

企業内保有情報からの知識抽出に着目した 事業継承支援システムの構築

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・実用新案を用いる. これらの情報を時間軸, 技術的キーワード, 関連するステークホルダーなどの観点から分析することで事業に関する知識を抽出し, 可視化する. 実験により, アプローチの有用性を評価する.

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

Development of A Business Succession Support System Based on Knowledge Extraction from Company Information

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

Abstract: Recently, experience and knowledge succession among small business managers lead to be serious problem, because of their aging. To solve this problem, we aim to support business succession. In order to accomplish this purpose, we 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's web blog and patent/utility model as information held within the company. We will extract knowledge about the business by analyzing information from the viewpoints of time lapes, technical keywords, stake holders, etc. 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, 高齢化などの背景により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 [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」とはつきりと明示されていない「暗黙知」が存在する.

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

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る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、アプローチに基づく実験システムについて説明する。6章では、アプローチの有用性を検証するための実験システムを用いた実験の結果と考察をまとめている。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

本章では、知識・経験の伝承についての関連研究とこれを踏まえた本研究における研究課題について述べる。

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。「行動」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を踏まえ、必要な情報要素を定義し直す必要がある。それらの情報の分析・活用方法も事業継承の観点から適切に修正しなければならない。どのような情報を活用し、それらの情報からどのような知識が得られればよいか検討しなければならない。

2.2 研究課題

2.1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の分析手法を検討する必要がある。本研究の課題を以下に示す。

- (1) 企業内保有情報のうち、どのような情報を活用すべきか明らかでない
- (2) 企業内保有情報から有用な知識を抽出するための分析手法が明らかでない

3. 提案アプローチ

2.2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。

本章では、以上の前提と2.2節の研究課題を踏まえたアプローチについて述べる。

3.1 経営者ブログおよび知的財産の活用

本研究では、1章で述べた背景と2.2節で述べた研究課題を踏まえ、企業内保有情報のうち活用する情報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する。

3.1.1 経営者ブログ

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2015年12月末時点までで5400を越える記事が投稿されている。図1は経営者ブログの記事の一例である。記事の内容は、経営者の観点からの記事と技術者の観点からの記事の大きく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の内容が含まれていることを確認している。経営者ブログは、経営者の様々な知識が経営者自身によって表出化された形式知として扱うことができる。

3.1.2 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



図 1 経営者ブログ

Fig. 1 Manager Web Blog

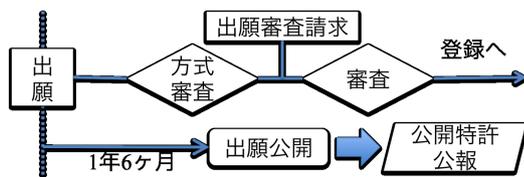


図 2 特許出願の流れ ([5] を参考に作成)

Fig. 2 Flow of The Patent Filing
(Based on [5])

的価値を有する情報である [6]。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利用対象とする。図 2 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、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。

3.2 知識抽出手法

本研究では、暗黙知の伝承の対象が個人であることや企業内の情報を活用することから、知識抽出の軸として検索性と既存知識の組み合わせに着目する。3.1 節で述べたように、経営者の持つ知識と企業の技術情報を紐付けることで事業に関するノウハウが表現できると期待できる。これを実現するため、知識の再発見を促す分析手法と新しい気

付きを促す分析手法という 2 つの側面から分析手法を提案する。知識の再発見については 3.2.1 項および 3.2.2 項、新しい気付きについては 3.2.3 項および 3.2.4 項で述べる。

