

環境関連記事データを対象とした文脈多様性対応型意味的連想検索と 5次元世界地図システムへの適用

佐々木 史織[†] 藤岡 華子[†] 渡邊 敏央[‡] 大塚 恭平[‡] 石井 昌之[‡] 清木 康[†]

[†] 慶應義塾大学政策・メディア研究科 〒252-0882 神奈川県藤沢市遠藤 5322

[‡] 日本経済新聞社デジタル事業 BtoB ユニット 〒100-8066 東京都千代田区大手町 1-3-7

E-mail: [†] {sashiori, hanako, kiyoki}@sfc.keio.ac.jp, [‡] {toshihiro.watanabe, kyohei.otusuka, masayuki.ishii}@nex.nikkei.com

あらまし 本論文は、大規模新聞記事群を対象とした文脈多様性対応型意味的連想検索と5次元世界地図システム（“5D World Map System”）への適用方式を示す。本方式は、社会・自然現象に関して記述した記事データ群を対象として、5D World Map System を適用することにより、時間、空間、意味的な連続性および意味的関連性の分析結果を一つのストーリーとしてダイナミックに生成、蓄積、可視化するシステムを実現する。本システムは、時空間・意味の多様なコンテキスト（文脈）のダイナミックな組み合わせにより、記事内に潜在的に含まれている事象間の相関関係／因果関係を発見的に抽出し、時間軸を持った世界地図上に可視化する新しい新聞記事閲読・分析環境である。本方式により、多様な文脈に応じて意味的に近い情報を検索するための検索環境、および、事象や文脈自体の時系列変化・空間的拡大を分析するための分析環境を提供することができる。本論文は、システムの実現方法を示し、環境分野、特に「地球温暖化」に関する新聞記事10年分を対象とした検索・可視化実験により、本方式の有効性を明らかにする。

キーワード 新聞記事, 意味検索, 意味計量, ドキュメント, テキスト, 時空間, 可視化, 環境, 地球温暖化, 気候変動

1. はじめに

本論文では、大規模新聞記事データベースを対象とした文脈多様性対応型意味的連想検索の実現方式と、その情報可視化システム「5次元世界地図システム（5D World Map System）」[5]-[9]への適用方式を示す。

新聞記事データは、一社から毎日数百件発行される大規模時系列データであると同時に、多様なトピック・事象と地理情報を含む分野横断型データである。また、新聞記事データベースは、各専門分野の知識に基づいて最新の社会・自然現象について「言葉」によって記述・解説された内容であることから、信頼性と理解可能性の高い情報源であり、「集合知（Collective Knowledge）」とも呼ぶべき知識ベースであることも特徴の一つである。

新聞記事の読者・利用者にとって、大量の記事データ集合の中から、関心・視点・問題意識に応じた適切な情報を獲得すること、また新聞記事の提供者にとっては、そのような知識獲得環境を提供することは重要な課題である。特に社会・自然現象について記述した記事集合の分析者にとって、共時的分析（ドキュメント内容の空間的な広がりに関する分析）や通時的分析（時間経緯に伴うドキュメント内容の変化の分析）は重要であり、ドキュメント間の意味的な関連性や時空間的関連性について分析・可視化表現を可能とするシステムは、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。

本研究の目的は、マルチデータベーススキーマ[10][14][15]、時空間データベースシステム[11][12][13]、意味的連想検索方式[1]-[4]、マルチデータベースシステム[16][17]、時空間的情報を伴ったマ

ルチデータベースシステム[18]といった異種のデータベース群間での相互運用性や動的統合を実現する手法を統合することにより、新たな知識表現メディアを生成することにある。

本方式は、社会・自然現象に関して記述した新聞記事データ群を対象として、5D World Map System を適用することにより、時間、空間、意味的な連続性および意味的関連性の分析結果を一つのストーリーとしてダイナミックに生成、蓄積、可視化するシステムを実現する。本システムは、時空間・意味の多様なコンテキスト（文脈）のダイナミックな組み合わせにより、記事内に潜在的に含まれている事象間の相関関係／因果関係を発見的に抽出し、時間軸を持った世界地図上に可視化する新しい新聞記事閲読・分析環境である。

2. 基本方式

「5次元世界地図システム（5D World Map System）」は、ドキュメントデータを含むマルチメディアデータ間の意味的・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、その分析結果を時系列で並べられた世界地図（5D World Map）という表現メディア上で可視化・表示するシステムである[5]-[9]。5D World Map を構成する要素は、空間の次元（3D）、時間の次元（4D）、そして意味の次元（5D）である。各対象ドキュメントの意味は、多次元ベクトル空間上において計量され、時系列で並べられた地図空間上に1次元のランキングとして表現される。

本システムの特徴は、多面的な内容を含むメディア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、多種多様な事象

