

**テキストマイニングによる地域ニーズの分析
手法の開発に関する研究 —モンズーンア
ジア地域の水資源インフラ整備—**

独立行政法人水資源機構総合技術センター

主幹 杉浦 政裕

平成21年9月

助成研究者紹介

研究代表者

すぎうら まさひろ

杉 浦 政 裕

現職：独立行政法人水資源機構総合技術センター参事

主な著書：

Masahiro Sugiura(2008), "Creation of appropriate technology on water resources management having linkage to society - Towards sustainable innovation based eco-industrial clusters -," *Eco-Industrial Clusters as an Inclusive Business Development Model : Policy Report 2008*, IGES, pp.131-144.

Masahiro Sugiura(2008), "Sustainable Management for Water Resources Infrastructure," *NARBO workshop on Managing Assets and Risks*. (NARBO workshop on Managing Assets and Risks was co-organized by Ministry of Public Works of the Republic of Indonesia, Asian Development Bank, Asian Development Bank Institute, and Japan Water Agency on 21th Feb. 2008 in Indonesia.)

杉浦政裕、落井康裕（2007）「インドネシア共和国チタルム川における PJT-II の水資源管理」『ダム技術』（財）ダム技術センター，No. 246(2007. 3)，pp23-31.

共同研究者

こもだ ふみお

菰 田 文 男

現職：埼玉大学経済学部教授（経済学博士）

主な著書：

菰田文男，松島三兒，高橋敏昭，垣内淳，矢賀部裕（2007）『技術と市場ニーズの探索・融合』税務経理協会.

Fumio Komoda(2006)" Bio-based Economy and the Opportunity for Developing Countries" , *Asian Economy and Social Environment*, Vol.1.

菰田文男(2003)『脳の外化と生命進化』多賀出版.

菰田文男(2003)『脳外革命』中経出版.

菰田文男(2000)『科学・技術と価値』多賀出版.

菰田文男(1991)『情報通信技術と世界経済』ミネルヴァ書房.

菰田文男(1987)『国際技術移転の理論』有斐閣.

**テキストマイニングによる地域ニーズの分析手法の開発に関する研究
— モンスーンアジア地域の水資源インフラ整備 —**

目次

1. はじめに.....	4
2. 研究の課題と方法.....	6
2-1 研究の課題.....	6
2-2 研究の方法.....	6
3. 先行研究の概説と簡易分析手法の提案.....	7
3-1 先行研究の概説.....	7
3-2 簡易テキストマイニングによる分析手法の提案.....	8
4. 簡易テキストマイニングによる地域ニーズの分析.....	9
4-1 時系列分析手順と利用データ.....	9
4-1-1 分析対象データの準備.....	9
4-1-2 単語の出現頻度の順位比較.....	9
4-1-3 数量化Ⅲ類による高出現頻度語の相対的位置関係の マッピング.....	9
4-2 時系列分析結果とその限界.....	10
4-2-1 分析対象データのプロフィール.....	10
4-2-2 単語の出現頻度の順位比較の結果.....	11
4-2-3 数量化Ⅲ類分析による日本国内発行新聞（1992, 1996, 2000年）のテキストマイニングによる時系列分析の結果... 14	14
4-3 単年分析手順と利用データ.....	22
4-3-1 分析対象データの準備.....	22
4-3-2 単語の出現頻度の順位.....	22
4-3-3 数量化Ⅲ類による高出現頻度語の相対的位置関係の マッピング.....	22
4-4 単年分析結果とその限界.....	23
4-4-1 分析対象データのプロフィール.....	23
4-4-2 単語の出現頻度の順位表示および数量化Ⅲ類分析によ るインドネシア国内発行新聞（2008/2009年）のテキ ストマイニングによる単年分析の結果.....	25
5. 簡易テキストマイニングによる分析の限界.....	28
5-1 重要な低頻度出現語と「気づき」の支援.....	28
5-2 効果的にテキストマイニングを実施するための文字情報データ の収集方法と分析のための辞書作成の留意点.....	31
6. 今後の研究の発展可能性.....	32
註.....	32
参考文献.....	32
附属資料.....	34

図表一覧

図 4-1	数量化Ⅲ類分析による 1992 年の社会状況のマッピング	15
図 4-2	数量化Ⅲ類分析による 1996 年の社会状況のマッピング	16
図 4-3	数量化Ⅲ類分析による 2000 年の社会状況のマッピング	17
図 4-4	数量化Ⅲ類による 1992 年、1996 年、2000 年の社会状況マッピングの時系列配置	18
図 4-5	数量化Ⅲ類分析による 2008/2009 年の社会状況のマッピング	26
図 4-6	副詞・形容や否定を表す単語と高出現頻度語の共起状況 (1992 年)	30
表 4-1	見出しキーワードにより選別された関連新聞記事数	10
表 4-2	分析対象キーワード数	10
表 4-3 (1/2)	高出現頻度語の順位比較によるニーズ発見支援	12
表 4-3 (2/2)	高出現頻度語の順位比較によるニーズ発見支援	13
表 4-4	数量化Ⅲ類分析結果 (1992 年)	19
表 4-5	数量化Ⅲ類分析結果 (1996 年)	20
表 4-6	数量化Ⅲ類分析結果 (2000 年)	21
表 4-7	キーワードにより選別された関連新聞記事	23
表 4-8	分析対象キーワード数 (2008/2009)	23
表 4-9	高出現頻度語の順位表示によるニーズ発見支援	25
表 4-10	数量化Ⅲ類分析結果 (2008/2009 年)	27
表 4-11	「ダム」と出現単語の単純な共起関係 (1992 年)	29
附表 4-1 (1/4)	置換辞書 (1992 年、1996 年、2000 年)	35
附表 4-1 (2/4)	置換辞書 (1992 年、1996 年、2000 年)	36
附表 4-1 (3/4)	置換辞書 (1992 年、1996 年、2000 年)	37
附表 4-1 (4/4)	置換辞書 (1992 年、1996 年、2000 年)	38
附表 4-2 (1/3)	置換辞書 (2008/2009 年)	39
附表 4-2 (2/3)	置換辞書 (2008/2009 年)	40
附表 4-2 (3/3)	置換辞書 (2008/2009 年)	41
附表 4-3 (1/12)	副詞・形容や否定を表す単語と高出現頻度の共起 (1992 年)	42
附表 4-3 (2/12)	副詞・形容や否定を表す単語と高出現頻度の共起 (1992 年)	43
附表 4-3 (3/12)	副詞・形容や否定を表す単語と高出現頻度の共起 (1992 年)	44
附表 4-3 (4/12)	副詞・形容や否定を表す単語と高出現頻度の共起 (1992 年)	45
附表 4-3 (5/12)	副詞・形容や否定を表す単語と高出現頻度の共起 (1992 年)	46
附表 4-3 (6/12)	副詞・形容や否定を表す単語と高出現頻度の共起 (1992 年)	47
附表 4-3 (7/12)	副詞・形容や否定を表す単語と高出現頻度の共起 (1992 年)	48
附表 4-3 (8/12)	副詞・形容や否定を表す単語と高出現頻度の共起 (1992 年)	49
附表 4-3 (9/12)	副詞・形容や否定を表す単語と高出現頻度の共起 (1992 年)	50
附表 4-3 (10/12)	副詞・形容や否定を表す単語と高出現頻度の共起 (1992 年)	51
附表 4-3 (11/12)	副詞・形容や否定を表す単語と高出現頻度の共起 (1992 年)	52
附表 4-3 (12/12)	副詞・形容や否定を表す単語と高出現頻度の共起 (1992 年)	53

1. はじめに

本研究は、社会インフラ整備事業の企画段階において、文字で書かれた情報（インタビュー結果、新聞記事など）をテキストマイニングすることにより、「曖昧なニーズ表現の中に含まれている真実」を発見することによって、対象地域が真に求めているニーズを抽出し、社会インフラ整備事業の発見／創出の支援を目指すことを目的としている。

社会インフラ整備事業予算の逼迫につれて、厳格な「事前・事後評価」が求められるようになりつつあるが、それと並行して「住民のニーズ」を正しく捉えることの重要性が高まっている。そのため、アンケート調査等が実施されているが、その多くは〇×式の回答であるため、回答の背後にある意味の構造を見出すことが困難であった。とりわけ、河川・水資源関係の事業は住民間や企業の利害関係が錯綜していて真の要求を読み取れず、調整も困難であり、結果として適正な事業の実施がおこなわれないと判断される場合もあった。この限界を克服して、ニーズを正しく把握し、利害を適正に調整するためには、〇×式のアンケートではなく関係者の要求や本音にかんする情報が切り捨てられることなく含まれやすい「フリー記述」のアンケート調査やインタビューをおこない、その回答の中に含まれている「事業に対する要望・不満」や改善要求を可能な限り正しく読みとることが必要である。フリー記述のテキストデータには、「非常に」「強く」「とても」のような副詞や、「美味しい水」「綺麗な河川」のような形容詞、形容動詞など、人の感情を含んだ情報、「思い」の強さを表現した情報が多く含まれている。このような情報をテキストマイニングと多変量解析という統計的手法を用いて解析することが、従来は不可能であった住民ニーズの解明・把握を可能にしてくれるものと期待できる。

そのため、本研究では事例として、現在あるいは将来的に日本経済にとって重要なアジア地域、中でも日本の建設産業にとって有望市場であり、日本と同じ稲作文化圏でもあるモンsoonアジア地域の水資源インフラの整備・管理事業のニーズ解明・把握の適応を試みる。

持続性のある水資源施設の建設・管理のためには、①流域のユーザー（企業、住民等）が何を求めているのか、②その実現のための技術的条件は整っているのか、③その実現のためにはどのような社会制度（法律、行政の介入）が必要か、等々を解明する必要がある。

これまで、当該部門なり当該地域なりのニーズから出発して、それを満たすために事業が必然的に案出されたり、すでに数十年存在していたものを取り上げられたにすぎず、隠れた要求を見落としていたり、必ずしも創造的とはいえない場合もある。

しかし、時代のニーズは変化し、たとえば洪水によって農地や家屋を失う農民の生産性を高め生活水準を改善することが、経済社会開発のもっとも緊急かつ重要な戦略であるということになれば、危険区域設定、災害復興支援などのソフト対策と、護岸や貯水池建設などのハード対策を組み合わせた洪水対策プロジェクトが創出されることになる。

そのため、状況に即した考え方、正確なインフラ需要分析や予測、効率的な運用のための制度設計などについて、地域発展と地域ニーズの両面から考えることが重要となる。現在、事業の発見あるいは創出には、こういった時代にニーズの変化に対応できる創造的思考がかなり必要となってきた。

このように事業計画段階において、「曖昧なニーズ表現の中に含まれている真実」を的確に把握することが可能となれば、事業計画者の経験や現場感覚を十分に引き出した創造的思考を支援することができる。それにより、効率的なインフラ整備の実施が可能となる。また、これはインフ

ラ整備後に必要とされる管理技術/社会制度の設計にも反映することが可能となり、事業の効率的な執行にも寄与する。したがって本研究は、事業計画者の意思決定支援、社会ニーズの解明、整備されたインフラを十分に活用するための受容条件の解明にも貢献できる。また本研究で開発・提案される手法は、単にインフラ整備事業にとどまらず、さまざまな分野の政策立案への支援を期待できる。

2. 研究の課題と方法

2-1 研究の課題

本研究の課題は、事業企画の実務担当者が事業企画段階において、大量の文書から「曖昧なニーズ表現の中に含まれている真実」を的確に把握し、事業計画の経験や現場感覚を十分に引き出すための「気づき」を支援する手法を提案することである。

近年、私たちは、情報通信技術の普及により、膨大な情報にアクセスできるようになった。しかし、膨大な情報は、情報が大量かつ多様という二面性をもっている。この二面性は、利用者にとって大変やっかいな性質である。利用者は求める情報を得るためには、文書の中身を精査しなければならないため、その煩雑さも膨大になるからである。また、求める情報を提示すれば、自動的に膨大な情報から精査して提供してくれるようなシステムは現段階では存在しない。人工知能が発達してきたとはいえ未だ創造力思考を行っているのは人間である。その想像力思考を支援するひとつの方法は、「気づき」を支援することであろう。

人間の「気づき」を支援するためには、大量の文書の構文解析や単語出現頻度の分析を行っただけでは役に立たない。単語のネットワークが作り出している意味を的確に把握することが求められる。そのためには、分析結果を可視化し、実務担当者の「気づき」を最大限に引き出すことが重要となる。

そこで本研究では、人間の経験や感覚を最大限に活用し創造するための「気づき」を支援するために、膨大な文書からの的確な情報を取り出し、分析結果の可視化を図る手法を提案する。

2-2 研究の方法

水資源インフラ整備ニーズを深く分析するためには、フリー記述のアンケートを住民、政府の両者に対して実施することが望ましいが、ここでは社会の中に潜む水資源インフラ整備ニーズを探る分析手法の開発に焦点を合わせるため、客観性が高く収集しやすい新聞記事の分析を採択する。

本研究において新聞記事を代用した理由は、①事実を客観的に伝えることにより発達してきたマスメディアの中でも特に影響力を持つ、②マスメディアを通じて膨大な情報が流通しており、マスメディアの動向を知ることは社会ニーズを考えるうえで重要となると判断したからである。新聞は、インドネシア国内で発行されている現地新聞『じゃかるた新聞』を選定する。分析対象は、水資源インフラ関連記事（じゃかるた新聞社のオンラインデータベースの制約上1年分）とする。

しかし、単年分の分析では蓄積された膨大な情報を生かし、ニーズの全体像と時系列的な変化に対して十分なアプローチは困難である。

そこで、日本の主要な新聞である『毎日新聞』における水資源インフラ関連記事（1992年、1996年、2000年）を用いて、社会状況の変遷をマッピングし分析することにより、社会状況変化を先取りしたインフラ整備計画策定のためのニーズの芽の発見支援の手法を確認する。

そして、この手法に基づき『じゃかるた新聞』1年分の記事によるニーズ分析の結果とその限界について考察する。

3. 先行研究の概説と簡易分析手法の提案

3-1 先行研究の概説

テキストマイニングによる分析手法は、記事の動向表現の分析、記事の因果関係の分析、潜在的意味解析、グラフ理論の応用、ニューラルネットワーク理論の応用などが研究されている。

まず、テキストマイニングにより記事の動向表現を分析する関連研究は、数詞に注目してその周辺の言語パターンを解析することにより情報を分析する方法（斉藤ら（1998））、係り受け関係を利用する方法（藤畑ら（2001））、統計量名を注釈付けするためのタグセットを定義してアノテーション付コーパスにより機械学習を使って自動抽出する方法（森（2007））、統計量表現に共通してよく出現する suffix に着目したパターンマッチングを利用した統計量表現抽出する方法（河合ら（2008））、そして、複数の記事に時系列に出現する様々な動向情報からデータを取り出し、データテーブルを作成し、Data Transaction, Visual Mapping, View Transformation の3つのプロセスからデータを可視化表現する方法（松下ら（2005））などがある。

次に、テキストマイニングにより記事の因果関係を分析する関連研究は、諸事象間の因果関係を有向グラフとして表し、事象の連鎖反応を分析する手法（佐藤ら（1999）、佐藤ら（2006））、因果関係を含む可能性の高い共起関係に着目し、共起ネットワークを構築・観測する方法（河合ら（2008））、そして、因果関係知識の自動獲得を目指して、任意に定めたテキスト集合に対して因果関係情報に注釈をつけることにより、因果関係の出現傾向を分析する手法（乾ら（2005））などがある。

