

(19) 日本国特許庁(JP)

(12) 公表特許公報(A)

(11) 特許出願公表番号

特表2009-510598
(P2009-510598A)

(43) 公表日 平成21年3月12日(2009.3.12)

(51) Int.Cl.	F I	テーマコード (参考)
G06F 17/30 (2006.01)	G06F 17/30 340A	5B075
G06Q 50/00 (2006.01)	G06F 17/60 132	
G06Q 10/00 (2006.01)	G06F 17/60 514	

審査請求有 予備審査請求有 (全 53 頁)

(21) 出願番号 特願2008-533302 (P2008-533302)
 (86) (22) 出願日 平成18年9月26日 (2006. 9. 26)
 (85) 翻訳文提出日 平成20年5月27日 (2008. 5. 27)
 (86) 国際出願番号 PCT/SG2006/000280
 (87) 国際公開番号 W02007/037764
 (87) 国際公開日 平成19年4月5日 (2007. 4. 5)
 (31) 優先権主張番号 PCT/SG2005/000320
 (32) 優先日 平成17年9月27日 (2005. 9. 27)
 (33) 優先権主張国 シンガポール (SG)
 (31) 優先権主張番号 PCT/SG2005/000321
 (32) 優先日 平成17年9月28日 (2005. 9. 28)
 (33) 優先権主張国 シンガポール (SG)
 (31) 優先権主張番号 11/440, 896
 (32) 優先日 平成18年5月24日 (2006. 5. 24)
 (33) 優先権主張国 米国 (US)

(71) 出願人 508094628
 サーカー ビーティーイー リミテッド
 シンガポール, シンガポール 0486
 95, ナンバー05-01 グランド
 ビルディング, フィリップ ストリート
 17
 (74) 代理人 100094318
 弁理士 山田 行一
 (74) 代理人 100123995
 弁理士 野田 雅一
 (74) 代理人 100107456
 弁理士 池田 成人

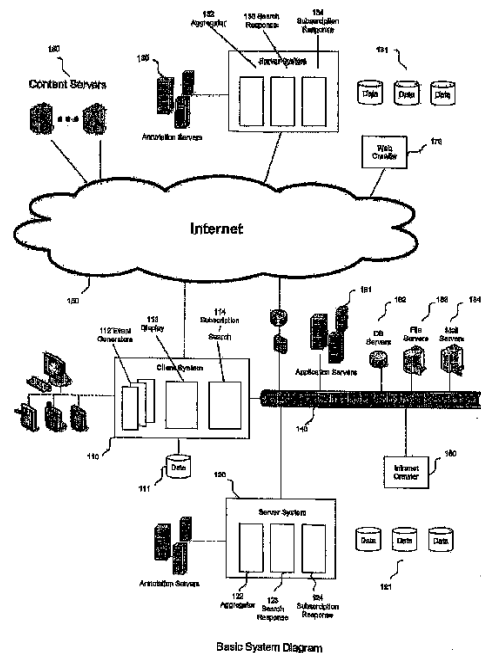
最終頁に続く

(54) 【発明の名称】 コミュニケーション及びコラボレーションのためのシステム

(57) 【要約】

アイテムがユーザ間で共有されることができ、アイテム及びユーザの両方が既存の情報検索ランク付け技術に基づいて検索及びランク付けされることができるようになる。汎用的なアノテーションベースのメカニズムを使用するコミュニケーション及びコラボレーションのためのシステム及び方法。新しい方法が、カテゴリコンテキストに基づいてユーザ及びアイテムを同時にクラスタ化するために導入される。これらのメカニズムは、コンテキストに基づいてアイテムを発行及びサブスクライブすることを可能にするメカニズムをもたらすために利用される。

【選択図】 図3



【特許請求の範囲】**【請求項 1】**

一意な識別子を有する複数のユーザ間で共有することができる、一意な識別子を有する複数のアイテムを特定するステップと、

各ユーザにその他のユーザと無関係に少なくとも1つの自然言語の少なくとも1つのキーワードを用いて複数の前記アイテムをアノテーションさせるステップであって、それぞれの前記アイテムが少なくとも1人のユーザによってアノテーションされ、それぞれの前記アノテーションが、アノテーションする前記ユーザの前記識別子と、アノテーションされている前記アイテムの前記識別子と、アノテーションする前記ユーザがアノテーションされている前記アイテムを記述するために選択する少なくとも1つのキーワードとを含むアノテーションイベントによって示され、それぞれの前記アノテーションイベントが少なくとも1つの種類の複数のイベントソースから生成される、ステップと、

特定のアイテムに関連する前記キーワードが該アイテムのための前記アノテーションイベントから収集され、かつ、特定のユーザに関連する該キーワードが該ユーザのための前記アノテーションイベントから収集されるように前記イベントソースから前記アノテーションイベントを収集するステップと、

少なくとも1人の前記ユーザにアイテム又はユーザをキーワードによって検索させるステップであって、該検索に使用したキーワードを前記収集されたキーワードの中に有するアイテム又はユーザが結果として返されるように検索させるステップとを含む、コラボレーションのための方法。

【請求項 2】

前記イベントソースの種類が、記憶ダイアログ、アイテムを発行すること、アイテムをタグ付けすること、検索のクリックストリーム、アイテム内の単語を強調し、前記単語を前記アイテムに関するキーワードとして働かせること、ファイルをファイルシステムに保存すること、リンクテキスト解析、オペレーティングシステム、及びソフトウェアモジュールからなる群のうちの少なくとも1つのである請求項 1 に記載の方法。

【請求項 3】

前記アイテムが、デジタルアセット、物理的資産、人、生物、テキスト広告、映像広告、イベント、場所、状態、プロセス、行為、グループ、概念、ファイル、電子メール、インスタントメッセージ、ブログのポスト、ポッドキャスト、ウェブページ、ウェブサイト、ウェブサービス、データ構造、ソフトウェアモジュール、ソフトウェアオブジェクト、アプリケーション、オペレーティングシステム、リレーショナルデータベース内のテーブルの行、XML データ、及び R D F で示されたリソースからなる群のうちの任意の 1 つである請求項 1 に記載の方法。

【請求項 4】

前記一意な識別子が、ハッシュ値、URL、URI、URN、UNC、バーコード、RFID、基準マーカ、電子メールアドレス、社会保障番号、車両登録番号、及び電話番号からなる群のうちのいずれか 1 つである請求項 1 に記載の方法。

【請求項 5】

前記アイテムの各々が最大で 1 つの一意な識別子を有する請求項 1 に記載の方法。

【請求項 6】

前記ユーザの各々が最大で 1 つの一意な識別子を有する請求項 1 に記載の方法。

【請求項 7】

前記識別子がグローバルに一意である請求項 1 に記載の方法。

【請求項 8】

前記ユーザにユーザ ID 及びパスワードを用いて認証させることによって一意なユーザ識別子を割り当てるステップをさらに含む、請求項 1 に記載の方法。

【請求項 9】

前記アイテムが、少なくとも 1 つのタイトルフィールド及び / 又は説明フィールドによってさらに記述される請求項 1 に記載の方法。

10

20

30

40

50

- 【請求項 10】
前記アイテム識別子がユーザ識別子である請求項 1 に記載の方法。
- 【請求項 11】
前記少なくとも 1 つのキーワードがアイテム識別子である請求項 1 に記載の方法。
- 【請求項 12】
前記少なくとも 1 つのキーワードがユーザ識別子である請求項 1 に記載の方法。
- 【請求項 13】
前記アノテーションイベントが前記ユーザによってデジタル署名される請求項 1 に記載の方法。
- 【請求項 14】 10
前記アノテーションイベントが収集のためにネットワークを介して少なくとも 1 つのサーバに送信される請求項 1 に記載の方法。
- 【請求項 15】
前記ネットワークがインターネットである請求項 14 に記載の方法。
- 【請求項 16】
ユーザ毎のアノテーションイベントの前記収集がクライアントにおいて行われる請求項 1 に記載の方法。
- 【請求項 17】
前記アノテーションイベントがアイテム毎に収集される前にユーザ毎の前記アノテーションイベントが正規化される請求項 1 に記載の方法。 20
- 【請求項 18】
情報検索ランク付けアルゴリズムを使用してクエリに対する関連性に基づいて前記検索結果をランク付けするステップをさらに含む請求項 1 に記載の方法。
- 【請求項 19】
複数の前記アルゴリズムに基づいて前記検索結果をランク付けさせ、次に前記ランクを収集して関連性を判定する請求項 18 に記載の方法。
- 【請求項 20】
前記クエリに関するキーワードベクトルを構築し、収集されたキーワード及び各結果に関するそれらのキーワードの発生の頻度のベクトルを構築するステップと、
TF-IDF、TF-IDF の変形形態、OKAPI、及びピボット正規化からなる群からのランク付けアルゴリズムを使用して前記ベクトルに基づいてランクを計算するステップと
をさらに含む請求項 18 に記載の方法。 30
- 【請求項 21】
前記アイテムの前記収集されたキーワードが前記アイテムの定義キーワードに制限される請求項 20 に記載の方法。
- 【請求項 22】
アイテムの結果集合を計算するステップと、
これらのアイテムをアノテーションしたすべての前記ユーザを前記結果集合に含めるステップと、
リンク解析ランク付けアルゴリズムのためにユーザをハブとして扱い、アイテムをオーソリティとして扱うステップと、
それぞれの前記ユーザから、前記ユーザによってアノテーションされたすべてのアイテムへの統合的なハイパーリンクを生成するステップと、
少なくとも 1 つのリンク解析ランク付けアルゴリズムを使用して、アイテムに関するランクを計算し、かつ、少なくとも 1 つのリンク解析ランク付けアルゴリズムを使用して、前記結果セットのユーザに関するランクを計算するステップと
をさらに含む請求項 18 に記載の方法。 40
- 【請求項 23】
前記リンク解析ランク付けアルゴリズムが、Indegree、HITS、ランダム化 50

HITS、サブスペースHITS、SALSA、HUBAVG、オーソリティ閾値システムのアルゴリズム、MAX、BFS、BAYESIAN、単純化BAYESIAN、Page Rank、パーソナライズされたPage Rank、Traffic Rank、TOPHITS、CubeSVD、PHITS、及びPLSA+PHITSからなる群のうちの1つである請求項22に記載の方法。

【請求項24】

前記結果集合内のユーザ及びアイテムの双方に関して収集された前記キーワードが、前記検索の前記キーワードを含む請求項22に記載の方法。

【請求項25】

前記結果集合を計算するために使用される各アノテーションイベントが、前記クエリのすべてのキーワードを含む請求項22に記載の方法。

10

【請求項26】

前記結果集合を計算するために使用される各アノテーションイベントが、前記クエリのすべてのキーワードを含み、少なくとも1つの前記キーワードが前記アノテーションイベントの前記アイテムに関する定義キーワードである請求項22に記載の方法。

【請求項27】

前記少なくとも1つのキーワードが、意味の機械表現である請求項1に記載の方法。

【請求項28】

前記少なくとも1つのキーワードが、意味メタデータである請求項1に記載の方法。

【請求項29】

少なくとも1つの辞書によって前記意味メタデータを記述し、前記辞書に基づいて前記クエリの前記キーワードと、アイテム及びユーザの前記収集されたキーワードとの曖昧性を除去するステップをさらに含む請求項28に記載の方法。

20

【請求項30】

アイテムの定義キーワード又はカテゴリーコンテキストのキーワードを集合として扱い、すべての前記集合におけるキーワードの共起に基づいてオントロジーを生成するステップをさらに含む請求項1に記載の方法。

【請求項31】

辞書内の概念の間の意味関係が前記共起データから生成される請求項30に記載の方法

30

【請求項32】

コンテキストによってアイテム及びユーザの両方を同時にクラスタ化するステップをさらに含む請求項1に記載の方法。

【請求項33】

コンテキストによってアノテーションイベントを収集するステップと、既定の最小数の一意なユーザ識別子及び既定の最小数の一意なアイテム識別子をコンテキストの収集されたアノテーションイベント内に有するコンテキストを判定するステップと、

前記コンテキストに基づいてアイテム及びユーザをクラスタ化するステップとをさらに含む請求項32に記載の方法。

40

【請求項34】

前記アノテーションイベントを所定の期間に制限するステップをさらに含む請求項33に記載の方法。

【請求項35】

前記クエリのサブコンテキストであるすべての前記コンテキストを判定するステップと、

前記判定されたコンテキスト内に存在するすべてのキーワードが前記検索クエリの一部である前記キーワード以外に存在する場合に、前記判定されたコンテキストから一意なキーワードの集合を計算するステップと、

それぞれの前記一意なキーワードを、前記キーワードを元のクエリに追加し、前記クエ

50

りを再発行することによってユーザがドリルダウンすることができる下位カテゴリとして提示するステップと

をさらに含む請求項 3 3 に記載の方法。

【請求項 3 6】

ユーザがキーワードの集合を用いてアイテムをアノテーションするときにアイテムを発行するステップをさらに含む請求項 1 に記載の方法。

【請求項 3 7】

前記ユーザに対応する一意な発行者識別子を用いて前記発行されたアイテムをアノテーションするステップをさらに含む請求項 3 6 に記載の方法。

【請求項 3 8】

前記発行するユーザにユーザ ID 及びパスワードを用いて認証させることによって一意な発行者識別子を割り当てるステップをさらに含む請求項 3 7 に記載の方法。

【請求項 3 9】

前記アノテーションイベントが発行アノテーションイベントである請求項 3 6 に記載の方法。

【請求項 4 0】

前記発行アノテーションイベントが一意なアイテム識別子を有するアイテムである請求項 3 9 に記載の方法。

【請求項 4 1】

前記アノテーションイベントのキーワードの前記集合がカテゴリコンテキストである請求項 3 6 に記載の方法。

【請求項 4 2】

前記発行されたアイテムが、所定の条件が満たされるまで前記コンテキストに関する前記ユーザのサブセットに対して前記発行されたコンテキストにおいて高くランク付けされる請求項 3 6 に記載の方法。

【請求項 4 3】

検索結果が前記発行者識別子に基づいてランク付けされる請求項 3 7 に記載の方法。

【請求項 4 4】

アイテムを発行する行為がワークフロープロセス内のステップの完了に対応する請求項 3 7 に記載の方法。

【請求項 4 5】

前記ワークフロープロセス内の前記ステップが、認可、要求、割り当て、終了、及び通知を含む群のうちの 1 つである請求項 4 4 に記載の方法。

【請求項 4 6】

ユーザプロファイルに基づいて自動的に周期的に検索するステップと、
情報検索ランク付けアルゴリズムを使用して、クエリに対する関連性に基づいて検索結果をランク付けするステップと、

ユーザ毎に上位にランク付けされた結果のサブセットを取り出し、記憶するステップと

、
前記結果を前記ユーザの要求で前記ユーザに提示するステップと
をさらに含む請求項 3 7 に記載の方法。

【請求項 4 7】

発行者識別子によって各コンテキストに関する前記記憶された結果を再ランク付けするステップをさらに含む請求項 4 6 に記載の方法。

【請求項 4 8】

前記発行されたアイテムが、所定の条件が満たされるまで前記コンテキストに関する前記ユーザのサブセットに対して前記発行されたコンテキストにおいて高くランク付けされる請求項 4 6 に記載の方法。

【請求項 4 9】

ユーザの前記サブセットが、高くランク付けされた発行者、高くランク付けされたユー

10

20

30

40

50

ザ、及びユーザの無作為の選択からなる群のうちの少なくとも1つのユーザから選択される請求項48に記載の方法。

【請求項50】

提示されたアイテムに関するアノテーションイベントのときに存在する発行者識別子が前記アノテーションイベントに含まれるように、ユーザが前記アイテムをアノテーションするときに前記アノテーションイベントを生成するステップをさらに含む請求項46に記載の方法。

【請求項51】

前記発行者が、前記ユーザ及び/又は元の発行者によって最も認められた発行者に制限される請求項50に記載の方法。

【請求項52】

ユーザプロフィールに基づいて自動的に周期的にコンテキストの集合を検索するステップと、

情報検索ランク付けアルゴリズムを使用してクエリに対する関連性に基づいて検索結果をランク付けするステップと、

各コンテキストに関して上位にランク付けされた結果のサブセットを取り出し、記憶するステップと、

前記結果を前記ユーザの要求で前記ユーザに提示するステップと

をさらに含む請求項1に記載の方法。

【請求項53】

所定の期間内に前記コンテキストに追加されたイベントに前記検索を制限するステップをさらに含む請求項52に記載の方法。

【請求項54】

前記コンテキストがカテゴリコンテキストに制限される請求項52に記載の方法。

【請求項55】

前記コンテキストが前記ユーザによって明示的に指定される請求項52に記載の方法。

【請求項56】

キーワード及びユーザ識別子からなる群からの少なくとも1つのパラメータに基づいてTFIDFの時間に基づく変形によって各コンテキストに関する前記記憶された結果を再ランク付けするステップをさらに含む請求項52に記載の方法。

【請求項57】

少なくとも1つの広告が、関連性のあるユーザを検索し、前記広告を前記ユーザに対して表示することによって少なくとも1人のユーザにプッシュされる請求項1に記載の方法。

【請求項58】

前記広告がユーザプロフィールに基づいて再ランク付けされる請求項57に記載の方法。

【請求項59】

前記広告が、テキスト、オーディオ及びビデオ、並びに部門別案内広告の群からの少なくとも1つからなる請求項57に記載の方法。

【請求項60】

前記検索が、コンテキストに基づいていくつかのアノテーション収集サーバに渡って連携される請求項1に記載の方法。

【請求項61】

アイテムが所定のキーワードの集合のうちの少なくとも1つを用いてアノテーションされる場合に前記結果から前記アイテムの集合を削除するステップをさらに含む請求項1に記載の方法。

【請求項62】

前記キーワードのアノテーションの頻度が所定のレベルを超える場合にのみアイテムが削除される請求項61に記載の方法。

10

20

30

40

50

【請求項 6 3】

前記検索が検索エンジンにおいて実行され、アノテーションイベントが前記検索エンジンの検索ログのクリックストリームから生成される請求項 1 に記載の方法。

【請求項 6 4】

一意な識別子を有する複数のユーザ間で共有することができる、一意な識別子を有する複数のアイテムを特定する手段と、

各ユーザにその他のユーザと無関係に少なくとも 1 つの自然言語の少なくとも 1 つのキーワードを用いて複数の前記アイテムをアノテーションさせる手段であって、それぞれの前記アイテムが少なくとも 1 人のユーザによってアノテーションされ、それぞれの前記アノテーションが、アノテーションするユーザの前記識別子と、アノテーションされている前記アイテムの前記識別子と、アノテーションする前記ユーザがアノテーションされている前記アイテムを記述するために選択する少なくとも 1 つのキーワードとを含むアノテーションイベントによって示され、それぞれの前記アノテーションイベントが少なくとも 1 つの種類の複数のイベントソースから生成される、手段と、

特定のアイテムに関連する前記キーワードが該アイテムのための前記アノテーションイベントから収集され、かつ、特定のユーザに関連する前記キーワードが該ユーザのための前記アノテーションイベントから収集されるように、前記イベントソースから前記アノテーションイベントを収集する手段と、

少なくとも 1 人の前記ユーザにアイテム又はユーザをキーワードによって検索させる手段であって、該検索に使用したキーワードを前記収集されたキーワードの中に有するアイテム又はユーザが結果として返されるように検索させる手段とを備える、コラボレーションのためのシステム。