についてのストーリー（解釈）を生成可能とする点にある。また、本システムに適用される意味的連想検索方式 [1]-[4]は、言葉・ドキュメント・事象間の多様な意味的関連性は「コンテキスト（文脈）」によって定まるといふコンセプトを実現している。本システムは、この意味的連想検索方式を適用し、意味の多次元空間において動的に計量・評価・分析されたメディアデータを、時間軸と空間軸を持った地図上へ写像する（図 1）。

本システムは、ユーザによって与えられた対象データベースと意味・時間・空間に関する検索条件（意味的・時空間的コンテキスト）を入力として、意味の多次元ベクトル空間上で設定されたコンテキストに応じた関連性計量を行い、時空間的な関連性による射影結果を 5D World Map 上へ統合的に写像し、可視化する。本システムのユーザは、対象データのコンテキストに対する意味的相関量の時系列変化を、空間的分布情報を伴いながら獲得することができる。

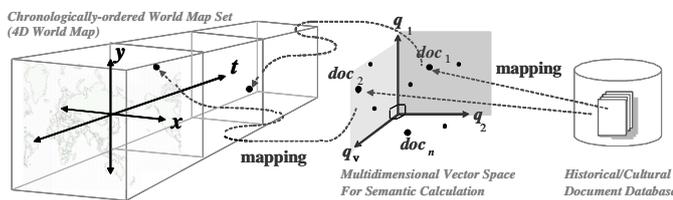


図 1 5次元世界地図システム（5D World Map System）の概念図 [6]

5D World Map System は、分析対象となるマルチメディアデータがドキュメントデータである場合、以下のデータを「コンテキスト」として処理し、分析する。

- 意味コンテキスト (CX1) : 単語または単語のセットにより指定される自然言語による意味表現。ユーザにより発行される意味的コンテキスト sm は $\{word_i : string \mid 0 < i < n\}$ と表現される。
- キーワード・コンテキスト (CX2) : キーワードまたはキーワードのセットにより指定される形式言語によるパターン表現。ユーザにより発行されるキーワード・コンテキスト b は $\{keyword : string\}$ と表現される。
- 空間コンテキスト (CX3) : ドキュメントに関連する場所に関する記述。ユーザにより発行される空間的コンテキスト sp は、緯度経度ポイント ($longitude : float, latitude : float$)、または、複数のポイント $\{(point)\}$ により指定される特定の地域として表現される。
- 時間コンテキスト (CX4) : ドキュメントに関連する年/月/日/時、または、開始時刻と終了時刻により指定される一定の期間。ユーザにより発行される時間的コンテキスト tm は、例えば年単位で指定する場合、($year : integer$)、または、($startYear : integer, endYear : integer$) と表現される。

これらのコンテキストに従って対象ドキュメントを分析するため、システムは以下の機構・機能群により構成される。

1. Metadata Extractor (メタデータ抽出機構) : 意味・

キーワード・時間・空間の各コンテキストに対応するメタデータをドキュメントから抽出する。各コンテキスト対応メタデータは、対象データベース内のドキュメントデータより、以下の機能群により抽出される：Semantic Metadata Generator（意味的メタデータ生成機能）、Keyword Extractor（キーワード抽出機能）、TempoCoder（時間符号化機能）、GeoCoder（地理符号化機能）。Semantic Metadata Generator と Keyword Extractor は、ドキュメント内の重要単語、該当する専門分野の専門用語、記号、文字列を、パターンマッチと出現頻度計算により言語メタデータとして抽出する。GeoCoder と TempoCoder は、ドキュメント内の地理情報および日付・時間情報を数値データに変換し、メタデータベースに格納する。

2. Evaluator（評価機構）：ユーザにより指定された意味・キーワード・時間・空間の各コンテキストとドキュメントから抽出されたメタデータとの関連性を計量する。各コンテキストとドキュメントの各コンテキスト対応メタデータは、以下の機能群により計量される：Semantic Calculator（意味的関連性計量機能：F1）、Boolean Evaluator（パターンマッチ機能：F2）、Spatial Evaluator（空間的距離計量機能：F3）、Temporal Evaluator（時間的距離計量機能：F4）、Global Evaluator（総合評価機能：F5）。分析対象ドキュメント・データベースが指定され、検索・分析条件クエリとして意味・キーワード・時間・空間の各コンテキストがユーザにより指定されると、Evaluator は全ての対象ドキュメントについてコンテキスト毎に計量を行い、結果を統合して出力する。Semantic Calculator には、意味的連想検索方式 [1]-[4]を適用する。

