

インターネット情報源からの栄養ロスフリーな調理法の抽出

竹村 望[†] 佐藤 哲司^{††}

[†] 筑波大学情報学群 〒305-8550 茨城県つくば市春日 1-2

^{††} 筑波大学図書館情報メディア系 〒305-8550 茨城県つくば市春日 1-2

E-mail: †{takemura20,satoh}@ce.slis.tsukuba.ac.jp

あらまし 厚生労働省では、1日の野菜の目標摂取量は350g以上と推奨しており、健康促進のために、栄養に気を付けた食生活を送ることは大切だと考えられている。野菜をただ摂取するのではなく、栄養価を変化させる調理法に着目し、栄養ロスフリーな調理法の支援が必要であると考え、野菜に含まれる栄養価を生かす調理法をインターネット上から抽出する。本研究では、野菜に含まれる栄養価を生かす調理法を発見することを目的として、インターネット上のウェブサイトから抽出する手法を提案する。野菜名を検索語に加えたウェブ検索で収集したウェブサイトの構造を解析や共起語に着目し、栄養ロスフリーに言及しているエレメントを抽出する。抽出したエレメントを野菜名、効果の2つにラベル付けし栄養ロスフリー辞書を作成する。4種類の野菜を対象に実験を行い、適合率、再現率、F値の算出と、2冊の書籍に記載されている栄養ロスフリーな調理法の件数比較を評価した。前者は1種類を除き栄養ロスフリーな調理法を抽出することができたが、適合率、F値は期待する数値は得られなかった。再現率は6割を超えることができた。後者では2種類が書籍以上の調理法を抽出することができ、書籍に記載されていない1種類は新しい調理法を抽出することができた。前者の抽出精度を向上させることで、この提案手法はインターネット情報源からの栄養ロスフリーな調理法の抽出をするのに有効であると言える。

キーワード 栄養ロスフリー、情報抽出、野菜摂取、栄養価、調理変化

1 はじめに

厚生労働省では、1日の野菜の目標摂取量は350g以上と推奨しており、健康促進のために、栄養に気を付けた食生活を送ることは大切だと考えられている。野菜摂取量350gの内訳は、120g以上の緑黄色野菜を摂取することが目標とされているため、残りの230gは淡色野菜で摂取することになる。日々の生活の中で様々な工夫を凝らし野菜の摂取をすとしても、上記のことを踏まえると緑黄色野菜と淡色野菜のバランスも考え、目標である350gを摂取することは難しい。平成30年度の厚生労働省の調査では、1日の野菜摂取量は男性が290.9g、女性が273.3g、平均値281.4gとなっており、約70gも少ない。ここ10年間の結果を見ても有意な増減は見られない[1]。私たちは野菜を食べているつもりが、実際は目標摂取量には達していない場合が多く存在していることが考えられる。また、平成30年国民健康・栄養調査でも、成人の野菜摂取目標量の350gに対して20~40歳代では約7割しかとれていない。事実、油脂類の消費が増加する一方で、野菜類の消費量は伸び悩み、野菜を食べない傾向が顕著になっている[2]。

さらに、私たちの体内環境は生活環境を取り巻く様々な要因により悪化し、免疫機能が低下すると言われている。中でも、栄養不足は免疫機能の低下を引き起こすことが知られており、食生活が免疫機能に与える影響が大きい。栄養価の不足がない食生活を送ることが免疫機能を正常に保つうえで重要である[3]。野菜には体に欠かせない栄養価が含まれている。それらには、

ビタミン類・鉄・カルシウム・カリウム・マグネシウム・食物繊維など、体調を整え、生活習慣病を予防する栄養価が豊富であり、野菜が栄養価の一番の供給源でもある。

以上のことより、野菜が健康促進に良いことを知っていても、意識しなければ目標摂取量の野菜を食べることは難しい。調理法によっては、多様な栄養を摂取できるという野菜本来の機能を調理の段階で損失してしまう。調理をすることで増減する栄養価があることを知ることは、野菜の栄養をよりよく摂取するために必要である。

一方、調理法によっては栄養価が増加する場合もある。間違った調理法で野菜の栄養価を捨てており、その結果栄養豊富な野菜を食べても栄養価を体が吸収していないのである。野菜をただ摂取するのではなく、調理法を工夫し同量の野菜から栄養価をよりよく摂取することが健康促進につながる。したがって、野菜の栄養を逃さない調理法を知ることで、野菜ごと調理を工夫し、野菜の栄養価を逃さず摂取できるのだ。これにより、野菜を食べることで栄養価を身体に取り込み健康促進へと繋げることができる。

栄養価を変化させる調理法に着目し、栄養ロスフリーな調理法の支援が必要であると考え、また、栄養ロスフリーな調理法が書かれた書籍もある。しかし書籍の情報は性質上、情報の劣化が避けられない。さらに、書籍に記載されている野菜は、一般的に普及している野菜のみ記載されており、栄養ロスフリーな調理法が得られる野菜の種類も限られてしまうという点が挙げられる。インターネット上の情報は、日々アップデートされ、多くの情報があるため栄養ロスフリーな調理法が得られ

