(19) **日本国特許庁(JP)**

(12) 特 許 公 報(B2)

(11)特許番号

特許第5222581号 (P5222581)

(45) 発行日 平成25年6月26日 (2013.6.26)

(24) 登録日 平成25年3月15日(2013.3.15)

(51) Int.Cl. F 1

GO6F 17/30 (2006.01) GO6F 17/30 380E

GO6F 17/30 22OC GO6F 17/30 35OC

請求項の数 3 (全 19 頁)

(21) 出願番号 特願2008-32612 (P2008-32612) (22) 出願日 平成20年2月14日 (2008. 2. 14)

(65) 公開番号 特開2008-204454 (P2008-204454A) (43) 公開日 平成20年9月4日 (2008.9.4)

審査請求日 平成23年2月8日 (2011.2.8)

(31) 優先権主張番号 60/890,464

(32) 優先日 平成19年2月16日 (2007. 2. 16)

(33) 優先権主張国 米国 (US) (31) 優先権主張番号 11/837, 942

(32) 優先日 平成19年8月13日 (2007.8.13)

(33) 優先権主張国 米国(US)

||(73)特許権者 502096543

パロ・アルト・リサーチ・センター・イン

コーポレーテッド

Palo Alto Research Center Incorporated アメリカ合衆国、カリフォルニア州 94 304、パロ・アルト、コヨーテ・ヒル・

ロード 3333

|(74)代理人 100075258

弁理士 吉田 研二

(74)代理人 100096976

弁理士 石田 純

最終頁に続く

(54) 【発明の名称】文書に注釈を付するシステムと方法

(57)【特許請求の範囲】

【請求項1】

コンピュータに

複数の文書識別子とタグのそれらの関連セットとをURL/TAGペアのセットに変換する<u>手</u>順と、

前記URL/TAGペアセットを使用して二次元グラフを構成する手順と、

前記二次元グラフを使用し、URL/TAGのセットの文書識別子とタグに対する活性化拡散を用いタグプロフィールと文書プロフィールとを計算する手順と、

選択されたタグまたは文書に対して、選択されたタグまたは文書に関連したタグプロフィールまたは文書プロフィールをそれぞれ識別する手順と、

選択されたタグまたは文書に関連した、前記識別されたタグプロフィールまたは文書プロフィールのタグまたは文書を適切性に従って配列する手順と、

文書プロフィールを、同様なプロフィールを識別するために他の文書プロフィールに対してランク付けする手順と、

既定の閾値以上にランク付けされたそれらの同様なプロフィールの文書を、アウトプット用に識別する手順と、

を実行させるための探索プログラム。

【請求項2】

請求項1<u>に記載の探索プログラム</u>において、タグプロフィールと文書プロフィールとが 、ベクトルAとして活性化拡散を繰り返し使用して、次式:

A[1] = E;

A[2] = M*A[1] + E;

.

A[n] = M*A[n-1] + E;

(式中、

10

20

A [1] , A [2] , ... A [n] は、URLとタグの繰り返し計算されたプロフィールベクトルで、

Eは、タグまたは文書を表す単位ベクトルで、

Mは、選択された<u>タグまたは文書</u>に従って列または行で配列された<u>二次元グラフ</u>のマト リックス表現で、

と は、活性化拡散を調整するためのパラメータである)

に従って計算されることを特徴とする探索プログラム。

【請求項3】

請求項<u>1</u>に<u>記載の探索プログラム</u>において、前記ランク付けが、選択された文書が二次元グラフにないときでも、二次元グラフを構成するのに使用された文書であって、<u>前記選択された文書と</u>同等な類似度を有する文書をそれに代替することによって、実行されることを特徴とする探索プログラム。

【発明の詳細な説明】

【技術分野】

[00001]

以下は、一般に、文書に注釈を付し、その後そのような注釈を共有し、注釈付文書コレクションを探索するための方法、装置、および製造品に関連する。

【背景技術】

[0002]

ウェブページのソーシャルタグ付けを可能にするウェブサービスが今日、インターネット上で利用可能である。例えば、ヤフーのマイウェブ(MyWeb)やデリシャス(del.icio.us)などである。そのようウェブベースサービスを使用すると、ユーザは、興味あるウェブ文書(ウェブページなど)にタグを付け、ウェブ文書を共有または後で再び呼び出すことが可能である。これは、ユーザがウェブ文書をブックマークに記憶しておき、自由に選択されたタグ(またはキーワード)のセットをそのウェブ文書に付加することによって行い得る。ユーザは、自分のブックマークまたはタグを他のユーザと共有すると決めることも可能で、それらを、後で他のユーザが探索したり、ブラウズしたりすることが可能である。

[0003]

ユーザが、他のユーザによって定義・共有されたタグを経由して、ブックマークに記憶されたページを発見し得ることに加えて、ソーシャルタグ付加から得られたデータは、また、文書探索を強化するのにも使用し得る。しかし、ソーシャルタグ付加システムは、タグ付けされたウェブページのコンテンツの性質(例えば、ウェブページのコンテンツがダイナミックであるとか、または、あるウェブページのコンテンツが別のウェブページのそれと同様であり得るとかの性質)を考慮しないので、限定的である。例えば、ユーザが、ウェブ文書に付いている、以前に定義され、そして他のユーザと共有されたタグをレビューし、更新しない限り、URL(ユニバーサル・リソース・ロケータ)に関連した各ユーザ指定タグは、同じままである。これは、URLで示されたウェブページの基底コンテンツの

40

30

サブ文書エレメントが変化するときでさえも同様で、何故そのタグが文書に付されたかの 理由が不正確になったり、あるいは全く反映されなくなったりする。

[0004]

さらに、利用可能なソーシャルタグ付加システムは、公開ウェブ文書間の類似性を考慮していない。例えば、相異なるウェブサイトが同一または極めて類似のニュース記事を公開することもあり得る。この場合、利用可能なソーシャルタグ付加システムは、相異なるウェブ文書の公開コンテンツ間の類似性が考慮しないので、タグを同様なコンテンツに伝搬するようには適応していない。タグを付けた情報をそのように伝搬していければ、同様なコンテンツで公開された情報にタグを付けるユーザの努力が、有利に簡素化される筈である。また、利用可能なソーシャルタグ付加システムは、ウェブブラウザ(またはリーダある。また、利用可能なソーシャルタグ付加システムは、ウェブブラウザ(またはリーダが入って統合が行われていない。その代わり、利用可能なソーシャルタグ付けシステムでは、ユーザは、現在読んでいるウェブ文書とは独立なウェブページにアクセスする必要がある。相互運用性をそのように欠いているので、ウェブ文書に対してタグを新たに付加したり、またはレビューしたりしているとき同時にそのウェブ文書を参照するというユーザ能力が損なわれている。

[0005]

