(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第5990178号
(P5990178)

(45)発行日　平成28年9月7日(2016.9.7)　(24)登録日　平成28年8月19日(2016.8.19)

(51)Int.Cl.　F I
G06F 17/30 (2006.01)　G06F 17/30 210A
G06F 17/27 (2006.01)　G06F 17/30 220Z
G06F 17/30 170A
G06F 17/27 665

請求項の数 20　(全 18 頁)

(21)出願番号　特願2013-537776 (P2013-537776)
(86)(22)出願日　平成23年11月2日 (2011.11.2)
(65)公表番号　特表2013-544397 (P2013-544397A)
(43)公表日　平成25年12月12日 (2013.12.12)
(86)国際出願番号　PCT/US2011/058899
(87)国際公開番号　WO2012/061462
(87)国際公開日　平成24年5月10日 (2012.5.10)
審査請求日　平成26年8月7日 (2014.8.7)
(31)優先権主張番号　61/410,392
(32)優先日　平成22年11月5日 (2010.11.5)
(33)優先権主張国　米国 (US)

(73)特許権者　399037405
楽天株式会社
東京都世田谷区玉川一丁目１４番１号
(74)代理人　100088155
弁理士　長谷川　芳樹
(74)代理人　100113435
弁理士　黒木　義樹
(74)代理人　100144440
弁理士　保坂　一之
(72)発明者　スタンキーウィッチ,　ゾフィア
アメリカ合衆国,　ニューヨーク州,　フォレスト　ヒルズ,　イエローストーン　ブルーヴァード　６８－３７,　アパートメント　シー６５

最終頁に続く

(54)【発明の名称】キーワード抽出に関するシステム及び方法

(57)【特許請求の範囲】
【請求項１】
（ａ）ウェブページからテキストを抽出して少なくとも候補キーワードの第１のセットを生成し、言語処理を適用して少なくとも候補キーワードの第２のセットを生成し、前記候補キーワードの第１及び第２のセットを第１の候補プールに結合する前処理部と、
（ｂ）少なくとも前記第１の候補プールを記述するデータを前記前処理部から受信して第２の候補プールを生成する候補抽出部と、
（ｃ）前記前処理部および前記候補抽出部に接続され、少なくとも前記第２の候補プールを記述するデータを受信し、一般的特徴及び言語的特徴について前記第２の候補プールを分析する特徴抽出部と、
（ｄ）少なくとも前記第２の候補プールを記述する前記データと関連データとを前記特徴抽出部から受信し、前記第２の候補プール内の各候補が１次又は２次キーワードである尤度を判定する分類部と
を備え、
各候補が１次又は２次キーワードである尤度の前記判定が、複数の注釈者からの注釈入力を結合することによって作成されたトレーニングデータに基づき、各注釈が１次キーワードと２次キーワードとの間の区別を含む、
コンピュータシステム。
【請求項２】
前記言語処理の少なくとも一部がトークナイザ及びパーサによって実行される、

請求項１に記載のコンピュータシステム。
【請求項３】
前記言語処理の少なくとも一部がトークナイザ、パーサ、品詞タグ付けプログラム、及び固有表現タグ付けプログラムによって実行される、
請求項１に記載のコンピュータシステム。
【請求項４】
前記言語処理の少なくとも一部がトークナイザによって実行される、
請求項１に記載のコンピュータシステム。
【請求項５】
前記言語処理の少なくとも一部がパーサによって実行される、
請求項１に記載のコンピュータシステム。
【請求項６】
前記言語処理の少なくとも一部が品詞タグ付けプログラムによって実行される、
請求項１に記載のコンピュータシステム。
【請求項７】
前記言語処理の少なくとも一部が固有表現タグ付けプログラムによって実行される、
請求項１に記載のコンピュータシステム。
【請求項８】
前記候補キーワードの第１のセットがメタデータテキストを含む、
請求項１～７のいずれか一項に記載のコンピュータシステム。
【請求項９】
前記第２の候補プールが名詞句及び名詞列を含む、
請求項１～８のいずれか一項に記載のコンピュータシステム。
【請求項１０】
前記第２の候補プールが名詞句、名詞列、及びｎグラムを含む、
請求項１～８のいずれか一項に記載のコンピュータシステム。
【請求項１１】
前記一般的特徴が頻度、文書中の位置、及び大文字使用のうちの一又は複数を含む、
請求項１～１０のいずれか一項に記載のコンピュータシステム。
【請求項１２】
前記言語的特徴が品詞、語句構造、及び固有表現情報のうちの一又は複数に関連する、
請求項１～１１のいずれか一項に記載のコンピュータシステム。
【請求項１３】
前記一般的特徴が頻度特徴を含み、前記頻度特徴が前記ウェブページ内の関連語出現頻度と語出現頻度のログとのうちの一又は複数を含む、
請求項１～１０のいずれか一項に記載のコンピュータシステム。
【請求項１４】
前記一般的特徴が頻度、文書中の位置、及び大文字使用を含み、前記言語的特徴が品詞、語句構造、及び固有表現情報に関連する、
請求項１～１０のいずれか一項に記載のコンピュータシステム。
【請求項１５】
前記一般的特徴が頻度特徴を含み、前記頻度特徴が前記ウェブページ内の関連語出現頻度と語出現頻度のログとのうちの一又は複数を含み、前記言語的特徴が品詞、語句構造、及び固有表現情報に関連する、
請求項１～１０のいずれか一項に記載のコンピュータシステム。
【請求項１６】
前記注釈が、少なくとも一人の注釈者によって前記１次キーワードとしてマークを付けられたことを示す第１の注釈と、前記複数の注釈者によって前記２次キーワードとして選択されたことを示す第２の注釈と、一人の注釈者により前記２次キーワードとして選択され、かつ他の注釈者により選択された前記１次キーワードに部分的に一致することを示す