る野菜の種類も書籍に比べると多いと考えられる。本研究では、野菜に含まれる栄養価を生かす調理法をインターネット上から抽出する。

2 関連研究

2.1 栄養に着目したレシピに関する研究

食の栄養バランスに着目し、レシピから栄養価に注目する研究は数多く行われている。

角ら [5] は、Web サイトのレシピへ栄養価情報を付加することで栄養価の値による検索を可能とするシステムを構築している。Web サイトの HTML の必要なレシピ情報が記載されている箇所を XPath で取得し、XML 文書に変換を行い栄養価情報を付加した。結果、XML 文書を編集することで栄養価情報を正確に付加することができ、非常に高い正答率を上げることができる。

菊米ら [6] は栄養素摂取を意識した献立作成に焦点を当て、単品のレシピでなく複数のレシピを組み合わせることで、栄養素のバランスやメニューの偏りを防ぐ検索システムを提案した。食材を6つの食品群に分類し、食品群ごと摂取量に基づいて好ましい料理の組み合わせを生成している。栄養素の充足率を図るために、食品群辞書とグラム変換辞書を作成し、計算精度を高めている。結果、実験によって栄養バランスを考慮した献立を提案することができ、有効であることがわかった。

岩上ら [7] は、ユーザーが入力した健康を意識する文章を元に、栄養素の専門知識をもたないユーザーに合った効果的で栄養素豊富なレシピ推薦システムを提案している。自然な文章を形態素に分解し、目的となる語を抽出後、その単語と共起回数の多い栄養素が記載されている共起辞書を用い、レシピ推薦を行っている。共起辞書は、栄養素名を検索エンジンで検索し、上位 100 件の Web ページにおいて、栄養素名と目的の単語が同時に出現する頻度で構築した。結果、自然な文章から正しい栄養素を高い数値で推薦でき、またレシピ推薦の F 値も平均で 0.65 を達成している。

植田ら [8] は、WEB 上の料理レシピから、調理加工による栄養素変化の自動計算システムについて提案している。調理手順を重視し解析を行い、栄養素計算へ反映させて計算精度を向上させている。調理法をゆでる、焼く、煮る、生の4つの調理加工動作に分類し、食材の栄養素の変化を意識したレシピの栄養計算を行った。

2.2 インターネット情報源からの情報抽出に関する研究

ブログやツイッターをはじめとする、インターネット情報元からの情報抽出に関する研究は多く行われており、内容も多岐にわたる。

花井ら [9] の健康を意識した代替食材の発見手法では、日常的に作る料理が健康を意識した料理になるように食材の栄養成分の効能に着目し、特定の食材と他の食材との類似度を算出している。具体的な手法として、一般的に使用頻度の高い食材の栄養成分とその効能、過剰摂取から悪影響を及ぼす成分、食材

の食感の抽出を元に、他の食材の類似度を算出する。その結果から、健康目的別に代替可能な食材を発見する手法を提案している。栄養成分による効能の抽出では、ブログの特徴を捉えた段落を用いている。ブログは、短い文章が段落ごとに分けられていることが多い。よって、同一段落中での食材名と栄養成分の共起、栄養成分と効用名との共起に着目することで、抽出の誤差が少ないことを明らかにしている。

佐野ら [10] は、陸上競技選手モチベーション維持支援を行うために、陸上競技選手のブログの記事を対象とした、活動内容と大会記録の抽出手法の研究を行っている。ブログ記事に類出する活動メニュー名を登録した陸上競技練習用語辞書を作成している。その辞書を使用して、演算子を用いてブログ上から陸上の活動内容抽出を正規表現で行い、形態素解析をしている。結果、活動内容は精度が 67.3%、再現率が 57.3%であり、大会記録は、精度が 80.9%、再現率が 73.9%という結果である。

湯沢ら [11] は災害時の SNS 上のやりとりに着目し、災害に関連する投稿を抽出する研究を行っている。災害時は人と人とのやりとりが増える点から、挨拶や応答に用いられる感動詞と共起する語に着目し、さらに災害語と共起する語の共起関連から共起関連語集を作成した。この2つの共起関係を利用して手がかり語集合を生成する手法を提案している。正解割合において従来手法と比較し、より高い値が得ることができている。

栗原ら [12] は、Twitter から、機械学習とルールベースを用いて不具合情報を自動的に抽出している。機械学習ではテストデータである Twitter を適用すると適合率が大きく低下した結果に対し、ルールベースではテストデータを適用しても適合率が高い抽出結果となっている。機械学習の精度向上には元のデータ数を増やすことが重要であるが、そのためには人手によるデータ収集が必要となりコスト面で現実的ではない。ルールベースでは、ドメインに着目し、段階を追った抽出ルールを作成した。パターンマッチング等を用いた抽出問題とアプローチが適していたため、高い適合率で不具合情報を抽出することができたとしている。

