

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4672726号  
(P4672726)

(45) 発行日 平成23年4月20日(2011.4.20)

(24) 登録日 平成23年1月28日(2011.1.28)

(51) Int. Cl.	F I
<b>G06F 17/30 (2006.01)</b>	G06F 17/30 210A
<b>G06Q 30/00 (2006.01)</b>	G06F 17/30 170Z
	G06F 17/30 419B
	G06F 17/60 326

請求項の数 9 (全 12 頁)

(21) 出願番号	特願2007-525710 (P2007-525710)	(73) 特許権者	506329306
(86) (22) 出願日	平成17年8月8日(2005.8.8)		アマゾン テクノロジーズ インコーポレ イテッド
(65) 公表番号	特表2008-509501 (P2008-509501A)		アメリカ合衆国 89507 ネバダ州
(43) 公表日	平成20年3月27日(2008.3.27)		レノ ビーオー ボックス 8102
(86) 国際出願番号	PCT/US2005/028148	(74) 復代理人	100115624
(87) 国際公開番号	W02006/020576		弁理士 濱中 淳宏
(87) 国際公開日	平成18年2月23日(2006.2.23)	(74) 復代理人	100120581
審査請求日	平成19年4月6日(2007.4.6)		弁理士 市原 政喜
(31) 優先権主張番号	10/914,722	(74) 代理人	100077481
(32) 優先日	平成16年8月9日(2004.8.9)		弁理士 谷 義一
(33) 優先権主張国	米国 (US)	(74) 代理人	100088915
			弁理士 阿部 和夫

最終頁に続く

(54) 【発明の名称】 キーワードターゲット広告を発行する際に、使用キーワードを識別する方法及びシステム

(57) 【特許請求の範囲】

【請求項1】

商品を示す情報と、前記商品の特徴を示す記述とを関連付けて記憶する記憶装置と、制御部とを備え、通信リンクを介してウェブサーバーと、サーチエンジンサービスとに接続される広告システムにおいて、広告対象となる商品について最適な句を特定する方法であって、

前記制御部が、

a) 前記広告対象となる商品に関連付けられた記述を前記記憶装置から取得し、取得した記述に対応するリンク情報を前記サーチエンジンサービスから取得し、当該リンク情報に基づいて前記ウェブサーバーから複数のドキュメントを取得するステップと、

b) 前記複数のドキュメントに含まれる単語ごとに、前記複数のドキュメントにおける出現頻度に基づいたスコアを算出するステップと、

c) スコアが算出された全ての前記単語のうち、上位のスコアを有する単語を関連する単語として、所定の割合で選択するステップと、

d) 前記関連する単語を含めた句を生成するステップと、

e) 前記ドキュメントにおいて、前記句に含まれる関連する単語の後の関連する単語について前記句に含めるか否かを判定するステップと、

f) ステップ e) において前記後の関連する単語を前記句に含めると判定した場合に、前記後の関連する単語を前記句に含めて前記句を拡張するステップと、

g) ステップ e)、f) を繰り返して、前記後の関連する単語を前記句に含めないと判

定されるまで前記句の拡張を実施するステップと、

h) 前記句に含まれる関連する単語から算出される句のスコアに基づいて、前記句から最善句を識別する

ことを特徴とする方法。

【請求項 2】

ステップ a) において、前記制御部は、前記記述を前記サーチエンジンサービスに送信して、前記サーチエンジンサービスによる検索の結果として前記リンク情報を取得するステップを含むことを特徴とする請求項 1 に記載の方法。

【請求項 3】

前記記憶装置は、前記商品を示す情報と関連付けて広告を記憶しており、  
ステップ h) において 識別された最善句と、前記広告対象となる商品に係る広告との組を前記サーチエンジンサービスに送信するステップと、  
ユーザからの検索クエリに係るワードと前記 識別された最善句とが一致する場合に、前記サーチエンジンサービスが前記広告対象となる商品に係る広告を表示するステップと、  
をさらに備えることを特徴とする請求項 1 に記載の方法。

10

【請求項 4】

前記記述は前記商品に関するタイトルであることを特徴とする請求項 1 に記載の方法。

【請求項 5】

ステップ b) のスコアの算出において、前記複数のドキュメントに含まれるノイズワードは、前記複数のドキュメントに含まれないものとして処理されることを特徴とする請求項 1 に記載の方法。

20

【請求項 6】

ステップ e) において、前記後の関連する単語が前記句に含まれる関連する単語と類似する場合、前記句に含めないと判定されることを特徴とする請求項 1 に記載の方法。

