語 f_r として抽出し、IR基本データ行列 $T\text{-}M_r$ とした。例えば、「arms control(軍備管理)」という項目については、「capability」「actor」「crisis management」「deterrence」「disarmament」「Cold War」「superpower」「non-proliferation」「ABC weapons」「security regime」などの関連特徴語に1という値が与えられる。この基本行列 $T\text{-}M_r$ は、IR-Dic. 内の項目間の関連性を表すものであり、基本データ数712、特徴語数712の712×712行列となった。なお、この行列のベクトルを基に生成した意味的連想検索空間(IR空間)は、次元数710となった。また、生成した行列 $T\text{-}M_r$ に、一般語辞書を用いて構築された既存の行列 $G\text{-}M_d$ を合成した。$G\text{-}M_d$ は、英英辞書である"Longman Dictionary of Contemporary English"[39]を用いて生成した。Longman-Dic. は約五六、〇〇〇語の一般語について約二、〇〇〇語の基本単語で説明した辞書である。$G\text{-}M_d$ はこの約、二〇〇〇語の基本単語を他の基本単語で定義づけて生成された約2000×約2000の行列であり、各行は一般語の定義特徴語 f_d によって特徴づけられた一般基本単語のベクトル

図 4 意味的・時空間的分析・情報可視化機能の基本アイデア



を表している。

統合空間生成の際は、専門語による一般語の関連づけとして部分 $G\text{-}M_r$ を生成した。例えば、一般語の基本語「arms」は、「arms control」「arms race」「arms sales」といったＩＲ関連特徴語で特徴づけられている。また、一般語による専門語の定義づけとして部分 $IR\text{-}M_a$ を生成した。例として、ＩＲの基本用語「arms control」は、「arms」「control」「reduce」「remove」「weapon」「threat」「force」などの定義特徴語で特徴づけられる。特徴づけは、IR-Dic. の説明文中から用語の定義の部分について動詞と名詞を抽出した。定義特徴語にない単語がある場合は Longman-Dic. で調べ、動詞と名詞を抽出した。さらに、その他の基本語の追加を行い、行列 $T\text{-}M_r$ と行列 $G\text{-}M_a$ のいずれにも存在しないが検索対象ドキュメント群に頻出する語、たとえば「democracy」「economy」「policy」といった重要単語を基本データに追加し、一般辞書およびＩＲ用語集を用いて定義特徴語と関連特徴語で特徴づけた。

以上の作業の結果、基本データ数約 2000＋712、特徴語数 2861 の行列が生成された。なお、この行列を基に生成された意味的連想検索空間（統合空間）の次元数は 2846 となっている。本空間を用いることにより、利用者は専門語と一般語の組み合わせを用いて専門語を直接含まない専門的なドキュメントを検索することが可能となった。

(二) 意味的・時空間的分析・情報可視化機能

言語分析において「共時態」「通時態」という概念があるように［46］、政策科学、とりわけ国際関係に関する政策科学においても事象や表象についての共時的分析と通時的分析が不可欠であると考える。ここでの共時的分析とは、出来事の固まりを特定の時間軸で切り取り、同時刻・同時代における特定の出来事の空間的な広がり・分布について分析すること、または、同時刻・同時代に起きた別の出来事について検証することであり、通時的分析とは、特定の空間・場所を固定して、その範囲における特定の出来事の時間的な変化を分析することである。これらの視点は、政策科学の提唱者H・ラズウェルが指摘するように、政策分析には時間的分析と空間的分析が不可欠という前提に立つものである［1］。

一方、データ工学の分野においては、マルチデータベース・スキーマ［45］［51］［52］［33］、空間データベースシステム［50］、メタデータベースシステム［32］［55］、時空間マルチデータベースシステム［54］［56］の研究において、異種データベース間の相互運用性（interoperability）と動的統合（dynamic integration）の方式について多くの研究がなされている。

本機能群は、これら時空間マルチデータベースシステム構築技術と先述の国際関係ドキュメントを対象とした意味的連想検索空間とを統合し、対象ドキュメント間の意味的関連性の時間的遷移について地理情報を伴って可視化するものである。本機能は、歴史や国際関係などの事象・現象における「ストーリー」は、意味的・時間的・空間的コンテクストの組み合わせに応じてほぼ無限に存在する、という認識論に立って設計されている（**図4**）。

本機能群は、意味、パターンマッチング（ブール演算）、空間、時間の各コンテキストに関する関数によってユ

ーザの発行した検索クエリとベクトル化されたドキュメントとの相関量計量を動的に行い、ユーザの発行した検索条件に応じて結果を絞り込み（projection）、アウトプットとしてのドキュメント集合をランキングに応じて地図上に表示する機能群である。

Semantic Calculator on Vector Spaces: $f_{semantic}$

対象ドキュメント集合を $D=\{d_1, d_2, \ldots, d_x\}$、意味空間を S, ユーザの指定した検索クエリを sm, 相関量の閾値を th, とすると、意味的連想検索関数 $f_{semantic}$ は以下の式で表される。

$$f_{semantic}(sm, D, S, th) \rightarrow \{d_j, v_{s,dj} > th\}$$

あるドキュメント d_j のクエリ sm に対する空間 S における相関量が閾値 th よりも大きい場合、d_j はアウトプットとして表示される。クエリ sm が与えられない場合、ドキュメント集合 D のすべてが表示される。

