

概念検索を用いた事業継承支援システムの提案

安藤 大岳^{1,a)} 大場 みち子^{2,b)}

概要：近年，中小企業における経営者の高齢化により技術・経験の伝承が課題となっている．この課題を解決するために，本研究では中小企業における事業継承の支援を目的とする．目的を達成するために，企業内保有情報を活用した新たな気づきの発見を支援するシステムを提案する．提案システムは，実在する企業をモデルケースに企業内保有情報として経営者が運用する Web ブログおよび企業が保有する特許・実用新案を利用する．これらをクラスタリングと特徴語を利用して類似の文書を検索する概念検索により提示する．今後は，提案システムの実装と評価により事業継承の促進が可能であるかを検証する．

キーワード：知識伝承，テキストマイニング，概念検索，経営者ブログ，特許・実用新案

Proposal of A Business Succession Support System Using Concept Search

ANDO DAIGAKU^{1,a)} OBA MICHIKO^{2,b)}

Abstract: Recently, problems of experience and knowledge succession among small business managers must be serious, because they are aging. To solve this problem, we aim to propose a business succession support system. In order to accomplish this purpose, our purpose are to propose a system to support new findings by using information held within the company. The proposed system is based on real company as a model case. This system use manager web blog and patent/utility model as information held within the company. Information is shown by concept search that search similar documents by using clustering and feature words. Moreover, the study shows effectiveness of the proposed method by a experiment system.

Keywords: knowledge succession, text mining, concept search, manager web blog, patent/utility model

1. はじめに

近年，団塊の世代の大量退職による退職した職員の技術・経験の喪失が大きな課題となっている [1]．特に中小企業では，高齢化などの背景から経営者から後継者への事業の円滑な継承が課題となっている [2]．中小企業の事業継承においては，その多くが親族内で行われている．親族内での事業継承において，経営者はトップダウン型のワンマンなリーダーシップを発揮していることが多いことが知られ

ている [2]．以上を踏まえ，事業の継承を技術・経験の伝承という観点から見ると，その大部分が経営者に属人化していると考えられる．実際，中小企業自身が事業継承において認識している課題として，経営の能力などを含む後継者の育成が第一に挙げられている [2] ことから，経営者の持つ思想や技術をどのように後継者へ伝承するかが課題であるといえる．

野中らはナレッジマネジメントにおいて，組織における知識創造のプロセスである SECI モデルを提唱している [3][4]．知識には，明確に言語などで表現される「形式知」とはっきりと明示されていないメンタル・モデルなどの「暗黙知」が存在する．両者は，人間の創造的活動において相互に変換される．組織の知は，互いに違う知識を持つ個人が相互に作用し合うことで創られる．以上の前提に

¹ 公立はこだて未来大学 大学院
Graduate School of Future University Hakodate

² 公立はこだて未来大学
Future University Hakodate

a) g2115002@fun.ac.jp

b) michiko@fun.ac.jp

基づき、SECIモデルでは知識創造を、表出化、連結化、内面化、共同化という4つのプロセスに区分し、その相互作用により知識創造が行われるとしている。中小企業においては、先述のとおり属人的なノウハウが暗黙知として存在しており、その一部は組織の知的財産として表出していると考えられる。しかし、経営者が暗黙的に持つ思想を形式知として表出させ、既に形式知として存在する知的財産と組み合わせて新たな知識を創造する連結化のための仕組みは不十分である。

本研究の目的は、中小企業における事業継承の支援とする。この目的を達成するために、企業内保有情報を活用した情報の連結化を支援するシステムを提案する。

本論文では、2章で暗黙知の表出化および形式知の連結化に関する関連研究から課題を見出す。3章および4章でこれらの課題を解決するためのアプローチとその実現方法を提示する。5章では、アプローチの妥当性を検証するための予備実験の結果と考察をまとめている。6章では、アプローチに基づく実験システムと評価手法について説明する。7章では、結論と今後の課題を説明する。

2. 関連研究と研究課題

知識の連結化を促すためには、そのもととなる形式知を効果的に提示する必要がある。事業継承における支援として、経営者の持つ暗黙知を形式知として表現するための手法および表現の適切な情報の提示手法を検討しなければならない。これらに対して、以下の関連研究がある。

