

SEO 対策向け文章校正システムの提案

西村祐輝¹ 芋野美紗子¹

概要：ビジネスにおける集客でインターネットは不可欠となっている。例えば自分が提供しているサービスのウェブページを検索結果に表示させることで集客に繋げることができる。ウェブページを検索結果に表示させるための対策を SEO と呼ぶが、これにはある程度の専門知識を要するほか、難しい場合には外注費用が発生することも考えられる。そこで、人間に代わって SEO 対策を支援するシステムを構築すれば、このような問題解決に繋がると考えた。ウェブページの文章は、人間のみならず検索エンジンにも読まれる。そこで本研究では、人間と検索エンジンの両方の視点から良質な文章であるかを判定するシステムを構築した。システムに入力した文章とその判定結果を提示してアンケート評価を行った。文字数や執筆者など異なる条件の文章で実施したところ、「的を射た判定である」と回答した割合は最大で 93.3%、最小で 13.3% となった。提示した文章の特徴とアンケート結果を比較すると、1000 文字以下の文章では判定精度が低下する傾向が見られた。

キーワード：SEO 対策, 文章校正, Web 集客

Proposal of Sentence Proofreading System for SEO

YUKI NISHIMURA^{†1} MISAKO IMONO^{†1}

Abstract: The Internet has become indispensable for attracting business customers. For example, companies can connect with customers by displaying their Web pages on search results. The measures to display Web pages on search results are called Search Engine Optimization (SEO). SEO requires a certain amount of technical knowledge, and outsourcing costs may be incurred in cases where it is difficult. Therefore, we thought that building a system to support SEO on behalf of humans would solve such problems. In this study, we constructed a system to judge whether a sentence is good from both human and search engine perspectives because Web page text is read not only by humans but also by search engines. The sentence input to the system and the judgment result were presented for evaluation. As a result of using texts with different conditions such as the number of characters and the author, the percentage of respondents who answered "adequate" was 93.3% at the maximum and 13.3% at the minimum. Comparing the presented text with the results of the questionnaire, the judgment accuracy tended to decrease for texts with less than 1000 characters.

1. はじめに

ビジネスにおける集客手法として、インターネットが欠かせないものとなっている。商売を成り立たせるためには顧客が必要であり、そのためには集客という工程を避けては通れない。集客方法は多種多様であるが、その 1 つとしてインターネットを活用する SEO 対策がある。

我々が情報収集する際、Google や Yahoo などの検索エンジンを活用している。もし自分の店舗やサービスを検索結果に表示すれば、より多くの人に存在を知ってもらえるので集客に繋がる。このようにウェブページを検索結果に反映させる作業を SEO (Search Engine Optimization: 検索エンジン最適化) [1] という。

しかし SEO は容易な作業ではなく、ある程度の専門知識が必要である。自社で行うことが難しい場合は専門の業者へ外注する必要がある、余計に経費が嵩むことになる。そこで、システムに文章を校正させることで人間に代わって SEO 対策の助言を行えないかと考えた。

本稿では、情報発信者側の視点で SEO 対策に特化した文章校正を行うためのシステムを提案する。

2. SEO 対策について

SEO 対策を行う目的は 2 つある。1 つ目は制作したページを検索結果に反映させるためである。検索エンジンは、インターネット上にクローラーと呼ばれるロボットを巡回させ、無数のウェブページを整理している。整理したウェブページは検索エンジンのサーバーに蓄積され、何かしらのキーワードで検索された際に、それに合致したページを検索結果に表示する。たとえば「卒業論文の書き方」を解説するページを作る場合、ここで書くべき文章は、「卒業論文」「書き方」の 2 つのキーワードを含めた文章である。クローラーが文章中に頻繁に出現する単語を用いて認識していることは検索結果から読み取れる。

2 つ目は人間にとって読みやすい文章を書くことである。最終的にウェブページに書かれた文字を読むのは人間である。もし読みにくければ、読まれずにページを閉じてしまう可能性が高く、検索結果に表示されても意味がない。文章を読む人の立場で考えてウェブページを制作することが健全な SEO 対策である。

¹ 大同大学情報システム学科
Dept. of Information Systems, Daido University

3. システムの概要

本システムは、制作したウェブページの本文とタイトルを入力すると、人間が読みやすく且つクローラーが内容を理解しやすい文章で書かれているかを判定し、制作者に助言を行う。図1にシステムの概略を示す。