【特許文献1】米国特許第6,360,951号

【発明の開示】

【発明が解決しようとする課題】

[0006]

従って、サブ文書エレメント(ウェブ文書のパラグラフなど)のその場(in situ)タグ付加と、そのようなタグ付加されたデータ(または、より一般的には注釈されたデータ)の共有とをサポートするためのシステムと方法に対するニーズが依然として存在する。ウェブ文書のサブ文書エレメントにタブを付加するソリューションであって、ウェブブラウザと統合化されたソリューションがあれば、ウェブページに注釈を付加するために必要とされる認知的対話的オーバーヘッドの量が有利に減少する筈である。さらに、ウェブページのソーシャルタグ付けを容易にする統合ソリューションが提供されれば、ユーザは、相互協力やタグ付加されたデータの共有が有利に行い易くなる筈である。また、同様なコンテンツを有するウェブページにタグを自動的に伝搬したり、ウェブページのダイナミック性を考慮したりすることによって、タグとウェブページのサブ文書エレメントとの間の関連性の一貫性が維持される。

[0007]

さらに、タグ付加された(または、より一般的には注釈が付加された)文書のコレクションを相互協力的タグ付加(または、より一般的には相互協力的注釈付加)によって探索するための改良されたシステムと方法に対するニーズが依然として存在する。従って、文書のタグベースのコレクションを探索するための改良されたシステムと方法を提供し、そのような文書コレクションを探索する確度および / または精度を増すことは、有利な筈である。

[0008]

本開示のこれらと他の態様は、添付図面と関連させて以下の説明を読めば、容易に明快になることと思われる。なお、同じ様な部分には同じ参照数字を付してある。

[0009]

A.用語の定義

以下で定義される用語は、特許請求の範囲と図面と含む本特許出願明細書を通じて、その示された意味を有する。

[0010]

「文書」または「ウェブ文書」は、電子データのコレクションが閲覧されるときどのようにフォーマットにされているかに依拠して可変数のページを規定し得る電子データのコレクション、例えば、ウェブブラウザを使って閲覧し得る文書(例えば、ウェブページ、イメージ、ワード文書、およびポータブル・ドキュメント・フォーマット(pdf)形式の

10

20

30

40

20

30

40

50

文書)を意味するのに本明細書では使用される。文書を構成する電子データは、静的および/または動的コンテンツから成り得る。

[0011]

「サブ文書エレメント」は、単独に取り出されるとき、文書全体より小さいサイズの文書構造のエレメントを意味するのに本明細書では使用される。サブ文書エレメントは、ワード、イメージ、フレーズ、センテンス、パラグラフ、ページ、節、および章のセットから選択されるタイプであり得る。

[0012]

「共同ユーザ」は、汎用コンピュータの「ユーザ」(アクセスコントロールおよび/またはブロックを使用して許可される)に認識または知悉され得るユーザ(例えば、個人または個人の代理)またはユーザのグループ(例えば、グループまたは組織内の配信リスト)であって、例えば、ユーザまたはログインネームのような識別子を使用して汎用コンピュータのユーザに知悉、または知悉させられ得るユーザを意味するのに本明細書では使用される。

[0013]

B. ダイナミック注釈と探索のための操作環境

図1は、本明細書に開示の実施の形態を実行するための例示の操作環境100を示す。この例示の操作環境において、汎用コンピュータ104は、インターネットのようなネットワーク102上で共に通信し得るが、クライアント側またはサーバ側のいずれかの操作を実行するようにプログラムされている。当業者なら分かることであるが、クライアントサーバ配置の別法が存在する。例えば、中央ターミナルベース配置またはこれらの組み合わせ配置である。クライアント側操作は、クライアント側アプリケーションモジュール106をロードした汎用コンピュータ104で実行される。モジュール106は、注釈プラグイン108とウェブブラウザ110(例えば、インターネットエクスプローラ、ファイアフォックス、またはサファリ)を備える。サーバ側操作は、一個または複数個のサーバ側アプリケーションモジュール112をロードした汎用コンピュータで実行される。モジュール112は、一個または複数個の注釈モジュール114、ページタグ付加モジュール116、ウェブページ(またはウェブ文書)サーバ118、およびタグベース探索サーバ120のうちの一個またはそれらの組み合わせを備える。

[0014]

図2は、本明細書に開示の実施の形態、例えば、図1に示される実施の形態を実行する ための汎用コンピュータ104を示す。汎用コンピュータ104は、ハードウェア212 とソフトウェア 2 1 4 とを備える。ハードウェア 2 1 2 は、プロセッサ(すなわち、CP U)216、メモリ218(ROM、RAMなど)、持続的記憶装置220(例えば、C D-ROM、ハードドライブ、フロッピー(登録商標)ドライブ、テープドライブなど) 、ユーザI/O222、およびネットワークI/O224で構成される。ユーザI/O2 2 2 は、カメラ 2 0 4 、マイクロフォン 2 0 8 、スピーカ 2 0 6 、キーボード 2 2 6 、ポ インティングデバイス228(例えば、ポインティングスティック、マウスなど)、およ びディスプレイ230を備え得る。ネットワークI/O224は、例えば、インターネッ トのようなネットワーク102に接続し得る。汎用コンピュータ104のソフトウェア2 1 4 は、オペレーティングシステムソフトウェア 2 3 6 とアプリケーションソフト 2 4 0 を備え、クライアント側アプリケーションモジュール106またはサーバ側アプリケーシ ョンモジュール112の命令を含み得る。ソフトウェア214は、一般に、メモリ218 に読み込まれ、プロセッサ216に指定された操作を実行させる。ソフトウェア214は クライアント側アプリケーションモジュール106またはサーバ側アプリケーションモ ジュール112の命令とともにアプリケーションソフト240を含む。

[0015]

C. ダイナミック注釈エレメントと操作

図3は、図1に示される例示の操作環境100における注釈サーバ113と注釈クライアント(複数を含む)111とによって実行される諸操作を示すフローチャートである。

20

30

40

50

注釈サーバ113と注釈クライアント(複数を含む)1111とにより、サブ文書エレメントのダイナミック注釈が可能となる。例示であって、限定目的ではないが、本明細書に記載の例は、ウェブページのパラグラフなるサブ文書エレメントの注釈に関する。当業者なら分かるように、相異なるサブ文書エレメントと文書、またはこれらの組み合わせも、本発明で可能となるダイナミック注釈の操作を行う時に代わりに使用し得る。

[0016]