【請求項 6 5】

前記コンテキストを、すべてのアイテムが前記アイテムをアノテーションする既定の数のユーザを有する既定の数の前記アイテムが存在するコンテキストに制限するステップをさらに含む請求項 3 3 に記載の方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、アイテム及びユーザの両方が既存の情報検索ランク付け技術で検索及びランク付けされることを可能にするように検索の概念を拡張するコミュニケーション及びコラボレーションのための方法に関する。本発明は、コンテキストベースのコミュニケーションメカニズムを提供するために利用される。

【発明の背景】

【0002】

[背景]

ワールドワイドウェブの登場とコンピュータのほぼユビキタスな存在とは、人々が情報を発見し使用する様式を劇的に変化させた。しかし、我々が 21 世紀を迎えたとき、我々は他に例を見ないジレンマに直面している。情報及び知識がこれまでになく重要になり、ますます多く生み出されているが、有用で適切な情報を発見することはますます難しくなりつつある。

【0003】

ウェブ検索技術は、当技術分野における重要な飛躍的進歩を示す。初期の検索エンジンは、ウェブ上のページのグラフを行き来し、それらを集中サーバにダウンロードするソフトウェアプログラムであるウェブクローラ又はスパイダを作成した。次に、これらは転置インデックスに変換され、情報検索 (IR) 方法に基づいて検索された。人は、単語又は単語の集合を含んだすべてのドキュメントを検索することができた。ページの小規模な集合に対して有用であったが、これはスケーリングできなかつた。検索の有用性は、クエリに関連しない多数の返却ドキュメントによって損なわれた。したがって、関連性のあるドキュメントを発見するためにすべての結果を調べるのが現実的ではないとき、重要なド

10

20

30

40

50

キュメントが発見されることはできなかった。

【0004】

この「多量さ」の問題は、Kleinberg、Page、及びBrinの初期の重要な論文において対処された。Kleinberg、Page、及びBrinは、ドキュメント間のハイパーリンクを関連性の判断の代用として使用した。多くのその他のページによってリンクされたページは通常よりも関連性があるであろうことが既に知られていた。Page及びBrinは、あるページをいくつかのページが指すかだけでなく、これらのページがどのような質であるかも重要であるという洞察によってこの概念を改善した。Page及びBrinは、Google検索エンジンにおいて使用されているPageRank法を提案した。これは、そのページにリンクするページに基づき、クエリに依存しないページのランク付けである。

10

【0005】

Kleinbergは、ウェブページの質のより改善された概念を提案した。Kleinbergは、質の高いページがその他の質の高いページ(Kleinbergはオーソリティと呼んだ)を指すことは必要でないと主張した。その代わりに、質の高いオーソリティへのリンクの集合を含むハブと呼ばれる特別なノードが存在する。HITSアルゴリズムは、質の良いハブは多くの質の良いオーソリティにリンクするハブであり、質の良いオーソリティは多くの質の良いハブによってリンクされるオーソリティであるとの洞察によってハブ及びオーソリティの両方に基づいて質を評価した。これはクエリに依存する態様で計算された。

20

【0006】

これらの方法は、ネットワーク中の集合的な知性を取り入れて、適切な及び有用なページを検索結果の上位にすることを助けることを可能にした。そのようなリンク解析及びランク付け(LAR)アルゴリズムの成功は、Googleの成功にはっきりと見られる。この成功は当該領域の多数の研究をもたらし、Hilltop、SALSA、ランダム化HITS、サブスペースHITSなどのような上記アルゴリズムの多くの変形が存在する。より最近、結果の質を向上するために3変数のテンソル分解を利用するTOPHITSのような3レベルアルゴリズムが提案された。HITSは、ハブ及びオーソリティを計算するために使用された初期ページがクエリのトピックにまったく関連しなかった可能性がある場合にトピックドリフトを被った。TOPHITSはリンクテキスト(ハイパーリンク内のテキスト)を使用して、ハブ及びオーソリティのクエリのトピックに対する関連性を割り当てることによってHITSを改善する。

30

【0007】

しかし、ウェブ検索技術は進歩したものの、まだ多くの重大な問題が残っている。通常のクエリは数百万の結果を返す。所望のページが上位の10件又は20件の結果の中に発見されない場合、検索は無駄である。この問題は、たとえユーザが多様な情報ニーズを持っている可能性があったとしてもすべてのユーザが同じクエリに対して同じ結果を得るという事実によっていっそう大きくされる。その一方で、ウェブに対する通常のクエリは2、3語である。このことは、検索エンジンがすべてのユーザの情報ニーズを満足することを非常に難しくする。この問題を軽減するために主要な検索エンジンが取っている1つのアプローチはパーソナライズされた検索である。CubeSVDは、ユーザのクリックストリーム(ユーザがクリックするクエリ結果)を使用してユーザに関する関連性を判定することによるパーソナライズされた検索に対する最近のテンソル分解アプローチである。しかし、パーソナライズが実装される態様によっては重大なプライバシーに関する懸念が残り、これらのアプローチが効果的であるかどうかはまだ分かっていない。

40

【0008】

会社のイントラネットのウェブページの検索又はファイル共有などのエンタープライズサーチ、及びデスクトップ検索における状況はさらに悪い。いくつかの研究は、ファイル、電子メールなどのすべての法人のデータの80%までもが構造化されていない(データベース又はアプリケーション内にはない)と見積もる。ドキュメント間にハイパーリンクの

50

接続性がないために、ウェブのL A Rアルゴリズムの進歩は適用されるようにされることができない。これらのシステムにおけるランク付けは、全文検索のT F I D Fスタイルのアルゴリズムにまだ限定され、より低い質をもたらす。近く登場するM i c r o s o f t W i n d o w s V i s t aオペレーティングシステムにおけるキーワードに基づくカテゴリの使用などの最近の充実したインデックス付けアプローチでさえも、以前のI Rシステムの同様の問題にまださらされている(キーワードに基づいてドキュメントを検索する人は検索回数の20%しか当該ドキュメント上のキーワードと同じキーワードを使用しない可能性が高いと推定されている)。このことは、インターネット上でドキュメントを発見することの方が、ユーザ自身のハードディスク上でそのドキュメントを発見することよりも容易であるという逆説的な状況を生じる。

10

【0009】

最近、フォークソノミと呼ばれる分担方法を通じてカテゴリーの領域で興味深い研究がなされてきた。集中的な態様でウェブを手動で分類するY a h o o !及びO D Pの初期の試みと異なり、それらは、ブックマーク、画像、及びウェブページを共有するためのコラボレーションによるタグ付けアプローチを使用するように試みる。この空間の主要な革新者はF l i c k r、d e l . i c i o . u s、及びT e c h n o r a t iを含む。タグ付けに対するそれぞれのアプローチは異なるが、それらのすべてはユーザのための共有空間を作成するための基礎としてキーワードを使用するように試みる。しかし、全般に、少数のユーザだけが実際にフォークソノミを使用する。これは、それらが検索よりも使用することが難しいという事実が原因の一部であり、フォークソノミのページの適用範囲が利用可能なページのうちのごくわずかであることも原因である。ウェブ又はディスクを自動的に巡回してすべてのドキュメントにインデックスを付けることができる検索とは異なり、フォークソノミに効率的にページを取り込むことができる同等のメカニズムは存在しない。

20

【0010】

同様の一連の問題が、概してブログ及びメッセージングシステムに存在する。ブログは、ユーザが単に情報の消費者であるだけでなく、オンラインでコンテンツをポストすることもできる「読み書き」ウェブを作成する試みである。ブログの真の期待は、誰かに発行されたブログのポストに興味を持っている可能性がある読者に効率的な態様で届くことができるときにのみ実現されることができる。必要とされているのは、人が関連性のあるブログのポスティングを検索する代わりにブログのポスティングが関連性のある人を検索する必要がある、検索とは反対のものである。これを実現する態様は現在存在しない。電子メールのようなその他のメッセージングシステムにおいては状況はさらに悪い。電子メール及びI Mは1対1のインタラクションのための効率的なメカニズムであるが、電子メール及びI Mはトピックに対して電子メールを送信する概念を効率的に扱わない。配布先リストはトピックに関してメールを送信することに対する最も近い代用であるが、動的に配布先リストを作成し、人をそれらの配布先リストに割り当てることはできない。多くの場合、ドキュメントの存在を他の人に伝える唯一の態様は電子メールである。このことはそのような配布先リストの乱用につながり、受信箱が、ユーザに対する関連性が低く、情報の過多及び損失をもたらすメールで溢れるようになる。

30

40

【0011】

組織が法人か、政府機関か、非政府機関か、軍隊か、それとも宗教団体かにかかわらず、概して組織はより大きく、より複雑になった。そのような組織が規模を増すにつれて、人が互いに知り合うことは次第に難しくなり、多くの場合組織は、ある人が他の人が何をしているのか分からない周囲が見えない部分に分割される。これは、組織が急速に変わる環境に直面するとき、及び組織の異なる部分が好機を物にするか又は脅威に立ち向かうために動的に協働するようにされなければならない場合に重大な問題である。電話又は電子メールのような2点間コミュニケーションを用いた階層及び部門などの従来の組織構造は、効果的な対応を行うための柔軟性を認めない。コンテキストベースのコミュニケーション及びコラボレーションメカニズムが重要な役割を果たすことができる。人が新しい好機

50

ノ脅威などの特定のコンテキストに基づいて集まり、その好機ノ脅威が過ぎ去るときに解散することを可能にするパラダイムは、変更により対応しやすい組織化に対するより有機的なアプローチの作成を可能にする。ウェブページの作成者及び最終ユーザが「コラボレーションする」ために互いを知る必要がないウェブ検索の基本的メタファは、組織が複雑性を管理することを可能にするようにコミュニケーション及びコラボレーションにまで拡張されることができる。

【 0 0 1 2 】

[本発明の背後にある基本的概念]

本発明は、コミュニケーション及びコラボレーションの観点で問題を定義することによって解決法を提供するように試みる。本発明は、i) 情報検索の進歩が適用されることができるといえるような汎用的なアノテーションベースのコラボレーションシステムの作成と、ii) ユーザ及びアイテムをクラスタ化するための方法と、iii) 人がコンテキストに基づいてメッセージを発行及びサブスクライブすることを可能にするコミュニケーション方法とを可能にするシステム及び方法を作成することに重点を置く。当技術分野にとって新規性のあるこれらのメカニズムには複数の特徴がある可能性がある。これらの方法はスタンドアロンで使用されることができただけでなく、互いに又はその他のシステムと関連して使用されることができる。

10

【 0 0 1 3 】

[I Rを用いた汎用的なアノテーションベースのコラボレーションシステム]

当技術分野に知られている多くの形態のアノテーションベースの分担システムが存在する。すべてのフォークソノミはそのようなシステムの例である。しかし、アノテーションシステムは、分担及びコラボレーションの観点でそれらのシステムの有効性に違いがある。常にMicrosoft WindowsのNTFSファイルシステムは任意のファイルにキーワードを追加する能力を有していたが、この機能は、主にそのようなキーワードがファイルを効率的に見つけるために使用されることができなかつたという理由で実際には使用されてこなかった。Technorati Tag (プロガが人がそれらのプロガのポストを発見できるようにそれらのプロガのポストにタグを付けることを可能にする技術)は、作者しかアイテムにタグを付けることができないので有効性が低い。Flickrは、ユーザがそれらのユーザの写真を他の人と共有することを可能にする写真共有サイトである。しかし、ユーザは、それらのユーザがいったんその他のユーザの写真を見てしまうとその写真に関する利用を制限され、したがって、Flickrの力はコラボレーションについてはより低く、共有についてはより高い。

20

30

【 0 0 1 4 】

しかし、Del.icio.usは、ブックマークに関する成功したコラボレーションシステムの例である。ある人のブックマークは場合によってはその他の人に対して独自の価値を有する可能性があるため、そのブックマークを共有する行為は例えばFlickrよりも重要な役割を負う。様々な人が同様の問題に直面する可能性があるため、その問題を解決することに役立つ関連情報を共有することはコラボレーションの側面を帯びる。ブックマークの本質的な価値は別にして、del.icio.usは、そのメカニズムの観点においてもその他のフォークソノミとは異なる。そのメカニズムは複数のユーザが同じアイテムにタグを付けることを可能にし、複数のユーザはそれらのユーザがタグを付けることから独立に有用性を導き出すので同じアイテムにやはりタグを付ける。

40

【 0 0 1 5 】

ウェブの多くの性質がべき乗則を示すことが当技術分野において知られている。べき乗則は、図2のような両対数グラフにプロットされるとき直線を形成する分布である。これは、フラクタルであることの基本指標とみなされ、スケール不変性を説明する(分布はどのようなスケールでも自己相似的に見える)。べき乗則は、ウェブページのコンテンツ、ウェブページ間のハイパーリンク、検索クエリ、ウェブサーバ上のファイルサイズ、トラフィックパターン、及びインターネットを構成する物理的配線において認められている。ページ間のハイパーリンクにおけるべき乗則は、Page Rank及びHITSのような

50

L A Rアルゴリズムが解に効果的に収束し、したがって実用的な方法になることを可能にする。ドキュメントのコンテンツにおいて使用される語は、ジップの法則としても知られるべき乗則にやはり従う。これは、T F - I D F (語の頻度 - ドキュメントの頻度の逆数)、及び関連性によって検索結果をソートするためのその他の方法などのI Rランク付け方法によって暗黙的に利用される。D e l . i c i o . u sも図1のようなべき乗則を示す。アイテム内のキーワードの頻度、キーワードに対するアイテム数、ユーザに対するアイテム数、アイテムに対するユーザ数などがすべてべき乗則の分布に従う。本発明は、d e l . i c i o . u sにおけるアイテムに対するキーワードの分布が(実際にはウェブのリンクテキストであるような)ウェブ検索エンジンにおけるクエリに対するキーワードの分布と似ていることに留意する。アイテム毎のユーザの分布は、ワールドワイドウェブ上のインリンク(ウェブページに入ってくるハイパーリンク)の分布に似ている。実際には、ウェブ上のハイパーリンクは通常はページナビゲーションの形態とみなされるが、本発明はそれらのハイパーリンクがアノテーションの形態としてもみなされ得ることに留意する。

10

20

30

40

50

【0016】

本発明はこれらのべき乗則の性質が、言語自体の性質、及び我々が我々の周りのアイテムの有用性を理解する態様の性質であると仮定する。これは、コミュニケーション及びコラボレーションの2つの相互に強化し合うプロセスの必然的結果である。したがって、これらのプロセスにおいて行われる創発的な自己組織化を好適に利用することができる任意のアノテーションシステムは、I R及びL A R研究の既存の団体によって有利に利用されることができると同様のべき乗則の特徴を示す。

【0017】

この基本的な考え方をを用いて、本発明は、アノテーションシステムの概念を少なくとも2つの重要な態様で一般化する。アノテーションは、ユーザがキーワードを用いてアイテムにタグを付ける行為であると考えられるだけでなく、ユーザに簡潔な態様でアイテムを記述するように求める任意の行為であるとも考えられる。検索におけるクリックストリームは、アノテーションのための少なくとも1つのその他の等価な方法である。また、それは、クエリのキーワードが所与のユーザに関するクリックされたU R Lに関連付けられるようにアノテーションを生成する。ウェブ上のハイパーリンク内のリンクテキストも、表現力は低いがそのようなアノテーションのための別のメカニズムである。ファイルをファイルシステムの階層内に置くことは、より柔軟性がなく、制限されてはいるがアノテーションの形態であり、当該ファイルよりも上の階層ツリー内のすべてのディレクトリ名を当該ファイルに関連付けることと同様である。リンクテキスト及びファイル名はタグ付け又はクリックストリームほど効率的でないが、十分な数があればそれらのリンクテキスト及びファイル名は本発明の形態のアノテーションの効果を近似する。

【0018】

次に、アノテーションは、グループのコミュニケーション及びコラボレーション活動内で特定されることができると任意のアイテムに適用されることができると。これは、ウェブのU R Lだけに制限されず、概念的に説明されることができるとあらゆるものを包含することができる。これは、企業L A N内のファイル、プロジェクト管理システム内のタスク及び課題、ブレインストーミングセッションにおけるアイデア、紙のドキュメント、スプレッドシート上の表、R D B内のデータ、ウェブサービス、R S Sフィードなどを含むことができるがこれらに限定されない。ユーザがアイテムを取り出すか又は使用することを可能にするための(オフライン又はオンラインのデジタル式か又はその他の形式の)メカニズムが存在すると仮定して、アイテムは(U R I、社会保障番号、又はバーコードのような)一意なI Dによって示されることができると任意のものであることができる。

【0019】

本発明の汎用的なアノテーションベースのコラボレーションシステムは、i)システムが互いに無関係な多数のユーザによって(アイテムが取り出されるか、評価されるか、見られるか、又は使用されることができると)アイテムが特定及び共有されることを可

能にし、ii) それぞれのそのようなユーザがアイテムを記述するのに有用であると感じるキーワードを用いてそれらのユーザがアイテムを独立にアノテーションすることができ、それぞれのそのようなアイテムが多数のそのようなユーザによってアノテーションされる可能性があり、iii) 各ユーザがそのようなキーワードに基づいてアイテムを独立に発見することができ、その結果、(各アイテムに関して複数のユーザに渡って収集された) 対応するキーワードを有するすべてのアイテムが取り出される、多数のアイテムを含む任意のアノテーションシステムとして定義され、ユーザ及びアイテムの十分に多様な集団に対してべき乗則の分布に近い自己組織化の性質を示す。ここで重要な概念は独立性の概念であり、この概念はユーザが互いの活動及び/又は存在を知ることなしに操作を行うことができることを示唆している。これは、ユーザ間のアイテムの組織化のための唯一のメカニズムが、様々なユーザの間のキーワードの共有された意味であることを意味する。

10

【0020】

より具体的には、アノテーションは、以下の形態のイベントを生成する任意のメカニズムによって生成されることができる。

[アイテムID][ユーザID][キーワード1, キーワード2, . . . キーワードN]

一意な識別子[ユーザID]を有するユーザが、アイテムを記述するキーワード[キーワード1, キーワード2, . . . キーワードN]を用いて一意な識別子[アイテムID]を有するアイテムを記述する度に、本発明は、そのようなアノテーションベースのシステムが

20

- ・キーワードに対するアイテム数
- ・アイテム内のキーワードの頻度
- ・アイテムに対するユーザ数
- ・ユーザに対するアイテム数

においておおよそべき乗則の分布を示す場合にそのようなメカニズムが以下の性質を示すことに留意する。

【0021】

・関連性に関して検索結果をソートすることにTFIDFスタイルのランク付けを使用することができる。各アイテムは、IRにおける通常のドキュメントと同様にキーワードベクトルによって示されることができる。さらに、ウェブ上のリンクテキストと同様に、各アノテーションはアイテムのコンテンツの異なる個人の判断を示し、収集されるときに当該アイテムの作者/作成者よりもさらに上手くコンテンツを記述する。したがって、収集されたキーワードに基づくランク付けは標準的な全文検索に比べて優れた質をもたらすことが多い。

