

## [ 論文 ]

# オープンソース・インテリジェンスの競争分析への 活用の戦略的枠組み

テキスト・マイニングによる日本の製薬業界の 2010 年問題における  
M&A 情報分析を事例として

## Strategic Framework of Open Source Intelligence for Competitive Analysis

A Case Study of Japanese Pharmaceutical Industry M&A  
on the 2010 problem through Text-Mining

中島庸介\*, 保井俊之\*\*, 神武直彦\*\*\*

### Abstract

Since the global financial crisis in 2008, the US and other intelligence communities realized the significance of the open source intelligence (OSINT). The OSINT is widely used for the intelligence for the national security. This paper is to illustrate opportunities and threats of the OSINT when it is applied to the competitive intelligence; and to put the text-mining as the core methodology for that application. For the case study of the application, this paper analyses the M&A developments of the Japanese pharmaceutical industry for the last decade, which holds the 2010 problem. We also validate the effectiveness of the text-mining as the OSINT methodology for the competitive intelligence.

### 要旨

2008 年のグローバルな金融危機以降、オープンソース・インテリジェンスの重要性が米国のインテリジェンス・コミュニティを中心に認識され、国家保障に関するインテリジェンスに積極的に活用され出している。本論文は、オープンソース・インテリジェンスをコンペティティブ・インテリジェンスの手法として戦略的に活用する際の好機と制約を明らかにしつつ、テキスト・マイニングをその中核に位置付ける。そして、その具体的な活用事例として、2010 年問題を抱える日本の製薬業界のこの 10 年間の M&A の動向を取り上げ、手法としての有効性を分析・検証する。



### \* Yousuke NAKAJIMA

- 慶應義塾大学システムデザイン・マネジメント研究科 後期博士課程
- 慶應義塾大学システムデザイン・マネジメント研究科修士課程修了
- 修士（システムデザイン・マネジメント学）
- 〒223-8526 横浜市港北区日吉4-1-1 協生館SDM 神武研究室
- y.n@z5.keio.jp
- Doctoral Program of Graduate School of System Design and Management, Keio University
- Graduate School of System Design and Management, Keio University
- Master of System Design and Management
- Collaboration Complex, 4-1-1 Hiyoshi, Kohoku-ku, Yokohama-City, 223-8526, Japan



### \*\* Toshiyuki YASUI

- 慶應義塾大学先端研究センター特任教授
- 東京大学教養学部教養学科国際関係論分科卒業
- 国際基督教大学 博士（学術）
- 〒223-8526 横浜市港北区日吉4-1-1 協生館SDM
- 045-564-2461
- t.yasui@z2.keio.jp
- Project Professor, Keio Advanced Research Center, Keio University
- Faculty of International Relations, School of Arts and Sciences, the University of Tokyo
- Ph.D., Arts and Science, International Christian University
- Collaboration Complex, 4-1-1 Hiyoshi, Kohoku-ku, Yokohama-City, 223-8526, Japan



### \*\*\* Naohiko KOHTAKE

- 慶應義塾大学システムデザイン・マネジメント研究科 准教授
- 慶應義塾大学大学院政策・メディア研究科 博士課程修了
- 慶應義塾大学 博士（政策・メディア）
- 〒223-8526 横浜市港北区日吉4-1-1 協生館SDM
- 045-564-2580
- kohtake@sdm.keio.ac.jp
- Associate Professor, Graduate School of System Design and Management, Keio University
- Graduate School of Media and Governance, Keio University
- Ph.D., Media and Governance, Keio University
- Collaboration Complex, 4-1-1 Hiyoshi, Kohoku-ku, Yokohama-City, 223-8526, Japan

## 1. 問題の所在と分析の視覚

### 1.1 問題の所在

オープンソース・インテリジェンス(以下、OSINT)は、米国のインテリジェンス・コミュニティのコンセンサスと考えられる米国の2006年度国防歳出権限付与法の定義によれば、「特定のインテリジェンス要求に対応するために、適切な依頼元に適時に、公表情報から収集され、取り出され、配布されるインテリジェンス」と規定されている(US Department of Army [1] 2-1頁)。HUMINTなどの他のインテリジェンス収集方法に比べ、OSINTはインテリジェンスの収集から配布までそれぞれの目的に対応したプロセス(end-to-end processes)になっていないため(Lowenthal [2] 77頁)、インテリジェンスの手法としては一段格下にみられる傾向にあった。

しかしIT技術の発達による高度情報化社会の到来を迎え、2001年の米同時多発テロの発生と2008年のグローバルな金融危機の発生とともに、普段はノイズとして無視されている公開情報によるインテリジェンスの重要性について、米国インテリジェンス・コミュニティを中心とする各国のインテリジェンス・コミュニティが認識する契機となった(保井俊之[3] 30-31頁)。

OSINTはこのように、この数年の間で米国を中心とするインテリジェンス・コミュニティに最も注目され、活用されるインテリジェンス手法となった。またOSINTは公開情報の活用を基本とする点(McGonagle & Vella [4] 3頁)において、コンペティティブ・インテリジェンスとも親和性の高いインテリジェンス手法である。他方、コンペティティブ・インテリジェンスの手法としてOSINTを活用する際に、確立した応用の方法論並びに応用を促進するツールの研究及び開発の蓄積は十分とは言い難い。

### 1.2 分析の視覚

OSINTの特徴である大量の情報収集と処理の標準的プロセス手法として、マーケティングの世界において、テキスト・マイニングが社会システムの文脈分析も行えるツールとして近年脚光を浴びている。例えば、坪井塑太郎・萩原清子[5] (327-328頁)は、ある情報収集テーマに関する考え方と事実を単位化し、グループ化と抽象化を繰り返すことで、問題を明確化・構造化する手法としてテキスト・マイニングの手法としての適性を挙げている。

このようなテキスト・マイニングの問題の明確化・構造化の能力は、大量の情報処理と分析を必要としながらも、インテリジェンス・サイクルの最初では鍵となる目標収集テーマ並びに標的の問題が必ずとも明らかになっていないOSINTに必須と考えられる。

したがって、本論文の分析の視角としては、OSINTの標準的ツールとしてテキスト・マイニングがどの程度有

効なのかについて、論考を進めることが必要である。

## 2. 本研究の対象と検証の方法

### 2.1 テキスト・データ分析とCIの活用の研究

製薬業界をテーマにしたCIの活用とデータ分析に関する研究として、高橋義仁[6]によるオープンソース(公開されている情報)の収集とインテリジェンスの構築(集められた情報からその「意味合い」を抽出するプロセス)およびその活用についてのコンペティティブインテリジェンス(CI)の重要性を説いたものや、調査票を用いてデータを収集し、製薬企業でのCI活用の能力と研究開発の関係というテーマで分析をおこなったものがある(高橋義仁[7])、(高橋義仁・下村博史[8])。また、インテリジェンスとテキスト・データ分析に関する研究として、富田準二[9]によるビジネスインテリジェンスをめぐるテキスト集約技術についての研究がある。さらに時間情報をもつテキスト・データの分析に関する研究として櫻井茂明、植野研ら[10]による営業日報のテキスト・データを用いた時系列のイベント発生パターンを抽出する手法を扱ったものがある。