3 栄養ロスフリーな調理法の抽出手法の提案

本研究では、インターネットを用い、栄養ロスフリーな調理法を抽出する手法を提案する。キーワードとなる「栄養ロス」は、書籍『その調理、9割の栄養捨ててます!』[13]と『その調理、まだまだ9割の栄養捨ててます!』[14]で広く認知され、これらの書籍に触発されて記載されたウェブサイトが多数存在する。一般的に認知されている食費ロスではなく、書籍の中の「栄養ロス」という観点を元に調理法の抽出を行う。

まず、インターネット上から調理法を収集するためのシードとなる語彙の選定について 3.1 で説明する。その後、決定したシード語彙を元にインターネット上からウェブサイトの収集を行う。収集したウェブサイトから目的箇所の抽出を行う処理について 3.2 で説明をする。ウェブサイトの構造や、特徴的な語の共起に着目し抽出を行う。最後に、抽出された目的箇所から栄養ロスフリー辞書の作成について 3.3 で説明をする。

3.1 シード語彙の選定

本研究では、Google 検索を用い、インターネット上から栄養ロスフリーな調理手法を収集するためのシードとなる語彙の選定を行う。Google 検索で「栄養ロス」という検索ワードを使用し検索を行った。検索結果の上位 150 件を閲覧し、人手で栄養ロスフリーな調理法の記載があるのかを調査した。結果、150 件中 11 件が栄養ロスフリーな調理法について書かれたウェブサイトだった。さらに、その 11 件中 9 件は、前述した 2 冊の書籍を参考にして書かれたウェブサイトであった。そのため、同じ内容の調理法しかウェブサイトに記載されていない。「栄養ロス」というキーワードでは、調理法の記載が少ないため、異なるキーワードで調査を行う。

次に「”野菜名” 栄養ロス」と「”野菜名” 栄養 下処理」の 2 つの検索ワードを元に google 検索で比較を行った。検索ワードに関しては、野菜名を入力する。“野菜名”は野菜の種類により異なり、検索ワードに野菜名を入力することによって、検索する野菜に特化し安定した抽出結果を得るためである。栄養の増減をウェブサイトからヒットさせるための検索ワードの候補として「下処理」「下ごしらえ」「下準備」を挙げ検討を試みた。共に似通った言葉であるが、下ごしらえ・下準備は調理の前段階として食材に加工を行うことだ。野菜の切り方、皮のむき方、水にさらす時間等、料理の食感や味または見た目を良くするための工夫である。一方下処理は、食材を加工する前段階の作業だ。野菜を切る前、皮をむく前、水にさらす前等の工夫である。最後に、検索ワード候補の組み合わせ比較を行った際、栄養をよりよく摂取する栄養ロスフリーな調理法が書かれていることが多かったのは、野菜の下処理のワードだった。

比較の結果から、本研究の主旨に添っている、栄養と下処理という言葉キーワードにした。10 種類の一般的に普及している野菜を上記の 2 つ検索ワードに当てはめ検索をし、上位 3 件のウェブサイト、計 60 件を閲覧した。その中から栄養ロスフリーな調理法の記載があるかを人手で調査した。結果、前者は 30 件中 18 件の同じウェブサイトがヒットし、後者はウェブサイトの重複がなかった。また、30 件中 6 件がキーワードに入力した野菜について書かれていないウェブサイトがヒットし、後者はキーワードに入力した野菜について書かれていなかった。よって、後者の方が多様な内容で栄養ロスフリーな調理法に該当するページを見つけることができていた。

以上のことから、本研究では「”野菜名” 栄養 下処理」をシード語彙に決定した。

3.2 目的箇所の抽出

ウェブサイトに記載されている栄養ロスフリーな調理法には食品会社を書いたものやコラム、個人で書いたブログなどの種類があり、多様な表現が用いられている。栄養ロスフリーな調理法は、「どのように調理すればよいのか」、「その調理をするるとどのような栄養面の効果があるのか」ということが記載されている。

本研究では、ルールベースで実装を行う。栄養ロスフリーな調理法の記載には特徴があるため正規表現で抽出をする。具体



図 1 HTML を div タグで区切った際の抽出例

加熱する,*,*,*,動詞,自立,*,*,サ変,基本形
加熱さ,*,*,*,動詞,自立,*,*,サ変,未然形
加熱し,*,*,*,動詞,自立,*,*,サ変,未然形
加熱せ,*,*,*,動詞,自立,*,*,サ変,未然形
加熱し,*,*,*,動詞,自立,*,*,サ変,連用形
加熱する,*,*,*,動詞,自立,*,*,サ変,終止形
加熱する,*,*,*,動詞,自立,*,*,サ変,連体形
加熱すれ,*,*,*,動詞,自立,*,*,サ変,仮定形
加熱しろ,*,*,*,動詞,自立,*,*,サ変,命令形
加熱せよ,*,*,*,動詞,自立,*,*,サ変,命令形