2.1 暗黙知の伝承支援

大野らは、東京消防庁における消防経験の伝承を支援するSNSを提案している[1]。この提案手法では、退職間際のベテラン職員が後輩に残すべき事柄を書き記した「心の伝承」に着目している。「心の伝承」を分析することで、伝承対象とする消防活動経験を「状況」「判断」「行動」から成り立つものと定義している。それぞれの要素について庁内資料から、「状況」には消防活動報告、「判断」には消防活動基準を採用している。「活動」については「心の伝承」が電子データではない点を加味して、SNSにより消防活動体験を職員が投稿できる仕組みを新設している。これらの要素をSNSの日記機能を利用し、職員が投稿した体験などと消防活動報告や消防活動基準と結びつけて出力することで、形式知の連結化を促すことが可能となっている。

本研究においては、組織内に存在する資料を活用し、それらを結びつける点においては有用であると考えられる。しかし、本研究では経営者の持つ暗黙知を伝承対象としている点や情報の検索性を鑑みて、SNSはシステム形態として適切ではないと考えられる。活用する資料についても、事業継承の観点から適切に修正しなければならない。どのような資料を活用し、どのようなシステムが適切か検討し



図1 新書マップ

Fig. 1 Shinsho Map

なければならない。

2.2 情報の連結化を促す検索手法

高野らは、次世代情報検索技術として汎用連想検索エンジンGETA(Generic Engine for Transposable Association)を開発しており、GETAを用いた情報検索サービスがいくつか実用化されている[5]。GETAでは、検索キーとして与えられた文書群から、最初に意味的に重要そうな単語セットを検索し、それらを検索要求にして文書検索を行うという連想検索を連想計算により可能にしている。具体的には、文書群と文書群に出現する全ての単語の文書ごとの頻度を用いた行列を計算対象とする。文書から単語を連想するような検索では、与えられた文書における全ての単語に対して計算を行う。単語に対して、それぞれの文書中での頻度に基づいて連想スコアと呼ばれる値を計算・集計し、値の大きい上位数個の特徴語群を出力する。単語から文書を連想するような検索では、これらの特徴語群を用いて、文書方向へ連想検索を行い関連文書を得る。

新書マップ[6]は、高野らがGETAを用いた連想計算を活用することで連想検索機能を構築したサービスの一つである。7000冊の新書・選書が1000のテーマに分類され、テーマごとに本の一覧と読書案内を読むことができる。図1は、新書マップの検索インターフェースである。画面左上の検索フォームに文章やキーワードを与えて検索すると、10のテーマとテーマに関連する12の関連語が星座図のように配置される。配置されたいずれかの関連語を選択することで、その語に関連するテーマを再検索することが可能である。新書マップでは、テーマ分類とユーザからの検索ワードを連想検索機能を用いることで、思いがけない新たな発見すなわち情報の連結化を促すような仕組みが構築されていると考えられる。

本研究において、連想検索機能を用いた情報の検索と提示は形式知の連結化を促すうえで有用であると考えられ

れる．しかし，本研究において新書マップのテーマ分類に相当する情報を検討しなければならないため，そのまま適応することはできない．

2.3 研究課題

2.1 節および 2.2 節を踏まえ，本研究では事業継承のために必要な情報とそれらの情報を分類するためのカテゴリールを検討する必要がある．本研究の課題を以下に示す．

- (1) 企業内の保有情報のうち，事業継承においてどのような情報を活用すべきか明らかでない
- (2) 企業内の保有情報を分類・整理するカテゴリールが存在しない

3. アプローチ

3.1 アプローチ方針

2.3 節の研究課題を解決するアプローチを検討するにあたって，本研究ではある工務店 F 社をモデルケースとして情報の提供やヒアリングなどにご協力いただいている．工務店 F 社は，工務店経営のほかにも高機能住宅の研究・開発を行っており，1 章で説明したような課題を抱えている．具体的には，現経営者が創業者であり経営者が中心となり技術開発を行ってきたが，高齢化により親族内での事業継承を考えている．しかし，後継者は事業内容を含め現経営者の持つ暗黙的な技術開発などの思想を十分に理解していないため，経営者の持つノウハウの伝承が急務であるというのが F 社の抱える課題である．

