

論文概要に対する色付きアンダーライン付与システムの運用・分析

松岡 有希 坂本 竜基 中田 豊久 伊藤 禎宣 武田 英明

国立情報学研究所 〒101-8430 東京都千代田区一ツ橋 2-1-2

総合研究大学院大学 〒240-0193 神奈川県三浦郡葉山町（湘南国際村）

株式会社国際電気通信基礎技術研究所 〒619-0288 京都府相楽郡精華町光台 2-2

北陸先端技術大学院大学 〒923-1292 石川県能美旭台 1-1

東京大学 人工物工学研究センター 〒277-8568 千葉県柏市柏の葉 5-1-5

E-mail: {m-yuki, takeda}@nii.ac.jp, {skmt, sito}@atr.jp, t-nakada@jaist.ac.jp

あらまし 本論文ではアノテーション形式の中でもアンダーラインに注目し、アノテーションによって付与された原文への2次情報（メタデータ）を使ってどのようなサービスに応用できるのかについて分析した。我々は明治大学の齋藤孝教授が提唱する三色ボールペン読書法を用いて Web 文書に色付きのアンダーラインを付与できるシステムを開発し、第19回人工知能学会全国大会で運用した。三色ボールペン読書法とは、客観的および主観的に重要な部分に色付きのアンダーラインを付与しながら読書をする方法である。分析の結果、ユーザが引いたアンダーラインは文章内の重要語を多く含むことや、ユーザ全体から見るとアンダーラインの色の効果を見出すことができないが、ユーザ個人に目を向けると色による個人の特性があることが分かった。これらの結果より、アンダーラインから得られる情報は、文書要約やユーザプロファイルの作成等に応用できることが分かった。

キーワード アノテーションシステム、個人知識抽出、重要文抽出、パーソナライゼーション

An Analysis of Web Annotation System Underlining Important Sentences of Abstracts with Colors

Yuki Matsuoka Ryuuki Sakamoto Toyohisa Nakada Sadanori Ito and Hideaki Takeda

National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430 Japan

The Graduate University for Advanced Studies, Shonan Village, Hayama, Kanagawa 240-0193 Japan

Advanced Telecommunications Research Institute International, 2-2 Hikoridai, Seikacho, Soraku-gun, Kyoto 619-0288 Japan

Japan Advanced Institute of science and technology, 1-1, Asahidai, Nomi, Ishikawa 923-1292 Japan

Research into Artifacts, Center for Engineering, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8568 Japan

E-mail: {m-yuki, takeda}@nii.ac.jp, {skmt, sito}@atr.jp, t-nakada@jaist.ac.jp

Abstract We analyzed the second information to Web pages given by underlining. We developed the system that is able to give the underline to the Web document by using the three colors ballpoint reading method that Prof. Takashi Saito at Meiji University advocated, and operated it in the 19th Annual Conference of The Japanese Society for Artificial Intelligence. The three colors ballpoint reading method is a technique giving the underline with color to objectively or subjectively important sentences. Result of analysis, the underline data involved important sentences and some personal characteristic. It will be able to apply to the document summary and making the user profile, etc.

Keywords annotation system, extraction of personal knowledge, extraction of important sentences, personalization

1. はじめに

アノテーションとは、古くから人間がおこなってきた原文へ二次情報を付加する行為である。情報が文書

の形態で流通し始めた時代からそれほど下ることなく、この行為は広く行われてきた。原文の解説、図解、翻訳、推敲、意見など目的は様々であるが、原文と付加

情報との関係性を空間的近さが担保するという表現上の制約と、原文に対して二次的な情報を持つことがアノテーションに共通する特徴と言える。一方、Webにおいては、その膨大なリソースの二次的利用を目指し、リソースに対するメタデータ付与とフィードの気運が高まりつつある。その原動力となっているのはメタデータの標準化(RSS)とBlogツールの普及である。しかし、メタデータとして記述される内容は、著者名や投稿日時といった単純なものに限定されており、主にリソースの配信のために使われているのが現状である。なお、メタデータはリソース(原文)に対する二次情報であることからアノテーションと呼ばれることがあるが、ここでのアノテーションとは、原文の著者を含むすべてのアノテーション行為をおこなう人間(アノテータ)が様々な目的で人為的に付与する二次情報を指し、機械的に付与される類の二次情報とは分けて考える。本研究では、Web利用者の膨大なマンパワーを背景としたアノテーションによるメタデータの流通を目標とする。アノテータの視点から記されたメタデータは、既存のメタデータだけでは叶えられない新しいサービスやコミュニケーション形態を生み出す可能性を秘めている。