Boolean Evaluator: $f_{boolean}$

対象ドキュメント集合 D 内に含まれる単語を $\{kw_1, kw_2, \ldots, kw_i\}$、ユーザ指定したキーワードを b と表すと、ブール演算関数 $f_{boolean}$ は以下の式で表される。

$$f_{boolean}(b, D) \rightarrow \{d_h \mid kw_{dh} = b\}$$

ユーザの指定したあるキーワードが b あるドキュメント d_h に含まれていた場合、そのドキュメントはアウトプットとして表示される。

Spatial Evaluator: $f_{spatial}$

緯度経度に変換されたドキュメント内の地理情報とユーザの発行した検索条件とのマッチングを行い、アウトプットのセレクションを行う。あるドキュメント d_p 内における地理情報を $\{s_1, s_2, \ldots, s_i\}$ とし、ユーザの

指定した地理的検索条件を sp とすると、空間評価関数 $f_{spatial}$ は以下の式で表される。

$$f_{spatial}(sp, D) \rightarrow \{d_r \mid s_{dp} \subseteq sp\}$$

地理的検索条件 sp が指定されない場合、全てのドキュメント D がアウトプットとして表示される。

Temporal Evaluator: $f_{temporal}$

数値データに変換されたドキュメント内の日付情報とユーザの発行した検索条件とのマッチングを行い、アウトプットのセレクションを行う。あるドキュメント d_r 内における時間情報を $\{t_1, t_2, \ldots, t_r\}$ とし、ユーザの指定した時間的検索条件を tm とすると、時間的評価関数 $f_{temporal}$ は以下の式で表される。

$$f_{temporal}(tm, D) \rightarrow \{d_r \mid t_{dr} \subseteq tm\}$$

時間的検索条件 d_r が指定されない場合、全てのドキュメント D がアウトプットとして表示される。

Global Analyzer: f_{global}

意味的連想検索関数、ブール演算関数、空間的評価関数、時間的評価関数の全ての結果について、AND/OR のオペレーション Op を行う。このグローバル評価関数 f_{global} は次の式で表される。

$$f_{global}(d_j, d_k, d_p, d_r) \rightarrow \{D' = Op(d_j, d_k, d_p, d_r)\}$$

$$Op = (\cup, \cap)$$

各関数 $f_{semantic}$, $f_{boolean}$, $f_{spatial}$, $f_{temporal}$ のアウトプットとして得られたドキュメント d_j, d_k, d_p, d_r について、AND/OR オペレーション Op の結果得られたドキュメント集合 D' がアウトプットとして表示される。

図５ ドキュメント検索結果（ロイター国際ニュース記事29件、網掛けは正解ドキュメント）

Keyword	EU (technical term)		economy (general word)		EU + economy	
Rank	Document ID	correlation	Document ID	correlation	Document ID	correlation
1	EU_leader	0.228731	OECD	0.314615	WTO	0.293941
2	EU_election2	0.223200	trade_barrier_poverty	0.300814	OECD	0.292742
3	WTO	0.215860	WTO	0.284417	trade_barrier_poverty	0.260793
4	EU_election1	0.198315	Brasil_IMF	0.245942	EU_election2	0.249190
5	OECD	0.180130	Mexico_migrants	0.224171	Brasil_IMF	0.225401
6	Russia_EU_summit	0.175681	G8_Arab_Israel	0.216797	EU_leader	0.218966
7	trade_barrier_poverty	0.149634	G8_Arab	0.212676	EU_election1	0.214392
8	G8_Arab_Israel	0.144973	EU_election2	0.211438	Russia_EU_summit	0.202394
9	Brasil_IMF	0.144374	Japan_tariff_cut	0.190317	Mexico_migrants	0.187978
10	NATO_Iraq	0.136771	Belgium_kidnapping	0.188365	G8_Arab	0.183492

Keyword	nuclear weapons (tech. term)		terror (general word)		nuclear weapons + terror	
Rank	Document ID	correlation	Document ID	correlation	Document ID	correlation
1	NorthKorea_missile	0.186406	G8_Arab_Islael	0.186206	NorthKorea_missile	0.196572
2	G8_nuclear_weapons	0.163908	Iraqi_prisoner_Geneva	0.178489	G8_nuclear_weapons	0.171631
3	Nepal_Maoist_rebels	0.127008	NorthKorea_nuclear	0.173509	Nepal_Maoist_rebels	0.137768
4	Nato_Iraq	0.109748	UN_Rwanda_genocide	0.173057	Afghan_Chinese_killed	0.117263
5	Iran_nuclear	0.105083	Madrid_bomb_suspect	0.169807	Olympic	0.113940
6	Afghan_Chinese_killed	0.104195	Iran_Nuclear	0.169089	Nato_Iraq	0.113749
7	Olympic	0.103030	Afghan_Chinese_killed	0.167931	Iran_nuclear	0.112816
8	NorthKorea_nuclear	0.091837	G8_nuclear_weapons	0.166116	NorthKorea_nuclear	0.111772
9	plane_crash_Gabon	0.086706	Olympic	0.165223	UN_Rwanda_genocide	0.094980
10	Afghan_redcross	0.081193	G8_Arab	0.165187	Madrid_bomb_suspect	0.091999