【請求項 7】

ステップ a) において取得される前記複数のドキュメントはそれぞれ類似していないことを特徴とする請求項 1 に記載の方法。

【請求項 8】

請求項 1 または 2 のうちのいずれか 1 項に記載の各ステップをコンピュータに実行させるための命令を記録したコンピュータ読取可能な記録媒体。

30

【請求項 9】

商品を示す情報と、前記商品の特徴を示す記述とを関連付けて記憶する記憶装置と、制御部とを備え、通信リンクを介してウェブサーバーと、サーチエンジンサービスとに接続され、広告対象となる商品について最適な句を特定するための広告システムであって、

前記制御部が、

前記広告対象となる商品に関連付けられた記述を前記記憶装置から取得し、取得した記述に対応するリンク情報を前記サーチエンジンサービスから取得し、当該リンク情報に基づいて前記ウェブサーバーから複数のドキュメントを取得する取得手段と、

40

前記複数のドキュメントに含まれる単語ごとに、前記複数のドキュメントにおける出現頻度に基づいたスコアを算出する算出手段と、

スコアが算出された全ての前記単語のうち、上位のスコアを有する単語を関連する単語として、所定の割合で選択する選択手段と、

前記関連する単語を含めた句を生成する生成手段と、

前記ドキュメントにおいて、前記句に含まれる関連する単語の後の関連する単語について前記句に含めるか否かを判定する判定手段と、

前記判定手段において前記後の関連する単語を前記句に含めると判定した場合に、前記後の関連する単語を前記句に含めて前記句を拡張する拡張手段と、

前記判定および前記拡張を繰り返して、前記後の関連する単語を前記句に含めないと判

50

定されるまで前記句の拡張を実施する句拡張手段と、

前記句に含まれる関連する単語から算出される句のスコアに基づいて、前記句から最善句を識別する最善句識別手段と、

を実行するように構成される

ことを特徴とする広告システム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、一般的に商品 (item) に関連する用語に関し、特に、商品に対して広告を発行 (place) する際に使用する用語を検索することに関する。

10

【背景技術】

【0002】

グーグルやオーバーチュア等の多くの検索エンジンサービスは、インターネットを経由してアクセス可能な情報に対する検索を提供する。これらの検索エンジンサービスは、ユーザがユーザにとって興味のあるウェブページ及び他のインターネットアクセス可能なリソースを検索できるようにする。ユーザが検索用語を含む検索要求をサブミットした後、検索エンジンサービスはそれらの検索用語に関連したウェブページを識別する。早く関連するウェブページを識別するために、検索エンジンサービスはウェブページへのキーワードのマッピングを保守する。このマッピングは、各ウェブページのキーワードを識別するために、ウェブ (すなわち、ワールドワイドウェブ) を「巡回する (crawling)」ことによって生成することができる。ウェブを巡回するために、検索エンジンサービスは、ルートウェブページを介してアクセス可能な全てのウェブページを識別するために、ルートウェブページのリストを使用することができる。任意の特定のウェブページのキーワードを、見出しの単語、ウェブページのメタデータにおいて供給される単語、ハイライトされた単語等、様々な周知の情報取得技術を用いて識別することができる。それでもなお、いくつかの検索エンジンサービスは、インターネットを介してアクセスできない検索情報源を有することができる。例えば、本の発行者は、検索エンジンサービスに利用できるその本のコンテンツを作成することができる。検索エンジンは、キーワードと本との間のマッピングを生成する。検索エンジンサービスは、1又は複数の検索語を含む検索結果を受信した場合、そのキーワードが検索用語に最も近似する情報ソース (例えば、ウェブページ又は本) を識別するために、そのマッピングを使用する。検索語に最も近似する情報源の集合を「検索結果」という。検索エンジンサービスは、各合致、ウェブページの人気度 (例えば、グーグルのページランク) 等の近似度に基づいて、検索結果の情報源をランク付けする。次に、検索エンジンサービスは、そのランキングに基づいた順でそれらの情報源へのユーザリンクを表示する。

20

30

【0003】

いくつかの検索エンジンサービスは、検索結果におけるウェブページへのリンクを含め、ウェブページのプロバイダに課金しない。むしろ、検索エンジンサービスは、検索結果とともに広告を発行することで、収益を得る。これらの有料広告は、一般的に「スポンサーリンク」、「スポンサーマッチ」又は「有料検索結果」と呼ばれる。ある検索結果とともに商品についての広告を発行したい広告主は、広告及び検索語とともに検索エンジンサービスを提供する。検索結果を受信すると、検索エンジンサービスは、検索語が検索要求のそれに最も近似している広告を識別する。検索エンジンサービスは、検索結果を伴う各広告の発行に対する課金 (すなわち、広告当りのコスト)、又はユーザが広告に関連するリンクを実際に選択した場合のみの課金 (すなわち、クリック当りのコスト) のいずれかを行うことができる。