最初に302で、クライアント側アプリケーションモジュール106のウェブブラウザ110が注釈プラグイン108で初期化される。これには、サーバ側アプリケーションモジュール112と通信するための操作と、文書を増補するための操作とが含まれ、(a)ウェブブラウザ110の使って表示される文書の注釈のユーザ指定と、(b)ユーザ注釈と共同ユーザ注釈のデータの表示とが双方行われ得る。一般に、304と306との間のクライアント側で実行される操作と、305と307との間のサーバ側で実行される操作は、注釈でウェブページを増補することとユーザ注釈のための準備とに関し、一方、306と308との間のクライアント側で実行される操作と、307と309との間のサーバ側で実行される操作は、増補されたウェブページのユーザ注釈に関する。別の一つの実施の形態では、注釈プラグイン108の機能は、文書をユーザに送る前に注釈機能で文書を増補するプロキシサーバ(ネットワーク102上で操作)で提供し得る。

[0017]

3 1 0 で、ウェブブラウザ 1 1 0 は、ユーザリクェストを受理し、ウェブページをロードしてディスプレイする。 3 1 2 で、リクェストされたウェブページが、例えば、そのウェブページを記憶しているウェブページサーバ 1 1 8 のようなサーバのネットワーク 1 0 2 上に位置を識別するURL(ユニバーサル・リソース・ロケータ)を使用してアクセスされる。 3 1 4 で、 3 1 2 でのウェブページリクェストに呼応した注釈プラグイン 1 0 8 が、ウェブサーバ 1 2 8 経由で注釈サブレット 1 2 6 とリクェストされたウェブページとを通信させ(例えば、そのURLを送信して)、サービスがサーバ側注釈モジュール 1 1 4 で実行されるようにリクェストし、リクェストされたウェブページのコンテンツと同様なコンテンツに対してユーザと識別された共同ユーザとが行った注釈を識別する。

[0018]

3 1 6 で、サーバ側注釈モジュール 1 1 4 は、ウェブページサーバ 1 1 8 から当該ウェ ブページのコピーを(例えば、314で提供されたURLを使用して)検索する。別法と して、312で検索されたウェブページのコピーを、314でのサービスリクェストに従 ってクライアント側アプリケーションモジュール106で提供し得る。318で、サーバ 側注釈モジュール114が、検索されたウェブページに一個または複数個のウェブページ サブ文書エレメントを識別する。これらのサブ文書エレメントは、パラグラフのような単 一のタイプでもよく、またはパラグラフと節のような諸タイプの組み合わせでもよい。3 20で、318で識別されたサブ文書エレメントであって、ユーザと選択された共同ユー ザに関連しているサブ文書エレメントに対して、類似度測定が、類似度計算器124を使 用して計算される。類似度は、一個または複数個の因子(ユニークな識別子または指紋と して作用する)、例えば、(a)サブ文書エレメントに出現するワードの長さ、(b)サ ブ文書エレメントに出現する最初の n 個のワードの最初の文字、(c)サブ文書エレメン トに出現する同様なノンストップのワードの頻度、および(d)暗号ハッシュ関数の一つ であるMD5(メッセージ・ダイジェスト・アルゴリズム5)に基づいて計算し得る。実 施の形態の一つでは、類似度(または指紋)は、関連サブ文書エレメントに対して計算さ れるハッシュ値である。計算後は、注釈サーバ113に記憶された注釈済みサブ文書エレ メントの指紋を、318で識別された、検索されたウェブページのサブ文書エレメントの 指紋に対して比較し得る。

[0019]

3 2 2 で、ユーザまたは共同ユーザが以前に注釈したウェブページサブ文書エレメント (関連した注釈と類似度を有する)であって、3 2 0 で計算された類似度と同等の類似度 (すなわち、同一または同様な指紋である可能性が高いと考えられる)でデータベース 1 22に記憶されているウェブページサブ文書エレメントが、識別される。324で、32 2で識別されたサブ文書エレメントの記憶された注釈データが、クライアント側アプリケーションモジュール106に提供される。

[0020]

326で、ウェブページが増補され、ディスプレイと(さらなる)注釈が行われるが、それは、326(A)で、ウェブページ上の各ワードを別々に選択可能とするステップで設別されたサブ文書エレメントの終わりにユーザネームのラベルと、322で説別されたサブ文書エレメントの関連する記憶されたユーザの注釈データであって、テキスト注釈(例えば、タグ、キーワード、またはコメント)、グラフィック注釈(例えば、オーディオまたはビデオクリップへのリンク)を含む注釈データとを挿入するステップと、326(C)で、32で識別された記憶・注釈されたデータと関連したウェブページのサブ文書エレメントのコンテンツに注釈(例えば、テキスト、グラフィックス、オーディオ、またはビデオを含み得るハイライトしたもの)を挿入するステップとで行われる。326(B)と326(C)とで、それぞれ、ウェブページに挿入された注釈は、サブ文書エレメントがごのサブ文書エレメントに類似しているで、これでで、それぞれ、ウェブページに挿入された注釈は、サブ文書エレメントがでいまたはそのマッチングする指紋を有している)とき、318で識別・検索されたウェブページのサブ文書エレメントの組み合わせで規定された文書の章)に付加し得る。

[0021]

326での実施の形態の一つで、注釈プラグイン108は、ウェブページのHTMLコンテンツを増補することによって(例えば、AJAXを使用して)ウェブページ上の各ワードを選択可能にし得る。それは、(a)326(A)で、ウェブページの文書オブジェクトモデル(DOM)ツリーを、ウェブページがウェブブラウザ110でロードされる前にウェブページの各ワードをHTML(ハイパーテキストマークアップ言語)タグ<スパン>で囲むことによって変更するステップと、(b)326(D)で、ウェブページでユーザが実行したワード選択を検出するため各サブ文書エレメントにイベントリスナ、例えば、マウスイベントリスナを付加するステップとで行われる。

[0022]

図4は、外接ボックス長方形を使用してサブ文書エレメントの諸ワードが個々に選択可能(および/または注釈可能)とされた例示的テキストコンテンツ402を示す。例えば、外接ボックス404内でマウスなどのポインティングデバイスで選択(および/または注釈)するユーザが、ワード「San」を選択したと決定したとする。ユーザの好みに依存するが、すでにワードに関連しているダイナミック操作を設定し、選択(および/または注釈)操作に超える優先性を持たせ得る。例えば、ワード「San」がハイパーリンクと関連していたイベントにおいて、上位優先性がハイパーリンクに与えられているときは、そのリンクが辿られる。そうでなく、より高い優先性が注釈プラグインに関連した操作に与えられているときは、選択(および/または注釈)操作が実行される。別の実施の形態では、フレーズやセンテンスのような、より高いレベルの構成物が、フレーズまたはセンテンスを囲むとき、外接ボックス(どんな数の相異なる形状をも取り得る)内で選択可能(および/または注釈可能)とされる。

[0023]