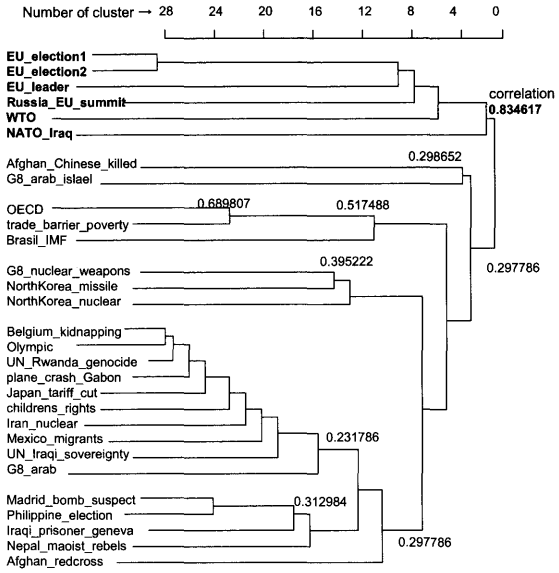
五　実ドキュメントデータを対象とした適用実験

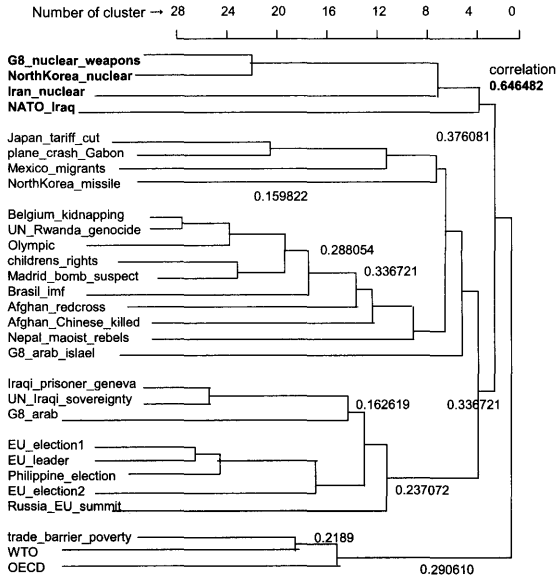
（一）　実験1：国際関係ドキュメントを対象とした意味のベクトル空間の精度検証

四章一節の方式により生成した意味のベクトル空間の精度および有効性を検証するため、複数のトピックを扱ったロイターの国際ニュース記事二九件［57］を用いてドキュメント検索の予備実験を行った。各ドキュメントには結果を見やすくするため、doc1,doc2,… といった番号ではなく、記事タイトルより "NorthKorea_missile", "trade_barrier_poverty" などの文字によるＩＤが与えられている。また、各ドキュメントには、空間生成に用いた辞書単語リストとのパターンマッチングにより自動抽出された単語が、単語の出現頻度に基づく重要度（TF*IDF: Term Frequency*Inversed Document Frequency）に応じて重みつきメタデータとして付与されている。たとえば、ドキュメント "NorthKorea_missile" には "test, missile, north, talk, engine, south, official…" といったメタデータが、"trade_barrier_poverty" には "country, poor, develop, trade, cut, industrial, tariff,…" といったメタデータが与えられている。**図5**は、検索クエリとして「ＥＵ」「経済」「ＥＵ、経済」を入力した場合と、「核兵器」「テロ」「核兵器、テロ」を入力した場合の結果を示している。専門家により予め正解として設定されたドキュメントのＩＤに網掛け処理をしてある。

図5に示すように、ＩＲ用語である「ＥＵ」、一般語「経済」のそれぞれのクエリについて、関連するドキュメント（正解）が上位を占めていることが分かる。また、「ＥＵ、経済」の専門語＋一般語の組み合わせによるクエリについては、両者の結果が混在した検索結果を得ていることが分かる。また、異なる文脈キーワード「核兵器」「テロ」「核兵器、テロ」についても同様に、専門語、一般語、専門語＋一般語のそれぞれによって妥当な検索結果が得られていることが分かる。

図6 ドキュメントのクラスタリング結果
（ロイター国際ニュース記事29件、階層的クラスタリングによる）





また、**図6**は、図5の検索結果について、動的クラスタリングと意味的データマイニング手法［58］を用いてドキュメント間の関係を分析・可視化したダイアグラムである。左の「EU、経済」をクエリとした場合のクラスタリング結果では、経済関連のドキュメントが高い相関量を持ってクラスターを形成しているのに対し、右の「核兵器、テロ」をクエリとした場合では、安全保障関連のドキュメントが相関の強いクラスターを形成しており、逆にそれぞれのクエリに対して各ドキュメントクラスターの位置が逆転していることも見てとれる。

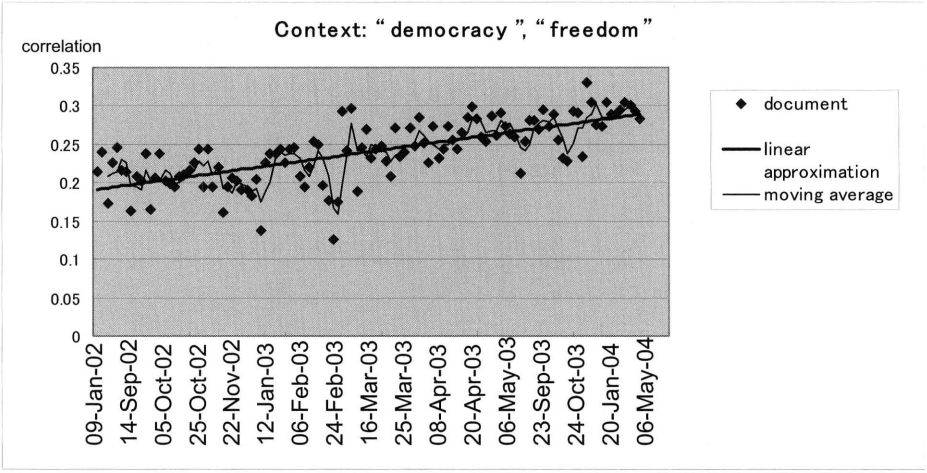
なお、ベクトル空間の精度について定量的に検証するため、空間を構成する全単語二八六一件をクエリとした単語間相関量計量を行った結果、上位二〇件における適合率（precision）は八五・六九％であった。また、BBCの国際ニュース記事一〇〇件［59］を対象として、空間を構成する全単語二八六一中、政治・経済・外交・安全保障・公共衛生・人権・環境・エネルギー・科学技術の各カテゴリーからランダムに選択したクエリ八六件についてドキュメント検索を行った結果、上位一〇件における適合率は七二・二二％であった。