30

【0022】

・各イベントをユーザからアイテムへの「統合的な」ハイパーリンクとして扱うことができる。ユーザをハブとして扱い、アイテムをオーソリティとして扱うことによって、HITS、Hilltop、SALSA、PHITS、ランダム化HITS、サブスペースHITS、TOPHITS、CubeSVDなどのような(Borodinらの文献に記載されている)LAR方法のようなIRアルゴリズムを結果セット及び検索結果のランク付けを決定することに取り入れることができる。そのようなアプローチによって達成される結果の質は、ウェブにおけるハイパーリンクに基づいたそれらの現在の性能と同等か、ときにはそれ以上である可能性がある。アイテムに対するユーザ及びユーザに対するアイテムの両方がべき乗則を示すので、これらのアルゴリズムはウェブのように急速に収束する。このことは、企業内のファイルのようにアノテーションされたアイテムがLARスタイルのアプローチの恩恵を受けることが現在不可能である場合に、アノテーションされたアイテムがLARスタイルのアプローチの恩恵を受けることを可能にすることができる。

40

【0023】

・ユーザがキーワードの集合を用いてアイテムをアノテーションするとき、アイテムを記述する多数の態様のうちから、通常のユーザはそれらのユーザが当該アイテムを定義す

50

るのに有用であると思うキーワードを自然に選択する。そのようにする際に、ユーザはアイテムを記述するだけでなく、同時にそれらのユーザが重要だと思ふことを記述する。このことは、ドキュメントと同様にキーワードによってユーザに対してクエリを行うことを可能にする。クエリに対するユーザのランク付けは、ユーザをハブとして扱う、L A Rアルゴリズムを含む上記のI Rアプローチのうちのいずれかによって行われることができる。

【0024】

・コミュニティは、人が場合によってはアイテム/ユーザを誰よりも効果的に検索できる可能性がある様々なキーワードをアノテーションする。しかし、キーワード-アイテム又はキーワード-ユーザ行列を単語-ドキュメント行列と同様の態様で使用することによってL S I又はP L S Aを使用して、たとえアイテム又はユーザをキーワードに明確に関連付けるアノテーションがないとしてもキーワードによってそれらのアイテム又はユーザに対してクエリを行うことを可能にすることもできる。これらは、高次共起データを利用してそのような単語を発見する。これは、イベントデータに対してP H I T S + P L S A、H O S V D、C u b e S V D、及びT O P H I T Sのような3レベルアプローチを使用することによってさらに向上されることができる。

10

【0025】

上記に基づいて、そのようなアノテーションシステムにおいてアイテム及びユーザに対してI Rに基づく多数の検索方法があり得ることが当業者には明らかであろう。特定の実施形態においてそれらが選択的に使用されるのか、それとも互いに連動して使用されるのかは、本発明の精神から逸脱しない。さらに、アイテム/ユーザに関するアノテーションキーワードベクトル、キーワード-アイテム行列、及びキーワード-ユーザ行列を構築し、アノテーションシステムがべき乗則を示すかどうかにかかわらずアノテーションを統合的なハイパーリンクとして扱うことがいつでも可能である。すべての上記の方法が任意のそのような場合において、ただし場合によっては低い有効性で適用されることができる。

20

【0026】

[アイテム及びユーザのクラスタ化]

過去に、ドキュメントをクラスタ化するための方法を生み出す試みがあった。分類法及び統制語彙が試され、1つの上部構造をすべてのアイテムが効果的にその上部構造に入れられることができるように作成することが現実的でないために失敗した。原文の類似性に基づく結果データに対するクラスタ化を使用する試みがなされたが、そのような自動化された方法によって生成された下位カテゴリーはユーザが理解することが容易でないことが多い。

30

【0027】

本発明はクラスタ化の概念を概してアイテムに、具体的には検索に拡張する。コンテキストが次第に狭まると共に検索結果にドリルダウンする能力は、通常のクエリの長さが2、3個のキーワードであるウェブ検索における問題に対する可能な解決策である。現在のユーザはファイルシステム内のフォルダにドリルダウンして、それらのユーザが探すファイルを発見することに慣れている。したがって、検索結果がキーワードによって下位カテゴリーにクラスタ化された場合(例えば、図10)、ユーザはファイルシステムと同様の閲覧の行動を示す可能性がある。これは、そのような下位カテゴリーに対応するキーワードを用いて元のクエリを増強することに等しい。これは、G o o g l e S u g g e s tなどのクエリ改善方法よりもユーザフレンドリーであり、より優れている可能性がある。そのような方法はウェブ検索に限定されず、デスクトップ検索及びエンタープライズサーチを含むがこれらに限定されない任意の形態の検索に適用されることができる。

40

【0028】

アイテムをクラスタ化することとは別に、ユーザをクラスタ化することに関する多くの潜在的な用途が存在する。これは、同じ興味を持つ人の集まりの動的な形成、又は類似性などに基づくのではなく興味に基づいたソーシャルネットワークの生成をもたらすことである可能性がある。本発明の汎用的アノテーションメカニズムは、キーワードに対してユーザをアイテムと同様に扱うことを可能にする。アイテム及びユーザをクラスタ化するた

50

めに使用されることができ、当技術分野において知られている多くのクラスタ化アルゴリズムが存在する。これらは、主成分解析及び多次元尺度構成法のような射影法、又は自己組織化マップ、K - 平均クラスタ化などのようなその他の方法を含む。アイテムは、それらのアイテムのアノテーションにおいて使用されるキーワード、又はそれらのアイテムをアノテーションするユーザ、又はその両方に基づいてクラスタ化されることができ、同様の態様で、ユーザはそれらのユーザのキーワード、又はアイテム、又はその両方に基づいてクラスタ化されることができ。

【0029】

すべてのクラスタ化方法が解決する必要がある重大な問題のうちの1つは複雑性の削減である。例として、アイテム及びユーザのキーワードに関連する高い複雑性がある。実際には、アイテムが存在するのと同じぐらい多くの一意なアノテーションのコンテキストが存在する可能性がある。したがって、クラスタ化の問題は、類似したアイテム及びユーザをまとめる目的に役立つ関連性のあるキーワードのサブセットを選択することになる。これは、とりわけ、可能な組合せの膨大な数と、キーワードのアイテム及びユーザに対する関連性を判定することの難しさが原因で困難な問題である。また、アイテムとは異なり、ユーザは時間の経過と共に変わる多くの側面を持ち、多くのキーワードを共有する可能性がある。

10

【0030】

LSIなどのパターン認識方法は、次元縮小をこの複雑性に対処するための方法として使用するが、それらの方法は実行され、更新され続けるために高いコストがかかり、さらにそれらの方法が実際に何をしているのかを理解することが難しい。

20

【0031】

本発明は、ユーザ及びアイテムの両方が同時にクラスタ化されるときに最も適切なクラスタ化が行われることに留意する。本発明は、大幅な複雑性の削減を実現し、ユーザ及びアイテムの両方に関して直感的で効果的なクラスタ化の結果を達成するアプローチを明らかにする。そのアプローチはコンテキストの概念に基づく。ここで、コンテキストはキーワードの集合として定義される。検索の場合、コンテキストはキーワードの論理積に基づく検索に対応する。アイテム又はユーザは、それらのアイテム又はユーザがコンテキストのすべてのキーワードにマッチする場合にコンテキストに属するとみなされる。コンテキストのサブコンテキストは、当該コンテキストのすべてのキーワードと少なくとも1つのその他のキーワードとを有するコンテキストである。したがって、サブコンテキスト内に存在するすべてのユーザ及びアイテムはコンテキスト内にも存在する。アノテーションイベント内のキーワードの集合はコンテキストである（アノテーションイベント自体が、ユーザがキーワードの定義をユーザID及びアイテムIDを含むように拡張する場合はコンテキストとみなされることができ、その場合、アノテーションイベントはキーワードの集合によって形成されるコンテキストに対するサブコンテキストである。また、ユーザはアイテムとみなされることができ。）。

30

【0032】

そのとき、アイテム及びユーザのクラスタ化は、ユーザ及びアイテムをまとめるために最も関連性のあるコンテキストの集合を決定することとして定義されることができ、複雑性の削減は、そのようなコンテキストの集合の大きさがシステム内のすべてのコンテキストの集合よりもずっと小さいときに達成されることができ、本発明は、関連性のある/有用なコンテキストは使用されるコンテキストであるという洞察を用いる。したがって、そのようなコンテキストの集合は、ユーザによってアイテムを記述することに使用されたアノテーションイベント内の実際のコンテキストから、並びに少なくとも特定の最小数のアイテム及びユーザを含むコンテキストを発見することによって決定されることができ、実際には、十分に多様なユーザ/アイテムの集団を有するアノテーションシステムに関して、小さな最小数でさえも大きな次元縮小と、興味に基づくカテゴリーに従ったユーザ/アイテムの効率的クラスタ化とをもたらすことができる。これは、ユーザが効果的にコミュニケーションできるようにどのようなトピックがそれらのユーザに対して有意義で

40

50

あるのかを判断することが難しいというフォーラムの実装の主な問題のうちの1つを克服する。本発明の創発的なコンテキスト又はトピックは、この問題に解決法を提供する、ユーザ及びアイテムに対する動的で適切なクラスタ化をもたらす。

【0033】

ウェブ検索を含む検索の場合、アノテーションは検索結果のクリックストリームに基づいてクエリログから得られることができる（それらはリンクテキストと有利に組み合わせられることもでき、そのようなリンクは異なるウェブホスト、又はブログ、又は独立したアノテータ及びその他のアノテーションソースのその他の代理からくる）。上で定義されたコンテキストの集合が計算され、カテゴリコンテキストと呼ばれることができる。検索における任意の所与のコンテキストに関して、検索クエリのサブコンテキストであるカテゴリコンテキスト内のコンテキストの集合が計算されることができ、結果として得られる集合内の（検索コンテキストのキーワードを取り出した後の）キーワードのそれぞれが上述のようにドリルダウン下位カテゴリとして提供されることができ、所与の下位カテゴリをドリルダウンすることは、クエリのコンテキストがドリルダウン下位カテゴリに対応するキーワードを含むようにそのクエリのコンテキストを変更することと等価である。したがって、それぞれのドリルダウンキーワードは、クエリのコンテキストのサブコンテキストに対応する。これは1語のキーワードに限定されず、ドリルダウンキーワードとみなされることができ単語同士の結びつき及び n -gram に基づく単語列を包含する。これらのドリルダウンキーワードは、特定のランク付け順に基づいてソートされて示されることができ、そのようなランク付け順は、そのようなサブコンテキストに関するイベント（又はユーザ若しくはアイテム）数から計算されることができ（さらに、それらの数は「今日」などの対象期間及び累計に基づいて計算されることができ）。さらに、カテゴリコンテキスト自体の計算が、所与の期間内のすべてのイベントがカテゴリコンテキストを計算するために使用される期間を定めた態様で行われることができる。

10

20

【0034】

ドリルダウン中の実際の検索プロセスは、そのようなドリルダウンを計算することによって使用されたアノテーションとは独立に行われることができ、全文検索などを含む、検索を実行するために検索エンジンが使用するどんな方法であってもよい。下位カテゴリの役割は単に関連性のあるキーワードをユーザに示すことであり、次にそのキーワードが検索クエリを増強するために使用される。

30

【0035】

[コンテキストベースのコミュニケーション方法]

コミュニケーションはすべてのコラボレーション活動の中核である。しかし、今日のほとんどのコミュニケーション技術は（電話、電子メール、SMS / IM などのような）1対1パラダイム又は（TV、ラジオ、ウェブなどのような）1対多パラダイムに限られている。最新技術において適切に対応されていないコミュニケーションの重要な種類、すなわち、多対多コミュニケーションが存在する。多対多によって示唆されるのは、電話会議の場合のような複数の人、又は電子メールの複数の受信者ではない。そうではなく、1対多の場合のように、受信者は、コミュニケーションの前はメッセージの送信者に知られていない。例として、企業ファイルシステム内にファイルを置く人は、そのファイルの存在をそのファイルを必要とする可能性がある場合によっては未知の人に伝えることができる必要がある。ブログはウェブ上で誰でもコンテンツをポストできるようにするが、対象のウェブユーザがブログのポスティングを発見することを可能にする有効なメカニズムは存在しない。

40

【0036】

問題は、「各アイテムに関してはそのアイテムのユーザであり、各ユーザに関してはそれらのユーザのアイテムである」と明確に表現されることができ、検索は人が関連性のあるアイテムを発見することを可能にするが、コミュニケーション及びコラボレーションの観点で恐らくより重要なのは、アイテムが関連性のある人を見出すための能力である。上述の方法がこの問題を解決するために有利に使用されることができ、

50

【0037】

多対多コミュニケーションシステムは2つの別個の部分、すなわち、関連性のあるアイテムを発見するユーザを検索する必要があるアイテムの発行者、及びユーザに関連するアイテムを求めてすべてのアイテムを検索するユーザとして実装されることができる。任意のそのようなメカニズムは実用的な実装をもたらすために包括性、検索能力、及びプライバシーに関する必要性のバランスを取らなければならないことが当業者には明らかであろう。

【0038】

本発明はコミュニケーションプロセスを3つの段階、すなわち、発行、コンテキスト設定、及びサブスクライブに分割する。本発明は、発行者が問題のアイテムが所望のサブスクライバに到達するために最も関連性のあるコンテキストを選択する必要がある、発行に対するコンテキストベースのアプローチを使用する。このコンテキストは、上述のアノテーションシステムのカテゴリコンテキストに制限されることが好ましい。発行者は、アイテムに加えて公開された形態の識別情報を明らかにする。これは、(発行者のユーザIDと同じであっても、又は同じでなくてもよい)発行者に対して一意な発行者IDを使用し、このIDを用いてアイテムをアノテーションすることによって行われることができる。発行の行為は、発行アノテーションイベントと呼ばれる特別な種類のアノテーションイベントを生成するか、アイテムに発行者IDを追加する通常のアノテーションイベントである可能性がある明確な行為に変換されることができる。

【0039】

サブスクライバは、ユーザがそれまでに興味を示すか、又は興味を明らかにしたカテゴリコンテキストの集合からのコンテキストに基づいてアイテムを周期的に取得する(引き出す)。次に、そのようなサブスクライブされたアイテムは、アイテムのキーワードと、ユーザがコンテキストに関して過去のアノテーションにおいて使用したキーワードとの間のマッチに基づいて「パーソナライズされる」か又は再ランク付けされることができる。これは、ドキュメントの発行者IDと、それまでにユーザが有用であると思った(又はアノテーションした)アイテムの発行者IDとの間のマッチを計算することによって有利に増強されることができる。発行者IDは、発行者に関する評判の分散された形態として働く。それまでに発行者からのアイテムを(選択又はブックマークするなど)アノテーションしたサブスクライバは、パーソナライズにおける再ランク付けの後で同じ発行者からの将来のアイテムを高くランク付けさせることができる。同様に、発行者がそれまでにユーザから比較的少ないアノテーションを受けている場合、当該発行者からの将来のアイテムは低くランク付けされる。フィードバックループを完成するためにアノテーションイベントは通常の見つけとは異なるように実装されることができ、アイテムをアノテーションするユーザはそれらのユーザが認める発行者IDをそれらのユーザのアノテーションに自動的に含めさせることができる。これらのアノテーションイベントは収集されることができ、ちょうどアノテーションイベントがアイテムをランク付けするようにコンテキスト内の発行者をランク付けすることを可能にすることができる。

【0040】

コンテキスト設定は、特定のコンテキスト内で初めて発行されたアイテムがそのコンテキスト内のユーザのサブセットにプッシュダウンされるプロセスである。このユーザのサブセットはアイテムに対するパイロットグループとして働き、それらのユーザがそのアイテムが適切と感じる場合、それらのユーザはそのアイテムをそのキーワードを用いてアノテーションする。これらのアノテーションイベントが収集され、次に、アノテーションされたアイテムがすべてのコンテキストサブスクライバに利用できるようにされる。コンテキスト設定フェーズは、既定の期間、アイテムがアノテーションを獲得するレートに対する下限、又はその他の好適な測定基準に制限されることができる。このプロセスは、特定の状況でアイテムの利用のプロセスを促進する。これは、アイテムが将来の検索又はサブスクリプションのためにそのアイテムを十分に分類するためのアノテーションの特定の最小レベルを受け取ることを可能にする。これは、新しいアイテムが関連性があり適切であ

10

20

30

40

50

ると感じた場合に当該新しいアイテムを奨励し、速やかな受け入れを助けることができるコンテキストに関して影響力のある発行者に当該新しいアイテムを紹介するために有利に使用されることができる。コンテキスト設定は、実装が実装することから価値を導き出せる場合に実装が実装することができる任意的なステップである。

【0041】

以下で検討されるようにこれらに従ってシステムをモデル化するいくつかの理由がある。1つの主要な理由はエンドユーザのプライバシーである。企業のワークグループなどのいくつかの場合、キーワードによってユーザを検索することは許容されるか、又は望ましい可能性さえある（「専門家を発見する」など）。そのような場合、人は検索された人と電話又は電子メールで直接連絡を取ることができる。しかし、多くの場合、ユーザはそれらのユーザの連絡先又は個人詳細が開示されることを望まない可能性がある。これらの場合、同様の目的が、送信者がそれらの送信者の識別情報を公開するが、しかしコンテキスト内のユーザに送信先を知らない状態でメッセージを送信することによって達成されることができ、そのようなユーザはそれらのユーザの個々の判断に基づいて返信することができる。送信者を1つのコンテキスト（又は、場合によっては少数のコンテキスト）に制限することは、送信者をメッセージを送るために妥当な人のグループを選択することに集中させる。これは、自動化されたプロセスがアイテムを収集する情報フィルタリング及び検索にはない人間の判断の重要な点である。

10

【0042】

公開された識別情報持つことによって、アイテムは、それらのアイテムの発行者IDにも基づいて検索及びサブスクリプションのためにランク付けされることができる。これは、エンドユーザが作者のポストに認めた価値に基づいて作者がサブスクライバの間で評判を確立することを可能にする。それは、発行者に関するアカウントビリティに関するメカニズムでもある。広く評価されている作者は、質の高いアイテムだけを発行することによってそれらの作者の評判を守ることに強い関心がある。知らないうちに又は故意にそれらの作者がそのようにすることができない場合、それらの作者は広く評価されなくなる。評判は築くのに時間がかかるので、広く評価された作者は、質の低いアイテムを奨励することによって得るものはほとんどなく、失うものは多い。これはスパム行為を働くことが難しい評判の分散された形態であり、そのような作者は、リンク解析、又はアノテーションベースの解析の統合的なハイパーリンクのいずれかから認められ得ること以外に重要な新しい特徴を情報検索アルゴリズムに加えることができる。本質的に、それは専門家の判断の要素をランク付けプロセスに含める。したがって、実装は、アイテム毎のアノテーション情報と発行者ID情報との両方を、コンテキスト内のアイテムの全体的なランク付けを計算することに取り入れることを選択することができる。

20

30

【0043】

コンテキスト設定フェーズは、アイテムが妥当なキーワードの集合を用いて速やかにアノテーションされるために重要である。上述のように、アイテムにコンテキストを設定するコミュニティの能力は、どんな個人の能力をもしのぐ。アイテム内のキーワードの分布がべき乗則に従うことが *del.icio.us* のようなコミュニティにおいて見られた。既定の割合を超えるユーザによってアノテーションされるキーワードの数は比較的一定していることが多く、特定の閾値を超えるイベント数に対してスケール不変性を示す。アイテムの定義特徴/キーワードと呼ばれるこれらの上位キーワードは、アイテムがコンテキスト設定などのコンテキストの影響を受ける態様でユーザに公開される場合に比較的早く取得されることができる。定義特徴は、アイテムのコンテキストのコミュニティの判断を適度に示し、サブスクリプションがより関連性があり、正確であることを可能にする。