第３の注釈とから選択される、
請求項１～１５のいずれか一項に記載のコンピュータシステム。
【請求項１７】
コンピュータ処理システムによって実装されるステップを含む方法であって、
（ａ）ウェブページからテキストを抽出して少なくとも候補キーワードの第１のセットを生成し、言語処理を適用して少なくとも候補キーワードの第２のセットを生成し、前記候補キーワードの第１及び第２のセットを第１の候補プールに結合するステップと、
（ｂ）少なくとも前記第１の候補プールを記述するデータを受信して第２の候補プールを生成するステップと、
（ｃ）少なくとも前記第２の候補プールを記述するデータを受信し、一般的特徴及び言語的特徴について前記第２の候補プールを分析するステップと、
（ｄ）少なくとも前記第２の候補プールを記述する前記データと関連データとを特徴抽出部から受信し、前記第２の候補プール内の各候補が１次又は２次キーワードである尤度を判定するステップと
を含み、
各候補が１次又は２次キーワードである尤度の前記判定が、複数の注釈者からの注釈入力を結合することによって作成されたトレーニングデータに基づき、各注釈が１次キーワードと２次キーワードとの間の区別を含む、
方法。
【請求項１８】
前記注釈が、少なくとも一人の注釈者によって前記１次キーワードとしてマークを付けられたことを示す第１の注釈と、前記複数の注釈者によって前記２次キーワードとして選択されたことを示す第２の注釈と、一人の注釈者により前記２次キーワードとして選択され、かつ他の注釈者により選択された前記１次キーワードに部分的に一致することを示す第３の注釈とから選択される、
請求項１７に記載の方法。
【請求項１９】
（ａ）ウェブページからテキストを抽出して少なくとも候補キーワードの第１のセットを生成し、言語処理を適用して少なくとも候補キーワードの第２のセットを生成し、前記候補キーワードの第１及び第２のセットを第１の候補プールに結合するステップと、
（ｂ）少なくとも前記第１の候補プールを記述するデータを受信して第２の候補プールを生成するステップと、
（ｃ）少なくとも前記第２の候補プールを記述するデータを受信し、一般的特徴及び言語的特徴について前記第２の候補プールを分析するステップと、
（ｄ）少なくとも前記第２の候補プールを記述する前記データと関連データとを特徴抽出部から受信し、前記第２の候補プール内の各候補が１次又は２次キーワードである尤度を判定するステップと
をコンピュータシステムに実行させ、
各候補が１次又は２次キーワードである尤度の前記判定が、複数の注釈者からの注釈入力を結合することによって作成されたトレーニングデータに基づき、各注釈が１次キーワードと２次キーワードとの間の区別を含む、
プログラム。
【請求項２０】
前記注釈が、少なくとも一人の注釈者によって前記１次キーワードとしてマークを付けられたことを示す第１の注釈と、前記複数の注釈者によって前記２次キーワードとして選択されたことを示す第２の注釈と、一人の注釈者により前記２次キーワードとして選択され、かつ他の注釈者により選択された前記１次キーワードに部分的に一致することを示す第３の注釈とから選択される、
請求項１９に記載のプログラム。
【発明の詳細な説明】

【序文】
【０００１】
キーワード抽出は、通常は、ウェブページの内容に合致する広告がページテキストから自動的に選択されたキーワードに基づいて選ばれる、文脈的広告システムの中核的な構成要素としての機能を果たす。そのウェブページに関連し、ユーザにとって興味深いと思われる広告を表示するためには、そのテキスト中に存在する多くの特徴が評価されて、どのキーワードがそのページの内容を的確に反映するかに関する決定を行う必要がある。
【０００２】
本明細書に記載の例示的な一実施形態では、キーワード抽出システムはページのｕｒｌを入力として取り、上位キーワード候補としてシステムによってランク付けされた１０個のキーワード語句を返す。本システムはまずウェブページテキストを処理し、その構造を使用して、キーワード候補プールの役割を果たす語句を抽出する。次いで、各語句はウェブページ上の頻度、テキスト中の位置、大文字使用、及びその言語構造（たとえば、それが名詞句を構成するかどうか）などの特徴のセットによって説明され得る。人間が注釈を付けたキーワードを有するサンプルウェブページのコレクションに基づき、本システムは、候補語句が「良い」キーワードである可能性が高いかどうかの決定にこれらの特徴がいかに寄与するかを学習する。この方式でトレーニングされた後には、本システムは、前に見ていない（すなわち、トレーニングセットになかった）ウェブページ上のキーワードを識別するために使用することができる。
【０００３】
既存のキーワード抽出システムの大多数は、ｔｆ－ｉｄｆなどの統計的頻度測定を使用する情報検索モデルに依存する。例示的なシステム実施形態は、性能を改善するために、自然言語処理技法を使用することによってこの手法を改善する。一又は複数の例示的な実施形態では、語句構造に敏感な新しいキーワード候補抽出方法を使用し、より優れた機械学習結果をもたらす追加の言語的特徴を含み得る。
【０００４】
ここで、ｔｆ－ｉｄｆ重み（語出現頻度－逆文書頻度）は、情報検索及びテキストマイニングでしばしば使用される重みである。この重みは、単語がコレクション又はコーパス中の文書にとってどの程度重要かを評価するために使用される統計的測定である。その重要性は、その単語が文書中に現れる回数に比例して増すが、コーパス中のその単語の頻度によってオフセットされる。
【０００５】
例示的な一態様は、（ａ）ウェブページからテキストを抽出して少なくとも候補キーワードの第１のセットを生成し、言語処理を適用して少なくとも候補キーワードの第２のセットを生成し、候補キーワードの第１及び第２のセットを第１の候補プールに結合する前処理部と、（ｂ）少なくとも第１の候補プールを記述するデータを前処理部から受信して第２の候補プールを生成する候補抽出部と、（ｃ）少なくとも第２の候補プールを記述するデータを受信し、一般的特徴及び言語的特徴について第２の候補プールを分析する特徴抽出部と、（ｄ）少なくとも第２の候補プールを記述するデータと関連データとを特徴抽出部から受信し、第２の候補プール中の各候補が１次又は２次キーワードである尤度を判定する分類部とを備えるコンピュータシステムを含む。
【０００６】
一又は複数の例示的な実施形態で、及びそれらの組合せで、（１）言語処理の少なくとも一部はトークナイザ及びパーサによって実行され、（２）言語処理の少なくとも一部はトークナイザ、パーサ、品詞タグ付けプログラム、及び固有表現タグ付けプログラムによって実行され、（３）言語処理の少なくとも一部はトークナイザによって実行され、（４）言語処理の少なくとも一部はパーサによって実行され、（５）言語処理の少なくとも一部は品詞タグ付けプログラムによって実行され、（６）言語処理の少なくとも一部は固有表現タグ付けプログラムによって実行され、（７）候補キーワードの第１のセットはメタデータテキストを含み、（８）第２の候補プールは名詞句及び名詞列を含み、（９）第２