図 2 動詞の活用例

的には、特徴的な語の共起に着目する。さらに、HTML の特徴的な構造にも着目した。

最初にウェブサイトの収集後、ウェブサイトごとに HTML を取得する。それらを、Beautiful Soup を使用して div タグで区切る。Beautiful Soup はスクレイピングに特化した python ライブラリである。栄養ロスフリーな調理法が記載されている箇所は文中にまとまって掲載されている。そのため、各コンテンツをグループ化の際に使用される div タグに着目し、内容がまとまっている箇所で分けていく。

div タグで区切った例を図 1 で示す。div タグで区切ると入れ子構造になり、図のように 5 つの要素ができる。そして、要素の中には重複する場所がある。その、重複している要素を抜き取ることにより、栄養ロスフリーな調理法が記載されている最小単位の div を取得することができる。

区切った div タグの要素に野菜名、その野菜に含まれている栄養名、調理法が含まれているもののみを取り出す。栄養名は、4 つの野菜ごとに食品成分データベース¹に記載されている成分名を用い、調理法は調理手法が記載されている調理法辞書を使用する。調理法辞書には 329 個の調理法が記載されている。

調理法辞書に記載されている動詞を活用して調理法辞書の抽

1: <https://fooddb.mext.go.jp/>(参照 2020-12-10)

張を行った。ウェブサイトの調理法は、「煮る」などのような動詞のみもあるが、「加熱」「する」「裏」「こす」のように名詞+動詞で表現されている場合が多くある。動詞の活用形には、未然形・連用形・終止形・連体形・假定形・命令形の6種類があるが、調理法の抽出過程で多くの調理法の抽出を可能にするため全て採用した。動詞の活用の例を図2に示す。

3.3 栄養ロスフリー辞書の作成

3.2の手法を用いて栄養ロスフリーな調理法の抽出を行う。

抽出した目的箇所から、野菜ごとに調理法、効果の2つに仕分けを行いラベル付けをする。3.2で行った調理法の動詞活用形は形態素解析を使用して動詞の原形に戻す。ラベル付けしたものを4種類の野菜ごと分類し、栄養ロスフリー辞書に格納する。

4 実験と評価

4.1 データセット

4.1.1 ウェブサイトの収集

Google検索を用い、ウェブサイトの収集を行う。ウェブサイトの収集は、CustomSearchAPI²を使用する。CustomSearchAPIはGoogleが提供しているAPIであり、ウェブ検索または画像の検索結果をJSON形式で取得することができる。

CustomSearchAPIを使用し、シード語彙のもとウェブサイトの収集を行う。それぞれの研究対象の野菜につきGoogle検索の上位100件のウェブサイトの検索順位、URLを収集した。

そして、CustomSearchAPIで収集したURLからHTMLを取得する。また、収集したURLには、PDF、ウェブサイトが削除されている、クライアントからのリクエストが無効であるという理由でHTMLの取得が不可能なものがある。さらに、重複しているウェブサイトがあったため、それらを削除した。

収集データは2020年10月6日に収集したものを使用する。

4.1.2 野菜の選定

本研究では、4種類の野菜を対象に栄養ロスフリーな調理法の抽出を行う。この野菜の選定は、日本食品標準成分表³に記載されている野菜類を対象とした。日本食品標準成分表は、野菜の目標摂取量である350gの基礎となっている、国民健康・栄養調査における食品の分類表である。

さらに、書籍『その調理、9割の栄養捨ててます!』、『その調理、まだまだ9割の栄養捨ててます!』の対象とする野菜が記載されているのか、されていないのかを選択基準にした。この2冊の書籍を野菜の選定基準にする理由としては、シード語彙の選定時に210件のウェブサイトを閲覧した結果、この2冊を元に書かれたウェブサイトが多数存在したからである。

書籍には一般的に普及している野菜が記載されてある。その中でより一般的で最も調理法数の記載が多い、にんじんとだい

こんを研究対象の野菜にした。一方で、書籍に記載されていない野菜は一般的に普及していないものを研究対象にした。

また、野菜が緑黄色野菜か、その他の野菜かということを選定基準にした。

以上のことを踏まえ、にんじん、だいこん、エンダイブ、コールラビを研究対象の野菜にした。表1のような分類分けになる。

表1 研究対象とする野菜

	書籍に記載されている	書籍に記載されていない
緑黄色野菜	にんじん	エンダイブ
その他の野菜	だいこん	コールラビ

4.2 提案手法の実行結果

3章の提案手法を用いて、4.1.2で選定した4種類の野菜を元にシステムを実行した。

にんじんは981件、だいこんは1265件、コールラビは452件、エンダイブは380件のデータを抽出することができた。以下の表2、表3、表4、表5は、にんじん、だいこん、コールラビ、エンダイブの栄養ロスフリー辞書の抽出結果の一部である。