また米デジタルグローブ社等が提供する商業衛星写真を他国が自国に及ぼす軍事的危険性に関わる動向の分析に利用した画像のオープンソースによる競争分析に関する事例があり(YOMIURI ONLINE[11])、この際にはその衛星写真が撮影された時間が特に重要な意味をもつことが知られている。このようにオープンソースを競争分析に用いる場合、その時間情報はとても重要であると考えられている(北岡元[12] 101-102頁)。

しかし、製薬業界の2010年問題といった時間情報をもつ具体的事例を題材にテキスト・データ形式のオープンソースを競争分析に用いた研究は過去に実施されていない。

### 2.2 本研究の対象とその検証方法

これまでの研究にテキスト・データからインテリジェンスを構築する、あるいは時間情報をもつテキスト・データを集約するといった個別にこれらを扱ったものはある、しかし実際に現実の事例にこれら技術を応用するには、時間情報をもつテキスト・データを集約するという先行研究を組み合わせた手法が必要であり、またその手法による結果の確かさと有用性を検証する必要がある。

そこで本稿では、時間情報をもつオープンソースによるテキスト・データを集約し、インテリジェンスとしての競争分析を実践するための手法を提案し、この手法による結果の確かさと有用性を検証する。検証は、同様の話題が詳しく簡潔に解説された専門誌の特集記事との比較をもっておこなうこととした。

### 3. 本稿で提案する分析手法

#### 3.1 分析に用いたデータ

本稿では、2種類のオープンソースを利用する。オープンソースAとして2010年の日経ビジネスの2010年問題に関する特集記事(日経ビジネス[13])および日経マネーに掲載された専門記事(日経マネー[14])を、オープンソースBとして国内の主要新聞3紙(朝日新聞, 毎日新聞, 読売新聞)の2011年5月を起点として過去10年分(2001年6月~2011年5月)のテキスト・データを分析用データとして用いる。

オープンソースAの特集記事からこれまでの経緯に注目して要約図を作成し、オープンソースBについては、テキスト・マイニング手法を用いて分析を行う。最後にこれらAとBの分析結果を比較しその内容の整合性を確認した。

#### 3.2 テキスト・マイニングによる用語の分析

はじめにオープンソースBに対するテキスト・マイニングの手法を用いた分析をおこなった。まず、先の主要3紙に対し「(製薬 or 医薬 or ファーマ) and M&A」という条件にて記事検索を行った。3紙合計で705の記事が検索条件に合致し、これらを分析に用いた。過去10年間の705の記事の1年ごとの発現頻度を確認した。各年によって40程度の記事数から100を超える記事数まである程度の幅があることが確認できる(図1)。

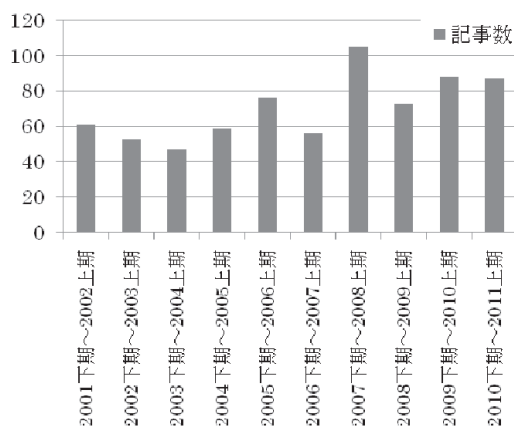


図1 1年ごとの記事数

次に、これら705の記事中に現れる単語の発現数を確認した。これら記事からの総抽出単語数は289,577で、異なる単語数は12,674、分析に使用可能な単語数は6,830であった。

次に、抽出した単語を組み合わせ「2010年問題」、「特許切れ」、「敵対的買収(TOB)」および「提携」の4用語を抽出する簡単なコーディング・ルールを作成し、時系列発現頻度を確認した(図2)。これら4語を対象とした理由は、今回の主要な分析対象用語であり、まずはこれらの用語の発現の分布状況を確認するためである。

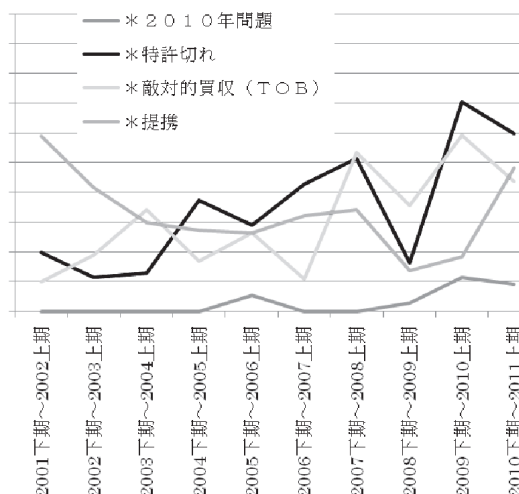


図2 分析対象となる主な用語の発現頻度

図2から「2010年問題」、「特許切れ」、「敵対的買収(TOB)」の3用語は、発現する用語の頻度には差はあるが、年代を追うごとに増加傾向にあり、「提携」は2000年代前期に多くみられ、その後各年ほぼ均等に発現していることがわかる。ここから「特許切れ」に起因する「2010年問題」や「敵対的買収(TOB)」に関する記事が実際の2010年に近づくにより増加し、一方、製薬企業間のM&Aや業務および製剤別の「提携」は過去10年間に亘って恒常的に行われているのではないかという推測が可能である。

本稿では、このような推測が確認をもって正しいと言えるのかどうかを確かめるため、以下においてこれら新聞記事からのオープンソースBのテキスト・データの分析をさらに進め、最後にオープンソースAの要約と比較し、本稿で提案する分析の手法を評価する。