の候補プールは名詞句、名詞列、及びｎグラムを含み、（１０）一般的特徴は頻度、文書中の位置、及び大文字使用のうちの一又は複数を含み、（１１）言語的特徴は品詞、語句構造、及び固有表現情報のうちの一又は複数に関連し、（１２）一般的特徴は頻度特徴を含み、頻度特徴はウェブページ内の関連語出現頻度と語出現頻度のログとのうちの一又は複数を含み、（１３）各候補が１次又は２次キーワードである尤度の判定は注釈付きトレーニングデータに基づき、（１４）各候補が１次又は２次キーワードである尤度の判定は、複数の注釈者からの注釈入力を結合することによって作成されるトレーニングデータに基づき、各注釈は１次キーワードと２次キーワードの区別を含み、（１５）一般的特徴は頻度、文書中の位置、及び大文字使用を含み、言語的特徴は品詞、語句構造、及び固有表現情報に関連し、そして／あるいは、（１６）一般的特徴は頻度特徴を含み、頻度特徴はウェブページ内の関連語出現頻度と語出現頻度のログとのうちの一又は複数を含み、言語的特徴は品詞、語句構造、及び固有表現情報に関連する。

【０００７】

もう一つの態様は、コンピュータ処理システムによって実装されるステップを含む方法を含み、当該ステップは、（ａ）ウェブページからテキストを抽出して少なくとも候補キーワードの第１のセットを生成し、言語処理を適用して少なくとも候補キーワードの第２のセットを生成し、候補キーワードの第１及び第２のセットを第１の候補プールに結合するステップと、（ｂ）少なくとも第１の候補プールについて説明するデータを受信して第２の候補プールを生成するステップと、（ｃ）少なくとも第２の候補プールについて説明するデータを受信し、一般的特徴及び言語的特徴について第２の候補プールを分析するステップと、（ｄ）少なくとも第２の候補プールについて説明するデータと関連データとを特徴抽出部から受信し、第２の候補プール内の各候補が１次又は２次キーワードである尤度を判定するステップとを含む。

【０００８】

もう一つの態様は、（ａ）ウェブページからテキストを抽出して少なくとも候補キーワードの第１のセットを生成し、言語処理を適用して少なくとも候補キーワードの第２のセットを生成し、候補キーワードの第１及び第２のセットを第１の候補プールに結合するステップと、（ｂ）少なくとも第１の候補プールについて説明するデータを受信して第２の候補プールを生成するステップと、（ｃ）少なくとも第２の候補プールについて説明するデータを受信し、一般的特徴及び言語的特徴について第２の候補プールを分析するステップと、（ｄ）少なくとも第２の候補プールについて説明するデータと関連データとを特徴抽出部から受信し、第２の候補プール内の各候補が１次又は２次キーワードである尤度を判定するステップとを含むステップ群を実行するように動作可能なソフトウェアを格納した、コンピュータ読み取り可能な記録媒体を備える。

【０００９】

本明細書で提供される説明及び図面から、他の態様及び実施形態が当業者には明らかとなろう。

【図面の簡単な説明】

【００１０】

【図１】例示的な一実施形態の処理の概要を示す図である。

【図２】例示的な一実施形態が実装され得るコンピュータシステムを示す図である。

【選ばれた例示的な実施形態の詳細な説明】

【００１１】

コンピュータで実装される例示的な実施形態の概要を図１に示す。各構成要素は本明細書の残りの節でさらに詳細に説明される。

【００１２】

例示的な前処理部

例示的な一実施形態では、潜在的なキーワード語句がそのページから選択され得る前に、そのページのプレーンテキストがＨＴＭＬ形式から抽出されてもよい。順に、このテキストをさらに処理することで、キーワード抽出システムに有用であり得るその構造に関す

る情報を取得することができる。本システムの前処理部は、好ましくは、ウェブページテキストの抽出並びにタグ付け及び書式付けを実行して、後に続く候補語句選択及び特徴抽出の段階のための適切な入力を提供する。

【００１３】

その前処理段階で、まず、定型的な内容を除去してそのページの主要なテキスト本体のみを保存するＢｏｉｌｅｒＰｉｐｅ（たとえば［９］を参照）を使用して、ウェブページから内容テキストを抽出してもよい。本体テキストに加えて、タイトル、メタ記述、メタキーワードなどのヘッダ情報を抽出し、ＢｏｉｌｅｒＰｉｐｅ出力と結合することで、さらなる処理のためのプレーンテキスト入力を作成してもよい。

【００１４】

次いで、ページテキストはトークン化され、そのトークナイザ出力は品詞タグ付けプログラム（たとえば［１８］を参照）及びパーサ（たとえば［１３］を参照）に渡される。キーワードは名詞句を構成する傾向があるので、そのテキスト中で名詞句を見つけるためにパーサ出力を使用してもよい。チャンカではなくてパーサの使用は、キーワード候補抽出を改良するために、基本的名詞句のかたまりに対するものとして、階層的語句構造のより肌理の細かい情報を取得したいという要望によって動機付けられ得る。

【００１５】

個人名又は組織名などの固有表現（「ＮＥ」）は有用なキーワードとなり得るので、固有表現もウェブページテキストから抽出してもよい。二つの異なるＮＥシステム（たとえば［１８］，［４］を参照）を、エンティティタイプの大きなセットを網羅するために用いるのが好ましい。

【００１６】

例示的な候補抽出部

候補抽出は、潜在的なキーワードであり、且つ所与の語句がキーワードである尤度を推定する分類器の入力として使用することができる語句を選択するために用いてもよい。加えて、例示的な分類器が語句をトレーニングしている間に、候補抽出の精度の向上は、キーワードではありそうにない単語の組合せをフィルタリングするのに役立ち、良くないトレーニングサンプルの量を減らし、ひいては、良いトレーニングデータと悪いトレーニングデータとの比率を改善する（キーワード抽出タスクは、良いラベルデータが殆どなく、良いサンプルと悪いサンプルの間の不均衡を有する）。

【００１７】

例示的な一実施形態で、キーワード抽出方法は以下のように行われる。まず、基底候補セットが、構文解析されたテキストからすべての名詞句を再帰的に抽出することによって形成される。次いで、名詞のみから成るすべての候補部分列（左から右に抽出された）が候補セットに追加される（たとえば、「最高のニクソン・カメラ・アクセサリ」が候補である場合、「ニクソン・カメラ・アクセサリ」、「カメラ・アクセサリ」及び「アクセサリ」がその候補セットに追加されることになる）。そして、候補セットは、候補語句から抽出されたすべてのユニグラム、バイグラム、及びトライグラムにより増補される。

【００１８】

候補セットについて、最頻出の英単語のストップリストに対するフィルタをかけてもよい。ストップワードを含むユニグラム又はバイグラムを候補セットから取り除くのが好ましい。しかし、その真ん中にストップリスト内の単語を含むより長い語句は保持してもよい。

【００１９】

例示的なキーワード分類部

どの候補語句がキーワードであるかを識別するために、例示的な一実施形態では、入力（候補語句の特徴）を使用してその語句がキーワードである確率を推定する分類器を使用し、出力ラベル（キーワード又は非キーワード）をその語句に割り当てる。特徴入力をキーワードラベルにマップする分類機能は、監視された機械学習を使用して取得してもよい。すなわち、本マッピングは、「正しい」出力ラベルが注釈者によって提供されたデータ