（二）実験2：意味的・時空間的分析・情報可視化機能の検証

四章二節の方式により生成した意味的・時空間的分析・情報可視化機能の実現可能性および有効性を検証するため、時系列を伴ったドキュメントデータ（日付つきドキュメント）を対象に、予備実験を行った。対象ドキュメントを二〇〇二年－二〇〇四年のイラク関連米国大統領演説一一四件［60］とし、検索クエリを「大量破壊兵器、脅威」「民主主義、自由」と設定し、それぞれの検索クエリに対して相関量の高いドキュメントがどの時期に集中しているかを比較する。

図7は、それぞれのクエリに対しての各ドキュメントの相関量（縦軸：相関量、横軸：時系列）、各ドキュメントの並びを時系列データとみなした場合の線形近似（太線）、および、三区間移動平均（細線）を表している。各クエリに対する各ドキュメントの相関量は部分空間が異なるため範囲も異なる（上図のクエリの場合は、相関量の

図７ 時系列を伴ったドキュメントデータを対象とした相関量計量結果
（2002年 - 2004年のイラク関連米国大統領演説114件）



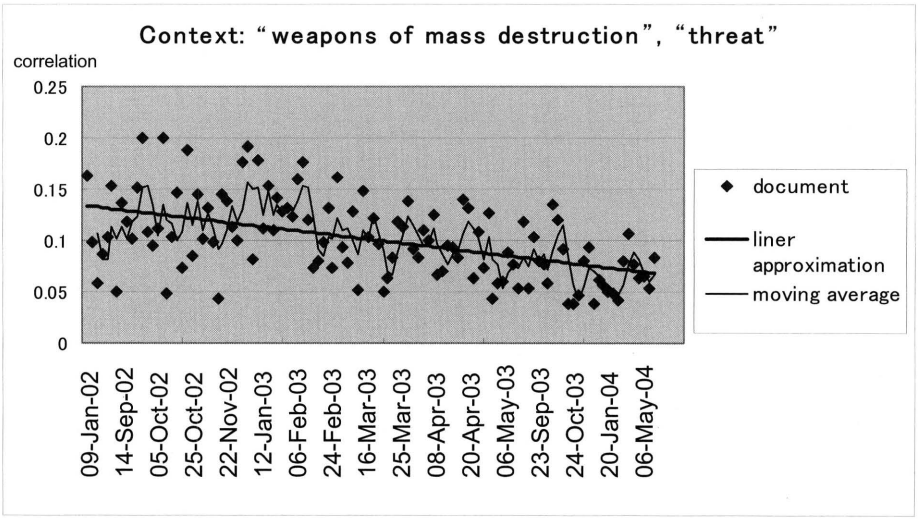


図８ 単語出現頻度計測結果
（2002年-2004年のイラク関連米国大統領演説１１４件）

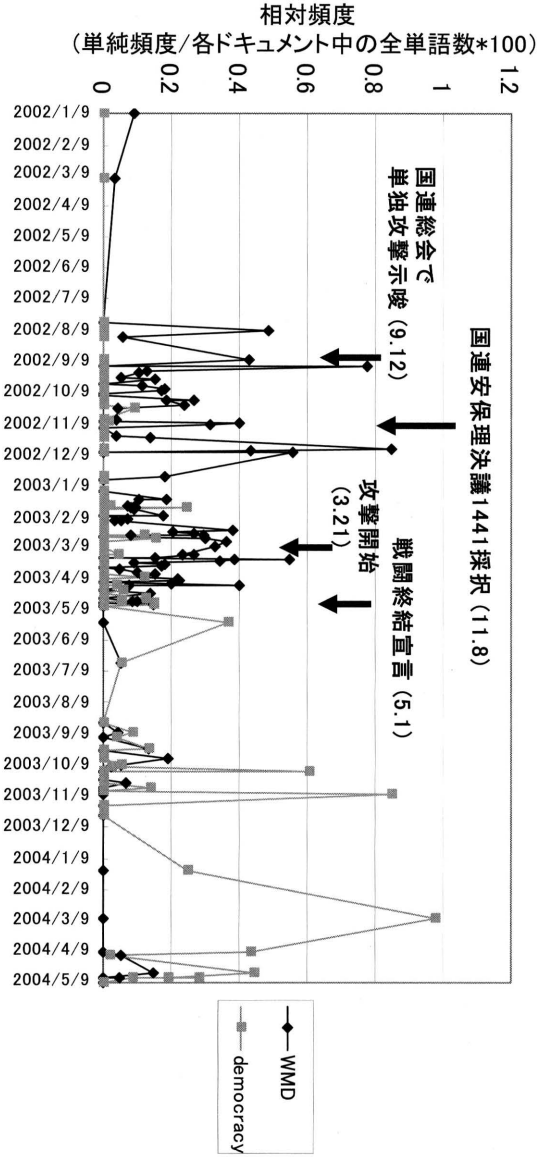
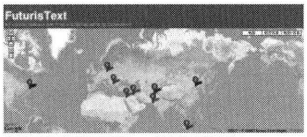
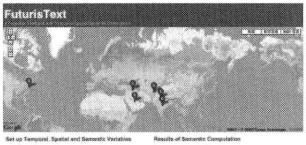