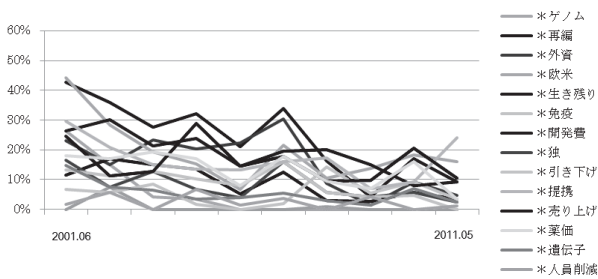
#### 3.2.1 過去10年間に発現した用語の発現傾向の把握

先に分析対象とした4用語に加え、オープンソースBから分析に意味のありそうな名詞、サ変名詞、固有名詞、組織名および地名を単語として抽出し、すべての用語を対象に1年ごとに発現率を求めた(図2)。分析対象を単語ではなく、単語と単語の組み合わせも含めた用語とした。これは単語と単語の組み合わせによって、それらひとつの単語の場合と意味が異なるものを区別するためである。例えば、技術分野のひとつを意味する「バイオ」とその成果物である「バイオ医薬品」では記事中での意味が異なるため、作成したコーディング・ルールを用いて区別した。具体的には、コーディング・ルールとして「バイオ」と「医薬品」という単語が連続して現れた場合のみ「バイオ医薬品」として認識するという指示を与えている。また、以下の分析では記事中に発現する用語の頻度を調べるため、ひとまず企業名は分析対象から除いている。

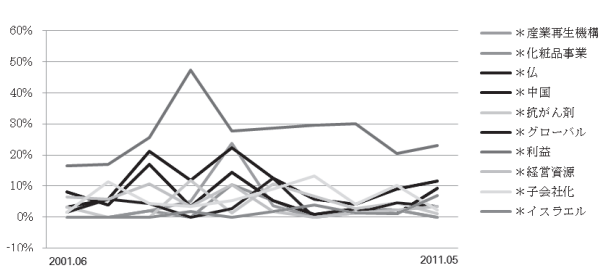
10年間のデータ中に発現した用語は、企業名を除いた

名詞、サ変名詞、固有名詞、組織名および地名の分析に意味のありそうなものだけでも 400 用語を超えていた。また 1 年という期間でみた場合、抽出された用語の発現数に大きな差みられたため、それぞれの期間毎に発現した用語の発現頻度からカイ二乗値を算出し、カイ二乗値の大きいものから上から順に 45 用語を以下の分析に使用することとした。これら 1 年間毎のそれぞれの用語の発現率を求めその発現傾向確認した。すると、これら用語には①過去 10 年間で発現傾向が減少にある用語の集団、②一時的に増加がみられた用語の集団および③発現が増加傾向にある用語の大きく分けて 3 つの傾向を示す集団があることがわかった。これらを時系列に図 3 としてグラフに示し、同様に表 1 としてリストとして示した。

① 発現率が減少傾向にある用語



② 一時的に増加がみられた用語



③ 発現率が増加傾向にある用語

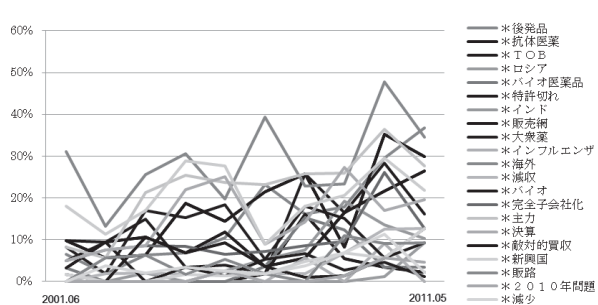


図 3 発現率の傾向

表 1 発現率の傾向別用語

傾向	用語
発現率が減少傾向にある用語	ゲノム, 再編, 外資, 欧米, 生き残り, 免疫, 開発費, 独, 引き下げ, 提携, 売り上げ, 薬価, 遺伝子, 人員削減
一時的に増加がみられた用語	産業再生機構, 化粧品事業, 仏, 中国, 抗がん剤, グローバル, 利益, 経営資源, 子会社化, イスラエル

発現率が増加傾向にある用語	後発品, 抗体医薬, TOB, ロシア, バイオ医薬品, 特許切れ, インド, 販売網, 大衆薬, インフルエンザ, 海外, 減収, バイオ, 完全子会社化, 主力, 決算, 敵対的買収, 新興国, 販路, 2010年問題, 減少
---------------	---

①減少傾向がみられた用語としては、「ゲノム」、「再編」及び「生き残り」などであった。②一時的に増加がみられた用語としては、「産業再生機構」、「化粧品事業」および「仏」などであった。③発現の増加傾向がみられた用語としては「後発品」、「抗体医薬」及び「TOB」などであった。

本分析で最初に注目した 4 用語においても、発現が 2010 年にむけて増加傾向にあるもの、発現がある程度一定であるものが存在した。そこで、分析対象データである記事においてその時々で話題となるトピックの潮流のようなものがあり、これが用語の発現率として分析結果に現れているのではないかと考えた。次節では本節とは異なる一定期間別にその潮流が確認できるのか分析を進める。

3.2.2 10 年間で 3 等分した期間に発現した用語の分析  
3.2.2.1 共起ネットワーク分析

前節までの分析から、対象とするデータ中に時系列に発現する用語の傾向があることがわかった。本節では、の傾向をつかむため過去 10 年間ぶんのデータを 3 等分し、共起ネットワーク図を用いてそれぞれの期間に発現する特徴語の確認とそれら用語間の関係を視覚化して確認する。尚、本分析にはテキスト・マイニング分析ツールである KH Coder [15] を用いた。

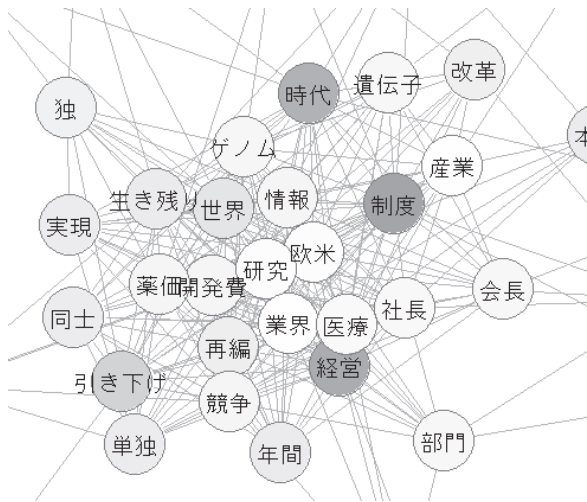
過去 10 年間 120 か月を 3 等分するとひと区分は 40 か月となる。それぞれの期間に呼び名をつけて以降で用いる。呼び名と期間は以下の通りである。

- ・2000 年代前期：2001 年 6 月～2004 年 9 月
- ・2000 年代中期：2004 年 10 月～2008 年 1 月
- ・2000 年代後期：2008 年 2 月～2011 年 5 月