40

【0044】

コンテキスト設定は新しいアイテムが速やかに認識されるためにも重要である。比較的知られていない発行者からのアイテムは、コンテキストに関する広く評価されている発行者に的を絞った態様で送信されることができ、それらの広く評価されている発行者が当該アイテムに価値を見出す場合、それらの広く評価されている発行者はそれらの広く評価さ

50

れている発行者の識別情報と共に当該アイテムを発行することができる（基本的に、それらの広く評価されている発行者の識別情報を元の作者と共にアイテムに追加する）。これは、新しいアイテムに関して集団全体の中での速やかな利用を可能にする。これは、有望な新しい才能が早く取り上げられることも可能にする。

【0045】

コンテキスト内のアイテムを処理するコミュニティの能力は、コンテキスト内のユーザ数に基づいて増す可能性がある。すべての人がすべてのアイテムを処理する必要はない。アイテムはコミュニティのサブセットの間で分割されることができ、並列的にコンテキスト設定されることができ、そのようなコラボレーションが行われるための有意義な場所を示す。例として、Googleにはインデックス付けされた50億を超えるページが存在し、1日当たり1千億を超える電子メール（残念ながらスパムを含む）が存在する。このコミュニケーションメカニズムの好適な実装はウェブ全体の大きさの領域を適正な期間内にコンテキスト設定することが可能である可能性がある。汎用的アノテーションメカニズムに関して、このコミュニケーション方法は、ウェブ検索におけるスパイダの役割の実用的な代替を示す。さらに、全般的なコンテキストレベルの統計が送信者に利用可能にされることができ、それらの送信者が適切なコンテキストを発見することを可能にすることができる。そのような統計は、コンテキスト内のユーザ数及びアイテム数を含むことができるがこれらに限定されない。ユーザ対アイテムの比が平均よりも高い場合、これは話題になっているトピックのよい指標である可能性がある。比が平均よりもかなり低い場合、送信者は、コンテキストがその送信者のメッセージに対して競争が激しすぎると判断することができる。これは、ウェブ上のコンテンツ生成に、又はより包括的には、任意のコラボレーションする組織がどのようにタスクにリソースを割り当てたいかに影響を与える可能性がある重要なフィードバックループを与える。

10

20

【0046】

アノテーションは最初のコンテキスト設定プロセスの後でさえも継続し、アイテムは時間の経過と共に異なる人によって異なるキーワードを用いて記述され続ける。コンテキスト設定フェーズは、関連性のある又は有望なアイテムが前面に出されるようにアイテムの最初の選別に寄与することができる。さらなるアノテーションは、定義キーワードのより関連性のある集合を通じてアイテムのさらなる特徴付けを可能にする。アイテムに関する定義特徴の中のキーワードは、コミュニティがアイテムの有用性を時間の経過と共に異なる態様で特徴付けるので時間に対して変化する（激しく変動する挙動を示す）可能性がある。アノテーションのプロセス全体は、送信先のユーザが未知であり、各ユーザがアイテムのコンテキストに対する関連性のそれらのユーザの判断に基づいて新しいコンテキストを通じてその他のユーザにアイテムを渡すソーシャルネットワークを介してアイテムを送信することに結びつけられることができる。効果的に、コンテキストの関連の意味ネットワークは、そのようなコンテキストに基づくインタラクションのソーシャルネットワークによって作成されるのみならず、そのようなコンテキストに基づくインタラクションのソーシャルネットワークを反映する。カテゴリコンテキストは、ユーザ間のアイテムの効率的な伝達を可能にする、そのようなネットワークにおける多くの接続を有するハブを示す。

30

40

【0047】

サブスクリプションプロセスは、コンテキストレベルで収集されたアイテムがユーザの興味に基づいて周期的に取り出されることを可能にする。コンテキストに対するユーザの興味は、ユーザによって明示的に指定されることができ、又はユーザのアノテーション、クリックストリーム、若しくはサブスクリプションプロセス内のアイテムの使用パターンに基づいて暗黙的に導出されることができ、ユーザの興味の明示的な指定は、特定のコンテキストに関するクエリを繰り返し行い、バックグラウンドでそのコンテキストに関するアイテムを継続的に取り出すことに相当する。しかし、検索プロセスと異なり、サブスクリプションに対するユーザの興味の明示的な指定は実用的な方法になりにくい。任意の所与の瞬間に、ユーザが気付いていない可能性があるユーザに潜在的に関連する多くの

50

コンテキスト内の多くのアイテムが存在する可能性がある。そのようなサブスクリプションシステムの暗黙的な目的はそのようなアイテムの発見を容易にすることである。本発明は、アノテーションイベントに基づいてユーザをプロファイリングして興味を推論するパーソナライズの形態を使用する。

【 0 0 4 8 】

当技術分野に知られているパーソナライズに対する多くのアプローチが存在する。本発明は、適切なパーソナライズを達成することにおいて重要な3つの点、カテゴリーコンテキストと、発行者IDと、再ランク付けのためのTFIDFの時間に基づく変形とを導入する。パーソナライズに対するほとんどのユーザプロファイリングに基づくアプローチは、これまでにユーザが価値があると思ったキーワードに基づいてアイテムをランク付けするように試みる。しかし、そのようなアプローチは、重要な新しい興味の領域を見逃し、キーワードの制限された集合を増強し続け、ユーザエクスペリエンスを損なう。カテゴリーコンテキストを使用することによって、コミュニティが興味深いと思うものに基づいた思いがけない発見の要素を導入することができる。概して、そのような思いがけなく発見されるアイテムは、ユーザを予期しない新しい領域に導き、ユーザプロファイルに対する関連性のある新しい興味の領域の発見を容易にする。これは、そのような新しいアイテムのアノテーションイベント（クリックなど）において暗黙的に、さらに、アイテムを読むこと／使用することが原因でユーザがそのような新しいコンテキストにおいて検索を実行することによって明示的に獲得されることができる。

10

【 0 0 4 9 】

サブスクリプションプロセスはそれぞれのカテゴリーコンテキストを、そのプロセスがサブスクライブする独立したアイテムのソースとみなす。サブスクリプションは、ユーザのプロファイル内のすべてのカテゴリーコンテキストからアイテムを取り出す。これは、興味の分布に基づいてそのような取り出しを分散することによってなされることができる。例として、我々がユーザが制限された注意範囲（1日当たりの特定の最大アイテム数など）を有すると仮定する場合、コンテキストから取り出されるアイテム数は、取り出されるアイテムの合計に対する比が、（読まれる／取り出されるアイテム数などのような）その特定のコンテキストにおいてユーザが費やす注意の量対合計の比と同じである可能性がある。この分布の計算は所与の期間に制限されることもでき、その期間中にユーザがプロファイリングされる。

20

30

【 0 0 5 0 】

様々なコンテキストのカテゴリーにまたがる興味の分布の形態のユーザプロファイルは、ユーザの検査のために要求に応じてユーザに示されることができる。ユーザは様々なメタファでそのような比を削除又は更新して、それらのユーザの趣向の調整を可能にすることができる。しかし、特定の比に対する変更がどのような意味を持つかはユーザに容易に明らかにならない可能性があるので、実施形態は、実際のユーザの使用に関わりなく継続的にダウンロードが行われるカテゴリーコンテキストをユーザが指定又は指定解除することを単に可能にすることができる。それは、ユーザがプロファイルから任意のカテゴリーコンテキストを取り除く／削除することも可能にすることができる。

【 0 0 5 1 】

そのようなコラボレーションによるランク付けと、ユーザが関連性があると思うものとの間のバランスが必要である。本発明は、特定のユーザに対する関連性を判定するために、ユーザプロファイルに基づく再ランク付けのためのTF-IDFアプローチの時間に基づく変形を導入する。特定のコンテキストの特定のユーザに関して、キーワード及びそれらのキーワードの実際の使用頻度が、そのコンテキストに関するアノテーションイベントから導出される。特定期間の使用頻度を計算するために、キーワードの使用回数と、ユーザがそのコンテキスト内のそのキーワードを最初に使ったときからの時間間隔とが測定される。本発明が特定期間の使用頻度と称するものを与えるために、この頻度を元に所定の期間に対する頻度が推定される。例として、所与のキーワードに関して、ユーザがそのキーワードの最初の使用以来2日の間にそのキーワードを2回使用した。これは、1年につ

40

50

き365回の特定期間の使用頻度を与える。この特定期間の使用頻度は、TF-IDFスタイルのアプローチでよくあるように頻度の対数値を使用することによって変動を小さくされ、コンテキストに対するユーザの興味のキーワードベクトルに対する重みとして使用される。アイテムのキーワードベクトルの重みは従来のTF-IDFの態様で行われることができ、このユーザに関するコンテキストに対するアイテムのランク(マッチ)が通常の態様で2つのベクトルの内積として計算されることができ、アイテムが、そのような計算されたランクに基づいてコンテキスト内で再ランク付けされる。

【0052】

上述のユーザの興味のプロファイルは扱いに注意を要する個人データを含む。したがって、ユーザは、記憶されているものを見て、それを変更するためのフルアクセスを有するそれらのユーザのPC上のクライアントサイドの実装の方が、そのような情報が集中サーバ上で管理されるよりも満足を感じやすい。しかしこれは、所与のコンテキストに関して、大量のデータをクライアントサイドにダウンロードすることなしに、コンテキストに対応するアイテムの集合全体に渡って再ランク付けを実行することができない可能性があることを意味する。あるいは、集中サーバ上でさえも、そのようなパーソナライズされた再ランク付けはコストがかかりすぎて実行できない可能性があるか、又はコンテキストのコミュニティのコラボレーションによるランク付けが失われる可能性があるので望ましくない可能性がある。これらの目的の間のバランスが、再ランク付けをコンテキストにおける上位のコラボレーションによってランク付けされた結果のサブセットに制限することによって得られる。これは、コンテキストからのいくつかのアイテムだけをそのコンテキストから取り出し、ユーザプロファイルに基づいてこれらの結果を再ランク付けすることによって達成されることができ、この数は、コラボレーションによるランク付けとパーソナライズの間の混合の制御を可能にする。

【0053】

TF-IDFの時間に基づく変形のようなレートに基づく計算が発行者IDと共に、興味深い作者の効率的検出に有利に適用されることができ、場合によっては多数のアイテムの一定のフローを有する可能性がある通常のキーワードとは異なり、ほとんどの作者は比較的少数のアイテムを生成する。発行者IDがアイテムに関するキーワードベクトル内のキーワードのように扱われる場合、通常のTF-IDFに似た方法でさえもその他のキーワードよりも発行者IDにより高い重みを与える。これは、発行者IDに関するユーザの特定期間の使用頻度を使用することによって好適に増強されることができ、例として、例えば発行者Aが全部で2つのブログのポストを発行しており、ユーザは当該発行者IDに遭遇してから過去2日の間に当該発行者のポストの両方を読んでいる。例えば、別の発行者Bが20のブログのポストを有し、ユーザはそれらのポストのうちの20個すべてをこの1年の間に読んでいる。発行者Aは新しいアイテムに関して発行者Bよりも高くランク付けされる。しかし、ユーザが発行者Aによるその後のアイテムを読まないか、又は発行者Aがその後アイテムを生成しない場合、ランク付けは時間の経過と共に落ちる。この方法は、最近有用だと分かった発行者IDが最初に比較的高くランク付けされることを保証するが、有用なアイテムを安定して生成するその他の発行者が有用なアイテムを安定して生成しない発行者を追い抜くことを可能にする。

【0054】

同様の時間に基づくTF-IDFアプローチが、システムのドリルダウンカテゴリにおいてカテゴリコンテキストをランク付けするために使用されることができ、本質的に、コンテキストのユーザの実際の使用、及びその最新性がランク付けプロセスに好適に組み込まれることができ、コンテキスト内のイベントの累積数又は所与の期間中の累積数だけではない可能性がある。

【発明の概要】

【0055】

本発明の広範な概念によれば、本発明はコラボレーションのための方法を提供し、この方法は、

10

20

30

40

50

一意な識別子を有する複数のユーザ間で共有することができる、一意な識別子を有する複数のアイテムを特定するステップと、

各ユーザにその他のユーザと無関係に少なくとも1つの自然言語の少なくとも1つのキーワードを用いて複数のそのようなアイテムをアノテーションさせるステップであって、それぞれのそのようなアイテムは少なくとも1人のユーザによってアノテーションされ、それぞれのそのようなアノテーションはアノテーションするユーザの識別子、アノテーションされているアイテムの識別子、及びアノテーションするユーザがアノテーションされているアイテムを記述するために選択する少なくとも1つのキーワードを含むアノテーションイベントによって示され、それぞれのそのようなアノテーションイベントは少なくとも1つの種類の複数のイベントソースから生成される、ステップと、

特定のアイテムに関連する前記キーワードが該アイテムのためのアノテーションイベントから収集され、かつ、特定のユーザに関連する該キーワードが該ユーザのためのアノテーションイベントから収集されるようにイベントソースからアノテーションイベントを収集するステップと、

少なくとも1人の前記ユーザにアイテム又はユーザをキーワードによって検索させるステップであって、該検索に使用したキーワードを前記収集されたキーワードの中に有するアイテム又はユーザが結果として返されるように検索させるステップとを含む。

【0056】

本発明の実施形態は、各アイテムが複数のユーザによってアノテーションされる可能性があり、各ユーザがアイテムの複数のアノテーションに渡って収集されたアイテムのキーワードに基づいてアイテムを検索することができるようにユーザが互いに独立して複数のアイテムをアノテーションすることを可能にするシステム及び方法を提供する。概して、本明細書において使用される用語「アノテーション」は、キーワードがユーザから収集され、その後、そのユーザの識別子と関連して記憶されるようなアイテムの任意の簡潔な記述を指す。アイテムは、一意な識別子によって特定されることができる任意のアイテム（ファイルシステム内のファイル、紙のドキュメント、プロセス管理システム内のタスク及び課題、リポジトリ内に記憶されたアイデアなどを含む）に対応する可能性がある。本発明の実施形態において、アノテーションは、発行、タグ付け、検索結果の集合内の結果のクリック、ファイルシステムパスからのディレクトリ及びファイル名、ハイパーリンクテキストなどを含む様々な方法で収集されることができる。

【0057】

本発明は、各イベントソースの種類に関して別々にクエリに対する関連性に基づいて検索結果をランク付けするステップと、各結果の最終的なランクを計算するためにそのようなランクを収集するステップと、関連性の順番で最終結果を示すためにすべてのイベントソースの種類にまたがって結果を収集するステップとをさらに含むことができる。

【0058】

一実施形態において、検索の結果集合はTF-IDFのような情報検索アルゴリズムを使用してランク付けされる。本発明の別の態様において、各アノテーションは、ハイパーリンクと、リンク解析ランク付けアルゴリズムに基づく結果集合の決定及び結果集合のランク付けとに等しいとみなされる。別の態様において、あらゆるアノテーションに関して、各ユーザはハブとみなされ、各アイテムはオーソリティとみなされ、結果及びランク付けがHITSなどのリンク解析アルゴリズムによって決定される。例として、そのような方法を通じて、企業のファイル共有におけるファイルなどのウェブページ以外のアイテムが、ウェブ検索に通常関連するより高い精度の恩恵を受けることができる。

【0059】

本発明の別の実施形態において、アノテーションは、ユーザ及びアイテムのコンテキストベースのクラスタ化を可能にするために処理される。アノテーションは、特定の既定のユーザ数及び既定のアイテム数を超えるユーザ数及びアイテム数を有するコンテキストがユーザ及びアイテム両方のクラスタを同時に示すようにキーワードに基づくコンテキスト

10

20

30

40

50

に従ってグループ化される。これは、検索結果に関するドリルダウンカテゴリを提供することによって検索プロセスを強化するために使用される。例として、ウェブ検索エンジンは、そのクリックストリームログからアノテーションを収集することができ、以下のクラスタ化方法を使用して結果を分類し、ユーザが意味のある態様で結果をさらに絞ることを可能にするあり得るクエリ修正を提供することができる。

【 0 0 6 0 】

本発明の別の実施形態において、キーワードに基づいてユーザを検索することができる。上述のように、これは、ユーザのアノテーションの収集されたキーワードを使用することによってか、又は H I T S、B F S、若しくは I N D E G R E E のようなリンク解析リンク付けを使用することを通じてかのいずれかで行われることができる。クエリに対するユーザのランク付けは、上述のようにアイテムと同様の態様で遂行されること

10

【 0 0 6 1 】

本発明の別の実施形態において、ユーザは、発行及びサブスクライブを通じてコンテキストベースの態様でその他のユーザとコミュニケーションすることができる。発行において、ユーザは、コンテキスト及び発行者識別子を用いて新しいアイテムをアノテーションすることによってその新しいアイテムをシステムに導入する。その他のユーザは、検索を通じて、又はサブスクライブによってそのようなアイテムを発見することができる。サブスクライブは、これまでにユーザが有用であると思ったコンテキストに基づいて上位の結果を自動的に検索し、取り出し、パーソナライズされた態様でそれらの結果を提示することを指す。そのようなコンテキストは、ユーザによって明示的に指定されるか、又はユーザのアノテーションに基づいて監視されることができる。本発明の一態様において、パーソナライズは、上位にランク付けされたサブスクライブされたアイテムのサブセットを T F - I D F の時間に基づく代替形態を用いて再ランク付けすることを通じて遂行される。本発明の別の態様において、サブスクリプションは指定期間内のアイテムに制限されることができる。本発明の別の態様において、発行及びサブスクライブの両方は、ユーザ及びアイテムのクラスタを示すコンテキストに制限されることができる。

20

【 0 0 6 2 】

別の実施形態において、発行は、各サブスクライバにおいてパーソナライズされた態様でアイテムをランク付けする基礎として使用される発行者識別子を用いた明示的な行為に変換される。これは発行者がユーザの集団中の利用に基づく分散された評判を獲得することを可能にし、高くランク付けされた発行者は質の高いアイテムを発行することに強い関心がある。これは、アイテムのランク付けで利用されることができる専門家の判断の形態をもたらす。

30

【 0 0 6 3 】

本発明の別の態様によれば、本方法は、コンテキストによってアイテム及びユーザの両方を同時にクラスタ化するステップと、コンテキストによってアノテーションイベントを収集するステップと、コンテキストの収集されたアノテーションイベントに既定の最小数の一意なユーザ識別子及び既定の最小数の一意なアイテム識別子を有するコンテキストを判定するステップと、そのようなコンテキストに基づいてアイテム及びユーザをクラスタ化するステップとをさらに含む。この方法は、クラスタ化アルゴリズムを使用することと、検索結果に対する下位カテゴリとしてクラスタを提示することとに基づいてアイテム又はユーザの結果集合をクラスタ化することをさらに含む。クラスタ化は、L S A、K - 平均、自己組織化マップ、主成分解析、多次元尺度構成法、及び射影法からなる群からの方法に基づいて遂行されることができる。クラスタ化は、キーワード、アイテム識別子、及びユーザ識別子からなる群からの少なくとも1つのデータの種類に基づいて実行されることができる。