記事の動向表現の分析や記事の因果関係の分析は、高出現頻度語の分析を中心とし、既知の傾向を定量的に把握には貢献している。しかし、これらの手法では、同様の意味をもつ低出現頻度重要語を取り込むことは困難であった。そこで、語句の背後にある意味を分析することにより、低出現頻出重要語を分析の対象に取り込むことを試みている手法が、テキストの潜在的意味解析（LSA: Latent Semantic Analysis）である。テキストのLSAの関連研究は、特異値分解に基づいたLSAを発展させたPLSI（Probabilistic Latent Semantic Indexing）、因子分析と情報理論に基づいた堅固な統計モデルによるSLSI（Statistical Latent Semantic Indexing）などがある。LSAは、中村（2008）が解説するとおり、すべての文章の背後には意味の構造が存在すると考え、これを行列の形で表現し、分析するところに特徴がある。また、LSAは、さまざまな言葉で表現される意味の豊かすぎる部分を、行列の分解という形でとり除き、複数の語句の背後に共通して潜在する意味構造を抽出している。それは、漠然とした意味の豊かさよりも、凝縮した構造の方が語句に留まらない意味の豊かさを効率的に表現できる可能性があるからである。

さらに、グラフ理論やニューラルネットワーク理論を応用することにより、既存の文書を分析から未来予測を試みる研究も取り組まれている。

3-2 簡易テキストマイニングによる分析手法の提案

ここでは、実務者が膨大な文字情報を前に、手軽に分析できる手法を提案する。まず、単語の出現頻度の順位を比較することが出発点である。これにより、

- (1) 高出現頻度語から、社会に定着した単語を発見できる。
- (2) 出現頻度が急増している単語から、現在の社会の変化を発見できる。
- (3) 出現頻度は低くて、目立たない単語から、将来の社会の変化の予兆を発見できる。

(1)(2)は比較的容易に知ることが出来るが、しかし重要な知見は得られない。(3)は知ることが難しいが、得られる知見は大きい。社会の変化を単語により把握するには「単語の共起関係」や「単語と単語のネットワーク」を知ることが重要となる。

次に、膨大な文字情報の中から高出現頻度語を抽出し、それらを数量化Ⅲ類により、抽出された各語の相対的位置関係をマッピングすることにより、社会状況の可視化を図る。そして、社会状況を可視化することにより、実務者が社会ニーズに「気づく」ことを支援する。

この手法により、①1992年・1996年・2000年の『毎日新聞』全記事、②2008/2009年(2008年8月1日から2009年7月31日まで)の『じゃかるた新聞』全記事を分析してみた。この手法で、社会に定着した問題、新たな課題として登場した社会的問題、社会的問題の背景の推移、社会的問題の内容の変化など多くの知見が得られることを示す。

4. 簡易テキストマイニングによる地域ニーズの分析

4-1 時系列分析手順と利用データ

4-1-1 分析対象データの準備

- 1) 『毎日新聞』全文電子データ（1992年、1996年、2000年）を使って、検索キーワード（水資源、洪水、渇水、水道、水質、水力、用水、地下水）で各年の記事見出しを検索し、記事を抽出する。以下、抽出された記事を分析対象とする。
- 2) 各年の頻出単語から、関係が深いと考えられる単語を抽出し、それを変数として選択する。
- 3) 置換辞書を作成するために、1992年、1996年、2000年の分析対象を分かち書きし、名詞を抽出する。
- 4) 抽出された名詞から、類似の意味をもつ名詞の表現を統一するために、置換辞書を作成する。例えば、「利根川」＝「河川－川系」、「荒川」＝「河川－川系」、「木曾川」＝「河川－川系」と表示する。
- 5) 作成した置換辞書を使って、1992年、1996年、2000年の分析対象の類似の意味をもつ名詞表現を統一する。

なお、①分かち書き処理、キーワード抽出、置換辞書作成には『Word Miner』（日本電子計算機社製）、②数量化Ⅲ類分析には『エクセル統計 2006』（社会情報サービス社製）を使用した。

4-1-2 単語の出現頻度の順位比較

- 1) 作成した置換辞書を用いて、再度、1992年の全分析対象を分かち書きし、記事毎に名詞を抽出する。
- 2) 抽出された名詞毎の出現数を数える。なお、対象記事の長さによる同一出現語の出現頻度のバイアスを除去するために、対象記事1つに複数回出現する同一語は、対象記事1つにつき1回の出現とした。
- 3) 高出現頻度語 100 を目安に、出現頻度数の多い順に並べる。
- 4) 1996年、2000年についても、1) から3) を行う。
- 5) 各年毎に出現単語の入れ替わり状況を分析する

この手順により、社会問題のキーワードとして長期化するキーワード、出現頻度が増加傾向のキーワード、短期間で出現頻度が減少傾向となるキーワードをとらえる。

4-1-3 数量化Ⅲ類による高出現頻度語の相対的位置関係のマッピング

- 1) 対象記事中に、「4-1-2 3)」で並べられた名詞の有無を調べ、その結果を年毎に行列データにする。
- 2) 行列データを元に、数量化Ⅲ類分析をする。
この手順により、新聞記事から社会状況をマッピングし、社会ニーズがどのように変化していくかを、経年変化を定量的にとらえる。
- 3) 分析にあたっては、統計学的な基準はないが、単相関係数 r が 0.5 以上または累積寄与率 50%以上となることを目安にした。

4-2 時系列分析結果とその限界

4-2-1 分析対象データのプロフィール

『毎日新聞』全文電子データ（1992年、1996年、2000年）を使って、検索キーワード（水資源、洪水、渇水、水道、水質、水力、用水、地下水）で各年の記事見出しを検索し、抽出された記事を表4-1に示す。1992年の分析対象記事数は12件、1996年の分析対象記事数は91件、2000年の分析対象記事数は64件であった。

表4-1 見出しキーワードにより選別された関連新聞記事数

見出しキーワード	見出しキーワードにより選別された関連新聞記事数		
	1992年	1996年	2000年
水資源	0	2	0
洪水	0	3	6
渇水	0	15	7
水道	0	11	8
水質	0	22	6
水力	0	0	2
ダム	7	22	22
用水	1	7	2
地下水	4	9	11

(筆者作成)

分析対象のキーワード数を、表4-2に示す。置換辞書使用後の全文を通じて異なるキーワード数は、1992年は843個、1996年は1,682個、2000年は2,851個であった。置換辞書使用後の全文を通じて異なるキーワード数は、①各年の高出現頻度語から関係が深いと考えられる単語を抽出、②置換辞書を作成し類似の意味をもつ名詞を統一表現にする、③年毎に全文を通じて異なる名詞の数を計数したものである。

今回は306語を置換辞書、記号・句読点・助詞を削除辞書として登録し、置換辞書および削除辞書により、埼玉県や茨城県など地方自治体名は「自治体-都道府県市」、利根川や荒川などの河川名は「河川-川系」、危機や非常時など危険や非常を表す表現は「危険-非常系」などへ表現を統一し、記号・句読点・助詞は削除した。なお、作成した辞書の内容については附属資料の附表4-2に示す。

表4-2 分析対象キーワード数

	1992年	1996年	2000年
置換辞書使用前の全文を通して異なるキーワード数	885	1,842	3,026
置換辞書使用後の全文を通して異なるキーワード数	843	1,682	2,851

(筆者作成)

4-2-2 単語の出現頻度の順位比較の結果

1992年、1996年、2000年の高出現頻度単語の経年変化を分析した結果を表4-3（1/2）および表4-3（2/2）に示す。社会問題のキーワードとして長期化する単語と出現頻度が増加している単語、短期間でキーワードとして存在しなくなる単語があることがわかる。長期化する単語または毎年出現する単語は、基本語となっており社会の変化を探るには困難である可能性が高い。その一方で、出現頻度が増加する単語は社会の変化を表している可能性が高い。

具体的には、1992年、1996年、2000年の各年に出現する「自治体-都道府縣市系」、「事業-工事系」、「調査」などの単語は、基本語として認識されるため、社会状況の変化を探ることは困難である。しかし、1996年から出現頻度が高くなった「議会-議会審議系」、「改善-改善系」や、2000年から出現頻度が高くなった「委員会-会議」「見直し」などは社会の変化を表しているといえるだろう。こうしたことから、出現頻度が急増している単語から、現在の社会の変化への気づきを支援できることがわかる。

しかし、表4-3の結果に分析において注意しなければならないことは、この分析手法は単語のカウントを実施しているだけであり、近いうちに大きな問題に発展しそうな予兆を示す語が切り捨てられている可能性があることである。

表4-3 (1/2) 高出現頻度語の順位比較によるニーズ発見支援

番号	1992年		1996年		2000年	
	関連類出名詞	サンプル 度数	関連類出名詞	サンプル 度数	関連類出名詞	サンプル 度数
1	自治体-都道府県市系	12	自治体-都道府県市系	59	自治体-都道府県市系	55
2	事業-工事系	9	河川-川系	33	地域-地域系	34
3	地域-地域系	7	地域-地域系	30	河川-川系	32
4	ダム	6	ダム	27	全国-日本系	30
5	全国-日本系	6	危険-非常系	24	事業-工事系	28
6	河川-川系	5	水	23	危険-非常系	27
7	調査	5	水質	23	水	27
8	危険-非常系	4	全国-日本系	21	環境	25
9	使用	4	対策-対策系	21	対策-対策系	25
10	市民-住民系	4	国土交通省-本省本局系	18	計画-予定系	23
11	水	4	水不足-水不足系	18	ダム	21
12	設置	4	環境	17	調査	21
13	地下	4	事業-工事系	17	降雨-雨系	19
14	地下水-地下水系	4	調査	16	水質	19
15	計画-予定系	3	降雨-雨系	15	影響	17
16	現場	3	議会-議会審議会系	14	国土交通省-本省本局系	16
17	山	3	影響	13	市民-住民系	16
18	指摘	3	制限	13	設置	16
19	人	3	水系	12	発表	16
20	対策-対策系	3	水道水-水道系	12	水不足-水不足系	15
21	中央-中心系	3	発表	12	水位	14
22	排水	3	取水	11	水道水-水道系	14
23	反対	3	地方	11	汚染-汚濁系	13
24	物質	3	汚染-汚濁系	10	指摘	13
25	利用	3	原因-理由系	10	地方	13
26	グループ	2	上流域-上流系	10	下流域-下流系	12
27	コンクリート	2	環境省-本省本局系	9	基準	12
28	プロジェクト	2	給水	9	原因-理由系	12
29	安定	2	貯水量-貯水率	9	周辺	12
30	井戸	2	保全-保護系	9	地下水-地下水系	12
31	汚染-汚濁系	2	市民-住民系	8	改善-改善系	11
32	開発	2	設置	8	対象	11
33	期待-希望系	2	必要	8	中央-中心系	11
34	掘削	2	改善-改善系	7	被害	11
35	経済	2	水位	7	確認	10
36	検討	2	水源	7	環境省-本省本局系	10
37	公園	2	知事	7	検出	10
38	公開	2	地下水-地下水系	7	公共	10
39	工業	2	農業	7	洪水	10
40	工場	2	琵琶湖	7	取水	10
41	降雨-雨系	2	利用	7	制限	10
42	高潮	2	開発	6	団体	10
43	作業	2	強化	6	農業	10
44	市内	2	計画-予定系	6	発生	10
45	自然	2	検査	6	会議	9
46	実際	2	市内	6	見直し	9
47	遮水壁	2	施設-設備系	6	中止	9
48	取水	2	場合	6	反対	9
49	出水	2	浄化	6	琵琶湖	9
50	上流域-上流系	2	水道局	6	物質	9

□ 1992年頻出単語71のうち
1996年には残らなかった単語

▨ 1992年頻出単語71のうち
1996年頻出単語85に残った
単語

▩ 1992年頻出単語71のうち2000
年頻出単語100に残った単語

▧ 1992年頻出単語71のうち
1996年頻出単語85に残った
単語

▦ 1996年新出単語85のうち新
出単語

▤ 1996年新出単語のうち2000年
頻出単語100に残った単語

▣ 2000年頻出単語100のうち新
出単語

(筆者作成)

表4-3 (2/2) 高出現頻度語の順位比較によるニーズ発見支援

番号	1992年		1996年		2000年	
	関連類出名詞	サンプル 度数	関連類出名詞	サンプル 度数	関連類出名詞	サンプル 度数
51	条例	2	大震災	6	保全-保護系	9
52	人々	2	発生	6	方針	9
53	水面	2	防止	6	用水	9
54	成長	2	要請	6	委員会	8
55	請求	2	シンポジウム	5	可能性	8
56	専門家	2	会長	5	関係	8
57	地価	2	管理	5	規則	8
58	地質	2	企業	5	検討	8
59	地盤	2	現場	5	生活	8
60	中心部	2	工業	5	発電	8
61	沈下	2	国土庁	5	防止	8
62	鉄道	2	通常-平常系	5	異常	7
63	土	2	特別	5	維持	7
64	土地	2	判断	5	運動	7
65	買収	2	物質	5	期待-希望系	7
66	発展	2	用水路	5	議会-議会審議会系	7
67	判断	2	流域	5	経済	7
68	復興	2	連絡	5	健康	7
69	噴出	2	下水道	4	市内	7
70	有害	2	回復	4	施設-設備系	7
71	用水	2	基準	4	場合	7
72			実施	4	水源	7
73			循環	4	地下	7
74			場所	4	通常-平常系	7
75			情報	4	土地	7
76			状態	4	排水	7
77			人口	4	批判	7
78			水資源	4	必要	7
79			生活	4	変化	7
80			団体	4	報告	7
81			地盤	4	目的	7
82			提出	4	沿岸	6
83			発がん性	4	開発	6
84			販売	4	完成	6
85			方針	4	管理	6
86					気象台	6
87					供給	6
88					教授	6
89					研究	6
90					自然	6
91					上流域-上流系	6
92					水系	6
93					説明	6
94					程度	6
95					土壌	6
96					濃度	6
97					発がん性	6
98					利水	6
99					利用	6
100					流域	6

□ 1992年頻出単語71のうち
1996年には残らなかった単語

▨ 1992年頻出単語71のうち
1996年頻出単語85に残った
単語

▩ 1992年頻出単語71のうち2000
年頻出単語100に残った単語

▧ 1992年頻出単語71のうち
1996年頻出単語85に残った
単語

▦ 1996年頻出単語85のうち新
出単語

▤ 1996年新出単語のうち2000年
頻出単語100に残った単語

▣ 2000年頻出単語100のうち新
出単語

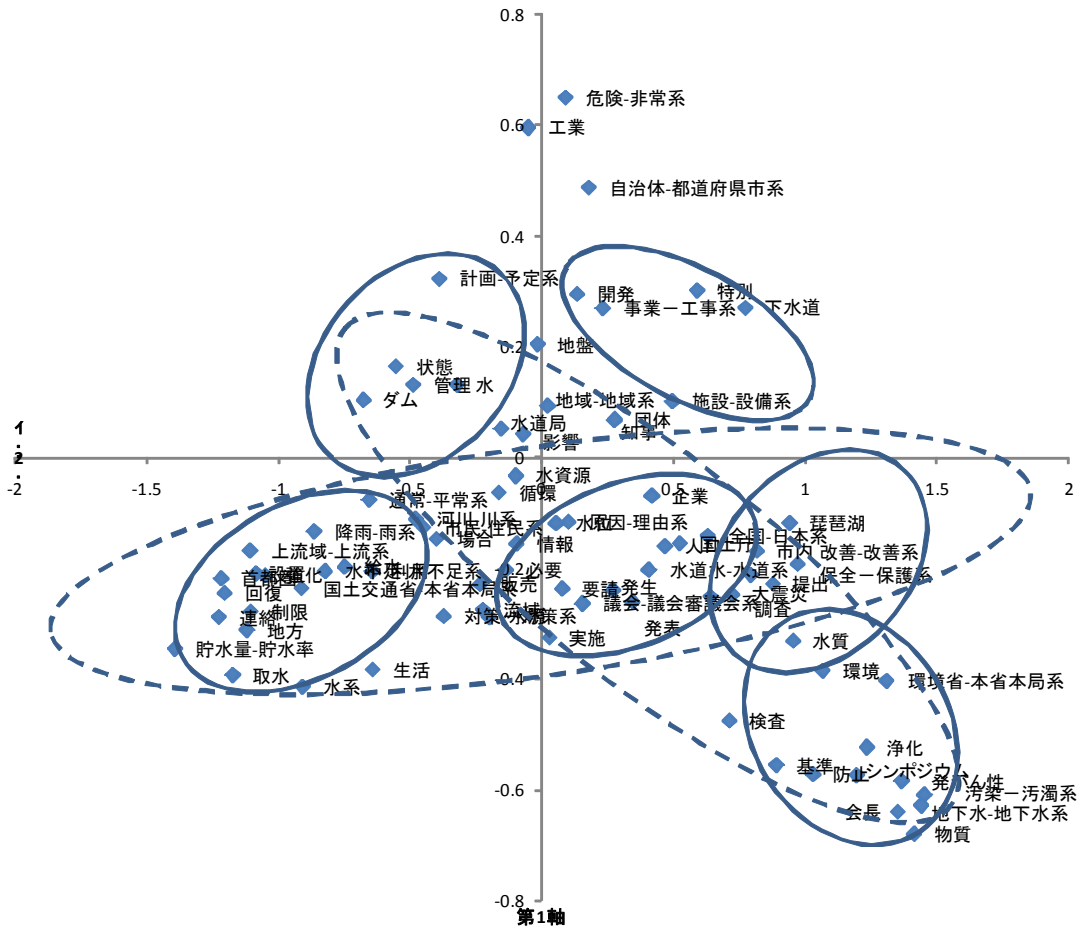
(筆者作成)