3. Visualizer（可視化機能：V）：Evaluator により出力されたドキュメント分析結果を、意味・時間・空間コンテキストに応じた表現として 5D 世界地図上に写像し、描画・可視化する。

4. Controller（クエリ管理機能：C）：ユーザが発行する意味・キーワード・時間・空間の各コンテキストをシステム内で処理可能な形に変換し、管理する機能。

3. ドキュメントデータを対象とした実現方式

3.1 意味・キーワード・時間・空間コンテキストについてのメタデータ抽出

分析対象データベース内の各ドキュメントについて、以下の機能群によりメタデータを抽出する：意味メタデータ抽出機能（Semantic Metadata Generator）、キーワード抽出機能（Keyword Extractor）、時間符号化機能（TempoCoder）、地理符号化機能（GeoCoder）。Semantic Metadata Generator と Keyword Extractor は、ドキュメント内の重要単語、該当する専門分野の専門用語、記号、文字列をメタデータとして抽出する。GeoCoder と TempoCoder は、ドキュメント内の地理情報および日付・時間情報を数値データに変換し、メタデータベースに格納する [6][7]。

3.2 意味・キーワード・時間・空間についての評価機構およびグローバル・アナライザ

Evaluator (F1, F2, F3, F4, F5) は、意味・キーワード・

時間・空間的計量評価機能の組み合わせにより、歴史・文化関連のドキュメント間の動的な分析・連結を行う。分析対象ドキュメント・データベースが指定され、検索・分析条件クエリとして意味・キーワード・時間・空間の各コンテキスト (CX1, CX2, CX3, CX4) がユーザにより指定されると、Evaluator は全ての対象ドキュメントについてコンテキスト毎に計量を行い、結果を統合して出力する。

3.3 動的マッピング・可視化機能: Visualyzer (V)

Visualyzer は、Global Analyzer の出力結果であるドキュメント集合について、Semantic Calculator によるドキュメントの意味的関連性計量のランキング結果を伴いながら、時系列順に並べられた 2 次元世界地図上へ写像・可視化する。本機能によって可視化された情報検索・分析の結果を得ることで、ユーザはより直観に近い形で情報源全体の概観を把握することが出来る。

3.4 クエリ処理・管理機能: Controller (C)

Controller は、ユーザが発行する意味・キーワード・時間・空間の各コンテキストをシステム内で処理可能な形に変換し、管理する。この機能により、ユーザは各自の関心領域に応じた対象ドキュメント群を選択し、各自の視点・問題領域に応じた意味的関連性計量結果を、各自が設定した地域と期間に応じて複数ビューに表示された世界地図として獲得することができる。

以上の機能群により実現された本システムを用いることにより、ユーザは、大量のドキュメントデータを対象として、情報源別・地域別・時間別の専門知識が反映された意味的分析結果を、多様な視点から動的に獲得することが可能となる。

4. 新聞記事データへの適用と実験システム構築

日本経済新聞社が所有する 2007 年～2017 年の 10 年分の環境関連記事[19]を用いて意味的連想検索[1]-[4]のためのメタデータ空間の生成[21]および記事検索実験システムを構築した。

第一に、記事に予め与えられている環境に関する分類コードおよびキーワードタグを用いて、分析対象の環境関連ドキュメント群を抽出した。

- 期間: 2007/09/01 ~ 2017/08/31
- 分類コード[20]: 自然災害(#W 5 0 2 0 4)/環境問題(#W 5 0 4 0 4)/エネルギー問題(#W 5 0 4 0 2)
- 抽出用キーワード: 214 語 (ヒアリ, デング熱, …, 干ばつ, 土砂崩れ, …, 大気汚染物質, 煤煙, 等)
- 該当件数: 289,825 件

第二に、環境関連ドキュメント全体 289,825 件を対象に、キーワードタグとして「地球温暖化」「温暖化」が予め付与されている対象ドキュメント群をメタデータ空間生成用ドキュメント(温暖化関連ドキュメント)として設定した。

第三に、統計的処理により、特徴語抽出を行った。「地球温暖化」「温暖化」に関連する対象ドキュメント(温暖化関連ドキュメント)全体のキーワードに対して、共起性の高く、偏りの少ないキーワードを特徴語

として抽出した。各プロセスは以下の通りである。

STEP 1: キーワード「地球温暖化」「温暖化」で対象ドキュメントを選択

結果⇒温暖化関連ドキュメント数 5,927 件

STEP 2: 温暖化関連ドキュメント群に出現するキーワードをカウント

結果⇒全出現キーワード数 9,925 語

STEP 3: 出現キーワードの出現頻度を計測

結果⇒出現回数 1-10 のキーワード数は 9,088 語

出現回数 11-1000 のキーワード数は 833 語

STEP 4: 出現回数に閾値を設定し、特徴語を抽出

結果⇒暫定的に、出現回数 11 を閾値として 833 語を抽出

STEP 5: 各特徴語と温暖化関連ドキュメント群内の他の特徴語との共起頻度を計測し、閾値以上の共起性がある語を選択し、語と語の関連を示すメタデータ空間生成用マトリクス M を生成[21] (図 3)

