

日本 MOT 学会による査読論文 (2018-1)

IT 業界におけるコンセプト用語の進化プロセス —クラウドコンピューティングの事例研究—

An evolution process of IT concept words
— Case study of cloud computing —

片岡利枝子、井川康夫、内平直志
Rieko Kataoka, Yasuo Ikawa, Naoshi Uchihira

要 旨

本論文では、IT 業界のコンセプト用語について、IT 雑誌の記事をテキストマイニング手法を用いて分析することによって、コンセプトが推移する段階を進化プロセスとして表出した。具体的にはクラウドコンピューティングの事例について、クラウドに至るまでのコンセプト用語の推移には、「集合」「補完」「特化」の3つの主要な形態が存在することを新たに発見した。

ABSTRACT

New concept words frequently appear and affect the business environment in IT industry. The core technology and the architecture are, in many cases, the same even if the words change in fashion. This paper analyzes the articles of popular IT magazine by using the text mining method, and proposes the concept evolution process of three phases, the gathering phase, the complementation phase and the specialization phase, focusing on the case of cloud computing.

キーワード：コンセプト進化，クラウドコンピューティング，コンセプト推移，テキストマイニング

投稿区分：研究論文

第1章 はじめに

1.1 背景と目的

IT (Information Technology) 業界では数年ごとに新しいコンセプト用語が登場しビジネスモデルが変化する。しかし、その基本技術や IT アーキテクチャ¹ は同じであることが多い。IT 関連の雑誌や新聞の見出しにも技術用語やサービス用語が時代と共に登場し、いつのまにか消えて行くことが繰り返される。そしてそれらは専門的な技術の関連性よりも企業のビジネス事由が優先され、販売のために目を引く「看板」とし

て用いられるケースも多い。その時々看板はどのような理由で選ばれ、架け替えられていくのだろうか。

例えば、IT 業界では展示会が盛んであり、毎年多くの総合イベントが開催される。中でも業界最大のイベントである「ITpro Expo」は、毎年、看板となるテーマが決められ、展示会やセミナーだけではなく、Web サイトや雑誌といった他メディアとも連携したクロスメディアイベントとなっている。「ITpro Expo 2008」でクラウドコンピューティングという言葉が注目され始め、2009 年から 2010 年にはテーマとして採用されている。

片岡 利枝子 (国) 北陸先端科学技術大学院大学
内平 直志 (国) 北陸先端科学技術大学院大学
受領日：2017 年 7 月 10 日、受理日：2018 年 7 月 20 日

井川 康夫 (国) 北陸先端科学技術大学院大学

こうしてIT業界のみならず一般にも浸透した「クラウド」であるが、その技術的本質は必ずしも新規のものではない。クラウドとはクラウドコンピューティングの略称であり、共用の構成可能なコンピューティングリソースに、どこからでも簡便に必要なに応じてネットワーク経由でアクセスすることを可能とし、ユーザに情報サービスやアプリケーションサービスを提供するという、コンピュータの構成・利用に関するコンセプトのことである。しかし、クラウドコンピューティングの出発点はグリッドコンピューティング技術であると言われており [17]、その後、SaaS (Software as a Service) など、さまざまな用語を経て今日のクラウドに至っている [25]。クラウドというキーワードが、それまでのグリッドコンピューティングや SaaS というキーワードと何が異なり、何が継承されて発展してきたのだろうか。クラウドに至るまでの過程を分析することによって、IT業界のキーワードの発展のパターンを見いだすことができるのではないかと考えられる。

本研究は、IT業界でトレンドとなったコンセプト用語をテキストマイニング手法を用いて分析し、その推移の過程を進化として捉えたものである。「コンセプト用語の進化」とは、「同じ対象に対しての捉え方(視点)が、時間の経過とともに何らかの理由で変化したことに伴う対象に対する用語の推移」と定義する。本論文では、クラウドコンピューティングの事例に焦点をあて、クラウドという用語に至るまでのコンセプト用語の進化の過程を分析し、IT技術用語がどのような変遷を経て一般大衆に広まる用語にまで発展していったのかを説明するメカニズムを明らかにすることを目的とする。さらに、テキストマイニングを用いてコンセプト進化を検証することによって、従来の定量的手法に比べて膨大なデータ量を扱うことが可能となるため、コンセプト推移についてのより高精度で有効な定量的分析手法を提示することを目的とする。

本研究の特徴は、技術的な成長過程を追うものではなく、進化パターンを抽出するための手がかりとなるコンセプトにつけられた名称(ラベル)である「用語の進化」を追うことによって、IT業界のコンセプトの変化を客観的に捉える点にある。

1.2 本論文で使用する用語の定義

最初に、本論文における主要な用語を次のように定義する。

「コンセプト」：物事の共通事項を包括し、思考活動の基盤となる基本的な形態として捉えたもの。本論文では特に、IT業界における技術やサービスを対象とする。

「キーワード」：コンセプトの中から主要なもの、特に影響力の大きいものを指す。

「コンセプトの進化」：同じ対象に対しての捉え方(視点)が、時間の経過とともに何らかの理由で変化することに伴う対象に対するコンセプト用語の推移のこと。

本研究では、対象に対する捉え方の特長に応じて、コンセプトの進化を集合、補完、特化3段階に分類した。この分類については第3章に詳しく記述する。

1.3 研究の方法

本研究は、グリッドコンピューティングの技術を基礎としてクラウドコンピューティングに至るまでの長期にわたるコンセプト用語の推移を対象とするものであり、本論文はクラウドの事例を取り上げている。後述するようにユニークさ、および、クラウドという広く一般的となった用語であることから典型性も有しており、これらを事例研究採用の正当性の論拠 [15] として事例研究設計を行った。

事例研究の手順は Eisenhardt に従い、事例の選定、プロトコルの生成、データ収集、データ解析、仮説モデルの生成とした [3]。プロトコルの作成等については第3章に詳しく記すが、本研究のためのデータ解析にはテキストマイニングを採用し、テキストマイニングの対象データとしては、日経BP社発行のIT雑誌「日経コンピュータ」の記事を採用した。日経コンピュータは、企業情報システムやネットワークに携わるITプロフェッショナルはもちろん、ITを経営に活用するための判断に迫られる経営者まで、情報技術にかかわるプロフェッショナルに向けた総合情報誌であり、ニュースだけでなく詳細な解説やコラム、ネットの双方向性を活用したコンテンツを提供しているものである [24]。従って、IT業界のトレンドを広く探るといふ本研究の目的に合致した内容であると考えられる。

インターネットの普及により、我々が日々扱う文書データは通常の知識情報に加え、論文や特許のような技術情報、メールやツイッターなど個人が発信する感情や意見など、多種多様な形態で増え続けている。その結果、せつかくの文書データが十分に活用できていない懸念が生じている。しかしながら、菰田・那須川によると、膨大な文書データのすべてに目を通すことは不可能であるが、テキストマイニングを使いこなすことにより、その中から有効な情報を見だし活用することが可能になる。但し、テキストマイニングは通常のアプリケーションソフトウェアと異なり、単に文書データを入力すれば有効な結果が得られるわけではない。どのように分析し、その出力特徴をどう読みと

るかによって結果が大きく変わってくる [19]。

これらの特徴を踏まえ本研究では、膨大な文書データからテキストマイニングによって、“話題の経時的な変化”という、時間軸を含んだ知見を有効に抽出する手法を提示することを目的とする。単に文書データにおける話題の時間的推移を取得するのみならず、話題間の関係性も取得する手法が本研究のテーマである。通常、テキストマイニングのアプリケーションプログラムは時間軸での検出が可能であるが、有効な結論を導くためには目的に合わせた活用方法が求められる。本論文ではまず、IT 業界の文書データからコンセプトの推移を取得する手法について事例を用いて説明する。これによって得られた分析結果をもとに、コンセプトの推移についての発見事項をまとめる。最後に IT 業界でさまざまな活躍する複数の有識者へのインタビューを実施し、結果について多面的に考察する。

第 2 章 先行研究レビュー

コンセプト用語の推移に関しては、「コンセプトが変化する事由」についての文献と、「クラウドに関するコンセプトの変化」についての文献を中心にレビューした。加えて、「テキストマイニングを用いた定量的分析について」レビューする項目を設け、テキストマイニングを駆使して文書データやインターネット記事を解析する手法についてレビューする。また、それらの中から特に「コンセプト推移の分析手法」を論じたものについてレビューを行った。

2.1 コンセプトが変化する事由

ビジネスシーンでは“新たな価値創造”を起こすために、内容はほとんど同じである場合でも、あえて看板となるコンセプト用語を変えることがある。一企業が戦略的にビジョンとして打ち出した用語が一般的になるケースや、冒頭で述べたようにビジネスショーのテーマとして掲げられたものが業界全体のトレンドとして広まっていくケースなどがある。いずれにしても、ビジネス事由によって新たなコンセプトが登場する。このため、経営コンサルタントがコンセプトのラベル付けを行うことについての研究、および価値創造をもたらすためのテクノロジー・インテリジェンスに関する研究について調査した。これらの先行研究は以下のように整理することができる。