4-2-3 数量化Ⅲ類分析による日本国内発行新聞（1992, 1996, 2000年）のテキストマッピングによる時系列分析の結果

表4-3に表示されている1992年、1996年、2000年の高出現頻度語を、年毎の記事毎に単語の有無を行列データにし、数量化Ⅲ類分析をした結果を図4-1、図4-2、図4-3に示す。また、年毎の結果に数量化Ⅲ類分析した結果を時系列に配置した図を図4-4に示す。

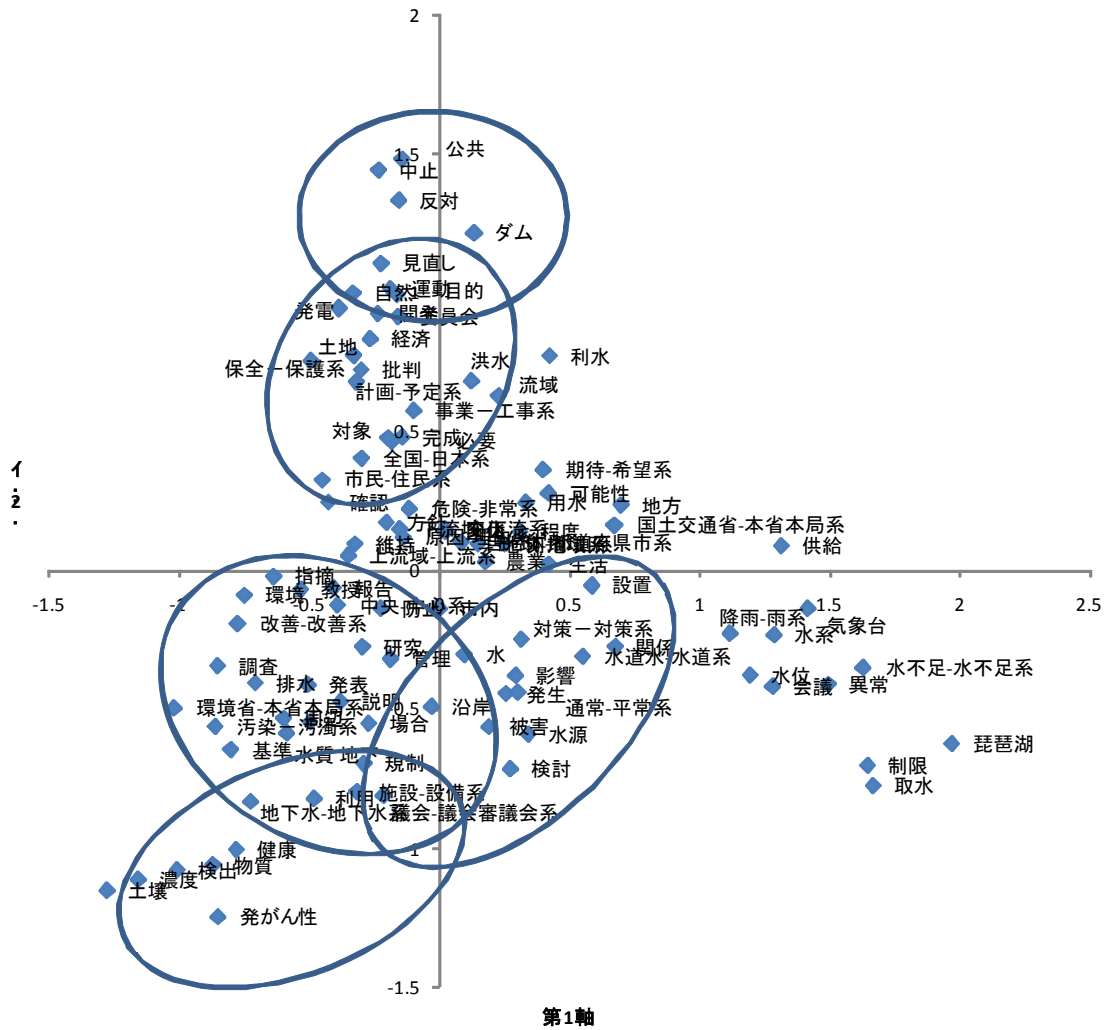
図4-4は、たとえば次のように解釈する。1992年の河川下流問題（河川汚染）は、市民に社会問題意識を高めさせた。そして1996年には、河川下流問題の対策がとられたものの、河川下流問題に対する市民意識はさらに高まり（下流問題クラスターと問題意識クラスターの接近）、それに加え環境保全も視界に入ってきた。これらの問題の解決に向け、政府委員会設置へ向け活動をはじめた。こうした流れの中で2000年には、河川下流問題は解決されたが、公共事業に対する市民意識はさらに高まり、ダム建設問題を政府委員会で議論させ、市民活動も活性化された。このように解釈すれば、市民活動を支える基盤作りという社会ニーズは、「下流問題クラスターと問題意識クラスターの接近」を引き金に、加速された。と解釈することができる。

このように新聞記事から社会状況がマッピングされ、マッピングされた固有語句の群れが、解析者にその意味を気づかせるのである。このことは、①語句の一致ではなく、語句の意味の近さを探ることが重要となること、②分析者がテキストデータの「読み込み」（意味の把握）作業を丁寧に行う必要があることを示唆している。つまり、固有単語のもつ潜在的意味に着目することが重要となるのである。



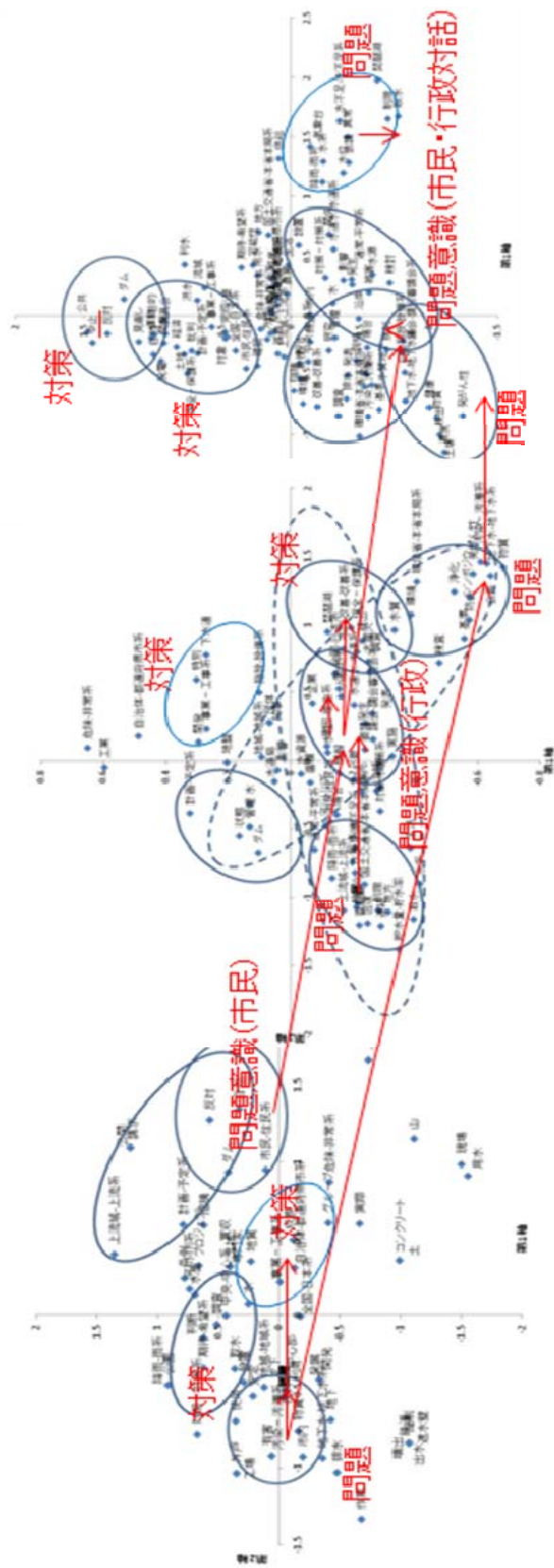
(筆者作成)

図 4-2 数量化Ⅲ類分析による 1996 年の社会状況のマッピング



(筆者作成)

図4-3 数量化Ⅲ類分析による2000年の社会状況のマッピング



(著者作成)

図4-4 数量化Ⅲ類による1992年、1996年、2000年の社会状況マッピングの時系列配置

表4-4 数量化Ⅲ類分析結果（1992年）

1992年

高出現頻度単語				高出現頻度単語			
	第1軸	第2軸		第1軸	第2軸		
1 グループ	0.6028	-0.4096	51 地価	-0.5373	-0.0439		
2 コンクリート	0.3551	-0.9931	52 地質	0.3489	0.2297		
3 ダム	0.9314	0.4169	53 地盤	-0.5373	-0.0439		
4 プロジェクト	0.3210	0.6576	54 中央-中心系	-0.0046	0.4355		
5 安定	-0.5316	0.2120	55 中心部	-0.5373	-0.0439		
6 井戸	-1.0360	0.3485	56 調査	-0.0627	0.5101		
7 汚染-汚濁系	-0.9231	0.0574	57 沈下	-0.5373	-0.0439		
8 河川-川系	0.1684	0.7346	58 鉄道	-0.8374	-1.0729		
9 開発	-0.4244	-0.3349	59 土	0.3551	-0.9931		
10 危険-非常系	0.8656	-0.4191	60 土地	0.3153	0.4017		
11 期待-希望系	-0.3453	0.6186	61 排水	-1.0303	-0.4801		
12 掘削	-0.8374	-1.0729	62 買収	0.3153	0.4017		
13 経済	-0.5373	-0.0439	63 発展	-0.4244	-0.3349		
14 計画-予定系	0.5899	0.7894	64 判断	-0.1498	0.6220		
15 検討	-0.5373	-0.0439	65 反対	1.2710	0.5680		
16 現場	0.9840	-1.5033	66 復興	-0.5373	-0.0439		
17 公園	-0.4583	0.9097	67 物質	-0.7537	-0.1586		
18 公開	1.0943	1.2221	68 噴出	-0.8374	-1.0729		
19 工業	-0.5373	-0.0439	69 有害	-0.9231	0.0574		
20 工場	-1.0360	0.3485	70 用水	0.9059	-1.5572		
21 降雨-雨系	-0.4583	0.9097	71 利用	-0.4978	-0.0556		
22 高潮	-0.5373	-0.0439					
23 作業	-1.3361	-0.6806					
24 山	1.1491	-1.1143					
25 使用	-0.6935	0.3556					
26 市内	-0.9287	-0.1984					
27 市民-住民系	0.9464	0.1059					
28 指摘	0.5862	0.6178					
29 事業-工事系	0.2159	-0.0108					
30 自治体-都道府県市系	0.3054	-0.1232					
31 自然	1.3425	0.3270					
32 実際	0.5972	-0.6655					
33 遮水壁	-0.8374	-1.0729					
34 取水	-0.3510	0.3623					
35 出水	-0.8374	-1.0729					
36 上流域-上流系	0.3943	1.3552					
37 条例	0.2417	0.7766					
38 人	1.6647	-0.7379					
39 人々	1.9331	-1.6319					
40 水	0.0723	0.2500					
41 水面	0.1313	0.6897					
42 成長	0.4899	-0.1186					
43 請求	1.0943	1.2221					
44 設置	-0.4413	0.2874					
45 専門家	0.3153	0.4017					
46 全国-日本系	-0.0089	-0.1780					
47 対策-対策系	-0.7813	0.6711					
48 地域-地域系	-0.4723	0.1198					
49 地下	-0.6845	-0.4304					
50 地下水-地下水系	-0.9367	-0.3622					

各軸の値は相関係数を掛けた後の値

	固有値	寄与率	累積寄与率	相関係数r
第1軸	0.5027	17.01%	17.01%	0.7090
第2軸	0.3718	12.53%	29.59%	0.6097

(筆者作成)

表4-5 数量化Ⅲ類分析結果（1996年）

1996年

	高出現頻度単語	第1軸	第2軸		高出現頻度単語	第1軸	第2軸
1	シンポジウム	1.1965	-0.5740	51	水道水-水道系	0.4086	-0.2038
2	ダム	-0.6724	0.1091	52	水不足-水不足系	-0.8192	-0.2066
3	影響	-0.0680	0.0416	53	制限	-1.1012	-0.2793
4	汚染-汚濁系	1.4538	-0.6097	54	生活	-0.6980	-0.3935
5	下水道	0.7748	0.2700	55	設置	-1.0795	-0.2109
6	河川-川系	-0.4744	-0.1122	56	全国-日本系	0.6924	-0.1420
7	会長	1.3528	-0.6409	57	対策-対策系	-0.3692	-0.2876
8	回復	-1.2003	-0.2460	58	大震災	0.7257	-0.2485
9	改善-改善系	0.9725	-0.1936	59	団体	0.2775	0.0663
10	開発	0.1376	0.2944	60	知事	0.2803	0.0690
11	環境	1.0679	-0.3857	61	地域-地域系	0.0252	0.0927
12	環境省-本省本局系	1.3102	-0.4043	62	地下水-地下水系	1.4418	-0.6289
13	管理	-0.4899	0.1306	63	地盤	-0.0133	0.2043
14	企業	0.4204	-0.0703	64	地方	-1.1153	-0.3120
15	危険-非常系	0.0932	0.6497	65	貯水量-貯水率	-1.3881	-0.3461
16	基準	0.8927	-0.5564	66	調査	0.6438	-0.2518
17	議会-議会審議会系	0.1570	-0.2646	67	通常-平常系	-0.6514	-0.0770
18	給水	-0.7447	-0.1981	68	提出	0.8793	-0.2292
19	強化	-1.0461	-0.2140	69	特別	0.5920	0.3010
20	計画-予定系	-0.3856	0.3224	70	農業	0.2605	4.4933
21	検査	0.7144	-0.4766	71	発がん性	1.3666	-0.5849
22	原因-理由系	0.1045	-0.1172	72	発生	0.2725	-0.2404
23	現場	-0.3272	1.3419	73	発表	0.3471	-0.2617
24	工業	-0.0484	0.5952	74	判断	0.2465	1.0771
25	降雨-雨系	-0.8611	-0.1341	75	販売	-0.2317	-0.2297
26	国土交通省-本省本局系	-0.9075	-0.2373	76	琵琶湖	0.9428	-0.1189
27	国土庁	0.5264	-0.1558	77	必要	-0.1343	-0.2035
28	市内	0.8186	-0.1694	78	物質	1.4148	-0.6811
29	市民-住民系	-0.4457	-0.1268	79	保全-保護系	0.7956	-0.2125
30	施設-設備系	0.4981	0.1002	80	防止	1.0312	-0.5729
31	事業-工事系	0.2340	0.2692	81	用水路	0.3959	5.4296
32	自治体-都道府県市系	0.1803	0.4870	82	要請	0.0806	-0.2381
33	実施	0.0318	-0.3261	83	利用	-0.6409	-0.2061
34	取水	-1.1692	-0.3933	84	流域	-0.2211	-0.2755
35	首都圏	-1.2132	-0.2196	85	連絡	-1.2209	-0.2387
36	循環	-0.1594	-0.0637				
37	上流域-上流系	-1.1093	-0.1688				
38	場合	-0.3962	-0.1482				
39	場所	0.7686	1.0342				
40	情報	-0.0924	-0.1563				
41	浄化	1.2359	-0.5241				
42	状態	-0.5500	0.1641				
43	人口	0.4697	-0.1613				
44	水	-0.3184	0.1305				
45	水位	0.0571	-0.1195				
46	水系	-0.9035	-0.4145				
47	水源	-0.1960	-0.2882				
48	水資源	-0.0954	-0.0338				
49	水質	0.9559	-0.3328				
50	水道局	-0.1515	0.0513				