結果として、以下の実験環境を構築し、SQLite と Python, PHP を用いて Web アプリケーションとして検索インタフェースを実装した。

- 検索に使用可能な文脈語: 9,925 語
- 特徴空間次元数: 約 833 次元
- 検索対象記事数: 5,927 件

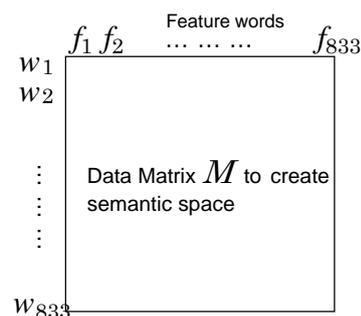


図 3 温暖化関連ドキュメント群を対象とした意味的連想検索のためのメタデータ空間生成用マトリクス M

出現キーワード, 出現頻度

地球温暖化,	5941	米国政府,	366
温暖化対策,	2673	首相,	366
温暖化ガス,	1930	ポスト京都議定書,	359
温暖化ガス削減,	1903	気候変動,	334
政府,	888	会合,	334
CO2削減,	753	経済産業省,	314
京都議定書,	728	EU,	314
温暖化,	664	環境保全,	312
国連気候変動枠組み条約締約国会議,	628	先進国,	301
米国,	610	環境対策,	298
CO2,	607	エネルギー,	292
温暖化ガス排出量,	574	欧州,	279
国連,	552	国際会議,	268
環境省,	535	排出量,	263
中国,	493	主要国首脳会議,	260
省エネルギー,	435	環境問題,	230
排出量取引,	433	大統領,	221
CO2排出量,	396	太陽光発電,	202
発展途上国,	379	産業界,	201
CO2排出量,	396	パリ協定,	200
発展途上国,	379	:	
		:	

図 4 温暖化関連ドキュメント群から抽出されたメタデータ空間先生用特徴語 (メタデータ) 一覧 (一部)

5. 検索実験と適用例

5.1 時空間マッピング・分析的可視化機能の検証

4. において構築した 5D World Map 実験システムを用い、日経新聞社が所有する 2007 年～2017 年の地球温暖化関連記事 5,927 件の記事データを対象として、本システムの時空間マッピング・分析的可視化機能の有効性を確認した。図 4～図 6 は、検索キーワードを「鳥インフルエンザ」に設定したパターンマッチ全文検索の結果、図 7、図 8 は「台風」に設定した同結果を示している。