40

【 0 0 6 4 】

本発明のさらに別の態様によれば、アイテムに対して定義キーワードが、アノテーションする集団の特定の割合を超える割合によって使用されるキーワードの集合にそれらのキーワードが対応するように、アイテムの収集されたキーワードから計算される。定義キー

50

ワードのこれらの集合は、意味、概念、及びそれらの意味関係の機械表現を決定するための基礎として使用される。意味関係は、パターン認識方法とLSAなどの相関解析方法とからなる群からの少なくとも1つの方法を用いて計算されることができ、オントロジが知識表現フォーマットで示されることができ、オントロジは、RDF、OWL、実体関連図、リレーショナルデータベーススキーマ、オブジェクト指向クラス、XML、及び表からなる群のうちの一つであるフォーマットで示されることができ、方法は、結果集合からのアイテムが特定のキーワードがそれらのアイテムの定義キーワード内に存在する場合に削除されることができ、フィルタリング方法をさらに含む。

【0065】

以下、本発明を添付図面を参照して説明する。

10

【好ましい実施形態の詳細な説明】

【0066】

[コンピューティング環境]

図3は、汎用的アノテーションシステムの全体の概略を示す。そのようなシステムは、イントラネットLAN/WAN140などのネットワーク接続することができ、さらに直接か又は当該LANを介してかのいずれかでインターネット150に接続する可能性があるクライアントシステム110を含む。そのようなクライアントシステム110は、インターネット150からのウェブページなどのコンテンツと、LAN140からのファイル及び電子メールなどのイントラネットのコンテンツとにアクセスし、受信し、取り出し、表示するために使用されることができ、コンテンツサーバ180は、コンテンツをクライアントシステム110にサービスすることができるウェブサーバ又はアプリケーションサーバなどのウェブ上で利用可能なサーバであることができる。メールサーバ184、ファイルサーバ183、データベースサーバ182、並びにサーバ181内に収容された多くの特別に作られたソフトウェアアプリケーション及びパッケージソフトウェアアプリケーションなどの、場合によってはクライアントシステム110にコンテンツをサービスする可能性があるイントラネットLANにおいて利用可能な多くの種類のサーバが存在する。場合によっては、クライアントシステム110は、エクストラネット、仮想プライベートネットワーク(VPN)、非TCP/IPベースのネットワークなどのその他の種類のネットワークを通じてコンテンツサーバに接続する可能性もある。

20

【0067】

30

クライアントシステム110は、処理ユニット202と、システムメモリ203と、システムメモリ及びその他のシステムコンポーネントを処理ユニット202に結合するシステムバス204とを含む通常のパーソナルコンピュータ201の形態の図4のような多目的コンピューティングデバイスで実装されることができ、システムバス204は、メモリバス又はメモリコントローラ、周辺バス、及びローカルバスを含むいくつかの種類のうちの一つかであるか、各種のバス構造のいずれかを使用することができる。システムメモリ203は、読み出し専用メモリ(ROM)205及びランダムアクセスメモリ(RAM)206を含む。ROM205に記憶される基本入出力システム(BIOS)207は、パーソナルコンピュータ201のコンポーネント間で情報を転送する基本的なルーチンを含む。BIOS205は、システムのための起動ルーチンも含む。コンピュータ201は、ハードディスク(図示せず)からの読み出し及びハードディスクに対する書き込みのためのハードディスクドライブ208と、リムーバブル磁気ディスク210からの読み出し及びリムーバブル磁気ディスク210に対する書き込みのための磁気ディスクドライブ209と、CD-ROM又はその他の光媒体などのリムーバブル光ディスク212からの読み出し及びリムーバブル光ディスク212に対する書き込みのための光ディスクドライブ211とをさらに含む。ハードディスクドライブ208、磁気ディスクドライブ209、及び光ディスクドライブ211は、それぞれハードディスクドライブインターフェース213、磁気ディスクドライブインターフェース214、及び光ドライブインターフェース215によってシステムバス204に接続される。ドライブ及びそれらのドライブに関連するコンピュータ可読媒体は、パーソナルコンピュータ201のためのコンピュー

40

50

タ可読命令、データ構造、プログラムモジュール、及びその他のデータの揮発性の記憶を提供する。コンピュータによってアクセス可能なデータを記憶するその他の種類のコンピュータ可読媒体が動作環境内で使用されることもできる。

【0068】

プログラムモジュールは、ハードディスク、磁気ディスク210、光ディスク212、ROM205、及びRAM206に記憶されることができる。プログラムモジュールは、オペレーティングシステム216、1つ又は複数のアプリケーションプログラム217、その他のプログラムモジュール218、及びプログラムデータ219を含むことができる。ユーザは、キーボード222及びポインティングデバイス221などの入力デバイスを通じてパーソナルコンピュータ201に命令及び情報を入力することができる。その他の入力デバイス(図示せず)は、マイクロホン、ジョイスティック、ゲームパッド、衛星通信用パラボラアンテナ、スキャナなどを含む可能性がある。これらの及びその他の入力デバイスは、システムバス204に結合されたシリアルポートインターフェース220を介して処理ユニット202に接続されることが多いが、それらの入力デバイスは、パラレルポート、ゲームポート、又はユニバーサルシリアルバス(USB)などのその他のインターフェースを介して接続されることができる。モニター228又はその他のディスプレイデバイスも、ビデオアダプタ223などのインターフェースを介してシステムバス204に接続する。ビデオカメラ又はその他のビデオソースが、ビデオ会議及びその他のアプリケーションのためのビデオ画像を提供するためにビデオアダプタ223に結合され、それらのビデオ画像は処理され、さらにパーソナルコンピュータ201に送信されることができる。さらなる実施形態において、別個のビデオカードが、衛星放送の符号化画像を含む複数のデバイスからの信号を受け取るために提供されることができる。モニターに加えて、概してパーソナルコンピュータは、スピーカ及びプリンタなどのその他の周辺出力デバイス(図示せず)を含む。

10

20

【0069】

パーソナルコンピュータ201は、リモートコンピュータ229などの1つ又は複数のリモートコンピュータへの論理接続を使用したネットワーク化された環境で動作することができる。リモートコンピュータ229は、別のパーソナルコンピュータ、サーバ、ルータ、ネットワークPC、ピアデバイス、又はその他の一般的なネットワークノードであってよい。概して、そのリモートコンピュータ229は、パーソナルコンピュータ201に関連して上述されたコンポーネントのうち多く又はすべてを含む。図4に示された論理接続は、ローカルエリアネットワーク(LAN)227及び広域ネットワーク(WAN)226を含む。

30

【0070】

LANネットワーク環境内に置かれるとき、PC201は、ネットワークインターフェース又はアダプタ224を介してローカルネットワーク227に接続する。インターネットなどのWANネットワーク環境において使用される場合、概して、PC201は、モデム225、又はネットワーク226を介して通信を確立するためのその他の手段を含む。モデム225はPC201に内蔵されるか又は外付けされることができ、シリアルポートインターフェース220を介してシステムバス204に接続する。ネットワーク化された環境において、201内に存在するように示されている、Microsoft Wordを含むプログラムモジュールなどのプログラムモジュール又はそのプログラムモジュールの一部は、リモートの記憶装置230に記憶されることができる。

40

【0071】

クライアントシステム110は、デスクトップパーソナルコンピュータ、ワークステーション、ラップトップ、携帯情報端末(PDA)、セル電話、又は任意のWAP対応デバイス、若しくはインターネットと直接的に又は間接的にインターフェースを取ることができる任意のその他のコンピューティングデバイスも含む可能性がある。クライアントシステム110は、MicrosoftのInternet Explorer(商標)ブラウザ、Netscape Navigator(商標)ブラウザ、Mozilla(商標

50

) ブラウザ、Opera (商標) ブラウザ、又はセル電話、PDA、若しくはその他の無線デバイスなどの場合にはWAP対応のブラウザなどのブラウジングプログラム内で実行されることができる。

【0072】

サーバシステム120はイントラネット環境内のアノテーションサーバに対応し、サーバシステム130はウェブ中からクライアントにサービスを提供することができるインターネット150上のアノテーションサーバに対応する。サーバシステム120は、アノテーションイベントを受信し、イベントを収集し、クライアントからの検索及びサブスクリプション要求を処理することによってクライアントシステムにサービスを提供する。これは、上述のようなPC、又はSun MicrosystemsからのUNIXサーバ、Linuxベース及びWindowsベースのIntelのサーバなどのサーバ構成上に実装されることができる。

10

【0073】

ソフトウェアは、C、C++、Java、C#、Visual Basic、PERL又はTCLなどのスクリプト言語を含む多くの様々な方法を使用して設計されることができる。クライアントシステムの態様は、HTML、XML、Java、JavaScript、ActiveXなどのコードなどのブラウザベースの供給、又は任意のその他の好適なスクリプト言語(例えば、VBScript)のために開発されることができる。一部の実施形態において、いかなるコードもクライアントシステム110にダウンロードされず、必要なコードがサーバによって実行されるか、又はクライアントシステム110に既に存在するコードが実行される。

20

【0074】

本発明は、ハンドヘルドデバイス、マルチプロセッサシステム、マイクロプロセッサベースのプログラム可能な家庭用電化製品、ネットワークPC、ミニコンピュータ、エンジニアリングワークステーション、メインフレームコンピュータなどを含むその他のコンピュータシステム構成を用いて実施されることができる。本発明は、デジタル電子回路で、又はコンピュータハードウェア、ファームウェア、ソフトウェアで、又はそれらの組合せで実装されることができる。好適なプロセッサは、例として、汎用マイクロプロセッサ及び専用マイクロプロセッサの両方を含む。上記のいずれも、特別に設計されたASIC(特定用途向け集積回路)によって補完されるか、又は特別に設計されたASIC(特定用途向け集積回路)に組み込まれることができる。

30

【0075】

[アノテーション]

上で説明されたように、アノテーションは、ユーザの判断に従ってアイテムを記述するキーワードが導出されることができるような、ユーザによるアイテムの任意の簡潔な記述を指す。これは、そのようなアノテーションが発生する度に対してイベントの形態で発せられる。イベントに含まれるデータ要素のサンプルが図14にXMLの形態で示される。そのようなイベントは、複数のイベントソース及びイベントソースの種類から複数の態様で生成されることができる。好ましい実施形態は、イベントジェネレータ112の概念を用いてこれを遂行する。これらのイベントジェネレータは、ツールバー、アドイン、共有ライブラリ、OSレベルのサポートなどの形態であることができる。それぞれのイベントソースの種類は、その種類独自のイベントジェネレータを有する。それぞれのイベントソースの種類に対して複数のイベントジェネレータが存在する可能性もある。好ましい実施形態はユーザインタラクションから直接キーワードを獲得するが、それらのキーワードは、ユーザの電子メール及びドキュメントの全文検索に対して自動化されたプロシージャを使用することによって取得されることもできる。また、メールメッセージ又はドキュメントに対するキーワードの指定は、強調されるべきテキストを指定することと同様の態様でテキストにおいてユーザによって行われることができる。当業者は、そのようなキーワードがユーザから取得されることができる多数の態様が存在することを認識するであろう。

40

【0076】

50

例として、ブラウザベースのアノテーションイベントは、図5のようなブラウザに対するツールバーアドインから生成されることができる。ファイルシステムに基づくイベントは、図6のようなエクスプローラアドインから生成されることができる。各イベントジェネレータ112は、そのイベントジェネレータ112がアノテーションイベント内にセットすることができるイベントソースID及びイベントソースラベル要素を通じてそのイベントジェネレータ112自身を記述することができる。ブラウザにおけるウェブページに対するアノテーションイベントは、企業LANのファイルシステムに基づくアノテーションイベントとは異なるイベントソースによって示されることができる。そのようなイベントは、イベントジェネレータ112に基づいて、LANなどの私設ネットワーク上の1つ若しくは複数のアノテーションサーバに伝えられることができるか、又はインターネット上の1つ若しくは複数のアノテーションサーバに伝えられることができる。

10

【0077】

アノテーションイベントは、アノテーションされているアイテムに関するアイテムIDと、アノテーションするユーザに関するユーザIDとを含む。最大の精度のために、すべてのイベントソースからのイベントに渡って一意なアイテムは同じ一意なアイテムIDに対応すべきであり、一意なユーザは同じ一意なユーザIDに対応すべきである。しかし、好ましい実施形態は、実装の要件の通りにこの目的を達成する最良の態様を決定するためにこの実装を離れる。これは、たとえユーザID及びアイテムIDが上記の要件に厳密に従わないとしても異種のイベントジェネレータと統合することから得られる利点があるので行われる。収集値は、アイテム及びユーザの適度な多様性のある集団が存在する多くの場合、精度よりも価値がある可能性がある。例として、ウェブ上の同じページが複数のURL(事実上複数のアイテムID)を有する可能性がある。しかし、多くの場合、それらのうちの1つだけを見つければ十分である。Googleのような検索エンジンは、ユーザの容易さのためにそのようなURLを収集するように試みるが、ウェブの基本的構造はそのような収集又は一意性を義務づけない。

20

【0078】

好ましい実施形態において、アイテムは、URIによって与えられる一意なアイテムIDによって特定されることができる任意のものであることができる。当然、これは、フォークソノミでよく見られるようにURLを使用するウェブコンテンツを含む。これはファイルシステム内のファイル及びフォルダ、メールサーバの電子メールメッセージも含むことができ、さらに、バーコードを有する紙のドキュメントなどの物理的対象、一意なIDを有するプロジェクト管理システム内のタスク/課題、URIを有するアプリケーション内にテキストの形態で記憶されたブレインストーミングセッションにおけるアイデアなどを含むことができる。そのような一意なアイテムIDを生成するための当技術分野において知られている多数の方法が存在し、それらの方法が実装の要件の通りに利用されることができる。好ましい実施形態は、電子メールがタイトル及び本文を許可する態様と同様に、アノテーションイベントに含まれるアイテムに対してタイトル及び説明を任意的に指定することを可能にする。これは、図14におけるようにアイテムIDと共に記憶される。好ましい実施形態は、ユーザがユーザのためにローカルに記憶される「記憶」ダイアログアノテーションイベント中にタイトル及び説明を指定することを可能にする。これは、図10のようなアノテーションブラウザウィンドウの「マイアイテム」オプションでアイテムを示すことに使用されることができ、ユーザがアイテムの記述をカスタマイズすることを可能にする。発行者は、アイテムがシステムに対して新しい場合はアイテムのタイトル及び説明を指定しなければならず、そうでなければ、図9のような「発行」ダイアログはサーバに記憶されたアイテムの情報からこの情報を予め記述する。クリックストリームアノテーションイベントは、タイトル及び説明情報を指定しない。その他の実施形態は、最新のタイトル及び説明情報を用いてサーバ内のアイテムの情報を更新するなど、いくつかの異なる態様でこれを処理することができる。一部の実施形態は、ユーザに基づくキーを用いてアノテーションイベントにデジタル署名を施して、必要に応じて認証及び否認防止を可能にすることができる。

30

40

50

【0079】

例として、図5、6、及び7は、イベントジェネレータ112をツールバーの形態で統合するアプローチを示す。ツールバーは、ウェブブラウザ、ファイルシステムエクスプローラ、及び電子メールアプリケーションなどの既存のアプリケーションに対するアドインとして働くプログラムである。これは当技術分野で知られており、現在、MSN、Yahoo!、及びGoogleからのツールバーなどの多くの例が存在する。ツールバーは、「記憶」ボタン及び「発行」ボタンを有する。「記憶」ボタンはユーザが図8のような記憶ダイアログウィンドウを起動することによってアイテムをアノテーションすることを可能にし、「発行」ボタンはユーザが図9のような発行ダイアログウィンドウを起動することによってアイテムを発行することを可能にする。図5は、ウェブブラウザのツールバーがユーザが現在表示されているURLを「記憶」ボタンを通じてアノテーションすること、又は「発行」ボタンを用いてアイテムをシステムに発行することを可能にすることができることを示す。そのようなツールバーは、ユーザが表示されたページ内の任意のハイパーリンクを右クリックし、記憶又は発行ダイアログウィンドウを起動するメニューアイテムを選択することも可能にすることができる。さらに、ユーザがツールバー内の検索を使用してページを検索し、結果内の返されたURLのうちの1つをクリックする場合、ツールバーはそのようなユーザの行為を監視し、図15のような、アイテムに対する検索において使用されたキーワードを使用するイベントを自動的に生成することができる。これは、生成されるイベントを決定する前に、ユーザがクリックされたページを読むかどうか、又はさらにはユーザがどこまでページを読むかを評価するなど、さらなるユーザの監視によって増強されることができる。本発明は、一意なユーザから生成されたイベントが収集されるときに最もよく機能する。したがって、アノテーションするユーザに対するユーザIDの割り当てが必要とされる。これは、ユーザにクッキーとして記憶される一意なIDを動的に割り当てるか、又はアノテーションの前にユーザにサーバにログインさせることによってインターネットベースのアノテーションサーバに対して遂行されることができる。ブラウザベースのイベントジェネレータは、そのイベントジェネレータのすべてのアノテーションイベントをそのようなイントラネットサーバに送信することができるが、公的なウェブのページに対して生成されたイベントをインターネットベースのアノテーションサーバに任意的に送ることができる。ユーザIDはURIの形態で示される。

10

20

【0080】

図6において、メタファの同様の集合がファイルシステムの場合に使用されることができる。ちょうどブラウザの場合のように、任意のファイル又はフォルダが、「記憶」又は「発行」ボタンのいずれかを使用してアノテーションされることができる。アイテムを右クリックすることは、ボタンと同じ機能にアクセスするためのコンテキストメニューアイテムを与えることができる。ファイルシステムのアイテムに対して実行される検索は、上述のようにバックグラウンドでアノテーションされることができる。ユーザID情報は、イントラネットLANにログインしたユーザなどの特定の場、オペレーティングシステムから有利に取得されることができる。そのようなイベントは、セキュリティ及びプライバシーなどの理由でイントラネットベースのアノテーションサーバにのみ送信される可能性がある。

30

40

【0081】

図7は、同じ概念がMicrosoft Outlookなどの電子メールソフトウェアに適用されることを示す。任意のメールが上述のようにキーワードを用いてアノテーションされることができる。同様に、ファイル保存及びファイルオープンダイアログボックスが、キーワードを用いてファイルをアノテーション又は発行し、キーワードに基づいてファイルを検索する能力によって増強されることができる。当業者は、そのような機能が任意の所与のアプリケーション内に実装されることができる多数の態様が存在することを認識するであろう。例として、ツールバー又はOSが、特別に作られたアプリケーションが任意のアイテムに対して記憶ダイアログウィンドウを起動することを可能にするAPIを提供することができる。そのような特別に作られたアプリケーションは、そのアプリケ

50

ーション独自のイベントソース識別情報を有する可能性があり、そのアプリケーション独自の要件の通りにアイテムIDを予め記述することができる。

【0082】

したがって、説明の目的で、すべてのアノテーションイベントが、上記のツールバーのような(図3の)イベントジェネレータ112によって生成されると仮定される。これらのイベントジェネレータ112は、ジェネレータによってイベントに追加される一意なイベントソース識別子を持つことができる。例えば、電子メールをアノテーションするツールバーは、`http://www.abc.tld/Email`のようなイベントソースIDと、「ABC Company Email」のようなラベルとを追加することができる。イベントソースIDにおける名前空間の競合を避けるために、好ましい実施形態はURIに基づく構文を使用する。しかし、一意性を保証する責任は実装に委ねられる。