各軸の値は相関係数を掛けた後の値

	固有値	寄与率	累積寄与率	相関係数r
第1軸	0.5060	7.06%	7.06%	0.7113
第2軸	0.4657	6.50%	13.57%	0.6824

(筆者作成)

表4-6 数量化Ⅲ類分析結果(2000年)
2000年

	高出現頻度単語	第1軸	第2軸		高出現頻度単語	第1軸	第2軸
1	ダム	0.137397	1.212962	51	場合	-0.26813	-0.55216
2	委員会	-0.1548	0.913457	52	水	0.101153	-0.30419
3	異常	1.498351	-0.41085	53	水位	1.197721	-0.37832
4	維持	-0.32106	0.095219	54	水系	1.291584	-0.23411
5	運動	-0.18467	1.013598	55	水源	0.345301	-0.59117
6	影響	0.297299	-0.38017	56	水質	-0.58123	-0.58733
7	沿岸	-0.02459	-0.49126	57	水道水-水道系	0.555958	-0.31064
8	汚染-汚濁系	-0.85665	-0.56289	58	水不足-水不足系	1.632673	-0.35187
9	下流域-下流系	-0.14942	0.148118	59	制限	1.649746	-0.7034
10	可能性	0.423416	0.276355	60	生活	0.423747	0.021544
11	河川-川系	0.254696	0.094536	61	設置	0.590302	-0.05545
12	会議	1.284167	-0.41887	62	説明	-0.37668	-0.47378
13	改善-改善系	-0.77227	-0.19278	63	全国-日本系	-0.29436	0.402702
14	開発	-0.23169	0.922576	64	対策-対策系	0.318134	-0.24979
15	確認	-0.42301	0.243769	65	対象	-0.17899	0.465197
16	完成	-0.13848	0.477942	66	団体	0.026804	0.136108
17	環境	-0.74592	-0.08949	67	地域-地域系	0.151727	0.095238
18	環境省-本省本局系	-1.01636	-0.49708	68	地下	-0.49242	-0.54701
19	管理	-0.18551	-0.31916	69	地下水-地下水系	-0.72204	-0.83484
20	関係	0.680932	-0.27533	70	地方	0.699956	0.234504
21	危険-非常系	-0.11321	0.22016	71	中央-中心系	-0.3881	-0.12534
22	基準	-0.79637	-0.64518	72	中止	-0.22938	1.44069
23	期待-希望系	0.402012	0.360547	73	調査	-0.84943	-0.3444
24	气象台	1.419866	-0.13925	74	通常-平常系	0.304857	-0.43909
25	規制	-0.28643	-0.6952	75	程度	0.311178	0.137174
26	議会-議会審議会系	-0.21187	-0.81232	76	土壌	-1.27356	-1.15409
27	供給	1.316878	0.087436	77	土地	-0.32516	0.773316
28	教授	-0.52852	-0.06978	78	濃度	-1.15356	-1.11383
29	経済	-0.26343	0.831325	79	農業	0.1797	0.029367
30	計画-予定系	-0.31729	0.677632	80	排水	-0.70364	-0.40586
31	健康	-0.77703	-1.00707	81	発がん性	-0.84783	-1.24374
32	検出	-1.00403	-1.0794	82	発生	0.260563	-0.4436
33	検討	0.276338	-0.71711	83	発電	-0.38199	0.942851
34	研究	-0.29242	-0.27525	84	発表	-0.50165	-0.41277
35	見直し	-0.22054	1.104909	85	反対	-0.15187	1.33147
36	原因-理由系	-0.13419	0.117739	86	批判	-0.29588	0.721179
37	公共	-0.1397	1.480133	87	被害	0.193642	-0.56523
38	洪水	0.126622	0.680654	88	琵琶湖	1.973795	-0.62537
39	降雨-雨系	1.120728	-0.22874	89	必要	-0.19226	0.476624
40	国土交通省-本省本局系	0.675944	0.162735	90	物質	-0.86702	-1.06105
41	市内	0.005346	-0.13886	91	変化	0.027105	0.149927
42	市民-住民系	-0.44645	0.325622	92	保全-保護系	-0.48987	0.753176
43	指摘	-0.63232	-0.02248	93	報告	-0.40219	-0.06362
44	施設-設備系	-0.31276	-0.79634	94	方針	-0.19891	0.172333
45	事業-工事系	-0.09586	0.574144	95	防止	-0.22054	-0.13835
46	自治体-都道府県市系	0.088631	0.099873	96	目的	-0.15912	0.992411
47	自然	-0.32764	0.998372	97	用水	0.336367	0.243727
48	取水	1.671038	-0.77866	98	利水	0.427665	0.772261
49	周辺	-0.59474	-0.53419	99	利用	-0.47868	-0.82387
50	上流域-上流系	-0.34625	0.051762	100	流域	0.232706	0.627388

各軸の値は相関係数を掛けた後の値

	固有値	寄与率	累積寄与率	相関係数r
第1軸	0.3926	8.30%	8.30%	0.6266
第2軸	0.3174	6.71%	15.00%	0.5634

(筆者作成)

4-3 単年分析手順と利用データ

4-3-1 分析対象データの準備

- 1) 『じゃかるた新聞』全文電子データ（2008年8月1日から2009年7月31日まで）を使って、検索キーワード（水資源、洪水、渇水、水道、水質、水力、用水、地下水）で各年の記事見出しを検索し、記事を抽出する。以下、抽出された記事を分析対象とする。
- 2) 頻出単語から、関係が深いと考えられる単語を抽出し、それを変数として選択する。
- 3) 置換辞書を作成するために、分析対象を分かち書きし、名詞を抽出する。
- 4) 抽出された名詞から、類似の意味をもつ名詞の表現を統一するために、置換辞書を作成する。例えば、「チタルム川」＝「河川－川系」、「ソロ川」＝「河川－川系」、「チリウン川＝河川－川系」と表示する。
- 6) 作成した置換辞書を使って、分析対象の類似の意味をもつ名詞表現を統一する。

なお、①分かち書き処理、キーワード抽出、置換辞書作成には『Word Miner』（日本電子計算機社製）、②数量化Ⅲ類分析には『エクセル統計 2006』（社会情報サービス社製）を使用した。

4-3-2 単語の出現頻度の順位

- 1) 作成した置換辞書を用いて、再度、全分析対象を分かち書きし、記事毎に名詞を抽出する。
- 2) 抽出された名刺毎の出現数を数える。なお、対象記事の長さによる同一出現語の出現頻度のバイアスを除去するために、対象記事1つに複数回出現する同一語は、対象記事1つにつき1回の出現とした。
- 3) 高出現頻度語 100 を目安に、出現頻度数の多い順に並べる。
この手順により、社会問題のキーワードの傾向をとらえる。

4-3-3 数量化Ⅲ類による高出現頻度語の相対的位置関係のマッピング

- 1) 対象記事中に、「4-3-2 3)」で並べられた名詞の有無を調べ、その結果を行列データにする。
- 2) 行列データを元に、数量化Ⅲ類分析をする。この手順により、新聞記事から社会状況をマッピングし、社会ニーズの状況を定量的にとらえる。
- 3) 分析にあたっては、統計学的な基準はないが、単相関係数 r が 0.5 以上または累積寄与率 50%以上となることを目安にした。

4-4 単年時系列分析結果とその限界

4-4-1 分析対象データのプロフィール

『じゃかるた新聞』全文電子データ（2008年8月1日から2009年7月31日まで）を使って、検索キーワード（水資源、洪水、渇水、水道、水質、水力、用水、地下水）で1年間の記事内容を検索し、抽出された記事を表4-7に示す。分析対象記事数は重複記時を除く177件であった。

表4-7 キーワードにより選別された関連新聞記事数

キーワード	キーワードにより選別された 関連新聞記事数
	2008年8月1日から 2009年7月31日まで
水資源	10
洪水	89
渇水	1
水道	27
水質	10
水力	12
ダム	24
用水	8
地下水	8
計	189
計（重複記事除く）	177

（筆者作成）

分析対象のキーワード数を、表4-8に示す。置換辞書使用後の全文を通じて異なるキーワード数は、1,732個であった。置換辞書使用後の全文を通じて異なるキーワード数は、①高出現頻度語から関係が深いと考えられる単語を抽出、②置換辞書を作成し類似の意味をもつ名詞を統一表現にする、③全文を通じて異なる名詞の数を計数したものである。

今回は220語を置換辞書、記号・句読点・助詞を削除辞書として登録し、置換辞書および削除辞書により、ジャカルタ市や東ジャワ州など地方自治体名は「自治体-州縣市系」、利根川や荒川などの河川名は「河川」、危機や非常時など危険や非常を表す表現は「危険-非常系」などへ表現を統一し、記号・句読点・助詞は削除した。なお、作成した辞書の内容については附属資料の附表4-2に示す。

表4-8 分析対象キーワード数（2008/2009）

	2008/2009
置換辞書使用前の 全文を通じて異なる キーワード数	4,539
置換辞書使用後の 全文を通じて異なる キーワード数	1,732

（筆者作成）

4-4-2 単語の出現頻度の順位表示および数量化Ⅲ類分析によるインドネシア国内発行新聞（2008/2009年）のテキストマイニングによる単年分析の結果

2008/2009年（2008年8月1日から2009年7月31日まで）の高出現頻度単語を単年分析し、出現頻度順に整理した結果を表4-9に示す。また、表4-9に表示されている高出現頻度語を記事毎に単語の有無を行列データにし、数量化Ⅲ類分析をした結果を図4-5に示す。なお、分析にあたっては、単相関係数 r が0.5以上または累積寄与率50%以上となることを目安にしたため、行列表データの行や列に1や0が多い場合は、その行や列を削除して、行列表を作成し直した。そのため、表4-9の高出現頻度語は、図4-4に表示されている語数よりも多い。

図4-5は、たとえば次のように解釈する。語群⑤は2008年3月下旬に発生したギントゥンダムが決壊により、百人以上の死者を出した事故であったが、一過性のトピックであるため、全体の語群から離れたところに位置している。しかし、中央政府による貯水池施設の安全点検、メディアによる災害時の避難体制への高まりがあることから、語群⑥、語群④へ派生していることがわかる。語群⑥では、事故原因の追究を行い、今後更にダム安全性を向上させ、国民が安全・安心して生活できるよう努力していることを示している。また、語群④では、メディアは、死者数、事故原因の報道を通じ、避難体制の整備やダム安全性の確保の必要性を社会に投げかけていることを示している。

語群①は中央政府のODA依存状況を示しており、ODAにより国土インフラの整備を推進しようとしている。特に、水道事業は大きな関心ごとである。また、語群②は中央政府は、同時に計画的な住宅整備を図っていることもわかる。

語群③は降雨による洪水被害は、都市部の道路が冠水することが多く、生活の支障をきたしていることがわかる。

これらの状況から、流域水ユーザー、水関連インフラ整備のために必要な技術的条件・法制度の視点から、次のようにニーズを捉えることもできるであろう。新規事業（水道施設建設など）、改築事業（道路改築、ダム施設点検・補修など）などのインフラ整備事業が望まれている。それと並行して、ダムなど河川構造物の安全基準の見直し、警戒避難のためのハザードマップ作成、自治体への災害復旧支援制度、市民への生活再建支援制度などの技術基準や法制度の策定ニーズがあると解釈することができる。

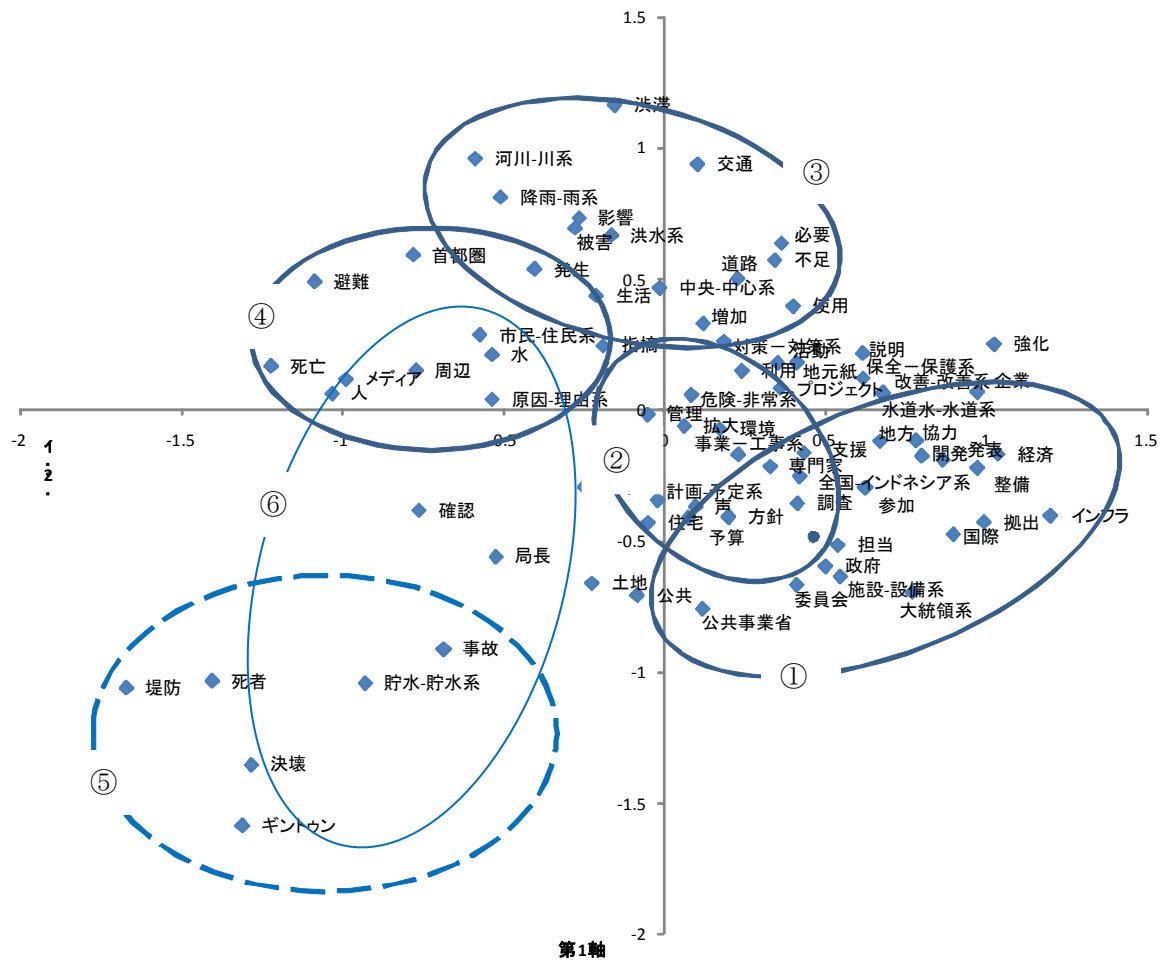
この分析は、単年のみの記事であるため先の予測は容易ではないが、過去の判例や行政の判断パターンを学ぶことにより、単年の少ない手がかりから、確からしさを高めた「気づき」を支援することが可能となるであろう。