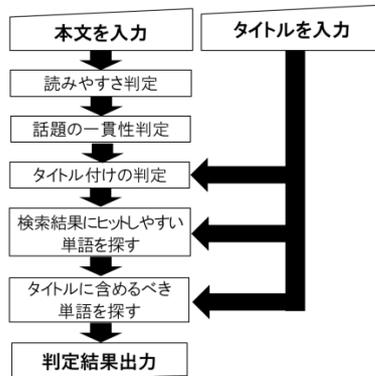


図1. システムの流れの概略

3.1 読みやすさ判定

読みやすい文章に仕上げるためには、適切な文法を意識して書くことは当然である。しかし、最終的にウェブページを読むのは不特定多数の人間である。本節では不特定多数の人間にとって読みやすい文章であるかを判定するための手法を述べる。

(1) TF 値と検索結果の相関について

システムで SEO 対策を行うにあたって、検索ワードの TF 値^[2]と検索結果の順位に相関があるかを調べた。例えば「マーケティング 集客」と検索し、1~20 位に表示されたページの文章で TF 値が高い単語を調べる。すると、検索ワードである「マーケティング」「集客」が抽出される。これらの単語の TF 値を求めると、0.01~0.03 の範囲に収まる傾向にあった。さらに下位になるにつれて TF 値が低くなる傾向もみられ、順位と TF 値に相関があることが推測される。順位と単語の TF 値を散布図にプロットし、近似線を引くと図2のグラフになる。

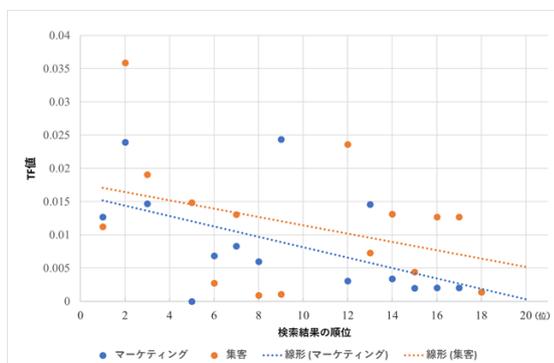


図2. 「マーケティング 集客」の検索順位と TF 値

同様の方法で「心臓 痛い」でのキーワードでも調べると、

図3に示す結果となる。同キーワードで表示されるのは医療関係のページであるが、業界によって TF 値が低くなるにせよ、同様の傾向がみられた。よって、文章中から TF 値が 0.01 以上の単語を抽出すれば、その文章がどのようなキーワードで検索される可能性があるかを推定できると考えた。

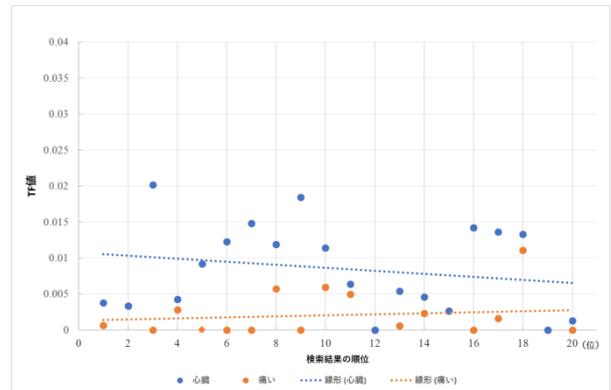


図3. 「心臓 痛い」の検索順位と TF 値

ただし、故意に同じ単語を過剰に使うことはクローラーがスパム行為と判断する恐れがあるため推奨されない。また人間視点での読みやすさにおいても、同じ単語を過剰に含めることは文の表現として不自然である。

(2) 読みやすさの判定手法

前述した通り、ウェブページの文章は人間が読みやすくしなくてはならない。そこでこの処理では文章の読みやすさを判定し、文章の改善を促す。処理の流れを図4に示す。

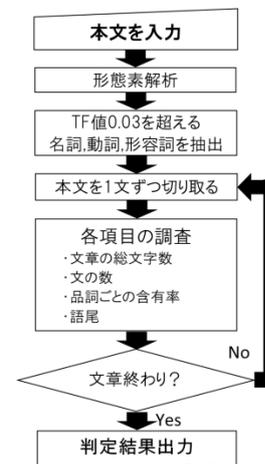


図4. 読みやすさ判定の流れ

はじめに入力した本文を形態素解析^[2]し、TF 値が 0.03 以上の名詞、動詞、形容詞を探す。過剰に同じ単語を繰り返すとスパムと判定される恐れがあることから、システムが別の言葉を使うよう促す。次に、句点、疑問符、感嘆符、空白の位置を調べて本文を1文ずつ切り取り、表1に示す判定条件を満たすかどうかを調べる。判定条件を満たした場合、システムが文章に対して指摘を行う。

