

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

藤岡 華子[†] 佐々木 史織[†] 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 〒252-0805 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: [†] {hanafuji, shashioir, kiyoki}@sfc.keio.ac.jp,

[‡] {toshihiro.watanabe, kyohei.otsuka, masayuki.ishii}@nex.nikkei.com

あらまし 本論文では、環境分野の新聞記事を対象としたドキュメント検索において、意味の数学モデルを適用して意味的連想検索を行う際のメタデータ空間の生成方式を示し、多様な文脈を持つ単語やドキュメント間の関連性を計量する本意味的連想検索実現方式の有効性を検証する。具体的には、地球温暖化について書かれた新聞記事 10 年間分を用いて、それらの新聞記事を持つ単語間の意味的関連性を計量するベクトル空間を生成する。本方式によって生成したメタデータ空間を用いることにより、特定の分野に関する情報群に対して、多様な文脈に応じて意味的に近い単語や新聞記事を検索することが可能となる。本論文ではまた、文脈による検索結果の違いを複数比較することで性能評価を行う。

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、環境、地球温暖化、気候変動

1. はじめに

現在、WWW 上および新聞記事等のデジタルアーカイブには、政治経済に関連する膨大な文書データが存在する。それらの情報群から必要な情報を適切に獲得するためには、フィルタリングや検索エンジン等のナビゲーションシステムが必要である。

一般的な検索エンジンで用いられるパターンマッチ検索方式は、キーワードの単純なパターン照合による検索であるため、1 つの単語に複数の意味がある場合や、異なる単語が同じ意味を持つ場合に、これらの意味的解釈を伴った検索は困難である。そのため、過去から蓄積された言葉と言葉の関係を記述した膨大なソーラスやオントロジーを用いて、関連データを引き出そうとしているのが現状である。

これに対し、意味の数学モデルによる意味的連想検索方式など、ベクトル空間を用いて言葉の意味を扱う検索方式が提案されている。これは、言葉と言葉の関係を直接記述するのではなく、言葉の意味をマトリクス上で数値化し、直行ベクトル空間を形成し、「部分空間の選択」によって、言葉間の関係や言葉の持つ文脈に応じた多様な意味を計算することができる。

意味の数学モデルを用いたベクトル空間を生成するためには、対象のデータ群の特徴を表したメタデータを選定する必要がある。

本論文では、地球温暖化の分野に限定した新聞記事を対象としたドキュメント検索において、意味の数学モデルを適用して意味的連想検索を行う際のメタデータ空間の生成方式を示し、多様な文脈を持つ単語やドキュメント間の関連性を計量する本意味的連想検索実現方式の有効性を検証する。

2. 基本方式

本論文では、環境分野の新聞記事を用いてメタデータ空間を作成し、意味的連想検索方式[1][2][3][4][5]を適

用することで、環境ニュース記事データベース検索システムを実現する。意味的連想検索方式は単純なパターンマッチではなく、情報の意味に着目することで、文脈に応じて意味的に近い情報を検索できる。また、本環境ニュース記事データベース検索システムは、大規模な環境関連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の実現[11]を目的として、「5次元世界地図システム (“5D World Map System”）」[6][7][8][9][10]との連結を実現する。

3. 実現方式

日本経済新聞が過去 10 年間に発行した新聞記事[12]と、各記事に対して日本経済新聞社があらかじめ付与した分類コード[13]およびキーワードタグを用いて、下記の 5STEP で「地球温暖化」メタデータ空間を作成する。

STEP1: 全新聞記事の中で、下記の条件を満たす新聞記事を抽出する。結果の新聞記事は 289,825 件であった。

- ・ 期間：2007 年 9 月 1 日～2017 年 8 月 31 日
- ・ 分類コード：自然災害(#W50204)/環境問題(#W50404)/エネルギー問題(#W50402)
- ・ 抽出キーワードタグ：214 語

実際の抽出クエリは下記。

ヒアリ or デング熱 or 鳥インフルエンザ or エボラ出血熱 or 感染症 or 院内感染 or ウイルス感染症 or インフルエンザ or エイズ or 風邪 or 肝炎 or 劇症肝炎 or 慢性肝炎 or B型肝炎 or C型肝炎 or 狂犬病 or はしか or 風しん or ヘルペス or ヤコブ病 or SARS or 細菌感染症 or 結核 or 敗血症 or 破傷風 or ハンセン病 or 真菌症 or 手足口病 or 伝染病 or コレラ or ジフテリア or 赤痢 or 天然痘 or 西ナイル熱 or 日本脳炎 or ポリオ or マラリア or 太陽光発電 or 太陽熱発電 or ソーラーシステム or 自然エネルギー or グリーンエネルギー or 地熱エネルギー or 太陽エネルギー or 風力エネルギー or EV or 低公害