表4-9 高出現頻度語の順位表示によるニーズ発見支援

2008/2009年(2008.8.1-2009.7.31)

	関連頻出名詞	サンプル度数		関連頻出名詞	サンプル度数
1	自治体-都道府県市系	144	51	プロジェクト	19
2	地域-地域系	140	52	会社	19
3	洪水系	100	53	拡大	19
4	降雨-雨系	75	54	首都圏	19
5	事業-工事系	66	55	企業	18
6	全国-インドネシア系	66	56	参加	18
7	被害	66	57	死者	18
8	発生	65	58	メディア	17
9	市民-住民系	54	59	悪化	17
10	危険-非常系	53	60	国内	17
11	計画-予定系	53	61	声	17
12	対策-対策系	51	62	利用	17
13	政府	45	63	交通	16
14	河川-川系	42	64	公共事業省	16
15	環境	35	65	国際	16
16	中央-中心系	35	66	市内	16
17	影響	34	67	住宅	16
18	公共	34	68	設置	16
19	指摘	34	69	専門家	16
20	水	34	70	土地	16
21	水道水-水道系	34	71	委員会	15
22	貯水-貯水系	34	72	監視	15
23	原因-理由系	33	73	機関	15
24	支援	31	74	強化	15
25	必要	31	75	経済	15
26	開発	30	76	国営	15
27	方針	30	77	渋滞	15
28	管理	29	78	大規模	15
29	死亡	28	79	大統領系	15
30	道路	28	80	中部	15
31	周辺	27	81	不足	15
32	協力	25	82	保全-保護系	15
33	生活	25	83	民家	15
34	決壊	24	84	インフラ	14
35	整備	24	85	ダム	14
36	説明	24	86	抛出	14
37	知事	24	87	災害	14
38	地元紙	24	88	実施	14
39	改善-改善系	23	89	人	14
40	使用	23	90	増加	14
41	施設-設備系	23	91	担当	14
42	事故	23	92	地方	14
43	避難	23	93	堤防	14
44	活動	22			
45	局長	22			
46	調査	22			
47	発表	22			
48	確認	20			
49	予算	20			
50	ギンタウン	19			

(筆者作成)



(筆者作成)

図 4-5 数量化Ⅲ類分析による 2008/2009 年の社会状況のマッピング

表4-10 数量化Ⅲ類分析結果（2008/2009年）

2008/2009年(2008.8.1-2009.7.31)

高出現頻度単語		第1軸	第2軸	高出現頻度単語		第1軸	第2軸
1	強化	1.0269	0.2518	51	周辺	-0.7690	0.1521
2	堤防	-1.6689	-1.0591	52	整備	0.9737	-0.2202
3	専門家	0.3312	-0.2152	53	道路	0.2296	0.5015
4	大統領系	0.7723	-0.6956	54	必要	0.3658	0.6358
5	拡大	0.0637	-0.0611	55	生活	-0.2095	0.4368
6	担当	0.5408	-0.5145	56	方針	0.2015	-0.4068
7	不足	0.3461	0.5715	57	影響	-0.2624	0.7330
8	プロジェクト	0.3616	0.0843	58	環境	0.1746	-0.0717
9	委員会	0.4132	-0.6666	59	公共	-0.0827	-0.7056
10	確認	-0.7597	-0.3833	60	指摘	-0.1885	0.2449
11	国際	0.9005	-0.4738	61	支援	0.4365	-0.1621
12	死者	-1.4021	-1.0337	62	水道水-水道系	0.6818	0.0659
13	地方	0.6720	-0.1176	63	貯水-貯水系	-0.9271	-1.0416
14	保全-保護系	0.6196	0.1220	64	原因-理由系	-0.5329	0.0402
15	拠出	0.9950	-0.4272	65	水	-0.5334	0.2115
16	土地	-0.2226	-0.6595	66	中央-中心系	-0.0124	0.4662
17	発表	0.8672	-0.1901	67	市民-住民系	-0.5702	0.2871
18	利用	0.2425	0.1500	68	被害	-0.2761	0.6941
19	インフラ	1.2013	-0.4031	69	政府	0.5025	-0.5958
20	局長	-0.5206	-0.5596	70	発生	-0.4000	0.5372
21	交通	0.1058	0.9382	71	計画-予定系	-0.0199	-0.3441
22	公共事業省	0.1205	-0.7588	72	降雨-雨系	-0.5074	0.8121
23	死亡	-1.2194	0.1675	73	対策-対策系	0.1868	0.2630
24	首都圏	-0.7771	0.5912	74	事業-工事系	0.2323	-0.1693
25	人	-1.0286	0.0625	75	全国-インドネシア系	0.4221	-0.2525
26	地元紙	0.4133	0.1811	76	危険-非常系	0.0848	0.0592
27	メディア	-0.9872	0.1184	77	洪水系	-0.1616	0.6661
28	活動	0.3541	0.1803				
29	参加	0.6256	-0.2948				
30	使用	0.4024	0.3969				
31	施設-設備系	0.5482	-0.6350				
32	住宅	-0.0482	-0.4286				
33	洪滞	-0.1520	1.1629				
34	声	0.0991	-0.3682				
35	増加	0.1236	0.3301				
36	予算	0.0764	-0.4092				
37	ギンタウン	-1.3087	-1.5858				
38	改善-改善系	0.6787	0.0634				
39	経済	1.0381	-0.1689				
40	説明	0.6178	0.2169				
41	調査	0.4157	-0.3556				
42	企業	0.9740	0.0698				
43	協力	0.7834	-0.1165				
44	決壊	-1.2806	-1.3534				
45	知事	-0.2465	-0.2928				
46	河川-川系	-0.5852	0.9589				
47	管理	-0.0494	-0.0159				
48	事故	-0.6836	-0.9110				
49	避難	-1.0838	0.4892				
50	開発	0.8018	-0.1752				

各軸の値は相関係数を掛けた後の値

	固有値	寄与率	累積寄与率	相関係数 r
第1軸	0.3434	9.03%	9.03%	0.5860
第2軸	0.2810	7.39%	16.42%	0.5301

(筆者作成)

5 簡易テキストマイニングによる分析の限界

5-1 重要な低頻度出現語と「気づき」の支援

低出現単語の重要性

提案した簡易テキストマイニングによる分析手法には、「4-2-1」で示すように、(1) 出現頻度は低くて、目立たない単語から、将来の社会の変化の予兆を発見するためには、個々の単語を他から切り離して分析していたのではできない、(2) 単語と単語を結びつけ、「単語のネットワーク」として捉えなければ、将来の予兆などの重要な知識は得られないという限界がある。

たとえば、「ダム建設」という単語だけでは「建設賛成」なのか「建設反対」なのか分からない。また、「反対」の理由が「生態系保全」のためなのか「水需要変化」という点からなのか、分からない。単語をネットワークとして捉えなければ、この問題は克服できない。さらに、高出現頻度の一般的な単語により具体的な意味を持たせることができなければ、重要な知見に結びつけることはできない。つまり、「単語のネットワーク」を発見し、高出現頻度の単語に具体的な意味を持たせることによって、発見が難しい低出現頻度の重要単語の発見が可能になるのである。

単純な高出現単語と出現単語の共起関係分析の限界

単語のネットワークは「単語」と「単語」の共起関係を見ることによって知ることができる。文章の構造分析により「係り受け関係」の解明が進みつつあるが、しかしこの現状は不十分であり、現段階では共起関係を分析することが望ましい。

しかし、表4-7に「高出現単語」の「ダム」と「出現単語」の単純な共起関係が示すように、この中の語句集団に潜在的な意味をもった単語が潜んでいる可能性は非常に高いが、このような単語列挙の状況から、潜在的意味に気付くことは困難である。

これは、テキストマイニングは語句をマイニングするのではなく、周辺語句の位置関係から意味の発見を支援することを示唆している。ここで、新たに開発しなければならない分析手法は、意味の共起を探り出す手法である。

単順な副詞・形容や否定を表す出現単語と他の出現単語の共起関係分析の限界

副詞・形容や否定を表す単語の共起による単語ネットワークへの「気づき」の支援することも考えられる。図4-6に、副詞・形容や否定を表す単語と高出現頻度語の共起状況を示す。この図は、1992年の毎日新聞全記事を対象として抽出した22種類の副詞・形容や否定を表す単語を横軸、1992年の毎日新聞全記事を対象として抽出した100種類の高出現頻度語を縦軸、そして22種類の副詞・形容や否定を表す単語と100種類の高出現頻度語の共起回数を高さ方向の軸に示している。なお、22種類の副詞・形容や否定を表す単語と100種類の高出現頻度語の共起関係については、附属資料の附表4-3に示す。

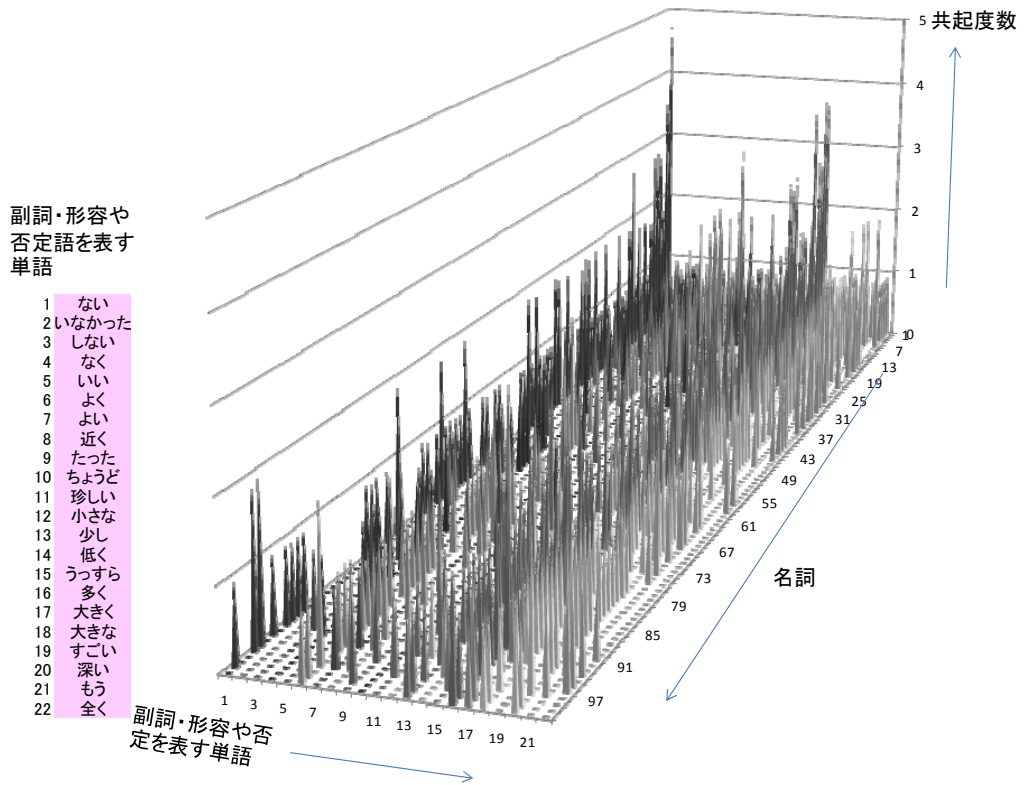
これは副詞・形容や否定を表す単語を含む記事には、潜在的意味を含んだ固有単語が潜んでいる可能性が高いという仮定に基づいて分析したものである。例えば「ない」という否定を表す語句が、高出現頻度語を多く共起させる傾向があるが、これだけでは重要な出現単語を分別することは困難である。なおこの分析は、単に共起関係を調べているだけであるので、これを正確に分析することは今後の課題として残されている。

表4-11 「ダム」と出現単語の単純な共起関係（1992年）

単語	共起数	単語	共起数	単語	共起数
自治体-都道府縣市系	6	地域-地域系	2	メートル	1
事業-工事系	5	全国-日本系	2	使用	1
河川-川系	3	中	2	設置	1
危険-非常系	3	調査	2	現場	1
反対	3	市民-住民系	2	今	1
		水	2	山	1
		計画-予定系	2	対策-対策系	1
		指摘	2	中央-中心系	1
		人	2	グループ	1
		公開	2	コンクリート	1
		高枝	2	プロジェクト	1
		自然	2	ミリ	1
		上流域-上流系	2	期待-希望系	1
		請求	2	恐れ	1
				公園	1
				降雨-雨系	1
				最高	1
				昨年	1
				実際	1
				取水	1
				手	1
				首	1
				女性	1
				条例	1
				人々	1
				水面	1
				成長	1
				専門家	1
				地質	1
				調べ	1
				土	1
				土地	1
				買収	1
				判断	1
				本社	1
				来年	1
				連続	1

註：この表は、「ダム」と「出現単語」の単純な共起関係を示したものである。

(筆者作成)



(筆者作成)

図4-6 副詞・形容や否定を表す単語と高出現頻度語の共起状況 (1992年)

5-2 効果的にテキストマイニングを実施するための文字情報データの収集方法と分析のための辞書作成の留意点

ここでは、これまで述べてきた一般化された情報源の分析の手法とその限界を踏まえ、テキストマイニングによる分析を前提として、目的を明らかにしたアンケート収集時の留意点と辞書作成上に留意することを述べる。

アンケートは、設問の方法を十分に検討することである。具体的には、分析目的に合致した文章を収集する工夫をする、回答を誘導しない、思うままに書いて頂く、事実を書いて頂くことである。

例えば、

(わるい例) あなたは、渇水の被害によく遭いますか？

(わるい例) あなたが必要と思う洪水や渇水の被害を軽減するためのインフラは何ですか？

(よい例) 最近5年間で、何回渇水の被害に遭いましたか？

(よい例) あなたは、洪水や渇水の被害を少なくするためにはどのような対策が必要と思いますか？

特に数量を含んだ辞書作成上に留意することは、松下ら（2005）も指摘しているが、

1) 統計量を必ずしも直接的に抽出できない。

[文例] 10月の雨量は前年の10月よりも50%上多かった。

雨量と時期を推定することが必要。

2) 記事間で単位が一致していない。

[文例] 月平均雨量 150mm、年平均雨量 1,800mm。

単位をそろえる処理が必要。

3) あいまいな数値を含んでいる。

[文例] 日雨量約 50mm。日雨量 40-60mm。

記事内容との整合性を考慮しながら値の推定が必要。

ということも必要となる。

6. 今後の研究の発展可能性

「単語のネットワーク」の発見、「高出現頻度の単語に具体的な意味をもたせること」によって、発見が難しい低出現頻度の重要単語の発見に近づける。しかし、本研究では、単語の出現数に着目し多変量解析による分析をしている段階で、まだ単語の意味分析までは到達していない。これは、現在多くの研究者が模索していることである。

多変量解析によるテキストマイニングの限界を克服するためには、例えばグラフ理論、ニューラルネットワーク理論、潜在的意味解析などの適用が必要となる。これらの取り組みにおいても、発見が難しい低出現頻度の重要単語の発見に近づくためには、分析目的に応じた情報源の加工作業である「辞書」を整備することは重要になる。分析対象に応じた辞書を整備することは、不要なノイズを切り捨て、重要な単語を抽出し、分析目的に合致した意味や評価を行うためには不可欠である。そのため、分析手法の開発と併せて辞書整備への取り組みが、益々重要となる。

今後、文字情報データベースや辞書の構築を図り、テキストマイニングによる単語ネットワーク分析の研究を進めることにより、様々な切り口から日本国内外の河川や道路に関する建設・管理事業に対するニーズやシーズの分析支援を期待できる。

註

中村健太郎 (2008) 「潜在的意味解析」 豊田秀樹監修『データマイニング入門—R で学ぶ最新データ解析』東京図書, pp. 271-272.