326でウェブページを増補した後、ウェブページが330でディスプレイされ、ユーザに利用可能にされるが、その目的は、(a)ユーザまたは識別された共同ユーザが同じようなコンテンツに対して以前に付加した注釈を見たり、(b)さらに注釈を加えたりするためである。332で、注釈プラグインに内蔵のイベントハンドラが、表示されたウェブページで示されたユーザから受け取ったインプットに関連したイベントを受理し、これに応答する。注釈されるべきサブ文書エレメントの部分を示すイベントにより、データベース122にある注釈を(注釈に関連したサブ文書エレメントの類似度(または指紋)とともに)、それぞれ、334で送信し、342で記憶し、そして340でこれらの注釈を

10

20

30

40

20

30

40

50

ユーザに対してウェブページ上にディスプレイすることが行われる。例えば、図5は、その後に挿入されたユーザネームラベル504とともにサブ文書エレメント502を示す。この例では、ユーザは、先ず、テキストの最初(すなわち「CHI」)にマウスのポインタを合わせ、次いでマウスのポインタがテキスト文の最後(すなわち「2007」)に移動するまで左マウスボタンを押し続けて、下線が引かれたテキストをハイライトすることによって文書の注釈付けを行った。

[0024]

ユーザによるワード選択を示すイベントにより、イベントハンドラは、ウェブページにタグコメントフィールドを開き、そのフィールド内に選択されたワードをタグとして挿入する(そして実施の形態の一つではネームラベルフィールドにも挿入する)。選択されたタグまたはキーワードを保存するということを示すイベントにより、イベントハンドラは、保存されるタグおよび/またはキーワードをサーバ側注釈モジュール114に送信し、336でエラーチェック(スペルおよび/または句読法エラーの訂正、矛盾点の解消、および/または特定の注釈の終端ワード削除を含む)を行い、その後に342で(ワード選択に関連したサブ文書エレメントの類似度(または指紋)とともに)記憶する。336で打正された注釈は、338で注釈クライアント111に送り戻され、その後340でイスプレイされる。例えば、図6は、サブ文書エレメント602と、その後のユーザネームラベル604(図5に示されるものと同じ)とを示し、その間にワード選択がサブ文書エレメント602にタグ付けするために行われる。

[0025]

具体的に説明すると、図6には、ビューアに表示されたウェブページのタグ付けをユー ザが実行する際の一連のイメージフレームが示される。フレーム600では、サブ文書エ レメント602の増補は、ユーザネームラベル604を後で挿入することと、サブ文書エ レメント602の個々のワードを選択可能とすることによって行われる。フレーム606 では、ワード「San」を選択するユーザインプットに応答し、ディスプレイされた文書 は、選択されたワード「San」をユーザネームラベル604の後に続いている(すなわ ち、選択されたワードに関連するサブ文書エレメントの後に続いている、またはこれに関 連する)タグコメントフィールド607に挿入することによってさらに増補されている。 フレーム 6 0 8 では、表示された文書は、選択されたワードである「Jose,」、「April」 「28th,」、「2007」、「May」、および「3rd,」でさらに増補されている。「save」ボ タン605がクリックされると、選択された(またはタイプされた)注釈文が、注釈サー バ113によって実施の形態の一つでは(または示されていないが、別の実施の形態では 注釈プラグイン108によって)図3の336で訂正され(例えば、最初はワードの一部 であると考えられていたコンマが除去され)、その後で選択されたワードが最終的に決定 され、フレーム610に示されるようにユーザネームラベル604の後に続いて示される 。別の実施の形態では(図示せず)、共同ユーザに関連したキーワードのリストと、最も 頻繁におよび/または最も最近に使用されたキーワードの選択可能のリスト(ウェブペー ジに、例えば、ユーザと共同ユーザのキーワードの後に続いて追加された)を、同様に選 択可能とし、タグコメントフィールド607に追加し得る。さらに別の実施の形態では、 選択れたワードを、最終決定が行われる前にタグコメントフィールド607に配置せず、 代わりに、直ちに、フレーム610に示される形式で配置する。

[0026]

有利なことであるが、ウェブページに注釈を加えた後、そのユーザ注釈は、データベース122にある注釈サーバ113に記憶され、同様なコンテンツを有するウェブページにその注釈を伝搬することが可能となる。換言すれば、注釈が伝搬するのは、注釈サーバ113が、記憶された注釈を注釈クライアント111に提供し、記憶されたコンテンツと同様なコンテンツを含む文書にその注釈を加える時である。さらに、注釈サーバ113は、クライアント111には、サブ文書エレメントのコンテンツの指紋とマッチしていない記憶された注釈についてはこれらを伝搬しない。たとえ、そのコンテンツが、ウェブページのサブ文書エレメントの注釈と指紋とが最初に記録されたときに閲覧中のウェブページに

以前存在していた場合でも、そうである。

[0027]

D.注釈の作成・共有のためのユーザインターフェイス

図7は、注釈を作成し、共有するための例示的ユーザインターフェイス700を示す。 ユーザインターフェイス700のトップには、「remove」と「add」コマンドを含む制御 エリア702があり、ユーザは、共同ユーザを指定して、その共同ユーザの注釈が、ユー ザが閲覧中の文書中にディスプレイされるのを見たいかどうかを決め得る。

[0028]

さらに、エリア702で、ユーザは、ユーザのページレベルタグ付けを指定し、指定された共同ユーザのページレベルタグを見ることが可能である。ページレベルのタグ付けは、サブ文書レベルでのタグ付けとは相異なって、単にタグ、キーワード、またはコメントをウェブページ(またはURL)と関連付けするものである。サブ文書エレメントのタグ付と同様に、ページレベルのタグ付けは、706に記録される。これは、実施の形態の一つでは、エリア702の「Save Page Tag」および「Update Page Tag」というコマンドを使い、ページタグサーブレット138によってウェブサーバ136経由でデータベース140にアクセスすることを含む(これは、別の実施の形態では、サーバ側注釈モジュール114と統合して、または一緒に操作し得る)。

[0029]

制御エリア702の下には、文書エリア704がある。文書エリア704では、ウェブページコンテンツが増補され、それは、この例では、ユーザ(「lichan_hong」)と選択された共同ユーザ(「edhchi」と「kooltag」)に関連した注釈を含む。従って、制御エリア702と文書エリア704とを調べることによって、ユーザは、ディスプレイされた文書がページレベルで、および/またはサブ文書レベルで注釈されているかどうかを知ることが可能である。

[0030]