現在、人為的に付与されるメタデータの代表として挙げられるのが、Folksonomyのタグである。Folksonomyはリソースに対して人々がタグ(キーワード)を付与することで、リソースの分類を共同で行うものである。リソースに付けられたタグや他人が付けたタグをたどることで、今まで知らなかったリソースを発見できることが目新しく、多くの人々がFolksonomyのサービスを利用している。これに対し、本研究ではメタデータの作成方法としてアノテーションを採用し、中でも人々が読書をする際に、文書に下線を引くという自然行為に注目する。下線を引くという行為は、文脈や個人の知識背景といった様々な"コンテキスト"を取り入れながら行われるため、下線が引かれた箇所には様々な情報が含まれている[2]。従って、これらの情報を使えば、新しいサービスに活かすことができるものと思われる。そこで本論文では、色付きのアンダーラインを付与できるシステムを開発し、第19回人工知能学会全国大会で運用して得られたデータを基に分析を行った。

以後、第2章において、アノテーションの先行研究や本研究の位置づけについて述べる。第3章では、色付きアンダーライン付与システムについて述べ、第4章ではシステムの運用で得られたデータの分析結果について述べる。第5章でまとめを行い、今後の課題について述べる。

2. アノテーションについて

アノテーションは、人間がおこなってきた原文へ二次情報を付加する行為である。アノテーションの形式としては、下線やハイライトマーカー、アスタリスク、囲み文字、余白へのメモ等いろいろなものがある[1]。その中でも本研究では、下線を引くという行為に着目する。

2.1. 下線引きについて

[1]によると、大学の教科書におけるアノテーション行為を調べた結果、人は他で引用するために重要な文章を記録したり、後で見直したりするために下線を引き、線の色は情報の種類をコード化するために使う。また、学習目的で文章を読む場合、下線引きは読み手が重要だと考える情報を探し出す探索・選択過程であるといえる[3]。認知、情意器官がどのようなきっかけによって探索・選択過程と結びつくのかについては依然はっきりしていない。しかし、これらのきっかけは個人によって異なるものであり、また慣用的なものであるため、それぞれの場合によって異なるということでははっきりしている[4]。従って、原文の文章を理解して下線を引く人もいれば、あらかじめ持っている知識を基に下線を簡単に引く人もいる。一方で、ただの精神運動活動として下線を引く人もいる。探索・選択過程に成功するかどうかは、下線を引いているときの認知的処理の深さおよび量によるといえる[5]。

2.2. Web アノテーション付与システムに関する先行研究

これまでに、Webページに対してアノテーションを付与するシステムは多く開発されている。これはWWW上で任意のWebページに対してユーザが意見を書き込んだり、ユーザ同士で意見交換をしたりと、ページの作者とその読者、あるいは読者間同士といった双方向による知識共有の需要が高まったためである。

iMarkupは、Webブラウザを通じてWebページにマウスで絵を描いたり付箋を貼り付けることができる。また、iMarkupでwebページに描いた絵などのデータを添付ファイルとしてメール送信する機能も備えている。一方、iMarkupのように自由に描画できるアノテーション機能はなく、コメントのみを付与することができるシステムが多々ある。ComMentor[6]やCoNote[10]、CritLink[8]、Third Voice、Yawas[7]は、webブラウザを通して、webページの一部を特定し、付箋のようにコメントを付与できる。ComMentorやCritLink、Third Voice、Yawasでは、コメントはそれ自身がドキュメントであるとし、ユーザはコメントに対してコメントを追加することができる。これにより、オリジナルのweb

ページに対してスレッド形式でディスカッションすることができる。このように、様々な Web アノテーション付与システムの実現により、ユーザ間で知識共有をしたり、効率的な知識の交換が可能になった。本研究では、アノテーションの形式として下線を採用していることが大きく異なる。

Annotea[9]は、上記のものと同様に Web ブラウザを通してコメントを付与ことができ、コメントは RDF フォーマット[11]のメタデータとして保存される。これは Semantic Web[12]の実現に向けて、メタデータの作成を支援する目的があるため、将来的にはこのメタデータを使って新たなサービスを提供するものと思われる。メタデータを再利用して新しいサービスを提供することは、本研究の目的と同じだが、コメントではなく下線引きを採用するところが異なる。

3. 色付きアンダーライン付与システム

本章では、2005年6月15日(水)から6月17日(金)に開かれた第19回人工知能学会全国大会で、学会支援システムの一環として開発した色つきアンダーライン付与システムについて述べる。学会支援システムとは、学会会期前後に Web を介して利用できる大会支援 Web システムのことで、複数のシステムから構成されている。学会参加者(発表者、共著者)が自由に使えるようになっており、参加者は ID を使って大会支援 Web システムにアクセスできる。

3.1. システム概要

色付きアンダーライン付与システムは、HTML 形式の発表論文の概要文書(297個)に対して、Web ブラウザを通して色付きの下線を付与できる Web ベースのシステムである。