図 4 対象期間：2007 年 9 月～12 月、検索キーワード「鳥インフルエンザ」によるパターンマッチ全文検索結果

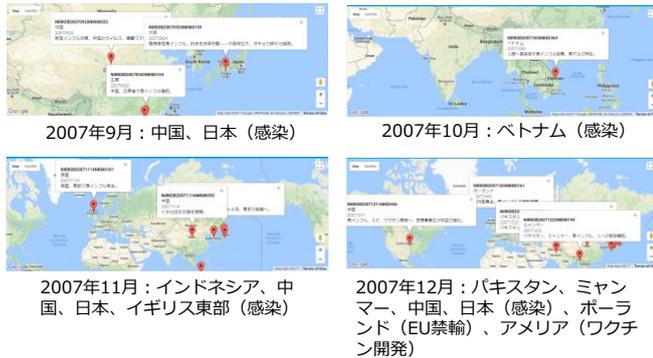


図 5 対象期間：2007 年 9 月～12 月の 1 ヶ月ごと、検索キーワード「鳥インフルエンザ」による検索結果

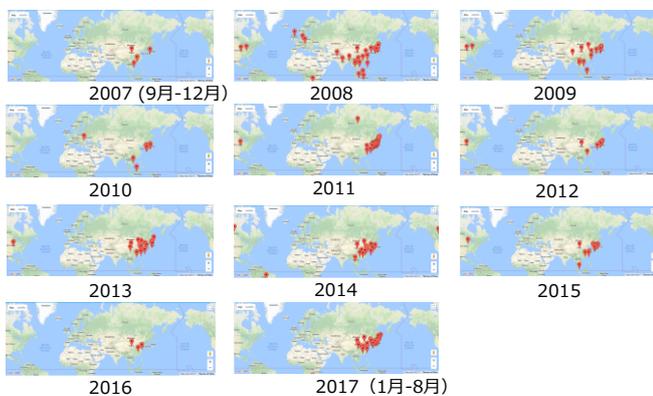


図 6 対象期間：2007 年～2017 年の 1 年ごと、検索キーワード「鳥インフルエンザ」による検索結果

図 5 は、2007 年 9 月～12 月の 1 ヶ月ごとの記事群を対象とした検索結果を示しており、この結果から 2007 年 9 月にその年最初の感染が中国と日本で確認され、10 月～11 月に東南アジアとイギリス東部、12 月には南アジアに感染が広がり、EU では欧州委員会が禁輸を決定、最終的に米国にてワクチン開発が始まった経緯が把握できる。図 6 は、2007 年～2017 年の 1 年ごとの記事群を対象とした検索結果を示しており、この結果からは、記事の地理的分布と年毎の記事数の増減傾向から、世界的な感染拡大が確認された年が 2008 年であること、東南アジア地域では 2011 年、2014 年、2017 年に多く感染現象が報告されていることが発見的に分かる。これらの結果から、本機能の“ビジュアルデータマイニングツール”としての有効性も確認できる。



図 7 対象期間：2017 年 1 月～8 月、検索キーワード「台風」(マーカの色相は記事に付与されているテーマ(企業・政治・経済・技術・社会)を表し、彩度は対象期間内での新規度を表す.)

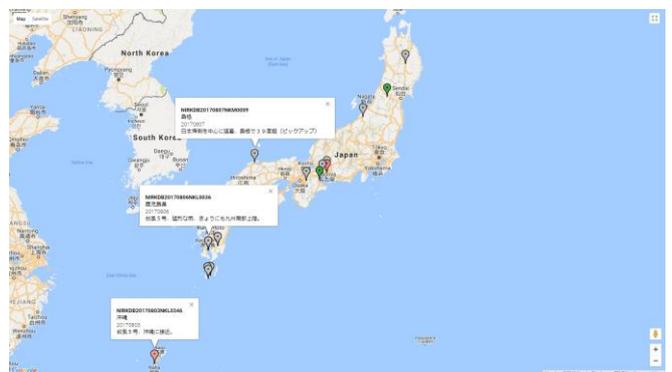


図 8 対象期間 2017 年 8 月 1 日～8 日、検索キーワード「台風 5 号」

図 7 の結果より、利用者は、2017 年 1 月～8 月の 8 か月間に、台風 3 号 (7 月)、5 号 (8 月) が日本本島に上陸したこと、また、1 月には前年の台風 10 号の

被害が残っていること等を見てとることができる。

さらに図8に示すように、台風5号に着目し、対象期間を2017年8月1日～8日の一週間に設定すると、台風5号が7月1日沖縄に接近し、8月6日九州南部に上陸、8月7日には日本海側を中心に39度超の猛暑を引き起こした現象について、その進路・経路と共に把握することができる。

以上のように、本システムは、時間的文脈、空間的文脈のダイナミックな切り替えと組み合わせにより、記事データを入口として、自然・社会事象の時間的遷移と空間的な移動の様子を俯瞰的に把握することを可能とする。特に、世界的な感染症や自然災害、異常気象といった時間的・空間的広がりを持つ事象の分析に本システムの時空間マッピング機能が有効であることを示している。

5.2 意味的連想検索機能と時空間マッピング機能の統合による効果の検証



図9 パターンマッチ全文検索による記事検索結果の5D World Map 上での可視化結果（対象期間：2015年1月～12月、キーワード「異常気象」、検索結果数全11件中、位置情報を持つ4件を地図上に可視化）