セットに基づく分類器システムによって学習され得る。

【００２０】

例示的なシステムの分類器をトレーニングするために、最大エントロピー（ＭＥ）モデルを用いてもよい（これはときにロジスティック回帰モデルと呼ばれる。その紹介については［１１］を参照）。ＭＥモデルは、トレーニングデータから制約条件を導出し、そのトレーニングセットによって網羅されない場合に最大エントロピーの分配を仮定する。そのＭＥ分類器の入力は、各特徴に関連付けられた重みを学習するためにそのモデルによって使用される、各キーワード候補の値のベクトルで構成される。新しい入力データが与えられると、トレーニングされた分類器は、語句が、候補語句の入力値を与えられたキーワードである確率を計算することができる。

【００２１】

入力値

【数１】

$$\vec{x}$$

を与えられたラベルｃの確率は以下の公式により計算することができる。

【数２】

$$P(c \mid \vec{x}) = \frac{exp(\sum_i \alpha_i f_i(\vec{x}, c))}{1 + exp(\sum_i \alpha_i f_i(\vec{x}, c))}$$

ここで、ｆは共同特徴（入力ベクトル及びラベルの関数）であり、αはその特徴に割り当てられた重みである。

【００２２】

最大エントロピー分類器をトレーニングするために、自然言語ツールキット（Ｎａｔｕｒａｌ　Ｌａｎｇｕａｇｅ　Ｔｏｏｌｋｉｔ（ＮＬＴＫ）。［１］を参照）で提供されるＰｙｔｈｏｎライブラリを使用することができる。ＣＧはトレーニング方法として用いることができる。しかし、そのアルゴリズムは一意解に収束するので、トレーニング方法の選択は分類器の性能に影響せず、他のトレーニング方法が本明細書に記載の本発明の範囲を逸脱することなく使用可能であることが当業者に理解されよう。たとえば、サポート・ベクトル・マシン（ｒｂｆカーネル）（［８］を参照）が使用可能であるが、ＭＥモデルを使用して得られる結果に比べた改善は本発明者によって発見されなかった。

【００２３】

ここで、ＣＧは共役勾配法のことであり、これは、分類器ライブラリ内のトレーニング方法の一つとして提供されるスパース１次方程式系を解くための標準反復法である。ＣＧはＰｙｔｈｏｎ及びＮＬＴＫでインストールされるサイパイ（ｓｃｉｐｙ）パッケージ（ｈｔｔｐ：／／ｗｗｗ．ｓｃｉｐｙ．ｏｒｇ／）を必要とする。

【００２４】

良いトレーニングデータと悪いトレーニングデータとの不均衡（すなわち、トレーニングデータ内の候補語句の大部分が通常はキーワードではない）のために、分類器によって割り当てられたラベル（キーワード又は非キーワード）を使用しないことを選択してもよいが、その代わりに、たとえば所与のウェブページ内で最も高い確率を有する１０個の候補を選択して、確率のスコアに直接基づいて候補をランク付けすることができる。

【００２５】

例示的な特徴

特徴値のセットは各キーワード候補について計算されて分類器への入力として使用することができる。特徴の選択は分類器の性能で重要な役割を果たす。その特徴は、（ａ）一般的、非言語的特徴と、（ｂ）言語的特徴の二つのタイプに分けることができる。一般的特徴は、［１７］に記載のシステムによって使用される特徴と同様でもよく、頻度、文書

中の位置、及び大文字使用などの情報を含む。言語的特徴は品詞、語句構造、及び固有表現情報を使用する。二つのタイプの特徴について以下にさらに詳しく説明する。

【００２６】

【表１】

例示的な分類特徴

特徴名	定義	タイプ	[17]での使用
TF	キーワード候補が文書中で出現する回数／文書中の候補の総数	実数値	はい
TFLog	Log(TF+1)	実数値	はい
DF	Google Ngramコーパスのユニグラム及びバイグラムデータに基づくキーワード候補の相対的頻度。候補語句が2単語より長い場合、その候補語句内のすべてのバイグラムの平均頻度が使用される。	実数値	はい(異なる文書コレクションに基づく)
DFLog	Log(DF+1)	実数値	はい
Title	キーワード候補が文書のタイトル中にある場合は１、そうでない場合は０。	2値	はい
IsCap	キーワード候補内のすべての単語が大文字で書かれている場合は１、そうでない場合は０。	2値	はい
HasCap	キーワード候補内の少なくとも1つの単語が大文字で書かれている場合は１、そうでない場合は０。	2値	はい
Location	文書中の単語の総数に対する、当該文書内のキーワード候補の最初の出現の相対的な位置。	実数値	はい
LocationLog	Log(Location)	実数値	はい
Length	キーワード候補内の単語の数	実数値	はい
Url	キーワード候補が文書のｕｒｌに含まれる場合は１、そうでない場合は０。	2値	はい
IsNoun	キーワード候補中のすべての単語が名詞である場合は１、そうでない場合は０。	2値	はい(しかし、固有名詞と総称的名詞の区別で異なるように定義される)
hasNoun	キーワード候補中の少なくとも1つの単語が名詞である場合は１、そうでない場合は０。	2値	はい
isNP	キーワード候補が名詞句である場合は１、そうでない場合は０。	2値	はい
hasNP	キーワード候補が名詞句を含む場合は１、そうでない場合は０。	2値	いいえ
POS sequence	キーワード候補の品詞タグの列	実数値	いいえ
IsNE_oak	キーワード候補がＯＡＫシステムによって見つけられた固有表現である場合は１、そうでない場合は０。	2値	いいえ
HasNE_oak	キーワード候補がＯＡＫシステムによって見つけられた固有表現を含む場合は１、そうでない場合は０。	2値	いいえ
TagNE_oak	ＯＡＫシステムによってキーワード候補に割り当てられたＮＥタグ	実数値	いいえ
IsNE_Stanford	キーワード候補がスタンフォードＮＥＲシステムによって見つけられた固有表現である場合は１、そうでない場合は０。	2値	いいえ
HasNE_Stanford	キーワード候補がスタンフォードＮＥＲシステムによって見つけられた固有表現を含む場合は１、そうでない場合は０。	2値	いいえ
TagNE_Stanford	スタンフォードＮＥＲシステムによってキーワード候補に割り当てられたタグ	実数値	いいえ
Pmi	候補語句の自己相互情報量のスコア	実数値	いいえ
Iswiki	キーワード候補がウィキペディアのタイトルである場合は１、そうでない場合は０。	2値	いいえ(しかし、[14]及び[17]で使用される探索照会特徴と同様)
WikiFreq	キーワードがウィキペディアのタイトルである場合はウィキペディアトラフィック頻度、そうでない場合は０。	実数値	いいえ(しかし、[14]及び[17]で使用される探索照会特徴と同様)