表 1. 判定する文章の特徴と判定条件

文章の指摘項目	判定条件
1文が長い	1文あたりの平均文字数が60字以上
句読点の打ちすぎ	句読点の含有率5%以上 かつ 句読点までの平均文字数15字以下
人間味がなく 無機質	代名詞の含有率1%以下 かつ 代名詞+副詞の含有率3%以下
接続詞が多すぎる	接続詞の含有率3%以下 かつ 接続詞を含む文が40%以上存在
ナルシストな 雰囲気が出ている	1文の文字数が平均60字以上 かつ 代名詞+副詞の含有率3%以上
くどい印象が強い	接続詞+副詞+係助詞の含有率7%以上 かつ 終助詞の含有率1%以上
同じ単語の多用	TF値が0.03を超える単語が存在
押しつけがましい	1文の文字数が平均31~49字 かつ 代名詞+副詞の含有率5%以上 かつ 終助詞の含有率1%未満
稚拙な印象がある	1文の文字数が平均31字未満 かつ 終助詞の含有率1%未満
内容が抽象的	形容詞の含有率2%以上
単調な印象がある	同じ語尾が4回連続している

表1に示す判定条件の決定にあたって、11人に対してアンケートを実施した。文章の基礎に関する書籍^[3]やウェブページ^{[4][5]}にある例文を用いて、9つの選択肢（「1文が長い」、「句読点が多い」、「接続詞が多い」、「曖昧な表現が多い」、「回りくどい」、「単調でつまらない」、「偉そうな雰囲気がある」、「押しつけがましい」、「稚拙な印象が強い」）から該当する特徴をすべて選択してもらった。そのアンケートから文章の特徴を判定する条件を調査・決定した。調査内容の一部を以降で説明する。

まず表2は人間が長いあるいは短いと感じる1文あたりの平均文字数を調べた結果である。

表 2. 1文あたりの平均文字数についてのアンケート結果

1文あたりの平均文字数	1文が長いと回答した割合	稚拙な印象が強いと回答した割合
23文字	0.0%	27.3%
28文字	18.2%	18.2%
32文字	9.1%	18.2%
33文字	18.2%	18.2%
34文字	27.3%	18.2%
36文字	9.1%	9.1%
43文字	18.2%	0.0%
53文字	18.2%	0.0%
66文字	45.5%	0.0%

次に文章を読んだ際の印象についてのアンケート結果を調査した。図5、図6、図7は、アンケートで用いた文章の一部である。

私は1つのことに真剣に取り組むことができる真面目な性格です。一心に作業や政策に打ち込める集中力も持っています。私の真面目さは、勉強はもちろん作品制作においても、小さいころから学校の先生や友人に認められてきました。その上、私には観察力があり、観察することを通じて何かを学ぶことが得意です。(略)

図 5. 自慢臭が強い文章の例^[3]

では、パラダイムによらない場合は、どうするか。この場合、自前で設定した目標を、自力で追求してゆくことになるが、その際、くれぐれも留意すべきは、独りよがりにならないということであろう。そのためには、様々なパラダイムの中で、達成されている知見に心を配り、目を見開いている必要がある。(略)

図 6. 句読点が多い文章の例^[4]

Aさんは子育てと仕事を両立する在宅Webライターです。朝5時に起きて1時間半ほど仕事をしてから朝食を作り、7時に子どもを起こします。着替えや登校の準備をさせて子どもが出かけた後、9時頃から本格的に仕事を始めます。

図 7. 単調な文章の例^[5]

図5は書籍から引用した文章で、自慢臭が強いと解説されている。アンケートを実施したところ、27.3%が「偉そうな雰囲気がある」、36.4%が「押しつけがましい」と回答した。図6は81.8%が「句読点が多い」と回答したが、ほかの選択肢において、27.3%が「偉そうな雰囲気」、「回りくどい」と回答した。図5、図6の対照となる例が図7である。図7は54.5%が「単調でつまらない」と回答した。機械的な文と捉えられたことが、単調でつまらないと回答した要因になったと思われる。

図5、図6、図7において、副詞、代名詞、終助詞の含有率を比較すると表3のようになる。回りくどいと感じた人が多かった図6は終助詞の含有率が多く、口語的な表現が回りくどい印象を与えていると考えられる。

表 3. 副詞・代名詞・終助詞の含有率比較

例文	副詞含有率	代名詞含有率	終助詞含有率
図5	1.4%	3.8%	0.0%
図6	3.7%	3.7%	1.47%
図7	0.0%	0.0%	0.0%

「強い自慢臭がする」と解説されている文章であっても、アンケート結果から実際にそのように感じた人は2~3割程度である。よって今回は、先述した9つの選択肢から回答率が20%以上だった選択肢を文章の特徴としてシステムの判定条件に利用することにした。

3.2 話題の一貫性判定

SEOを意識してウェブページを制作するには、話題ごとに1ページを設けることが有効である^[6]。卒業論文の書き方を調べたい人にとっては、それ以外の話題には興味がな

いので余分な話題は削除することが望ましい。本節では文章中に使われている単語を使って cos 類似度を算出することにより、話題を絞って書かれているかを判定する手法を述べる。