3.1.1. 下線引きの機能

アノテーション機能として明治大学の齋藤孝教授が著書等において提唱している三色ボールペン読書法[13]による下線機能を提供した。三色ボールペン読書法は、客観的にとても重要だと思う箇所は赤色で、客観的にまあ重要だと思う箇所は青色で、主観的に重要だと思う箇所は緑色で下線を引ながら読書をする方法である。こうすることで、集中しながら読書ができるようになり、要約力やコメント力が向上すると齋藤教授は述べている。本システムでは、この三色ボールペン読書法と同じように、客観的または主観的に赤・青・緑の三色を使って下線を引くことができる。

3.1.2. システムの操作

システムの利用者が、論文概要に下線を引くときの

操作は下記のとおりである。

1. 下線を引きたい箇所をマウスカーソルでなぞり、指定されたボタンをクリックする(図1)
2. 下線の色を選択して、“色線を引く”というボタンをクリックする(図2)
3. 色線が引かれた画面が表示される(図3)

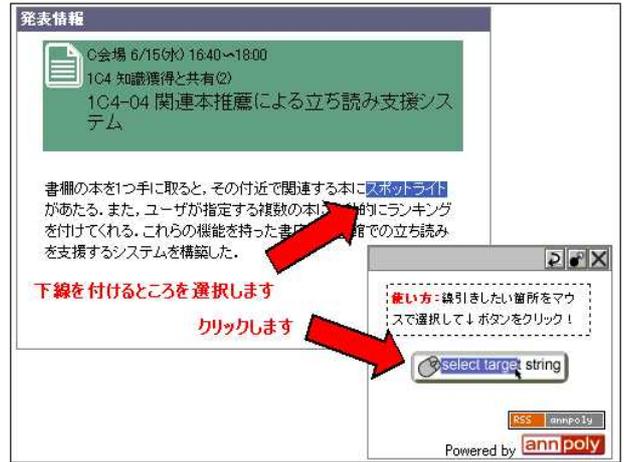


図1 下線を引く箇所を選択する画面

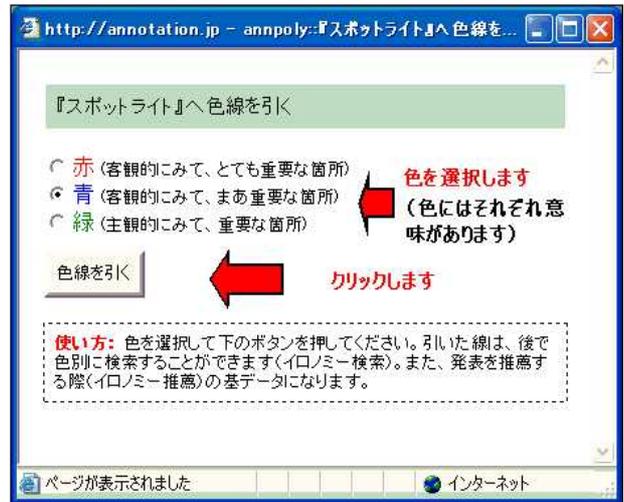


図2 下線の色を選択する画面

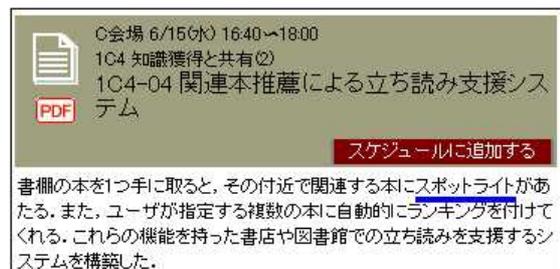


図3 色線を引いた後の画面

3.2. 運用結果

大会支援 Web システムの利用者は 175 人おり、そのうち色つきアノテーション付与システムを使用したのは 27 人だった。すなわち、15.4%の利用者がいたことになる。また、論文概要に引かれた下線に関して表 1 にまとめた。下線が引かれた論文概要は 88 個あり、下線の本数は 265 本で、三色の中では青線の数が一番多かった。論文概要はそれぞれ 50 字から 250 字の間で書かれており、平均文字数は 140 文字だった。下線はそれぞれ 2 字から 33 字の間で書かれており、平均文字数は 14 文字だった。

表 1 下線について

下線が引かれた論文概要の数	88 本
下線の本数	265 本
赤線の本数	83 本
青線の本数	101 本
緑線の本数	81 本
論文概要の平均文字数	140 字
下線の平均文字数	14 字

4. 分析

システムの運用によって得られた下線のデータを基に分析を行った。

4.1. はじめに

システムの運用によって得られた下線に関するデータは以下のものがあり、分析で使用できるのは、下記のデータと論文概要のデータである。

- 下線が引かれた日付
- 下線を引いたユーザ ID
- 下線が引かれた論文概要 ID
- 下線の色
- 下線が引かれた文章

ここでは、分析対象として論文概要および下線文における名詞に注目した。名詞を採用する理由は、下線が引かれた文章内で一番多く使われていた品詞だからである(表 2)。

また、今回アノテーションの対象となった文書は人工知能に関する研究発表の論文概要であるため、専門用語が多く使われていた。そこで、文書から名詞を抽出する際、専門用語(キーワード)自動抽出システム[14]を使用し、294 個の論文概要の文書(3 個の論文概要は英文のため排除した)と下線が引かれた文章からキーワードを抽出した。以後、このキーワードを使って分析を行うこととする。