表2 にんじんの栄養ロスフリー辞書

調理法	効果
熱す	可食部100gに9100μg(マイクログラム)も(後略)
炒める	ニンジン・ナスは油で炒めてβカロテン、ナスニ(後略)
防ぐ	緑黄色野菜に含まれ、美肌がいいβカロテンは、(後略)
炒める	緑黄色野菜に含まれ、美肌がいいβカロテンは、(後略)
含む	緑黄色野菜に含まれ、美肌がいいβカロテンは、(後略)

表3 だいこんの栄養ロスフリー辞書

調理法	効果
おろす	プロテアーゼ(たんぱく質分解酵素)効果を期待(後略)
もみ	プロテアーゼ(たんぱく質分解酵素)効果を期待(後略)
含む	大根の葉はβ-カロテンを多く含む緑黄色野菜。ま(後略)
含む	大根の根の部分は淡色野菜、葉の部分は緑黄色野(後略)
捨てる	大根の根の部分は淡色野菜、葉の部分は緑黄色野(後略)

表4 コールラビの栄養ロスフリー辞書

調理法	効果
まとめ	コールラビの簡単人気レシピを大公開!栄養豊富(後略)
出す	コールラビに含まれているカリウムの特徴は、体(後略)
足す	コールラビに含まれているカリウムの特徴は、体(後略)
含む	コールラビに含まれているカリウムの特徴は、体(後略)
含む	淡色野菜類の中では上位に上がるほどビタミンC(後略)

2: <https://developers.google.com/custom-search/v1/overview>(参照2020-12-21)

3: https://www.mext.go.jp/a_menu/syokuhinseibun/1365297.htm(参照2020-12-10)

表5 エンダイブの栄養ロスフリー辞書

調理法	効果
持ち上げる	エンダイブ [Endive] という洋野菜があるの (後略)
合わせる	エンダイブ記事をご覧の皆様エンダイブに関 (後略)
含む	エンダイブにはカリウムやカルシウムがたく (後略)
取る	3つのテーマから献立を選ぼう!「簡単」献立 (後略)
揚げる	3つのテーマから献立を選ぼう!「簡単」献立 (後略)

にんじんとだいこんは一般的に普及している野菜ということもあり、多くの調理法を抽出できた。反対に一般的に普及していない野菜であるコールラビとエンダイブはにんじん、だいこんに比べると抽出できた調理法は少ない。

3.3の手法のように、拡張した調理法辞書の動詞を形態素解析により原形に戻した。しかし、表3の「もみ」や表4の「まとめ」のように本来なら「もむ」、「まとめる」になるはずが動詞の原形でない形になっている調理法がある。これらは動詞の活用形にした場合に、「もむ」の連体形が「もみ」になる。この「もみ」を形態素解析にかけると名詞の「もみ」と判定される。同じく、「まとめ」も形態素解析の結果名詞と判定されたためである。そのため、動詞の原形でない形の調理法が存在する。

上記の表の中には、異なった調理法で効果が同一の結果が抽出された。効果の欄には同一のデータが並んでいるのに対して、調理法が2行から数行に渡り異なっている。その理由は、抽出した1つのdiv要素の中に、調理法辞書に記載されている調理法が複数含まれている場合があったためである。

4.3 精度評価

4.3.1 正解データの作成

抽出した目的箇所を手で作成した正解データと比較し、精度評価を行うための正解データを作成する。

本研究のシステムでは、3.2に示した手法の通り図3のようにdivタグで区切り抽出を行う。そして、黄色のハイライト部分が本来抽出したい正解データである。この正解データは人手で判定をする。抽出したい箇所が含まれている最小単位のdiv (図中で緑色の部分) を正解とする。

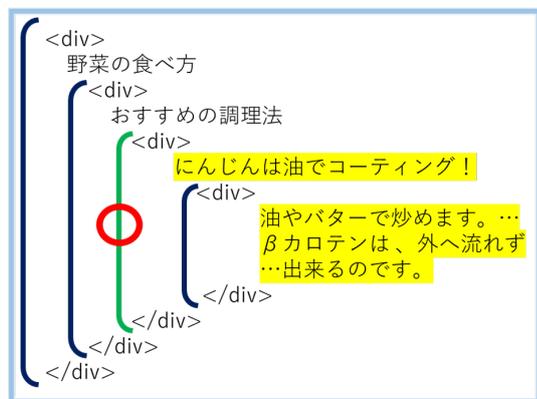


図3 精度評価に基づく抽出箇所の基準

1つの野菜に対し100件のウェブページを収集したため、合

計400件のウェブページを手で確認した。

4.3.2 正解データとの照合結果

適切に栄養ロスフリーな調理法が抽出できたかどうかを確認するため、4.3.1で作成した正解データを用いて、混同行列を行った。そして、適合率と再現率とF値の計算精度を確認した。本研究では、4.3.1で述べたように抽出したい箇所を含む最小単位のdivが正しく抽出できたのかということの評価をする。そのため、調理法のデータは使わず、栄養ロスフリー辞書の効果のデータのみを使用する。また、栄養ロスフリー辞書では同一のデータの効果が重複しているが、重複しているデータは1つのデータとして扱う。