車 or アルコール車 or エタノール車 or メタノール車 or 水素自動車 or 天然ガス自動車 or 電気自動車 or ソーラーカー or 燃料電池車 or ハイブリッド車 or LPG車 or 気象 or 梅雨 or 梅雨前線 or 異常気象 or 寒波 or 暖冬 or 熱波 or 猛暑 or 冷夏 or エルニーニョ or オーロラ or 季節風 or 春一番 or 気圧 or 高気圧 or 気候変動 or 黄砂 or 桜前線 or 湿度 or 竜巻 or 熱帯夜 or 白夜 or フェーン現象 or 真夏日 or 初雪 or 乱気流 or 震度 or マグニチュード or 高潮 or 潮流 or 暖流 or 流水 or 自然災害 or 液状化現象 or 火山災害 or 火砕流 or 火山灰 or 土石流 or 噴火 or 干ばつ or 地震 or 群発地震 or 震災 or ジャワ島地震 or スマトラ沖地震 or 東海地震 or 新潟県中越地震 or 阪神大震災 or 福岡県西方沖地震 or 地割れ or 雪害 or 豪雪 or 津波 or インド洋大津波 or ジャワ津波 or 土砂災害 or 地滑り or 土砂崩れ or 土石流 or 雪崩 or 日照不足 or 病虫害 or 風水害 or 洪水 or 集中豪雨 or 台風 or サイクロン or ハリケーン or 暴風雨 or 床上浸水 or 床下浸水 or 水不足 or 山火事 or 落石 or 落雷 or 冷害 or 環境破壊 or 油流出 or 原油流出 or オゾン層破壊 or オゾンホール or 海洋汚染 or 海洋投棄 or 砂漠化 or 森林破壊 or 大気汚染 or 酸性雨 or 地球温暖化 or 土壌汚染 or ヒートアイランド現象 or 環境問題 or ごみ問題 or 不法投棄 or 公害 or アイドリング or 悪臭 or 塩害 or 光化学スモッグ or 残留農薬 or 食品公害 or 地盤沈下 or 水質汚濁 or 赤潮 or 富栄養化 or 騒音 or 航空機騒音 or 交通騒音 or 薬品公害 or 公害訴訟 or 公害病 or カネミ油症 or 水俣病 or 地球環境 or 廃棄物 or 医療廃棄物 or オガクズ or 産業廃棄物 or 家畜廃棄物 or 鉱さい or 産業排水 or 工場排水 or 事業系ごみ or スラッジ or 廃材 or 廃船 or 廃タイヤ or 排土 or 廃プラ or 廃油 or フェロシルト or ヘドロ or 焼却灰 or ごみ焼却灰 or 生活廃棄物 or 空き缶 or 空き瓶 or ごみ or 粗大ごみ or 生ごみ or 不燃ごみ or し尿 or 生活排水 or 廃家電 or もみがら or 六価クロム or 有害物質 or 環境ホルモン or ダイオキシシン or PCB or SO₂ or 大気汚染物質 or 排煙 or 排ガス or 粉塵 or NO_x or NO₂ or SO_x or bunruisub=#W50204 or bunruisub=#W50404 or bunruisub=#W50402

STEP2: STEP1 で抽出した新聞記事の中で、キーワードタグに「地球温暖化」または「温暖化」を持つ新聞記事を抽出する。結果の新聞記事は 5,927 件であった。

STEP3: STEP2 で抽出した新聞記事が持つ全キーワードタグを抽出し、「地球温暖化」メタデータ空間を表現するメタデータ群として設定する。結果のメタデータは 9,925 語であった。

STEP4: STEP3 で抽出した各メタデータが、STEP2 で抽出したいくつの新聞記事に対して付与されているかをカウントした。暫定的にカウントが 11 以上 1000 以下のメタデータを、「地球温暖化」メタデータ空間を表現する特徴語群として設定する。結果のメタデータは 833 語であった。

STEP5: 縦軸に STEP3 で設定したメタデータ群を、横軸に STEP4 で設定した特徴語群を持つ「地球温暖化」メタデータ空間を作成し、各新聞記事が持つキーワードの中で縦軸の語と横軸の語が 1 件でも共起していれば 1 を、していなければ 0 を立てる。

4. 実験

本論文で提案した方式に基づいて、日本経済新聞の過去 10 年分の記事を用いて構築された「地球温暖化」メタデータ空間における意味的連想検索の実験結果について、あらかじめ設定した正解セットとの再現率を評価し、その有効性を検証する。