表 2 下線に含まれる品詞

	名詞	助詞	動詞	助動詞	未知語	副詞
赤線	296	99	73	14	14	10
青線	306	88	58	17	19	13
緑線	232	76	55	11	8	12
全下線	834	263	186	42	41	35

4.2. 下線について

本節では、学会参加者(以後、ユーザと呼ぶ)が引いた下線の特徴について、およびユーザが客観的に重要な箇所と主観的に重要な箇所とで下線の色を区別して使っているのかについて調査した。

4.2.1. 下線文の特徴について

まず初めに、対象文書となった論文概要と下線が引かれた文章とで比較を行い、下線にはどのような特徴があるかについて調べた。ここでは、論文概要に含まれるキーワードと下線文に含まれるキーワードの差を調べるために、tfidf[15]を用いて検証した。tfidf は下記のような特徴がある。

- tfidf 値が高い語は、対象文書内で出現頻度が高く、他の文書には現れにくいので、対象文書内における重要語と言える
- tfidf 値が低い語は、対象文書内での出現頻度が低く、他の文書に頻繁に出てくる語であるため、一般語である可能性が高い

分析にあたり、図 4 のように処理を行った。論文概要と論文概要に引かれた下線の文章からキーワードを抽出し、キーワードに対する tfidf 値 $t_{ij}(1)$ を求める。

$$t_{ij} = tf(i, j) * \log\left(\frac{N}{df(j)}\right) \quad (1)$$

$tf(i, j)$ は論文概要 A_i におけるキーワード w_j の出現頻度数を、 $df(j)$ は w_j が含まれる論文概要の出現頻度数を表す。 N は、論文概要の総個数を表し、今回は 294 個である。このように求められた tfidf 値は論文概要ごとに上下限値にばらつきがあるため、値を基準化(2)する。

$$z_j = \frac{t_{i,j} - \bar{t}_i}{s_i} \quad (2)$$

$$\bar{t}_i = \frac{t_{i,1} + \dots + t_{i,j}}{j}$$

$$s_{i'} = \sqrt{\frac{(t_{i,1} - \bar{t}_i)^2 + \dots + (t_{i,j} - \bar{t}_i)^2}{j}}$$

論文概要 A_i におけるキーワード $w_j (1 < j < n)$ に対応する tfidf 値の t_{ij} の平均値が \bar{t}_i で、標準偏差が $s_{i'}$ である。

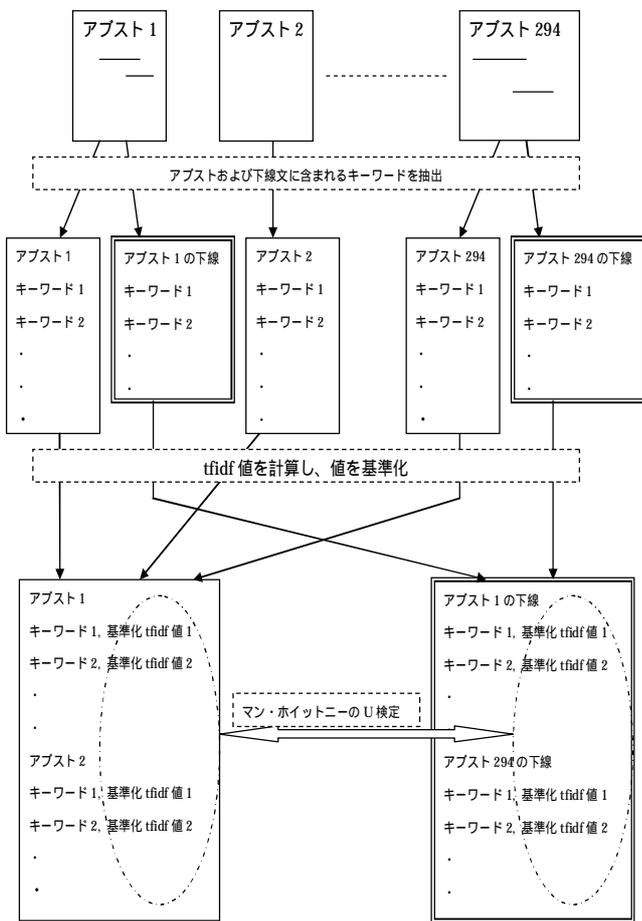


図 4 分析処理手順

表 3 は全ての論文概要と全ユーザが引いた下線文に含まれるキーワードに対する基準化 tfidf 値群の平均値と分散とそれぞれに含まれるキーワード数を示している。全論文概要に含まれるキーワードの基準化 tfidf 値群と全ユーザが引いた下線文に含まれるキーワードの基準化 tfidf 値群とで平均値の差がないかを調べるために、マン・ホイットニーの U 検定を行った。