Giroux は、コンセプトのラベル付け（すなわちネーミング）を行う場合は、単なる興味からではなく、社会の変化や革新的な要素が重要視されるべきであると述べている [6]。Alvesson らは知識をタイプ別に分類することによって、単なるアイデアから流行（ファッション）を意識したコンセプト用語に進化させる手法を提示している [1]。Porter は、企業は顧客に長時

間持続可能な価値をもたらすことができたときのみ競合相手を押さえることができる、と述べる [12]。これを受けて菅澤は、顧客が価値を見いだすためには競争相手とは異なる価値が提供されている必要があるとして、テクノロジー・インテリジェンスを提唱する [20]。La らは、競争優位を確保するためにはライバル企業、すなわち競争相手といかに差別化するかが重要であり、その戦略のひとつとして新しいコンセプトを使うことがあるとし、企業の思惑どおりにコンセプトが定着するかどうかは世の中の受け手側に委ねられる、と述べている [10]。

要するにビジネスの世界ではライバル企業といかに差別化するかが重要である。そのための戦略として新たなコンセプト用語を導入することによって、競合他社に対して違いを際立たせ自社のポジションを確立しようとするところがあると論じている。しかしながらこれらの先行研究は、いずれもコンセプト用語が具体的にどのようにして推移していくのかについて数年にわたって連続して注目したものではない。

2.2 クラウドに関するコンセプトの変化

技術の発展にともなうコンセプト用語の違いを取り上げた先行研究は広く存在するが、本論ではクラウドに関連する部分に焦点を当てて調査を行った。クラウド技術が専門である Villegas によると、現在もっとも注目されるクラウドコンピューティングの最初の出発点はグリッドコンピューティングであるという [17]。同じく Foster らは、クラウドは決して新規のコンセプトではなく、グリッドコンピューティングの理論や従来技術との深い関連によるものであるとし、クラウドとグリッドの両者を多方面から比較した。そして、ビジネスモデルやセキュリティなどにおいて異なる面も多いが、ビジョンやアーキテクチャ、根本技術は共通であると結論付けている [4]。また、Mell らによると、米国 NIST² のクラウドコンピューティングの定義の中では、クラウドのサービスモデルの定義のひとつに SaaS が含まれており、クラウドコンピューティングは従来から存在する SaaS などを言い換えたもの、あるいはさらに発展させたもの、と説明されている [11]。

しかし一方で、日経 BP 社のクラウドの集大成ではクラウドコンピューティングに厳密な定義を与えていない [25]。IT 企業で長年、クラウドの上級エンジニア職にあった三崎が指摘するように、IT 業界においても「クラウドは従来の IT 活用と何が違うのか、技術的には同じではないか」という声もあがっており、クラウドの定義は明確にされていない [27]。

すなわち、これまでのクラウドに関する先行研究はコンセプトの定義付けやコンセプト間の比較に留まっている。

2.3 用語の推移の分析手法について

テキスト中に現れる用語の推移をテーマにした研究がある。Arinoらは約10年間にわたりWorld Wide Web (www)学会の研究発表のタイトルを分析し、それらの推移とタイトル間の関係性をネットワーク化して表した。これにより、ソーシャルネットワークとeコマースの急速な発展トレンドを認識することが可能となり、その手法がR&Dの意思決定戦略に役立つことを示唆している [2]。

また、藤垣・永田による、科学技術政策の答申文書を36年間にわたってデータベース化し分析した例がある。これは、科学技術政策コンセプトの推移を定量的に把握したものであり、定量分析が政策研究の方法論として有効であることを示唆している [5][26]。しかしこの例では、コンセプト推移の形態までは見いだしてはいない。また、文章から用語を抽出するという方法論は同じであるが、テキストマイニング手法は用いていない。このため、扱うデータ量に限界があること、および、用語の登場頻度が分析の主流となっている。一方、本研究ではテキストマイニング手法を用いることによって、ビッグデータと呼べる膨大なデータ量を分析対象とすることを可能とした。その結果、用語の推移について、頻度分析だけでは読み取ることが難しかった用語間の相関を高精度で抽出することが可能となり、定量的分析の新たな手法を提示している。しかしながら、同様にテキストマイニング手法を用いた先行研究もあるので、それについては次節で詳しく述べる。

Tsumotoらは、新しいコンセプトは毎年のように現れるが、それが残るための条件は世の中に発表された文書の中にそのコンセプトがどれだけ採用されたかに依存すると述べ、実際にバイオ医療関係の研究論文から専門用語を抽出し構造化したものに独自の方程式を当てはめて、専門用語間の関係を分析している [14]。竹岡らは、デジタルカメラに関する技術用語の関連ネットワークを特定時点におけるスナップショットで捕らえ、それを2年ごとに繰り返すことによって、デジタルカメラという商品が属するまわりの環境設定のカテゴリーが変化する過程を分析している [22]。他にも同様に、技術用語について取り上げた論文は存在するが、単純にある特定の用語間の関係性を述べたものが一般的である。

Tsengらは、特許の出願文書が多くの重要な技術研究の成果を含んでいることに着目し、出願書類の記載事項をテキストマイニングによってマッピングする手法を示した。出願書類の記載は構造化されたものであるから、それらの抽出ルールを設定し自動的にデータ収集することによって分析を容易に実現することを目的とした。特許のパターンや、特許同士の関連や

トレンドを押さえることによって、特許の重要度や影響度を把握できる [13]。この他Yoonらのように、特許の出願文書をテキストマイニングすることによって特許のネットワーク化を行う手法を示したものもある [16]。これらは、それぞれが用いているテキストマイニングのアプリケーションソフトウェアを、いかにして構造化された特許書類の記載内容に当てはめ有効な情報を取り出すかという、メソドロジー（方法論）の研究である。

2.4 テキストマイニングを用いた定量的分析について

次に、テキストマイニング手法を用いて各種の文書データやインターネット上の記事を分析している先行研究を調査して、本研究でのテキストマイニングの使い方の参考にすると共に、本研究で示す手法との違いを述べる。

Guptaらは、インターネットの世界では文書データはますます膨大なものになるからテキストマイニングによる自動分析が重要であると主張し、テキストマイニングによって文書データを分析する段取りをわかりやすく解説している。そしてテキストマイニング分析がもたらす種々の効用を取り上げているが、その中のひとつに本研究のテーマである“Concept Linkage”が存在する。Guptaらはテキストマイニングをコンセプト間の関連性を分析することに用いることによって、例えば、人間が気が付かないような、病気と処置方法との関連性を表出することができるとし、大量の文書データから異なるトピック間の関連を調べることの可能性を示唆している [7]。

森脇は、話題の変化を文書データで語られている単語の経時的な変化ととらえ、単語の出現する頻度の増減から消費者トレンドの変化を検知する試みを行っている。さまざまな単語の出現件数を時間軸と共に記述し、出現頻度の変化の大きい単語に着目すれば話題が変化したことを察知できるとしている [28]。本文献ではトレンドが動いたと判断するための単語出現の増減を検知する「感度基準」が議論されているが、トレンドを代表する単語間の関係性までは問われていない。

白井らは、文書データからトレンド情報を抽出するためには、重要なキーワードがあらかじめ抽出されることが必要であるとし、キーワードの抽出方法を目的に応じて選択することによってマイニング環境の整備を行っている。テキストの中から名詞と認定された候補から専門用語を抽出する形態素解析による抽出手法、および、特定の言い回しが頻繁に用いられるよう

なテキストや、特定のフォーマットが定義されたテキストを対象にする場合に有効な、パターンマッチングによる抽出がある [21]。また、奥和田らは、キーワード抽出において、前出の形態素解析を用いて実施し、さらにそれらを分野別特徴によって分類することによって、自由記述の文書データから自動的に効率よく目的とする結果を取得する手法を示している [18]。

これらの手法を参考に、本研究でもキーワードの抽出方法を目的に応じて選択することによってマイニング環境の整備を行った。これについては第3章で詳しく述べるが、専門用語の抽出方法については独自の手法を採用している。

2.5 本研究の意義

これらの先行研究では以下の点が指摘されてきたと言える。

- コンセプトのラベル付け（すなわちネーミング）を行う際には、社会の変化や流行（ファッション）を意識して行われる。
- 競争優位を確保するために、競争相手とは異なる価値が提供されている必要があるから、差別化戦略のひとつとして新しいコンセプトを使うことがある。
- IT業界のコンセプトの推移の特徴は、ビジネスモデルの視点で変化する面が多いが、それらのビジョンやアーキテクチャ、根本技術は共通である場合が多い。
- 用語の推移と関係性をネットワーク化して表すことは、発展トレンドを認識することが可能となり、その定量的な分析手法は R&D の意思決定戦略に役立つ。
- コンセプト用語の推移を分析した事例もあるが、コンセプト用語間の関係を示したもの、あるいは時系列に列挙したものに留まっている。

このように先行研究では、コンセプトの推移を時間的に連続したものと捉えた上でどのように発展して「クラウド」に至ったかという、推移の形態を明らかにしたものはない。この点を明らかにしようとするところに本研究の意義がある。また、先行研究において、用語の推移を登場頻度から定量的に分析を行った例は存在するが、本研究ではテキストマイニングを用いることで膨大な量のデータを分析対象とすることを可能にし、これによって用語の推移について用語間の相関を多面的、かつ高精度に抽出し分析するという定量的手法を提示したことにもうひとつの意義がある。