(1) cos 類似度

異なる単語間の意味の近さを調べるために cos 類似度^[2]を用いる。単語 A と単語 B で比べたとき、意味が近いほど最大値の 1.0 に近い値が算出される。本研究ではウェブページの文章ではあらゆる話題や専門用語を取り扱うため、朝日新聞単語ベクトル^[7]を利用した。これは 33 年分、約 800 万記事(延べ 23 億単語)を用いて生成された単語ベクトルである。算出例が表 4 である。

表 4. cos 類似度の算出例

単語 A	単語 B	cos 類似度
研究	専門	0.49048486
研究	遊び	0.01082329

「研究」といえば専門的や学術的といったことが想起されるように、2 つの単語の間に関連性がある場合には cos 類似度が高くなる。逆に 2 つの単語間で関連性が想起されない場合は cos 類似度が 0 に近づく。

(2) 話題の一貫性判定の手法

図 8 に話題の一貫性を判定するための流れを示す。



図 8. 話題の一貫性を判定する流れ

本文を形態素解析し、5 つの TF 値 0.009, 0.007, 0.005, 0.003, 0.001 を閾値に用いて名詞、動詞、形容詞を抽出する。5 つの閾値を用いる理由は、本文中において出現頻度が低くても話題に関係する単語が含まれている可能性があるためである。

具体例を用いて判定手法を説明する。図 9 の文章は、社内業務の話題に絞って書かれている例である。図 9 から先述した閾値を用いると図 10 に示す単語が抽出される。

業務の生産性の品質の向上のためには、スキルや能力の向上もさることながら、多くは本人の「やる気」が結果を左右することがあります。そうかといって「やる気を出して頑張れ!」とただ指示して出るのなら苦勞はありません。むしろ人に言われてやらされるよりも、自分の意思で能動的に取り組んだ方が全然結果が違うなんてことはよくあることではないでしょうか。(略)

図 9. 話題を絞って書かれている例^[8]

仕事 やる気 する

図 10. 図 9 に含まれる TF 値 0.009 以上の単語

閾値 t において本文から抽出された n 個の単語を (A_1, A_2, \dots, A_n) , cos 類似度の平均を \overline{X}_t とすると、以下の計算式で示すことができる。

$$\overline{X}_t = \frac{1}{C_2} \sum_{i=1}^{n-1} \sum_{j=1}^{n-i} \cos(\overline{A}_i, \overline{A}_{i+j})$$

同様の手順で、閾値 0.007, 0.005, 0.003, 0.001 の場合の \overline{X}_t も求め、最も高い値を \overline{X}_{max} として表 5 に示す判定基準に沿ってシステムが判定を行う。図 9 の文章をシステムに入力したときの出力結果を図 11 に示す。

表 5. 話題の一貫性を判定する基準

\overline{X}_{max} の値	判定
$\overline{X}_{max} \geq 0.05$	話の筋が通っている(話題を絞っている)
$0.025 \leq \overline{X}_{max} < 0.05$	話がそれている(余談が含まれている)
$\overline{X}_{max} < 0.025$	何が言いたいのかわからない

◆本文一貫性チェック◆
 cos類似度平均(TF=0.009) : 0.09972048675020535
 cos類似度平均(TF=0.007) : 0.06213620470629798
 cos類似度平均(TF=0.005) : 0.03728172282377879
 cos類似度平均(TF=0.003) : 0.06460411488990654
 cos類似度平均(TF=0.001) : 0.06822930953713993
 ⇒話の筋が通った文章になっています。合格です。

図 11. 図 9 の文章を判定した結果

図 12 は、あるブログ記事から抽出した TF 値 0.009 以上の単語である。プライバシーの観点から文面は割愛する。図 12 の単語でシステムが判定を行うと図 13 に示す結果が出力された。話題が混合し要旨が不明瞭になっている文章では、cos 類似度の算出において関連性が低い組み合わせが存在することが考えられる。その場合、 \overline{X}_{max} が低下するので話題に一貫性があるかの判定をするために有効だと考えた。

する 注射 A 受け させ 有様 満喫

図 12. 要旨が分からない文章に含まれる単語例

◆本文一貫性チェック◆
 cos類似度平均(TF=0.009) : 0.00022769666143826077
 cos類似度平均(TF=0.007) : 0.00011384833071913038
 cos類似度平均(TF=0.005) : 7.589888714608692e-05
 cos類似度平均(TF=0.003) : 5.692416535956519e-05
 cos類似度平均(TF=0.001) : 4.553933228765215e-05
 ⇒何が言いたいのかわからない文章になっています。

図 13. 図 12 の単語でシステムが判定した結果

3.3 SEO 対策を意識したタイトル付けであるかの判定

検索結果に反映させるためには、タイトルの付け方^[6]が重要になる。例えば、「卒業論文 書き方」と検索した時に表示されるタイトルを見ると、必ず検索ワードが含まれて