context: 異常気象

the number of results: 100

results:	ID	Location	Description
1	0.896095	N1R1KDB201411104NKE0258	気象研究所 気象庁 気温 気象 温暖化]
2	0.895956	N1R1KDB20170321NKE0198	WMO 異常気象 気象 温暖化]
3	0.895804	N1R1KDB20150722NKM0023	気象庁 気象 温暖化]
4	0.890289	N1R1KDB20160410NKM0280	北極 北極海 氷 気象 温暖化]
5	0.889751	N1R1KDB20150710NKM0185	エルニーニョ 予測 海洋研究開発機構 気象庁 気温 温暖化]
6	0.887503	N1R1KDB20170621NKE0491	ヨーロッパ WMO アメリカ合衆国 気象 温暖化]
7	0.887186	N1R1KDB20130505NKM0022	進行 気象研究所 予測 気象庁 温暖化]
8	0.884202	N1R1KDB20131225NKM0033	平均気温 気象庁 温暖化]
9	0.883592	N1R1KDB20130523NKS0075	海洋研究開発機構 温暖化]
10	0.883592	N1R1KDB20140911NKS0058	海洋研究開発機構 温暖化]
11	0.883133	N1R1KDB20150403NKE0576	南極 気温 温暖化]
12	0.882687	N1R1KDB20111117NKE0368	原因 南極 温暖化]
13	0.882403	N1R1KDB20140527NKM0025	大気汚染 気象庁 二酸化炭素 温暖化]
14	0.881698	N1R1KDB20121124NKM0028	南極 温暖化]
15	0.881698	N1R1KDB20140630NKE0236	南極 温暖化]
16	0.881592	N1R1KDB20150128NKS0065	海水 研究開発 温暖化]
17	0.880971	N1R1KDB20130714NKM0019	気象庁 気温 温暖化]
18	0.880860	N1R1KDB20121118NKM0042	気象庁 温暖化]
19	0.880860	N1R1KDB20150501NKM0491	気象庁 温暖化]
20	0.878855	N1R1KDB20150215NKM0060	魚 北極海 海洋研究開発機構 温暖化]
21	0.878368	N1R1KDB20120531NKE0393	海水 北極海 チーム 北極海 海洋研究開発機構 温暖化]
22	0.878205	N1R1KDB20140514NKS0075	研究開発 NASA 海面上昇 氷 温暖化]
23	0.878039	N1R1KDB20170831NKS0066	農業 食品産業技術総合研究機構 研究チーム 調査統計 温暖化]
24	0.877626	N1R1KDB20160406NKS0060	ネイチャー 調査統計 チーム 温暖化]
25	0.877032	N1R1KDB20110127NKE0448	地図 北極海 氷 温暖化]
26	0.876454	N1R1KDB20140201NKM0356	二酸化炭素 温暖化]
27	0.876145	N1R1KDB20111204NKM0358	気象庁 気温 温暖化]
28	0.874639	N1R1KDB20130507NKE0026	今世紀 気象研究所 チーム 温暖化]
29	0.874289	N1R1KDB20111116NKM0294	気温 温暖化]
30	0.873451	N1R1KDB20120423NKM0179	北極 温暖化]

図10 意味的連想検索を用いた文脈語「異常気象」による記事検索結果、上位100件

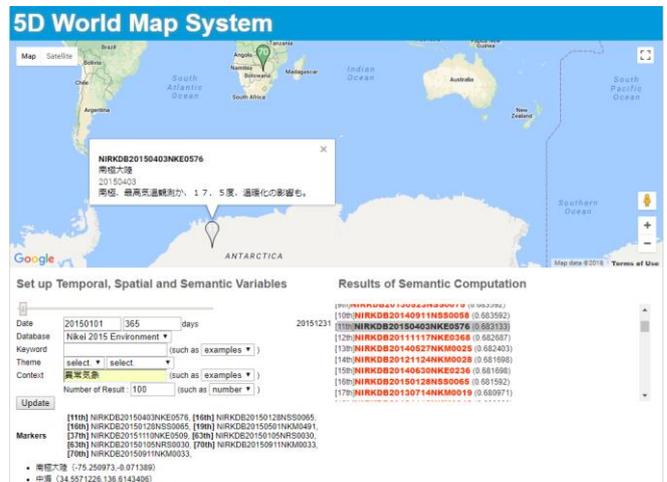


図11 図10の意味的連想検索による記事検索結果と時空間マッピングの統合による5D World Map 上での可視化結果（対象期間：2015年1月～12月、文脈語「異常気象」、上位100件）

図10は、意味的連想検索を用いた文脈語「異常気象」による記事検索結果、上位100件を示し、図11は、その結果の5D World Map 上での可視化結果を示している。意味的連想検索において11位に検索された記事(N1R1KDB20150403NKE0576)は、「南極大陸における最高気温観測」に関する記事であり、「異常気象」という語そのものをメタデータとして持っていないにも関わらず、異常気象に意味的に関連する記事として上位に検索されたことを示している。この記事は、図9に示したパターンマッチ検索では検索できない。

また、意味的連想検索にて上位100件に検索された記事のうち、「異常気象」をメタデータとして持っている記事は4件のみであったが、気象関連のメタデータを持っている記事の件数は、「気象」については29件、「海（海面上昇・北極海・海洋など）」29件、「北極」15件、「気温」13件、「氷」12件、「南極」7件であったことから、「気温の上昇による海面上昇やそれに伴う気象の変化」に関する記事群が発見的に検索されることが分かる。