図 9 意味的相関の高いドキュメントの空間的分布可視化結果
（1993 年 - 2002 年の米国大統領と各国首脳会談におけるプレス・ブリーフィング記事 40 件、検索クエリ：「自由貿易」「人権」「天然資源」、表示時期：2001 年）



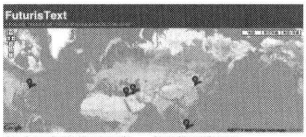
Context
= "free trade"

[1] 0.249874 doc17.txt
[2] 0.235563 doc8.txt
[3] 0.222802 doc32.txt
[4] 0.212001 doc29.txt
[5] 0.210511 doc20.txt
[6] 0.205174 doc22.txt
[7] 0.204091 doc21.txt
[8] 0.199376 doc3.txt
[9] 0.193422 doc7.txt
[10] 0.193174 doc4.txt



Context
= "human rights"

[1] 0.189881 doc15.txt
[2] 0.180994 doc22.txt
[3] 0.178667 doc14.txt
[4] 0.177363 doc19.txt
[5] 0.175750 doc1.txt
[6] 0.175325 doc20.txt
[7] 0.174758 doc25.txt
[8] 0.174301 doc6.txt
[9] 0.174190 doc35.txt
[10] 0.173730 doc13.txt



Context
= "natural resources"

[1] 0.189358 doc15.txt
[2] 0.183708 doc14.txt
[3] 0.181338 doc40.txt
[4] 0.177446 doc16.txt
[5] 0.176407 doc19.txt
[6] 0.176381 doc25.txt
[7] 0.175910 doc20.txt
[8] 0.174764 doc22.txt
[9] 0.174748 doc1.txt
[10] 0.174430 doc24.txt

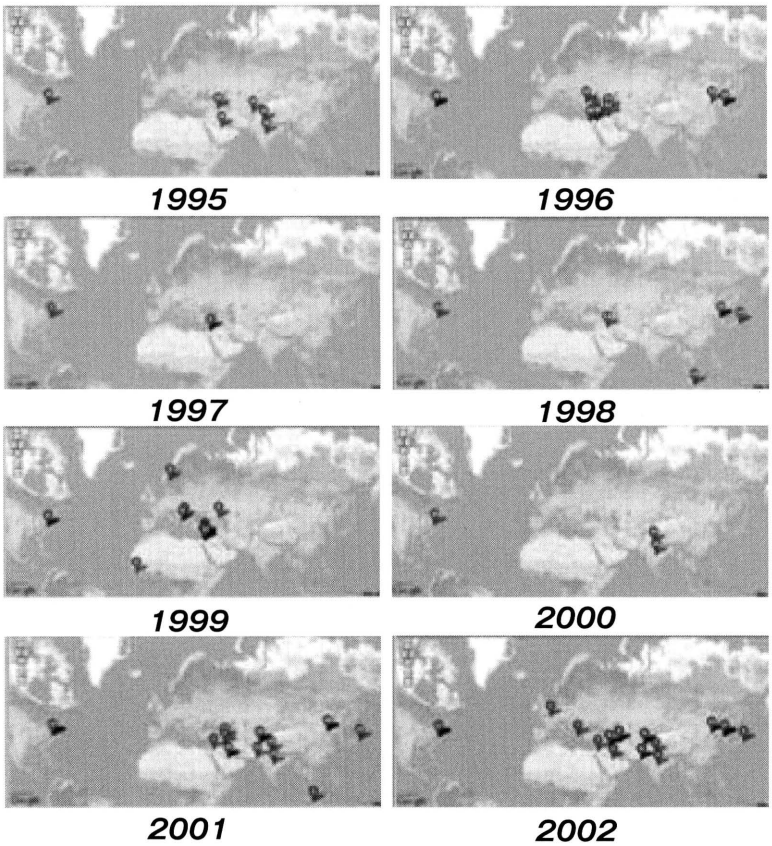
最大値〜最小値は約〇・八三〜約〇・一三、下図のクエリの場合は、約〇・二〜〇・〇四)。比較のために最も大きい値で正規化を行い、グラフを重ね合わせてみると、近似線の交差点が二〇〇三年三月付近であることが分かる。この結果をさらに詳細に検討するため、「大量破壊兵器(WMD)」と「民主主義」の単語出現頻度(単純出現回数を各ドキュメントの全単語数で割った相対頻度)を調べた。**図8**に示すように、米国によるイラク攻撃戦闘終結宣言の時期を境に、それぞれの頻度が逆転していることが分かる。(二〇〇四年三月の「民主主義」の値はほぼ一・〇に近いため、外れ値と考えられる。)