40

【発明の開示】

【発明が解決しようとする課題】

【0004】

広告主は、検索結果とともに発行された広告に対する支払いに使用された広告料の効果

50

を最大化することを望む。これらの広告主は、広告主にとって最も高い利益 (benefit) (すなわち、最も高い利潤 (profit)) となる広告対象の商品に対する検索語を特定しようとする。より目的に適った又は広告対象の商品に関連する検索語を特定することによって、広告主が広告料の効果の最大限とすることを可能とする技術が望まれる。

【課題を解決するための手段】

【0005】

検索結果とともに広告を発行するための検索語を識別するための方法及びシステムを提供する。一実施形態では、広告システムは、広告対象となる商品の記述 (description) を選択する。例えば、商品が本の場合、記述は本のタイトルとすることができ、又は、商品が電子装置の場合、記述は装置の短い要約とすることができる。次に、広告システムは、情報源のコーパスから、選択された記述に合致する (例えば、その主題に最も関連する) 文書又は他の情報源を取得する。例えば、広告システムは、取得された文書である検索結果のウェブページとともに、選択された記述をサーチエンジンサービスにサブミットする (submit) ことができる。次いで、広告システムは、広告対象の商品に対する単語の関連を示す、取得された文書の各単語に対するスコアを生成する。一実施形態として、広告システムは、情報源のコーパス内よりも取得された文書内で頻繁に使用される単語に対する高いスコアを生成することができる。例えば、商品がハリーポッターの本の場合、「Hogwarts」、「Fluffy」、「three-headed」、「dog」、「Hermione」及び「Granger」等の単語は、比較的高いスコアを有する。というのは、これらの単語は、ハリーポッターの議論において、関連のない議論よりもより頻繁に現れるからである。単語に対するスコアを生成した後、広告システムは、商品に関連しそうな文書内の単語の句を識別する。例えば、広告システムは、句「Fluffy the three-headed dog」及び「Hermione Granger」は、その本に関連しそうであることを識別する。そして、広告システムは、識別した句から広告対象の商品に対する検索語を生成する。広告システムは、商品についての有料広告の発行のために、検索語及び広告をサーチエンジンサービスにサブミットする。例えば、広告システムは、検索語「Hermione Granger」を用いて、ハリーポッターの本についての広告を発行することができる。ある人がサーチエンジンサービスに「Hermione Granger」の検索要求をサブミットした場合、検索結果とともにその広告を表示する。このようにして、広告システムは、広告対象の商品への関連が知られている情報源の中で使用されている句に基づいて、検索語を認識することができる。

【0006】

一実施形態では、広告システムは、広告対象の商品に関連しそうな句を識別する。 $n$  が文書内の単語数である場合、文書内の句の数は  $O(n^2)$  であり、 $k$  が異なる単語の数であり、 $l$  が句の長さである場合、文書のコーパス内の可能な句の数は  $k^l$  であるため、可能な各句をコンピュータで判断し追跡することは非常にコストがかかる。判断する句の数を減らすために、広告システムは相互に近似する高いスコアの単語のグループを判断する。広告システムは、まず、商品に関連する文書内の単語に対するスコアを生成する。そのスコアは、単語が商品に関連する可能性を示す。次に、広告システムは、極めて関連する単語と関連する単語とを特定する。極めて関連する単語は、上位 10% のスコア等の非常に高いスコアを有し、関連する単語は上位 25% のスコア等の高いスコアを有する。広告システムは、極めて関連する単語についての文書を検索する。文書内の極めて関連する単語各々は、句の「アンカーワード (anchor word)」とみなされる。広告システムは、近くの関連する単語を含むように句を拡張しようとする。一実施形態では、広告システムはアンカーワードに続く任意の近くの関連する単語によって句を拡張することができる。例えば、「Hermione」が極めて関連する単語であって、「Granger」が関連する単語である場合に、文書中で「Hermione」の次に「Granger」が続く場合、句「Hermione Granger」は句として識別される。代替として、広告システムは、アンカーワードの前の単語も含むように句を拡張することができる。例えば、「Granger」が極めて関連する単語であり、「Hermione」が単に関連する単語である場合、句「Hermione Granger」はそれでも識別される。広告システムは句のスコアを計算し、句の全ての単語が関連する単語か否かに関わ

