

タイトルの単語を本文で強調表示する読書支援技術の実験

丸山 大輔*

村田 真樹*

徳久 雅人*

馬 青**

*鳥取大学 工学部 知能情報工学科

**龍谷大学 理工学部 数理情報学科

*{s062057, murata, tokuhisa}@tottori-u.ac.jp

**qma@math.ryukoku.ac.jp

1 はじめに

近年はインターネットの普及にともない、電子文書を読む機会も増えてきている。電子文書の読書の効率化を計る技術の構築が望まれている。そこで、本研究では電子文書における読書支援に関する研究を行う。本研究では特に、タイトルに出現した単語を重要単語として本文中で強調表示することにより読書の効率を向上させることを目的とする。

本研究の手法の効果として以下の点を期待する。

- 読書速度の向上

タイトルに出現した単語が重要なことは種々の研究 [1, 2] により指摘されており、本手法ではその重要な単語を本文で強調表示することにより、本文のどの箇所にどういった内容の記述があるかを概観できる。よって、読者は自分の興味のある内容のみを選択的に読むことができ、読書時間を減少させることができると思われる。

このようにして、本研究で有効性を示すことができればタイトルの単語が重要な意味をもつことが再確認できる。よってこれにより自然言語処理の分野においてタイトルの単語を利用した研究手法の体系化にも繋がると考えられる。

本研究は以下の手順に沿って行う。

1. 強調表示方法の調整
2. 強調表示の人手による考察
3. 被験者実験

関連研究・関連技術を示す。タイトルにある単語が重要であるという特徴を利用した関連研究として、以下のものがある。内山ら [1] はテキストの自動要約に各種重要度を用いたものを比較しており、この中では

特にタイトルとの類似度に基づく方法が高精度であったとしている。村田ら [2] も情報検索の研究において位置情報を利用して(記事の先頭に出現するタイトルなどに重みを付ける)、高精度を出している。

文章の表示方法を変えることにより読者の読書支援を行う関連研究として、以下のものがある。西原ら [3] は文章の主題(テーマ)と各文の関連度を評価し、文章中の各文の表示方法を変更することで読者のテキスト内容把握に対して有効に働くことを確認している。タイトルに出現した単語を重要単語として本文中で強調表示するシステムを村田ら [4] が提供しているが、評価実験がなされておらず、有効性が確認されていない。

関連技術としてはウェブブラウザの検索機能などがある。これらの技術は広く普及・活用されており、強調表示が読者のテキスト作成・読解の支援方法として有効であることが確認できる。

2 読書支援の方法

まず、以下の手順で強調表示を行う。

1. タイトルを形態素解析する。
2. 1の結果から名詞と未知語である形態素を重要単語とする。
3. 本文を同様に形態素解析する。
4. 2の重要単語が3の結果と一致すればそれを強調表示する。

なお、形態素解析には ChaSen を用いる。強調表示は html タグを付与することにより実現する。重要単語の背景色を変更し、重要単語以外の語と差別化して読者の注意を引いて読みやすくする。本手法の出力の一例を図1に示す。

対ソ金融支援打ち出さず 英首相が書簡、海部首相も同意—ロンドン、サミット

七月のロンドン・サミット(先進国首脳会議)の議長を務めるメジャー英首相が六日、他の参加六カ国首脳に送った親書の中で、焦点の対ソ金融支援について、七カ国が協調して実施することはないとの方針を伝えていたことが明らかになった。これに対し、海部首相は七日、メジャー首相に返書を送り、この方針に同意するとともに、日本は金融支援を行わないことを伝えた。米国も金融支援は約束しない考えであることから、ソ連の期待している金融支援は昨年のヒューストン・サミットと同じく、共同歩調を打ち出さず、各国の判断にゆだねる公算が大きくなった。

関係筋によると、メジャー首相の提案は(1)ソ連経済が混乱している現状では、金融支援を大規模に行っても効果的ではない(2)英国としては金融支援を行う考えはない(3)サミットの場で金融支援を約束したり、支援策を取りまとめることはしない—というもの。

また、ゴルバチョフ・ソ連大統領の招待に関連して、サミットの公式会議終了後に七カ国首脳が大統領と協議する方針を明らかにしたうえ、大統領との協議では金融支援問題を議題としない考えを伝えている。

図 1. 本手法の出力の一例

3 強調表示が有効な例 (人手による考察)