本研究では，F 社の経営者ブログおよび F 社が保有する特許・実用新案を活用し，文書間の関連と関連ごとの特徴語から類似の文書を検索する概念検索により情報の連結化を支援する手法とする．

3.2 経営者ブログおよび特許・実用新案の活用

本研究では，1 章で述べた背景を踏まえ，事業継承に必要な情報要素を経営者の持つ知識と企業が保有する技術情報の 2 種類と定義した．これは，中小企業が事業において開発してきた技術と経営者の持つ知識を紐付けることで，事業に関するノウハウを表現することができると考えたためである．これらの要素を補完する媒体として，経営者の持つ知識として経営者ブログ，企業が保有する技術情報として知的財産の一部である特許および実用新案を採用した．

3.2.1 経営者ブログ

図 2 は経営者ブログの記事の一例である．経営者ブログは，F 社現経営者が掲載・更新する Web ブログであり，2006 年 2 月から 2016 年 1 月時点までで 5400 を越える記事が投稿されている．記事の内容は，大きく経営者の観点からの記事と技術者の観点からの記事の 2 種類に分類することができる．技術的な記事には，当時の開発技術について触れられている内容も含まれていることも確認してい



図 2 経営者ブログ

Fig. 2 Manager Web Blog

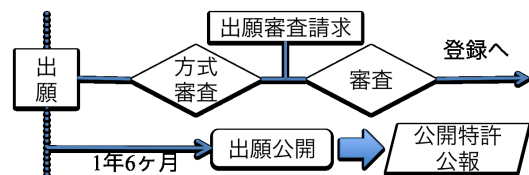


図 3 特許出願の流れ ([7] を参考に作成)

Fig. 3 Flow of The Patent Filing
(Based on [7])

る．このことから，経営者ブログには経営者の様々な知識が断片的に潜在していると考えられる．

3.2.2 知的財産

知的財産とは，知的創造活動によって生み出された財産的価値を有する情報である [8]．本研究においては，知的財産の中でも「営業秘密その他の事業活動に有用な技術上又は営業上の情報」に着目する．知的財産として，F 社が権利を保有する特許・実用新案を本研究における活用の対象とする．図 3 は，特許出願の流れを示している．特許は，出願公開制度により出願から 1 年 6 ヶ月後にその内容が公開特許公報として公開される．実用新案については，特許と異なり審査を必要としない権利であるため，登録から約 1 ヶ月後に登録実用新案掲載公報という形でその内容が公開される．これらの公報に関して，F 社においては 2015 年 4 月の時点で F 社名義もしくは F 社現経営者名義のものは 81 件（うち特許 74 件，実用新案 7 件）であった．本研究では，これらの公報を情報の媒体として利用する．

3.3 概念検索

高野らの研究では，連想計算による連想検索機能を利用して，新書・選書に紐づくテーマを検索ワードから提示し

ていた。本研究では、高野らの研究を参考に文書間の関連をクラスタリングにより推測し、推測したクラスタから特徴語を抽出することで類似の文書を検索して提示する概念検索を提案する。概念検索のアルゴリズムを事前処理と検索時処理に分けて説明する。

- 事前処理

- (1) 経営者ブログ、特許・実用新案をそれぞれ形態素解析し、単語に分解する。
- (2) 文書群のクラスタリングを行う。
- (3) TFIDF を用いて各クラスタごとの特徴語を抽出する。文書頻度は同じクラスタに属する文書数を基準に算出することとする。

- 検索時処理

- (1) 与えられた検索ワードを各クラスタの特徴語と比較し、一番検索ワードに近いクラスタを選択する。
- (2) (1) において選択したクラスタから検索ワードに関連する関連語を抽出する。
- (3) (2) において抽出した関連語の上位数個を用いて各関連語ごとに紐づく文書を検索し、提示する。