E. タグベースの探索

ウェブベースの注釈サービス、例えば、クライアント側注釈モジュール106で利用可能となったものを使用して、ユーザは、ウェブ文書をブックマークに記憶し、ページルで文書エレメントに(そして、別の実施の形態では、サブ文書エレメントレー般的には、カリーを付けたりするによる注釈付加する)能力がでは、ハイライトしたりコメントを付けたりすることによる注釈付加する)能力がクラークには、ユーザの夕グのセットを使用して、ユーザの個人的ブックマークにそして、関連タグまたはキーワード)をすべてのユーザは、カーケックマーク(そして、関連タグまたはキーワード)をすべてのユーザに完全公開して、マークに記憶された共同ユーザに半公開して共有するかを決定し得る。共有されたブックマークは、そうすれば他のユーザがブラウズしたり、探索したりするのが可能となる。ユーザがようすれば他のユーザが対方ウストリカションをこのように協力的に共有するによって、ユーザは、ユーザが共有する(すなわち、協力的に開発された)ブックマークコレクションを使用して諸々のブックマークに記憶された文書を発見するという便益を享受し得る。

[0031]

図1に示された実施の形態の一つでは、データベース122と140に格納された協力的に共有されたブックマークコレクションは、ウェブブラウザ110上で作動するユーザインターフェイスを通じてタグベースの探索サーバ120を使用して探索し得る。ブックマークに記憶された文書を発見するための照会がユーザによって行い得る前に、ユーザのグループに関連したブックマークコレクションが処理され、キャッシュメモリに記憶され、ブックマークコレクション中のブックマークに記憶された文書に対するタグプロフィールと文書プロフィールが識別される。一般に、選択されたタグオブジェクト(例えば、TAG2)と選択された文書オブジェクト(例えば、URL1)に対するタグプロフィールと文書プロフィールは、エントリノードとしてバイグラフ(bigraph)中のその選択されたタグまた

10

20

30

40

は文書オブジェクトを使用する活性化拡散を行って構成される。

[0032]

図8は、タグベースの探索サービスで実行される操作を示すフローチャートで、タグと文書のプロフィールを識別し(802,804,806,808,810,および812で)、タグベースの照会に結果を提供する(814,816,818,および820で)ものである。最初に802で、ブックマークコレクションが識別される。これは、実施の形態の一つでは、文書識別子(例えば、URL)のリストとすることができる。この場合、ブックマークコレクション中の各ブックマークは、タグのセット(例えば、[URL1,TAG2,TAG6,...,TAGN])に関連したURL付の文書を識別するのである。804で、ブックマークコレクションはURL/TAGペアのセット(例えば、[URL1,TAG2]、[URL1,TAG6]、...[URL1,TAGN])に分解(すなわち、変換)される。[URL,TAG]タプルの形式のURL/TAGペアが与えられると、タグに対するバイグラフ(すなわち、二次元グラフ)が、806で構成される。バイグラフが構成されると、バイグラフはマトリックスとして表現される(808で)。別法としては、バイグラフは、例えば、隣接リストを使用して示し得る。

[0033]

図9は、902の箇所でブックマークコレクションの例を示す。これは、904の箇所で識別され(図8の802で規定されたように)、そしてURL/TAGペアに処理され(図8の804で規定されたように)、906の箇所でバイグラフに処理され(図8の806で規定されたように)、そして908の箇所で、906に示されるバイグラフに対して、URLを識別する列と夕グを識別する行とから成るマトリックスに処理される(図8の808で規定されたように)。

[0034]

8 1 0 で、バイグラフ行列(または、より一般的にはn次元行列)を使用して、バイグラフのノード範囲に対してタグプロフィールと文書プロフィールとを計算する。これらのプロフィールは、ベクトルAとして活性化拡散を繰り返し使用してプロフィールを計算することによって得られる。具体的には、以下の通りである。

〔数1〕

A[1] = E;

A[2] = M*A[1] + E;

•

A[n] = M*A[n-1] + E;

式中、

A [1] , A [2] , ... A [n] は、URLとタグの繰り返し計算されたプロフィールベクトルで、

Eは、タグまたは文書エントリノードを表す単位ベクトルで、

Mは、選択されたエントリノードに従って列または行で配列されたバイグラフ(または、より一般的にはn次元グラフ)のマトリックス表現で、

と は、活性化拡散を調整するためのパラメータである。

[0035]

活性化拡散を「n」ステップだけ繰り返して実行した後(ステップ「n」の数は、精度および / または性能に依存して変え得る)、活性化拡散は、バイグラフのタグ側またはバイグラフの文書側で停止され、従って、タグまたは文書エントリノード E に対するタグプロフィールベクトルまたは文書プロフィールベクトルを提供する。タグプロフィールベクトルと文書プロフィールベクトルにおける得られた重みパターンは、「タグプロフィール」と「文書プロフィール」をそれぞれ規定する。図10は、「n」=4ステップ、エントリノード E として文書1(すなわち、URL1)、パラメータ = 1、パラメータ = 0、文書D1に対して得られたタグプロフィールベクトルA[4]_{TAG}および得られた文書プロフ

10

20

30

40

20

30

40

50

ィールベクトルA[3]_{DOC}を有する活性化拡散の例を示す。

[0036]

図10に示されるように、各繰返し計算(すなわち、n=2、n=3、n=4)の後、得られた文書とタグのプロフィールベクトルが、初期エントリベクトルに対する最も興味あるタグまたは文書を識別する。この場合、ユーザは文書 DOC_1 に興味を示したのである。すなわち、初期エントリベクトルは、 DOC_1 に1の重みを割り当てたのである。ステップ n=2 で、図10は、 DOC_1 が TAG_1 、 TAG_2 、および TAG_3 と関連していることを示す。図10はさらにステップ n=3 で、活性化拡散計算は、興味ある文書 DOC_1 、 DOC_2 、 DOC_3 、および TAG_4 で、活性化拡散計算は、興味ある文書 TAG_4 で、 TAG_4 で TAG_4 で

[0037]

タグプロフィールと文書プロフィールとが、810で計算された後、それらは、実施の形態の一つではキャッシュメモリに記憶され(812で)、後で検索・探索目的に供される。別の実施の形態では、そのようなタグと文書のプロフィールは、オンデマンドでリアルタイムに計算される。ユーザは、ウェブブラウザ110を通して入手可能の図11に図示の探索インタフェース1100を使用して文書コレクションを探索し得る。探索インタフェース1100を使用して文書コレクションを探索し得る。探索インタフェース1100を使用して文書コレクションを探索し得る。探索インタフェース1100を使用して、関連したタグと文書(1104で)を識別する(1106で)。これは、1102でキーワードおよび/またはタグを取り、キャッシュメモリに記憶された活性化拡散計算(図8の812で)で、どんな他の文書とタグが関連し、興味あるものであるかを探し出すことによって行われる。文書プロフィールベクトルに規定された活性化拡散されたタグは、1106にディスプレイされる。これらは、活性化拡散の重みに従って類似度でランク付けされた順序でディスプレイされる

[0038]

