

思わず読みたくなるタイトルを文書に付与するシステムの開発

Development of the system which gives a title to come to read involuntarily to a document

土井 祥吾
Shogo Doi

1 研究概要

1. 作りたいもの
思わず読みたくなるタイトルを文書に付与するシステム
2. 誰が使うか
ブログをしている人である。また、SNS (facebook, Twitter など) による情報配信者 (Webライター) である。
3. どのような場面で使うか
情報発信をする人が記事を作成した後に、システムを利用する。
4. どのように使うか
情報配信者とは記事を書くプロではなく、老若男女、だれもができる一つの職であると考え。ゆえに、誰が、どこで、いつ使うか特定できないので、誰もが、どこでも、いつでも使えることを前提とする。一度記事を作成し、その記事の内容を解析し、文書にふさわしいタイトルを付与する。文書を解析しタイトルを付与するときに、必ずしも解析結果で生成されたタイトルが、思わず読みたくなるタイトルかどうか分からないので、2, 3 個候補となるタイトルを生成する。
閲覧者によって、そのタイトルが思わず読みたくなるタイトルなのかが違って来る。例えばお年寄りの場合、わかりやすいタイトルを付与したほうが閲覧されやすいと考え、若者の場合は、ネット用語を用いたタイトルなどのほうが閲覧されやすいと考える。こういった場合も考慮し、候補としていくつかの形のタイトルを生成する。
5. 何を材料として作るか
ブログのタイトルと閲覧数と購読者数
情報検索サイトのトップニュースタイトル

1.1 どういう記事に対してこのシステムが扱われるかを考える。

ブログの記事に対してこのシステムを利用する。ブログ記事を対象とした理由は、「ブログ」というメディアによって、記事を書いてタイトル付けをするということが圧倒的に多いからである。

タイトルをつける人が多いということは、このシステムを作成するにあたって、素材となる記事も多くなり、このシステムも多くのユーザに利用されるからである。

1.2 人の心理と記事のタイトルとの関係を調べる。

人には無意識のうちに心理現象が働く。思わず読みたくなるタイトルを生成するには、人の心理を上手く利用することで良いタイトルが生成できると考える。心理現象とは多く存在し、その中でタイトルと関係する現象をいくつかあげる。

- 認知的不協和
認知的不協和とは、個人のもつある認知と他の認知との間に不一致・不調和が生じること。その結果、不協和を解消あるいは低減しようとして行動や態度に変化が起こる現象である。
例えば、「タバコを吸うとガンになるリスクが高くなり長生き出来ないのでやめなさい」と言った時に、喫煙者は「タバコを吸いたい」けれども、「タバコは体に悪いのでやめた方がいい」という不調和が生じる。その不調和を何とか解消したいと思い、「タバコを吸っても長生きしている人がいる」や「タバコを吸わないとイライラして逆に体に悪い」という答えを出し、不調和を解消していこうとする。
例タイトル「ヘビースモーカーが100歳まで生きる方法」
解説：ヘビースモーカーが長生きできるという違和感に強く気になってしまう。
- カリギュラ効果
カリギュラ効果とは、禁止されると、かえって気になり余計にその行為をやってみたくなる心理のことである。
例えば、「絶対にこの封筒はあけないで」と言われると、かえって気になって開けたくなくなる心理現象である。
禁止と言われたものにこそ、人は強く惹かれるという特性である。(否定命令効果)
例タイトル「羊を数えるな！すぐに寝れる快眠方法！」
解説：寝るときに人は羊を数えるという事を知っている。それをするのが当たり前という常識を否定命

令することで人は強く惹かれる。

補足：否定されたことをやりたくなるように促すことではなく、当たり前だという概念を否定命令し、その違和感で気になるのである。

● リフレーミング

リフレーミングとは、ある枠組みで捉えられている物事を枠組みをはずして、違う枠組みで見ること。また否定的な考え方に、肯定的な考え方を与えることである。

例えば、試験で残り時間が 15 分あった場合、悲観的に考えた人は「もう 15 分しかない」と思う。逆に楽観的に考えた人は「まだ 15 分もある」と思う。

リフレーミングを生かすなら、悲観的な考え方を楽観的に変えて提示することで魅かれるタイトルを生成する。

例タイトル「まだ 1 か月もある！東大に合格する 10 のポイント！」

解説：もう時間がない、という悲観的な考え方をまだ時間がある、という楽観的な考え方に転換している。

● ツァイガルニック効果

ツァイガルニック効果とは、人は達成できなかった事柄や中断している事柄のほうを、達成できた事柄よりもよく覚えているという現象である。

例えば、「4 時間以内に 42.195 キロを完走する」という目標を掲げた人がマラソンを走るとする。この場合、4 時間以内に走りきれた時と、走りきれなかった時とでは、後者が後々印象として大きく頭に残るとされている。

(追記) 違う見方をしてみる。ツァイガルニック効果とは、完成されたものより、未完成のものに興味をひかれることである。

例えば、ドラマはいつも中途半端な場面で「続きは次週！」となる。

前回の「100 社失格！就活で失敗したことで気付いた 8 つの掟」という例タイトルを

「100 社失格！就活で したことで気付いた 8 つの掟」というタイトルにすることでより一層惹かれるタイトルになる。

解説「重要度の高い単語をあえて伏字にする。」

2 今回の課題

2.1 ブログ記事のタイトルを抜き出して、思わず読みたくなるものと、そうでないものに分けて、特徴を調べる。

ブログの記事を 200 個抜き出す。

抜き出したタイトルの特徴と頻度を抽出する。

一つのタイトルから複数の特徴を抽出しているのので、100 を超えている。

にほんブログ村 引用

カテゴリ「IT 技術」新着記事のタイトル(100 件)