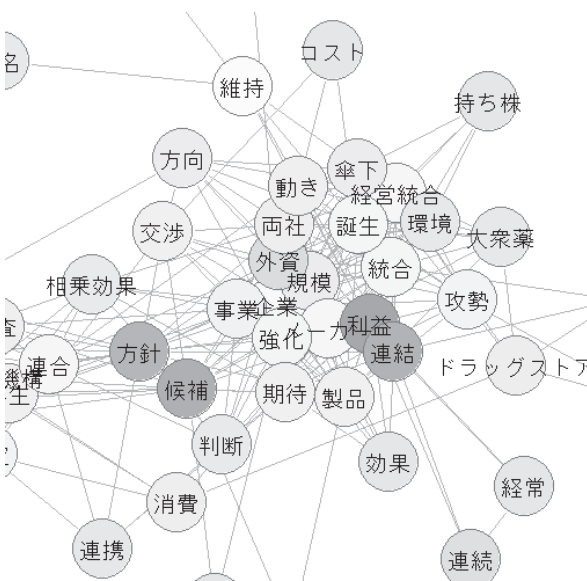
これら 40 か月間のデータをひとまとまりとし、それら期間に発現した特徴的な用語をある定義に従い特定し、発現した用語間の関係を共起ネットワークにて視覚化し確認した (図 4)。特徴的な用語の条件は、各期間でそれぞれ 10%以上の発現率であり且つカイ二乗検定の 1%水準で有意な差がみられた用語と定義した。これらの発現頻度を用いた相関係数から共起ネットワークを作成し、各用語の間のつながりを確認した。

これらの共起ネットワーク図では、分割した 3 期間それぞれの用語の繋がりを確認することができる。2000 年代前期の共起ネットワーク図からは、発現している用語の多くが「開発費」、「薬価」、「再編」、「生き残り」、

2000年代前期



2000年代中期



2000年代後期

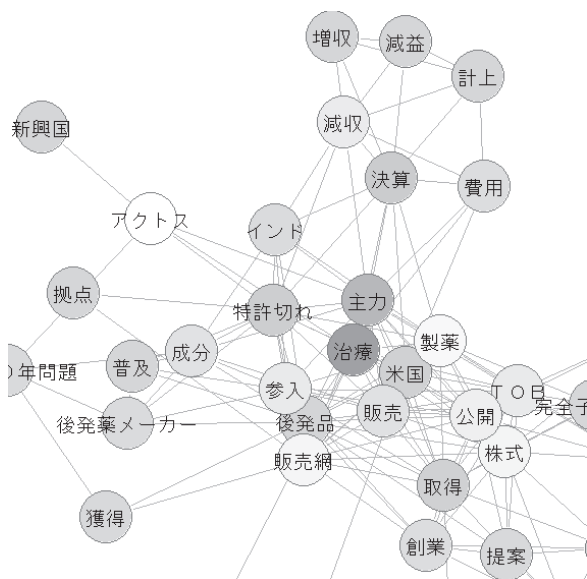


図4 共起ネットワーク(用語)の結果

「世界」, 「経営」および「引き下げ」などといった用語を経由して繋がっていることがわかる。2000年代中期では、「外資」, 「利益」, 「連結」, 「方針」, 「候補」および「連続」などの用語をハブとした用語のネットワーク図が示された。2000年代後期では、「決算」, 「主力」, 「治療」, 「特許切れ」, 「後発品」, 「米国」, 「株式」および「取得」などの用語をハブとした用語のネットワーク図が示された。

以上の共起ネットワーク図から、各期間別に特徴的に発現する用語とその用語間の関係が視覚的に確認することができた。しかし、本稿の目的である製薬業界におけるM&Aの傾向として実際の企業名が分析結果に表れていないことでこのままでは解釈が難しいと考えた。

そこで次に、ひとまず保留にしておいた企業名のデータを用い同様の3期間別に共起ネットワーク図を作成することとした。しかし、この分析を企業名のみで行った場合、結果として得られるネットワーク図に示される企業と企業の関係がどのような関係にあるのかは直感的に理解することはできない。そこで先の共起ネットワーク図にて各期間でハブとして発現していた用語を、この企業名の分析に組み入れて分析をおこなうこととした(図5)。

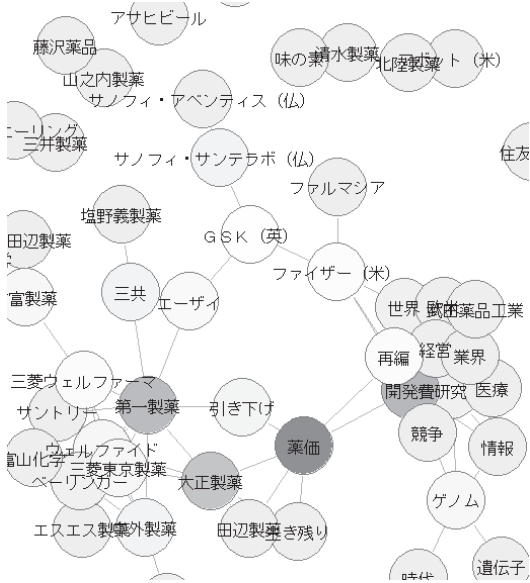
図5では、2000年代前半のデータを用いた分析では企業間の関係が強く示されているが、先の分析でハブとなっていた「開発費」, 「薬価」および「引き下げ」などの用語がこの企業名データの分析でもハブとなっていることが確認できた。2000年代中期の分析でも、その時点での企業間の関係が主に示されており、ここでも「方針」という用語などがハブとなり関係のある企業集団を繋いでいた。2000年代後期の分析でも傾向は同じであった。ここでは「後発品」, 「販売網」および「参入」といった用語がハブとなりネットワークを形成していた。また、それぞれの期間に投入した用語がハブとなっているだけでなく、その周辺の企業名や用語の関係を理解するのを助けている働きがあることがわかる。

このように企業名データに投入した企業名以外の用語が、企業名データの共起ネットワーク図を用いた分析でもハブとなって示され、さらにこれら用語がその周辺に現れる用語や企業名の関係性の理解を助ける機能があることも確認できた。

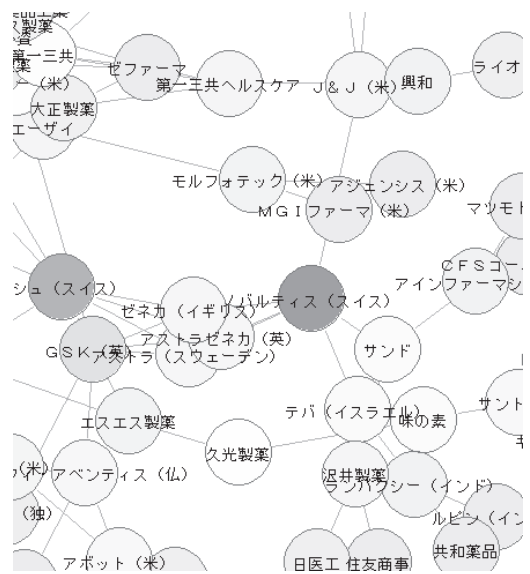
3.2.2 階層別クラスター分析

次に、分析対象データから用語の共起ネットワーク分析に用いた用語と企業の共起ネットワーク分析に用いた企業名データの両方をすべて入れて3期間別に階層別クラスター分析をおこなった。階層別クラスター分析では、分析データに含まれる用語を用語間の相関係数に基づきクラスター化された形でみることができる。ひとつのクラスターと関係の強いクラスターが隣り合う形になって