【００２７】

例示的な一般的特徴

例示的な頻度特徴

頻度特徴は、ＴＦｘＩＤＦと同様の情報を提供する。頻度特徴は、文書内の関連語出現頻度と語出現頻度のログとの他にＤＦ（文書コレクション中の頻度）とログＤＦ値とを含んでもよい。ＤＦ値は、Ｇｏｏｇｌｅ　Ｎｇｒａｍコーパスからの頻度を使用して概算してもよい。好ましくは、ユニグラム及びバイグラムの頻度情報のみが、ＤＦを計算するために使用される。２単語よりも長い候補語句については、その語句内のすべてのバイグラムのＤＦの平均値をＤＦ値として用いてもよい。平均値は、異なる長さの語句についての値の似た範囲を得るために使用することができる。また、ブログのコレクション全体について計算されたＤＦ値をＧｏｏｇｌｅ　Ｎｇｒａｍコーパスからの頻度の代わりに用いてもよい。

【００２８】

ここで、ＴＦｘＩＤＦは語出現頻度－逆文書頻度のことをいい、用語の相対的重要性を評価するために情報検索で使用される標準スコアである。これは、文書のコレクション内のその全体的頻度によって相殺された所与の文書内の用語の頻度に基づく。ｔｆ及びｉｄｆの標準的公式は以下の通りである。

$tf_{i,j} = n_{i,j} / \Sigma_k n_{k,j}$

ここで、$n_{i,j}$は、文書ｊ内の考察される用語ｉの出現回数である。

$idf_i = log(|D| / |d : t_i \in d|)$

これは、コレクション内のすべての文書の数のログを、用語ｉを含む文書の数で割ったものである。

【００２９】

２．タイトル

候補語句がその文書のタイトル中にあるかどうか。

【００３０】

３．大文字使用

大文字で書かれた単語は、所与の文書中の重要な用語としてマークされた固有名詞又は単語を含む。例示的な大文字使用の特徴は、キーワード候補内のすべての単語が大文字で書かれているかどうか、及び、候補語句内の少なくとも一つの単語が大文字で書かれているかどうか、である。

【００３１】

４．位置

文書内でキーワード候補が最初に出現する相対的な位置であり、単語の数で数える。たとえば、文書が２０個の単語を有し、候補語句の最初の単語がその文書中の５番目の単語である場合には、位置＝５／２０＝０．２５である。

【００３２】

５．長さ

候補語句中の単語の数。

【００３３】

６．ＵＲＬ

候補語句がページｕｒｌ内にあるかどうか。

【００３４】

７．Ｗｉｋｉトラフィック

ウィキペディア・トラフィックの統計値に基づく特徴を、頻出する探索／照会項目としてキーワード候補の人気を反映するために用いてもよい。この特徴のセットは、候補語句がウィキペディアのタイトル（リダイレクトを含む）であるかどうかと、その候補語句のトラフィック数字（その候補がウィキペディアのタイトルでない場合には０）とを含み得る。トラフィック統計値は、一定期間（たとえば、２０１０年６月中の２０日間）に亘って集められた１時間当たりのウィキペディアのログに基づくものであってもよい。

【００３５】

例示的な言語的特徴

１．名詞句
その候補が名詞句である又は名詞句を含むかどうか。
２．名詞
その候補語句が少なくとも一つの名詞を含むかどうか、及び、その候補語句が名詞のみから成るかどうか。
３．ＰＯＳタグ
その候補語句に割り当てられた品詞タグの列。
４．固有表現
キーワード候補が固有表現であるかどうか、キーワード候補が、固有表現及びその候補に割り当てられた固有表現タグを含むかどうか（その候補語句がＮＥでない場合には「０」）。
５．ＰＭＩ
自己相互情報量（Ｐｏｉｎｔｗｉｓｅ　ｍｕｔｕａｌ　ｉｎｆｏｒｍａｔｉｏｎ（ＰＭＩ））は、語句がコロケーションである可能性が高いかどうかを反映する。候補語句のＰＭＩスコアは以下のように計算することができる。
バイグラムについては、
$PMI(w_1, w_2) = \log\{P(w_1, w_2) / P(w_1) * P(w_2)\}$
ただし、Ｐ（ｗ）は単語又は語句の相対的頻度である。
単一の単語については、
$PMI = \log\{1 / P(w)\}$
２単語よりも長い候補語句については、ＰＭＩは、その語句内のすべてのバイグラムのＰＭＩスコアの平均値に設定してもよい。

【００３６】
例示的な評価及びデータ
例示的なトレーニングデータ
トレーニングデータは、たとえば５００個のウェブページ（ブログページのコーパスから無作為に選択されたもの。［３］を参照）を含んでもよい。注釈者は、ブログページから抽出されたプレーンテキストを提示され、そのページの内容を最もよく表すキーワードを選択するように指示され得る。ヘッダからのメタ情報は注釈付きテキストに含まれないのが好ましい。単一のページについて選択することができるキーワードの数には制限を設けないのが好ましい。追加のページについても、注釈を付け、トレーニングに使用されないテストセットとして除外することができる。

【００３７】
注釈者間の合意及び最も理想的な基準（ｇｏｌｄｅｎ　ｓｔａｎｄａｒｄ）
各ページについて、キーワードは２人の注釈者によって選ばれるのが好ましい。しかし、このタスクへの注釈者間の合意は高くなくてもよい（たとえば、一実装形態では、注釈者のカッパスコアは０．４９であった）。低いカッパスコアの要因としては次のことが考えられる。第１に、注釈者が、部分的にのみ一致する同様の語句にタグを付けることがある。第２に、選択可能なキーワードの最大数が指定されていないときに、一人の注釈者が、所与のテキストについてもう一人よりも多くのキーワードを選択することを選ぶことがある。

【００３８】
ここで、コーエンのカッパ係数は、分類タスクでの注釈者間の合意を測定するために一般に使用される統計的測定である。カッパは、｛Ｐ（Ａ）－Ｐ（Ｅ）｝／｛１－Ｐ（Ｅ）｝で計算され、ここで、Ｐ（Ａ）はコーダ間で観測された合意であり、Ｐ（Ｅ）はコーダが偶然合意した確率である。０．６～０．７を上回るカッパスコアは「実質的合意」として考えられる。

【００３９】
単独の注釈者に依存しない最も理想的な基準（Ｇｏｌｄｅｎ　Ｓｔａｎｄａｒｄ（ＧＳ））を作成するために、両方の注釈者からの出力を結合してもよい。キーワードに注釈を