10

【0083】

ユーザのユーザIDは、実装の要件の通りにいくつかの態様でイベントジェネレータ112によって決定されることができる。これは、ユーザにユーザID(電子メールIDなど)及びパスワードを用いてアノテーションサーバにログインさせ、それによってアノテーションサーバが一意なユーザIDを生成することを可能にすることによって獲得されることができるか、又はそれは、オペレーティングシステムのログイン情報、連携型識別情報ソリューション、シングルサインオンデータを使用することができるか、又はそれは自動的に一意なIDを生成し、そのIDをブラウザにクッキーとして保持することができる。

20

【0084】

一部の実施形態において、クローラの形態のその他の形態のアノテーションイベントジェネレータを有することができる。ウェブクローラ170は当技術分野においてよく知られており、インデックス付けのためにウェブ上のページを取得するために検索エンジンによって使用される。そのようなクローラが、ウェブページの間のハイパーリンクを発見し、アノテーションイベントを生成するために使用されることができる。リンクテキストがキーワードの代わりに使用されることができ、ウェブホスト又はブログの情報がユーザIDの代わりに使用されることができる。これは、概してリンクテキストが少数の関連性のあるキーワードを決定するのに有用であり、判断の独立したソース(ハイパーリンクを作成する実際のユーザ)を特定することが容易でないので上述の記憶ダイアログ又はクリックストリームに比べて質が低い可能性があるイベントを生成する。Technoratiのrel="tag"スタイルのタグが、ページに関するアノテーションイベントを生成するために使用されることができる。しかし、これは、ページの作者だけがそのようなタグを割り当てることが可能なのでやはり制限される。

30

【0085】

その他の実施形態において、イントラネットベースのクローラ160を生成することもでき、ファイルシステム内のすべてのファイル、メールサーバ内のメール、又はその他のデータの種類が自動的にアノテーションされることができる。例として、ファイルシステムクローラは、個人の及び共有のドライブを含むファイルシステム全体をスキャンすることができる。そのようなクローラは当技術分野に知られており、デスクトップ検索ソフトウェアにおけるファイルをインデックス付けするプログラムと同様である。そのようなプログラムは、ディレクトリパス内のディレクトリ及びファイルのラベルをファイルに対するキーワードとして使用するように修正されることができる。これは、(暗号的ハッシュなどの)ファイルのコンテンツに基づく一意なハッシュを計算し、異なるユーザの個人フォルダ内の同じファイルを発見し、異なるユーザID及び異なるキーワードの集合を使用して同じファイルに対して異なるイベントを生成することができる。しかし、ユーザがファイルを直接アノテーションするアノテーションメカニズムと比較してそのような方法には制限が存在する。これらは、「スタッフ(Staff)」のような誤解を招きかねない名前を有するファイル、関連性のある情報をほとんど又はまったく持たない電子メール及びその他の形態のデータなどを含む可能性がある。したがって、そのようなクローラは

40

50

、アイテムのコンテンツに基づいてキーワードを生成する自動アノテーション方法も使用することができる。アイテムをアノテーションサーバに取り込む好ましい方法は、後で詳細に説明される発行及びサブスクリプトパラダイムを使用することである。

【0086】

[収集]

アノテーションイベントが、本発明の機能を達成するために収集される。これは、情報検索におけるドキュメントのインデックス付けに幾分似ている。ドキュメントのインデックス付けにおいて、各単語がその単語が現れたドキュメントにマッピングされる単語の転置インデックスが生成される。本発明において、イベントは、各ユーザIDをそのイベントにマッピングし、アイテムIDをそのイベントにマッピングし、各コンテキストをそのイベントにマッピングする3つの別個のマッピングに収集される。ここで、コンテキストは、少なくとも1つのキーワードからなるキーワードの集合を指す。このマッピングを実現するために利用されることができる、当技術分野で知られている多くのインデックス付け又はハッシュ方法が存在し、そのようなプロセスを詳細に説明する必要はない。参考として、1つのそのような方法が、Google, Inc. のDeanらによる論文「Map Reduce: Simplified Data Processing on Large Clusters」に見られる。

10

【0087】

第1に、ユーザに関してすべてのイベントがマッピングされる。これは、同じユーザIDを有するすべてのイベントが同じデータ構造で利用できることを示唆する。これは、イベントの正規化及びユーザプロファイリングのために行われる。ユーザに基づくマッピングは、実装の要件に応じてクライアントシステム110又はサーバシステム120において行われることができる。好ましい実施形態は、ユーザに基づくマッピングをクライアントシステム110において使用することがプライバシー、セキュリティ、及びネットワークの末端の計算機能力を使用することに関してより望ましい可能性があるためユーザに基づくマッピングをクライアントシステム110において使用する。これは、すべてのイベントジェネレータ112からのイベントを受信し(ステップ300)、それらのイベントを111のような永続的記憶装置のローカルに記憶するクライアントシステム110上に存在するソフトウェアを用いて実現されることができる。

20

【0088】

ほとんどのフォークソノミ及びその他のアノテーション方法は、ユーザがアイテムを一度だけアノテーションすると仮定する。しかし、本発明の汎用的アノテーションメカニズムなどの汎用的アノテーションメカニズムにおいて、ユーザは同じアイテムを複数回アノテーションすることができる。これは、異なるコンテキストにおいてアイテムをアノテーション/使用すること、又は異なるイベントジェネレータにおいてそのアイテムを使用することが原因である可能性がある。所与のアイテムに関するすべてのアノテーションイベントがユーザに基づくマッピングにおいて収集され、各イベントジェネレータに対して一意なコンテキストを示すイベントの集合が計算され、ローイベント(raw events)と呼ばれる。次に、ローイベントのこの集合が正規化される(図11のステップ310又は405)。正規化は、各ユーザに関するアイテムに対するすべてのキーワード及びそれらのキーワードの使用回数を累積し、ユーザに関するアイテムに対するキーワード使用の合計で割ることを指す。例として、ユーザがアイテムに対して2つのキーワード(例えば、キーワード1及びキーワード2)を使用する場合。ユーザがアイテムに対してキーワード1を使用する合計回数が3であり、キーワード2を使用する回数が7である。そのとき、正規化の後、キーワード1の数は $3 / (3 + 7)$ で0.3となり、同様に、キーワード2の数は0.7となる。正規化の最後に、正規化されたイベントが、特定のユーザに関する所与のアイテムに対して生成される。正規化は、実装によって要求されるようにアイテム毎にイベントソース毎か又はすべてのイベントにまたがって実行されることができる。イベントを正規化する多数の態様が存在する可能性があるが、好ましい実施形態はすべてのイベントソースにまたがって正規化を行い、アイテム毎にキーワードを収集すると

30

40

50

きに各ユーザがアイテム毎に1つのキーワードの「投票」を事実上得ることを保証する。

【0089】

そのような正規化されたイベントとそのイベントに対応するローイベントとが、120のようなイントラネットベースのサーバ又は130のようなインターネットベースのサーバなどの関連するアノテーションサーバに送信される(ステップ320)。これは、漸次的な態様で、状態変化があるときにのみ行われる。正規化されたイベントをどのサーバに送信するかを選択はアイテムに基づくことができる。アイテムがインターネットのウェブページのような公的資産である場合、イベントは、インターネットベースのアノテーションサーバ及びイントラネットベースのアノテーションサーバの両方に伝達されることができる。アイテムがイントラネットのドキュメントである場合、イベントはイントラネットベースのサーバにだけ送信されることができる。クライアントシステム110がイベントを伝達することができる複数のインターネット又はイントラネットサーバが存在してもよい。この情報は、クライアントシステム110内に設定プロファイルとして保存されることができる。

10

【0090】

次に、イベントは、収集モジュール122又は132によって120又は130などのアノテーションサーバで受信される。はじめに、正規化されたイベントがアイテムに基づいてマッピングされる(ステップ406)。これは、特定のアイテムIDに対応するすべてのイベントが同じデータ構造に収集されることを意味する。これらは正規化されたイベントなので、各アイテムは特定のユーザIDからの最大で1つのイベントを有する。キーワードの数は、アイテムをアノテーションしたすべてのユーザからのキーワードによって収集される。キーワードの数の合計は、ユーザ数に等しいか、又は実質的にアイテムをアノテーションした判断の独立したソースの数に等しい。上述のように、キーワードの分布はおおよそべき乗則の分布に近い。これは、アイテムをアノテーションするユーザのうちのあるわずかな数を超えるユーザが使用するキーワードの数がほぼ一定であり、アノテーションイベントの数に対してスケール不変であることを意味する。例として、アイテムに関するアノテーションする集団のうちの5%超によって使用されるキーワードの数が、アイテムが50回のアノテーションを受けたか、500回のアノテーションを受けたかによらずほぼ一定である場合。5%又は10%などの実装の必要に応じて好適な割合のユーザによって使用されるキーワードが、アイテムの定義特徴又はキーワードとみなされることができる。定義キーワードは、アノテーションの収集から生じるグループの記述であり、アイテムをこれらのキーワードに潜在するトピックに割り当てるための信頼できるガイドである。アイテムに関する定義キーワードは、イベントに基づいて上述のように更新される(ステップ407)。定義集合内のキーワードは激しく変動する挙動を確かに示し、すなわち、異なる単語が時間の経過につれて定義集合に入り、出て行く可能性があることに留意されたい。

20

30

【0091】

次に、収集モジュールがコンテキストに基づいてローイベントをマッピングする。コンテキストは単一のキーワード又は2つ以上のキーワードの集合である可能性がある。コンテキストのすべてのキーワードをそのイベント内に有するイベントは、コンテキストの一部とみなされる。したがって、イベントからコンテキストへのマッピング408は、ユーザがコンテキストの一部であるすべてのイベントを単一のデータ構造に収集することを可能にする。実際には、アイテムと同じか又はそれ以上の一意なコンテキストが恐らく存在し得る。上述のように、本発明は、カテゴリコンテキストの概念を使用してコンテキスト空間の複雑性を削減し、ユーザ及びアイテムの両方の同時のクラスタ化を可能にする。好ましい実施形態において、カテゴリコンテキストがローイベント(正規化されていないアノテーションイベント)から計算され、特定の最小数の一意なアイテム及びユーザを有するコンテキストに対応する。一部の実施形態は、すべてのアイテムが特定の最小数のユーザによってアノテーションされる特定の最小数のアイテムを有するカテゴリコンテキストにカテゴリコンテキストの定義をさらに制限することができる。一部の実施形態

40

50

は、カテゴリコンテキストの計算のために正規化されたイベントを使用することを好む可能性がある。その他の実施形態はローイベントを使用することができるが、アイテム及びユーザに対するそれぞれの一意なコンテキストに関して1つのイベントだけが送信されるように、サーバに送信されるローイベントを制限することができる。

【0092】

カテゴリコンテキストは、イベントが受信されるときに生成的に計算されることができる(ステップ408)。ユーザは、一意なキーワードに基づいてイベントを収集することからはじめることができる。イベントが特定のキーワードに累積されるとき、これらは元のキーワードを削除し、それぞれがそれ自体及び元のキーワードのコンテキストを示す一意なキーワードの第2の集合を取得することによって再びハッシュされることができる。これは、各コンテキストがカテゴリコンテキストに関する所定の基準を満たすときにカテゴリコンテキストを繰り返し生成し続けることができる。好ましい実施形態において、この方法は、トピックドリフトを防止するように2つのさらなる制限によってさらに増強される。アイテムの定義キーワードに対応するキーワードだけが、イベントがコンテキストの一部であるかどうかを判定するために使用される。さらに、特定のレベル未満のイベントのレート有するカテゴリコンテキストはカテゴリコンテキストであることから除外される(又は、最も最近使用されたカテゴリコンテキストのみを保持するキャッシュメカニズムが使用されることができる)。上記の制限は、より大規模な実装においてより意味を持つ可能性があるかなり厳格な要件である。これらは、そのような制限から有用性を得ない実装に対して少なくとも以下の態様で緩和されることができる。

- ・すべての受信されたイベントが計算のために使用される
- ・たとえコンテキストのキーワードがアイテムに関する定義キーワードに対応しないとしてもコンテキストのキーワードを含むイベント
- ・コンテキストの少なくとも1つのキーワードがアイテムに関する定義キーワードに対応する場合にそのようなキーワードを含むイベント
- ・イベントを使用する代わりにそれぞれユーザ及びアイテムの収集されたキーワードに基づいてコンテキストに対応するユーザ及びアイテムを判定することによる
- ・コンテキストのキーワードに基づいてアイテムを判定し、そのようなアイテムをアノテーションしたユーザを判定することによる
- ・コンテキストのキーワードに基づいてユーザを判定し、それらのユーザがアノテーションしたアイテムを判定することによる

【0093】

さらに、好ましい実施形態は、「統合的な」カテゴリコンテキストを生成する。コンテキストは、有向非循環グラフ(又はDAG)を形成すると考えられることができる。例として、コンテキスト「キーワード1 AND キーワード2(キーワード1+キーワード2)」は、キーワード1に対応するアイテム及びキーワード2に対応するアイテムのサブセットであるアイテムを表す。キーワード1及びキーワード2は個々にコンテキスト「キーワード1+キーワード2」の親であると考えられることができる。キーワード1+キーワード2がカテゴリコンテキストである場合、好ましい実施形態は、(たとえその親コンテキストが所定の基準に基づいてそれらのイベントを収集することに基づいてカテゴリカテゴリとして適格でない可能性があるとしても)その親コンテキストもカテゴリコンテキストとして「統合する」。

【0094】

好ましい実施形態は、カテゴリコンテキストに基づいてイベントを収集する。これは、コンテキストがカテゴリコンテキストになるときにイベントデータがその親コンテキストとは別個に管理されるように生成的に行われる。これは、各カテゴリコンテキストに関するユーザ及びアイテムのランク付けの別個の計算を可能にし、それぞれの別個の計算を発行及びサブスクライブのための別個の目的として扱う。これは遅延された態様で行われることができ、イベントは、検索クエリのように要求されるか、又はそれに関して発行/サブスクライブ要求が受信されるときにのみカテゴリコンテキストに関して収集さ

10

20

30

40

50

れることができる。

【0095】

当業者は、アノテーションの収集の異なる要素に渡るアノテーションイベントの挿入、更新、及び削除の実装はそれぞれの実装の要件に依存するが、実装が上述の基本的な収集の要件に従うように実装することは比較的簡単であることを理解するであろう。収集されたアノテーションデータは、Luceneなどの検索エンジンのインデックス、又はリレーショナルデータベースなどのいくつかの異なる方法で記憶されることができる。イベントは、リアルタイムで、又は所定の間隔で若しくは検索クエリなどのユーザからの動作に応じて実行されるパッチモードで収集されることができる。厳密な方法は特定の実装の要件に基づいて決定されることができ、その方法の選択は本発明の基本的な意図を変更しない。特定のイベントソースの種類が、ある状況においてその他の種類よりも優れたランク付け結果をもたらす可能性がある。例として、実装に応じて、「記憶」ダイアログからのイベントがその他のイベントよりもユーザの興味の優れた指示をもたらす可能性がある。したがって、実施形態は、必要に応じてそのようなソースの種類に基づいた別個のランク付け及びその他の計算を可能にする収集データ構造を有することができる。最終的なランクが、異なるイベントソースの種類からのランクの収集に基づいて計算されることができる。

10

【0096】

[検索]

ユーザは、キーワードをクライアントシステム110の検索モジュール114に送信すること(ステップ500)によって検索を開始することができる。これはいくつかの態様で行われることができる。例として、図5、6、及び7のツールバー内の検索フィールド。ユーザは、図10のような専用アノテーションブラウザウィンドウを起動し、検索フィールドに入力することができる。これを実装する態様がクエリ文字列を検索モジュール114に伝達する限りこれを実装する多くの態様が存在する可能性がある。概して、検索はキーワードの形態であり、ウェブ上の検索エンジンにおいてよくある検索と同じフォーマットに従う。クエリは上述のように実質的にコンテキストを示す。

20

【0097】

サーバの検索応答モジュール123又は133は、マッチするアイテム又はユーザ(ヒット)を判定し、そのようなヒットの関連性(ランク付け)を判定する責任を負う。本発明の根幹をなす革新は、アノテーションに基づいてヒットのランク付けにおいて情報検索技術を使用する能力の認識である。これは、(Amit Singhalによる「Modern Information Retrieval: A brief overview」に記載されているような)従来のTF-IDFスタイルのアプローチ及び(Borodinらによる「Link Analysis Ranking Algorithm, Theory, and Experiments」に記載されているような)LARスタイルのアプローチを含む。クエリコンテキストがカテゴリコンテキストに対応する場合、LARスタイルのアプローチがランク付けの好ましい形態である。上述のように、それぞれのアノテーションをユーザとアイテムの間の統合的なリンクとして使用することによってLARスタイルのアプローチを取り入れることができる。概して、ユーザ及びアイテムの両方は、ユーザからアイテムに向かう有向リンクを有するグラフのノードとみなされることができる。より具体的には、これは、ユーザがハブのように扱われ、アイテムがリンク解析アルゴリズムにおけるオーソリティとして扱われることを可能にする。

30

40

【0098】

好ましい実施形態は、プライバシーに関する懸念のためにクエリを行うユーザがキーワードに基づいてユーザを検索することを許さず、その代わりに発行/サブスクライブ方法を生成する。基本的に、好ましい実施形態はユーザ情報を検索結果として返すことを許さないが、個人が(サーバにおいて行われたユーザに基づく検索によって決定される)関連性のある人にそれらの人がそれらの人の私的情報を明かす必要なしにメッセージを送信し、それらの人に送信者に折り返し連絡するか否かの選択を与えることを可能にする。これ

50

はカテゴリコンテキストに制限される。

【0099】

好ましい実施形態において、カテゴリコンテキストはHITSアルゴリズムに基づいてランク付けされる504。説明されたLARアルゴリズムなどの任意のLARアルゴリズムが使用されることができる。この分野は広範に研究されており、様々な欠陥を対象とする多くの変形が存在する。好ましい実施形態は、カテゴリコンテキストの一部であるイベントを初期集合として使用する。上述のように、これらのイベントは、コンテキストのすべてのキーワードが存在し、それらのキーワードがアイテムに関する定義キーワードに対応するように選択される。これは、HITSアルゴリズムに通常関連するトピックドリフト問題を解決するために行われる。これらのイベントはユーザとアイテムの間の統合的なリンクを生成するために使用され、HITSアルゴリズムがそれに適用される。これは、ハブのランクの形態のユーザに関するランクと、オーソリティのランクの形態のアイテムに関するランクとの両方を与える。これらのランクは、カテゴリコンテキストに対して生成時(ステップ408)に計算され、イベントが収集されるときに更新され続ける(ステップ409)。カテゴリコンテキストに対するクエリに基づくユーザ及びアイテムのランク付けがこれらのランクを使用することによって有利に行われることができる。ユーザのランク付けは、アイテムのランク付けとは異なるアルゴリズムを用いて行われることができる。例として、アイテムはHITSアルゴリズムを用いてランク付けされること
10
20