2000年代前期



2000年代中期



2000年代後期



図5 共起ネットワーク (企業) の結果

表示されるため、各クラスターに含まれる用語と隣り合うクラスターとの関係を容易に確認することができる。また、クラスター分析の結果は、大量のデータを分析することで共起ネットワーク図のように結果の表示データが重なり合ってしまうということがないため表示上のデータの制限がないというメリットがある。以下ではデータ量が多いため分析の結果をすべて表示することはしないがその一部を図6として、またこの要約の例を表2として示す。

2000年代後期

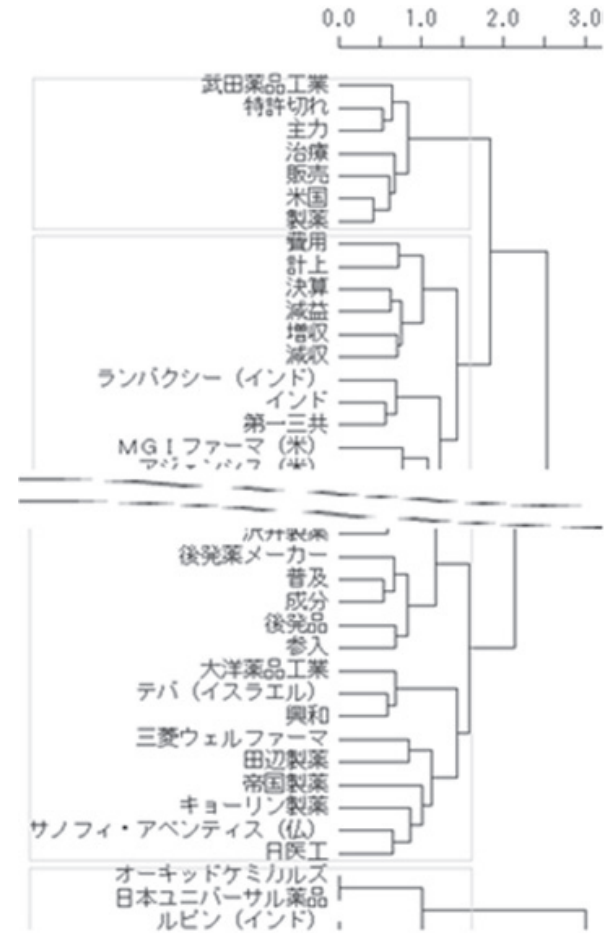


図6 階層的クラスター分析の結果 (抜粋)

表2 クラスター分析の要約の例 (2000年代後期)

区分	用語
特許切れによる決算への影響を意識した国内大手新薬メーカーのM&A	武田薬品工業, 特許切れ, 主力, 治療, 販売, 米国, 製薬, 費用, 計上, 決算, 減収, 増収, 減収, ランバクシー, インド, 第一三共など
TOB関連	TOB, 公開, 株主, 上場, 株式, 取得, 完全子会社化, 発行
新薬メーカーの後発品市場への参入をめぐる動き	沢井製薬, 後発薬メーカー, 普及, 成分, 後発品, 参入, 大洋薬品工業, テバ (イスラエル), 興和, 三菱ウェルファーマ, 田辺製薬, 帝国製薬, キョーリン

	製薬、サノフィ・アベンティス(仏)、日医工
バイオ、抗体医薬などの技術背景を絡めた国内外のM&A	バイオ、バイオ医薬品、抗体医薬、ロシュ(スイス)、ファイザー(米)、ワイス(米)、協和発酵キリンなど
2010年問題と販売網/拠点の獲得	アミリン(米)、2010年問題、アクトス、新興国、ナイコメッド(スイス)、拠点、販売網、獲得など

図6のようにクラスター分析では、投入した企業名と記事中の用語が関係の強い順にクラスター化され可視化された。そこには企業名と用語が組み合わさって示されたクラスター、企業名だけで構成されたクラスターおよび用語だけで構成されたクラスターの3種類をみることができる。しかし、より高次のクラスターでみることで、そこに含まれる企業名と用語の組み合わせからその高次のクラスターの意味が解釈できるまとまりとして見ることができた。例えば、表2で示した「区分」にあるような「特許切れによる決算への影響を意識した国内大手新薬メーカーのM&A」に関する企業名と用語のまとまり、「TOB関連」に関する用語のまとまりおよび「新薬メーカーの後発品市場への参入をめぐる動き」に関するまとまりと解釈することが可能であった。

### 3.2.2.3 共起ネットワークと階層別クラスター分析結果の考察と話題の推移

これまで共起ネットワークと階層別クラスター分析を行った。共起ネットワークの分析では、用語と企業名データをそれぞれ3期間別に分析を行い、用語のデータで共起ネットワーク図を作成した場合にハブとなっていた用語の幾つかは、企業名による共起ネットワーク図のデータに組み入れた場合でもハブとして機能していることが確認できた。またこれら用語を企業名の共起ネットワ

ーク分析に組み入れることにより、これら用語が繋がりが合う企業間の関係を理解する補足情報としての役割を担っていた。

次に共起ネットワークの用語別、企業名別の分析に用いたすべてのデータをまとめてそれぞれの期間別に企業名とそれぞれの用語との関係を確認するため階層別クラスター分析を実施した。階層的クラスター分析でもそれぞれの用語と企業名が関係性をもとにクラスター化され、さらに高い水準でクラスターを眺めることでそれらが意味するものが解釈可能な用語の集合として示されていた。

このように過去10年間の記事を3つの期間に区切り、それぞれの期間別にクラスター分析を行ったことで、それぞれの期間で記事となった内容の要約をみることができた。しかし、時間情報を利用した過去10年間の推移をおしなべて比較するという要望に対しては、同じカテゴリで3期間それぞれのデータを要約し、並べて比較することが必要だと考えた。

そこで、オープンソースBの分析のまとめとして3期間分のデータをそれぞれ同じカテゴリで要約し、それら結果を並べて見比べることでその変化の傾向を確認することとした。まず3期間分の共通のカテゴリの作成であるが、今回は3期間のそれぞれのクラスター分析の結果を注意深く検討し、図7にあるような9つ区分を作成し、それぞれの期間毎に発現する用語の頻度を求めることとした。各期間における企業名と用語のデータを9つの区分に対しコード化し集計を行い、要約グラフを作成した(図7)。