第3章 クラウドの事例分析

3.1 分析の手法

分析対象となる母集団には、日経 BP 社の IT 情報誌「日経コンピュータ」[24] の記事を取り上げた。月2回の発行であるが、2002年以降のすべての記事が電子化されており、長年にわたりほぼ同じ目次構成であるため定点観測が可能である。本研究では2002年1月より、クラウドという用語が十分に定着した2012年12月までを分析の対象とした。収集された記事の数は全部で2万2千本であり、これはテキストマイニングで傾向を分析するのに十分大きい数である。通常、ニュース記事の構成は、冒頭にリード文と呼ばれる記事の概要を掴み易くするための本文のサマリー（要約）がある。本研究では、記事のアブストラクト部分、もしくは冒頭の500文字までのサマリー（要約）部分を対象データとした。もしも記事全体を含めるとすると、他のトピックや参考文献情報などがノイズとして入り込み、かえって混乱を招くため十分と判断する。

IT業界の場合、キーワードのほとんどは米国からの輸入であるが、インターネットの発達により日本の雑誌の分析でもタイムラグも無く世界の動向を見ることが出来る。

テキストマイニングのソフトウェア・アプリケーションは、IBM Content Analytics Version 3.0（以下、ICA）を採用した。ただし近年、新たにデータサイエンティストという職業が登場したように、テキストマイニングによって有効な結果を導くためには、その手法を使いこなす人間の技量も重要な要素である。筆者らはこの点を解決すべく、ICAの開発者である那須川の提唱する「那須川メソッド」(図1)に従って分析を進めた。那須川メソッドとは、那須川が多くの分析プロジェクトの経験から得た知見を基に、テキストマイニングによって何らかの有効な結果を得るための技法をまとめたものである [23]。

ICAのアプリケーションの分析結果の表示形式を図2に示す。これらは最終結果を示すためのものであるが、ステップ1、ステップ2で特徴を把握するためにも使うものである。個々の形式については IBM Content Analytics with Enterprise Search [9] に詳しく記載されているが、本研究ではステップ1、ステップ2を進める際に、主に文書分析、ファセット分析、偏差分析、コネクション分析を用いて観察することによって記事の特徴をつかめることが分かった。

3.2 関連の定義

ICAのRedBooks³によると、相関は2つの項目間について次のように定義される。

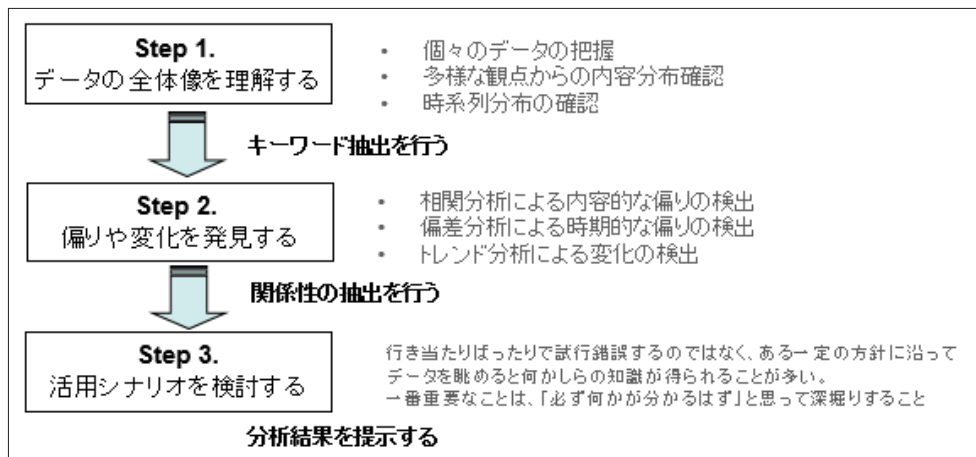


図1 「那須川メソッド」フローチャート（那須川の文献 [23] を基に筆者らが作成）

対話的テキスト・マイニングのための、ユニークな分析ビュー

文書一覧
現在の検索条件を満たす文書を一覧して現象を把握

ファセット分析
現在の文書セットと関連の高いファセット値を分析

時系列分析
現在の文書セットの頻度を時系列に並べて分析

コネクション分析
現在の文書セットでファセット値間の繋がりを分析

トレンド分析
現在の文書セットで急増しているファセット値を分析

ダッシュボード
複数のファセットの頻度・相関を一度に確認

図2 ICA3.0 表示形式（文献 [9]）

2つの文書集合AとBがある場合、Dは全文書集合、#は文書集合中の文書数とすると、相関値は以下の計算式で表される。ここで、左辺も右辺も値は同じである。

例えば、2つの文書集合A, Bが

$$\frac{\#(A \cap B) / \#A}{\#B / \#D} = \frac{\#(A \cap B) / \#D}{(\#A / \#D)(\#B / \#D)}$$

A={「商品」カテゴリのキーワード「パソコン」に該当する文書}

B={「名詞 ... 要望」カテゴリのキーワード「マニュアル...入手する...たい」に該当する文書}

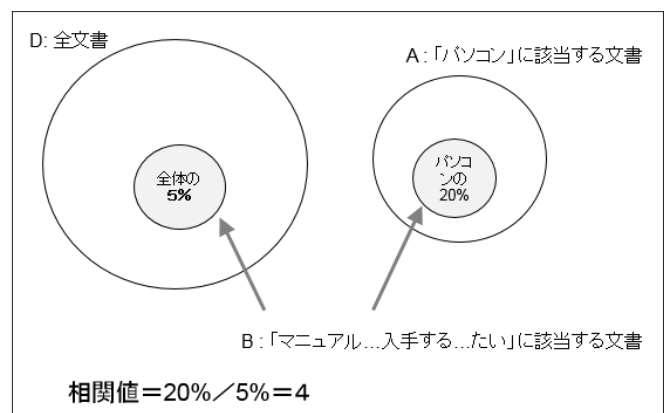


図3 ICA3.0 相関値のしくみ

であったとすると、上記の式の左辺は、

$$\frac{\text{パソコンに関する文書に限った時のマニュアル入手要望の割合}}{\text{全文書中のマニュアル入手の要望の割合}}$$

に相当する。この集合を図示すると図3のようになる。例として、マニュアル入手に関する文書は全文書中の5%、一方で、パソコンに関する文書に限るとマニュアル入手に関する文書が20%の場合、「パソコン」と「マニュアル…入手する…たい」は相関値が4であり、関連が強いと言える。

また、上記の式の右辺は、AとBの密度の掛け算 $(\#A / \#D) (\#B / \#D)$ と実際の $(A \cap B)$ の密度 $\#(A \cap B) / \#D$ の比で、「AとBの独立性からの逸脱」を表しているから、右辺の式の方が直感的と言える。

単純に言うと上記のような論理であるが、実際のICA3.0では信頼性の無い値は小さめに補正するなど、区間推定法⁴を用いた補正をおこなっている[8]。

3.3 分析の手順と結果

3.3.1 ステップ1

ここでは「データの全体像を把握」し、辞書を作成することを目的とする。テキストマイニングを行うためには、何について集計するかを抽出するための「辞書」と呼ばれる分析用頻出単語リストを作成する必要がある。分析用頻出単語は対象とする母集団テキストから自動抽出する方法もあるが、ステップ1で個々のデータを一通り把握した結果、頻繁に登場する「部長」「会議」…などの一般的なビジネス用語もキーワードとして集計されてしまうことがわかった。このため本事例研究では、手動でキーワードを辞書登録することによって分析用頻出単語リストの設定を行うこととした。

本研究では毎号必ず設けられている用語解説のため

の1ページの記事に着目した。ページタイトルは「今さら聞けないITキーワード」「知っておきたいキーワード」など、時代によって変遷はあるが、その時々で押さえておくべきキーワードを取り上げていることには変わらない。日経コンピュータは情報システムとネットワークの総合誌であるから、そこで取り上げられた約500語の「キーワード」に、コンセプト用語とされるものは一通り含まれていると考える。なぜなら、このページのキーワードの選定基準として、(1)技術用語としてきちんと説明しておく必要があるもの、(2)技術トレンドとして押さえておくべきもの、(3)いま、世の中で大事な言葉として提案するもの、であると定義されており、本研究の分析を行うためには十分であると判断する。この約500語について、同類の用語(例えば、クラウドとCloud)や複数回採用されたものを名寄せし、最終的に残った約350語をキーワードとして抽出した。同類の用語は、辞書登録機能の中で、代表されるキーワードの同義語として登録することによって、それらも有効なデータとして分析対象とする。

そうして完成した辞書をもとに、さらにさまざまな表示形式を確認しながらクラウド事例についての全体像を把握した。文章分析(図4)では全部で2万2千本の記事のなかから、辞書登録を行ったキーワードを含む記事を表出することができる。ひとつの記事の中に同じキーワードが複数回登場するものも多くあるが、この場合のカウント数は1件となる。