らず、拡張句のスコアが増加する限り、句を拡張し続ける。当業者は、かかる句を識別するための技術を、広告に対する検索語を生成する以外にコンテキスト中で使用できることを理解するであろう。例えば、サーチエンジンサービスは、ユーザに提示するための追加の関連する情報源の配置についての検索要求があった場合、検索結果の中で識別される句を使用することができる。代替として、広告システムは、追加の関連する情報源からより多くの句を識別することができる。より一般的には、情報源のコーパスを与える場合、句の識別技術を使用して、情報源のトピックスを識別することができる。例えば、情報源がチャットの議論である場合、認識された句はそのチャットの議論の最も人気のあるトピックスを現すことができる。

【発明を実施するための最良の形態】

【0007】

図1は、一実施形態における広告システムの構成を示すブロック図である。広告システム110は、通信リンク103を介してサーチエンジンサービスコンピュータシステム101及びウェブサーバコンピュータシステム102に接続されている。広告システムは、サーチエンジンサービスコンピュータシステムに商品の記述をサブミットし、ウェブサーバコンピュータシステムによって提供される、合致するウェブページに対するリンクを受信する。次いで、広告システムは、合致するウェブページをウェブサーバコンピュータシステムから取得する。広告システムは、合致するウェブページから句を識別し、識別した句から検索語を導く(derive)。次いで、広告システムは、商品に対する広告とともに検索語をサーチエンジンサービスにサブミットする。サーチエンジンサービスは、検索語に合致する検索クエリに対する検索結果とともに、広告を表示する。

【0008】

広告システムは、広告セット生成コンポーネント111、単語スコアコンポーネント112、最善句識別コンポーネント113、句発見コンポーネント114、共通句スコアコンポーネント115、商品データ格納部116、検索結果格納部117及びスコア格納部118を含む。商品データ格納部は、商品の記述とともに広告対象の各商品の識別子(例えば、SKU)を含む。例えば、商品データ格納部を、広告対象の本の電子カタログとすることができる。各カタログのエントリは、商品識別子、タイトル、アンカーネーム、要約等を含む。検索結果格納部は、検索語が識別される商品に合致するウェブページを含む。スコア格納部は、単語に対するスコア及び検索結果格納部の句を含む。広告セット生成コンポーネントは、商品識別子を提供され、その商品を広告する際に使用すべき検索語(例えば、キーワード)を識別する。広告セット生成コンポーネントは、検索結果の提供をサーチエンジンサービスに要求し、その検索結果のウェブページを取得し、単語スコアコンポーネント及び最善句識別コンポーネントを起動し(involve)、そして、広告セットを生成する。単語スコアコンポーネントは、単語が商品に関連する可能性を示す検索結果の各単語に対するスコアを生成する。最善句識別コンポーネントは、句発見コンポーネント及び共通句スコアコンポーネントを起動し、商品に関連する可能性のある句を識別する。

【0009】

中央処理装置、メモリ、入力装置(例えば、キーボード及びポインティングデバイス)、出力装置(例えば、表示装置)及び格納装置(例えば、ディスク装置)を含むコンピュータシステム及びサーバに、広告システムを実装することができる。メモリ及び格納装置は、広告システムを実現する命令を含むことができるコンピュータ読取可能な媒体である。加えて、データ構造及びメッセージ構造は、通信リンク上の信号等のデータ送信媒体を介して保存され送信される。インターネット、ローカルエリアネットワーク、ワイドエリアネットワーク又はポイントツーポイントダイヤルアップ接続等、様々な通信リンクを使用することができる。

【0010】

図2は、一実施形態における広告セット生成コンポーネントの処理を示すフロー図である。そのコンポーネントは商品の識別子を渡され、商品と関連する可能性がある句から導かれた検索語とともに広告セットを返す。ブロック201において、商品はその商品の記

10

20

30

40

50

述を取得する。例えば、その記述を本のタイトル、又は製品名（例えば、「ソニー（登録商標）DVDプレイヤー」と組み合わせられた商品名とすることができる。ブロック202において、そのコンポーネントは、検索要求時に取得した記述を使用して、サーチエンジンサービスに検索を実行するよう要求する。そのコンポーネントは検索結果を受信する。検索結果がウェブページへのURL等のリンクの場合、そのコンポーネントはリンクされたウェブページを取得し、検索結果格納コンポーネントにそれを格納する。そのコンポーネントは、検索結果の最も合致するウェブページのみ（例えば、上位15）を使用する。ブロック203において、そのコンポーネントは単語スコアコンポーネントを起動し、検索結果における各単語に対するスコアを生成する。その起動されたコンポーネントは、スコア格納部にスコアを格納する。ブロック204において、そのコンポーネントは最善句識別コンポーネントを起動し、商品に最も関連する句を識別する。その起動されたコンポーネントは、スコア格納部に句のスコアを格納する。ブロック205において、そのコンポーネントは最善の句を用いて商品に対する広告セットを生成する。次に、そのコンポーネントは終了する。そして、これらの広告セットを1又は複数のサーチエンジンサービスにサブミットすることができる。