4. 実装

3章で述べたアプローチの具体的な実現方法を述べる。

4.1 情報媒体の取得・蓄積

3.2.1 項および 3.2.2 項で述べた情報は全て Web 上で取得可能である。これらの情報を HTML で取得したのち、DOM を利用して必要な情報を取得・蓄積する。経営者ブログ、特許・実用新案はいずれも定型データであるため必要な情報の取得は容易である。具体的な取得情報については、経営者ブログは記事のタイトル、投稿日時をメタデータとして本文を取得する。取得した本文は、2 行以上の改行を挟む段落ごとに話題が変わることがあるため分割して取得することとする。本文の最後には固定リンクや定型文を含む事もあるため、これらを取得時に除去する。特許・実用新案は、本文中に出現する項目が決められており特許に関する書誌情報も本文中に含んでいる。書誌情報をメタデータとして、各項目の本文を繋げて一つの文書として取得する。

4.2 概念検索の実現手法

3.3 節で示した概念検索を実現する手法を順に説明する。

形態素解析には Mecab を使用し、ブログの特性を踏まえ、様々な固有表現を正しく分割するために mecab-ipadic-NEologd[9] という新語辞書を採用する。

クラスタリングには、コサイン距離を用いたワード法による階層的クラスタリングを採用する。コサイン距離の文書表現には、PV-DM(Distributed Memory Model of

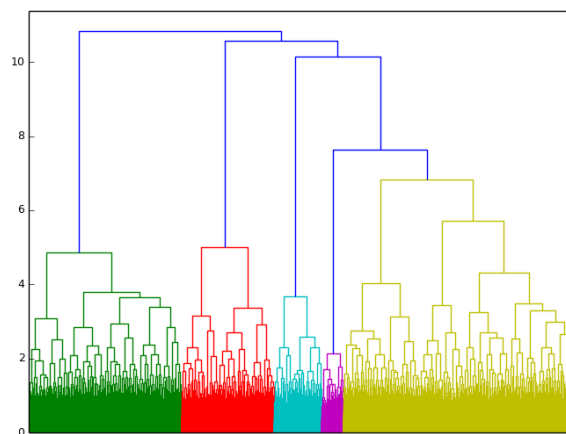


図 4 ブログクラスタリング結果

Fig. 4 Result of Rhe Web Blog Clustering

Paragraph Vector)[10] と呼ばれる教師なし学習の手法を採用する。ブログの解析においては同じ文脈を指す文書をまとめることが重要であると考えられる。しかし、一般に文書表現に用いられる Bag of Words モデルでは単語の順序や語の意味が失われてしまうという欠点がある。この欠点を克服したとされる PV-DM を採用する。

検索ワードと特徴語の比較には、Word2Vec[11] と呼ばれる教師なし学習の手法を採用する。Word2Vec は先述の PV-DM にも利用されている手法であり、同様の理由でこれを採用する。

PV-DM および Word2Vec のアルゴリズムは、Python で使用可能なフリーウェア gensim[12] に実装されているため、これを使用する。

5. 予備実験

5.1 目的

3.3 節および 4.2 節で説明した概念検索のアルゴリズムとその実装の妥当性の検証を目的とする。

5.2 実験データ

経営者ブログの本文を対象とする。全 5450 の各記事を 4.1 節に基づき取得した 24157 の文書を実験データとして扱う。本来は、特許・実用新案もデータとして扱う必要があるが予備実験では対象外とした。

5.3 手順

3.3 節および 4.2 節の手順で事前処理を行い特徴語を抽出する。クラスタは 10 程度に分割される閾値で統合し、それぞれのクラスタに対して TF-IDF 値の高い上位 100 単語を特徴語として抽出する。各クラスタに含まれる文書をサンプリングし、不自然なものがないか検証する。各クラスタから抽出した特徴語に関しても妥当性の検証を行う。

表 1 抽出した特徴語
Table 1 Extracted Feature words

	クラスタ 1	クラスタ 2	クラスタ 3	クラスタ 4	クラスタ 5	クラスタ 6	クラスタ 7	クラスタ 8	クラスタ 9
特徴語 1	の	の	明日	今日	家	さん	の	家	の
特徴語 2	よう	よう	今日	よう	の	社長	人	工務店	さん
特徴語 3	今日	事	東京	の	よう	工務店	事	の	よう
特徴語 4	事	iPhone から送信	移動	写真	熱	(企業用語)	よう	さん	今日
特徴語 5	写真	人	写真	気温	事	写真	自分	事	事
特徴語 6	さん	私	よう	北海道	暖房	今日	もの	よう	写真
特徴語 7	こと	今日	札幌	北斗市	エアコン	訪問	私	家づくり	工務店
特徴語 8	私	さん	の	さ	気温	私	こと	建	家
特徴語 9	さ	こと	北斗市	今朝	湿度	家	さ	主	(企業用語)
特徴語 10	人	達	北海道	雪	性能	左	さん	人	私
文書数	1730	5046	1624	2481	2096	969	3392	1967	4852