参考文献

- Makoto Suzuki (2008) Text Categorization Using the Maximum Ratio of Term Frequency, *J Jpn Ind Manage Assoc*, 58, pp. 438-444.
- 石田基広 (2008) 『Rによるテキストマイニング入門』森北出版.
- 市村由美, 長谷川隆明, 渡部勇, 佐藤光弘 (2001) 「テキストマイニング—事例紹介」『人工知能学会誌』人工知能学会, No. 16, Vo. 2, pp. 192-199.
- 乾孝司, 奥村学 (2005) 「文書内に現れる因果関係に出現特性調査」, 『情報処理学会研究報告』, 情報処理学会, pp. 81-88.
- 大澤幸生 (2003) 『チャンス発見の情報技術』東京電機大学出版局.
- 大澤幸生 (2006) 『チャンス発見のデータ分析』東京電機大学出版局.
- 河合英紀, 齋藤悠, 土田正明, 水口弘紀, 國枝和雄, 山田敬嗣 (2008) 「新聞記事における統計量表現の共起ネットワーク」『第22回人工知能学会全国大会論文集』人工知能学会, 3K3-10.
- 川前徳章, 青木輝勝, 安田浩 (2002) 「統計的潜在的意味空間の抽出」『自然言語処理(148-4)』, 情報処理学会, pp. 25-30.
- 斉藤公一, 迫田昭人, 中江富人, 岩井禎広, 田村直良, 中川裕志 (1998) 「数値情報をキーとした新聞記事からの情報抽出」『情報処理学会研究報告(98-NL-125)』情報処理学会, vol. 98, No. 48, pp. 63-69.
- 佐藤浩史, 笠原要, 松澤和光 (1999) 「テキスト上の表層的因果知識の獲得とその応用」『電子情報通信学会技術研究報告 (TL98-23)』電子情報通信学会, pp. 27-34.
- 佐藤岳文, 堀田昌英 (2006) 「Web マイニングを用いた因果ネットワークの自動構築手法の開発」『社会技術研究論文集』社会技術研究会, Vol. 4, pp. 66-74.

- 白井康之, 小関悠, 小池亜弥(2009)「テキストマイニングによるトレンド情報抽出環境の構築技術」
『三菱総合研究所所報』三菱総合研究所, No. 51, pp. 110-123.
- 新納浩幸(2007)『Rで学ぶクラスター分析』オーム社.
- 高橋由光(2005)「ネットワーク分析を利用したテキスト・マイニング-2004年の朝日新聞を事例にして-」『2005年人文科学とコンピュータシンポジウム』情報処理学会, pp. 35-40.
- 那須川哲哉(2006)『テキストマイニングを使う技術/作る技術』東京電機大学出版局.
- 野村総合研究所(2008)『顧客の声分析・活用術』リックテレコム.
- 田村正紀(2006)『リサーチデザイン』白桃書房.
- 藤畑勝之, 志賀正裕, 森辰則(2001)「係り受けの制約と優先規則に基づく数量表現抽出」『情報処理学会研究報告(2001-FI-64, 2001-NL-145)』情報処理学会, pp. 119-125.
- 松下光範, 加藤恒昭(2005)「動向情報に基づく情報可視化の基礎検討」『第19回人工知能学会全国大会論文集』, 人工知能学会, IE3-03.
- 森辰則, 藤岡篤史, 村田一郎(2007)「動向情報編纂のためのテキストからの統計量の自動抽出」『第21回人工知能学会論文集』人工知能学会, 3H9-4.
- 和多太樹, 関隆宏, 田中省作, 廣川左千男(2005)「単語の出現頻度に着目した病院評判情報の分析」
『研究報告-音声言語情報処理(SLP)』情報処理学会, vol.2005, No. 50, pp. 15-20.

以上

附属資料

本文記載を省略したデータおよび分析結果を附属資料としてとりまとめた。

附表4-1 (1/4) 置換辞書 (1992年、1996年、2000年)

番号	置換後	置換前	番号	置換後	置換前
1	議会-議会審議会系	協議会	41	河川-川系	犀川
2	議会-議会審議会系	市議会	42	河川-川系	糸魚川
3	議会-議会審議会系	審議会	43	河川-川系	手取川
4	議会-議会審議会系	国会	44	河川-川系	小瀬川
5	議会-議会審議会系	協議	45	河川-川系	石狩川
6	計画-予定系	計画	46	河川-川系	川
7	計画-予定系	予定	47	河川-川系	川辺川
8	計画-予定系	将来	48	河川-川系	相模川
9	水道水-水道系	上水道	49	河川-川系	多摩川
10	水道水-水道系	水道	50	河川-川系	太田川
11	水道水-水道系	水道水	51	河川-川系	筑後川
12	水道水-水道系	飲料水	52	河川-川系	猪名川
13	貯水-貯水系	貯水	53	河川-川系	長良川
14	貯水-貯水系	貯水池	54	河川-川系	柘植川
15	貯水量-貯水率	貯水量	55	河川-川系	田川
16	貯水量-貯水率	貯水率	56	河川-川系	道頓堀川
17	保全-保護系	保全	57	河川-川系	那賀川
18	保全-保護系	保護	58	河川-川系	白子川
19	汚染-汚濁系	汚染	59	河川-川系	薄川
20	汚染-汚濁系	汚濁	60	河川-川系	飛騨川
21	下流域-下流系	下流	61	河川-川系	姫川
22	下流域-下流系	下流域	62	河川-川系	服部川
23	下流域-下流系	川下流	63	河川-川系	平作川
24	下流域-下流系	川下	64	河川-川系	名取川
25	河川-川系	ライン川	65	河川-川系	名張川
26	河川-川系	綾瀬川	66	河川-川系	木曾川
27	河川-川系	伊丹生川	67	河川-川系	木津川
28	河川-川系	引地川	68	河川-川系	野川
29	河川-川系	宇陀川	69	河川-川系	淀川
30	河川-川系	河川	70	河川-川系	利根川
31	河川-川系	吉井川	71	河川-川系	鴨川
32	河川-川系	吉野川	72	河川-川系	安威川
33	河川-川系	球磨川	73	改善-改善系	改修
34	河川-川系	隅田川	74	改善-改善系	改正案
35	河川-川系	桂川	75	改善-改善系	改善
36	河川-川系	江戸川	76	改善-改善系	改善策
37	河川-川系	荒川	77	改善-改善系	改定
38	河川-川系	黒部川	78	改善-改善系	改定案
39	河川-川系	沙流川	79	改善-改善系	改良
40	河川-川系	最上川	80	環境省-本省本局系	環境省

(筆者作成)

附表4-1 (2/4) 置換辭書 (1992年、1996年、2000年)

番号	置換後	置換前	番号	置換後	置換前
81	環境省-本省本局系	環境庁	121	国土交通省-本省本局系	地方建設局
82	環境省-本省本局系	保全局	122	国土交通省-本省本局系	国土交通省
83	危険-非常系	危険	123	国土交通省-本省本局系	整備局
84	危険-非常系	危機	124	国土交通省-本省本局系	地方整備局
85	危険-非常系	危険性	125	市民-住民系	住民
86	危険-非常系	危惧	126	市民-住民系	市民
87	危険-非常系	非常時	127	市民-住民系	庶民
88	危険-非常系	問題	128	施設-設備系	施設
89	危険-非常系	深刻	129	施設-設備系	設備
90	危険-非常系	重大	130	事業-工事系	事業
91	危険-非常系	訴訟	131	事業-工事系	工事
92	危険-非常系	捜査	132	事業-工事系	建設
93	危険-非常系	課題	133	自治体-都道府縣市系	愛知県
94	危険-非常系	応急	134	自治体-都道府縣市系	岡山県
95	期待-希望系	期待	135	自治体-都道府縣市系	沖繩県
96	期待-希望系	希望	136	自治体-都道府縣市系	岐阜県
97	期待-希望系	夢	137	自治体-都道府縣市系	京都府
98	結果-結論	結果	138	自治体-都道府縣市系	熊本県
99	結果-結論	結論	139	自治体-都道府縣市系	群馬県
100	原因-理由系	原因	140	自治体-都道府縣市系	県
101	原因-理由系	理由	141	自治体-都道府縣市系	県庁
102	降雨-雨系	わか雨	142	自治体-都道府縣市系	広島県
103	降雨-雨系	雨	143	自治体-都道府縣市系	高知県
104	降雨-雨系	雨期	144	自治体-都道府縣市系	埼玉県
105	降雨-雨系	雨水	145	自治体-都道府縣市系	三重県
106	降雨-雨系	雨不足	146	自治体-都道府縣市系	山形県
107	降雨-雨系	雨模様	147	自治体-都道府縣市系	山口県
108	降雨-雨系	雨量	148	自治体-都道府縣市系	山梨県
109	降雨-雨系	雨量計	149	自治体-都道府縣市系	滋賀県
110	降雨-雨系	降雨	150	自治体-都道府縣市系	秋田県
111	降雨-雨系	降雨量	151	自治体-都道府縣市系	新潟県
112	降雨-雨系	降水量	152	自治体-都道府縣市系	神奈川県
113	降雨-雨系	豪雨	153	自治体-都道府縣市系	静岡県
114	降雨-雨系	秋雨	154	自治体-都道府縣市系	石川県
115	降雨-雨系	小雨	155	自治体-都道府縣市系	千葉県
116	降雨-雨系	大雨	156	自治体-都道府縣市系	大阪府
117	降雨-雨系	長雨	157	自治体-都道府縣市系	長崎県
118	降雨-雨系	梅雨	158	自治体-都道府縣市系	長野県
119	国土交通省-本省本局系	建設省	159	自治体-都道府縣市系	鳥取県
120	国土交通省-本省本局系	建設局	160	自治体-都道府縣市系	島根県

(筆者作成)

附表4-1 (3/4) 置換辞書 (1992年、1996年、2000年)

番号	置換後	置換前	番号	置換後	置換前
161	自治体-都道府縣市系	東京都	201	自治体-都道府縣市系	川崎市
162	自治体-都道府縣市系	徳島県	202	自治体-都道府縣市系	川西市
163	自治体-都道府縣市系	奈良県	203	自治体-都道府縣市系	泉南市
164	自治体-都道府縣市系	富山県	204	自治体-都道府縣市系	前橋市
165	自治体-都道府縣市系	福井県	205	自治体-都道府縣市系	草津市
166	自治体-都道府縣市系	福島県	206	自治体-都道府縣市系	草津市沖
167	自治体-都道府縣市系	兵庫県	207	自治体-都道府縣市系	大宮市
168	自治体-都道府縣市系	北海道	208	自治体-都道府縣市系	大月市
169	自治体-都道府縣市系	六都県	209	自治体-都道府縣市系	大阪市
170	自治体-都道府縣市系	和歌山県	210	自治体-都道府縣市系	大津市
171	自治体-都道府縣市系	茨城県	211	自治体-都道府縣市系	大東市
172	自治体-都道府縣市系	岩手県	212	自治体-都道府縣市系	長岡市
173	自治体-都道府縣市系	ソウル市	213	自治体-都道府縣市系	長野市
174	自治体-都道府縣市系	阿南市	214	自治体-都道府縣市系	田辺市
175	自治体-都道府縣市系	伊那市	215	自治体-都道府縣市系	藤沢市
176	自治体-都道府縣市系	宇治市	216	自治体-都道府縣市系	徳島市
177	自治体-都道府縣市系	横須賀市	217	自治体-都道府縣市系	奈良市
178	自治体-都道府縣市系	横浜市	218	自治体-都道府縣市系	南市
179	自治体-都道府縣市系	桶川市	219	自治体-都道府縣市系	南市籍
180	自治体-都道府縣市系	貝塚市	220	自治体-都道府縣市系	尼崎市
181	自治体-都道府縣市系	岸和田市	221	自治体-都道府縣市系	八幡市
182	自治体-都道府縣市系	岩国市	222	自治体-都道府縣市系	八尾市
183	自治体-都道府縣市系	京都市	223	自治体-都道府縣市系	富士宮市
184	自治体-都道府縣市系	橋本市	224	自治体-都道府縣市系	福岡市
185	自治体-都道府縣市系	金沢市	225	自治体-都道府縣市系	北本市
186	自治体-都道府縣市系	広島市	226	自治体-都道府縣市系	枚方市
187	自治体-都道府縣市系	堺市	227	自治体-都道府縣市系	箕面市
188	自治体-都道府縣市系	三田市	228	自治体-都道府縣市系	名古屋市
189	自治体-都道府縣市系	市	229	自治体-都道府縣市系	紋別市
190	自治体-都道府縣市系	市原市	230	自治体-都道府縣市系	和歌山市
191	自治体-都道府縣市系	志木市	231	自治体-都道府縣市系	自治体
192	自治体-都道府縣市系	小野市	232	自治体-都道府縣市系	西淀川区
193	自治体-都道府縣市系	松本市	233	自治体-都道府縣市系	江戸川区
194	自治体-都道府縣市系	神戸市	234	自治体-都道府縣市系	江東区
195	自治体-都道府縣市系	神戸市灘区	235	自治体-都道府縣市系	此花区
196	自治体-都道府縣市系	人吉市	236	自治体-都道府縣市系	西区
197	自治体-都道府縣市系	清水市	237	自治体-都道府縣市系	福島区
198	自治体-都道府縣市系	西宮市	238	自治体-都道府縣市系	麻生区
199	自治体-都道府縣市系	摂津市	239	自治体-都道府縣市系	港区
200	自治体-都道府縣市系	千葉市	240	自治体-都道府縣市系	千代田区

(筆者作成)

附表4-1 (4/4) 置換辞書 (1992年、1996年、2000年)

番号	置換後	置換前	番号	置換後	置換前
241	自治体-都道府縣市系	中央区	281	地域-地域系	東京
242	自治体-都道府縣市系	北区	282	地域-地域系	大阪
243	自治体-都道府縣市系	墨田区	283	地域-地域系	名古屋
244	自治体-都道府縣市系	世田谷区	284	地域-地域系	埼玉
245	自治体-都道府縣市系	川崎区	285	地域-地域系	福岡
246	自治体-都道府縣市系	多摩区	286	地域-地域系	仙台
247	自治体-都道府縣市系	大田区	287	地域-地域系	横浜
248	自治体-都道府縣市系	伏見区	288	地域-地域系	岐阜
249	自治体-都道府縣市系	目黒区	289	地域-地域系	水源県
250	上流域-上流系	上流	290	地域-地域系	都道府県
251	上流域-上流系	上流域	291	地域-地域系	二十三区
252	上流域-上流系	川上流	292	地域-地域系	八区
253	上流域-上流系	川上	293	地域-地域系	府県
254	水不足-水不足系	渇水	294	地域-地域系	管区
255	水不足-水不足系	渇水時	295	地域-地域系	区域
256	水不足-水不足系	水不足	296	地域-地域系	区画
257	水不足-水不足系	水不足ぎみ	297	地域-地域系	区間
258	水不足-水不足系	節水	298	地下水-地下水系	井戸水
259	水不足-水不足系	節水型	299	地下水-地下水系	地下水
260	水不足-水不足系	大渇水	300	中央-中心系	中央
261	全国-日本系	全国	301	中央-中心系	中心
262	全国-日本系	日本	302	通常-平常系	平常
263	全国-日本系	国	303	通常-平常系	通常
264	対策-対策系	対策	304	通常-平常系	平常時
265	対策-対策系	対策案	305	通常-平常系	常時
266	対策-対策系	対策課	306	通常-平常系	安全
267	対策-対策系	対策室			
268	対策-対策系	対策用			
269	対策-対策系	対応			
270	対策-対策系	対応策			
271	地域-地域系	県南部			
272	地域-地域系	堺市草			
273	地域-地域系	地元			
274	地域-地域系	地域			
275	地域-地域系	2県			
276	地域-地域系	3都県			
277	地域-地域系	4県			
278	地域-地域系	5県			
279	地域-地域系	7県			
280	地域-地域系	都県			