精度と再現率とF値は式(1)と式(2)と式(3)を用いて求める。

$$\text{適合率} = \frac{\text{システムが正しく抽出したデータの数}}{\text{システムが抽出したデータの数}} \quad (1)$$

$$\text{再現率} = \frac{\text{システムが正しく抽出したデータの数}}{\text{正解データの数}} \quad (2)$$

$$F \text{ 値} = \frac{2 * \text{適合率} * \text{再現率}}{\text{適合率} + \text{再現率}} \quad (3)$$

これらの式を用い、4つの栄養ロスフリー辞書の結果を表6に示す。

表6 栄養ロスフリーな調理法の抽出の評価結果

	にんじん	だいこん	コールラビ	エンダイブ
適合率	0.38	0.11	0.27	0.00
再現率	0.67	0.65	0.62	0.00
F 値	0.48	0.19	0.37	0.00

にんじん、だいこん、コールラビの適合率は全体的に高くなかった。間違えて抽出してしまったデータの多くは、クックパット⁴やクラシル⁵など、レシピ紹介サイトからの料理献立やレシピであった。さらに、栄養名を含む言葉に反応してしまい不適切に抽出されてしまったデータがある。一例として、にんじんの栄養の1つに「リン」がある。このリンが「スポンサーリンク」、「マリン」、「リンゴ」という言葉に反応してしまった。

また、コールラビに関しては、一般的に普及していない野菜のため、コールラビ自体の紹介のウェブサイトが多かった。その中では、コールラビの旬や食べ方・栄養に関する紹介はされていたが、栄養ロスフリーな調理法の記載はされていなかった。しかし、調理法辞書に基づいて抽出を行っているため、「免疫力を高めるビタミンCと、体内の余分な塩分を排出するカリウムを、多く含んでいます。」という文があった場合、文中に「含む」という調理法が記載されているため抽出されてしまっていた。

エンダイブも一般的に普及していない野菜であり、ウェブサイトにも栄養ロスフリーな調理法の記載があまり見受けられなかった。実際、エンダイブの正解データの数は他の3種類に比

4 : <https://cookpad.com/>(参照 2020-12-21)

5 : <https://www.kurashiru.com/>(参照 2020-12-21)

べると少ない。正解データ数が少ないため、は期待する数値が得られなかった。

にんじん、だいこん、コールラビの再現率はすべて 6 割を超えることができた。抽出できなかった多くのデータは、ウェブページから HTML を取得する際にクライアントエラーが出てしまい、ウェブページから HTML を収集することができなかったためである。本研究で収集した栄養では対応できず、抽出不可能だった栄養ロスフリーな調理法もあった。

4.4 抽出した調理法の件数比較

4.4.1 書籍内に記載されている調理法の件数の調査

2 冊の書籍から、本研究で対象としている野菜の栄養ロスフリーな調理法の件数とシステムが正しく抽出した調理法の件数比較を行うために正解データの作成を行う。

また、書籍内の調理法の内容には重複がある。例えば、「人参のβ-カロテンは脂溶性なので、油炒めや揚げ物、バターソテーなどの調理方法でいただくと、吸収率を高めることができ、必要量だけビタミン A に変換され、残りは体内に蓄積されます。」と「にんじんのβ-カロテン（ビタミン A）は「油（オリーブオイルを含む）」を使った料理でいただくと、栄養の吸収が良くなります。」が抽出された場合、どちらも「にんじんは油を使って調理するとβカロテンの吸収が良くなる」という調理法である。このように、重複している場合は 1 つの調理法としてカウントをする。

にんじんとだいこんは書籍に調理法が記載されているため、書籍に記載されている調理法の件数を人手で数えた。

4.4.2 書籍と提案手法との比較

4.1.1 で作成したデータを元に、書籍に記載されている調理法の件数と、システムから抽出できた調理法数の比較を行った。システムから抽出できた調理法にも 4.1.1 のように重複している内容があったため、人手で確認をした。実験の結果を表 7 に示す。

表 7 書籍とシステムに記載されている調理法数の結果

野菜	書籍	システム
にんじん	12	31
だいこん	9	25
コールラビ	0	4
エンダイブ	0	0

にんじんとだいこんは、2 冊の書籍の中に栄養ロスフリーな調理法が記載されており、結果は書籍に記載されている調理法の 2 倍以上の調理法を抽出することができた。一方コールラビとエンダイブに関しては、2 冊の書籍の中に栄養ロスフリーな調理法の記載がない。コールラビに関しては、システムでは多くの調理法を抽出することができた。エンダイブはシステムで正しい栄養ロスフリーな調理法を抽出できなかったため、このような結果になった。