今回は 45 件のテキストに関して強調表示が有効であるかどうかを手で考察し、どのような種類のテキストに関して有効であるかを調べた。結果として本手法が理論立てて有効であると説明できるものを 10 件見つけることができた。その 10 件は具体的に表 1 のように分類できる。

表 1. 強調表示が有効な例の分類

分類の種類	個数
A. ある事柄に関する詳細内容の把握	6
B. ある事柄に関する詳細記述箇所の把握 (ある事柄について詳細記述箇所の把握)	1
C. 複数の事柄について詳細記述箇所の把握	1
D. タイトルの (重要) 単語の補完	2

3.1 詳細内容の把握の例

図 2 は分類「A. ある事柄に関する詳細内容の把握」にあたる。

エジプトのアトロス・ガリ副首相を次期国連事務総長に任命—国連総会

【ニューヨーク3日松本照雄】第四十六回国連総会は三日午前十時(日本時間四日午前零時)から本会議を開き、安全保障理事会の勧告に基づき、全会一致の拍手で、エジプトのアトロス・ガリ副首相(69)を次期国連事務総長(任期五年)に任命した。中東アフリカ地域から事務総長が誕生するのは初めて。

ガリ氏は承認後の演説で「国連は国際平和を維持するだけでなく、平和をつくり出すために、機能を強化しなければならない。さらに南北の富の差を狭めるよう努力したい」と語った。

ガリ氏は年末で任期が切れるデクヤエル事務総長(ペルー出身)に代わって、来年一月一日から正式に第六代事務総長に就任、国際社会最大の組織のかけ取り役を務める。

ガリ氏は、エジプトでは少数派のコプト派キリスト教徒。パリ大学で国際法の博士号を取り、数々の国際機関で多くの場数を踏み、英語、仏語に堪能。夫人はユダヤ系エジプト人で、第三世界だけでなく、欧米にも受けがいいのが強みだ。冷戦の終結で、国連は地域紛争を平和解決するための調停者としてこれまでにない権威を持ち始めている。

民族主義の台頭が世界政治の火種になるうとしているだけに、外交官として定評のあるガリ氏の手腕への期待が高まっている。

図 2. 詳細内容把握の例

この強調表示を用いた記事は国連事務総長に関する事柄を伝える記事であるが、「事務総長」の強調表示が含まれる文を注目して抜き出すと表 2 のようになる。

表 2. 強調表示が出現した文

文 1	……を次期国連事務総長(任期五年)に任命した。
文 2	中東アフリカ地域から事務総長が誕生するのは初めて。
文 3	……任期が切れるデクヤエル事務総長(ペルー出身)に代わって……
文 3	……正式に第六代事務総長に就任……

このことから以下の事柄が容易に分かる。

- 国連事務総長の任期は 5 年
- 中東アフリカから国連事務総長が誕生するのは初めて
- 当時の国連事務総長はデクヤエル氏(ペルー)
- 次の事務総長で六代目となる(当時)

このようにタイトルに基づいた強調表示が有効な例の 1 つ目としては重要単語の詳細内容を把握しやすいということが挙げられる。

3.2 ある事柄に関する詳細記述箇所の把握

図 3 の記事について、イスラエルとパレスチナの問題についての記事であるが、長文であり全てを読もうとすると大変時間が掛かる。文全体でも「イスラエル」の強調表示が目立つが、唯一 5 段落目に「パレスチナ」の強調表示が集中しており、ここを重点的に読むことでパレスチナ側の意見を読み取ることができる。

- 米国の提案したパレスチナ暫定自治機関の受け入れを表明
- パレスチナ国家とヨルダンとの連合を進める
- イスラエルによる南部レバノン占領を非難
- 欧米はキリスト教徒で、この問題について無関心と西側を批判

以上のようなパレスチナ側の意見が短い時間で読み取ることができる。このような例も強調表示が有効であると思われる。

個別演説で対立浮き彫り イスラエル首相とパレスチナ代表―中東和平会議2日目

【マドリード31日日本特派員団】中東和平会議は二十一日、全体会議二日目に入り、イスラエルのシャミル首相演説を皮切りにアラブ、イスラエル各代表団が本格論議を開始、両陣営の主張の隔たりを浮き彫りにした。シャミル首相は、アラブ諸国が求める領土的妥協を一切拒否し、イスラエルの生存権承認を求め強硬演説を展開。これに対しヨルダンのジャビル外相、パレスチナのシャファイ代表、シリアのシャラ外相らが、平和と交換に占領地からの撤退と入植地の凍結をイスラエルに求めることを主張するなど、同日夕（日本時間一日未明）までアラブ各代表の演説が続き、二日目の会議を終了した。（7、9面に関連記事）