約350のキーワードすべてについて、同様の出力結果を画面で確認することができる。

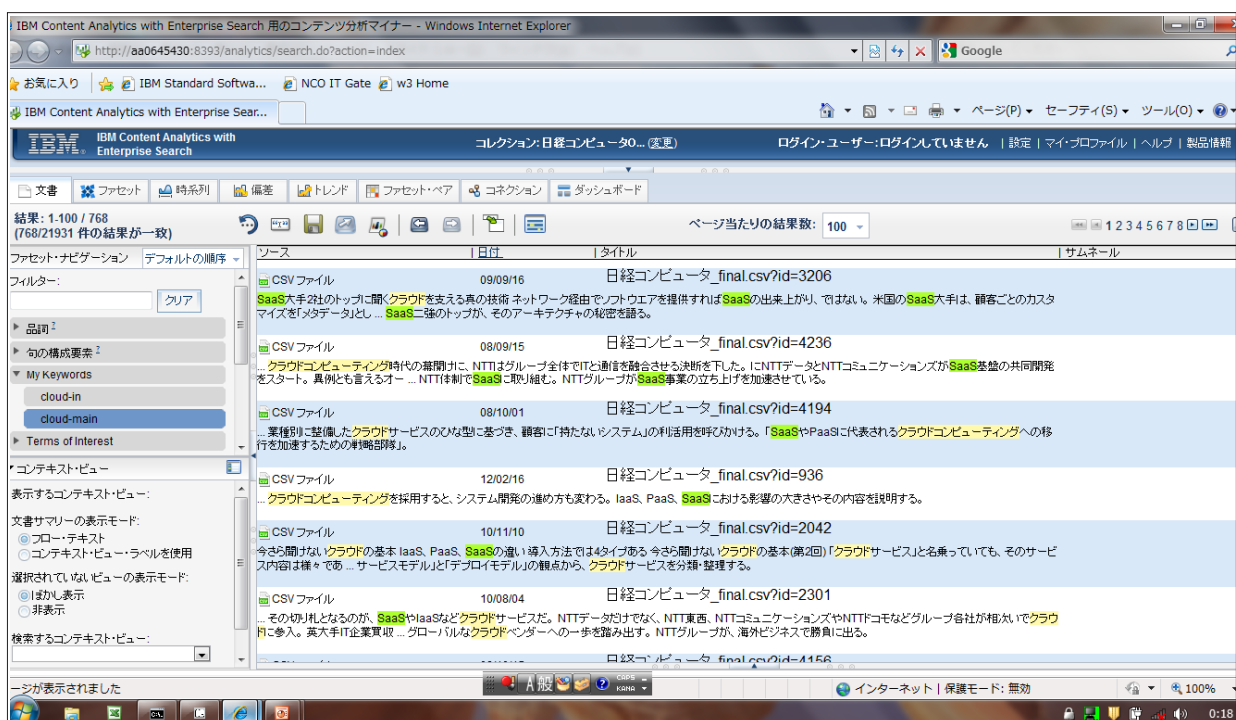


図4 文書分析の例(クラウドとSaaSが同時に含まれた記事)

3.3.2 ステップ2

次に「偏りや変化を発見する」ことを目的に、「ファセット分析」、「偏差値分析」にてグリッドコンピューティング、SaaS、クラウドコンピューティングなどの用語に関連した相関を確認した。

ファセット分析では、出現件数により分析する「絶対頻度」に加え、文字列同士の関連性の強さで分析する「相対頻度分析」を提供する。相対頻度分析では、関連性の強さを数値化した「相関値」を用いている。相関値によって、“ある条件下で出現するキーワ

ードは、他のキーワードに比べて特徴的である”といったことが分かり、分析者が気づいていない関連性が発見できる。例えば 図5 は、SaaS というコンセプトについての各キーワードの相関の強さを表している。SaaS というコンセプト用語は、クラウドコンピューティングと共に登場する頻度が高いが、相関という観点では実は、PaaS (Platform as a Service)、IaaS (Infrastructure as a Service) の方と強い相関を持っていることがわかる。

また、偏差分析でキーワード推移についての時系列分布を図6のように確認した。これによって、ファセ

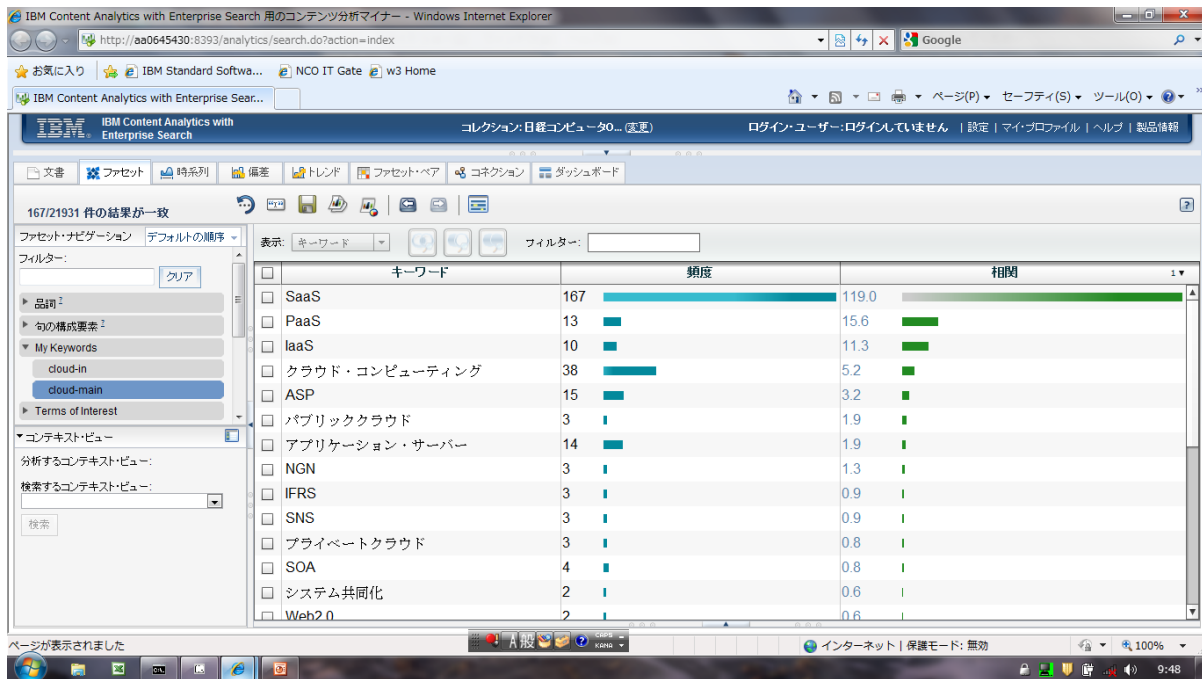


図5 ファセット分析の例 (SaaS)

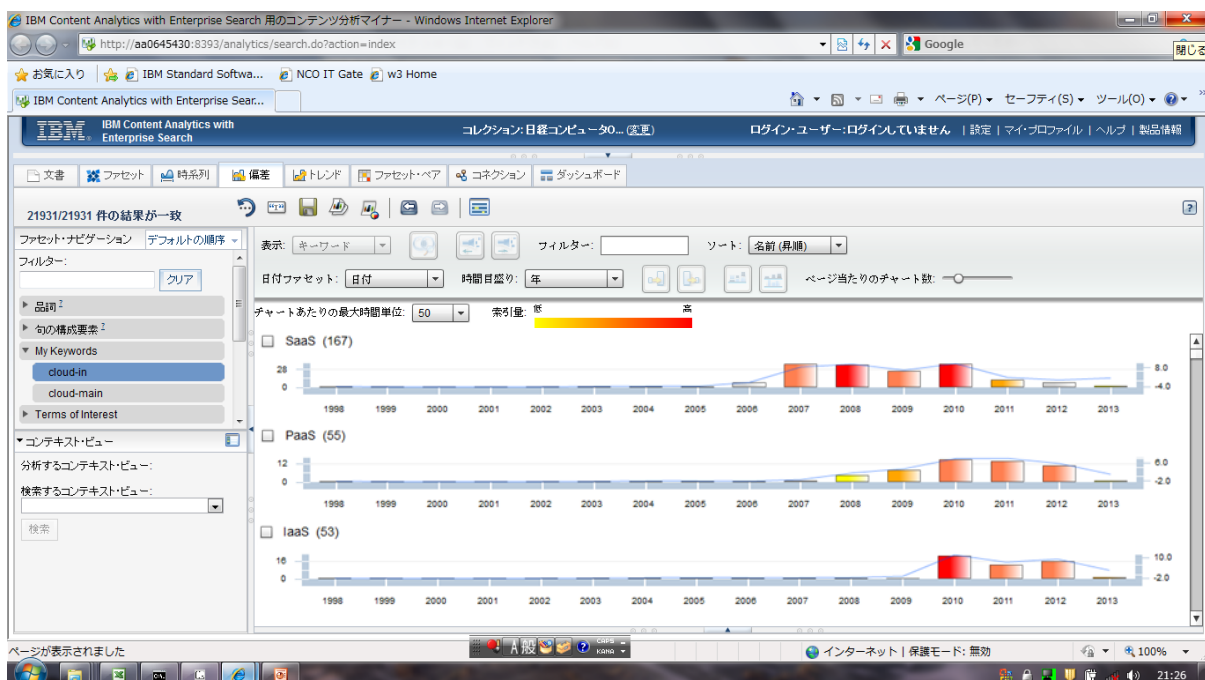


図6 偏差分析の例 (SaaS)