context: 国際枠組み

the number of results: 100

rank	id	title
1	0.957904	NIRKCE01161202NS0067
2	0.957458	NIRKCE01170731NK0137
3	0.952100	NIRKCE01170623NK0544
4	0.950010	NIRKCE01170611NK0133
5	0.950010	NIRKCE01170529NK0057
6	0.844873	NIRKCE01170510NK0592
7	0.842231	NIRKCE01170623NK0149
8	0.842277	NIRKCE01170724NK0132
9	0.841574	NIRKCE01170601NK0527
10	0.841158	NIRKCE01170611NK0100
11	0.838978	NIRKCE01170529NK0044
12	0.838915	NIRKCE01170623NK0058
13	0.835720	NIRKCE01160529NK0072
14	0.831629	NIRKCE01160904NK0037
15	0.829008	NIRKCE01170715NK0054
16	0.822437	NIRKCE01170602NK0588
17	0.822272	NIRKCE01170723NK0112
18	0.814157	NIRKCE01170623NK0149
19	0.813759	NIRKCE01170709NK0027
20	0.808417	NIRKCE01171129NK0566
21	0.805422	NIRKCE01170529NK0104
22	0.595482	NIRKCE01170602NK0531
23	0.595423	NIRKCE01170529NK0579
24	0.592249	NIRKCE01170603NK0542
25	0.587424	NIRKCE01170513NK0069
26	0.587414	NIRKCE01170531NK0522
27	0.586386	NIRKCE01161129NK0483
28	0.584957	NIRKCE01171129NK0144
29	0.576375	NIRKCE01160130NK0178
30	0.573890	NIRKCE01161129NK0074

図 12 意味的連想検索を用いた文脈語「国際枠組み」による記事検索結果，上位 100 件

図 12 は，意味的連想検索を用いた文脈語「国際枠組み」による記事検索結果，上位 100 件を示し，図 13，図 14 は，その結果の 5D World Map 上での可視化結果を示している。上位 100 件に検索された記事のうち，「国際枠組み」をメタデータとして持っている記事は 4 件のみであったが，国際合意・協定関連のメタデータを持っている記事の件数は，「米（米国・米大統領・米国政府・アメリカ合衆国など）」については 78 件，「パリ協定」41 件，「気候変動・気候変動枠組み条約」41 件，「離脱」16 件であったことから，上位に検索された記事群は，“米国大統領の方針により米国が気候変動枠組み条約パリ協定を離脱した”事象に関する記事が多いこと，またその多くは 2016 年から 2017 年にかけて発行された記事が多いことが分かる。



図 13 図 12 の意味的連想検索による記事検索結果と時空間マッピングの統合による 5D World Map 上での可視化結果（対象期間：2017 年 1 月～12 月，文脈語「国際枠組み」，上位 100 件）

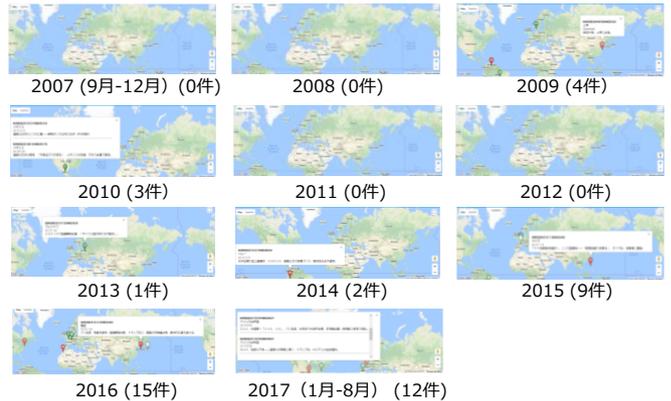


図 14 図 12 の意味的連想検索による記事検索結果と時空間マッピングの統合による 5D World Map 上での可視化結果（対象期間：2007 年～2017 年の 1 年ごと，文脈語「国際枠組み」），および，地図上での表示記事数

表 1 パターンマッチによる記事検索結果と意味的連想検索による記事検索結果（上位 100 件）の比較（キーワードおよび文脈語「国際枠組み」）

発行年	パターンマッチによる検索結果件数 (位置情報あり記事)	意味的連想検索による検索記事数 (上位100件)	意味的連想検索による検索記事数 (上位100件中、位置情報あり記事)
2007	0	1	0
2008	0	0	0
2009	0	12	4
2010	0	3	3
2011	3	2	0
2012	1	2	0
2013	2	3	1
2014	0	9	2
2015	3	16	9
2016	3	21	15
2017	4	31	12

表 1 は，キーワードを「国際枠組み」としてパターンマッチによって検索された記事数と，文脈語を「国際枠組み」として意味的連想検索によって検索された記事数（上位 100 件）を比較したものである。パターンマッチでは，各年通じて検索数 0 件や数件のみがヒットしているのに対し，意味的連想検索を用いた場合，関連する記事が多く検索されていることが分かる。また，新しい気候変動枠組みについての議論が始まった 2009 年に記事が多いこと，さらに，経年に従って関連記事が増加している傾向から，パリ協定が採択された 2015 年末から 2016 年の発効，2017 年の米国離脱という流れから国際的議論が加速した現象変化が把握できる。

6. 結論と今後の課題

本論文は，大規模新聞記事群を対象とした文脈多様性対応型意味的連想検索と 5 次元世界地図システム（5D World Map System）への適用方式を示した。本システムは，時空間・意味の多様なコンテキスト（文脈）のダイナミックな組み合わせにより，記事内に潜在的