10

## 【0011】

図3は、一実施形態における単語スコアコンポーネントの処理を示すフロー図である。単語スコアコンポーネントは、検索結果格納部のウェブページに格納された各単語に対するスコアを生成する。そのコンポーネントは、スコア格納部にスコアを格納する。ブロック301-308において、そのコンポーネントは検索結果の中から各単語を選択し、そのスコアを計算することを繰り返す（loop）。ブロック301において、そのコンポーネントは検索結果の中から次の単語を選択する。判定ブロック302において、検索結果の中の全ての単語が選択された場合、そのコンポーネントはリターンし、そうでない場合、コンポーネントはブロック303に進む。当業者は、コンポーネントがノイズワード（noise word）（例えば、「of」、「a」、「the」等）をスキップできることを理解するであろう。ブロック303において、そのコンポーネントは、検索結果の文書（例えば、ウェブページ）内で選択された単語の平均頻度を計算する。単語の「頻度」は、文書内の単語の全発生数によって割り算された、文書内における単語の発生数である。例えば、200語を含む文書内で単語が10回発生した場合、この頻度は0.05（すなわち、10/200）であり、これは、それが文書内で単語の5%を占めることを意味する。検索結果内の単語の「平均頻度」は、各文書に対するその単語の平均頻度である。例えば、4つの文書を有する検索結果において、単語の頻度が0.05、0.04、0.02及び0.01の場合、その単語に対する平均頻度は0.03（例えば、 $(0.05 + 0.04 + 0.02 + 0.01) / 4$ ）である。平均頻度は次式で表される。

20

30

## 【0012】

## 【数1】

$$\bar{f} = \frac{\sum_{i=1}^n f_i}{n} \quad (1)$$

40

## 【0013】

ここで、

## 【0014】

## 【数2】

$$\bar{f}$$

## 【0015】

は、単語の平均頻度であり、 $f_i$ は文書*i*内の単語の頻度であり、*n*は文書の数である。

50

ブロック304において、コンポーネントは単語についての「通常の頻度」を取得する。通常の頻度は、全てのウェブページ等の非常に大きな文書のコーパスの中の単語の平均頻度を表す。ブロック305において、そのコンポーネントは選択された単語についての「頻度のスコア」を計算する。選択された単語の平均頻度が、選択された単語の通常の頻度よりもずっと大きい場合、その単語は商品に非常に関連する。頻度のスコアは、通常頻度に対する平均頻度のスコアを提供する。頻度のスコアは、次の式で表現される。

【0016】

【数3】

$$S_f = .5 + \frac{\text{atan}\left(\frac{\bar{f} - \tilde{f}}{10 * \tilde{f}}\right)}{\pi} \quad (2) \quad 10$$

【0017】

ここで、 $S_f$ は単語に対する頻度のスコアであり、

【0018】

【数4】

$\tilde{f}$

20

【0019】

は、単語の通常の頻度であり、 $\text{atan}$ はアークタンジェントの関数である。当業者は、この式は、頻度のスコアを生成するために使用可能な多くの式のうちの1つに過ぎないことを理解するであろう。単語の平均と通常との頻度の違いに与えられる重みに基づいて、使用する特別な式を選択することができる。ブロック306では、コンポーネントは、選択された単語を含む検索結果の文書の数を計算する。ブロック307では、コンポーネントは、選択された単語を含む検索結果の文書の割合(fraction)を示す「コンテンスコア(contain score)」を計算する。次の式でコンテンスコアを表すことができる。

【0020】

【数5】

$$S_c = \frac{n'}{n} \quad (3) \quad 30$$

【0021】

ここで、 $S_c$ はコンテンスコアであり、 $n'$ は選択された単語を含む検索結果の文書の数である。ブロック308では、そのコンポーネントは、選択された単語に対するスコアを計算する。一実施形態において、単語のスコアは頻度のスコア及びコンテンスコアの線形組合せである。頻度のスコア又はコンテンスコアが、単語が商品に関連する可能性の正確な表現と見なされるか否かを反映するために、頻度のスコア及びコンテンスコアの重みを設定することができる。次の式で単語のスコアを表現することができる。