なお本意味的連想検索システムは、検索語句と検索コンテキストを指定して単語を検索する機能と、検索コンテキストを指定してドキュメントを検索する機能を有している。

4.1. 実験 1: 単語検索(コンテキスト固定)

ここでは、コンテキストを CO₂ に関する単語で固定し、複数の事なる検索語で検索を行う事で、メタデータ空間が正しく生成されていることを検証する。

「地球温暖化」メタデータ空間を、CO₂ に関するコンテキストで部分空間を切り取ることで、それぞれの検索語と近い意味を持つと期待される単語(正解セット)が、検索結果として出現することを検証する。正解セットは、それぞれ 15 件とした。

正解セット A

温暖化に関する災害系の単語(自然災害、温度、猛暑、熱中症、干ばつ、洪水、豪雨、ゲリラ豪雨、集中豪雨、海面上昇、高潮、寒波、台風、ハリケーン、エルニーニョ)を正解セットとして設定する。

正解セット B

温暖化に関するエネルギー系の単語(エネルギー政策、エネルギー開発、代替エネルギー、クリーンエネルギー、自然エネルギー、電力業界、火力発電、石炭火力、化石燃料、太陽光発電、原子力発電、原子力、地熱発電、スマートグリッド、代替フロン)を正解セットとして設定する。

正解セット C

温暖化に関する製品系の単語(商品開発、省エネルギー、家電、エアコン、太陽電池、燃料電池、燃料電池車、電気自動車、ハイブリッド車、バイオマス、バイオ燃料、バイオエタノール、ソーラーシステム、ヒートポンプ、CSR)を正解セットとして設定する。

それぞれの検索語で上位 100 位までを検索し、正解セットの再現率を計算して性能を評価した。

4.1.1. 検索例 1: 検索語「温暖化」、コンテキ

スト「CO₂、CO₂対策、CO₂削減、CO₂抑制、CO₂排出、CO₂排出権、CO₂排出量、CO₂排出量取引、CO₂放出量、CO₂濃度、CO₂貯留、CO₂地下貯留」

はじめに、抽象度の高い「温暖化」という検索語を用いて、単語を意味的検索した結果を示す。

コンテキスト「CO₂、CO₂対策、CO₂削減、CO₂抑制、CO₂排出、CO₂排出権、CO₂排出量、CO₂排出量取引、CO₂放出量、CO₂濃度、CO₂貯留、CO₂地下貯留」における、検索語「温暖化」と近い単語を検索した結果を Table 1 に示す。この結果から、CO₂ 関連の文脈における検索語「温暖化」は、

「異常気象」「再生可能エネルギー」「エコビジネス」のいずれの正解セットとも重複しないことが分かる。すなわち、「温暖化」という抽象度の高い検索語では、「異常気象」「再生可能エネルギー」「エコビジネス」といった単語から連想される単語を連想検索することが難しいことがわかる。

Table 1: 検索語が「温暖化」、コンテキストが「CO₂、CO₂対策、CO₂削減、CO₂抑制、CO₂排出、CO₂排出権、CO₂排出量、CO₂排出量取引、CO₂放出量、CO₂濃度、CO₂貯留、CO₂地下貯留」の時の、各正解セットの重複率

正解セット	A	B	C
重複率	0%	0%	0%

4.1.2. 検索例 2：検索語「異常気象」、コンテキスト「CO₂、CO₂対策、CO₂削減、CO₂抑制、CO₂排出、CO₂排出権、CO₂排出量、CO₂排出量取引、CO₂放出量、CO₂濃度、CO₂貯留、CO₂地下貯留」

実験4.1.1を踏まえ、より具体的な検索語を用いることで、特定の文脈を定めた中で、検索語から連想される単語を連想検索できていることを確認する。

コンテキスト「CO₂、CO₂対策、CO₂削減、CO₂抑制、CO₂排出、CO₂排出権、CO₂排出量、CO₂排出量取引、CO₂放出量、CO₂濃度、CO₂貯留、CO₂地下貯留」における、検索語「異常気象」と近い単語を検索した結果を Table 2、Table 3 に示す。4.1.1と比較すると、検索語を「異常気象」で指定したことで、期待する正解セット A の再現率が 0% から 60% に上がっていることが分かる。また、この文脈では関係のない「再生可能エネルギー」と「エコビジネス」の文脈から想定される正解セットの単語はほとんど含まれておらず、部分空間がシャープに切り取られた結果となっている。