5.4 結果

クラスタリングの結果を図 4 に示す。閾値を 4.5 とし、9 クラスタを生成した。以下では生成したクラスタを図 4 の左から順にクラスタ 1~クラスタ 9 として扱う。各クラスタから抽出した特徴語のうち上位 10 語を表 1 に示す。F 社に直接関係する単語については「(企業用語)」と表記し、F 社の特徴を表すのに有用であると考えられる単語をゴシック体で示している。

5.5 考察

表 1 の結果から、特徴語の抽出精度が高くないことがわかる。しかし、図 4 に基づくクラスタごとの文書を確認するとクラスタ 1,2 は世情に関する文書、クラスタ 3 は都市間の移動、クラスタ 4 は移動+その日の気温など一定のトピックに分類されていることが確認できた。この理由として、文書の意味的な近さと語句的な近さの違いが考えられる。すなわち、人間が目で見ても意味が近いと感じる文書でも、使われている語句にばらつきがあるため特徴語として抽出されにくいと考えられる。特徴語の算出方法自体にも問題がある。表 1 から、全てのクラスタにおいて「の」や「よう」など特に意味のなさない単語や「今日」や「さん」など一般的すぎる単語が上位にきてしまっていることがわかる。対策としては、ストップワードの除去処理が考えられる。以上の結果を踏まえ、特徴語の算出方法の改善や代替手法を検討する必要がある。代替手法としては、例えば Word2Vec を活用することでクラスタごとの特徴を単語ではなく単語のベクトルで表す手法が考えられる。単語をベクトルで表現することで、複数の単語のベクトルを足し合わせたベクトルに近い単語ベクトルの推測も可能になる。すなわち、キーワードの AND 検索とは異なる「単語の意味の足し算」が可能になると考えられる。

6. 実験システムと評価方法

3 章で述べたアプローチに基づき実験システムを構築し、

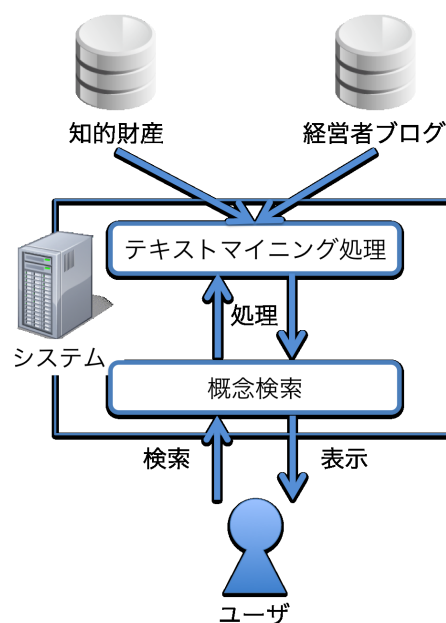


図 5 システム構成図

Fig. 5 System Configuration Map

有用性を評価する予定である。実験システムの概要および評価方法を以下に示す。

6.1 システム概要

実験システムは、経営者ブログと知的財産を活用し概念検索により情報を提示することで情報の連結化を支援する。システム構成図を図 5、システムの UI イメージを図 6 に示す。実験システムは、ユーザから与えられた検索ワードをもとに概念検索を行い、図 6 のような検索結果を返す。具体的には、図 6 にあるように「エアコン」と「効率」を含む検索ワードに対して「エアコン」や「エコキュート」などの関連するキーワード (概念) とキーワードに紐づく情報が提示される。例えば、「エアコン」であればエアコンの仕組みについて言及されたブログ記事やエアコンを効率的に使用するための特許などが提示される。「エコキュート」