前記の探索結果は、1104に表示されたどの文書が適切であるかを、例えば、「適切性フィードバック」として技術に既知の方法を使用して規定することによって、ユーザがさらに改良し得る。実施の形態の一つでは、ユーザは、選択ボックスをクリックすることによって自分の興味を示す。選択される文書は、キャッシュメモリに記憶された計算結果(図8の812で)中に探し出し得るが、プロフィールベクトルにある対応する関連タグと文書も、将来の探索を通知するのに使用し得る。別法として、ユーザは、関連タグを指定し、それらのタグに対応するキャッシュメモリに記憶された計算結果(812での)を探し出し、現在の興味プロフィールに追加し、次いでこれを使用して、将来の探索を通知することもし得る。

[0039]

より一般的には、キャッシュメモリに記憶された(812で)後は、タグプロフィールと文書プロフィールとは、検索、探索、および推薦事項のための相異なる類似度計算とルックアップの基礎を形成する。例えば、図8に示される実施の形態では、選択されたタグまたは文書に誘わる関連したタグまたは文書を識別するようにという探索リクエストを受け取ると(814で)、選択されたタグまたは文書に対応するタグプロフィールまたは文書プロフィールがキャッシュメモリから取り出される(816で)。別法として816で、選択されたタグまたは文書が、812でキャッシュメモリに記憶されたデータにあるプロフィールに存在しない場合は、812でキャッシュメモリに記憶されたデータにあるタグおよび/または文書であって、行方不明のタグまたは文書と最も類似のもののプロフィールが選択される。ステップ816で、文書(およびタグ)の間の類似性は、例えば、ユークリッド距離または他のベクトル距離計量を使用して決定し得る。

20

30

40

50

[0040]

820で、識別されたタグプロフィールおよび / または文書プロフィールが分類され、関連したタグおよび / または文書が重要度に従ってランク付けされる。ランク付けされた、これらの非常に類似性の高いタグおよび / または文書は、例えば、図11に示されるユーザインターフェイス1100でリターンして提示される。図示のユーザインターフェイス1100でリターンして提示される。図示のユーザインターフェイス1100は、各文書リンク(例えば「Dive into Greasemonkey」)とともに、最も高頻度で文書に付加されたタグのリスト(例えば、greasemonkey 65, firefox 54, javascript 33, programming 27, reference 18、およびweb 18)を含む。別の一つの実施の形態では、816で識別されたタグプロフィールまたは文書プロフィールが、818でシステムに含まれる他のすべての文書またはタグのプロフィールと対照して比較され、類似性の高いプロフィールが見出される。従って、選択されたタグおよび / または文書と対応または同様な識別されたタグおよび / または文書のプロフィールを使用して、816で識別されたものに加えて、または別のものとして、820で同様なタグまたは文書が識別される。

[0041]

換言すれば、814,816,818,および820で、ユーザがある文書に関連した 情報を見出したいと思えば、タグベースの探索サーバは以下のことを実行し得る。すなわ ち、(a)文書がそのキャッシュに記憶された(812で)活性化拡散計算によるプロフ ィールに既に存在すると仮定して、対応する文書プロフィールをルックアップし、そして 適切性に従って、例えば、最高から最低までの重み付けで配列されたそのプロフィールで 選択された文書を選択し、リターンし、ディスプレイすること(文書の配列は、さらに、 ある閾値以上に重み付けされたもののみを残すようにフィルタ処理され得る)、(b)対 応する文書プロフィールを使用し、これをシステムに含まれる他のすべての文書プロフィ ールと対照して比較し、類似性の高い文書プロフィールを見出し、そして、その後、適切 性に従って配列されたそれらのプロフィールで選択された文書を選択し、リターンし、デ ィスプレイすること、および/または(c)情報検索技法、例えば、文書ワードベクトル の従来のコサイン類似性測定の計算を使用して、ユーザ指定文書がキャッシュに記憶され た(812で)活性化拡散計算によるプロフィールに既に存在しないときは、URL/TAGバ イグラフに最も類似度の高い文書を見出し、そしてその後に、(a)または(b)のいず れかの手順を使用して、関連する文書を選択し、リターンし、ディスプレイすることを実 行し得る。別法として、814,816,818,および820で、ユーザがあるタグと 関連した情報を見出したいと思えば、文書に対して実行された同様な操作を、タグと対応 するタグプロフィール対して代わりに実行する。

[0042]

別の実施の形態では、キーワード探索を使用し、キャッシュされた(812で)活性化拡散計算によるプロフィールに文書またはタグを識別し得る。これらのプロフィールを、後で使用し、関連した文書とタグを識別するのである。さらに他の実施の形態では、複数のキーワードが探索基準として提供されるとき、複数のキーワードに対応するタグと文書に関連したプロフィールベクトルが、一緒に合計され、ランク付けが行われ、その後、興味ある文書とタグを識別するために使用される。別法として、相異なるキーワードに対応するプロフィールベクトルが相異なるように重み付けられ、その後で一緒に合計され、ランク付けが行われ、その後、興味ある文書とタグを識別するために使用し、結果を一層改良し得るが、それは、追加的キーワードおよび/または文書をプロフィールベクトルのセットに加えることによって行われ、当該ベクトルのセットは、最終的に一緒に合計され、ランク付けが行われ、その後、興味ある文書とタグを識別するために使用されるのである。

[0043]

別のさらなる実施の形態では、本明細書に示された例示的実施の形態は、文書とタグとの間という二つの変数関係を越えて拡張し得る。そのような別法の実施の形態では、グラフは、例えば、二つの相異なる変数の間(例えば、文書とユーザの間)、または三つ以上の変数の間(例えば、文書、タグ、およびユーザの間)にて定義し得る。本技法は、三つ

以上の変数に対しても有効である。なぜなら、活性化拡散技法を、これらの相異なる変数に対して実行し、キャッシュメモリに記憶し得るからである(例えば、タグプロフィール、文書プロフィール、およびユーザプロフィールが得られる)。

[0044]

さらに、前述の実施の形態においては、ブックマークコレクションの共有について議論が行われたが、当業者ならば、別法として、ブックマークコレクション共有の代わりに、個々のブックマークの好みを記録するメタデータを備える文書のコレクションを各文書と共有することについても議論が成り立つことが分かるであろう。

[0045]

F . 雑

上記の説明を考慮すれば、ウェブページに注釈を付するための実施の形態は、以下の諸ステップ、すなわち、複数の文書識別子と夕グのそれらの関連セットとをURL/TAGペアのセットに変換するステップと、前記URL/TAGペアセットを使用して二次元グラフを構成するステップと、前記二次元グラフを使用し、前記URL/TAGペアセットの文書識別子と夕グに対する活性化拡散を用いタグプロフィールと文書プロフィールとを計算するステップと、選択された夕グまたは文書に対して、選択された夕グまたは文書に関連したタグプロフィールまたは文書プロフィールをそれぞれ識別するステップと、選択された夕グまたは文書に関連した、前記識別された夕グプロフィールまたは文書プロフィールの夕グまたは文書の切性に従って配列するステップとを備える方法によって実施し得る。