4.5 考察

システムを実行し栄養ロスフリー辞書を作成した結果と、そ

れらを元に行った精度評価と、調理法の件数の比較を踏まえ、本研究の手法の有効性を考察する。

システムの実行結果では、4 種類の野菜を 3.3 で提案した通り、野菜ごとの調理法・野菜ごとの効果の 2 つに仕分けを行い、ラベル付け後に分類し、栄養ロスフリー辞書に格納することができた。しかし、表 3、表 4 の調理法のように、動詞が原形に戻っていない調理法が数箇所ある。この問題を解決するための手段として、以下のようなことが考えられる。本研究で提案した調理法辞書の拡張は、「加熱する」・「裏ごす」などの名詞+動詞の調理法には有効だったが、「もむ」・「まとめる」など動詞だけの調理法には有効に働かなかった。この結果から、名詞+動詞の調理法と動詞だけの調理法でアプローチを変えることが必要であると考えられるため、以下に手順を示す。動詞だけの調理法は、抽出した div に形態素分析で動詞があるか判定させ、その動詞を原形にし、拡張してない動詞の原形のみが記載されている調理法辞書に適合するかを判断する。それにより、名詞+動詞の調理法と動詞だけの調理法の双方に対応することができ、この問題の解決につながると考えられる。

次に、精度評価については期待する数値が得られなかった。特に、エンダイブに関しては正しい栄養ロスフリーな調理法を抽出することができなかった問題点がある。理由としては、エンダイブ自体が一般的に普及していない野菜であるため、栄養ロスフリーな調理法の記載がウェブサイトには少ないことが考えられる。本研究では、100 件のウェブサイトを対象とし実験を行ったが、このような場合を含めさらに多くのウェブサイトを対象とすることが必要であると考えられる。続いて、エンダイブ以外の野菜について、適合率が低い理由が 3 点考えられる。大きな要因は、献立やレシピに関するデータが多く抽出された点である。この点に関しては、決まったレシピ紹介サイトから抽出されていることが原因であり、そのようなウェブサイトを抽出外にすることが必要である。実際、正解データを作る際に 400 件のウェブサイトを閲覧したが、料理紹介サイトには栄養ロスフリーな調理法の記載が見当たらなかった。よって、レシピ紹介サイト等は抽出の対象外にしても影響は少ないと考えられる。2 点目は、栄養名が本来の栄養ではない言葉に反応して不適切に抽出されてしまう点である。カタカナの栄養名が他の言葉に反応してしまう場合があり、この問題を解決するためには形態素分析を行うことが必要であると考えられる。本研究では栄養名が div 要素に含まれている要素を抽出する手法だったが、div 要素を形態素分析で名詞が含まれているか判定し、その名詞が栄養名と完全一致するもののみを抽出することで、不適切な抽出が減少すると考えられる。3 点目は、一般的に普及していない野菜は、認知度の低さからその野菜の紹介がされているデータが誤って抽出されるケースである。この問題に関しては、ウェブサイトの紹介文に頻出する特徴的な言葉をストップワードとして設定することで、適合率が上がると考えられる。

再現率に関しては、6 割を超えることができた。さらに再現率を上げるために 2 点のことを提案する。1 点目は、ウェブサイトから HTML を取得する際にクライアントエラーが出てしまう点である。これについては、取得方法を再検討する必要が