この要約グラフから以下のようなことが考察できる。M&Aや業界再編の解説や業界再編における国内大手製薬メーカーの動向の解説に関する記事は3期間を通じて常に一定以上の割合で存在する。また、2000年代前期には、海外製薬メーカー同士のM&Aが多くみられ、2000年代中期には国内製薬メーカー同士のM&Aが多くみら

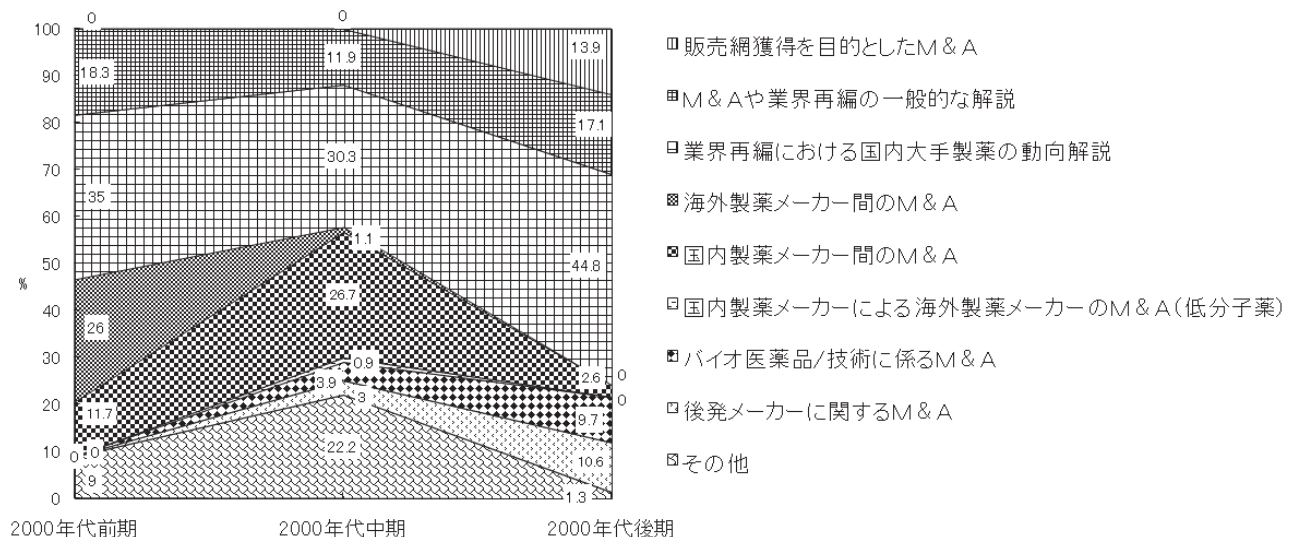


図7 期間別要約グラフ



れ、2000年代中期から2000年代後期にかけてバイオ医薬品/技術関連と後発薬メーカーに関するM&Aが増加する傾向がみられる。2000年代後期には、販売網/販路拡大を目的としたM&Aに関するものが特異的に多くみることができる。尚、2000年代中期で「その他」が多くなっているが、これはカネボウ再建のトピックスに関連した企業名や用語の発現が多かったためである。因みに、これは先の図3で一時的に増加していた用語の発現傾向と一致していた。

以上から、オープンソースBの分析から製薬業界のM&Aに関するもので業界再編の解説のような常に新聞記事として掲載されるものと、その時々で変化する業界の潮流や企業の戦略が、M&Aや提携のような実際の動きとなり記事となりそれが記事中の用語に反映され、テキスト・マイニングの手法を用いることで定量的に視覚的に確認することが可能であることを確認した。次節以降では、このオープンソースBの分析結果を他の情報、オープンソースAと比較し、その関係と内容の合致性を確認する。

### 3.3 特集記事の要約

前節までオープンソースBによる分析結果を検証する目的で、2010年問題に関する日経ビジネス特集記事と日経マネーの掲載記事をもうひとつのテキスト・データ(オープンソースA)として用意した。この2つの記事は、現在の2010年問題とその経緯を幾つかの企業の事例と共に解説しており、これらの記事から事例をもとに製薬業界再編の概要と経緯を時系列に知ることができる。

本稿では、これら特集記事の内容を要約し図8を作成した。この要約図では、製薬業界の変遷に関する大きな流れをトピックスとして枠で囲み、それらトピックスが時系列で関係するものを矢印でつないだ。さらに、それぞれのトピックスに関連して記事中に登場した企業名をそのトピックスの枠の近くに配置した。これにより、その当時のトピックとそのトピックの変遷とそれに係わる企業を一度に確認することができる。

図8の要約図から、1990年代は海外製薬メーカーによる分子量が小さく主に化学合成によって製造される低分子医薬品開発の最盛期であることがわかる。2000年前後には新技術によるバイオ医薬品やワクチンを追求した企業と低分子医薬品による規模拡大を目指した企業への分化があり、同時に新興国の後発薬メーカーの台頭をみることができる。2000年代中期は、日本の新薬メーカーは低分子医薬品での年間1,000億円以上を売上げるブロックバスターと呼ばれる大型製品の開発を追求し、すでにブロックバスター戦略で成功していた海外の大手製薬はその売上で時間稼ぎをしながら、次の成長機会としてバイオ医薬品が成長を窺っている。2000年代後期では、各国政府による推進が後押しとなって後発薬の普及が目に見えた形になると、新薬メーカー大手は一転し、後発薬市場への参入、またはバイオ医薬品メーカーとの提携や買収といった動きを活発化させている。つまり2000年代後半には新薬メーカー大手は、新薬開発の上流にある新技術の獲得とその下流にある後発薬の市場、成長著しい新興国への販売網の獲得という動きを示しているということがわかる。