いる。ここではタイトルと本文の単語を解析し、SEO 対策を意識したタイトルであるかを判定する手法を述べる。判定の流れを図 14 に示す。

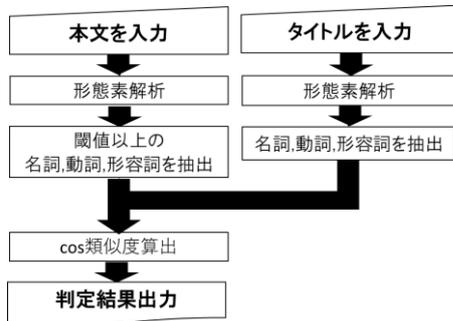


図 14. タイトル付けのよし悪しを判定する流れ

本文とタイトルを形態素解析し、名詞、動詞、形容詞を抽出する。タイトルから抽出した単語と本文から抽出した単語で意味に関連性があるかを、cos 類似度により判定する。本文から単語を抽出する際には 5 つの TF 値、0.009, 0.007, 0.005, 0.003, 0.001 を閾値に用いる。出現頻度が低くても話題に関係する単語が含まれている可能性があるためである。図 15 はタイトルと本文が合っている例である。タイトルと本文を形態素解析すると、図 16 に示す単語が抽出される。

タイトル:【5 つの施策例付】生産性向上に取り組むには、何かからどう始めればいいのか?
労働力不足が加速する中、企業では少ない労働量でも成果を生み出せるよう、生産性向上に向けた取り組みの必要性が高まっています。企業ができる生産性を上げるための施策にはどんなものがあるのか? (略)

図 15. タイトルが本文の話題と合っている例^[9]

(タイトルから抽出した単語)
施策 例 付 生産 性 向上 取り組む 始め
(本文中の TF 値 0.009 以上の単語)
生産 性 労働 向上 企業 業務 化 減少

図 16. 図 15 の文章から抽出した単語

cos 類似度を求めるには、タイトルから 1 語、本文から 1 語を選ぶ。すべての組合せで cos 類似度を求め、これらの平均をとる。タイトルから抽出された単語 A の cos 類似度の平均値を \overline{x}_A 、閾値 t において本文から抽出された単語を $(B_{t1}, B_{t2}, \dots, B_{tn})$ 、とすると計算式は次のように表される。

$$\overline{x}_A = \frac{1}{n} \sum_{i=1}^n \cos(\vec{A}, \vec{B}_{ti})$$

ただし、Web ページのタイトルと本文両方に単語が含まれている場合、検索エンジンの性質から検索結果に反映されやすくなる。そこで \overline{x}_A に 1 を加算してポイント付けを行う。タイトルから抽出された単語それぞれで \overline{x}_A が算出されるので、さらにこれらの平均値を求める。タイトルを文単位で適当であるかを判定するためである。タイトルから単

語が m 個抽出されたとき、 $(\overline{x}_1, \overline{x}_2, \dots, \overline{x}_m)$ の平均値を \overline{X}_t とすると計算式は次のように表される。

$$\overline{X}_t = \frac{1}{m} \sum_{i=1}^m \overline{x}_m$$

ここでは閾値を 5 つ用いているので、閾値ごとに \overline{X}_t が求まる。そのうち最も高い値を \overline{X}_{max} として、表 6 に示す基準に沿って判定を行う。

表 6. タイトル付けの判定基準

\overline{X}_{max} の値	判定
$\overline{X}_{max} \geq 0.5$	適切なタイトルである
$0.25 \leq \overline{X}_{max} < 0.5$	もう少し改善した方がよい
$\overline{X}_{max} < 0.25$	タイトルと本文が合っていない

表 6 の基準を決定するにあたって、インターネット上から、情報発信の初心者が書いた文章、検索結果の 1~10 位に表示された文章の 2 種類で \overline{X}_{max} の値を算出した。

図 15 の文章をシステムが解析した結果、 \overline{X}_{max} は 0.730 と算出されたので、タイトルは本文の話題と合っていると判定された。システムが出力した判定結果を図 17 に示す。

```

施策: 1.1668208174479122
例: 0.08699531996479401
付: -0.06809894575809057
生産: 1.1392769562796905
性: 1.1709436990249042
向上: 1.1634983397984446
取り組む: 1.1190435413199549
始めれ: 0.0627566782059148
単語cos類似度平均(TF=0.001): 0.7301545507854404
⇒適切なタイトルです。本文の内容と合っています。
cos_max_word=0.7301545507854404
    
```