表 3 全論文概要と全下線文に含まれるキーワードの基準化 tfidf 値の平均値と分散とキーワード数

	全論文概要	下線文
平均値	-1.38719E-05	0.444256283
分散	0.926784507	1.030486513
キーワード数	3879	394

検定の結果、P 値は $8.93E-14 (<.05)$ となり、両者に有意差があるという結果が得られた。図 5 は全論文概要に含まれるキーワードに対する基準化 tfidf 値のヒストグラムで、図 6 は全下線文に含まれるキーワードに対する基準化 tfidf 値のヒストグラムである。図 5, 6 において、基準化 tfidf 値のデータ区間が 1.6 から 3 の間のキーワードの出現頻度を見比べると、論文概要に含まれるキーワードの出現頻度よりも下線が引かれた文章に含まれるキーワードの出現頻度のほうが高いことが分かる。これは、ユーザに対して重要だと思ふ箇所に下線を引いてもらったため、論文概要内の重要語を選ぶ頻度が高かったものと思われる。

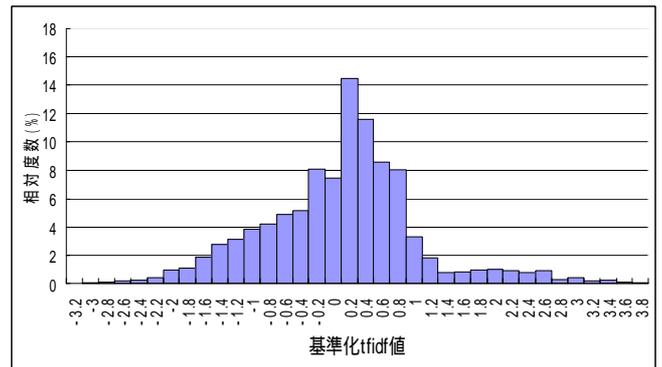


図 5 全論文概要に含まれるキーワードに対する基準化 tfidf 値のヒストグラム

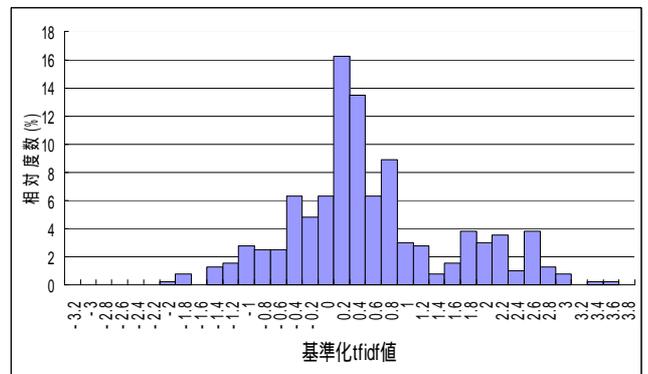


図 6 全下線文に含まれるキーワードに対する基準化 tfidf 値のヒストグラム

4.2.2. 主観と客観の区別について

システムでは、ユーザに対して、客観的に重要だと思う箇所には赤色と青色の下線を引いてもらい、主観的に重要だと思う箇所には緑色の下線を引いてもらった。従って、ユーザは論文概要の内容に対して主観的に重要か客観的に重要かを明示的に示した上で下線を引いたものと思われる。本節では、ユーザが客観的に重要か主観的に重要かを区別して下線を引いているのかについて調べた。

図7は、全ての赤・青・緑線に含まれるキーワードの基準化 tfidf 値の相対度数を積み上げ表示したものである。表4は、各色の下線に含まれるキーワードの基準化 tfidf 値の平均値と分散、およびそれぞれに含まれるキーワード数を示している。赤・青・緑線に含まれるキーワードの基準化 tfidf 値群の平均値に差があるかどうかを、クラスカル・ウォリス検定で調べた。

その結果、P値は0.2936 (> .05)となり、群間に有意差はないという結果が得られた。従って、ユーザ全体で見ると、tfidfで重み付けしたキーワードを色で区別していなかった、あるいはできていなかったと言える。

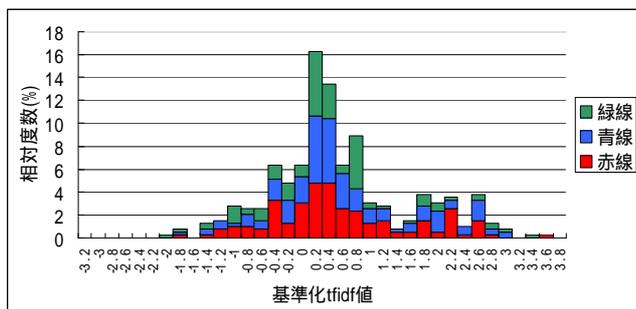


図7 赤・青・緑線に含まれるキーワードに対する基準化 tfidf 値のヒストグラム

表4 各色の下線に含まれるキーワードの基準化 tfidf 値の平均値と分散とキーワード数

	赤線	青線	緑線
平均値	0.43785473	0.527936226	0.307501387
分散	0.989936648	1.044014834	0.99108555
キーワード数	145	145	104