- [1] 健康局健康課栄養指導室. 国民健康・栄養調査. 令和元年「国民健康・栄養調査」の結果を公表します, 2019. https://www.mhlw.go.jp/stf/newpage_14156.html, (参照 2020-12-10).
- [2] 下橋淳子. 野菜摂取による健康増進. 駒沢女子大学研究紀要, No. 22, pp. 135-142, 2015.
- [3] 田中沙智. 食品成分による免疫制御メカニズム. 信州大学農学部紀要, Vol. 51, pp. 1-8, 2015.
- [4] 松戸市役所. 毎食、食べてほしい野菜の量について. 毎食、食べてほしい野菜の量について 医療都市まつど | 松戸市, 2020. https://www.city.matsudo.chiba.jp/iryoutoshi/healthcare/shokuseikatsu_point/veg-ryou.html, (参照 2020-12-10).
- [5] 角沙月, 塚本享治. 栄養価計算機能を有する料理レシピ検索サービスの構築. 情報処理学会 全国大会講演論文集, 第 70 回, データベースとメディア, pp. 703-704, 2008.
- [6] 苺米志帆乃, 藤井敦. 栄養素等摂取バランスを考慮した料理レシピ検索システム. 電子情報通信学会論文誌, Vol. 92, No. 7, pp. 975-983, 2009.
- [7] 岩上将史, 安藤哲志, 伊藤孝行, 田中雅章. 栄養情報を活用した目的指向料理推薦システムの試作. 情報処理学会 第 73 回全国大会講演論文集, Vol. 2011, No. 1, pp. 581-582, 2011.
- [8] 植田嗣也, 高橋淳, 吉村卓也, 伊藤孝行. 料理レシピの栄養素自動計算システムの試作: 「焼く, 煮る」などの調理加工による栄養素の変化の自動計算手法. 情報処理学会 第 74 回全国大会講演論文集, Vol. 2012, No. 1, pp. 435 - 436, 2012
- [9] 花井俊介, 難波英嗣, 灘本明代. 健康を意識した代替食材の発見手法. DEIM Forum 2015, No. G6-6, pp. 1-8, 2015.
- [10] 佐野正和, 福原知宏, 増田英孝, 山田剛一. 陸上競技ブログからの活動記録抽出. 人工知能学会全国大会 (第 29 回) 論文集, Vol. 29, pp. 1-4, 2015.
- [11] 湯沢昭夫, 小林亜樹. 感動詞の共起に着目した災害 tweet 抽出手法. 情報処理学会 第 80 回全国大会講演論文集, Vol. 2018, No. 1, pp. 387-388, 2015.
- [12] 栗原光平, 嶋田和孝. ルールと機械学習を用いた twitter からの不具合情報の抽出. 電子情報通信学会技術研究報告. 一般社団法人電子情報通信学会, Vol. 114, No. 81, pp. 1-6, 2014.
- [13] 東京慈恵会医科大学附属病院栄養部監修. その調理、9 割の栄養捨ててます! 世界文化社, 2017. 144p.
- [14] 東京慈恵会医科大学附属病院栄養部監修. その調理、まだまだ 9 割の栄養捨ててます! 世界文化社, 2019. 144p.

ある。2 点目は、野菜に含まれる栄養の種類を拡張する点である。本研究で収集した栄養の種類では対応ができず、抽出不可能だった栄養ロスフリーな調理法が存在するため、拡張を行うことで再現率が高くなる可能性が大きい。

F 値は期待する結果が得られなかったが、上記したことを踏まえることで、適合率と再現率の上昇が見込め、合わせて F 値も向上すると予測される。

最後に、調理法の件数の比較については、にんじん、だいこんは書籍以上の栄養ロスフリーな調理法が取得達成できた。また、コールラビに関しても新たな栄養ロスフリーな調理法の取得ができたため、本研究の手法は有効であると考えられる。

エンダイブに関しては、栄養ロスフリーな調理法が抽出できなかったが、上記したことを踏まえてシステムの実行を行うことで、栄養ロスフリーな調理法が抽出でき、新たな栄養ロスフリーな調理法の取得ができると考えられる。

5 結 論

本研究では、野菜に含まれる栄養が調理によって増減することに着目し、インターネット上から栄養ロスフリーな調理法を抽出する手法を提案した。

具体的には、Google 検索を使用し、インターネット上からウェブサイトの収集を行った。収集の際には「”野菜名” 栄養下処理」をシード語彙に定めた。野菜ごとに 100 件のウェブサイトを収集後、ウェブサイトの構造や、特徴的な語の共起に着目し抽出をした。そして、抽出された目的箇所を調理法と効果にラベル付けをし、栄養ロスフリー辞書を作成した。本研究で研究対象とする野菜は、にんじん、だいこん、コールラビ、エンダイブの 4 種類に設定した。結果、栄養ロスフリー辞書の調理法である動詞が原形になっていない箇所があった。また、精度評価では、エンダイブに関しては栄養ロスフリーな調理法の抽出はできなかったが、他の野菜は栄養ロスフリーな調理法を抽出することができた。しかし、期待する数値は得られなかった。続いて、システムで抽出した栄養ロスフリーな調理法の件数と 2 冊の書籍に記載されている栄養ロスフリーな調理法の件数比較を行った。2 冊の書籍に記載されているにんじん、だいこんは書籍の 2 倍以上の調理法を抽出することができ、書籍に記載されていないコールラビは新しい調理法を抽出することができた。

今後の課題として 2 点挙げられる。1 点目は、栄養ロスフリーな調理法の抽出精度を向上させることである。本手法では、書籍に記載されている件数以上の栄養ロスフリーな調理法は抽出できたが、精度評価が低かった。これについては、抽出の対象としないウェブサイトやストップワードの設定を行うことで改善されると考えられる。2 点目は、本研究では 4 種類の野菜を研究対象として実験を行ったが、他の野菜でも栄養ロスフリーな調理法が抽出可能かを調査することである。本手法は、野菜の種類に依存しない抽出を行ったため、4 種類以外の野菜でも栄養ロスフリーな調理法の抽出が行えることが期待できる。