図 17. 図 15 の文章における判定結果

図 18 はタイトル付けが不適切な文章から抽出した単語である。プライバシーの観点から文面は割愛する。図 18 に示す単語では、本文に使われている単語がタイトルに 1 つも含まれていないので \overline{X}_{max} の値が 0.079 と低くなり、図 19 に示す結果が出力された。

(タイトルから抽出した単語)
具体的 言い
(本文中に含まれる TF 値 0.009 以上の単語)
社内 朝礼 業務 化

図 18. タイトル付けが不適切である場合の単語例

```

具体: 0.08136094791384843
的: 0.1133492560340808
言: 0.04322395359094326
単語cos類似度平均(TF=0.001): 0.07931138584629083
⇒タイトルと本文が合っていない。再検討してください。
cos_max_word=0.07931138584629083
    
```

図 19. 図 18 の単語でシステムが判定した結果

3.4 検索結果にヒットしやすい単語の提案

文章から検索結果にヒットする可能性がある単語をシステムが提案する。提案する処理の流れを図 20 に示す。

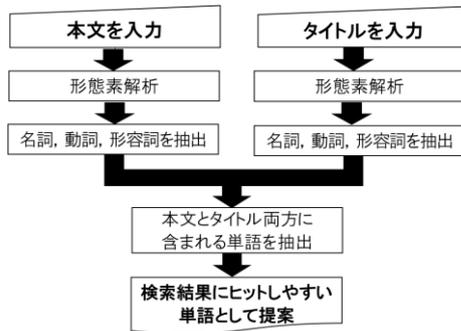


図 20. 検索結果にヒットしやすい単語を提案する流れ

本文とタイトルを形態素解析し、両方に含まれている単語をシステムが提案する。3.3 節の図 15 の文章をシステムに入力した場合を例にすると、まず図 21 に示す単語が抽出される。

(本文から抽出した単語)

生産性 労働 **向上** 企業 業務化 減少 効率 激化的
あり 力 生み出せる 高まっ 上げる ある 要素 み 注目
さ 国際 競争 人口 加速 する 少ない 成果 取り組み 必要
できる **施策** 取り組む

(タイトルから抽出した単語)

施策 例付 **生産性** **向上** **取り組む** 始めれ

図 21. 図 15 の文章から抽出された単語

ここで、本文から抽出した単語とタイトルから抽出した単語を比較し、両方に含まれている単語を探す。この場合、「施策」「生産」「性」「向上」「取り組む」が該当する。これら 5 つの単語を検索結果にヒットしやすい単語としてシステムが提案する。3.3 節の図 15 の文章をシステムに入力した結果を図 22 に示す。

◆検索されそうなキーワード◆

- ・ 施策
- ・ 生産
- ・ 性
- ・ 向上
- ・ 取り組む

図 22. 図 15 の文章を入力したときの単語提案

もし図 20 で示した処理で検索結果にヒットしそうな単語が 1 つも見つからなかった場合、検索結果に表示されない可能性があることをシステムが助言し、タイトルの変更や文章の見直しを促す。

3.5 タイトルに含めるべき単語の提案

検索キーワードとなる単語は 3.1 節で示した通り、TF 値が 0.01~0.03 の範囲に収まる傾向がある。もしそのような単語がタイトルに含まれていない場合、検索結果にヒットする可能性が低くなり、検索エンジン経由のアクセスが見込めない。そこで、図 23 に示す処理により、本文からタイトルに含めるべき単語を提案する。

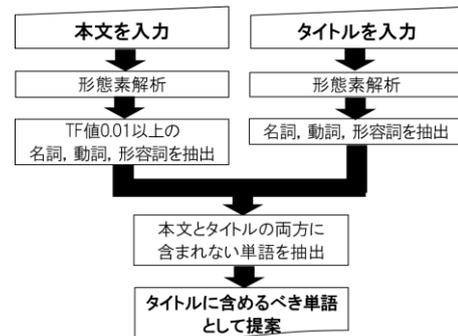


図 23. タイトルに含めるべき単語を提案する流れ

3.4 節の図 20 と類似したものであるが、異なる処理が 2 つある。1 つは本文を形態素解析した後に TF 値 0.01 を閾値として単語を抽出する点である。文章の軸を表し、かつ検索結果にヒットする可能性が高い単語を探す。もう 1 つは、本文とタイトルの単語を比較し、両方に含まれていない単語を抽出する点である。

3.3 節の図 15 の文章を例にする。「タイトルと本文は合っている」と判定されたが、さらに良いタイトルを考案できる可能性もある。そこで、図 23 に示した処理で単語を提案することで、タイトルの改善を行うことができる。実際にシステムが提案した結果を図 24 に示す。

◆この単語もタイトルに含めましょう◆

- ・ 労働
- ・ 企業

図 24. 図 15 の文章を入力したときの単語提案

タイトルに含めた方がよい単語として「労働」「企業」が提案された。これらの単語を含めてタイトルを再考すると、例えば「中小企業の方へ。労働者の生産性向上に取り組むための施策は考えていますか?」とすることができる。