4.3. ユーザについて

本節では、ユーザ個人についてどのように下線を利用したのかについて調べた。本システムは、27人のユーザが使用した。図8は、ユーザが下線を引いた数を示しており、各色の線の数を積み上げグラフにしたものである。27人中、線を引いた数が多かった上位8人を載せている。

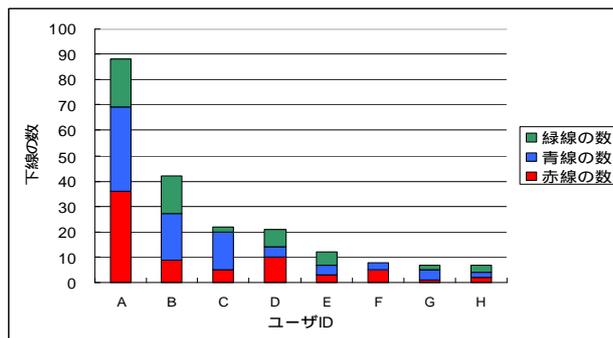


図8 ユーザが引いた下線の数

今まで使用してきた tfidf は、文書内におけるキーワードの重要度を求めるためのもので、それが客観的あるいは主観的に重要を示すものではない。よって、ある文書において、tfidf 値の高いキーワードを客観的に重要と思う人もいれば、主観的に重要だと思う人がいるだろう。そこで、本節ではユーザ個人に注目し、キーワードの重要度によって色の使い分けを行っているかどうかについて考察する。

図9から14は、各ユーザの下線に含まれるキーワードの基準化 tfidf 値のヒストグラムである。

ユーザAは、基準化 tfidf 値の高い区間である1.4から3.4の区間において、赤と青色の線を緑色の線よりも多く選んでいる。逆に基準化 tfidf 値が小さいほど(-0.8よりも下のデータ区間)緑色の線の割合が若干多い。ユーザAは、文書内の重要語を客観的に重要な語としてとらえ、一般語を主観的な語としてとらえる傾向があるとと言える。

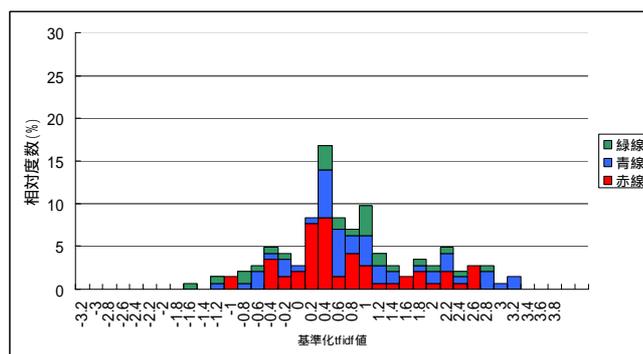


図9 ユーザAが引いた下線に含まれるキーワードに対する基準化 tfidf 値のヒストグラム

ユーザBは、青色の線が他の色よりも少し多く選ばれている。また、データ区間が0.2から0.4の間のキーワードの出現頻度が飛び抜けて高いのが特徴である。これは、全論文概要に含まれるキーワードの基準化 tfidf 値の出現頻度が一番多いデータ区間と同じであり

(図5), ユーザ B は重要度が中程度のものを好むことが分かる。もう1つの特徴は、基準化 tfidf 値の一番高い値から低い値まで、広範囲に渡って緑色の線を選んでいることである。ユーザ B が主観的に重要なキーワードを選ぶときに、tfidf のキーワードへの重み付けは何の影響も与えていないことを示している。

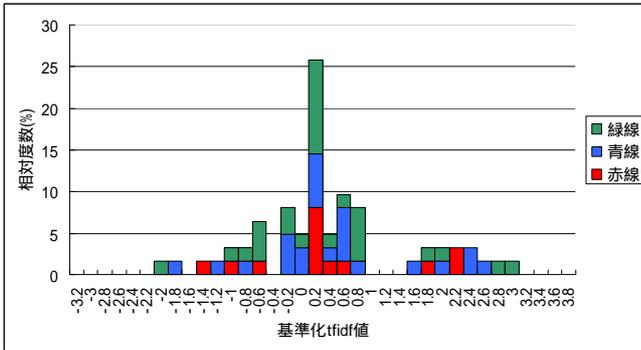


図 10 ユーザ B が引いた下線に含まれるキーワードに対する基準化 tfidf 値のヒストグラム

ユーザ C は、赤と青色の線ばかり選んでいるのが特徴である。また、データ区間が 2.2 から 2.4 の間のキーワードの出現頻度が他のユーザよりも高い。従って、ユーザ C は文章を読むときに、客観的な視点で読む傾向があることが分かる。しかし、単にユーザ C が三色ボールペン方を使いこなしていないだけ、ということも考えられる。これは長期的に観察しなければ分からない問題である。