付けるとき、注釈者はそのキーワードが「１次キーワード」であるか「２次キーワード」であるかも選択するように指示され得る。１次キーワードは、文書の主題又は中心思想をとらえたキーワードとして規定することができる。２次キーワードは、その文書に関する追加のキー情報（たとえば、その事象が起きた位置、追加のものではあるが重要な、記述された数字など）を提供する重要な語句として規定することができる。両方の注釈者の選択を正確に反映するキーワードセットを作成するために、ＧＳで以下のキーワードを保持することができる。

１．（一人の注釈者又は双方によって）１次としてマークを付けられたすべてのキーワード。

２．両方の注釈者によって選択された２次キーワード。

３．一人のみによって選択されたが、他方の注釈者によって選択された１次キーワードに部分的に一致する、２次キーワード。

【００４０】

ＧＳを使用する例示的な一実施形態では、各注釈者と標準との間のカッパスコアは、注釈者１については０．７５で注釈者２については０．７４であった。１次及び２次キーワードの詳細な合意統計値を下記の表２に示す。

【００４１】

【表２】

注釈者1/注釈者2	1次	2次	キーワードではない
1次	1652	319	1796
2次	437	264	1777
キーワードではない	1069	264	//////////

【００４２】

例示的な候補抽出部

上記のように、例示的な一実施形態では、基底候補セットとして名詞句を使用するが、その名詞句から抽出された名詞列とユニグラム、バイグラム、及びトライグラムで候補プールを増補する。

【００４３】

起こり得るすべての候補語句をテキストから取得する一つの従来の手法では、その候補セット内の長さｎ（通常は３～５）までのすべてのｎグラムを含む。このｎグラム方法の重大な欠点は、それが、意味のある語句でない及び／又は潜在的キーワードでなさそうな単語列の形で、かなりのノイズをもたらすことである。したがってｎグラム方法は低い精度に悩まされる。

【００４４】

従来の一代替方法では、候補を抽出するために言語構造のキューを使用する。キーワードは名詞句である傾向があるので、テキストからのすべての名詞句が候補プールを形成するために使用され得る。しかし、この方法では再現率がｎグラム抽出方法よりも著しく低く、これは多くの潜在的なキーワードが候補セットに含まれないことを意味する。

【００４５】

ｎグラム及び名詞句戦略の精度、再現率、及びＦ測定が本発明者により例示的な一実施形態の抽出方法と比較された。言い換えれば、本発明者は、さらなる分類段階なしに、キーワードを選択するための唯一の方法としてそれぞれの手法が使用された場合にそれがどの程度効果的であるかを評価した。結果は以下の表３に要約される。

【００４６】

【表３】

候補抽出方法の比較

方法	総数	候補にないキー	候補にあるキー	精度	再現度%	Fスコア
Nグラム	365,779	786	4839	1.3	85.9	2.6
名詞句	14,441	4160	1465	10.4	26	14.6
例示的実施形態	85,059	1008	4617	5.4	81.95	10.2

【００４７】

表３に示すように、ｎグラム手法の再現率は８０％を超えるが、それはまた、その３つの方法のうちで精度が最も低い（すなわち、候補セットがかなりの量のノイズを含む）。候補として名詞句を抽出することは精度が上がる点で有利だが、この方法では再現率が非常に低く（わずか２６％）、潜在的なキーワードを見落とす蓋然性が高い。

【００４８】

対照的に、本発明の方法の例示的な一実施形態は、名詞句を抽出することに比べて再現率を改善する。この手法の再現率はｎグラム方法と同程度であるが、精度はより高くなる。異なる手法が分類器の性能とどのように結合するかの評価結果を以下に述べる。

【００４９】

分類器の性能

システム性能全体を評価するために、本発明のシステムによって達成される結果が［１７］に基づいて基準値と比較された。基準値システムで、候補抽出方法はｎグラム方法であり、特徴は（ＮＰ／名詞特徴の単純なセットを加えた）一般的非言語的特徴で構成される。（ａ）結合された候補抽出方法の使用と（ｂ）分類段階での言語的特徴の追加とでシステム性能がどのように変わったかを分析した。

【００５０】

基準値に対して本発明のシステムを比較する際には以下の二つの評価測定を用いた。

１．Ｒ精度（上位ｎ個の結果のうちいくつの候補がキーワードであるか。ここで、ｎはページ上に有り得るキーワードの総数）。

２．上位１０個のスコア（Ｒ精度に似ているが、上位１０個の結果で切り捨てる。すなわち、すべてのｎ＞１０が１０にセットされる）。

【００５１】

上位１０個のスコアを有する候補がキーワード出力として選択されるときにどのように分類器が抽出システムとして機能するかの推定を上位１０個の測定が提供するので、その測定を評価に用いた。システム性能は、分類器トレーニングでは決して使用されなかった１００個のウェブページの提出されたテストセットでテストされ（表４を参照）、そして、相互検証テストは、５００ページのトレーニングセットで行われた（それぞれが約５０個の文書の１０倍。表５を参照）。

【００５２】

【表４】

提出されたセットの上位10個のスコア結果

方法	一般的特徴	一般的+言語的特徴
Nグラム	43.71	47.24
本発明	48.28	49.84

【００５３】

【表５】
相互検証テストの上位10個のスコア結果

方法	一般的特徴	一般的+言語的特徴
Nグラム	45.97	49.01
本発明	48.21	51.74

【００５４】
基準値と本発明のシステムとの結果の差は統計的に重要である（相互検証結果への対応のある両側ｔ検定によれば、ｐ＝０．０００１）。基準値に対する相対的な改善は１２．５５％である。

【００５５】
関連実施形態
所与のウェブページについて文脈的に適切な広告を選択する二つの好ましい手法がある。一つの手法は、広告プールへのウェブページテキストの直接の突き合わせを含む。もう一方の手法では、そのページと広告の間の突き合わせが中間キーワード抽出ステップを含む。各手法の例を以下に示す。

【００５６】
キーワード抽出
１．ＫＥＡ［５］
２．ＧｅｎＥｘ［１５］、［１４］
３．Ｙｉｈ他［１７］
４．Ｈｕｌｔｈ［７］、［６］
５．その他：［１０］、［１６］

【００５７】
文脈的広告
１．Ｂｒｏｄｅｒ他［２］
２．Ｒｉｂｅｉｒｏ－Ｎｅｔｏ他［１２］。

【００５８】
本明細書に記載のいくつかの例示的なシステム及び方法の実施形態では、キーワード抽出は（ａ）ウェブページからのテキスト抽出、並びに、品詞タグ付け及び構文解析などの言語的処理を含む前処理と、（ｂ）キーワード候補語句の抽出と、（ｃ）監視された機械学習を使用する候補分類とを含むのが好ましい。