図 6 概念検索 UI イメージ

Fig. 6 Concept Search UI Image

や「省エネ」といった関連する概念から給湯器の仕組みやエアコンの効率的な使用方法などの周辺知識を得ることも可能である。「エコキュート」から関連する概念を再検索することで、更に広い範囲で周辺知識を得られると考えられる。関連する概念に基づいて紐づく情報を提示することで、検索ワードには直接関連しない周辺知識も得やすくなると考えられる。

6.2 評価方法

構築したシステムを F 社協力の下、試用していただくことでシステムの有用性を評価する。評価はアンケートやインタビューによる評価とシステムによる概念検索の妥当性の検証を検討している。アンケート、インタビューと併せてどのような手順で検索が行われたかをアクセスログから解析し、情報の検索性の評価に利用する。アンケート、インタビューの項目は以下を検討している。

- システムは事業継承に役立つか
 - － 求める情報にたどり着きやすくなったか
 - － 新たな気づきや発見はどの程度あったか
- システムがあれば使いたいと思うか
- システムは使いやすいか

7. おわりに

中小企業における経営者の高齢化により事業の継承が急務である。事業継承において、経営者の持つ暗黙知をどのように伝承するかが課題となっている。この課題を解決するために、企業内保有情報を活用した情報の連結化を支援するシステムを提案した。具体的には、企業が保有する特許・実用新案および経営者ブログを活用し、文書間の関連と関連ごとの特徴語から類似の文書を検索する概念検索により提示することでこれを実現する。アプローチの妥当性を検証するために予備実験を行った。予備実験の結果、各クラスタごとの特徴語がうまく抽出できていないことが明

らかになった。

今後は、特徴語の算出方法の改善や代替手法を中心にアプローチの改善を行う。その後、実験システムを構築し、実際に企業で試用していただくことで有用性の検証を行っていく。

参考文献

- [1] 大野光太郎ほか：東京消防庁における消防活動経験の伝承を支援する SNS の提案，情報処理学会論文誌，Vol. 54, No. 1, pp. 284–294 (2013).
- [2] 久保田典男：事業承継に際しての組織改革：親族内承継の中小企業におけるケーススタディ，日本経営診断学会全国大会予稿集，Vol. 10, pp. 182–185 (2010).
- [3] 野中郁次郎，竹内弘高：知識創造企業，東洋経済新報社 (1996).
- [4] 野中郁次郎，梅本勝博：知識管理から知識経営へ：ナレッジマネジメントの最新動向，人工知能学会誌，Vol. 16, No. 1, pp. 4–14 (2001).
- [5] 高野明彦，西岡真吾，丹羽芳樹：連想に基づく情報アクセス技術：汎用連想計算エンジン GETA を用いて，情報の科学と技術，Vol. 54, No. 12, pp. 634–639 (2004).
- [6] 高野明彦：新書マップ～テーマで探す新書ガイド～，NPO 法人連想出版（オンライン），入手先 <<http://shinshomap.info/search.php>>（参照 2016-01-28）.
- [7] 特許庁：出願の手続，特許庁（オンライン），入手先 <<https://www.jpo.go.jp/cgi/link.cgi?url=/shiryou/kijun/kijun2/syutugan.tetuzuki.htm>>（参照 2016-07-30）.
- [8] 特許庁：知的財産権について，特許庁（オンライン），入手先 <http://www.jpo.go.jp/seido/s_gaiyou/chizai02.htm>（参照 2016-01-28）.
- [9] NEologd: [neologd/mecab-ipadic-neologd](https://github.com/neologd/mecab-ipadic-neologd), (online), available from <<https://github.com/neologd/mecab-ipadic-neologd>> (accessed 2016-01-28).
- [10] Le, Q. V. and Mikolov, T.: Distributed Representations of Sentences and Documents (2014).
- [11] google.com: Google Code Archive - Long-term storage for Google Code Project Hosting., google.com (online), available from <<https://code.google.com/archive/p/word2vec/>> (accessed 2016-01-29).
- [12] RadimRehurek: gensim: Topic modelling for humans, RadimRehurek.com (online), available from <<http://radimrehurek.com/gensim/index.html>> (accessed 2016-01-31).