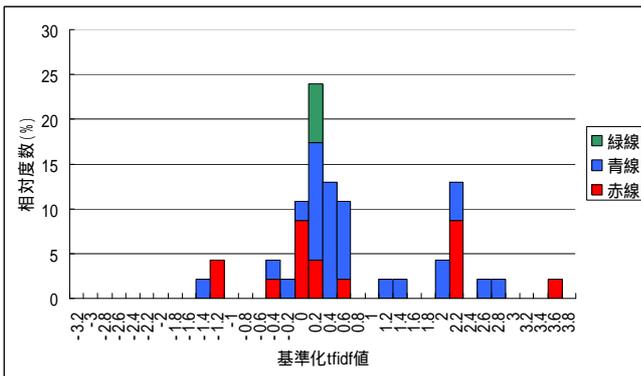


図 11 ユーザ C が引いた下線に含まれるキーワードに対する基準化 tfidf 値のヒストグラム

ユーザ D は、赤色を一番多く選んでいる。もとの論文概要に含まれるキーワードの基準化 tfidf 値の出現頻度が高い区間である 0.2 から 0.4 の区間の出現頻度が他のユーザと比べても低い。また、どの色に関しても幅広いデータ区間で選ばれている。従って、ユーザ

D にとって客観的および主観的に重要なキーワードというのは、tfidf の重み付けとは関係がない、と言える。

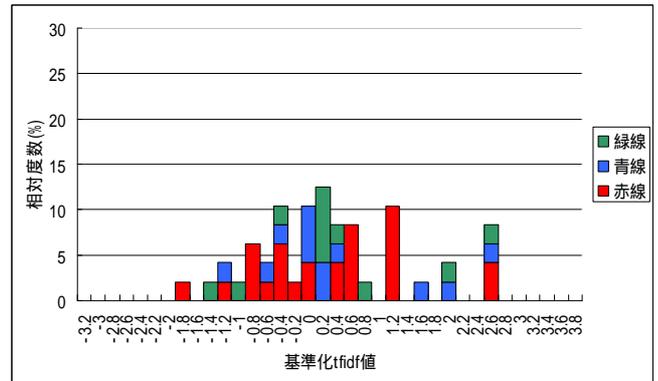


図 12 ユーザ D が引いた下線に含まれるキーワードに対する基準化 tfidf 値のヒストグラム

ユーザ E は、赤や青色の線よりも緑色の線のほうが多い。基準化 tfidf 値の出現頻度が一番高いデータ区間は、0.8 から 1 の区間で、他のユーザと比べて高いのが特徴である。赤色と緑色の使い方が似ているため、ユーザ E は客観的および主観的に重要な箇所を区別していないことが分かる。

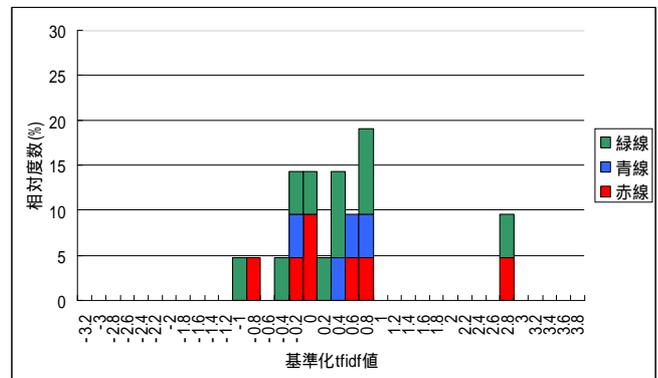


図 13 ユーザ E が引いた下線に含まれるキーワードに対する基準化 tfidf 値のヒストグラム

ユーザ G は、基準化 tfidf 値の出現頻度の一番高い区間が他のユーザと比べて低い。これは、ユーザ F が長い下線を引く傾向があるためである。全ユーザの下線文に含まれるキーワード数の平均値を示した図 15 によると、ユーザ G は 2 番目に高いことが分かる。論文概要の内容を表す特徴的な語を 1 つ選ぶのではなく、一般語から構成される長い文章を好んで選んでいるため、赤色がデータ区間の低いところにしか出現していないことが、他のユーザと大きく異なる。