Table 2: 検索語が「異常気象」、コンテキストが CO₂ 関連語句の時の、期待する正解セット A の単語の再現率

正解セット	A
再現率	60%

Table 3: 検索語が「異常気象」、コンテキストが CO₂ 関連語句の時の、期待する正解セット以外の単語との重複率

正解セット	B	C
重複率	7%	0%

4.1.3. 検索例 3: 検索語「再生可能エネルギー」、コンテキスト「CO₂、CO₂対策、C

CO₂削減、CO₂抑制、CO₂排出、CO₂排出権、CO₂排出量、CO₂排出量取引、CO₂放出量、CO₂濃度、CO₂貯留、CO₂地下貯留」

コンテキスト「CO₂、CO₂対策、CO₂削減、CO₂抑制、CO₂排出、CO₂排出権、CO₂排出量、CO₂排出量取引、CO₂放出量、CO₂濃度、CO₂貯留、CO₂地下貯留」における、検索語「再生可能エネルギー」と近い単語を検索した結果を Table 4、Table 5 に示す。4.1.1と比較すると、検索語を「再生可能エネルギー」で指定したことで、期待する正解セット B の再現率が 0% から 87% に上がっていることが分かる。また、この文脈では関係のない「異常気象」と「エコビジネス」の文脈から想定される正解セットの単語は全く含まれておらず、部分空間がシャープに切り取られた結果となっている。

Table 4: 検索語が「再生可能エネルギー」、コンテキストが CO₂ 関連語句の時の、期待する正解セット B の単語の再現率

正解セット	B
再現率	87%

Table 5: 検索語が「再生可能エネルギー」、コンテキストが CO₂ 関連語句の時の、期待する正解セット以外の単語との重複率

正解セット	A	C
重複率	0%	0%

4.1.4. 検索例 4：検索語「エコビジネス」、コンテキスト「CO₂、CO₂対策、CO₂削減、CO₂抑制、CO₂排出、CO₂排出権、CO₂排出量、CO₂排出量取引、CO₂放出量、CO₂濃度、CO₂貯留、CO₂地下貯留」

コンテキスト「CO₂、CO₂対策、CO₂削減、CO₂抑制、CO₂排出、CO₂排出権、CO₂排出量、CO₂排出量取引、CO₂放出量、CO₂濃度、CO₂貯留、CO₂地下貯留」における、検索語「エコビジネス」と近い単語を検索した結果を Table 6、Table 7 に示す。4.1.1と比較すると、検索語を「エコビジネス」で指定したことで、期待する正解セット C の再現率が 0% から 100% に上がっていることが分かる。また、この文脈では関係のない「異常気象」と「再生可能エネルギー」の文脈から想定される正解セットの単語はほとんど含まれておらず、部分空間がシャープに切り取られた結果となっている。

Table 6: 検索語が「エコビジネス」、コンテキストが CO₂ 関連語句の時の、期待する正解セット C の単語の再現率

正解セット	C
再現率	100%

Table 7: 検索語が「エコビジネス」、コンテキストがCO2 関連語句の時の、期待する正解セット以外の単語との重複率

正解セット	A	B
重複率	0%	7%

4.1.5. 考察

これらの結果から、コンテキスト固定して検索語を変えることにより、その部分空間内で関連のある単語が検索できたことが分かる。

4.2. 実験 2: 単語検索 (検索語固定)

ここでは、検索語を「温暖化」に固定し、複数の異なるコンテキストで検索を行うことで、本意味的連想検索が切り取られた部分空間の中で単語の文脈依存性を計量できていることを示す。

「温暖化」は文脈により多様な意味を持っている。次の3つの特徴的なコンテキストで「地球温暖化」メタデータ空間の部分空間を切り取ることで、それぞれのコンテキストにおいて「温暖化」と近い意味を持つと期待される単語 (正解セット) が、検索結果として出現することを検証する。正解セットは、それぞれ 15 件とした。

正解セット A

温暖化に関する災害系の単語 (自然災害、温度、猛暑、熱中症、干ばつ、洪水、豪雨、ゲリラ豪雨、集中豪雨、海面上昇、高潮、寒波、台風、ハリケーン、エルニーニョ) を正解セットとして設定する。

正解セット B

温暖化に関するエネルギー系の単語 (エネルギー政策、エネルギー開発、代替エネルギー、クリーンエネルギー、自然エネルギー、電力業界、火力発電、石炭火力、化石燃料、太陽光発電、原子力発電、原子力、地熱発電、スマートグリッド、代替フロン) を正解セットとして設定する。

正解セット C

温暖化に関する製品系の単語 (商品開発、省エネ機器、家電、エアコン、太陽電池、燃料電池、燃料電池車、電気自動車、ハイブリッド車、バイオマス、バイオ燃料、バイオエタノール、ソーラーシステム、ヒートポンプ、CSR) を正解セットとして設定する。