3.2.1 類似文書の提示

企業内保有情報から形式知を効率よく提示するための方法として、類似文書の提示が考えられる。類似文書を求めるためには文書を定量的表現する必要がある。古典的な文書の定量表現手法として Bag-of-Words(BOW)がある。しかし、BOW 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り、様々な単語が様々な文脈で使われるブログでは適していない。この欠点をクリアするため、本研究では深層学習による類似文書の提示を行う。形式知を効率よく提示するためには、単純に類似文書を提示するだけでは情報の検索性の観点からあまり効果的とはいえない。本研究では、高野らが提唱する連想検索 [7] を参考に、深層学習によりキーワードの類似単語を抽出し、これらの単語をもとに関わりの深い文書を検索する概念検索を提案する。類似文書の提示手法をまとめると、以下のとおりである。

(1) 類似文書検索 文書の主題の類似

(2) 概念検索 単語の意味の類似

3.2.2 特徴語の抽出

企業内保有情報から有用な知識を得るためには、適切なキーワードを知っている必要がある。キーワードの推測は、一般的に特徴語が利用される。抽出・活用することで、知識の再発見に繋げる。

特徴語の抽出に際して、3.1.1 項で述べたように経営者ブログの内容は一通りではないため、以下の 2 つの軸の組み合わせにより情報を分割したうえで抽出する。

(1) 時間軸 (年、月)

(2) 分類による軸

ここで、分類による軸について述べる。分類は、3.2.1 項で述べた類似文書検索を利用して経営者ブログ全体の文書をいくつかのクラスターに分類し、軸として利用する。

特徴語は、TF-IDF を用いて式 (3.1)、式 (3.2)、式 (3.3) の計算式で算出する。ここで、 n_t^d は索引語 t の文書 d における出現回数、 $|D|$ は経営者ブログの全文書、 $|\{d : d \ni t\}|$ は経営者ブログの全文書中索引語 t が出現する文書数とする。なお、索引語 t は表 1 に示す条件を満たす単語とする。

$$tf_t^d = \frac{n_t^d}{\sum_k n_k^d} \quad (3.1)$$

$$idf_d = \log_{10} \left(\frac{|D|}{|\{d : d \ni t\}|} \right) + 1 \quad (3.2)$$

$$tfidf_t^d = tf_t^d \cdot idf_d \quad (3.3)$$

表 1 索引語条件

Table 1 Conditions of Index Words

1	名詞であり、数、特殊、接尾、非自立、代名詞、副詞可能、形容動詞語幹、ナイ形容詞語幹のいずれかでない
2	英数字でない
3	一般的な単語 [8] でない

3.2.3 時間軸に着目した分析

新しい気付きを促すための方法として、時間軸に着目した分析が考えられる。具体的には、時期ごとによる単語の使用頻度の推移や企業内保有情報のタイムラインによる提示などが挙げられる。

時間軸に着目した分析は、これまで述べたアプローチを利用する。具体的には、企業内保有情報の持つ投稿・申請日時や抽出した特徴語の頻度情報を利用して情報の提示を行う。

3.2.4 ステークホルダーに着目した分析

工務店 F 社へのヒアリングにより、ステークホルダーとステークホルダーに関連する技術の分析が求められていることが明らかになった。本研究では、経営者ブログから取得した人名とその所属からステークホルダーを抽出し、関連する技術情報を知的財産と紐付けることでこれを実現する。