ペーカー米國務長官のあいさつの後、登壇したシャミル首相は、ユダヤ人の四千年の歴史を引用しながら、イスラエルの地がユダヤ民族の「祖国」であると強調。アラブ諸国がこのイスラエルの生存権を認めず、戦争によるイスラエル壊滅に失敗すると「イスラエルに対するボイコットやテロ攻撃で対抗している」と激しくアラブ諸国を非難した。

さらに「問題は領土ではない」と、アラブ側の最大の要求である占領地からの撤退をキツバリ拒否する姿勢を明確にし、その上で統一された地域社会の実現を希望、周辺アラブ諸国指導者をイスラエルに招待し、代わりに自らヨルダン、レバノン、シリア訪問の用意があると、一方的にアラブ側に譲歩を迫る強い内容の演説を行い、約三十分間の演説を終了した。

この後登壇したヨルダンのジャビル外相は「この会議がイスラエルの手前勝手な生存権に終止符を打たねばならない」と対抗。イスラエルの占領地からの撤退を求めた国連決議242に基づき、ヨルダン川西岸・ガザ両地区、ゴラン高原などイスラエル占領地からの撤退、平和のシンボルとして東エルサレムのアラブ主権の回復などを求めた。

またシャファイ・パレスチナ代表は米国の提案したパレスチナ暫定自治機関（政府）受け入れを表明し、パレスチナ国家とヨルダンとの連合に向け、主権確立の戦いを進めることを明らかにした。また、レバノンのブズ外相はイスラエル軍による南部レバノン占領を非難、シリアのシャラ外相も「イスラエルは占領地のパレスチナ人の人権を継続して侵害している。この事実は欧米では知られていない。なぜなら彼らはキリスト教徒だから」と西側記者を批判。ゴラン高原など占領地の返還を求めた。

アラブ側代表がイスラエルの占領地からの撤退と引き換えにイスラエルとの共存を受け入れ、これまでの「領土と平和の交換」を基本として演説を展開しているのに対し、シャミル演説は領土的妥協を一切拒否する強い調子で貫かれており、両陣営の主張する和平プロセスの違いと、溝の深さが改めて浮き彫りになった。

同日の会議は各代表団が一通り演説を終了、全体会議最終日の一日は再度イスラエル、アラブ各代表が演説を行う。

図 3. 集中記述箇所把握の例

3.3 複数の事柄についてそれぞれ記述されている箇所の把握の例

この記事は元銀行員の不正融資事件について報じている。

端田泰三・富士銀行会長が辞任 不祥事で引責 東海銀行、協和埼玉銀行も社内処分

元行員による不正融資事件が起きた富士、東海、協和埼玉の三銀行は三日、後員の引責辞任を含む行内処分をそれぞれ決めた。富士銀行の処分は端田泰三会長をはじめ、三人が辞任する厳しいもの。東海銀行、協和埼玉銀行でも事件に関連する部門担当の後員三人が辞任することになった。（10、11面に関連記事）

富士銀行の端田泰三会長は三日、一連の不祥事の責任を取って辞任した。赤坂支店の元課長らが同日起訴され、取締役会が決定した。端田氏は事件発生当時、経営の最高責任者だったことから責任を取った。

辞任後の記者会見で端田氏は「バブル経済の下で収益至上の風潮を作り、それが事件の土壌になった」と述べ、事件当時のトップとしての経営姿勢の問題から引責したことを明らかにした。

富士銀行は同時に、事件当時の赤坂支店長だった篠塚淳一取締役の辞任▽全従業員の報酬を十月から三〜六カ月間、一〇〜一五％カット▽などの役員処分を決めた。篠塚氏の後任の元赤坂支店長（現人事部参事役）の論旨解職など二十八人の行員について停職、減給などの処分を行った。

東海銀行は三日、新井永吉副会長（東京地区担当）が不祥事の責任を取って今月三十一日付で辞任することを決めた。新井氏は特別顧問に就く。また専務以上の役員五人の報酬を十月から六カ月間、二〇〜三〇％カットする処分も決めた。