それぞれのコンテキストで上位 100 位までを検索し、正解セットの再現率を計算して性能を評価した。

4.2.1. 検索例 1: 検索語「温暖化」、コンテキスト「温暖化」

コンテキスト「温暖化」における、検索語句「温暖化」と近い単語を検索した結果を Table 8 に示す。この結果から、「温暖化」というコンテキストにおける単語「温暖化」は、「再生可能エネルギー」や「エコビジネス」の文脈よりも「異常気象」の文脈から想定される正解セットの単語と重複する傾向にあることが分かる。

Table 8: 検索語が「温暖化」、コンテキストが「温暖化」の時の、各正解セットの重複率

正解セット	A	B	C
重複率	67%	7%	7%

4.2.2. 検索例 2: 検索語「温暖化」、コンテキスト「異常気象」

ここでは、コンテキスト「異常気象」における、検索語「温暖化」と近い単語を検索した結果を Table 9、Table 10 に示す。4.2.1 と比較すると、コンテキストを「異常気象」で指定したことで、期待される正解セット A の再現率が 67% から 73% に上がっていることが分かる。また、この文脈では関係のない「再生可能エネルギー」と「エコビジネス」の文脈から想定される正解セットの単語はほとんど含まれておらず、部分空間がシャープに切り取られた結果となっている。

Table 9: 検索語が「温暖化」、コンテキストが「異常気象」の時の、期待する正解セット A の単語の再現率

正解セット	A
再現率	73%

Table 10: 検索語が「温暖化」、コンテキストが「異常気象」の時の、期待する正解セット以外の単語との重複率

正解セット	B	C
重複率	7%	7%

4.2.3. 検索例 3: 検索語「温暖化」、コンテキスト「再生可能エネルギー、太陽光発電、風力発電」

コンテキスト「再生可能エネルギー、太陽光発電、風力発電」における、検索語「温暖化」と近い単語を検索した結果を Table 11、Table 12 に示す。4.2.1 と比較すると、コンテキストを「再生可能エネルギー、太陽光発電、風力発電」と複数指定したことで、「再生可能エネルギー」の正解セットの再現率が 7% から 20% に上がっていることが分かる。また、この文脈では関係のない「異常気象」と「エコビジネス」の文脈から想定される正解セットの単語はほとんど含まれておらず、部分空間がシャープに切り取られた結果となっている。

Table 11: 検索語が「温暖化」、コンテキストが「再生可能エネルギー、太陽光発電、風力発電」の時の、期待する正解セット B の単語の再現率

正解セット	B
再現率	20%

Table 12: 検索語が「温暖化」、コンテキストが「再生可能エネルギー」の時の、期待する正解セット以外の単語との重複率

正解セット	A	C
再現率	0%	7%

4.2.4. 考察

これらの結果から、コンテキストを変えることにより、そのコンテキストによって切り取られた部分空間上で意味的に絞られた単語が検索できることが分かる。

4.3. 実験 3:ドキュメント検索

ここでは、生成したメタデータ空間と意味的連想検索を用いることにより、特定のコンテキストを指定して検索することで、ドキュメントの持つ文脈多様性を計量できることを検証する。

4.3.1. 検索例 1: コンテキスト「異常気象」

コンテキスト「異常気象」における、温暖化関連ドキュメントの意味的連想検索結果上位 20 件を Table 13 に示す。「異常気象」と近い意味を持つキーワードタグを太字で表している。

「エルニーニョ」や「海面上昇」といった自然災害系や、南極北極といった氷に関する記事が多く含まれていることが分かる。

4.3.2. 検索例 2: コンテキスト「再生可能エネ

ルギー」

コンテキスト「再生可能エネルギー」における、新聞記事の意味的連想検索結果上位 20 件を Table 14 に示す。「再生可能エネルギー」と近い意味を持つキーワードタグを太字で表している。

「パリ協定」が頻出している理由として、議題に温室効果ガス排出量の削減が盛り込まれていたため、その手段として再生可能エネルギーが話題に上がっていることが考えられる。

4.3.3. 検索例 3: コンテキスト「エコビジネス」

コンテキスト「エコビジネス」における、新聞記事の意味的連想検索結果上位 20 件を Table 15 に示す。「エコビジネス」と近い意味を持つキーワードタグを太字で表している。