ステークホルダーの抽出は係り受けの分析により以下の手順で行う。

- (1) 人名+接尾語を含む文章を抽出する
- (2) 1 を文節に分け、人名の直前の「の」が機能語である文節を繋げ所属とする

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

3.1.1 項および 3.1.2 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。

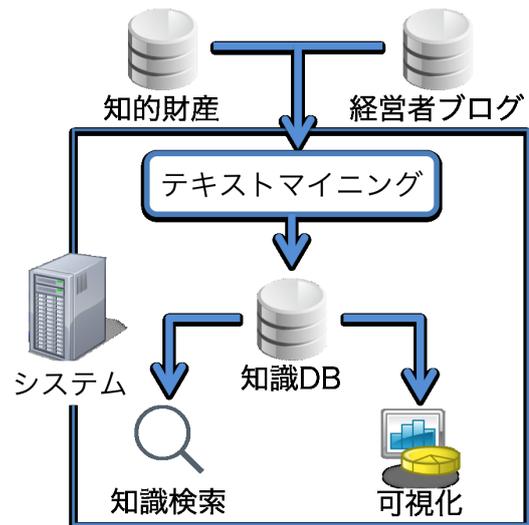


図 3 システム構成図

Fig. 3 System Configuration Map

4.2 知識抽出の実現手法

3.2 節で述べた知識抽出を実現する手法を順に説明する。形態素解析には Mecab[9] 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[10] という新語辞書を採用する。係り受けの分析には、以上に加えて CaboCha[11] を採用する。

概念検索には Word2Vec[12] を、類似文書検索には Doc2Vec[13] を採用する。これらは、Python で使用可能なフリーウェア gensim[14] に実装されているため、これを使用する。

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類似文書検索と同様に Doc2Vec を利用する。

5. 実験システム

本章では、3章で述べたアプローチを踏まえた実験システムについて述べる。

5.1 システム概要

システム構成図を図 3 に示す。実験システムは、企業内保有情報を 3 章のアプローチに従い処理し、データベースに格納しており、ユーザの入力に応じて情報の検索と可視化を行うシステムとなっている。図 4 に実験システムの画面例を示す。図 4 は概念検索の実際の使用例を表している。3.2.1 項で述べたように「エアコン」の類似語として「蓄熱」「輻射熱」「暖房」「暖房機」という単語が抽出され、それぞれの単語に関わりの深いブログエントリーが一項目最大 10 記事表示されるものとなっている。実験システムでは、ブログエントリーの表示とともにその記事の類似記事やキーワードを表示することで周辺知識の網羅的な検索を可能としている。

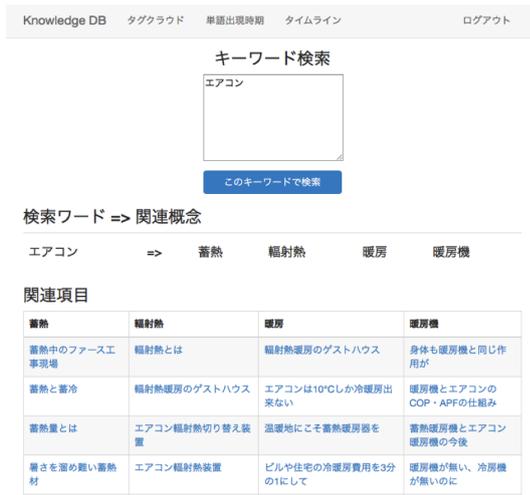


図 4 システム画面例

Fig. 4 System Screen Example



図 5 タグクラウド

Fig. 5 Tag Cloud



図 6 単語出現頻度グラフ

Fig. 6 Word Appearance Frequency Graph

5.2 情報の可視化

事業継承という特性を踏まえ、システムは使用者が直感的に操作でき、感覚的に重要な情報を把握できるような手法を考慮する必要がある。実験システムの構築に際し、これらの点を踏まえ、3.2.2 項で述べた特徴語の提示には図 6 に示すようなタグクラウドを採用し、3.2.3 項で述べた頻度情報はグラフ表示を採用した。タグクラウドでは、3.2.2 項でのアルゴリズムを用いて求めた TF-IDF を指標にこの