4. 評価および考察

本システムの判定精度を評価するために、30 人に対してアンケートを行った。サンプル文章には実際にインターネットで掲載されている文章を 7 つ使用した。表 7 がアンケートに使用した文章の特徴である。

表 7. アンケートに用いた文章

No.	文字数	文章の特徴	備考
1	424 文字	日記感覚のブログ記事	
2	987 文字	検索結果 10 位以内の文章	図 15 参照
3	398 文字	要旨がわからないブログ記事	
4	1292 文字	検索結果 10 位以内の文章	図 9 参照
5	1027 文字	備忘録感覚のブログ記事	
6	277 文字	日記感覚のブログ記事	
7	1314 文字	検索結果 10 位以内の文章	

被験者にサンプルの文章全文と、サンプル文章をシステムに入力した際の出力結果を提示する。システムが文章に対して適切にアドバイスができていないかを、適切であれば○、おおむね適切であれば△、適切でなければ×と評価してもらった。

4.1 読みやすさ判定の評価

評価の結果を表 8 に示す。なお 3.1 節の表 1 の判定条件にいずれも当てはまらなかったものは、読みやすさに問題はないとして「特になし」と判定する。

表 8. 読みやすさ判定に対するアンケート評価結果

No.	判定結果	○	△	×
1	・無機質で人間味がない	23.3%	30.0%	46.7%
	・同じ単語の多用	20.0%	26.7%	53.3%
	・稚拙な印象がある	86.7%	10.0%	3.3%
2	・同じ単語の多用	16.7%	30.0%	53.3%
3	・(特になし)	23.3%	23.3%	53.3%
4	・(特になし)	93.3%	3.3%	3.3%
5	・単調な印象が出ている	56.7%	30.0%	13.3%
6	・くどい印象が強い	89.7%	6.9%	3.4%
7	・1文が長い	63.3%	16.7%	20.0%
	・無機質で人間味がない	36.7%	50.0%	13.3%

1000 文字以下の文章では△および×の回答率が高い結果となった。文字数が少なくなることにより、判定精度が低下する傾向がみられる。2 番目の文章については TF 値 0.03 を超える単語が存在したため、「同じ単語の多用」と判断されたが、半数以上が×と回答した。TF 値 0.03 程度では、人間が読むには特に違和感がなかったと考えられる。

4.2 話題の一貫性判定の評価

判定結果を被験者に提示したうえで文章を読んでもらい、被験者とシステムの判定が合致しているかを評価してもらった。評価の結果を表 9 に示す。

表 9. 一貫性判定に対するアンケート評価結果

No.	判定結果	○	△	×
1	特定の話題に絞っている	63.3%	23.3%	13.3%
2	特定の話題に絞っている	79.3%	10.3%	10.3%
3	余談が含まれている	60.0%	16.7%	23.3%
4	特定の話題に絞っている	93.3%	6.7%	0.0%
5	特定の話題に絞っている	66.7%	20.0%	13.3%
6	話題が混合し要旨が不明瞭	82.8%	17.2%	0.0%
7	特定の話題に絞っている	83.3%	16.7%	0.0%

一貫性の判定では、どの文章も○の回答率が 60%以上となった。文字数による影響も少なく、cos 類似度による判定は有効な手法だと考察する。

本システムでは、話題が混合していないかを判定するまでに留まっている。段落や見出しなど範囲を絞った判定手法を加えることで、ストーリー性の有無を判定が可能になるのではないかと考えられる。

4.3 タイトル付け判定の評価

判定結果を被験者に提示したうえで、記事のタイトルと本文を読んでもらい、被験者の評価とシステムの判定が合致しているかを評価してもらった。評価結果を表 10 に示す。

表 10. タイトル付け判定に対するアンケート評価結果

No.	判定結果	○	△	×
1	もう少し改善した方がよい	65.5%	27.6%	6.9%
2	一目で話題がわかる	50.0%	33.3%	16.7%
3	話題がまったくわからない	90.0%	6.7%	3.3%
4	一目で話題がわかる	93.3%	3.3%	3.3%
5	一目で話題がわかる	83.3%	13.3%	3.3%
6	話題がまったくわからない	89.7%	10.3%	0.0%
7	一目で話題がわかる	70.0%	23.3%	6.7%

一貫性判定と同じく cos 類似度を用いているので、文字数が判定精度に与える影響は少ないとみられる。実際の SEO 対策においては、検索者の目を引くタイトルを付けなくてはならない。意味の類似以外に、感情面での判定を取り入れることで、実際に人間が行う SEO 対策に近くなると考察する。