具体的な企業名や、CO2 排出枠に関する話題が上位にあることが分かる。

4.3.4. 考察

これらの結果から、本項で提案したメタデータ空間を用いた意味的連想検索により、直接メタデータを持たない新聞記事も、連想的に検索することができることが示された。

Table 13: コンテキストに「異常気象」を指定した時のドキュメント意味的連想検索結果の上位 20 位

順位	ノルム	記事 ID	キーワードタグ	記事タイトル
1	0.713644	NIRKDB20141104NKE0258	気象研究所 気象庁 気温 気象 温暖化	気温高いと強い雨増加、気象庁、温暖化の影響裏付け。
2	0.711513	NIRKDB20150722NKM0023	気象庁 気象 温暖化	猛暑日増えてカエデ紅葉遅く、気象庁まとめ。
3	0.710766	NIRKDB20170321NKE0198	WMO 異常気象 気象 温暖化	異常気象多発恐れ、今年、温暖化の進行やまず。
4	0.709637	NIRKDB20150710NKM0185	エルニーニョ 予測 海洋研究開発機構 気象庁 気温 温暖化	エルニーニョ、強まる傾向—気象庁予測、温暖化ペースに影響も、梅雨が長引き雨量増、夏は比較的気温低め（ニュースな科学）
5	0.706706	NIRKDB20130505NKM0022	進行 気象研究所 予測 気象庁 温暖化	激しい竜巻発生、2075年以降倍増、気象研、温暖化で予測。
6	0.706596	NIRKDB20170621NKE0491	ヨーロッパ WMO アメリカ合衆国 気象 温暖化	欧州や中東、記録的高温、ポルトガル・スペイン40度超、5～6月「温暖化の影響」。
7	0.705772	NIRKDB20160410NKM0260	北極 北極海 氷 気象 温暖化	日本の北極戦略航路は一開き手から、研究体制の充実急務（創論）
8	0.703848	NIRKDB20130523NSS0075	海洋研究開発機構 温暖化	成層圏、空気の流れ変化、海洋機構など、温暖化響く。
9	0.703848	NIRKDB20140911NSS0058	海洋研究開発機構 温暖化	温暖化予測、精度向上へ、海洋機構など、関連物質の分布調査。
10	0.703107	NIRKDB20150403NKE0576	南極 気温 温暖化	南極、最高気温観測か、17.5度、温暖化の影響も。
11	0.70296	NIRKDB20131225NKM0033	平均気温 気象庁 温暖化	世界の平均気温、史上2位の高さ、今年、平年0.2度上回る、日本はプラス0.4度。
12	0.700921	NIRKDB20130714NKM0019	気象庁 気温 温暖化	温暖化で猛暑日増加、10年で0.5日、都市部はさらに暑く。
13	0.700833	NIRKDB20111117NKE0368	原因 南極 温暖化	国立極地研、温暖化が原因？、南極の雪増える。
14	0.700666	NIRKDB20121124NKM0028	南極 温暖化	南極レーダーで大気の動き分析、第54次観測隊、あす出発、高精度の温暖化予測期待。
15	0.700666	NIRKDB20140630NKE0236	南極 温暖化	温暖化でペンギン2割減、今世紀後半、絶滅危惧種に。

16	0.699864	NIRKDB20121118NKM0242	気象庁 温暖化	キーワード--データレスキュー、災害の爪痕、防災に活用（日曜版）
17	0.699864	NIRKDB20150501NKM0491	気象庁 温暖化	梅雨の降雨量、温暖化の影（でーたクリップ）
18	0.698966	NIRKDB20150215NKM0060	魚 北極海 海洋研究 開発機構 温暖化	北極海、温暖化進むと大漁場に？--海流・水温 …複雑に絡み合う（ナゾ謎かがく）
19	0.698898	NIRKDB20120531NKE0393	海水 北極圏 チーム 北極海 海洋研究開発 機構 温暖化	海洋機構、海水減少で雲が変化、北極圏、温暖 化を加速か。
20	0.696841	NIRKDB20140514NSS0075	研究結果 N A S A 海面上昇 氷 温暖化	南極の氷、未崩壊の限界超える（フラッシュ）