(筆者作成)

附表4-2 (1/3) 置換辞書 (2008/2009年)

番号	置換後	置換前	番号	置換後	置換前
1	議会-議会審議会系	協議会	41	改善-改善系	改善
2	議会-議会審議会系	市議会	42	改善-改善系	改善策
3	議会-議会審議会系	審議会	43	改善-改善系	改定
4	議会-議会審議会系	国会	44	改善-改善系	改定案
5	議会-議会審議会系	協議	45	改善-改善系	改良
6	計画-予定系	計画	46	改善-改善系	修復
7	計画-予定系	予定	47	改善-改善系	修理
8	計画-予定系	将来	48	環境省-本省本局系	環境省
9	水道水-水道系	上水道	49	環境省-本省本局系	環境庁
10	水道水-水道系	水道	50	環境省-本省本局系	保全局
11	水道水-水道系	水道水	51	危険-非常系	危険
12	水道水-水道系	飲料水	52	危険-非常系	危機
13	貯水-貯水系	貯水	53	危険-非常系	危険性
14	貯水-貯水系	貯水池	54	危険-非常系	危惧
15	貯水-貯水系	ダム湖	55	危険-非常系	非常時
16	貯水-貯水系	ギントウン湖	56	危険-非常系	問題
17	貯水量-貯水率	貯水量	57	危険-非常系	深刻
18	貯水量-貯水率	貯水率	58	危険-非常系	重大
19	保全-保護系	保全	59	危険-非常系	訴訟
20	保全-保護系	保護	60	危険-非常系	捜査
21	汚染-汚濁系	汚染	61	危険-非常系	課題
22	汚染-汚濁系	汚濁	62	危険-非常系	応急
23	下流域-下流系	下流	63	期待-希望系	期待
24	下流域-下流系	下流域	64	期待-希望系	希望
25	下流域-下流系	川下流	65	期待-希望系	夢
26	下流域-下流系	川下	66	結果-結論	結果
27	河川-川系	河川	67	結果-結論	結論
28	河川-川系	川	68	原因-理由系	原因
29	河川-川系	河川	69	原因-理由系	理由
30	河川-川系	ソロ川	70	降雨-雨系	わか雨
31	河川-川系	チタルム川	71	降雨-雨系	雨
32	河川-川系	チウジュン川	72	降雨-雨系	雨期
33	河川-川系	チリウン川	73	降雨-雨系	雨水
34	洪水系	洪水	74	降雨-雨系	雨不足
35	洪水系	浸水	75	降雨-雨系	雨模様
36	洪水系	冠水	76	降雨-雨系	雨量
37	本部系	本部	77	降雨-雨系	雨量計
38	本部系	本社	78	降雨-雨系	降雨
39	改善-改善系	改修	79	降雨-雨系	降雨量
40	改善-改善系	改正案	80	降雨-雨系	降水量

(筆者作成)

附表4-2 (2/3) 置換辞書 (2008/2009年)

番号	置換後	置換前	番号	置換後	置換前
81	降雨-雨系	豪雨	121	自治体-州縣市系	ボジョネゴロ県
82	降雨-雨系	秋雨	122	自治体-州縣市系	セラノ県
83	降雨-雨系	小雨	123	自治体-州縣市系	西カリマンタン州
84	降雨-雨系	大雨	124	自治体-州縣市系	ワリククン村
85	降雨-雨系	長雨	125	自治体-州縣市系	ルバック県
86	降雨-雨系	梅雨	126	自治体-州縣市系	マラン県
87	降雨-雨系	雨	127	自治体-州縣市系	マドゥラ島
88	公共事業省	事業省	128	自治体-州縣市系	ボウォ州
89	市民-住民系	住民	129	自治体-州縣市系	プロラ県
90	市民-住民系	市民	130	自治体-州縣市系	ファウジ州
91	市民-住民系	庶民	131	自治体-州縣市系	パブア州
92	施設-設備系	施設	132	自治体-州縣市系	バンダアチェ市
93	施設-設備系	設備	133	自治体-州縣市系	バリ州
94	事業-工事系	事業	134	自治体-州縣市系	チャレナン郡
95	事業-工事系	工事	135	自治体-州縣市系	ジョクジャカルタ
96	事業-工事系	建設	136	自治体-州縣市系	ジュンブル県
97	自治体-州縣市系	州	137	自治体-州縣市系	シドアルジョ県
98	自治体-州縣市系	特別州	138	自治体-州縣市系	カリマンタン島
99	自治体-州縣市系	バンテン州	139	自治体-州縣市系	アガム県
100	自治体-州縣市系	西ジャワ州	140	上流域-上流系	上流
101	自治体-州縣市系	東ジャワ州	141	上流域-上流系	上流域
102	自治体-州縣市系	東ジャカルタ	142	上流域-上流系	川上流
103	自治体-州縣市系	北ジャカルタ	143	上流域-上流系	川上
104	自治体-州縣市系	南ジャカルタ	144	水不足-水不足系	渇水
105	自治体-州縣市系	州政府	145	水不足-水不足系	渇水時
106	自治体-州縣市系	ジャワ州	146	水不足-水不足系	水不足
107	自治体-州縣市系	西ジャカルタ	147	水不足-水不足系	水不足ぎり
108	自治体-州縣市系	ジャワ島	148	水不足-水不足系	節水
109	自治体-州縣市系	タンゲラン県	149	水不足-水不足系	節水型
110	自治体-州縣市系	アチェ州	150	水不足-水不足系	大渇水
111	自治体-州縣市系	東カリマンタン州	151	全国-インドネシア系	全国
112	自治体-州縣市系	西スマトラ州	152	全国-インドネシア系	インドネシア
113	自治体-州縣市系	ボゴール県	153	全国-インドネシア系	国家
114	自治体-州縣市系	北スマトラ州	154	全国-インドネシア系	国
115	自治体-州縣市系	南スマトラ州	155	対策-対策系	対策
116	自治体-州縣市系	タンゲラン郡	156	対策-対策系	対策案
117	自治体-州縣市系	南カリマンタン州	157	対策-対策系	対策課
118	自治体-州縣市系	東ジャワ	158	対策-対策系	対策室
119	自治体-州縣市系	西スラウェシ州	159	対策-対策系	対策用
120	自治体-州縣市系	マンダラ県	160	対策-対策系	対応

(筆者作成)

附表4-2 (3/3) 置換辞書 (2008/2009年)

番号	置換後	置換前	番号	置換後	置換前
161	対策-対策系	対応策	201	地域-地域系	スラウェシ
162	地域-地域系	県南部	202	地域-地域系	スナヤン
163	地域-地域系	地元	203	地域-地域系	西ジャワ
164	地域-地域系	地域	204	地域-地域系	バンダアチェ
165	地域-地域系	水源県	205	地域-地域系	マラン
166	地域-地域系	管区	206	地域-地域系	南カリマンタン
167	地域-地域系	区域	207	地域-地域系	デンバサール
168	地域-地域系	区画	208	地域-地域系	クバヨランバル
169	地域-地域系	区間	209	地域-地域系	カンブン
170	地域-地域系	各地	210	大統領系	大統領
171	地域-地域系	ジャワ	211	大統領系	ユドヨノ
172	地域-地域系	スマトラ	212	地下水-地下水系	井戸水
173	地域-地域系	スマトラ島	213	地下水-地下水系	地下水
174	地域-地域系	クバヨラン	214	中央-中心系	中央
175	地域-地域系	クタ	215	中央-中心系	中心
176	地域-地域系	パダン	216	通常-平常系	平常
177	地域-地域系	ジャカルタ	217	通常-平常系	通常
178	地域-地域系	スラバヤ	218	通常-平常系	平常時
179	地域-地域系	アチェ	219	通常-平常系	常時
180	地域-地域系	ポンドック	220	通常-平常系	安全
181	地域-地域系	チャワン			
182	地域-地域系	チブタット			
183	地域-地域系	タンジュン			
184	地域-地域系	ボゴール			
185	地域-地域系	タンゲラン			
186	地域-地域系	メンテン			
187	地域-地域系	ムラユ			
188	地域-地域系	プロガドゥン			
189	地域-地域系	ブラバタン			
190	地域-地域系	ブサンガン			
191	地域-地域系	ブカロンガン			
192	地域-地域系	サヌール			
193	地域-地域系	ファトマワティ			
194	地域-地域系	スマラン			
195	地域-地域系	サマリンド			
196	地域-地域系	プカシ			
197	地域-地域系	バンドン			
198	地域-地域系	バリ島			
199	地域-地域系	ソロ			
200	地域-地域系	メダン			

(筆者作成)

附表4-3 (1/12) 副詞・形容や否定を表す単語と高出現頻度の共起 (1992年)

	1	2	3	4	5	6	7	8
	ない	多く	近く	もう	よく	深い	珍しい	しい
1 自治体-都道府県市系	5	4	3	2	2	2	2	1
2 事業-工事系	4	4	2	1	1	1	1	1
3 メートル	2	3	1	1	2	2	1	0
4 地域-地域系	3	3	1	0	1	1	0	0
5 ダム	3	1	2	1	0	0	1	1
6 全国-日本系	3	4	0	1	1	1	1	1
7 河川-川系	2	2	1	0	1	1	1	0
8 中	1	1	2	1	2	2	1	0
9 調査	2	2	1	0	1	1	1	0
10 危険-非常系	1	2	2	1	1	1	0	1
11 使用	2	2	0	0	1	1	0	0
12 市民-住民系	1	1	2	1	1	1	2	0
13 水	3	3	0	1	0	0	0	1
14 設置	2	3	0	0	1	1	0	0
15 地下	1	3	0	0	1	1	0	0
16 地下水-地下水系	1	2	0	0	1	1	0	0
17 計画-予定系	2	1	1	0	0	0	1	0
18 結果-結論	1	2	0	0	1	1	0	0
19 現場	0	1	2	1	2	2	1	0
20 今	1	1	0	0	1	1	0	0
21 山	2	2	1	2	1	1	1	1
22 指摘	1	1	1	0	1	1	1	0
23 人	1	1	2	2	1	1	1	1
24 対策-対策系	2	1	0	0	0	0	0	0
25 中央-中心系	1	2	0	0	1	1	1	0

(筆者作成)

附表4-3 (2/12) 副詞・形容や否定を表す単語と高出現頻度の共起 (1992年)

		1	2	3	4	5	6	7	8
		ない	多く	近く	もう	よく	深い	珍しい	いい
26	排水	2	1	0	0	0	0	0	0
27	反対	2	1	1	1	0	0	1	1
28	物質	2	2	0	0	1	1	0	0
29	利用	1	3	0	0	1	1	0	0
30	JR	0	1	0	0	1	1	0	0
31	キロ	1	2	0	0	1	1	0	0
32	グループ	2	2	0	1	0	0	0	1
33	コンクリート	0	1	1	0	1	1	0	0
34	プロジェクト	2	1	0	0	0	0	1	0
35	ミリ	2	0	0	0	0	0	0	0
36	安定	1	2	0	0	0	0	0	0
37	以前	1	1	0	0	1	1	0	0
38	井戸	1	1	0	0	0	0	0	0
39	一部	1	1	0	0	1	1	0	0
40	汚染-汚濁系	2	1	0	0	0	0	0	0
41	開発	1	2	0	0	1	1	0	0
42	期待-希望系	2	1	0	0	0	0	0	0
43	恐れ	1	0	1	0	0	0	0	0
44	空間	1	2	0	0	1	1	0	0
45	掘削	0	1	0	0	1	1	0	0
46	形	0	1	1	1	2	2	1	0
47	経済	0	2	0	0	1	1	0	0
48	検討	0	2	0	0	1	1	0	0
49	公園	1	1	0	0	0	0	0	0
50	公開	1	0	1	0	0	0	1	0

(筆者作成)

附表4-3 (3/12) 副詞・形容や否定を表す単語と高出現頻度の共起 (1992年)

		1	2	3	4	5	6	7	8
		ない	多く	近く	もう	よく	深い	珍しい	いい
51	工業	0	2	0	0	1	1	0	0
52	工場	1	1	0	0	0	0	0	0
53	降雨-雨系	1	1	0	0	0	0	0	0
54	高校	0	0	2	0	0	0	0	0
55	高潮	0	2	0	0	1	1	0	0
56	今年	1	1	0	0	0	0	0	0
57	最高	2	0	0	0	0	0	0	0
58	最大	1	2	0	0	0	0	0	0
59	作業	1	0	0	0	0	0	0	0
60	昨年	1	0	1	0	0	0	0	0
61	市内	1	1	0	0	1	1	0	0
62	自然	2	1	0	1	0	0	1	1
63	実際	1	2	0	1	1	1	0	1
64	遮水壁	0	1	0	0	1	1	0	0
65	取水	1	1	0	0	1	1	0	0
66	手	1	0	1	1	1	1	2	0
67	首	0	0	2	1	1	1	1	0
68	出水	0	1	0	0	1	1	0	0
69	所	0	2	0	0	1	1	0	0
70	女性	0	0	2	1	1	1	1	0
71	上流域-上流系	2	0	0	0	0	0	1	0
72	条例	0	1	1	0	0	0	0	0
73	人々	1	1	1	2	1	1	1	1
74	水面	0	1	0	0	0	0	0	0
75	成長	1	2	0	1	0	0	0	1

(筆者作成)

附表 4-3 (4/12) 副詞・形容や否定を表す単語と高出現頻度の共起 (1992年)

		1	2	3	4	5	6	7	8
		ない	多く	近く	もう	よく	深い	珍しい	しい
76	請求	1	0	1	0	0	0	1	0
77	専門家	1	1	0	0	1	1	1	0
78	戦後	0	2	0	0	1	1	0	0
79	大量	0	1	0	0	1	1	0	0
80	地価	0	2	0	0	1	1	0	0
81	地質	0	1	1	0	1	1	0	0
82	地盤	0	2	0	0	1	1	0	0
83	中心部	0	2	0	0	1	1	0	0
84	調べ	0	1	1	0	0	0	0	0
85	沈下	0	2	0	0	1	1	0	0
86	鉄道	0	1	0	0	1	1	0	0
87	土	0	1	1	0	1	1	0	0
88	土地	1	1	0	0	1	1	1	0
89	買収	1	1	0	0	1	1	1	0
90	発展	1	2	0	0	1	1	0	0
91	判断	1	0	1	0	0	0	0	0
92	復興	0	2	0	0	1	1	0	0
93	幅	1	1	1	1	1	1	1	0
94	噴出	0	1	0	0	1	1	0	0
95	本社	2	1	0	0	0	0	0	0
96	有害	2	1	0	0	0	0	0	0
97	用水	0	1	1	1	2	2	1	0
98	来年	0	0	0	0	0	0	0	0
99	連続	1	1	0	0	1	1	0	0
100	連絡線	0	1	0	0	1	1	0	0

(筆者作成)

附表4-3 (5/12) 副詞・形容や否定を表す単語と高出現頻度の共起 (1992年)