データマイニングの結果は、それ自体で特定の意味を表すものではない。利用者(分析者)は単に総当たり的に得られた結果を眺めるのではなく、作業仮説を立て、大凡の見当をつけた上で結果を解釈する必要がある。例えば図7の例では「ブッシュはイラク戦争前まで戦争の大義として『大量破壊兵器』という言葉を多用していたが、戦闘終結後はそれに代わり『民主主義』を多用しているようである。これは、この時期の米国の対外戦略(対テロ戦争)におけるロジックの転換を意味しているのではないか」といった作業仮説である。また、これらの作業仮説を裏付けるためには当然データマイニングの結果のみならず、図8に示すような大統領支持率、米軍死者数、「有志連合」死者数、経済指標の時系列推移など外部データを参照する必要がある。

次に、意味的・時空間的分析結果の地図上への写像・可視化機能を検証するため、日付情報(発表日)と地理情報(ドキュメント内で言及されている地域・国)を持つドキュメントデータを対象に写像実験を行った。一九九三年-二〇〇二年の米国大統領と各国首脳会談におけるプレス・ブリーフィング記事四〇件[60]を対象に、検索クエリを「自由貿易」「人権」「天然資源」、表示時期を二〇〇一年と設定し、意味的相関の高いドキュメントの空間的分布の違いを比較した。

図9は、異なる検索クエリに対してドキュメント検索を行い、相関量の高い順に地図上にマッピングした結果

図 10　意味的相関の高いドキュメントの空間的分布可視化結果
(1993 年 - 2002 年の米国大統領と各国首脳会談におけるプレス・ブリーフィング記事 40 件、検索クエリ：「紛争」、表示時期：1995 年 - 2002 年)



である。ドキュメント検索結果の相関量順ランキングがスクリーンショット画面の右下に表示され、各ランキングの順位が数字つきマーカーとしてスクリーン地図上に表示されている。「自由貿易」については二〇〇一年に発行されたdoc32（ランキング三位）、doc29（四位）doc31（一五位）が言及された地域・都市に応じて地図上に表示され、「人権」についてはdoc34（一六位）が、「天然資源」についてはdoc20（二〇位）が表示されている。この結果は、本機能の意味的なダイナミズム（動的計量）を表しているといえる。

さらに、図9の実験で用いたものと同じドキュメント群を対象に、意味的ドキュメント検索結果の空間的分布を時系列で表示した結果を**図10**に示す。検索クエリは「紛争」とし、表示時期を「一九九三年－二〇〇二年」と設定した例である。地図上にマッピングされている地域は必ずしも紛争発生地域ということはできないが、少なくとも「紛争というコンテクストに関連の強いドキュメント内で、言及されている地域の分布とその時系列変化」を見てとることができる。一九九五年にはイスラエル－パレスチナ紛争、一九九六年にはボスニア紛争、一九九九年には第二次チェチェン紛争、二〇〇一年には九・一一同時多発テロと米国によるアフガン攻撃が勃発した時期であるという事実に照らしてみると、この結果の妥当性が確認できる。

図9に示した実験結果は、意味的コンテクストを固定し、関連するドキュメントの地理的・空間的な広がりと時系列変化を示した結果であり、図10に示した実験結果は、時間軸を固定し、意味的コンテクストを変化させることによるドキュメントの地理的・空間的分布の動的なマッピングを示したものである。これらの結果は、本機能が多面的な内容を含む国際関係ドキュメント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「ストーリー」を生成可能とすることを示している。

六　まとめ

本稿では、言葉の集合から「雲」の広がりと動きを描くための一方法論の可能性について述べた。雲とは、国際政治や政策科学が対象とするものであり、人間の認知・認識や行動によって形成される社会的価値・社会構造・制度・プロセスでもある。人間は言葉の意味や内容を考え、その理解に応じて行動を起こす。したがって、政策科学の対象とする問題領域は、言葉で表される思想・主義・方針・意図・計画がどのような時間的変化をしているか、どのような空間的な広がりを持っているのか、どのように人の認知や認識に影響を及ぼしているのか、といった点にあるといっていい。本稿で述べた方法論は、膨大かつ多種多様なドキュメントデータの中から何らかのパターンや知識を抽出し、空間的な分布と時系列的遷移に関する分析を行い、「境界のはっきりしない」「ぼんやりとした」「刻々とその姿を変える」雲についてその「おおまかな」流れと概観を描き、起源と行方についての推測・予測を可能とするための試みの一つである。

参考文献

[1] H. D. Lasswell, "The Policy Orientation," D. Lerner and H. D. Lasswell, eds., *The Policy Sciences: Recent Developments in Scope and Method*, Stanford Univ. Press, 1951.