Table 14: コンテキストに「再生可能エネルギー」を指定した時のドキュメント意味的連想検索結果の上位 20 位

順位	ノルム	記事 ID	キーワードタグ	記事タイトル
1	0.569095	NIRKDB20170608NKE0457	離脱 知事 国際枠組 み 防止 パリ協定	ハワイ州、パリ協定履行へ。
2	0.535584	NIRKDB20161129NKM0161	積水ハウス 住宅 削 減目標 パリ協定	積水ハウス、マンションも「ゼロエネ」--パリ 協定の削減目標、住宅が達成のカギ。
3	0.528809	NIRKDB20170307NSS0029	商機 パリ協定	炭素価格「企業内で」500社、「パリ協定」で 取り組み加速、コスト削減・商機にらむ。
4	0.526542	NIRKDB20170313NKE0296	国際枠組み パリ協定	パリ協定「悪い契約」、米環境保護局長官が発言、 温暖化対策後退。
5	0.525569	NIRKDB20160830NKM0075	排出規制 商機 採択 国際枠組み パリ 規制 パリ協定	パリ協定後の排出規制--逆転の発想で商機に (GLOBAL EYE)
6	0.52089	NIRKDB20130227NKL0463	条例 防止	三重県、森林税導入へ、土砂災害防止など目指 す。
7	0.515936	NIRKDB20161019NSS0018	代替フロン 途上国 パリ 規制 パリ協定	パリ協定発効へ、分野別温暖化規制が前進、航 空会社、国際線、20年上限に、代替フロン、途 上国も80%削減。
8	0.506827	NIRKDB20161128NKM0164	パリ協定	温暖化対策「パリ協定」--異例のスピード発効 実現（ニュースのトリセツ）
9	0.504367	NIRKDB20151217NSS0172	省エネ 採択 防止 パ リ協定	パリ協定、競争力の糧に（眼光紙背）
10	0.500132	NIRKDB20161018NSS0087	旭硝子 モントリオール 議定書 パナソニック 閉幕 ダイキン工業 採択 代替フロン 日本 企業 規制	日本企業、新冷媒開発急ぐ、代替フロン規制採 択受け。
11	0.497625	NIRKDB20141127NKL0185	埼玉県 削減目標	温暖化ガス削減目標、埼玉県、21%に引き下 げ、原発の停止影響。
12	0.495708	NIRKDB20161105NKM0066	パリ パリ協定	パリ協定発効で開く温暖化対策の新時代（社説）
13	0.494823	NIRKDB20150127NKM0106	環境 国際交渉 交渉	第3章国際交渉の考え方（3）一橋大学教授岡 田章氏--全ての国の合意難航（やさしい経済学環 境と向き合う）
14	0.494502	NIRKDB20110530NKM0067	クールビズ 防止	2005年6月1日、クールビズ開始（アーカ イブ）
15	0.491365	NIRKDB20161016NKM0035	モントリオール議定 書 閉幕 エアコン 冷 媒 代替フロン 規制	締約国会議、代替フロン規制採択、先進国生産 85%減、19年に発効。
16	0.489427	NIRKDB20150501NKL0419	官公庁 スーパークー ルビズ 軽装 クールビ ズ 防止	クールビズ、今年も開始、官公庁や企業。
17	0.483692	NIRKDB20160424NKM0050	採択 削減目標 途上 国 COP21 パリ パリ協定	パリ協定--途上国含め削減目標（キーワード）
18	0.483565	NIRKDB20161216NSS0020	石油業界 ヨーロッパ 非政府組織 再生可能 エネルギー	欧州石油大手が再生エネ--欧州が先行、米は慎 重、業界に温度差。
19	0.472677	NIRKDB20161108NKE0542	衆議院 パリ協定	パリ協定、午後承認、衆院。
20	0.463262	NIRKDB20080421NKM0056	韓国政府 韓国 規制	韓国、車両通行に総量規制、温暖化ガス排出抑 制で。