【0100】

カテゴリコンテキストではないコンテキスト、好ましい実施形態は、イベントのアイテムに基づくマッピングにおいて正規化されたイベントに基づいて単純なTF-IDFに基づくランク付けを利用する(ステップ502)。ユーザの検索を可能にするその他の実施形態は、ユーザ毎の収集されたキーワードに基づいてヒットを生成することができる。コンテキストのキーワードが収集されたキーワード内に存在する場合、ユーザはヒットとして返されることができる(ステップ505)。そのようなメカニズムは、ランク付けのためのTF-IDFスタイルのメカニズムを使用することができる(ステップ506)か、IRからの適用可能なその他の方法のうちのいずれかを使用することができる。

【0101】

高速なクエリ応答を容易にするために、ユーザ及びアイテムの両方に関するそのようなヒット及びランク付け情報は、Luceneなどの従来の検索エンジンの転置インデックスに記憶されることができるか、又はOracleなどのリレーショナルデータベース内に記憶されることができる。

【0102】

その他の実施形態において、カテゴリコンテキスト内の発行者IDをアイテムと同様の態様でランク付けすることもできる。各アイテムは、場合によってはいくつかの発行者IDを持つ可能性がある。これらのIDのそれぞれは、アイテムと同様の態様で統合的なリンクを用いてユーザに関連付けられることができる。所与の発行者IDに関連するユーザがすべてのアイテムに渡って収集される。ユーザはハブとしてモデル化され、発行者IDはオーソリティとしてモデル化される。アイテムをランク付けするための同じアルゴリズムが発行者IDをランク付けするために使用されることができる。これらのランクは、発行者がコンテキスト内のユーザの間で有する権威のレベルを示す。ランク付けはカテゴリコンテキストではないコンテキストにおいても可能であるが、データのまばらさは、TF-IDFスタイルのアプローチ、CubeSVD、LSI、又はPLSA及びPHITSなどのテンソル分解アプローチなどを優れた代替にする。アイテムのランク付けはそのアイテムの発行者IDのランク付けに部分的に基づくこともでき、そのようなランクはクエリに対するアイテムの最終的なランクの計算で収集されることができる。

【0103】

ランク付けに関して、ユーザとアイテムの間に基本的な違いが存在する。概して、アイ
40
50

テムは、アイテムの定義特徴の決定を可能にするべき乗則の挙動を示す。これは、TF - IDFスタイルのアプローチが収集され正規化されたイベントにおけるキーワードの数を利用することができるのでTF - IDFスタイルのアプローチを関連性の弁別において効果的にする。その一方、ユーザは、アイテムよりも多くの側面を有する傾向があり、そのような側面は時間の経過と共に変わる。LARスタイルのアプローチ、又はCubeSVDなどのテンソル分解方法は、所与の実装においてTF - IDFアプローチよりも優れた代替を生み出す可能性がある。当業者が気付くであろうように、IR技術を適用することにおいて可能な多くの変更形態が存在する。異なる実施形態は、それらの要件の通りにランク付けのための異なるIR技術を実装することを選択することができる。これはこのメカニズムの基本的な意図から逸脱しない。

10

【0104】

また、検索の結果はさらなるドリルダウンカテゴリーを返す(ステップ502、504、及び506)。実質的に、これらは、クエリをさらに増強するための関連性のあるキーワードの提案である。好ましい実施形態は、カテゴリーコンテキストからドリルダウンカテゴリーを計算する。具体的には、これは、クエリコンテキストの子であるすべてのカテゴリーコンテキストに関して、クエリコンテキストのキーワードを削除し、次のレベルの一意なキーワードを生成し、それらのキーワードの累積のイベント数によってそれらのキーワードをソートし、上位(例えば、20)のキーワードを返すことを意味する。特定の実施形態は、コンテキストに対応するアイテムのキーワードからこれらを計算することができる。その他の実施形態は、検索のクリックストリームを計算のための基礎として使用することができる。一部の実施形態は、所与の期間内のイベント数の累積数に対応する「最近の」イベント数を使用することを好む可能性がある。上述のように、一部の実施形態は、ドリルダウンカテゴリーのよりパーソナライズされた集合を得るためにカテゴリーコンテキストのユーザの特定期間の使用頻度に基づく時間に基づくTF - IDFアプローチも使用することができる。

20

【0105】

ランク付けに対する多くのその他のアプローチが上記のアプローチと協調して使用されることができる。例として、テキストコンテンツに関して、アノテーションに基づくランク付けを増強するために全文インデックス付けが使用されることができる。ウェブページに関して、ハイパーリンクの接続性が従来のLARアプローチによって利用されることができる。アイテム又はユーザに関するランク付けのすべてのこれらの異なるソースは、実装の要件に基づいて選択されることができる当技術分野に知られているいくつかのランク収集アルゴリズムを用いて有利に収集されることができる。

30

【0106】

次に、結果及びカテゴリーが検索応答モジュール133からクライアントシステム110に返され、次にディスプレイモジュール113を使用してユーザに表示される。これらの結果は、図10のアノテーションブラウザなどの専用のウィンドウに表示されることができる。クエリされたコンテキストにおけるユーザ及びアイテムの総数などのその他のデータが検索結果において増強されることができる。一部の実施形態は、クエリ指定に使用されることもできる、ヒットに対応するイベントソースの種類又はイベントソースのリストを提供することができる(基本的にそれらをカテゴリーコンテキストのように扱う)。

40

【0107】

[パーソナライズ、サブスクリプション、及び発行]

すべてのイベントに基づいて検索を行い、検索結果を返すことの望ましくない結果は、べき乗則が、コンテキストに対する上位にランクされるヒットが少数のアイテムなどによって満たされはじめ、新しく入ってくるものが見られることが次第に難しくなる「リッチゲットリッチャー(rich get richer)」減少を示すことである。これは、システム全体の有用性にとって有害であり、本質的にインタラクションの生態系を少数の参加者だけに集中させる。これは、時間に基づくアプローチを使用することによって幾分改善されることができる。例として、上述のように収集全体とは別個の(ここ1時間、

50

又は今日、又は今週などの) 期間に対してイベントが収集されることができる。これは、最近のイベントが公開されやすくなることを可能にする。

【0108】

図10は、そのような時間に基づく結果を「最新」タブに表示することができるアノテーションブラウザを示す。表示の観点からすると、クライアントシステム110は、結果のそのような時間に基づく変化を処理することを可能にされる必要がある。概して、電子メールのような現在のほとんどの時間に基づくシステムは、それらが受信される時間によってメッセージをソートする。これは、このシステムにおいてよくあるように多数のそのようなメッセージが存在する場合には不可能である可能性がある。したがって、「最新」タブは、関連性に基づいてアイテムを表示する必要がある。これは、ユーザが(電子メールのようにリストを順次下に移動するなど)それらのユーザが見たものを追跡し続け、それらのユーザが見ていないものを効率的に発見することがもはやできないので、基本的なユーザインターフェースのメタファの変更を要求する。これは、ユーザによって既に見られたアイテムが「マイアイテム」と呼ばれる別個のリストに保持されることができるメタファを実装することによって行われることができる。これは、「最新」タブにある間にコンポボックス内の「マイアイテム」メニューアイテムを選択することによってアクセスされることができる。これは、ユーザが「最新」によって与えられる期間内に見る、アノテーションする、発行するなどしたすべてのアイテムを表示する。また、「マイアイテム」は、「すべて」タブを選択するときと同様の意味を有するが、期間にかかわらずすべてのアイテムを含む。

10

20

【0109】

そのような機能は、真に新しいアイテムと、既存のアイテムに関する新しいイベントとを区別する必要がある。人気のあるアイテムは、新しいユーザがそれらを見つけたときに継続的にアノテーションされる。収集モジュール132は、アイテムがシステムにとって新しいのか、それともアイテムがコンテキストに対して新しいのかを評価することによってイベントがそのような時間に基づく記憶に置かれるべきかどうかを検出することができる(ステップ402)、その場合はアイテムが追加され、そうでない場合アイテムは追加されない。代替として、一部の実施形態は、イベントが所与の期間アイテムに関して受信されなかったか、又はアイテムに対するイベントのレートが規定のレート未満に落ちた場合にアイテムに関するイベントを新しいイベントとしてシグナリングすることができる。

30

【0110】

これは、デフォルトの「最新」タブウィンドウが常に最新の及び最も関連性のあるアイテムを表示することに集中することを可能にする。これは、任意の時間の関連性によって継続的にソートされることができる。この表示メタファは、関連性が高いアイテムが受信されたときのデスクトップ上の新しいアイテムのチッカーテープ又はシステムトレイ上のアラートメッセージなどのその他のパラダイムによって増強されることができる。

【0111】

この時間に基づくアプローチは、新しい及び関連性のあるコンテンツの発見しやすさを高めるためにコンテンツの的を絞った配信を可能にする発行及びサブスクライブメカニズムを用いて有利に増強される。これは、以下のメカニズムを利用することによって達成される。

40

- ・ カテゴリーコンテキストを使用することによって発行者及びサブスクライバが集まることを可能にすること

- ・ 発行者がアイテムに対して最も関連性のあるコンテキストを選択することを可能にすることによる

- ・ 発行をアカウントビリティを伴う明示的な行為にすることによる

- ・ 分散された態様のコミュニティによる迅速な処理を可能にするコンテキスト設定フェーズを持つことによる

- ・ 発行者に関する評判の分散された形態を可能にするパーソナライズされたサブスクリプションプロセスを持つことによる

50

【 0 1 1 2 】

アイテムは発行を通じてシステムに導入されることができる。発行において、発行者は、図9のようにカテゴリコンテキストにアイテムを割り当て、次にそのアイテムをシステムに発行する。本来、発行はアノテーションの形態である。好ましい実施形態において、発行はアノテーションとは別個の明示的な行為に変換される。これは、図5、6、7のように「発行」ボタンを押すことによって遂行されることができる。発行者は、ユーザID及びパスワードを用いてシステムに認証する(ステップ600)必要がある。証明されたとき、システムは、発行者に発行者のすべての発行されるアイテムと共に使用される一意な発行者IDを割り当てる。これは、実装の要件に応じて発行者のユーザIDと同じであっても又は同じでなくてもよい。次に、発行者は、アイテムをそれらの発行者が当該アイテムに最も関連性があると感じる一意なカテゴリコンテキストに割り当てる(ステップ601)。これは発行者の判断によって決定されるが、発行者が意図されるカテゴリコンテキストのアイテム並びに総ユーザ数及び総アイテム数を見ることができるという事実によって支援される。意図されるカテゴリコンテキストがアイテムよりも比較的かなり多くのユーザを有する場合、それはアイテムにかなりの興味が存在する場合のコンテキストを示す可能性があり、アイテムが関連性がある場合、それは受け入れられるより大きな機会を有する。コンテキストがユーザと比べて比較的多くのアイテムを有する場合、アイテムがコンテキスト内のユーザの関心を引くためにその他のアイテムと競争する必要がある可能性があり、発行者がその他のアイテムと比較された当該アイテムの相対的な有用性のそれらの発行者の判断に応じてそのコンテキストが又は別のコンテキストを使用するように判断する可能性がある。

10

20

【 0 1 1 3 】

いったん発行者が(例えば、図9の発行ボタンを押すことによって)アイテムを発行すると、アイテムは発行イベントの形態でサーバシステム(120又は130)に送信される。そのような発行イベントの例が図16にXMLフォーマットで示される。発行イベントは、発行イベントがイベントの発行者の一意な発行者IDを必ず含むことを除いてアノテーションイベントと同様である。既存のアイテムは、アイテムをシステムに最初に導入した元の発行者だけでなく任意の発行者によって発行されることができる。これは、単にアイテムに新しい発行者IDを付け加えることと等価である。それぞれのそのような発行イベントは、正規化とアイテム及びコンテキストのマッピングとに関してその他のアノテーションイベントと同様の態様で処理される(ステップ602)。発行者IDは、これがキーワードの記述を歪曲するのでキーワードに関する正規化の計算において使用されない。しかし、発行者IDはアイテムに関するメタデータであり、すべてのそのようなアノテーションに渡ってアイテムレベルで収集されることができる。これらのIDは、アイテムに関する定義キーワードを決定することには含まれないが、ユーザがアイテムを発行者に関するそれらのユーザの熟知度に基づいて再ランク付けすることを可能にするために検索からの結果に含まれることができる(ステップ502、504、及び506)。発行者IDは、それらのサブスクリバによる発行者に関する評判(又はアカウントビリティ)の分散された形態を可能にする。

30

【 0 1 1 4 】

アイテムがコンテキストにとって新しいか、又はサーバシステム(120又は130)にとって新しい場合、好ましい実施形態はアイテムにコンテキスト設定するように試みる(ステップ603)。これは、カテゴリコンテキスト内のユーザのサブセットである可能性があるユーザの集合にアイテムがプッシュされるプロセスである。そのようなユーザは、コンテキストに関する上位ユーザを決定するためのランク付け方法を使用することによってサーバシステムによって決定されることができるか、又はコンテキストに対して影響力のある発行者を含むことができるか、又はコンテキストのユーザの無作為のサブセットであるか、又はコンテキスト内のすべてのユーザにアイテムを送信することを含む、実装の要件に応じたその他の態様による。そのような特定のユーザに関するプッシュメカニズムは、サブスクリバプロセスを通じてサーバによって実行されることができる。基本

40

50

的に、サブスクライバは時々サーバからアイテムを引き出す。サーバは、サブスクライバのユーザIDを使用して、ユーザに対する結果にコンテキスト設定のためのアイテムを追加するか否かを決定する。アイテムがコンテキスト設定されるまで、そのアイテムはユーザが通常のサブスクリプションプロセスの一部としてダウンロードするために利用できない。コンテキスト設定の主要な目的は、アイテムがコンテキストのコミュニティによってより速やかに利用されることができるよう（検索及びサブスクリプションに関するより高いランク付け）迅速な態様で人の小さいが代表的なグループがそれらの人が関連性があると感じたアイテムを「記憶」ボタンを使用してアノテーションするか又は「発行」を使用して発行することを可能にし、アイテムがそのアイテムの定義キーワード及びその他の人が当該アイテムを検索するために使用することができるその他のそのようなキーワードの両方を決定することを可能にすることである。このプロセスは、影響力のある発行者がコンテキストに関するユーザのサブセットに含まれる場合に促進されることができる。当業者は、コンテキスト設定プロセスが受け入れを迅速にするためだけにあり、高トラフィックのコンテキストにおいて有用であるが、必須の条件ではないことを認めるであろう。このステップは、コンテキストが低いレベルのトラフィック、又はこのプロセスのオーバーヘッドが相応の価値をもたらさないその他の状況にある場合は省略されることができる。

【0115】

検索に関する1つの問題は、ユーザがそれらのユーザが興味があるコンテキストを指定して結果を取り出す必要があることである。これは、ユーザが関連性のあるアイテムが利用可能になるときにそれらのアイテムを発見するための効率的な態様ではない。（上で説明されたように）好ましい実施形態は、サブスクリプションプロセス114を使用して自動化された態様で関連性のあるアイテムを取り出し、それらを図10の「最新」タブに表示し113、それらを関連性の降順で示し、関連性のあるカテゴリコンテキストをドリルダウンカテゴリとして示す。そのようなコンテキストは、「最も読まれているコンテキスト」、「最も最近のコンテキスト」などの広く有用なコンテキスト、及び関心のあるユーザのプロファイルに基づいたコンテキストを含むことができる。サブスクリプションプロセスは、カテゴリコンテキストに基づいてユーザプロファイルを生成する。これは、明示的及び暗黙的の両方で行われることができる。ユーザは、それらのユーザが興味があるカテゴリコンテキストを繰り返しクエリの形態で明示的に指定することができ、その結果、それらのカテゴリコンテキストからのアイテムがバックグラウンドで継続的にダウンロードされる。また、これは、（クリックすることなどによる）各カテゴリコンテキストに関するアノテーションイベントのユーザの相対的頻度を観測し、その割合でアイテムを取り出すことによって暗黙的に行われる。好ましい実施形態は、すべてのイベントソースの種類からのアノテーションイベントを使用してこれらの割合を計算する。その他の実施形態において、異なるイベントソースの種類からのアノテーションイベントは実装の要件の通りに異なるように重み付けされることができる。例えば、「記憶」ダイアログからのアノテーションイベントは、クリックストリームからのアノテーションイベントよりもユーザの興味をよりはっきり示すとみなされることができ、したがってより重く重み付けされることができる。

【0116】

ユーザのプライバシーを守るために、好ましい実施形態はクライアントシステム110にこのプロファイルを記憶し、その結果、ユーザはそれらユーザのプロファイルの完全な制御を握り、それらのユーザの望むようにそのプロファイルを見る又は編集することができる。クライアントシステム110はカテゴリコンテキストに基づいて匿名でアイテムを取り出すことができるが、コンテキスト設定プロセスが機能するために、サブスクリプションプロセス114は、ユーザに関するプロファイルが取り出されるためにユーザID/パスワードに基づくログインプロセスを用いてシステムに認証する（ステップ610）ようにユーザに要求する。一部の実施形態は、的を絞った広告の配信などの、コンテキスト設定とは別のその他のプロセスに対してこの形態のプッシュプロセスを使用することができる。

10

20

30

40

50

【 0 1 1 7 】

「最新」タブ内のドリルダウンカテゴリーの処理は、「すべて」タブ内のアイテムの集合全体に渡る通常の検索のための処理とは異なり、イベントのために使用される期間と、カテゴリーコンテキストのユーザの特定期間の使用頻度に基づく時間に基づくTF-IDFとの両方に対して調整される。基本的に、そのようなドリルダウンカテゴリーのランク付けは、期間の累積のイベント数、並びにユーザの使用及び使用の最新性を反映する。これは、ユーザが最近の「話題になっている」及び関連性のあるトピックを容易に見つけるようにする。

【 0 1 1 8 】

ユーザに提示されるアイテムの関連性を高めるために、サブスクリプションプロセス 1 1 4 は、ユーザに表示する 1 1 3 前にアイテムをパーソナライズする。当技術分野において知られているパーソナライズに対する多くのアプローチが存在するが、そのようなアプローチの有効性はまだよく理解されていない。好ましい実施形態は、コンテキストに関してサーバシステム (1 2 0 又は 1 3 0) において上位にランク付けされたアイテムを取り出し、次にユーザプロファイルに基づいてクライアントシステム 1 1 0 においてアイテムを再ランク付けするアプローチを取る。これは、向上されたプライバシー及びセキュリティなどの有利に利用されるいくつかの特徴を有し、コラボレーションによるランク付け及びコンテンツに基づくランク付けの両方 (それぞれサーバ及びクライアントベースのランク付け) を利用し、ネットワークの末端の計算機能力を使用する。アイテムはユーザプロファイルに基づいてクライアントにダウンロードされる (ステップ 6 1 1) 。これは、1 日のような所与の期間にユーザが通常読む / 使用するアイテム数をサンプリングし、次にその数の好適な倍数を取得し、ユーザのプロファイル内のコンテキスト中にそれを分配することによって行われることができる。例として、ユーザが 1 日に 1 5 0 個のアイテムを読み、すべてのそのような読んだもののうちの 1 0 % がコンテキスト「 Ajax 」内にある。サブスクリプションシステムは、アイテムのうちの 1 0 % がキーワード「 Ajax 」に対応するか、又は利用可能なだけ多くのいずれか少ない方であるようにその日中に 1 5 0 0 0 個のアイテムをダウンロードすることができる。これらの割合は、ユーザがシステムとインタラクションするときリアルタイムで、又は 1 日に 1 回など所定の間隔の後にバッチ的に再計算されることである。倍数は、 (アイテムに対する関連性のコミュニティの意見の代理である) サーバにおけるランク付けに基づく順序と、 (ユーザの興味に基づいて決定される) クライアントにおけるランク付けとの間の混合を可能にする。好適な倍数は、ユーザ毎若しくはコンテキスト毎に計算されることであるか、又はスライダコントロールのような視覚的なメタファを通じてインタラクティブな態様でユーザによって設定されることさえできる。