ット分析で SaaS と相関が強いとされた PaaS、IaaS のそれぞれのキーワードが登場した年代 (横軸目盛) や頻度 (縦軸目盛) を読み取ることができる。したがって、ステップ 1 で抽出されたキーワードは一時期に同時に現れたのではなく数年の時間を要しており、その間に時系列に関係構築されながら、ひとつひとつが出現していったと理解される。

ステップ 2 の結論として、ステップ 2 で明らかになった相関を基準にしてキーワード間を結ぶことによって、コンセプトの時間的な変遷が分かる図式として表出させることを考案した。そこで最終的な分析結果には、コネクション分析をベースに用いることとした。コネクション分析⁵は抽出した分析用頻出単語間の関係性をわかりやすく示すものであり、抽出したキーワードの出現頻度や相関を示すファセット分析によって得られる相関値をもとに自動生成される。ここにさらに、偏差分析で得られた時系列の要素を付加するのである。

偏差分析からは、そのコンセプト用語が急激に普及し始めたタイミングを読み取ることができる。コネクション分析の出力画面の横軸を時間軸として利用し、それぞれのコンセプト用語が普及し始めた年度に移動調整することによって、ステップ 3 で述べるような本研究が求める出力結果が見えてくる。

3.3.3 ステップ 3

最後に、ステップ 1 と 2 で把握したキーワードの関係性から、最終的な出力イメージとしてキーワード間の相関の強さを時系列でマップしたものを採用し、表出された結果を基に、進化プロセスの確認に至るまでの「活用シナリオ検討」を行った。

自動生成された相関関係を示す出力結果を時間軸でプロットすることにより、キーワード間の相関の強さを図 7 のようにマップすることができる。楕円の大きさはそのキーワードが登場した頻度を表す。相関が強いキーワードを太線の矢印のように時系列に追っていくことにより、時代と共に注目されたキーワードの推移を見ることができるのである。下端の目盛りは、年代の目安であるが上下方向の位置関係には特に意味はない。

この手法を実際に辞書登録したキーワードに適用したものの図 8 である。コンセプト用語は時系列に従って左から右にプロットされる (横軸の年度は目安を示し

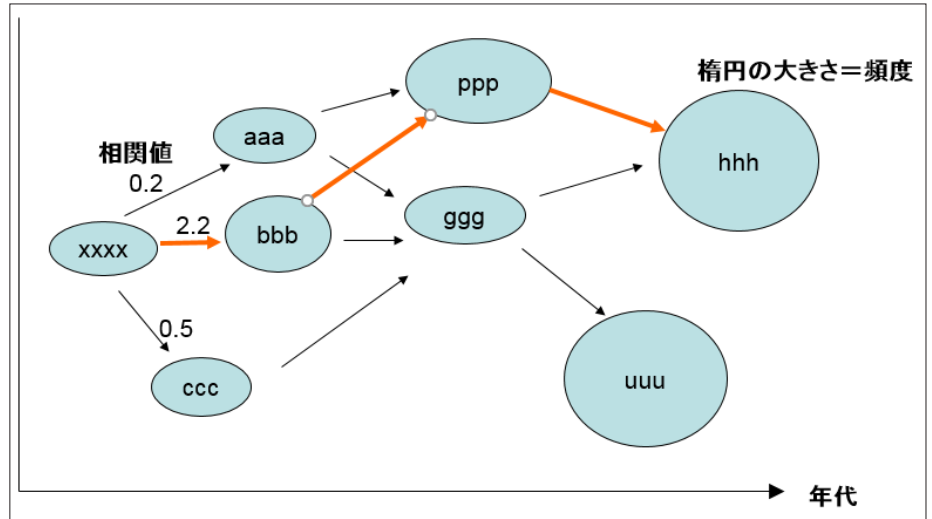


図 7 出力イメージ (コネクション分析の横軸に時系列を採用したもの)

たもの)。図上にキーワードが 350 語もプロットされていないのは、不要と判断される (1) 極端に登場頻度の少ないキーワード (対象期間を通じて 5 回以下のもの)、(2) 他のキーワードと相関が無いキーワード (他のキーワードとつながらないキーワード)、例えば「2007 年問題」など年度特有のもの、が表示されていないことによる。加えて、(3) 長年にわたり幾つものキーワードと強い相関を持つキーワード、具体的には「仮想化」は、2006 年ごろに現れ 2012 年になっても多くの技術やサービスの基本となるものであるから多くのキーワードとつながってしまい、コネクション分析で表示させた場合、全体の推移の様子を見えづらくする要因となる。このため、あらかじめステップ 2 でデータの傾向を確認した結果、偏差が確認されないものも非表示とした。

さらにクラウドの事例研究に相関がある部分のみを残して簡素化したものが図 9 である。グリッドコンピューティングから SaaS、クラウドコンピューティングまでの推移を中心に、それに付随するキーワードの発生と共に抽出したものを時系列に見ることができる。この流れを追っていくと、グリッドの時代にはハードウェアやソフトウェアの技術キーワードや製品が多く登場しているが、2007 年に SaaS が登場すると、そのあとは PaaS、IaaS が現れ、それらがクラウドコンピューティングにつながりさらに発展したあとで、今度はエンタープライズクラウド、プライベートクラウド、パブリッククラウドという、技術を直接示さないコンセプト用語が強い相関を持ったキーワードとして現れている。クラウドへの推移と並行して、下方にはソーシャルネットワークに関するコンセプトの推移が存在しているのが分かる。これらの一連の推移をわかりやすくするために、図上に矢印を加筆した。