Table 15: コンテキストに「エコビジネス」を指定した時のドキュメント意味的連想検索結果の上位 20 位

順位	ノルム	記事 ID	キーワードタグ	記事タイトル
1	0.604938	NIRKDB20100308NSS0099	半導体 事業所 設備投資	S I I の高塚事業所、温暖化ガス排出、抑制設備を導入。
2	0.591628	NIRKDB20090402NKM0216	プラント 発電所 共同事業 コストダウン 実験	温暖化ガス、仏社と地下貯留実験、ダウ・ケミカル、発電所向け。
3	0.586031	NIRKDB20080104NKM0119	環境ビジネス 植林 対外提携 業務提携 寄付 エコビジネス カーボンオフセット CO2	CO2 排出、寄付で「相殺」、カーボンオフセット、英大手が日本進出。
4	0.575226	NIRKDB20071025NSS0080	インターネット サイト サービス開始 消費者 CO2 排出権 CO2	PEARカーボンオフセット、CO2 排出権、個人向け販売、ネットでキロ 3 円。
5	0.569345	NIRKDB20090124NKM0233	ロケット 受注 衛星 打ち上げ 人工衛星 宇宙航空研究開発機構 三菱重工業 コストダウン	H2A 打ち上げ成功、三菱重、海外受注に弾み、信頼性アピール、大型機、次の焦点に。
6	0.565954	NIRKDB20090217NKM0195	法律改正 実験 CO2	CO2 回収・貯留へ法整備、政府指針、4 月にも一〇九年度内、大規模実験。
7	0.565078	NIRKDB20100429NKL0177	茨城県 首都圏 木材 茨城 建設 補助金 地方自治体 CO2	茨城県森林組合連合会（水戸市）一原木市場新たに整備（地の利を生かす）
8	0.5637	NIRKDB20071003NKL0106	鉄道 新潟 事業所 CO2	新潟トランス新新潟事業所、聖籠町一「ニッチ分野」幅広く（我が社の戦略拠点）
9	0.563478	NIRKDB20091110NKE0408	CO2	CO2 回収・貯留システム一油田などの地下に封じ込め（これも知りたい）
10	0.563478	NIRKDB20100219NSS0026	CO2	隠れた対象ガス（4）非エネルギー起源の CO2（温暖化対策の論点）終
11	0.563121	NIRKDB20081216NKL0143	緑化 新潟 商品開発 リサイクル	ノジマ（三条市）一廃瓦使い舗装ブロック（中堅企業挑む拓く）
12	0.561901	NIRKDB20090407NSS0021	カーボンフットプリント 情報開示 製造業 CO2	カーボンフットプリント試行開始、「川上」メーカーに警戒感。
13	0.560071	NIRKDB20071124NKM0019	WMO CO2	温暖化ガス過去最高、世界気象機関、「CO2 増加目立つ」。
14	0.559991	NIRKDB20100406NKL0286	CO2 排出枠 カーボンオフセット CO2	平和建設グループ、CO2 排出枠の購入先を地元。
15	0.557977	NIRKDB20081020NSS0017	新製品 CO2 排出枠 カーボンオフセット CO2	カーボンオフセット、日立建機、1 台 1 トン分相殺一建機に CO2 排出枠。
16	0.557317	NIRKDB20081201NSS0025	半導体 商品開発 新製品 共同開発 コストダウン	東京エレクト、PFC 除害装置に参入。
17	0.555782	NIRKDB20090123NKE0417	ロケット 地球観測衛星 宇宙 衛星打ち上げ 宇宙航空研究開発機構 三菱重工業 CO2	「いぶき」打ち上げ、CO2 大気濃度を観測、まいど 1 号も宇宙へ。
18	0.555605	NIRKDB20090820NSS0001	半導体 ダイキン工業 CO2	ダイキン、PFC ガス、完全回収一半導体製造用、冷凍機で空気と分離。
19	0.555208	NIRKDB20100816NRS0039	イオン CO2 排出枠 CO2	イオン、CO2 排出枠、1.5 万トン購入。
20	0.55481	NIRKDB20091107NKL0456	発電所 事業所 CO2	三菱化学黒崎事業所、亜酸化窒素の削減着手、CO2 換算、年最大 8 万トン。

5. 結論・今後の展望

本論文では、新聞記事を対象としたドキュメント検索において、意味の数学モデルを適用して意味的連想検

索を行う際のメタデータ空間の生成方式を示し、多様な文脈を持つ単語やドキュメント間の関連性を計量する本意味的連想検索実現方式の有効性を検証した。

今回は、正解セットに対する再現率に着目した実験結果となったが、今後は特定のコンテキストでのみ現れる単語に着目した差分の考察や、より再現率を上げるためのメタデータ空間の作成方法の改善を目指す。

また、本研究において実現した本意味的連想検索メタデータ空間を5次元世界地図システム「5D World Map System」[6][7][8][9][10]と統合することにより、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[11]。

参 考 文 献

- [1] T. Kitagawa and Y. Kiyoki, "A mathematical model of meaning and its application to multidatabase systems," Proceedings RIDE-IMS '93: Third International Workshop on Research Issues in Data Engineering: Interoperability in Multidatabase Systems, Vienna, 1993, pp. 130-135.
- [2] Yasushi Kiyoki, Takashi Kitagawa, and Takanari Hayama. 1994. A metadatabase system for semantic image search by a mathematical model of meaning. *SIGMOD Rec.* 23, 4 (December 1994), 34-41.
- [3] Kiyoki, Y., Kitagawa, T., & Hitomi, Y. (1995). Fundamental framework for realizing semantic interoperability in a multidatabase environment. *Integrated Computer-Aided Engineering*, 2(1), 3-20.
- [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] 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(TOD 2), pp.15-28, 1999.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] 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.

[11] 佐々木 史織, 藤岡 華子, 渡邊 敏央, 大塚 恭平, 石井 昌之, 清木 康: "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", 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

[12] 日経テレコン: <http://telecom.nikkei.co.jp/>

[13] 日経テレコン分類語一覧: <http://t21help.nikkei.co.jp/reference/cat845/post-563.html>