読みたくなるもの	タイトルが問いかけになっている	3
	検索ワードにかかるような工夫をしている	1
	タイトルが呼びかけになっている	1
	筆者の感想を提示する	1
	知りたい情報の難易度がわかる	1
	引きのある文言になっている	5
	伏字になっている	1
	自分の肩書を提示している	1
	検索する人にとって役立つ記事だとわかる	28
そうでないもの	タイトルが短すぎて内容が (or 結果しか) わからない	18
	文書がおかしい	3
	抽象的すぎる	3
	長すぎる	1
	情報量が少なすぎる	1
	購読者にしか伝わらない	1
	単語が連続していて読む気が失せる	5
	タイトルが普通過ぎて魅力を感じない	8
	タイトルを 2 つに提示している	1
	タイトルから記事の内容が推測できる	21

Table1 :読みたくなるものとそうでないものの特徴 (にほんブログ村)

生活情報配信サイト「All About」 引用
 カテゴリ「ビジネス・学習」新着記事のタイトル(100 件)

読みたくなるもの	具体的な数字が入っている	8
	タイトルが問いかけになっている	4
	引きのある文言になっている	28
	ターゲットを絞った呼びかけ (カクテルパーティ効果)	6
	楽観的に提示している	1
	検索する人の背景を理解した書き方	3
	ハロー効果を活用してタイトル	2
	結果を提示し、説得力のあるタイトル	1
	筆者の肩書を提示し、説得力のあるタイトル	2
	問題提示し、検索する人に疑問を生ませる	1
	心理的に損失回避を促すタイトル	2
	擬声語を入れている	2
	「～しても～できる」認知的不協和なタイトル	1
	検索する人にとって役立つ記事だとわかる	43
そうでないもの	単語が連続していて読む気が失せる	2
	漢字が連続していて読む気が失せる	1
	タイトルが普通過ぎて魅力を感じない	2
	検索する人にとってなんとなくしか内容がわからない	9
	普段使わない単語を多用する	1

Table2 :読みたくなるものとそうでないものの特徴 (All Anout)

2.2 黒田知宏君らの論文を読んで

論文名「文書クラスタリングにおけるクラスアタイトルの自動生成」[1]

検索エンジン「CiNii」、クエリ「タイトル 生成」

第 2 章 " タイトルの自動生成手法 "

2 節 " 形態素解析 "

文に対して形態素解析を行い名詞を抽出する．ここで

名詞とは形態素解析によって名詞と判断された単名詞の他に、単名詞やアルファベットによって成り立つ複合名詞も含む。

3 節 " 重要語の決定 "

形態素解析で抽出された名詞に対して重要度を計算する。

4 節 " タイトルパタンの適用 "

重要語を含む文に対して係り受け解析を行い、タイトルパタンを適用することでタイトル候補を生成する。係り受け解析には「南瓜」を用いる。

5 節 " タイトルの決定 "

名詞の重要度などを用いてタイトル候補の重要度を算出し、上位 1 件をタイトルとして出力する。

2.3 思わず読みたくなるタイトルが生成されるプロセス

ただタイトルを生成するだけでなく、思わず読みたくなるタイトルを生成することが本研究の趣旨である。

文献 [1] と特徴を参考に考えられる手法

- ・心理に基づいたタイトル生成
例「～しても～できる (認知的不協和)」、「重要度の高い語を伏字にする (ツァイガルニック効果)」、「女性必見、～のための～ (カクテルパーティ効果)」
- ・記事が箇条書きの場合、具体的な数字をタイトルに挿入
- ・魅かれる文言 (語句) をタイトルに挿入
例「必須、絶対、驚愕、成功、幸福、アップ、定番、便利、効率化、流行の、要注意、納得、誰でも、役立つ、必見、必要、簡単」
- ・感嘆符を挿入する
- ・体言止め
例「Java 言語はスマートフォンのアプリ開発でも重要です。」 「Java 言語はスマートフォンのアプリ開発でも重要」

上記のいくつかの手法は、文書に出現しない語を用いなければならない。

入力には仮タイトル + 記事本文である。

1. 形態素解析

仮タイトルを含めて行う。

2. 重要語の決定

仮タイトルを含めて行う。

最大重要値の 1/4 未満を対象外とする

タイトル候補の生成

重要語の含まれる文を元にする

その文の重要度の高い語順に心理に基づいてそれぞれの効果に合ったタイトルを生成

文 1 つに対して複数のタイトルを生成

強調する語句を感嘆符をつけて 1 つ付与する

長すぎると検索エンジンにタイトル全体が表示されないので 30 字以内とする。タイトルの決定
タイトル候補の重要度を計算し、上位 3 件をタイトルとして出力する。

3 次の発表までにすること

1. 文書に出現しない語をどう表現するか調査する。
2. 魅かれる文言を調査し、データとして蓄積する。
3. 心理を利用するための品詞を調査する。

参考文献

[1] 黒田知宏, 内海彰 (2008) . 文書クラスタリングにおけるクラスあタイトルの自動生成 . 情報科学技術フォーラム講演論文集 7(2),279-281

All About, 新着記事一覧 | ビジネス・学習 All About, <http://allabout.co.jp/r/business/latest/>, 2014/12/11 参照

にほんブログ村, IT 技術ブログ村新着記事, <http://it.blogmura.com/arrival.html>, 2014/12/11 参照