に含まれている事象間の相関関係／因果関係を発見的に抽出し、時間軸を持った世界地図上に可視化する新しい新聞記事閲覧・分析環境である。本方式により、多様な文脈に応じて意味的に近い情報を検索するための検索環境、および、事象や文脈自体の時系列変化・空間的拡大を分析するための分析環境を提供することができることを示し、環境分野、特に「地球温暖化」に関する新聞記事 10 年分を対象とした検索・可視化実験により、本方式の有効性を示した。

今後は、本研究の対象となる環境関連ドキュメント全 289,825 件を対象として文脈語全 9,925 語を用いた定量的な精度実験を行うと共に、複数の被験者を対象とした実証実験を行うことで、本システムの有効性・実用性を検証したい。さらに、画像・動画などの環境関連マルチメディアデータ、および、気象・統計データなどのセンシングデータとの統合による総合的な新聞記事閲覧・分析環境の実現を目指す。

参 考 文 献

- [1] Kitagawa, T. and Kiyoki, Y. : 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.
- [2] Kiyoki, Y. Kitagawa, T. and Hayama, T. : A metadata system for semantic image search by a mathematical model of meaning, *ACM SIGMOD Record*, Vol. 23, No. 4, pp.34-41, 1994.
- [3] Kiyoki, Y., Kitagawa, T. and Hitomi, Y. : 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
- [4] 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.
- [5] Kiyoki, Y. & Chen, X., "Contextual and differential computing for the multi-dimensional world map with context-specific spatial-temporal and semantic axes", *Information Modelling and Knowledge Bases XXV*. IOS Press, Vol. 260, p. 82-97, 2014.
- [6] Kiyoki, Y., Chen, X., Sasaki, S. and Koopipat, C., "Multi-Dimensional Semantic Computing with Spatial-Temporal and Semantic Axes for Multi-spectrum Images in Environment Analysis", *Information Modelling and Knowledge Bases*, Vol. XXVIII, IOS Press, pp. 14-30, May 2016.
- [7] 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, IOS Press, pp. 1 - 18, May 2010.
- [8] Sasaki, S. and Kiyoki, Y., "Real-time Sensing, Processing and Actuation Functions of 5D World Map System: A Collaborative Knowledge Sharing System for Environmental Analysis" *Information Modelling and Knowledge Bases*, Vol. XXVIII, IOS Press, pp. 220-239, May 2016.
- [9] Fujioka, H., Sasaki, S. and Kiyoki, Y., A Realtime Sensing-Data Triggered News Article Provision System with 5D World Map, *Proc. of Int'l Electronics Symposium (IES2017)*, IEEE, Surabaya, Indonesia, 8 pages, Sep. 26-27, 2017.
- [10] Batini, C. et al, 1986. A Comparative analysis of methodologies for database schema integration. *ACM Computing Surveys*, Vol. 18, No.4, pp.324-364.
- [11] Egenhofer, M. J. et al, 1998. Metric Details for Natural-Language Spatial Relations. *ACM Trans. Information Systems*, Vol.16, No. 4, pp.295-321.
- [12] Egenhofer, M. J., 1994. Spatial SQL: A Query and Presentation Language. *Knowledge and Data Engineering*, IEEE Transactions on Volume 6, Issue 1, pp. 86 - 95.
- [13] Guting, R. H., 1994. An Introduction to Spatial Database Systems. *VLDB Journal*, 3, pp. 357-399.
- [14] Litwin, W., 1985, An overview of the multidatabase system MRDSM. *Proceedings of the 1985 ACM annual conference on The range of computing*, pp.524-533.
- [15] Litwin, W. et al, 1990. Interoperability of multiple autonomous databases. *ACM Computing Surveys (CSUR)*, Volume 22 Issue 3, pp.267-293.
- [16] Hosokawa Y. and Kiyoki Y., 1999. Functional and parallel query processing and query optimization for multidatabase systems. *Proc. the 17 th IASTED International Conference on Applied Informatics*, pp.101-106.
- [17] Kiyoki, Y. et al, 2000. A Metadata 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.
- [18] Ishibashi, N. and Kiyoki, Y., 2004. 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.
- [19] 日経テレコン : <http://telecom.nikkei.co.jp/>
- [20] 日経テレコン分類語一覧 : <http://t21help.nikkei.co.jp/reference/cat845/post-563.html>
- [21] 藤岡 華子, 佐々木 史織, 渡邊 敏央, 大塚 恭平, 石井 昌之, 清木 康: "コンテキスト解釈機構を有する意味的計量系の適用による環境ニュース記事データベース検索システムの実現", 第 10 回データ工学と情報マネジメントに関するフォーラム (DEIM2018), 福島県あわら市, 8 ページ, 2018 年 3 月 3-6 日.