40

【0022】

【数6】

$$S = \alpha * S_f + (1 - \alpha) * S_c \quad (4)$$

【0023】

ここで、 $S$ は単語のスコアであり、

【0024】

## 【数7】



## 【0025】

は、0から1まで変化し、頻度のスコアに与えられる重みを表現する。次に、コンポーネントは、ブロック301に戻り、検索結果の中から次の単語を選択する。

## 【0026】

図4は、一実施形態において、最善句識別コンポーネントの処理を示すフロー図である。ブロック401において、そのコンポーネントは、検索結果に極めて関連する単語を選択する。極めて関連する単語は、そのスコアが上位15%の単語とすることができる。極めて関連する単語は、句に対するアンカーワードとして使用することができる。ブロック402において、そのコンポーネントは、検索結果の関連する単語を選択する。関連する単語は、そのスコアが上位40%の単語とすることができる。関連する単語は極めて関連する単語を含む。その句を拡張して、アンカーワードに近い関連する単語を含むようにすることができる。当業者は、様々な基準を使用して、極めて関連する単語及び関連する単語を選択することができる。例えば、極めて関連する単語は上位スコア内の10語とすることができる、関連する単語は上位スコア内の50語とすることができる。加えて、極めて関連する単語及び関連する単語は、同じ単語のセット(例えば、上位スコア内の20語)とすることができる。ブロック403-405において、そのコンポーネントは、検索結果の中の文書を選択しそれらの文書内で句を見つけるのを繰り返す。ブロック403において、そのコンポーネントは検索結果の中の次の文書を選択する。判定ブロック404において、検索結果の中の全ての文書が既に選択された場合、そのコンポーネントはブロック406に進み、そうでない場合、そのコンポーネントはブロック405に進む。ブロック405において、そのコンポーネントは句発見コンポーネントを起動し、選択された文書内の句を見つける。次に、そのコンポーネントはブロック403に戻り、次の文書を選択する。ブロック406において、全ての文書内で句を見つけた後、そのコンポーネントは共通句、すなわち、文書内で頻繁に発生する句を選択する。例えば、文書内で5回以上発生するものを共通句とすることができる、又は、文書内のある割合で発生するものを共通句とすることができる。ブロック407において、そのコンポーネントは、共通句スコアコンポーネントを起動し、各共通句に対する句のスコアを生成する。次いで、そのコンポーネントはリターンする。広告システムは、共通句から検索語を取得する。

## 【0027】

図5は、一実施形態において句発見コンポーネントの処理を示すフロー図である。このコンポーネントは文書を渡され、その文書内で句を識別する。ブロック501-509において、そのコンポーネントは、極めて関連する単語をアンカーワードとして有する文書内で句の識別を繰り返す。ブロック501において、そのコンポーネントは文書内で次に極めて関連する単語を選択する。判定ブロック502において、その文書の全ての極めて関連する単語が選択された場合、そのコンポーネントは完了し、そうでない場合そのコンポーネントはブロック503に進む。ブロック503では、そのコンポーネントは、選択された極めて関連する単語を有する句をアンカーワードとして初期化する。ブロック504-509では、そのコンポーネントは、近い関連する単語を含む句の拡張を繰り返す。ブロック504では、そのコンポーネントは、文書内の次の単語を選択する。判定ブロック505では、選択された単語が関連する単語の場合、そのコンポーネントはブロック506に進み、そうでない場合、そのコンポーネントは句の拡張を終了し、ブロック501に戻り、文書内の次の句を識別する。判定ブロック506において、選択された単語が既に句の中にある単語に類似している場合、そのコンポーネントは句の拡張を終了し、ブロック501に戻り次の句を識別し、そうでなければそのコンポーネントはブロック507に進む。判定ブロック507において、選択された単語が句のスコアを向上(improve)させた場合、そのコンポーネントはブロック509に進み、そうでない場合はそのコンポーネントはブロック508に進む。判定ブロック508において、選択された単語及び選

10

20

30

40

50



扱された単語の後の次の単語が句のスコアを向上させた場合、そのコンポーネントはブロック509に進み、そうでない場合、そのコンポーネントは句の拡張を終了し、ブロック501に戻り次の句を識別する。ブロック509において、そのコンポーネントは選択された単語を句に追加し、ブロック504に戻り、句を拡張するために次の単語を選択する。