[2] 薬師寺泰蔵『公共政策』（現代政治学叢書：10）、東京大学出版会、一九八九年。

[3] 薬師寺泰蔵「安全学とプログラム型国際政治学」、『安全科学』国際高等研究所報告書一九九八-〇一〇、一九九九年。

[4] 榊原英資、薬師寺泰蔵『社会科学における理論と現実—実証分析における一つの試論』日本経済新聞社、一九八一年。

[5] G. A. Almond and S. J. Genco, "Clouds, Clocks, and the Study of Politics," *World Politics*, Vo. XXIX, No. 4, pp. 489–552, 1977.
[6] Gary King, Robert O. Keohane and Sidney Verba, *Designing Social Inquiry*, Princeton University Press, 1994.
[7] 薬師寺泰蔵「政策分析におけるモデリングの諸問題」『オペレーションズ・リサーチ』一九七九年八月号。
[8] Paul Smoker, "Trade Defense and the Richardson Theory of Arms Races: A Seven Nations Study," *Journal of Peace Research*, Vol. 2, No. 2, 1965.
[9] Bruce M. Russett, *International Regions and the International Systems*, 1967.,
[10] David Singer et al., "Capability Distribution, Uncertainty, and Major Power War, 1820–1995", Bruce M. Russett ed., *Peace, War and Numbers*, 1972.
[11] 山本吉宣「混沌の中の法則性—戦争の確率論的なモデル」、山本吉宣・薬師寺泰蔵・山影進編『国際関係理論の新展開』、東京大学出版会、一九八四年。
[12] 薬師寺泰蔵「政治学における近代的モデリング—リチャードソン・モデルを中心として—」、山本吉宣・薬師寺泰蔵・山影進編『国際関係理論の新展開』、東京大学出版会、一九八四年。
[13] H. D. Lasswell et al., *The Comparative Study of Symbols*, 1952.
[14] D. A. Zinnes, "A comparison of Hostile Behavior of Decision-Makers in Simulated historical Data," *World Politics* 18, 1966, pp474–502.
[15] O. R. Holsti, "Content Analysis," Gardner Lindzey and Elliot Aronson eds., *The Handbook of Social Psychology*, 1968, pp. 596–632.
[16] O. R. Holsti and Robert C. North, "Comparative Data from Content Analysis: Perception of History and Economic Variables in the 1914 Crisis," Richard L. Merritt and Stein Rokkan eds., *Comparing Nations: The Use of Quantitative Data in Cross-National Research*, 1966, pp. 169–190.
[17] 猪口孝『国際関係の数量分析—北京・平壌・モスクワ、一九六一－一九六六年』巌南堂、一九七〇年。

[18] 武者小路公秀『行動科学と国際政治』東京大学出版会、一九七二年。
[19] 高木誠一郎「文革前中国の対外関心―『人民日報』社説の内容分析一九五〇－一九六五」、『国際関係論のフロンティア―国際関係理論の新展開』東京大学出版会、一九八四年。
[20] 浦野起央『国際関係理論史』勁草書房、一九九七年。
[21] R. Axelrod ed., *The Structure of Decision: The Cognitive Maps of Political Elites*, Princeton U. P., 1976.
[22] 山本吉宣・谷明良「認知構造図」『オペレーションズ・リサーチ』、一九七九年。
[23] C. Jonsson ed., *Cognitive Dynamics and International Politics*, London: Frances Printer, 1982.
[24] ユージン・B・スコルニコフ著、薬師寺泰蔵、中馬清福監訳『国際政治と科学技術』NTT出版、一九九五年。
[25] 薬師寺泰蔵編『アジアの環境文化』慶應義塾大学出版会、一九九九年。
[26] ジム・マクニール、ピーター・ヴィンゼミウス、薬師寺泰蔵著、日米欧委員会日本委員会訳『持続可能な成長の政治経済学：エコノミーとエコロジーの統合』、ダイヤモンド社、一九九一年。(Jim MacNeil, Pieter Winsemius, Taizo Yakushiji, *Beyond Interdependence: The Meshing of The World's Economy and the Earth's Ecology*, Oxford University Press, Newyork, U.S.A., 1991.)
[27] 徳永健伸『情報検索と言語処理』（言語と計算：5）、東京大学出版会、一九九九年。
[28] 福田剛志、森本康彦、徳山豪『データマイニング』（データサイエンス・シリーズ：3）共立出版、二〇〇一年。
[29] 増永良文『データベース入門』サイエンス社、二〇〇六年。
[30] Nello Cristianini, John Shawe-Taylor 著、大北剛訳『サポートベクターマシン入門』(Nello Cristianini, John Shawe-Taylor, *An introduction to support vector machines*, Cambridge University Press, 2000.) 共立出版、二〇〇五年。
[31] T. Kitagawa, and Y. Kiyoki, "The mathematical model of meaning and its application to multidatabase systems," *Proceedings of 3rd IEEE International Workshop on Research Issues on Data Engineering: Interoperability in Multidatabase Systems*, pp. 130–135, April 1993.
[32] Y. Kiyoki, T. Kitagawa, and T. Hayama, "A metadatabase system for semantic image search by a math-