[0046]

さらに、上記の説明を考慮すれば、前述の実施の形態は、以下の機能を備える。すなわち、タグプロフィールと文書プロフィールとは、活性化拡散を使用して計算し得る。。文書プロフィールは、同様なプロフィールを識別するために他の文書プロフィールに対してフィールの文書は、フロフィールの文書は、フロフィールの文書は、フロフィールの文書は、フロフィールの文書は、フロフィールの文書は、フロフィールの文書は、フロフィールの文書は、フロフィールの文書は、フロフィールは、フロフィールは、フロフィールは、フロフィールは、フロフィールは、フロフィールは、同様な文書をそれに代替することによって、実行し得る。タグプロフィールは、同様な文書を表別するに他のタグプロフィールに対して、重みによって、でで表記の文書を関連したタグプロフィールまたは文書と関連したタグプロフィールまたは文書ので表記の文書を識別子は、ブックマークコレクションとし得るが、ブックマークは、タグのセットと関連付けられた識別子で文書を識別する。前記二次でグラフは、複数のユーザで規定され得る3番目の次元をさらに含み得る。

[0047]

前述の仕様明細を使用して、本明細書に開示の諸実施の形態は、マシン(またはシステム)、プロセス(または方法)、または製造品として実装され、標準プログラミングおよび/またはエンジニアリング技法を使用することによって、プログラミングソフトウェア、ファームウェア、ハードウェア、またはこれらの組み合わせを製造し得る。当業者ならば分かることであるが、前記仕様明細に記載された諸フローチャートは相異なる可能な実施の形態の理解に便ならしめる意図で提供されたものである。従って、諸ステップの順序を変えること、一つまたは複数個のステップを並行的に実施すること、および/またはステップ数を追加したり、省略したりして実施することは、別の実施の形態で行い得る。

[0048]

得られたプログラム(複数を含む)は、コンピュータ可読プログラムコードを有し、一種または複数種のコンピュータ使用可能の媒体、例えば、メモリ装置または伝送装置内に内蔵することによって、開示の実施の形態に基づいてコンピュータプログラム製品または製造品を作製し得る。従って、本明細書に使用される「製造品」および「コンピュータプログラム製品」は、どのようなコンピュータ使用可能の媒体、例えば、どのようなメモリ装置または伝送装置上に実装されるコンピュータプログラム存在品(恒久的、一時的、または過渡的)を網羅するものとする。

10

20

30

40

[0049]

開示の実施の形態を具現化するマシンは、CPU、メモリノ記憶装置、通信リンク、通 信/伝送装置、サーバ、I/O装置を含む一個または複数個の処理システムを備え得るが 、これらに限定されない。または、ソフトウェア、ファームウェア、ハードウェア、また はこれらのコンビネーションまたはサブコンビネーションを含む一個または複数個の処理 システムのサブコンポーネントまたは個々のパーツであって、特許請求の範囲に示された 開示の実施の形態を具現化するものも含む。当業者なら理解することであるが、前記メモ リ装置は、固定(ハード)ディスクドライブ、フロッピー(登録商標)ディスク(または ディスケット)、光ディスク、磁気テープ、RAM、ROM、Promのような半導体メモリを含む が、これらに限定されない。伝送装置は、インターネット、イントラネット、電子掲示板 およびメッセージ/ノート交換、電話/モデムベースのネットワーク通信、通常有線/ケ ーブル通信ネットワーク、携帯電話通信、電波通信、衛星通信、および他の固定または移 動ネットワークシステム/通信リンクを含むが、これらに限定されない。

【図面の簡単な説明】

[0050]

【図1】例示の操作環境における注釈サーバと注釈クライアント(複数を含む)の実施の 形態の説明図である。

【図2】図1に示される実施の形態のような本明細書に開示の実施の形態を実行するため の汎用コンピュータの説明図である。

【図3】図1に示される例示の操作環境における注釈サーバと注釈クライアント(複数を 含む)とによって実行される諸操作を示すフローチャートである。

【図4】サブ文書エレメントのワードが、外接ボックス長方形を使用して個々に選択可能 とされたテキストコンテンツの例を示す図である。

【図5】ユーザのネームラベルが後で挿入されたサブ文書エレメントを示す図である。

【図6】ビューアに表示されたウェブページにユーザがタグ付けするとき一連の画像フレ -ムを示す図である。

【図7】ウェブ文書に付加する注釈を作成し、共有するためのユーザインターフェイスの 例を示す図である。

【図8】タグベースの探索サービスで実行される操作を示すフローチャートである。

【図9】図8に概略示されたように処理されたブックマークコレクションの例を示す図で ある。

【図10】4ステップに等しい「n」ステップで行われる活性化拡散の例を示す図である

【図11】ウェブブラウザを通して利用可能とし得るタグベースの探索インタフェースを 示す図である。

【符号の説明】

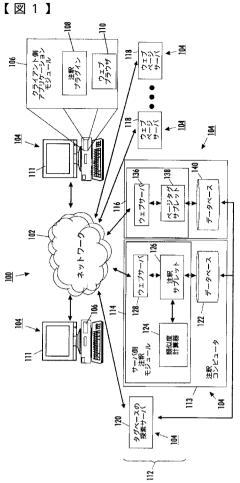
[0051]

100 操作環境、102 ネットワーク、104 汎用コンピュータ、106 クラ イアント側アプリケーションモジュール、108 注釈プラグイン、110 ウェブブラ ウザ、112 サーバ側アプリケーションモジュール、114 注釈モジュール、116 ページ - タグ付加モジュール、118 ウェブページサーバ、120 タグベースの探 索サーバ、212 ハードウェア、214 ソフトウェア、216 プロセッサ、218 メモリ、220 固定記憶装置、222 ユーザ I/O、224 ネットワーク I/O 204 カメラ、208 マイクロフォン、206 スピーカ、228 ポインティン グ装置、230 ディスプレイ、226 キーボード。

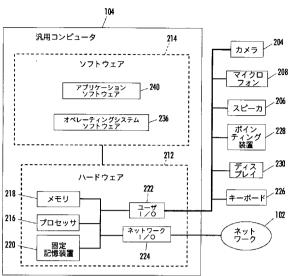
10

20

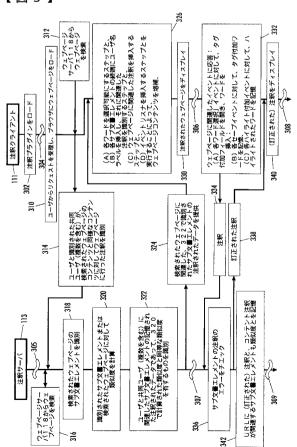
30



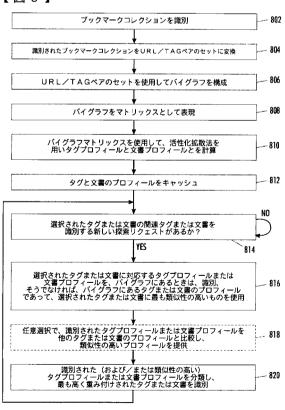




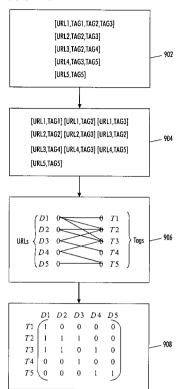
【図3】



【図8】







【図10】

$$\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} A[1]_{\text{ox}} \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\mathbf{n} = \mathbf{1}$$