#### 【0028】

図6は、一実施形態において共通句スコアコンポーネントの処理を示す図である。そのコンポーネントは、共通句に対する句のスコアを計算する。代替として、各共通句が識別された時に、句のスコアを計算してもよい。ブロック601において、そのコンポーネントは次の共通句を選択する。判定ブロック602において、共通句が既に選択された場合、そのコンポーネントはリターンし、そうでない場合コンポーネントはブロック603に進む。ブロック603では、コンポーネントは、選択された共通句に対する句のスコアを初期化する。ブロック604 - 607において、そのコンポーネントは、共通句の単語の単語スコアを句のスコアに織り込む (factor) ことを繰り返す。判定ブロック605において、選択された共通句の全ての単語が既に選択された場合、そのコンポーネントはブロック607に進み、そうでない場合は、そのコンポーネントはブロック606に進む。ブロック606では、そのコンポーネントは選択された単語の単語スコアを句のスコアに追加し、次いでブロック604に戻り、選択された共通句の次の単語を選択する。当業者は、多くの異なる技術を使用して句のスコアの計算が可能であることを理解するであろう。例えば、極めて関連する単語の2倍の単語スコアを句のスコアに追加して、極めて関連する単語の重要性を強調し、単語のスコアの非線形の組合せを使用等することができる。ブロック607において、そのコンポーネントは、検索結果内で選択された共通句の発生数で句のスコアを乗算し、次に、そのコンポーネントはブロック601に戻り、次の共通句を選択する。

#### 【0029】

ここでは例示のために広告システムの具体的な実施形態を記述したが、当業者は本発明の思想及び範囲から逸脱することなく様々な変形が可能であることを理解するであろう。用語「商品」は、広告すべき任意の製品、サービス又は概念を含む。例えば、政党は、特定の候補又は大義 (cause) に関連する広告を発行することができる。加えて、広告セットは、それに関連するリンクを有さなくてもよい。広告主は、特定の検索語を用いて要求をサブミットするユーザに、広告の情報を単に表示することを望む。例えば、ユーザが対抗者の名前を検索語として検索要求をサブミットした時に、候補者は広告を表示することを望む。当業者は、スコアを計算するための様々な均等物や技術を使用できることを理解するであろう。また、検索結果が重複する (又はとても似ている) 文書を含む場合、広告システムは、重複文書を無視することができる。広告システムは、全てのウェブページに極めて共通の単語等 (例えば「次のページ」又は「プライバシーポリシー」) の、句に追加すべきでない単語のリストを保持することができる。したがって、本発明は特許請求の範囲によらない限り限定されない。