値が大きい単語ほど大きく表示されるという方式を採用している。グラフでは、表示される値の最大値を基準に軸を調節することで頻度の変位をわかりやすくしている。

6. 実験

本章では、3 章で述べたアプローチの有用性を検証するために行った実験と結果について述べる。

6.1 実験概要

実験は、F 社専務および研究開発室長の 2 名に被験者をお願いした。実験手順は以下のとおりである。

- (1) 1 回 10 分程度システムを利用する
- (2) システム試用時の新たな気づきを記録する
- (3) 3 回程度試用し、アンケートを実施する

評価は、アンケートやアクセスログを用いてアプローチの有用性やシステムの使いやすさを評価する。

6.2 実験結果

アンケートの結果から、情報の検索性に関して、過去の情報を瞬時に検索できる、どんな情報を発信してどのようなことが伝えたいことなのかかわかるという評価が得られた。システムの使いやすさに関して、感覚的に操作できた、グラフでの表示がわかりやすい、使った文字の頻度が大きさを表現されているのが面白いという評価が得られた。一方で、書いてあるはずの情報が検索されない、概念検索で元のキーワードの検索結果が表示されないなどの改善点も明らかとなった。

記録された気づきとアクセスログから、被験者は何らかの意図を持ってシステムによる検索を行い、気づきを得ていることが明らかになった。例えば、東日本大震災以降での露出単語の変化を「原発」を始めとした複数の単語で調べ、気づきを得るといった使い方がなされていた。事業の根幹となるキーワードを用いて、過去から発信されている経営者の伝えたい情報の根本的な思想を確認するという使い方もなされていた。一方で、被験者の予想しない新たな気づきは得られていない。

6.3 考察

6.2 節の実験結果から、アプローチの有用性とシステムの使いやすさについてそれぞれの考察を述べる。

アプローチの有用性に関して、提案アプローチは既知の知識を再発見する際には十分有用であると言える。新たな気づきに関して、現状では全く新しい気づきに繋がる情報の提示には至っていない。しかし、複数の情報から新たな事実を推測するという使い方は可能であるといえる。被験者からのシステムへのコメントにより、提案アプローチが

事業継承以外の側面として利用できる可能性も示唆されている

システムの使いやすさについて、企業の一般的な社員が直感的に使用できることが示唆された。可視化に関して、タグクラウドやグラフを用いることで効果的な情報提示に繋がっているといえる。

7. おわりに

中小企業における事業継承を支援するため、企業内保有情報を活用した分析手法とその手法に基づくシステムを提案した。提案手法の有用性を検証するため、協力企業の社員を被験者とした実験を行い評価した。実験の結果、提案手法は事業継承のための一助となっはいるものの、現状では十分な情報提示を実現できていない。今後は、言語処理の精度向上や分析手法の追加、様々な中小企業への適応方法を検討することでシステムの実用性を高める。

参考文献

- [1] 大野光太郎, 小川祐樹, 諏訪博彦, 太田敏澄ほか: 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] 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=/shiryoku/kijun/kijun2/syutugan.tetuzuki.htm>> (参照 2016-07-30).
- [6] 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).
- [7] 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).
- [8] 大島裕明, 中村聡史, 田中克己: SlothLib Web 検索研究のためのプログラミングライブラリ, 日本データベース学会 letters, Vol. 6, No. 1, pp. 113-116 (2007).
- [9] 工藤拓, 山本薫, 松本裕治ほか: Conditional Random Fields を用いた日本語形態素解析, 情報学 NL 研報, pp. 161-173 (2004).
- [10] Toshinori, S.: Neologism dictionary based on the language resources on the Web for Mecab (2015).
- [11] 工藤拓, 松本裕治ほか: チャンキングの段階適用による日本語係り受け解析, 情報処理学会論文誌, Vol. 43, No. 6, pp. 1834-1842 (2002).
- [12] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. and Dean, J.: Distributed representations of words and phrases and their compositionality, *Advances in neural information processing systems*, pp. 3111-3119 (2013).
- [13] Le, Q. and Mikolov, T.: Distributed Representations of Sentences and Documents, *Proceedings of The 31st International Conference on Machine Learning*, pp. 1188-1196 (2014).
- [14] RadimRehurek: gensim: Topic modelling for hu-

mans, RadimRehurek.com (online), available from <<http://radimrehurek.com/gensim/index.html>> (accessed 2016-01-31).