【００５９】
本発明のシステム及び方法は、候補選択及び特徴抽出段階の両方で、言語的情報の使用による性能の改善を達成することができる。例示的な一実施形態は、階層的語句構造を使用する候補選択を含み、よりノイズの少ない候補プールをもたらす。分類に使用できる特徴は品詞や固有表現情報などの言語的特徴も含み、分類器の性能の改善をもたらす。

【００６０】
実施形態は、当業者には明らかであろうコンピュータ構成要素及びコンピュータ実装ステップを含む。たとえば、計算及び通信は電子的に実行でき、結果をグラフィカルユーザインターフェースを用いて表示することができる。

【００６１】
そのような例示的なシステムを図２に示す。コンピュータ１００はネットワーク１１０を介してサーバ１３０と通信する。複数のデータソース１２０，１２１もネットワーク１１０を介してサーバ１３０、プロセッサ１５０、及び／又は、情報を計算及び／又は送信するように動作可能な他の構成要素と通信する。一又は複数のサーバ１３０は、一又は複数の記憶装置１４０、一又は複数のプロセッサ１５０、及びソフトウェア１６０に結合され得る。

【００６２】
本明細書に記載の計算及び同等のものは、一実施形態では、完全に電子的に実行される

。当業者には明らかなように、他の構成要素及び構成要素の組合せもまた、本明細書に記載の処理データ又は他の計算をサポートするために使用できる。サーバ１３０は、一又は複数のプロセッサ１５０とのデータ通信と、記憶装置１４０からのデータ通信と、コンピュータ１００への通信とを円滑に進めることができる。プロセッサ１５０は、一時的な情報又は他の情報を記憶するために使用することができるローカル又はネットワークストレージ（図示せず）を任意で含んでもよいし、それと通信してもよい。ソフトウェア１６０は、コンピュータ１００またはプロセッサ１５０でローカルにインストールされてもよいし、及び／又は、計算と適用とを容易にするために中央でサポートされてもよい。

【００６３】

説明を容易にするために、本発明のあらゆるステップ又は要素がコンピュータシステムの部分として本明細書に記載されているわけではないが、各ステップ又は要素は対応するコンピュータシステム又はソフトウェア構成要素を有してもよいことは当業者に理解されよう。したがって、そのようなコンピュータシステム及び／又はソフトウェア構成要素は、それらの対応するステップ又は要素（すなわち、それらの機能性）について説明することによって可能にされ、本発明の範囲内にある。

【００６４】

さらに、コンピュータシステムが特定の機能を実行するためのプロセッサを有するものとして説明又は特許請求される場合、そのような使用は、単一のプロセッサが、たとえば、さまざまなプロセッサに委託されたタスクのいくつか又はすべてを実行するシステムを除外するものとして解釈されるべきではないことが当業者に理解されよう。すなわち、本明細書及び／又は特許請求の範囲で指定されるプロセッサの任意の組合せ又はすべては同一のプロセッサでもよい。そのようなすべての組合せは本発明の範囲内である。

【００６５】

別の方法として、又は組み合わせて、処理及び意思決定はデジタル信号プロセッサ回路又は特定用途向け集積回路などの機能的に同等の回路によって実行可能である。

【００６６】

ループ及び変数の初期化と一時的数値変数の使用などの多数のルーチンプログラム要素は本明細書には記載されていない。さらに、特段の指示のない限り、記載されるステップの特定の順番は説明のみを目的とし、一般に、本発明の範囲から逸脱することなしに変更することができることが、当業者に理解されよう。特段の記述のない限り、本明細書に記載のプロセスは順序付けされていない。すなわち、そのプロセスは任意の妥当な順番で実行することができる。

【００６７】

本明細書に記載のすべてのステップは、実行可能な場合、ソフトウェアによる実装が可能であることが当業者に理解されよう。さらに、そのようなソフトウェアは非一時的なコンピュータ可読媒体に格納可能であり、一又は複数のコンピュータプロセッサによって実行可能であることが、当業者に理解されよう。

【００６８】

本発明は、本明細書で概説される例示的な態様の実施形態とともに説明されているが、多数の代替、修正、及び変更が当業者には明らかとなろう。したがって、本明細書に記載するような本発明の例示的な態様及び実施形態は例示的なであって限定ではない。さまざまな変更本発明の趣旨及び範囲を逸脱することなしに行われ得る。

【００６９】

参考文献

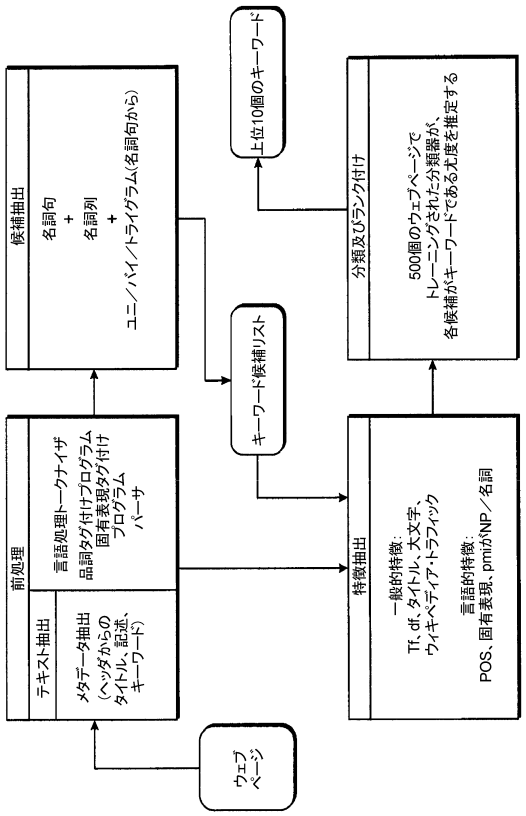
［1］Bird, Steven, Edward Loper及びEwan Klein。Natural Language Processing with Python。O’Reilly Media Inc.、２００９。

［2］Broder, Andrei及びFontoura, Marcus及びJosifovski, Vanja及びRiedel, Lance。A semantic approach to contextual advertising。SIGIR’07