一方、協和埼玉銀行も、根岸英治専務（事務本部長）と門田三郎取締役（前東京営業部長）の二人が三日付で辞任、役員十七人の報酬を五〜三〇％、一〜三カ月間カットすることを決めた。

図 4. 複数の事柄の記述の例

記事の中では富士銀行、東海銀行、また協和埼玉銀行の3銀行それぞれの処分内容が強調表示により分かりやすくなっている。以下に、記事から読み取れる3銀行の処分内容を示す。

- 富士銀行
 - － 端田会長の引責辞任
 - － 事件当時の赤坂支店長、篠塚取締役の辞任

－ 全従業員の報酬カット

- 東海銀行
 - － 新井副会長の引責辞任
 - － 専務以上の役員五人の報酬カット
- 協和埼玉銀行
 - － 岸根専務と門田取締役の引責辞任
 - － 役員十七人の報酬カット

このように、各銀行の個別の事柄について調べたい場合には有効である。

3.4 タイトルの単語の補完の例

新聞記事の見出し（タイトル）などは読者に一目で分かるように短く書かれているものが多い。そのような制限の中では単語数が限られてしまう。例えば以下の例などである。

天皇、皇后両陛下、あすタイへ出発 東南アジア3国歴訪へ

天皇、皇后両陛下は二十六日、十一日間の日程でタイ、マレーシア、インドネシア三国訪問に出発される。天皇の外国訪問は昭和天皇の訪米（一九七五年）以来十六年ぶり。在任中の天皇がアジアを訪問するのは初めて。各国で歓迎行事や文化施設、遺跡のご視察が予定されている。

各国でも第二次大戦で日本に侵攻された歴史を持つだけに、初の天皇訪問が国民の対日感情をいかに和らげ、今後の友好・親善関係を進める契機となるか注目される。

図 5. タイトルの単語の補完の例

この記事のタイトルでは「東南アジア3国歴訪」とあるが、タイ以外の国名は省略されている。そこで強調表示の「タイ」に注目すると本文中で「……タイ、マレーシア、インドネシア……」という文が見つかる。この記事は本文自体が短いため、強調表示の有効性が低いかもしれないが、本文が長くなるにつれて、強調表示により注目した単語の人手による検索速度が向上するのは他の例と同様である。

4 評価実験

4.1 実験方法

本手法が読書支援に対して、特に読書速度の向上に繋がるかを確認するための評価実験を行った。なお、実験に用いたテキストは毎日新聞91年データベースの記事を使用している。また今回の実験では被験者2人に対して、本手法を適用したものとそうでないものを3件ずつ、計6件与え、それを比較した。評価実験では、タイトルに出現した重要単語のうち、特に内容に関連があると思われるものを1つ選び、それを被験

者に本文中から全て探し出してもらい、その重要単語が出現した段落番号をすべて回答してもらう。その回答時間の平均を求める。実験に用いた実際の記事の一例を図6に示す。この記事では「食管法」という重要単語を本文中から探して、その段落を全て回答してもらう。

コメ関税化へ最終条件「税率700%、10年固定、食管法維持」—政府検討

関税貿易一般協定(ガット)新多角的貿易交渉(ウルグアイ・ラウンド)が大詰めを迎える中で、政府は二十三日、コメの関税化受け入れは避け難い状況との判断を強め、法的処理、条件などについて最終的な検討に入った。「受け入れ条件」では、コメの関税率を七〇〇%程度の高率で十年間固定し、輸入量をミニマム・アクセス(最低限の輸入枠)部分に実質的に限定、これにより政府によるコメ輸入の一元管理を維持し、食糧管理法を改正せずに関税化を受け入れるようにしていることが最大の特徴。食管法を改正しないことで、「コメ自給体制の骨格は不変」とし、国会決議との抵触を避ける狙いもある。(9面に関連記事)

検討中の受け入れ条件は(1)関税化した場合の税率を七〇〇%程度で十年間固定する(2)ミニマム・アクセスは当初国内消費量の三%とし、上限は五%に抑える(3)コメ生産農家への補償や生産性向上などへの補助金は十年間、新ラウンドの中で削減対象としない—など。米国と欧州共同体(EU)の農業分野での交渉で対象期間を当面、五年間にする案も浮上しているため、その場合には高率関税を五年間固定。その後の事後点検(レビュー)には応じるが、六年目以降、十年目の税率は削減幅を極力抑える。また、五年間での削減を認めるにしても、五〇〇%以下への低下には絶対に応じない方針。ミニマム・アクセスによる輸入量も三十万—五十万トンに絞り込めれば、国内生産農家への影響も軽微と判断。関税化を受け入れでも高率関税が維持できれば事実上、ミニマム・アクセス以上に輸入量は増えず、輸入数量制限を維持した部分開放とはほぼ同じ結果になるとみている。