		9	10	11	12	13	14	15	16
		いなかった	うっすら	しない	すごい	たった	ちょうど	なく	よい
1	自治体-都道府県市系	1	1	1	1	1	1	1	1
2	事業-工事系	1	0	1	1	1	1	1	0
3	メートル	0	1	0	0	0	0	1	1
4	地域-地域系	0	0	1	0	1	0	1	0
5	タム	1	1	1	1	1	1	0	0
6	全国-日本系	1	0	0	1	0	1	1	0
7	河川-川系	0	0	1	0	1	0	1	0
8	中	0	0	1	0	1	0	0	1
9	調査	0	0	1	0	0	0	1	0
10	危険-非常系	1	0	1	1	0	1	0	0
11	使用	0	0	0	0	1	0	1	0
12	市民-住民系	0	0	1	0	0	0	1	1
13	水	1	0	0	1	1	1	1	0
14	設置	0	0	0	0	1	0	1	0
15	地下	0	0	0	0	0	0	1	0
16	地下水-地下水系	0	0	0	0	0	0	1	0
17	計画-予定系	0	0	1	0	0	0	0	0
18	結果-結論	0	0	0	0	0	0	1	0
19	現場	0	0	0	0	0	0	0	1
20	今	0	1	0	0	0	0	0	0
21	山	1	0	0	1	0	1	0	1
22	指摘	0	0	1	0	0	0	0	0
23	人	1	0	1	1	0	1	0	1
24	対策-対策系	0	0	0	0	1	0	1	0
25	中央-中心系	0	0	0	0	0	0	1	0

(筆者作成)

附表4-3 (6/12) 副詞・形容や否定を表す単語と高出現頻度の共起 (1992年)

		9	10	11	12	13	14	15	16
		いなかった	うっすら	しない	すごい	たった	ちょうど	なく	よい
26	排水	0	0	0	0	0	0	0	0
27	反対	1	0	1	1	0	1	0	0
28	物質	0	0	0	0	0	0	0	0
29	利用	0	0	0	0	0	0	1	0
30	JR	0	0	0	0	0	0	0	0
31	キロ	0	0	0	0	0	0	0	0
32	グループ	1	0	0	1	0	1	0	0
33	コンクリート	0	0	0	0	0	0	0	0
34	プロジェクト	0	0	0	0	0	0	0	0
35	ミリ	0	0	0	0	1	0	0	0
36	安定	0	0	0	0	0	0	1	0
37	以前	0	0	0	0	0	0	0	0
38	井戸	0	0	0	0	0	0	1	0
39	一部	0	0	0	0	0	0	0	0
40	汚染-汚濁系	0	0	0	0	0	0	0	0
41	開発	0	0	0	0	0	0	0	0
42	期待-希望系	0	0	0	0	1	0	0	0
43	恐れ	0	0	1	0	0	0	0	0
44	空間	0	0	0	0	0	0	0	0
45	掘削	0	0	0	0	0	0	0	0
46	形	0	0	0	0	0	0	0	1
47	経済	0	0	0	0	0	0	1	0
48	検討	0	0	0	0	0	0	1	0
49	公園	0	0	0	0	1	0	1	0
50	公開	0	0	1	0	0	0	0	0

(筆者作成)

附表4-3 (7/12) 副詞・形容や否定を表す単語と高出現頻度の共起 (1992年)

		9	10	11	12	13	14	15	16
		いなかった	うっすら	しない	すごい	たった	ちょうど	なく	よい
51	工業	0	0	0	0	0	0	1	0
52	工場	0	0	0	0	0	0	1	0
53	降雨-雨系	0	0	0	0	1	0	1	0
54	高校	0	0	1	0	0	0	0	0
55	高潮	0	0	0	0	0	0	1	0
56	今年	0	0	0	0	0	0	0	0
57	最高	0	0	0	0	1	0	0	0
58	最大	0	0	0	0	0	0	1	0
59	作業	0	0	0	0	0	0	0	0
60	昨年	0	0	0	0	0	0	0	0
61	市内	0	0	0	0	0	0	0	0
62	自然	1	0	0	1	0	1	0	0
63	実際	1	0	0	1	0	1	0	0
64	遮水壁	0	0	0	0	0	0	0	0
65	取水	0	0	0	0	1	0	0	0
66	手	0	0	0	0	0	0	0	1
67	首	0	0	0	0	0	0	0	1
68	出水	0	0	0	0	0	0	0	0
69	所	0	0	0	0	0	0	1	0
70	女性	0	0	0	0	0	0	0	1
71	上流域-上流系	0	0	0	0	1	0	0	0
72	条例	0	0	1	0	0	0	1	0
73	人々	1	0	0	1	0	1	0	1
74	水面	0	1	0	0	0	0	1	0
75	成長	1	0	0	1	0	1	1	0

(筆者作成)

附表4-3 (8/12) 副詞・形容や否定を表す単語と高出現頻度の共起 (1992年)

		9	10	11	12	13	14	15	16
		いなかった	うっすら	しない	すごい	たった	ちょうど	なく	よい
76	請求	0	0	1	0	0	0	0	0
77	専門家	0	0	0	0	0	0	0	0
78	戦後	0	0	0	0	0	0	1	0
79	大量	0	0	0	0	0	0	0	0
80	地価	0	0	0	0	0	0	1	0
81	地質	0	0	1	0	0	0	0	0
82	地盤	0	0	0	0	0	0	1	0
83	中心部	0	0	0	0	0	0	1	0
84	調べ	0	0	0	0	0	0	1	0
85	沈下	0	0	0	0	0	0	1	0
86	鉄道	0	0	0	0	0	0	0	0
87	土	0	0	0	0	0	0	0	0
88	土地	0	0	0	0	0	0	0	0
89	買収	0	0	0	0	0	0	0	0
90	発展	0	0	0	0	0	0	0	0
91	判断	0	0	1	0	0	0	0	0
92	復興	0	0	0	0	0	0	1	0
93	幅	0	0	0	0	0	0	0	1
94	噴出	0	0	0	0	0	0	0	0
95	本社	0	0	0	0	1	0	0	0
96	有害	0	0	0	0	0	0	0	0
97	用水	0	0	0	0	0	0	0	1
98	来年	0	1	0	0	0	0	0	0
99	連続	0	0	0	0	1	0	0	0
100	連絡線	0	0	0	0	0	0	0	0

(筆者作成)

附表4-3 (9/12) 副詞・形容や否定を表す単語と高出現頻度の共起 (1992年)

		17	18	19	20	21	22
		小さな	少し	全く	大きく	大きな	低く
1	自治体-都道府県市系	1	1	1	1	1	1
2	事業-工事系	0	1	1	1	1	0
3	メートル	1	1	1	1	1	1
4	地域-地域系	0	1	1	1	1	1
5	ダム	0	0	0	0	0	0
6	全国-日本系	0	1	1	1	1	0
7	河川-川系	0	1	0	1	1	0
8	中	1	1	0	1	1	0
9	調査	0	1	0	1	1	1
10	危険-非常系	0	1	0	1	1	0
11	使用	0	1	0	1	1	1
12	市民-住民系	1	0	0	0	0	0
13	水	0	0	1	0	0	0
14	設置	0	1	1	1	1	0
15	地下	0	1	1	1	1	0
16	地下水-地下水系	0	1	0	1	1	1
17	計画-予定系	0	0	1	0	0	0
18	結果-結論	0	1	0	1	1	1
19	現場	1	1	0	1	1	0
20	今	0	1	0	1	1	1
21	山	1	0	1	0	0	0
22	指摘	0	1	0	1	1	0
23	人	1	0	0	0	0	0
24	対策-対策系	0	0	0	0	0	1
25	中央-中心系	0	1	0	1	1	0

(筆者作成)

附表4-3 (10/12) 副詞・形容や否定を表す単語と高出現頻度の共起 (1992年)

		17	18	19	20	21	22
		小さな	少し	全く	大きく	大きな	低く
26	排水	0	0	1	0	0	1
27	反対	0	0	0	0	0	0
28	物質	0	1	1	1	1	1
29	利用	0	1	1	1	1	0
30	JR	0	1	0	1	1	0
31	キ口	0	1	1	1	1	0
32	グループ	0	0	1	0	0	0
33	コンクリート	0	1	0	1	1	0
34	プロジェクト	0	0	1	0	0	0
35	ミリ	0	0	0	0	0	1
36	安定	0	0	1	0	0	0
37	以前	0	1	0	1	1	1
38	井戸	0	0	0	0	0	1
39	一部	0	1	0	1	1	1
40	汚染-汚濁系	0	0	1	0	0	1
41	開発	0	1	1	1	1	0
42	期待-希望系	0	0	1	0	0	0
43	恐れ	0	0	0	0	0	1
44	空間	0	1	1	1	1	0
45	掘削	0	1	0	1	1	0
46	形	1	1	0	1	1	0
47	経済	0	1	0	1	1	0
48	検討	0	1	0	1	1	0
49	公園	0	0	0	0	0	0
50	公開	0	0	0	0	0	0

(筆者作成)

附表4-3 (11/12) 副詞・形容や否定を表す単語と高出現頻度の共起 (1992年)

		17	18	19	20	21	22
		小さな	少し	全く	大きく	大きな	低く
51	工業	0	1	0	1	1	0
52	工場	0	0	0	0	0	1
53	降雨-雨系	0	0	0	0	0	0
54	高校	0	0	0	0	0	0
55	高潮	0	1	0	1	1	0
56	今年	0	0	1	0	0	0
57	最高	0	0	0	0	0	1
58	最大	0	0	1	0	0	0
59	作業	0	0	0	0	0	1
60	昨年	0	0	0	0	0	1
61	市内	0	1	0	1	1	1
62	自然	0	0	0	0	0	0
63	実際	0	1	0	1	1	0
64	遮水壁	0	1	0	1	1	0
65	取水	0	1	0	1	1	0
66	手	1	0	0	0	0	0
67	首	1	0	0	0	0	0
68	出水	0	1	0	1	1	0
69	所	0	1	0	1	1	0
70	女性	1	0	0	0	0	0
71	上流域-上流系	0	0	0	0	0	0
72	条例	0	0	0	0	0	0
73	人々	1	0	0	0	0	0
74	水面	0	0	0	0	0	0
75	成長	0	0	0	0	0	0

(筆者作成)

附表4-3 (12/12) 副詞・形容や否定を表す単語と高出現頻度の共起 (1992年)

		17	18	19	20	21	22
		小さな	少し	全く	大きく	大きな	低く
76	請求	0	0	0	0	0	0
77	専門家	0	1	0	1	1	0
78	戦後	0	1	0	1	1	0
79	大量	0	1	0	1	1	0
80	地価	0	1	0	1	1	0
81	地質	0	1	0	1	1	0
82	地盤	0	1	0	1	1	0
83	中心部	0	1	0	1	1	0
84	調べ	0	0	0	0	0	0
85	沈下	0	1	0	1	1	0
86	鉄道	0	1	0	1	1	0
87	土	0	1	0	1	1	0
88	土地	0	1	0	1	1	0
89	買収	0	1	0	1	1	0
90	発展	0	1	1	1	1	0
91	判断	0	0	0	0	0	1
92	復興	0	1	0	1	1	0
93	幅	1	0	1	0	0	0
94	噴出	0	1	0	1	1	0
95	本社	0	0	1	0	0	0
96	有害	0	0	1	0	0	1
97	用水	1	1	0	1	1	0
98	来年	0	0	0	0	0	0
99	連続	0	1	0	1	1	0
100	連絡線	0	1	0	1	1	0

(筆者作成)

Study on Development of Social Needs Analysis by Text Mining -Development and Management of Water Resource Infrastructures in the Monsoon Asia Area-

Masahiro Sugiura¹, Fumio Komoda²
¹Japan Water Agency, ²Saitama University

This report will propose methods to support ‘awareness’, by which the practitioners involved in project planning, in the phase of project planning, can assess ‘truth encompassed in ambiguous expressions of needs’ from a pile of papers and documents to extract their experiences and practical sense fully. To help ‘awareness’ of individuals, syntax analysis of vast amount of documents and analysis of words appearance frequency will be of no use. It is necessary that implication of word network should be understood accurately. For this, maximizing ‘awareness’ of practitioners is essential.

Therefore, in this study, the simple analysis will be proposed with the below two points in mind to help ‘awareness’ of practitioners: 1)Text data should be extracted accurately from great volumes of that by text mining; 2) Analysis result should be visualized by quantification method 3 to support the detection of word network from the text data extracted.

KEYWORD: *text mining, quantification method 3, social needs, water resource infrastructures, project planning*

研究成果の要約

助成番号	助成研究名	研究者・所属
第 2008-6 号	テキストマイニングによる地域ニーズの分析手法の開発に関する研究 —モンsoonアジア地域の水資源インフラ整備—	杉浦政裕 独立行政法人水資源機構 総合技術センター
<p>【研究目的】 本研究の目的は、事業企画の実務担当者が事業企画段階において、大量の文書から「曖昧なニーズ表現の中に含まれている真実」を的確に把握し、事業計画の経験や現場感覚を十分に引き出すための「気づき」を支援する手法を提案することである。人間の「気づき」を支援するためには、大量の文書の構文解析や単語出現頻度の分析を行っただけでは役に立たない。単語のネットワークが作り出している意味を的確に把握することが求められる。そのためには、分析結果を可視化し、実務担当者の「気づき」を最大限に引き出すことが重要となる。</p> <p>そこで本研究では、人間の経験や感覚を最大限に活用し創造するための「気づき」を支援するために、膨大な文書からの的確な情報を取り出し、分析結果の可視化を図る手法を提案する。</p> <p>【研究手順】 水資源インフラ整備ニーズを深く分析するためには、フリー記述のアンケートを住民、政府の両者に対して実施することが望ましいが、ここでは社会の中に潜む水資源インフラ整備ニーズを探る分析手法の開発に焦点を合わせるため、客観性が高く収集しやすい新聞記事の分析を採択する。</p> <p>本研究において新聞記事を代用した理由は、①事実を客観的に伝えることにより発達してきたマスメディアの中でも特に影響力を持つ、②マスメディアを通じて膨大な情報が流通しており、マスメディアの動向を知ることが社会ニーズを考えるうえで重要となると判断したからである。新聞は、インドネシア国内で発行されている現地新聞『じゃかるた新聞』を選定する。分析対象は、水資源インフラ関連記事（じゃかるた新聞社のオンラインデータベースの制約上1年分）とする。</p> <p>しかし、単年分の分析では蓄積された膨大な情報を生かし、ニーズの全体像と時系列的</p> <p>な変化に対して十分なアプローチは困難である。そこで、日本の主要な新聞である『毎日新聞』における水資源インフラ関連記事（1992年、1996年、2000年）を時系列分析することにより、手法の妥当性を確認する。</p> <p>そして、この手法に基づき『じゃかるた新聞』1年分の記事によるニーズ分析の結果とその限界について考察する。</p> <p>【研究成果】 提案した分析手法は、①的確に情報を取り出して社会状況を大まかに可視化し、②人間の経験や感覚を最大限に活用するための「気づき」を支援できることが明らかになった。</p> <p>その一方でこの手法は、単語の出現頻度に着目し、多変量解析による分析をしているため、次のような限界も明らかになった。①出現頻度が低くて目立たない単語から将来の社会の変化の予兆を発見することができない、②単語をネットワークとして捉えなければ将来の予兆など重要な知識は得られない。</p> <p>【今後の発展】 多変量解析によるテキストマイニングの限界を克服するためには、例えばグラフ理論、ニューラルネットワーク理論、潜在的意味解析などの適用が必要となる。これらの取り組みにおいても、発見が難しい低出現頻度の重要単語の発見に近づくためには、分析目的に応じた情報源の加工作業である「辞書」を整備することは重要になる。分析対象に応じた辞書を整備することは、不要なノイズを切り捨て、重要な単語を抽出し、分析目的に合致した意味や評価を行うためには不可欠である。そのため、分析手法の開発と併せて辞書整備への取り組みが、益々重要となる。</p> <p>今後、文字情報データベースや辞書の構築を図り、テキストマイニングによる単語ネットワーク分析の研究を進めることにより、様々な切り口から日本国内外の河川や道路に関する建設・管理事業に対するニーズやシーズの分析支援を期待できる。</p>		