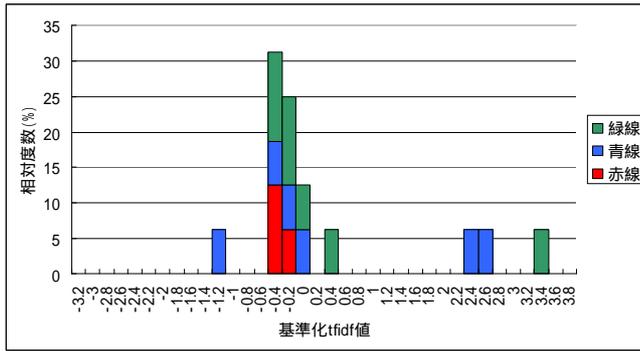


図 14 ユーザ G が引いた下線に含まれるキーワードに対する基準化 tfidf 値のヒストグラム

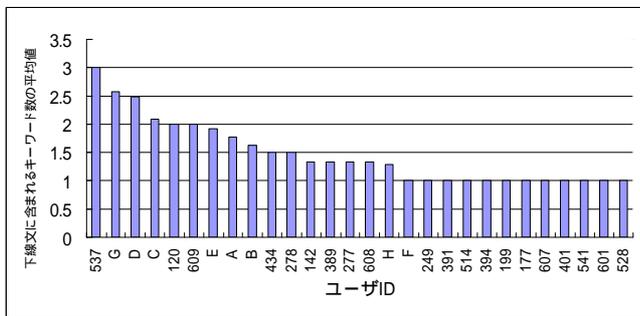


図 15 下線文に含まれるキーワード数の平均値

このように、ユーザごとに線の色やキーワードの選び方は様々である。また、下線の長さによって、キーワードの基準化 tfidf 値の色別の分布が異なることが分かった。このユーザごとの違いを特徴としてとらえれば、ユーザプロフィールの作成に活かすことができそうである。

5. まとめと今後の課題

本論文では、色つきアンダーライン付与システムを開発・運用した結果から得られたデータの分析を行った。その結果、ユーザが引いた下線は、論文概要内の重要語を多く含むという結果が得られた。これは、「重要な箇所を引く」という行為がユーザ間で共有できたことを示している。従って、下線文を文書内の重要箇所として利用することができ、文書要約のサービスへの応用が期待できる。

一方、ユーザ全体で見ると下線の色差を見出すことができないという結果が得られた。しかし、ユーザ個人に目を向けると、下線の色やキーワードの選び方がユーザごとに異なることが分かった。これは、個人の特徴がユーザが大勢集まることによって埋もれてしまったことを意味する。この結果より、各ユーザの下線

データを使って、Web 検索・推薦の際に用いるユーザプロフィールを作成することができ、情報獲得におけるパーソナライゼーション化に応用できるものと考えられる。

今後は、より多くのユーザに使用してもらえるようにシステムを改良し、メタデータの量を増やすことを課題とする。さらに、下線から得られたユーザ個々の特性をモデル化し、個人に適応したサービスの提供を目指していきたい。

文 献

- [1] Marshall, C., Annotation: from paper books to the digital library, in Proceedings of the ACM Digital Libraries '97 Conference.
- [2] Schilit B. N., Golovchinsky, G and Price M. Beyond Paper: Supporting Active Reading with free-form digital ink annotations. Proceedings of CHI'98
- [3] Gyunn, S. Capturing reader's attention by means of typographical cueing strategies, Educational Technology, 1978.
- [4] Blanchard, J. and Mikkelsen, V. Underlining performance outcomes in expository text. Journal of Educational Research, 80: 197-201
- [5] Morris, C., Bransford, J. and Franks, J., Levels of processing versus transfer appropriate processing, Journal of Verbal Learning and Verbal Behavior, 16 : 519-533
- [6] Roscheisen, M., Mogensen, C., and Winograd, T. Shared Web Annotations as a Platform for Third-Party Value-Added, Information Providers: Architecture, Protocols, and Usage Examples, Technical Report CSDTR/DLTR (1997), Stanford University.
- [7] L. Denoue and L. Vignollet. An annotation tool for web browsers and its applications to information retrieval. In In Proceedings of RIAO2000, Paris, April 2000.
- [8] Ka-Ping Yee. CritLink: Better Hyperlinks for the WWW, 1998.
- [9] Marja-Riitta Koivunen, Eric Prud'Hommeaux, Ralph R. Swick, Annotea: An Open RDF Infrastructure for Shared Web Annotations, in Proc. of the WWW10 International Conference, Hong Kong, May 2001
- [10] J.R. Davis and D.P. Huttenlocher. Shared annotation for cooperative learning. In Proceedings of the Computer Support for Cooperative Learning Conference, 1995 .
- [11] Lassila O., Swick R. (Editors), Resource Description Framework (RDF) Model and Syntax Specification, World Wide Web Consortium, 22 February 1999.
- [12] Berners-Lee, T., Hendler, J. and Lassila, O. The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities, Scientific American, May 2001.
- [13] 齋藤 孝, 三色ボールペン情報活用術, 角川書店, ISBN:4047041351 (2003)
- [14] 中川裕志他, 「専門用語(キーワード)自動抽出システム」, <http://gensen.dl.itc.u-tokyo.ac.jp/>
- [15] Salton, G.: Developments in automatic text retrieval, Science, Vol. 253, pp. 974-980 (1991).