4.4 検索結果にヒットしやすい単語の提案の評価

被験者にシステムが提案した単語と、表 7 で示した文章を提示する。提案された単語で検索した際に、提示した文章が検索結果に出てくる可能性があるかを回答してもらった。なお、検索結果にヒットされそうな単語が見つからなかった場合も、その判定が適切であるかを評価してもらった。評価の結果を表 11 に示す。

1 番目と 2 番目の文章が○の回答率が低い。1 番目の文章は抽象度が強いキーワードが提案されてしまったことが原因として考えられる。2 番目の文章は、「取り組む」が検索ワードとして違和感を覚える人が多かったことが要因と考えられ、動名詞で提案する必要があると考えられる。

3 番目と 6 番目の文章は単語が 1 つも提案されなかった。3 番目の文章は、本文は話題を絞っているがタイトルが良くない特徴がある。単語を雑多に使用しており、TF 値で重要単語を抽出できなかった。6 番目の文章は、本文の要旨が不明瞭になっている特徴がある。文の要旨が不明瞭になっていなければタイトル付けも難しくなるため、単語を提案しない処理でも被験者は適当だと回答したと考えられる。

表 11. 検索結果にヒットしやすい単語の提案処理の評価

No.	提案された単語	○	△	×
1	息子	13.3%	30.0%	56.7%
2	施策 生産性 向上 取り組む	41.4%	48.3%	10.3%
3	(提案された単語なし)	53.3%	26.7%	20.0%
4	生産性 向上 マネジメント 管理職	73.3%	26.7%	0.0%
5	暗記 世界史	80.0%	20.0%	0.0%
6	(提案された単語なし)	89.7%	10.3%	0.0%
7	週 労働 制 生産性 向上 企業 支出 削減 メリット	76.7%	23.3%	0.0%

4.5 タイトルに含めるべき単語の評価

4.4 節と同様の方法で、適切な単語が提案されているかを評価してもらった。なお単語が提案されなかった場合は、必要な単語がタイトルにすべて含まれているとする。もし、ほかにも含めるべき単語があると思ったら△または×と回答してもらった。評価の結果を表 12 に示す。

表 12. タイトルに含めるべき単語の提案処理の評価

No.	提案された単語	○	△	×
1	授業 あり 英文 テスト 学習	34.5%	51.7%	13.8%
2	労働 企業	46.7%	36.7%	16.7%
3	社内 朝礼 業務 化	46.7%	30.0%	23.3%
4	仕事 やる気 する	65.5%	20.7%	13.8%
5	(提案された単語なし)	60.0%	33.3%	6.7%
6	する 注射 A 受け させ 有様 満喫	17.2%	31.0%	51.7%
7	(提案された単語なし)	70.0%	23.3%	6.7%

5. おわりに

本稿では人間に代わって機械が SEO の助言ができないかを検証した。SEO に特化した文章とは、検索結果に反映されやすく、かつ繰り返し読んでもらえる文章であると考えられる。このことから、SEO 対策に特化した文章校正システムを構築するに至った。SEO、自然言語処理、日本語の文章について追究すれば、判定精度の高いアルゴリズムを考案できる可能性がある。最終的に本システムがコンサルタントの代人として活用できれば、ビジネスの強い味方になることが期待できる。

謝辞

本研究の一部は、JSPS 科研費 16K00311 の助成を受けて行ったものです。

参考文献

- [1] 鈴木良治, “Web プロジェクトを成功に導く戦略的 SEO 思考”, 朝日新聞出版 (2018).
- [2] Christopher D. Manning, Hinrich Schütze, “統計的自然言語処理の基礎”, 共立出版 (2017).
- [3] 阿部紘久, “文章力の基本の基本”, 日本実業出版社 (2009).
- [4] “読点がいっぱい 【文章技術：句読点の打ち方】”, オルタナティブブログ, <https://blogs.itmedia.co.jp/editech/2011/12/post-c109.html> (2019/12/4 参照).
- [5] “稚拙な文章って言われた…5 つの事例と対策”, 記事作成代行屋, <http://xn--3kq3hlnz13dlw7bzic.jp/chisetsu-no-taisaku/> (2019/12/4 参照).
- [6] 染谷昌利, “小さな会社のための Web 担当者のためのコンテンツマーケティングの常識”, ソシム株式会社 (2015).
- [7] 朝日新聞単語ベクトル, https://cl.asahi.com/api_data/wordembedding.html (2019/11/16 参照).
- [8] “働き方改革「生産性向上マネジメント (管理職編)」⑩”, 井上労働法務相談所, <https://soudansyo.net/2019/12/686/> (2020/2/27 参照).
- [9] “【5 つの施策例付】生産性向上に取り組むには、何からどう始めればいいのか?”, d's JOURNAL, https://www.dodads.com/content/181120_productivity/ (2020/2/27 参照)