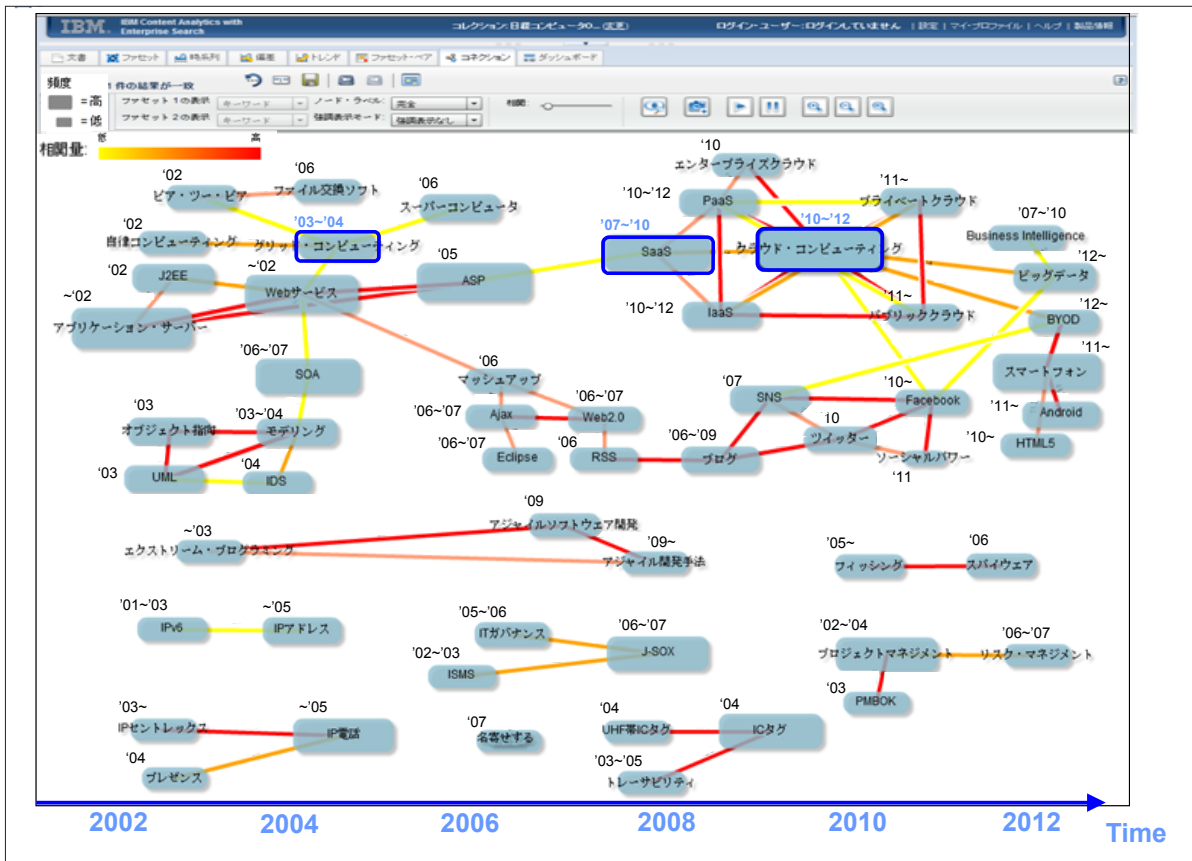


図8 クラウド事例分析
 相関値色別 黄色：2.2～4.0，オレンジ：4.1～6.0，ピンク：6.1～15.0，赤：15.1～

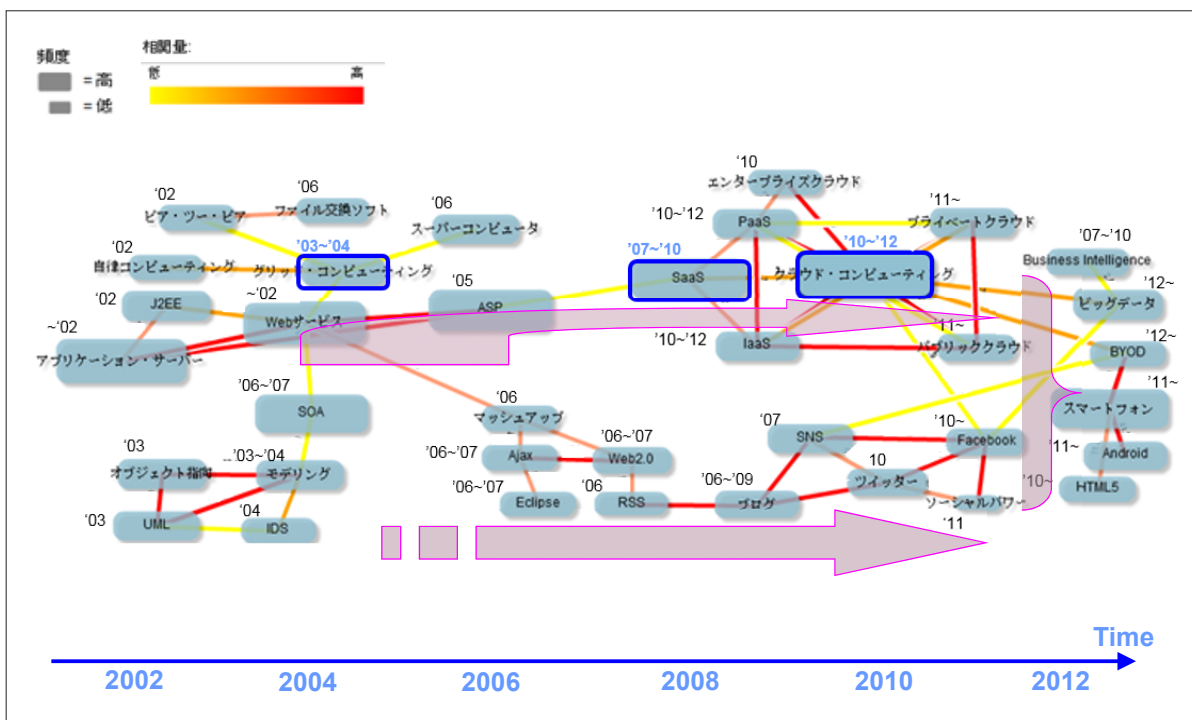


図9 クラウド事例分析 (簡素化バージョン)

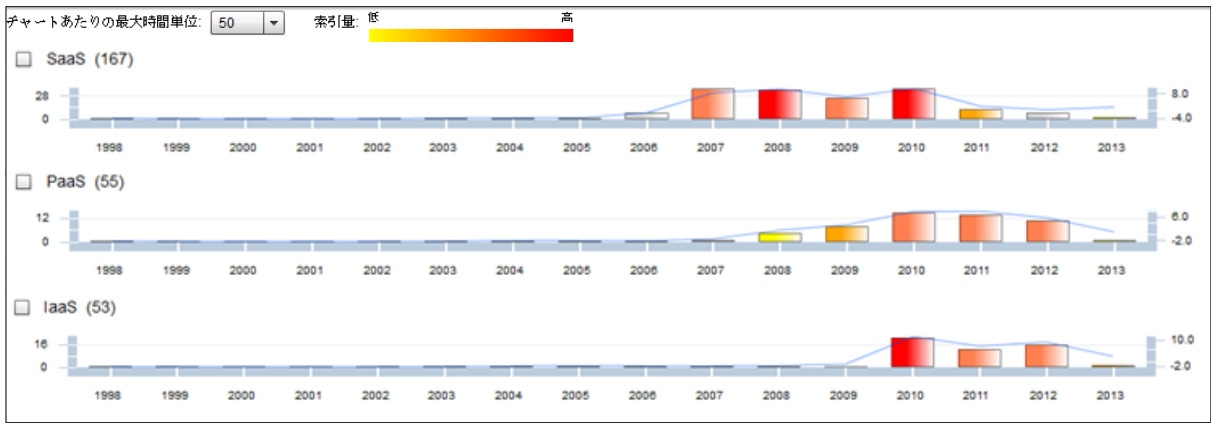


図 10 偏差分析 (SaaS, PaaS, IaaS)

3.4 コンセプト進化の形態

本節ではステップ3の続きとして、さらに「活用シナリオ検討」を進め、得られた分析結果をもとに、クラウドに至るまでのITコンセプトについて、ITビジネスのサービスの視点から進化の様子をまとめる。

2002年から2005年にかけては、サービスの構成要素となるキーワードがそれぞれ相関を持ちながら頻りに記事の中に登場した時代である。技術コンセプトであるグリッドコンピューティングは、Webサービス、ASP (Application Service Provider) と強い相関を示す。後段のSaaSやクラウドが「XaaS」「xxクラウド」などと用語の変形を取りながら推移していったのに比べて、グリッドの時代には関連するソフトウェアやハードウェアの技術キーワードや製品名が多く登場している。特にWEBサービスとASPは、グリッドコンピューティングと一体になると、将来のSaaSの前身ともいえるサービスを形成する。すなわち、それぞれがサービスの一要素である。従って、この段階をサービスの構成要素となるキーワードの集合段階の形態であると位置づける。

2007年からは、SaaSに追随する形でPaaS、IaaSが登場している。サーバの世界では上位層とされるアプ

リケーションレイヤーに位置するSaaSに対して、PaaSとIaaSは中間層のミドルウェア、下位層のインフラにそれぞれ対応する造語である。PaaSとIaaSは、Step2で用いた偏差分析の出力の図10によって時系列で見られるように、SaaSに続いて現れた概念である。従って、これを先に述べたサーバの技術的な階層構造に関連づけて解釈すると、この段階のキーワードは中間層と下位層の概念部分を補完するために新たなコンセプト用語を創出して当てはめている。すなわちコンセプトは、垂直方向に補完するための発展を遂げたと言える。

2010年になるとコンセプトはクラウドコンピューティングに発展するが、これに対して相関の強いキーワードとして現れるものはプライベートクラウドとパブリッククラウドであり、それらの出現の時系列は図11に示される。この他にもIT雑誌の記事上には「xxクラウド」という、クラウドコンピューティングを細分化するための用語が出現しているのが散見される。例えば、官公庁向けの自治体クラウド、病院向けの医療クラウドなど、それぞれの業界に特化したクラウドシステムの用語が出現するなど、いわゆるカスタマイズが起こるのである。SaaSに見られたような技術レイヤーの面での発展は起こっていないため、この段階のキーワードは水平的発展を遂げていると言える。

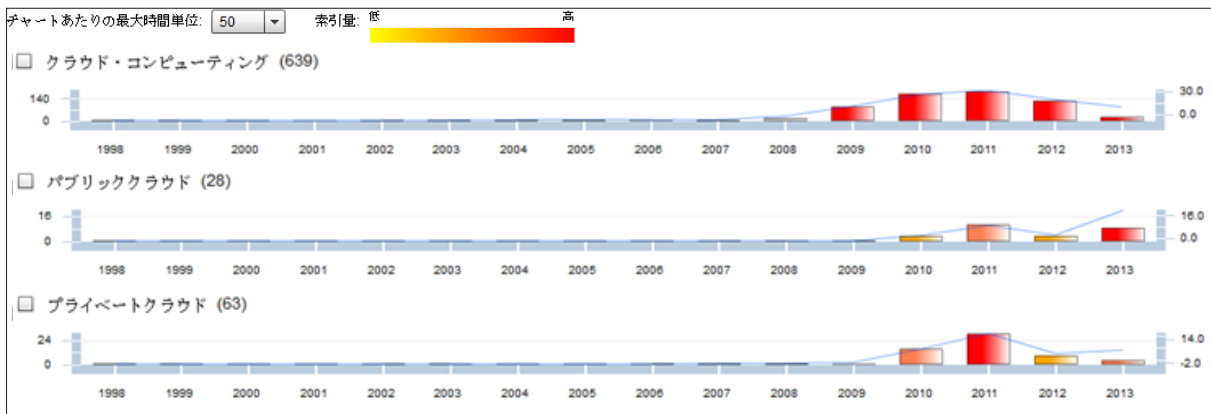


図 11 偏差分析 (クラウドコンピューティング)