この場合、輸入量が把握できるため、現行の小麦輸入と同じ政府によるコメ輸入の一元管理が可能で、量的管理を基本とした食管法の改正は不要になる。

関税化を受け入れると打撃を受ける国内生産農家を救済する必要が生じるため、政府は補助金の支出や生産性向上を目指した農業基盤整備事業への補助拡大を検討している。しかし現時点の新ラウンド農業交渉では、それら国内補助金も削減対象となる可能性が高く、政府が自由に支出できなくなる公算が大きい。このため政府は、関税化を受け入れる場合には、国内対策向けの補助金は削減対象から除外するよう、強く求めることとしている。

政府が条件付きながら、関税化受け入れの検討に着手したのは、新ラウンド農業交渉の中で、コメを含めた例外なき関税化の動きが大勢を占めるようになったことが大きい。ドングル・ガット事務局長が二十一日に提示した作業文書でも、「この文書は例外なしの関税化の概念をもとに作られている」と明記。日本は極めて苦しい立場に追い詰められた。

新ラウンドでコメの関税化受け入れを求める圧力が高まる中で、政府はこれがギリギリ譲れる最後の線と判断しているが、関税化受け入れと食管法維持が両立するの微妙な点もあり、法的な整合性について政府部内での検討作業を急いでいる。

図6. 実験に用いた記事の一例

4.2 実験結果

前節の方法により実験を行った。その結果は以下の通りである。

表4. 実験結果

問題	強調あり [秒]	強調なし [秒]
1	12.19	18.23
2	19.31	33.85
3	17.79	20.90
4	12.08	23.20
5	36.92	59.29
6	11.48	19.84
平均	18.295	29.218

タイトルに出現した重要単語の出現箇所の把握において、本手法による強調表示をした方が回答時間が短いことが示された。6件とも強調ありの方が回答時間が短く、有意差水準5%の二項分布にもとづく片側符号検定で有意に本手法の強調表示手法の方が出現箇所の把握において有効であることが確認できる。

4.3 考察

この実験で確認できる本手法の有効性としては、重要単語の本文中における出現箇所の把握速度の向上である。特に長文になるほどタイトルと本文の関連性が薄れ、タイトルの単語(重要単語)を本文中から探し出すことは困難だと思われる。よって、本手法により重要単語の出現箇所をより早く把握できるようになれば、読書支援方法として有効であることにも繋がるとと思われる。

5 おわりに

本研究では、まずタイトルの単語を本文で強調表示する読書支援技術を考察し、45件中10件の割合で本手法が有効に働く場合についての分類を行った。次に評価実験によると、本技術による重要単語の本文中での出現箇所の把握速度向上が確認できた。今後の課題として、重要単語出現箇所の把握速度向上が実際に読書速度向上に繋がっているかを確認する評価実験への発展が求められる。また、今回は4~5件に1件の割合で本手法が有効であると考察で述べたが、それ以外の本手法が有効でない場合には逆に本手法が読者にとって読みにくさの原因となっている可能性もある。本手法を実際にインターフェイスとして実装する場合にはユーザがオプションをにより本手法による読書支援を行う、または行わないを選択できるようにして極力ユーザにとって利便性を高めた形で提供したいと考えている。

参考文献

- [1] 内山将夫, 井佐原均. 自動要約のための文重要度の比較. 自然言語処理, Vol. 7, No. 2, pp. 91–116, 2000.
- [2] 村田真樹, 馬青, 内元清貴, 小作浩美, 内山将夫, 井佐原均. 位置情報と分野情報を用いた情報検索. 自然言語処理, Vol. 7, No. 2, pp. 141–160, 2000.
- [3] 西原陽子, 佐藤圭太, 砂山渡. 光と影を用いたテキストのテーマ関連度の可視化. 人工知能学会論文誌, Vol. 24, No. 6, pp. 479–487, 2009.
- [4] 村田真樹, 金丸敏幸, 白土保, 馬青, 井佐原均. 種々の重要表現強調表示ツールバーの開発. 言語処理学会年次大会発表論文集, Vol. 13, pp. 1042–1045, 2007.