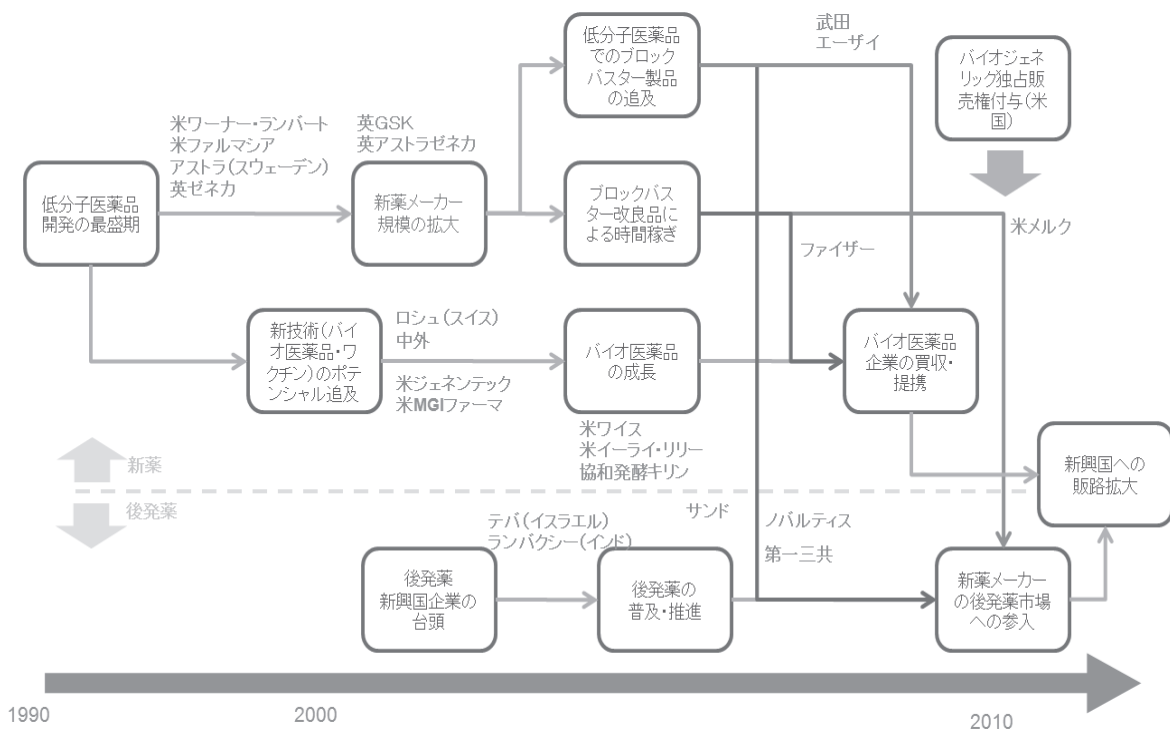


図8 オープンソースAの要約

### 3.4 テキスト・マイニングの結果と特集記事要約の比較

本節ではテキスト・マイニングの手法を用いて作成したオープンソースBの要約の確かさを確認するため、特集記事のオープンソースAから作成した要約図と比較を行い、検証を試みる。

オープンソースAのテキスト・データの内容は、過去数年間の幾つかの企業の動向を解説したものであり、これを図として要約することで製薬業界自体の変遷を時系列に概要としてみる事ができた。

オープンソースBを用いた共起ネットワーク分析では、オープンソースBに発現した用語や企業名の関係を視覚的にみることができた。しかし、この要約からなにか意味するものを解釈することは容易ではない。また3期間分のデータを階層的クラスタ分析し、それぞれの期間の分析結果を要約し、おしなべて見ることで業界の変遷の概略をみることができたものの、しかしここでも業界の知識や薬剤開発およびM&Aなどが起きた背景の情報なしでは、その解釈もまた容易ではないと考えられた。

反対にオープンソースAは、製薬業界再編の経緯を解説する内容のものであり、これを図として要約することで変遷の概略を明示的に時系列にみることができたが、オープンソースAはもともと特集記事や専門家が業界に詳しい読み手向けにある程度内容をまとめたものである。これは特集記事の性格によるものであるが、本稿で要約図として作成する前の段階で、つまり特集記事の作成段階でなんらかの幾つかの情報がすでに省略されている可能性が考えられる。

このような考察を踏まえて、オープンソースAの要約を作成した後に再度オープンソースBの分析結果を見直した。すると、オープンソースAから得た情報が業界の概要を説明する背景情報となり、オープンソースBの分析結果の解釈を助ける働きをしていることがわかった。たとえばAの要約図からは、2000年後期では幾つかの製薬企業のM&A戦略が国外や自社以外の販売網や販路の拡大のために行われているということが読み取れるが、オープンソースBのクラスタ分析の同時期の分析結果においても「販売網/拠点」として幾つかの用語と企業名が発現しており、共起ネットワーク図においても、「販

売網」がハブとなりその周辺に関係する用語と企業が名を連ねている。これは2000年代中期、2000年代前期にはみられない関係であり、時間的にも内容的にも2つの分析結果が合致していた。

また反対にオープンソースBがオープンソースAを補足する関係にあることもわかった。オープンソースBの要約では、2000年代前期に海外の製薬企業間のM&Aに関する話題が多く発現している。これに対しオープンソースAでは1990年代の低分子医薬品の最盛期から2000年前後の規模拡大を目指した業界の再編を示唆する文脈となっており、実際にオープンソースAでは登場しない海外企業同士のM&Aを示唆するデータをオープンソースBの共起ネットワーク図にみることができ、このような関係からこの2つのオープンソースの要約内容が関連しており、オープンソースAでは記述されなかった情報をオープンソースBの要約でより詳細に見ることができ、該当時期の業界の動きをより理解するのに役立っている。

以上から特集記事などの要約と新聞記事をテキスト・マイニングの手法を用いた分析結果の内容が一致していることが確認できただけでなく、これらが互いに補足し合い、これら2つの結果を突き合わせることで分析対象とする話題をより深く理解するのに役立つことが確認できた。

## 4. 結び

これまでの分析の手順を図9としてまとめた。本稿では製薬業界の2010年問題におけるM&Aに関する分析を事例に、時間情報をもつテキスト・データとして主要3紙の新聞記事と経済誌などで特集としてまとめられた記事の2つを用意した。まず新聞記事に対しテキスト・マイニングの手法を用いて分析を行い、この分析結果を経済誌などで特集としてまとめられた記事の要約と比較し内容の合致性を検証した。検証ではこれら2つの結果の一致を確認し、加えて特集記事の要約とテキスト・マイニング手法を用いた分析結果の相互補完的な関係性を確認した。経済誌の特集記事などの要約を、テキスト・