502

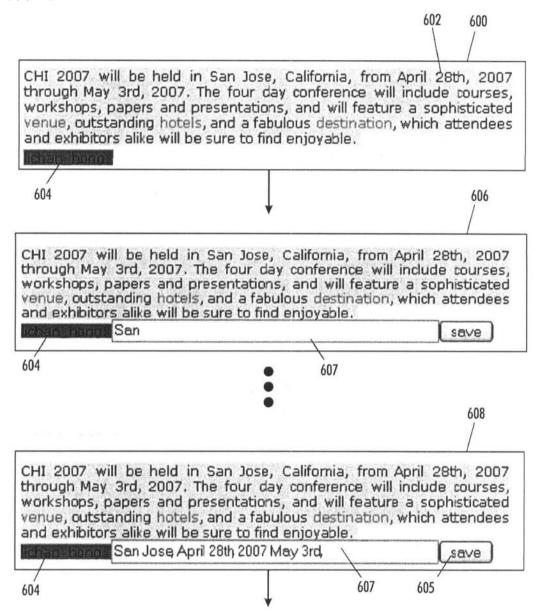
【図4】

CHI 2007 will be held in San Jose California from April 28th 2007 through May 3rd 2007 The four day conference will include courses workshops papers and presentations and will feature a sophisticated venue cutstanding hotels and a fabulous destination which attendees and exhibitors alike will be sure to find enjoyable.

【図5】

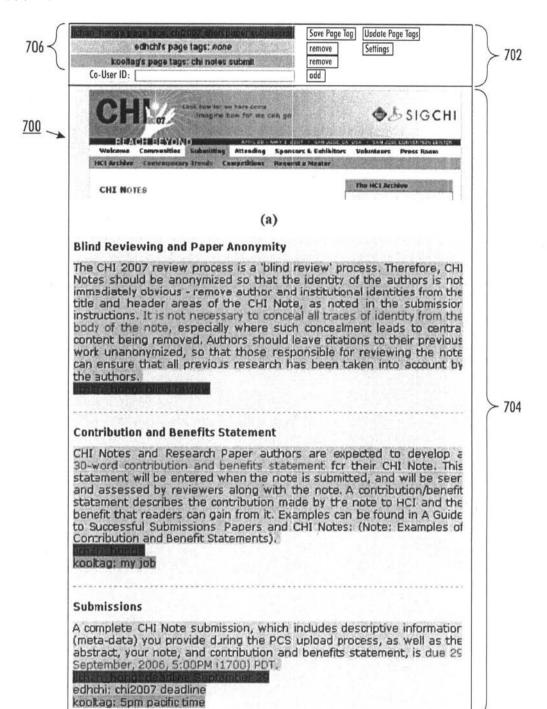
504

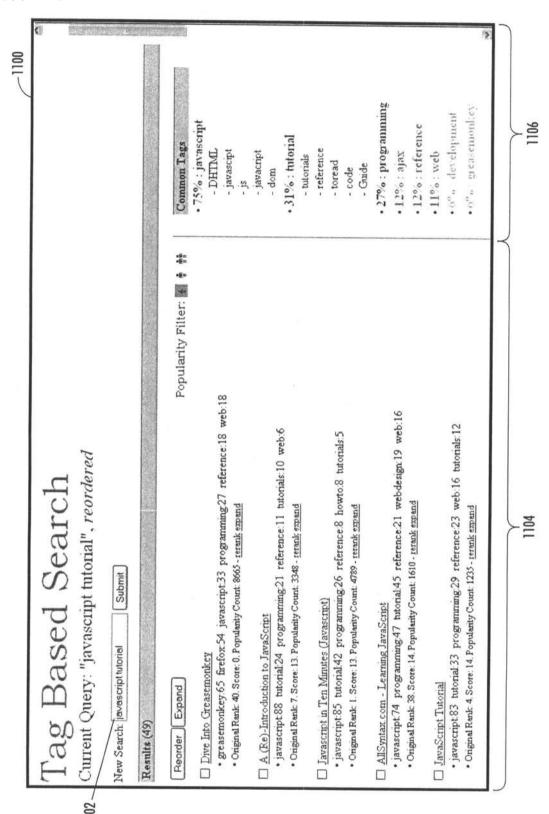
CHI 2007 will be held in San Jose, California, from April 28th, 2007 through May 3rd, 2007. The four day conference will include courses, workshops, papers and presentations, and will feature a sophisticated venue, outstanding hotels, and a fabulous destination, which attendees and exhibitors alike will be sure to find enjoyable.



CHI 2007 will be held in San Jose, California, from April 28th, 2007 through May 3rd, 2007. The four day conference will include courses, workshops, papers and presentations, and will feature a sophisticated venue, outstanding hotels, and a fabulous destination, which attendees and exhibitors alike will be sure to find enjoyable.

【図7】





フロントページの続き

(72)発明者 リチャン ホン

アメリカ合衆国 カリフォルニア マウンテン ビュー モンロー ドライブ 216

(72)発明者 エド エイチ チー

アメリカ合衆国 カリフォルニア パロ アルト グリーンメドー ウェイ 143

(72)発明者 ピーター ライ

アメリカ合衆国 マサチューセッツ ケンブリッジ メモリアル ドライブ 362

審査官 打出 義尚

(56)参考文献 特開2002-358330(JP,A)

Sheung-On Choy, Web Information Retrieval in Collaborative Tagging Systems, 2006 IEEE/WIC/ACM International Conference on Web Intelligence, 2 0 0 6 年 1 2 月, pp.352-355, URL, http://dx.doi.org/10.1109/WI.2006.191

Harris Wu , Harvesting social knowledge from folksonomies , Proceedings of the seventeen th conference on Hypertext and hypermedia , ACM , 2 0 0 6 年 , Pages 111-114 , U R L , htt p://dx.doi.org/10.1145/1149941.1149962

丹羽智史, Folksonomyマイニングに基づくWebページ推薦システム,情報処理学会論文誌,日本,社団法人情報処理学会,2006年 5月15日,第47巻,第5号,pp.1382-1392

(58)調査した分野(Int.CI., DB名)

G06F 17/30