これらの現象をまとめて、全体を集合・補完・特化の三段階の形態として表したものが、図12である。上段はテキストマイニングの結果から読み取れる発展形態を概略化したものであり、下段はさらに垂直と水平の階層概念を取り入れて表したものである。第一段階のサービスの構成要素の集合が第二段階のSaaSに発展したものであるから、SaaSの構成要素はひとつの集合体で表現される。PaaSとIaaSも同様の要素を必要とするので同様の構成概念を持つものである。上層階のSaaSに対応する中間層、インフラ層のサービス要素を補完している。さらにクラウドについては、米国NISTの定義によると、SaaS、PaaS、IaaSの3つのサービス概念で構成されていると説明されている。よって、クラウドコンピューティングは従来から存在するそれらの概念を言い換えたもの、あるいはさらに発展させたもの、というのがIT業界においての一般的な見方である。SaaS、PaaS、IaaSは単にその状態を表しただけでコンピューティングという呼び方はしていないが、同様にコンピュータの構成・利用に関するコンセプトのことである[11]。よって、第三段階のクラウドは、第二段階を構成要素とするクラウドの総称の下に、サービスを提供する分野に準じるように特化が進んでいると言える。

このようにクラウドの事例においては、そのコンセプトの進化には3つの形態が確認される。まず、サービス事業の構成要素が「集合」する形態である。技術であったり、事業形態であったり、主体であったり、サービス事業を構成するために必要な個々の要素がキーワードとして登場し、それらが集合することによって、ひとつのサービス事業が形成される。次にそれらの集合を単位として、さらにひとつの新しいコンセ

プト、すなわちSaaSに進化する。そしてそのコンセプトを「補完」するように、技術レベルの異なるキーワードが続いて発生する。補完形態の段階が終わると、それらのコンセプトをひとつの単位としてクラウドという新しいコンセプトが登場するのだが、その発展の形はこれまでと異なりより細分化し、サービスを提供するユーザーに応じて、あるいはユーザー分野別に「特化」の形態を示すものとなる。

第4章 考察

前章で導き出したコンセプト進化の形態について考察を行う。分析結果についての知見を深めるために、IT業界の第一線で活躍する3名⁷の有識者にインタビューを実施し、コンセプトの進化に関して意見をもらった。進化の各段階に関係するコメントの部分、および、その他のコメントを表1にまとめた。これらのインタビューの結果も加味して、進化モデルについて総合的に次のように考える。

クラウドはマーケティング用語ではなく、グリッドコンピューティングという技術用語から出発したものである。それが時代の変遷を経て、次の段階でマーケットニーズによって拡張されSoftware as a Serviceとして定着した。PaaSとIaaSは、SaaSに続いてIT業界のサービスにとって必然的に現れた概念であるから、すなわち補完に至ったとの解釈をすることができる。

SaaSはSoftware as a Serviceという状態そのものを表しているキーワードであり専門的すぎるが、クラウドというと自分勝手に解釈できる余地がある。クラウドは、経済の問題やセキュリティの問題など、あらゆる角度に概念が広がっているというのがクラウドというコンセプトが大きく浸透した理由と考えることが

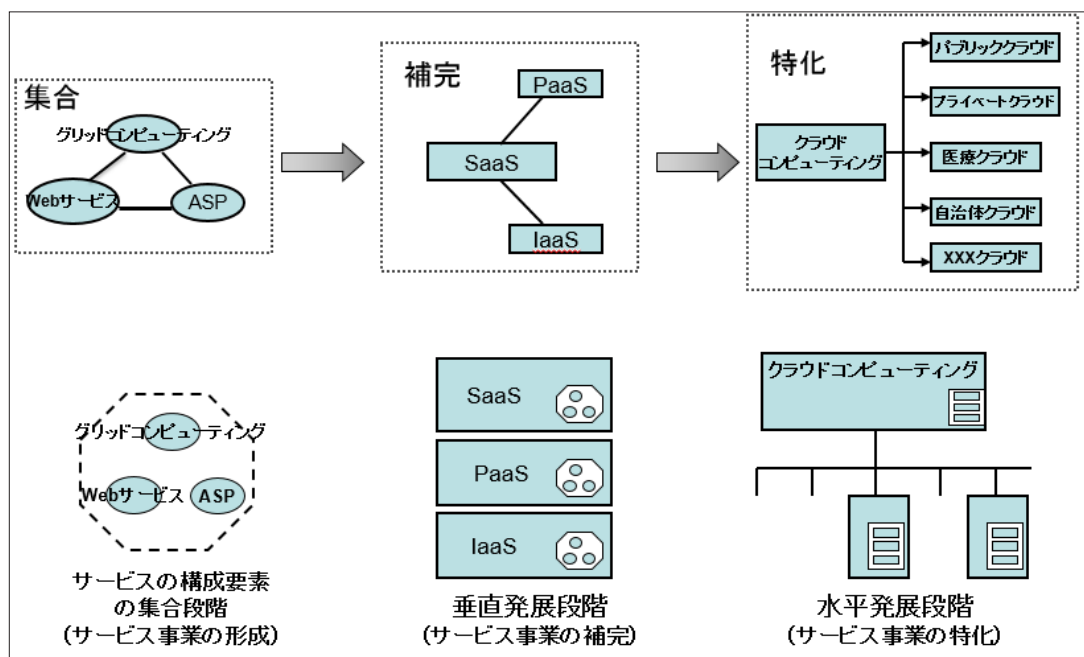


図12 コンセプト進化の形態 (クラウド事例)

できる。そして、業界が異なれば規則やスタンダードが異なるため、その規範の中でクラウドを構築する必要があるから、クラウドは特化したのである。

すでにコンセプトのトレンドはすっかりクラウドに移っているが、いつまでもクラウドを SaaS のことだろうと語る声も根強くある。流行言葉を語っても要するに SaaS と同様ではないか？という意見である。しかし、SaaS がクラウドになるということは、実はク

ラウドは SaaS ではないことを意味する。「クラウドのことをちゃんと分からずに、それを同じと言った瞬間に進化論的には置いてきぼりになる」(井芹氏)と指摘されるように、何かが拡張されることによって他の概念を帯びて行き、使われる単語は変わっていくのである。

表1 インタビューコメント

	進化モデルについての分析コメント
竹尾	<p>第一段階: Grid の時代は、用語を掛け合わせるとか変化させるというのではなく、<u>単にいくつかの関連するワードが出現した。</u></p> <p>第二段階: SaaS は技術用語なので、専門的すぎて、自由な解釈には至らない。<u>技術的には登場人物をそろえないと語れないため、PaaS, IaaS が出たと解釈する。</u></p> <p>第三段階: クラウドは細分化でなく、業界ごとのセキュリティなどの制約があるため<u>業界独自のライフサイクルを通じて、それぞれ特化したのであろう。</u></p> <p>その他: 分析結果をみると、SW に関する用語と、Social メディアに関する用語とが交錯している部分があるが、両者を比較すると、Social の方が世の中に与えるインパクトが大きい。よりインパクトの大きいほうが残っていくのではないかな。</p>
井芹	<p>第一段階: <u>最初は技術背景が必要な</u>のであって、それらの名前が登場する。</p> <p>第二段階: SaaS という<u>名称はそのものを述べているだけ</u>なので、PaaS, IaaS も順番として、そのあとの位置にハマったということだ。</p> <p>第三段階: クラウドは2年たっても、新しい用語に置き換わらないで、xx クラウドとなっている。これは概念定着したこと。<u>クラウドが概念定着している</u>ので、クラウドが何かと論じる必要がない。</p> <p>その他: IT 系のキーワードは米国からくる。イベントだったり、WEB 記者会見、プライベートコンファレンスで企業メッセージとして出てくる。基調講演のクラウドも意図的に出てきた。インターネットが始まった昔からネットワークを雲で書いていた。その概念はあった。ターミノロジーの元はあった。それが時代の変遷を経て、マーケットニーズがあって、できることが拡張されて、単語が定着する。</p>
谷島	<p>第一段階: 定義がしっかりある言葉なので、ワーディングがシャープで他の言葉が便乗しにくい。そのため、用語を並べようか、というイメージなのは。</p> <p>第二段階: 編集側の意図にプラスして、ビジネス上の事由が加わって用語が拡大されていったもの。</p> <p>第三段階: x xクラウドはひとつの成熟の表れである。この言葉が長持ちするから応用して使っている。定義が広くて関係ないものまで包含している。</p> <p>その他: 展示会のテーマについては事業構造が広告モデルで動くため、皆が出展したくなるようなテーマを掲げる。これとは別に、世の中全体にどれだけ影響があるのはやはり新聞であり、新聞が使い出すと広がり大きい。新聞の場合トレンドを創るつもりはないが、構成は仕組みが決まっていて、記事の大きさとかを決めるのは IT の素人=ゼネラリストである。極めて主体性がないメカニズムであるが、繰り返し掲載することによって大衆化が始まる。書き手側も大きく乗せたいために看板をつける。</p>