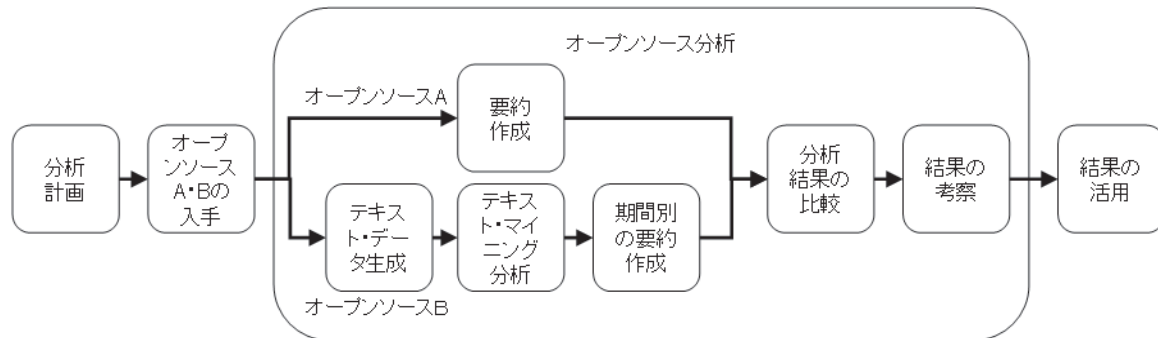


図9 オープンソース分析の流れ

マイニング手法を用いた新聞記事の分析結果を解釈する際に背景情報として用いることで新聞記事の分析結果の解釈が容易にできることがわかった。逆に新聞記事のテキスト・マイニング手法による分析結果が、特集記事の要約を補足する関係にあることで分析対象をより理解するのに役立つということも確認できた。

分析を行う手順は図8に示した通りであるが、その過程で注意すべき点もあった。今回はテキスト・マイニングの手法としてテキスト・データからの企業名を含む単語の抽出から分析対象とする用語の特定、次にこれら用語を用いて共起ネットワーク図の作成と階層的クラスタ分析および要約のための発現頻度集計をグラフとして作成し、もともと大量のテキスト・データを要約すると同時に用語同士の関係性を可視化する目的でこのような分析をおこなったが、これら分析の過程で分析対象とする各期間における特徴的な用語の選択では、客観性の確保と再現性を担保するために今回はカイ二乗検定をおこないその値を基準に用いた。これらにより選択した特徴的な用語に対し共起ネットワーク分析をおこない、また企業名に幾つかの特徴語を組み入れて分析をおこなったがいずれの場合にもそれらの関係性を視覚的に確認することができ、その関係もテキスト・マイニング手法による新聞記事の要約を支持するものであったが、一方で、選択した特徴的な用語すべてと企業名すべてを入れておこなった階層的クラスタ分析では、用語と企業名が同じクラスターに入り、そのクラスターが意味するものを容易に解釈することができるものがあつたものの、これとは別に企業名のみで構成されたクラスターもみられた。この企業名のみで構成されたクラスターの意味するものを解釈するには、経済誌の特集記事などの要約から背景情報を得ることで解釈可能であったが、今後さまざまな分析をおこなうには、その分析の性格に応じて用語の抽出方法の工夫が必要かもしれない。これは今後の課題である。

以上から、新聞記事などのオープンソースによる時間情報もつたテキスト・データを用いてインテリジェンスとする分析は総じて可能であり、今回はこれを特集記事などから得ることができる別の情報の要約と比較し、その内容が一致していたことでこの手法の確かさを確認した。さらに、この手法を用いることで2つの分析結果が互いに補足し合い、分析対象の話題をより深く理解するのに役立つという効果があることもわかった。

本稿では、製薬業界の2010年問題の主要3紙の新聞記事10年分120カ月のデータを40ヶ月分の3つに分割し、それぞれの期間を分析対象としたが、同様な手法でこれよりも短い期間あるいは長い期間の分析を行うことも可能であり、同じデータを用いた場合でも期間を変えて詳細あるいは概略の分析結果を得ることが期待できる。また、本稿で提案した時間情報をもつテキ

スト・データの分析手法はその手法のひとつとして他の業界や話題の分析にも利用が期待できる。

## 謝辞

本論文の作成にあたっては、匿名の2名の査読者から大変有益かつ貴重なコメントを頂戴した。記して深謝申し上げる。なお、本研究の一部は、文部科学省グローバルCOEプログラム「環境共生・安全システムデザインの先導拠点」の援助により行われた。この場を借りて謝意を表す。

## 参考文献

- [1] US Department of Army, Headquarters, Open Source Intelligence, the Federation of American Scientists Website (<http://www.fas.org/irp/doddir/army/fmi2-22-9.pdf>) (2011年7月14日アクセス)
- [2] Mark. M. Lowenthal, Intelligence: From Secrets to Policy, Fourth Edition, Washington, DC; CQ Press (2008)
- [3] 保井俊之, 金融インテリジェンスの誕生と発展: 9-11 テロとグローバルな金融危機を触媒に, インテリジェンス・マネジメント, 1(1), 23-34(2009)
- [4] John J. McGonagle and Carolyn M. Vella, Protecting Your Company Against Competitive Intelligence, West Point, Connecticut, QUORUM BOOKS (1998)
- [5] 坪井聖太郎・萩原清子, テキストマイニングによる自由回答の構造分析: 京都鴨川流域懇談会を事例として, 環境情報科学論集, 22, 327-338(2008)
- [6] 高橋義仁, 製薬とインテリジェンス: CI能力は製薬企業の研究開発に影響を与えるか, 研究技術計画, 23(1), 36-42(2008)
- [7] 高橋義仁, 主要医薬品企業における研究開発能力と研究開発特性: 情報収集・分析・活用能力の重要性(技術経営②, 一般講演, 第22回年次学術大会), 年次学術大会講演要旨集, 22, 86-89(2007)
- [8] 高橋義仁, 下村 博史, 製薬企業とバイオベンチャーの戦略的提携-R&D モジュール化の視点から, ベンチャーズ・レビュー, 5, 81-88(2004)
- [9] 富田 準二, ビジネスインテリジェンスをめぐる展望: 意思決定を支援するテキスト集約技術(ビジネスインテリジェンス及び一般), オフィスインフォメーションシステム, 103(707), 51-58(2004)
- [10] 櫻井茂明, 植野研, 酢山明弘, 折原良平, 「時系列イベントパターンマイニングにおける時間制約の導入」, DEWS, 6C(1), (2005).
- [11] 北朝鮮ミサイル基地の衛星写真・・・「発射可能」の分析も, YOMIURI ONLINE, (<http://www.yomiuri.co.jp/world/news/20090605-OYT1T00395.htm>) (2011年7月18日アクセス)
- [12] 北岡元, インテリジェンス入門, 慶應義塾大学出版会

(2003)

- [13] 特集 武田も揺るがす「2010年問題」大激変, 日経ビジネス 2010/07/05号, 94-108(2010)
- [14] 第1特集 第一三共V.S. エーザイ 収益ダウンに向けた取り組みの行方は? 正念場を迎える 2010年問題, 日経マネー 2010/10号, 56-57(2010)
- [15] 樋口耕一, テキスト型データの計量的分析-2つのアプローチの峻別と統合, 理論と方法, 19(1), 101-115(2004)
- [16] B.Gilad, Early Warning: Using Competitive Intelligence to Anticipate Market Shifts, Control Risk, and Create Powerful Strategies, Amacom Books (2003)
- [17] 厚生労働省, 新医薬品産業ビジョン(2007)