#### 【図面の簡単な説明】

#### 【0030】

【図1】一実施形態における広告システムの構成を示すブロック図である。

【図2】一実施形態における広告セット生成コンポーネントの生成処理を示す図である。

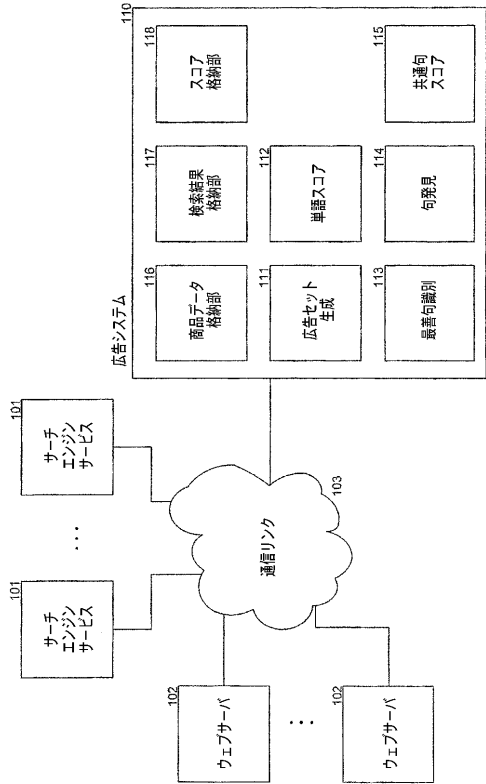
【図3】一実施形態における単語スコアコンポーネントの処理を示す図である。

【図4】一実施形態における最善句識別コンポーネントの処理を示す図である。

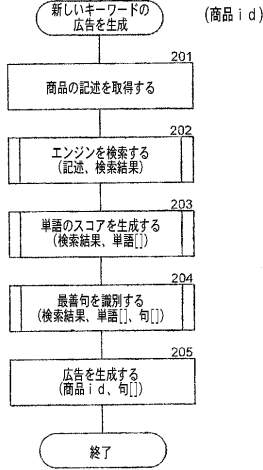
【図5】一実施形態における句発見コンポーネントの処理を示す図である。

【図6】一実施形態における共通句スコアコンポーネントの処理を示す図である。

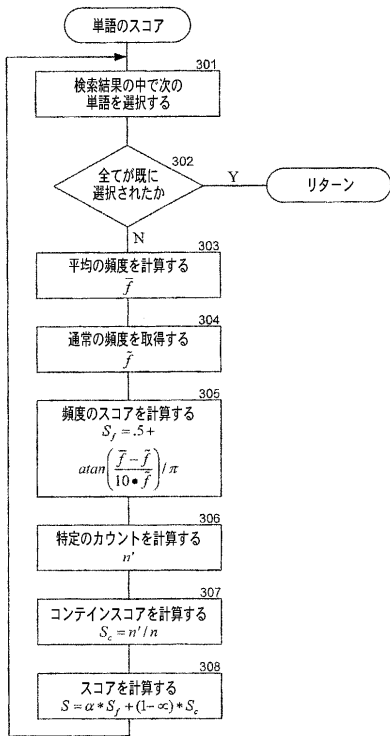
【図 1】



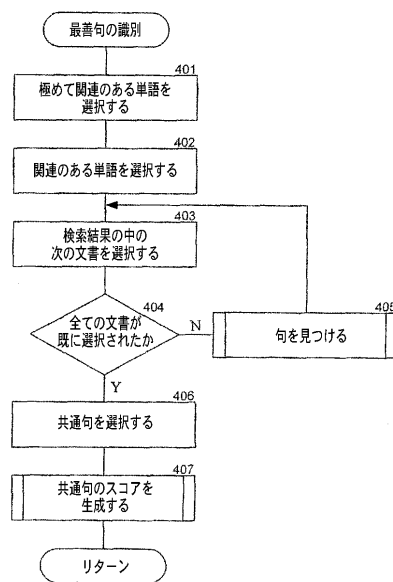
【図 2】



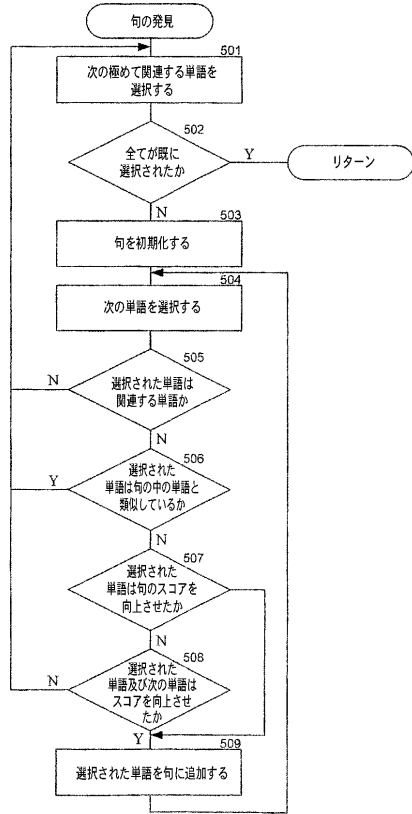
【図 3】



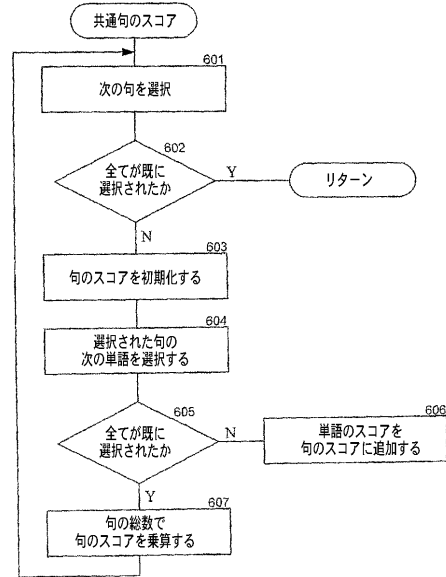
【図 4】



【図5】



【図6】



---

フロントページの続き

- (72)発明者 ナサニエル ビー・ショール  
アメリカ合衆国 98144 - 2734 ワシントン州 シアトル 12 アベニュー サウス  
1200 스위트 1200
- (72)発明者 アレクサンダー ダブリュ・デノイ  
アメリカ合衆国 98144 - 2734 ワシントン州 シアトル 12 アベニュー サウス  
1200 스위트 1200

審査官 岩間 直純

- (56)参考文献 特開2002 - 278949 (JP, A)

- (58)調査した分野(Int.Cl., DB名)

G06F 17/30

G06Q 30/00