第5章 まとめ

本研究ではIT業界のクラウド事例に注目し、テキストマイニングの手法を使って用語間の相関を抽出することによって、クラウドに至るまでのコンセプトがどのように推移したのかを分析し時間軸と共に提示した。この結果をもとに、コンセプトが進化する過程には、集合・補完・特化の3つの形態が存在することを発見し、新たに提示したことに学術的意義がある。これまでも競争優位を確保するために差別化戦略のひとつとして新しいコンセプトを使うことがあったが、コンセプトの推移を時間的に連続したものとして捉えて述べるものではなかった。

すでにクラウドは2010年に頻繁に取り上げられるようになってから数年が経過し、今日ではすっかり定着している。クラウドが何かということをもはや論じる必要はなく、さらに細分化したxxクラウドについて直接、論じることができるようになった。クラウドコンピューティングは最終的にそれぞれのケースや業界に合わせた特化が起こり、これによってコンセプトは成熟したと言えよう。

このように本研究によって提示したコンセプトの進化の分析手法を使えば、コンセプト用語が時間と共にどのような変化をたどったかを明らかにすることができる。藤垣・永田が提唱するように、文書についての定量的アプローチを行うことは、コンセプトが実際にどう変化しているかを分析すること、および、次のトレンド作りに応用することが可能となる [5][26]。しかし、従来の手法では、扱うデータ量にも限界があるためコンセプト用語の登場頻度をもとに分析を進めざるをえなかったが、本研究ではこの定量的分析手法をさらに発展させテキストマイニング手法を導入することによって、膨大な量のデータを扱うことを可能とした。その結果、コンセプト進化の分析において、図5のようにファセット分析により相関値を抽出したり、図6のように偏差分析を活用することによって時系列要素を取り入れることによって、多面的でより精度の高い定量的分析手法が有効であることを新たに提示したことに実務的意義がある。

今後もクラウド以降に出現したコンセプト用語の分析を継続して行うことによって、コンセプトの進化パターンを複数事例で考察したい。本研究のアプローチは、ビッグデータやIoTなどの現在進行形のコンセプト用語の推移にも適用可能である。今後再び、集合・補完の段階をたどり、特化をもって用語が成熟したと判断することができるであろう。あるいは、「集合」「補完」「特化」の3つの形態の順番が異なる場合、3つの形態のうちのいくつかのみ、あるいは全く新たなコンセプトの発展形態の出現が期待できるかもしれな

い。いずれにしてもコンセプトが形成される条件や、どのような形態で推移するのかなど、コンセプトの特性に関する理解を得ることが重要である。こうした研究を重ねることによって、現在の業界で注目されているコンセプトがビジネスの発展段階のどこの位置にあるのかを理解し、その後のIT業界が向かっていく方向を予測できる。例えば、補完の段階が終了していない集合段階のコンセプトである場合は、実際のビジネスでの応用はまだ数年先だと予想される。一方で特化が始まっている場合には、早々にさまざまな分野に水平展開され応用されていくことになるであろう。こうしてコンセプトの発展段階を通して業界内を把握することによって、企業として近い将来、どの分野に投資しビジネスを拡大していくべきであるのか、というような戦略立案に役立つことを期待する。

謝辞

インタビューにご協力いただき貴重なご意見を賜りました日経BP社 谷島宣之氏、株式会社インプレス R&D 井芹昌信氏、日本IBM株式会社 竹尾さつみ氏に深く感謝します。テキストマイニング手法についてご指導をいただいた日本アイ・ビー・エム株式会社 東京基礎研究所 那須川哲哉博士、吉田一星博士にお礼を申し上げます。

(かたおか りえこ、いかわ やすお、うちやま なおし)

《注釈》

1. ITアーキテクチャ：コンピュータの論理的構造や機能面から見たときの構成方式のこと。
2. National Institute of Standards and Technology, 米国国立標準技術研究所。
3. <http://www.redbooks.ibm.com/abstracts/sg247568.html>
4. 区間推定法は、「文書が無限にある場合の真の相関値 α を未知数とした際、一定確率以下の偶然を除いて、現状の相関値を実現し得る最小の α 」を算出するものである。
5. 「IBM Content Analytics with Enterprise Search (ICA)」より引用した。
6. 図中の各線には相関値が併記されているが、画面の解像度の関係上、相関値色別の値を目安とすることで対応する。
7. 有識者インタビュー：役職はインタビュー当時のもの。
谷島宣之氏：日経BPイノベーションICT研究所 上席研究員、「日経コンピュータ」元編集長、2013年5月28日、日経BP本社にて実施。
井芹昌信氏：株式会社インプレス R&D 代表取締役社長、2013年6月14日、インプレス R&D 本社にて実施。

竹尾さつみ氏：日本 IBM 株式会社、グローバル・テクノロジー・サービス事業部 技術理事

《参考文献》

1. Alvesson, M. and et.al (2001) "Odd Couple: Making sense of the curious concept of knowledge management", *Journal of Management Studies* 28, 7 November, pp.995-1018
2. Arino, K. and et.al (2011) "Temporal Network Analysis of Emerging Technologies", *Topic Transition in World Wide Web (WWW) Conferences*, National Institute of Science and Technology Policy
3. Eisenhardt, E. (1989) "Building Theories from Case Study Research", *the Academy of Management Review*, 14(4), pp.532-550
4. Foster, I. and et.al (2008) "Cloud Computing and Grid Computing 360-Degree Compared", *Grid Computing Environments Workshop*, Conference Publications
5. Fujigaki, Y. and Nagata, A. (1998) "Concept evolution in science and technology policy: the process of change in relationships among university, industry and government", *Science and Public Policy*, Volume 25, Issue 6, 1 December 1998, Pages 387-395
6. Giroux, H. (2006) "It Was Such a Handy Term; Management Fashions and Pragmatic Ambiguity", *Journal of Management Studies* 43, 6 September, pp.1227-1260
7. Gupta, V. and Lehal, G. (2009) "A Survey of Text Mining Techniques and Applications", *Journal of Emerging Technologies in Web Intelligence*, Vol.1, No.1, August 2009
8. IBM (2006) "Introducing OmniFind Analytics Edition", *Customizing Text Analytics*, an IBM Redbooks publication, <http://www.redbooks.ibm.com/abstracts/sg247568.html>
9. IBM SWG (2006) "IBM Content Analytics with Enterprise Search (ICA)" *IBM ソフトウェア 非構造化データの分析*, IBM-SWG-ICA-3
10. La, H.J. and et.al (2009) "Technical Challenges and Solution Space for Developing SaaS and Mash-Up Cloud Services", *The IEEE International Conference on e-Business Engineering (ICEBE)*, pp. 359-364
11. Mell, P. and Grace, T. (2011) 「NIST によるクラウドコンピューティングの定義」 *National Institute of Standards and Technology*, NIST Special Publication 800-145
12. Porter, E. M. (1980) "Competitive Strategy: Techniques for Analyzing industries and Competitors", Free Press
13. Tseng, Y. and et.al (2007) "Text mining techniques for patent analysis", *Information Processing and Management* 43, pp.1216-1247
14. Tsumoto, S. and Abe, H. (2012) "Similarity of Concepts Identified by Temporal Patterns of Terms in Biomedical Research Documents", *Rough Sets and Knowledge Technology*, 7th International Conference, Proceedings, pp.232-241
15. Yin, R. (2008) "Case Study Research: Design and Methods (4th Edition)", Sage Publications
16. Yoon, B and Park, T. (2003) "A text-mining-based patent network: Analytical tool for high-technology trend", *Journal of High Technology Management Research* 15, pp.37-50
17. Villegas, D. (2010) "The Role of Grid Computing Technologies in Cloud Computing", *Handbook of Cloud Computing 2010*, Part 1, pp.183-218
18. 奥和田久美、白井康之、小関悠 分野別の自由記述から科学技術政策上意味ある意見を自動抽出する試み (メトリクス, 一般講演, 第 22 回年次学術大会) 研究・イノベーション学会 (旧: 研究・技術計画学会) 年次学術大会講演要旨集 22 巻 pp.692-695.
19. 菟田文男、那須川哲哉 (2003) 『技術戦略としてのテキストマイニング』 中央経済社
20. 菅澤喜男 (2008) 「テクノロジー・インテリジェンス」 研究・技術計画学会誌, Vol.23, pp.28-35
21. 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」, MRI 技術レポート 5. 14, p110-123 (2009)
22. 竹岡志朗、井上雄介、高木修一、高柳直弥 (2016) 『イノベーションの普及過程の可視化』 日科技連、pp.90-103
23. 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
24. 日経コンピュータ編集局 (2002-2012) 『日経コンピュータ』 日経 BP 社
25. 日経 BP 社出版局編 (2011) 『クラウド大全』 日経 BP 社
26. 藤垣裕子、永田晃也 (2000) 「科学技術政策コンセプトの進化プロセス」 科学技術庁 科学技術政策研究所, POLICY STUDY No.5
27. 三崎文隆 (2010) 「クラウドの適材適所の活用に向けて」 *IBM Provision*, No.64, pp.64-67
28. 脇森浩志 (2013) 「ビッグデータに対するテキストマイニング技術とその適用例」 *UNISYS TECHNOLOGY REVIEW* 第 115 号, MAR. 2013 pp.337-349