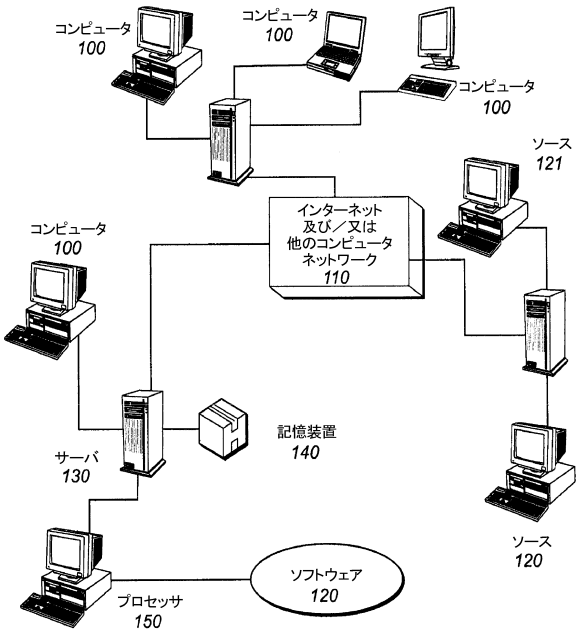
：Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval、ページ559～566、New York, NY, USA、2007。ACM。
［3］Kevin Burton及びAkshay Java及びIan Soboroff。ICWSM 2009 Spinn3r Dataset。San Jose, CA、2009。
［4］Finkel, Jenny Rose及びGrenager, Trond及びManning, Christopher。Incorporating non-local information into information extraction systems by Gibbs sampling。ACL '05：Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics、ページ363～370、Morristown, NJ, USA、2005。Association for Computational Linguistics。
［5］Frank, Eibe及びPaynter, Gordon W.及びWitten, Ian H.及びGutwin, Carl及びNevill-Manning, Craig G.。Domain-specific keyphrase extraction。IJCAI'99：Proceedings of the 16th international joint conference on Artificial intelligence、ページ668～673、San Francisco, CA, USA、1999。Morgan Kaufmann Publishers Inc.。
［6］Hulth, Anette。Automatic Keyword Extraction。Combining Machine Learning and Natural Language Processing。Verlag Dr. Muller、2008。
［7］Hulth, Anette。Improved automatic keyword extraction given more linguistic knowledge。Proceedings of the 2003 conference on Empirical methods in natural language processing、ページ216～223、Morristown, NJ, USA、2003。Association for Computational Linguistics。
［8］Joachims, Thorsten。Making large-scale support vector machine learning practical。：169～184、1999。
［9］Kohlschiitter, Christian及びFankhauser, Peter及びNejdl, Wolfgang。Boilerplate detection using shallow text features。WSDM '10：Proceedings of the third ACM international conference on Web search and data mining、ページ441～450、New York, NY, USA、2010。ACM。
［10］Matsuo, Y.及びIshizuka, M.。Keyword Extraction from a Document using Word Co-occurrence Statistical Information。Transactions of the Japanese Society for Artificial Intelligence、17：217～223、2002。
［11］Adwait Ratnaparkhi。A Simple Introduction to Maximum Entropy Models for Natur

al Language Processing。Technical report, IRCS、1997。
[12] Ribeiro−Neto, Berthier及びCristo, Marco及びGolgher, Paulo B.及びSilva de Moura, Edleno。Impedance coupling in content−targeted advertising。SIGIR '05：Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval、ページ496〜503、New York, NY, USA、2005。ACM。
[13] Sekine, Satoshi及びRalph Grishman。A corpus based probabilistic grammar with only two non−terminals。Fourth International Workshop on Parsing Technology, 1995。
[14] Turney, Peter D.。Coherent keyphrase extraction via web mining。IJCAI'03：Proceedings of the 18th international joint conference on Artificial intelligence、ページ434〜439、San Francisco, CA, USA、2003。Morgan Kaufmann Publishers Inc.。
[15] Turney, Peter D.。Learning Algorithms for Keyphrase Extraction。Inf. Retr.、2(4)：303〜336、2000。
[16] Wu, Xiaoyuan及びBolivar, Alvaro。Keyword extraction for contextual advertisement。WWW '08：Proceeding of the 17th international conference on World Wide Web、ページ1195〜1196、New York, NY, USA、2008。ACM。
[17] Yih, Wen−tau及びGoodman, Joshua及びCarvalho, Vitor R.。Finding advertising keywords on web pages。WWW'06：Proceedings of the 15th international conference on World Wide Web、ページ213〜222、New York, NY, USA、2006。ACM。
[18] OAK System, http：／／nlp.cs.nyu.edu／oak／。

【図１】


候補抽出
名詞句
＋
名詞列
＋
ユニ／バイ／トライグラム（名詞句から）
上位10個のキーワード
分類及びランク付け
500個のウェブページで
トレーニングされた分類器が、
各候補がキーワードである尤度を推定する
キーワード候補リスト
前処理
テキスト抽出
メタデータ抽出
（ヘッダからの
タイトル、記述、
キーワード）
言語処理トークナイザ
品詞タグ付けプログラム
固有表現タグ付け
プログラム
パーサ
特徴抽出
一般的特徴：
Tf、df、タイトル、大文字、
ウィキペディア・トラフィック
言語的特徴：
POS、固有表現、prn/NP／名詞
ウェブ
ページ

【図２】


コンピュータ
100
コンピュータ
100
コンピュータ
100
ソース
121
インターネット
及び／又は
他のコンピュータ
ネットワーク
110
コンピュータ
100
記憶装置
140
サーバ
130
ソース
120
ソフトウェア
120
プロセッサ
150

フロントページの続き

(72)発明者　関根　聡
アメリカ合衆国,　ニューヨーク州,　スカースデール,　ネルソン　ロード　２３２

審査官　齊藤　貴孝

(56)参考文献　米国特許出願公開第２００７／０１１２７６４（ＵＳ，Ａ１）
特開２００４－１３９５５３（ＪＰ，Ａ）
特開２００８－０６５４１７（ＪＰ，Ａ）
特開２００７－２５７３９０（ＪＰ，Ａ）
特開２００５－０８５２８５（ＪＰ，Ａ）
特開平０７－０６５０１８（ＪＰ，Ａ）
特開２００９－２７１７９４（ＪＰ，Ａ）
特開２００６－１４６７０５（ＪＰ，Ａ）
特開２０１０－２０４８６６（ＪＰ，Ａ）
上村　卓史、外2名，ウェブ閲覧における効率的なキーワード抽出とその利用，データベースとＷｅｂ情報システムに関するシンポジウム　情報処理学会シンポジウムシリーズ，日本，社団法人情報処理学会，２００７年１２月２０日，第２００７巻，第3号，ｐ．１－９
斎藤　一、外３名，キーワード地図構造モデリングによるグループ学習支援システムの構築，電子情報通信学会技術研究報告，日本，社団法人電子情報通信学会，２００３年　7月１７日，第１０３巻，第２１７号，ｐ.１－４

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ　１７／３０
Ｇ０６Ｆ　１７／２７