ematical model of meaning," *ACM SIGMOD Record*, Vol. 23, No. 4, pp. 34–41, 1994.
[33] Y. Kiyoki, T. Kitagawa, and Y. Hitomi, "A fundamental framework for realizing semantic interoperability in a multidatabase environment," *Journal of Integrated Computer-Aided Engineering*, Vol. 2, No. 1, pp. 3–20, John Wiley & Sons, Jan. 1995
[34] 清木康、金子昌史、北川高嗣「意味の数学モデルによる画像データベース探索方式とその学習機構」、電子情報通信学会論文誌、D-Ⅱ, Vol. J79-D-Ⅱ, No. 4, pp. 509–519, 1996.
[35] 宮川祥子、清木康「特定分野ドキュメントを対象とした意味的連想検索のためのメタデータ空間生成方式」。情報処理学会論文誌：データベース、Vol. 40, No. SIG5 (TOD2), pp. 15–28, 1999.
[36] 吉田尚史、図子泰三、清木康、北川高嗣「ドキュメントデータ群を対象とした文脈依存動的クラスタリングおよび意味的データマイニング方式」情報処理学会論文誌：データベース、Vol. 41, No. SIG 1 (TOD5), pp. 127–139, 二〇〇〇年。
[37] 石原冴子、清木康「異分野データベース群を対象とした意味的検索空間統合方式とその実現」情報処理学会論文誌、Vol. 43, No. SIG5 (TOD14), pp.37–53, 二〇〇二年。
[38] Graham Evans, and Jeffrey Newnham, *Dictionary of International Relations*, Penguin Books, 1998.
[39] *Longman Dictionary of Contemporary English*, Longman, 1987.
[40] 佐々木史織、清木康、薬師寺泰蔵「国際関係分野ドキュメント群を対象とした意味的連想検索のための空間生成方式」、日本データベース学会 Letters, Vol. 2, No. 1, pp. 39–42. 二〇〇三年。
[41] S. Sasaki, Y. Kiyoki, and T. Yakushiji, "Semantic Space Creation and Associative Search Methods for Document Databases of International Relations," *Proceedings of the 7th IASTED International Conference on Internet and Multimedia Systems and Applications*, Hawaii, U.S.A., pp. 399–405, August 2003.
[42] S. Sasaki, and Y. Kiyoki,, "Space Creation and Evaluation Method using Specialized and General Knowledge for Semantic Associative Search," *Proceedings of the IEEE International Symposium on Applications and the Internet* (*SAINT 2005*) *-the International Workshop on Cyberspace Technologies and Societies* (*IWCTS 2005*),

Trento, Italy, pp. 434–437, February 2005.

[43] S. Sasaki, Y. Kiyoki and H. Akutsu, "An application of Semantic Information Retrieval System for International Relations," *Information Modelling and Knowledge Bases*, Vol. XVIII, pp. 62–79, May 2007.

[44] S. Sasaki, Y. Takahashi, and Y. Kiyoki, "FUTURISTEXT: THE 4D WORLD MAP SYSTEM WITH SEMANTIC, TEMPORAL AND SPATIAL ANALYZERS," *Proceedings of the International Association for Development of the Information Society(IADIS) International Conference e-Society 2008*, Algarve, Portugal, pp. 162–170, April 9–12th, 2008.

[45] S. Sasaki, Y. Takahashi, and Y. Kiyoki, "The 4D World Map System with Semantic and Spatio-temporal Analyzers," *Information Modelling and Knowledge Bases, Vol. XXI*, (to appear), 20 pages, May 2010.

[46] 丸山圭三郎『ソシュールの思想』、岩波書店、一九八一年。

[47] C. Batini, et al., 1986. A Comparative analysis of methodologies for database schema integration. *ACM Computing Surveys*, Vol. 18, No. 4, pp. 324–364.

[48] M. J. Egenhofer, et al., "Metric Details for Natural-Language Spatial Relations." *ACM Trans. Information Systems*, Vol. 16, No. 4, pp. 295–321, 1998.

[49] M. J. Egenhofer, "Spatial SQL: A Query and Presentation Language. Knowledge and Data Engineering," *IEEE Transactions on Volume 6*, Issue 1, pp. 86–95, 1994.

[50] R. H. Guting, "An Introduction to Spatial Database Systems." *VLDB Journal*, 3, pp. 357–399, 1994.

[51] W. Litwin, "An overview of the multidatabase system MRDSM." *Proceedings of the 1985 ACM annual conference on The range of computing*, pp. 524–533, 1985.

[52] W. Litwin, et al., "Interoperability of multiple autonomous databases." *ACM Computing Surveys* (*CSUR*), Volume 22 Issue 3, pp. 267–293, 1990.

[53] N. Yoshida, et al., "An associative search method based on symbolic filtering and semantic ordering." *Proceedings of the 7th IFIP 2.6 Working Conference on Database Semantics*, pp. 215–237, 1997.

[54] Y. Hosokawa, and Y. Kiyoki, "Functional and parallel query processing and query optimization for multi-database systems." *Proc. the 17 th IASTED International Conference on Applied Informatics*, pp. 101–106, 1999.

[55] Y. Kiyoki, et al., "A Metadatabase System Architecture for Integrating Heterogeneous Databases with Temporal and Spatial Operations." *Advanced Database Research and Development*, Series Vol. 10, Advances in Multimedia and Databases for the New Century, A Swiss/Japanese Perspective, pp. 158–165, 2000.

[56] N. Ishibashi, and Y. Kiyoki, "Meta-Chronicle: A Spatial and Temporal Multidatabase System and its Application to Histories." *Proceedings of IEEE International Symposium on Applications and the Internet (SAINT 2004) -the International Workshop on Cyberspace Technologies and Societies* (IWCTS 2004), pp. 515–522, 2004.

[57] Reuters.com: http://www.reuters.com/

[58] N. Yoshida, T. Zushi, Y. Kiyoki, T. Kitagawa, "A Context Dependent Dynamic Clustering and Semantic Data Mining Method for Document Data," *Information Processing Society of Japan Transactions on Databases*, Vol. 41, No. SIG 1(TOD5), pp. 127–139, 2000.

[59] BBC News: http://news.bbc.co.uk/

[60] http://usinfo.state.gov/

[61] S. Sasaki, and Y. Kiyoki, "A Media-Query Creation Method for Representing Distinctive Impressions by the Combiations of Multiple Media Data Sets," *the Institute of Electronics Information and Communication Engineers (IEICE) Technical Report*, DE2008-32(2008-09), pp. 5–6, 2008.