【 0 1 1 9 】

それぞれのそのようなコンテキストに関してサーバからダウンロードされたアイテムが、ユーザプロファイルに基づいて再ランク付けされる (ステップ 6 1 2) 。これは、コンテキストに関するユーザプロファイルのキーワードベクトルを各アイテムのキーワードベクトルと比較することによって遂行される。好ましい実施形態は、ユーザプロファイルからのコンテキストのキーワードベクトルと各アイテムのキーワードベクトルとを厳格な態様で決定する。ユーザに関するコンテキストの一部であるアノテーションイベント内のアイテムに関する定義キーワードの一部であるキーワードのみが、キーワードベクトルを計算するために使用される (そのようなイベントは、プロファイル内の、再ランク付けされるべきアイテムに基づかないすべてのイベントを含む) 。「マイアイテム」タブのクリックストリームに対応するアノテーションイベントは計算から除外される。このベクトルに対する重みは既に説明されたように時間に基づくTF-IDFの態様で計算され、そのコンテキスト内のユーザに関するキーワードの特定期間の使用頻度がベクトル内のキーワードに対する重みとして使用される。再ランク付けは、コンテキストに関するユーザプロファイルのキーワードベクトルを各アイテムのキーワードベクトルと比較することによって行われる。アイテムに対するキーワードの頻度が、アイテムに関する収集され正規化され

10

20

30

40

50

たイベントから判定される。次に、これは、対数に基づく減衰 $\log(N/d)$ を用いる従来のTF-IDFアプローチのようにドキュメントの頻度の逆数によって乗算され、ここでNはアイテムの総数に対応し、dはそのうちの当該キーワードを有するアイテム数に対応する。ランクは、アイテムに関するキーワードベクトルとそのコンテキスト内のユーザに関するキーワードベクトルとの内積に基づいて各アイテムに対して計算される。

【0120】

上記の計算は、キーワードと同様の態様で発行者IDを用いて増強される。それぞれのそのような発行者IDはキーワードベクトルに含まれることができ、内積によって生成される最終的なランクに影響を与える。発行者IDはキーワードと比べて比較的出現頻度が低いので、それらの発行者IDは最終的な重み付けに対して大きな影響を有する。発行者IDの重み付けの効果はユーザが発行者を有用だと思ったコンテキストに制限され、1つのコンテキスト内の高くランク付けされた発行者は別のコンテキスト内のアイテムのランク付けに影響を与えない可能性があることに留意することが重要である。再ランク付けは、これらの計算されたランクの減少する値に基づいてアイテムをソートすることに等しい。当技術分野に習熟した読者は気付くであろうように、すべての上記の条件は厳格な制限であり、実装に対する必要性の通りに多数の異なる態様で緩和されることができる。実装は、イベントのキーワードがアイテムに関する定義キーワードの一部であるための条件を緩和することができる。実装は、アイテムのすべての定義キーワードをユーザがそれらのユーザのアノテーションにそれらのキーワードを使用したかどうかにかかわらず使用することができる。実装は、それらの実装の必要に応じてランク付けに対する発行者IDの効果 10
20
を好適に小さくすることができるか、又はランク付けに発行者IDをまったく使用しない可能性がある。使用されることができるランク付けのようなTF-IDFの多数の変態形態が存在する。アイテムに関する生成及び消費レートが計算の基礎として使用されることができる。

【0121】

次にこれらは、上述のように図10の「最新」タブ上に表示される613。アノテーションブラウザに関して、「最新」タブにおける検索は、システム内のすべてのアイテムを検索することと対称的に期間に基づいたアイテムのみを検索する。これは、システムに対して知られているすべてのアイテムをカバーする「すべて」タブにおける検索とは異なる 30

【0122】

サブスクリプションの表示は、発行者IDを認識し、それらをサーバに送り返すためにも使用される。好ましい実施形態は、アイテム毎に許可される発行者ID数を10などの所定の数に制限し、IDをアイテムと共にその所定の数まで発行順に記憶する。サブスクリプションは、アイテムを取り出すときに各アイテムに関するすべての知られている発行者IDをダウンロードする。サブスクリプションプロセスはすべてのそのような発行者IDを記憶し(基本的に、ユーザプロファイル内のそれぞれのそのようなIDに関する使用を更新する)、発行者IDに基づく再ランク付けの計算にそれを使用する。これらが上述のように再ランク付けのためだけに使用されるのではなく、アイテムの元の発行者ID及び最もマッチする発行者IDもアイテムに関するユーザからのアノテーションイベントに 40
追加される。次に、これは上述のアノテーション収集方法を通じてサーバに返送される(ステップ614)。また、これはコンテキストに対してユーザプロファイルを更新する(ステップ615)。これは、発行者IDに対するフィードバックループを閉じ、それらが将来のコンテキスト設定及びその他の目的のためにサーバサイドでランク付けされることを可能にする。最もマッチする発行者IDは、アノテーションの功績を認められるように、ユーザによく知られている発行者を与える。元の発行者IDは、アイテムをシステムに導入した発行者が功績を認められるように含まれる。当業者はそのようなフィードバックループを実装するための多数の異なる方法が存在することを認識し、所与の実装の要件に対して有利に使用されることができる異なるシステムの特徴を与える。しかし、これらは、中央サーバに戻る発行者IDに関するフィードバックループを提供する基本的意図から 50

逸脱しない。特定の実施形態は、ユーザが高くランク付けされた発行者を検索することを可能にすることができる。その他の実施形態は、検索のヒットを評価することにアイテムに関する発行者のランク付けを含むことによってアイテムのランク付けを増強することができる。特定の実施形態は、上述のように発行者IDと同様の態様で所与のアイテムに関するユーザIDを使用することができる。特定の実装が、上記のようなサブスクリプションと同様の態様で検索結果に関する再ランク付け507を使用することができる。検索結果に関するドリルダウンカテゴリーも、期間の基準又はユーザの特定期間の使用頻度の基準に従ってランク付けされることができる。好ましい実施形態は、純粹にサーバにおける累積のイベントの合計に基づいた検索結果を保持し、それを使用するすべてのユーザに対して同じである。サブスクリプションアイテムだけが再ランク付けされる。これは、ユーザがそれらの個々のビューとは別のデータの純粹なグループのビューを見ることが可能であるように行われる。これは、すべてのユーザに渡り共有されるアイテムに対する少なくとも1つのビューを可能にする。

【0123】

アノテーションシステムは、任意の所与のカテゴリーコンテキストに関する多数のアイテムを受信することができる。クライアントへのカテゴリーコンテキストに基づいたアイテムを受信するサブスクリプションプロセスは、そのようなフローに対応することができない可能性がある。好ましい実施形態は、周期的にコンテキストの所定の数の最も関連性のあるアイテムを取り出す。これは、再ランク付けのためにクライアントにおいて利用可能でない多くのアイテムが存在する可能性があることを示唆する。しかし、任意の所与の瞬間、クライアントは最も関連性のあるアイテムを有する可能性が高い。その他の実施形態は、すべてのアイテム又はすべてのイベントさえもが取り出され、ある期間のアイテムのクライアントイメージがサーバイメージとの同期を維持されるなど、当該アプローチに対する変更形態を取る可能性がある。再ランク付けにおいて使用されるキーワード及び発行者IDの重要な統計がサーバからクライアントに取得される。これは、サブスクリプションのときに行われ、周期的に最新に保たれる。同様の方法が、アイテムに関する定義キーワードが変わるときにその定義キーワードを取得するために使用される。すべてのそのようなデータは、検索などのユーザによる情報の実際の要求に便乗する態様で取り出されることができるか、又は規則的間隔で保守される可能性がある。すべてのユーザプロフィールデータも規則的間隔でサーバにバックアップされることができ、及び/又はネットワークベースの記憶から利用できる可能性がある。これは、アノテーションサーバを管理するエンティティから独立したエンティティによって記憶されることができる。当業者は気付くであろうように、上記の基本機能を変更することなく上記発明が実装されることができる多くのシステム構成が存在する。サブスクリプションは、クライアントサーバアーキテクチャにおいてアノテーションサーバがクライアントと通信することを暗黙的に仮定する。しかし、アノテーションサーバの処理は、ロードバランシング、3層アーキテクチャ、RPC/ウェブサービスベースのアプローチ、ピアツーピアアプローチなどのいくつかの従来態様で分散されることができる。処理はアイテム及びコンテキストに基づいて行われるので、計算がハッシュに基づいて分散されることができる。アイテムの場合、各サーバは、ハッシュ関数の基礎において決定されるアイテムのサブセットのみを処理することができる。コンテキストベースのサーバの場合、処理はコンテキストに基づいてサーバに割り当てられることができる。RESTのようなアプローチが、性能を高速化するためにローカルキャッシュサーバを実装することができるようになることができる。処理は、コンテキストベースのサーバのための少なくとも1つの新規性のある態様で分散されることができる。DNSのようなアプローチが使用されることができ、処理はコンテキストに基づいていくつかのサーバに渡って階層的な態様で連携される。例として、コンテキスト「Programming」に関するイベントに対するすべての処理は、このコンテキストを専門に扱うサーバに中継されることができる。コンテキスト「JavaScript Programming」に関するイベントは「Programming」を専門に扱うサーバに送信されることができ、そこからコンテキスト「JavaScript

10

20

30

40

50

Programming」に基づくさらなるサーバに送信されることができる。

【0124】

好ましい実施形態は単に本発明の基本的概念を使用するシステムの例にすぎず、可能な、及び本発明の基本的な意図から逸脱しない多数の変更形態が存在する。好ましい実施形態は、検索エンジンにおいてよくあるような検索コンテキストに関するAND、OR、及びNOTなどのブール論理式を使用するように拡張されることができる。コラボレーションの範囲を設けるか、又は「Spam」若しくは「Adult」などのような特別な目的のキーワードを実装することによってコンテンツのフィルタリングを取り入れることができる。これは、ボタンの形態で、又はその他の好適な形態でユーザに提示されることができる。フィルタリングは、ユーザがそのようなキーワードのキーワード数に関する値を設定すること可能にすることによってクライアントシステムにおいて制御されることができ、サブスクリプション取りだしプロセス及び再ランク付けプロセスは、ユーザが指定した量を超えるこれらのキーワードに関するキーワード数によって任意のアイテムを除去することができる。このテーマに関する可能な変更形態は、そのようなキーワードがアイテムに関する定義キーワードである場合にアイテムを除去することである。実施形態が発行を既存のアイテムに関するアノテーションイベントの代わりに特別なアイテムの種類として扱い、それによってそれぞれのそのようなアノテーションが別個のアイテムIDを持つことを可能にすることができる。これは、アノテーションがそのアノテーションのメタデータ内で別のアノテーションを指すことを可能にし、そのようなアノテーションのチェーンの作成を可能にする。これは、対話のスレッドの作成を可能にし、アイテムがアノテーションサーバに対して別個に記憶され、発行イベントが単にその存在を知らせるための方法として働く説明された好ましい実施形態におけるフォーマットの代わりにアノテーションサーバ自体の中にこれらのメッセージが保存されることを可能にするフォーラムのような機能を可能にする。

10

20

【0125】

重大な新しい機能を達成するための小さな調整を用いて既存のソフトウェアアプリケーション内で使用されることができる本発明の実施形態が存在する。1つのそのような実施形態において、カテゴリコンテキストの概念が、それらのカテゴリコンテキストを現在のウェブ検索エンジンにおいてクリックストリームログから導出することによってそのような検索エンジンに有利に組み込まれることができる。これは、任意の検索エンジンに容易に追加されることができ、より大きな平均の数のキーワードを有するクエリを生成することにおいて重要な役割を担うことができる。これは、現在のパーソナライズのアプローチよりも関連性のある結果を達成することに対するより有効な態様である可能性がある。

30

【0126】

別の実施形態において、電子メールが、キーワードをメールアドレスとして受け入れる特別に設計されたメールサーバを使用することによってキーワードのようなコンテキストを用いてアノテーションされることができる。例として、キーワードKey1がKey1@specialserver.tldとして入力されることができる。Outlookのような既存の電子メールクライアントに対するアドインモジュールが、メールアドレスをオートコンプリートすることによってそのようなキーワードのシームレスな入力を可能にするインタラクションパラダイムをもたらすために修正されることができる。この電子メールが、To:、Cc:、及びBccを使用することによってそのようなアドレスに送信されることができる。これは、基本的なプロトコルを変更する必要なしに電子メールのカテゴリを可能にする。メールが転送又は返信される度に、そのようなアノテーションが行われ、送信者の電子メールIDがこのシステムのユーザIDなどとして使用され、すべてのそのようなアノテーションがサーバにおいて収集されることができる。そのとき、フロントエンドのアドインモジュールは、電子メールのカテゴリコンテキストへのカテゴリ、及び検索などの好ましい実施形態のその他の機能を提供することができる。別のあり得る方法は、ストップワードが取り除かれ、それから電子メールが本発明のメッセー

40

50

ジと同様に処理されることができるよう「題名」行のテキストを使用してキーワードを導出することである。

【0127】

発行者IDに対するTF-IDFの時間に基づく変形に基づいてアイテムを再ランク付けする概念は、今でさえ電子メールクライアントソフトウェアに有利に実装されることができる。発行者IDは、受信箱内の受信電子メールの送信者電子メールIDとなる。メールソフトウェアは、各送信者に関する特定期間の使用頻度を計算するために、どの送信者からのどの電子メールがユーザによって読まれるかを見るためにユーザを監視することができる。特定の送信者から受信された電子メール数は、送信者IDに関連するアイテム数の代用として働くことができる。再ランク付け機能が、関連性に基づいてユーザの受信箱をソートするために有利に実装されることができる。

10

【0128】

同様のアプローチが、好適に開発されたソフトウェアを用いてユーザの集団全体に対してコンテキストの影響を受ける態様で発見され得るブログのポスト、ポッドキャスト、及び任意のRSSフィードに基づくアイテムを作成するために有利に利用されることもできる。すべてのそのようなポストは発行イベントとして好適なアノテーションサーバに送信されることができ、RSSリーダソフトウェアが、本発明によって説明されたようにサブスクリプション及びアノテーションを可能にするために好適に修正されることができる。発行者IDは、RSS url から合成されることができ、又はアノテーションサーバにおいて特別なログイン手順によって割り当てられることができる。

20

【0129】

当技術分野に習熟した読者は、本発明及び説明された実施形態が、企業又はデスクトップのファイル、ウェブ上のブログ、及び本発明に関するその他の好適な用途に様々な形態で適用されることができると認めるであろう。インスタントメッセージ通信ソフトウェア及びチャットソフトウェアは、これを使用してコンテキストベースのリアルタイムのメッセージ通信を実装することができる。また、本発明は、的を絞った広告の配信に特定の関連性を有する。ユーザの検索は、広告がユーザに対して的を絞った態様でプッシュされることを可能にする。これは、プルに基づく現在のキーワードベースの広告モデルの代わりにプッシュベースの新しい形態の広告を提供するためにウェブ検索エンジンのプロバイダによって利用されることができる。サブスクリプションは現在のウェブ検索と比較してユーザがさらされるコンテキストの数を増加させるので、サブスクリプションは、対応するより多くの高い関連性を有する広告機会をもたらす。ユーザプロフィールはクライアントに保持されることができ、これは、コンテンツがテレビなどのように無料で与えられるコンテンツ配信の新しいモデルを可能にする可能性があるパーソナライズされた態様でユーザにストリーミングされるコマーシャルと共にTV又は映画/ビデオを提供することなど、広告配信の新規性のある形態で利用されることができる。この形態のコミュニケーションは、多対多コミュニケーションパラダイムがより低コストでより効率の良いそのような広告の的を絞った配信を可能にする分類された広告の配信に理想的に適する可能性がある。例として、中古カメラの販売人は、対応するカテゴリコンテキストの使用を通じて中古カメラの潜在的な買い手とコミュニケーションすることができる。

30

40

【0130】

本発明の実施形態は、同じ発明者による2つの特許出願、「System for semantically disambiguating text information」(米国特許出願第10/954964号及び国際出願PCT/SG2005/000321号)並びに「A method and system for organizing items」(国際出願PCT/SG2005/000320号)において提供された意味論メカニズムを有利に利用することができる。これらの従来出願の内容は参照により本願に援用される。出願番号US10/054064は、テキストが意味の一意な機械表現に変換されることを可能にするユーザインターフェース方法を開示する。したがって、blog、blogs、weblogなどのようなキーワードが、意味「we

50

blog」を表す単一のIDにマッピングされることができる。これは、本発明のアノテーション、検索、サブスクリプション、及びその他のメカニズムがより正確であることを可能にする。アイテムをそれらのアイテムをより発見されにくくする上記のキーワードによって3つの別個のコンテキストに分割させる代わりに、それらは容易な想起及び比較のために1つのコンテキストにグループ化される。本発明のメカニズムは任意の言語からのキーワードを認めることができるので、意味のそのような機械表現が言語をまたがる態様で実装され、同じ曖昧性除去を達成することもできる。出現番号PCT/SG2005/000320は、そのような意味メタデータが「related-To」関係を通じて制限された階層の形態で組織化されることを可能にする知識表現方法を開示する。これは、ちょうど出願番号US10/054064の方法がテキストの曖昧性を除去するようにコンテキストの曖昧性を除去するために有利に利用されることができる。例として、以下のコンテキスト{“JavaScript”, “Programming”}及び{“JavaScript”}は事実上アイテムの同じ集合を指す。基本的に、JavaScriptはプログラミング言語であるので、JavaScriptに関連するアイテムの多くはプログラミングにも関連する。したがって、キーワード「Programming」は、新しい情報又は区別能力をアイテムの集合に加えない。2つの別個のコンテキストを持つことによって、コンテキスト空間が断片化される。これは、「JavaScript」から「Programming」に向かう、当該特許出願に記載の「related-To」関係を持つことによって改善されることができる。これは、任意のアイテムに関して、「JavaScript」がキーワードである場合に「Programming」キーワードが存在すると仮定される可能性があることを示唆する。そのような意味関係を前もって計画させることによって、2つのコンテキストを同じコンテキストに曖昧性除去することができる。当該特許出願は、より直感的なユーザエクスペリエンスのためのドリルダウンキーワードに関するカテゴリコンテキストと共に有利に使用されることができる「ブラウザパスベヘビア(Browse Path Behaviour)」と呼ばれるメカニズムも記載する。本発明のメカニズムと有利に組み合わせることができる、両方の特許出願に開示されたようなこれらの発明のいくつかのその他の能力が存在する。

10

20

【0131】

本発明は、意味メタデータ又は概念、及びそれらの関係を生成するための重要な方法として働くこともできる。本発明におけるアイテムの定義キーワードが、候補概念を生成するために利用されることができる。例として、同じ意味に割り当てられることができるキーワードが、各アイテムに関する定義キーワード内の同様の単語を調べることによって(自動的に又は手動で)生成されることができる。同じ語幹形を有する異なるキーワードは、自動的に通常の意味にマッピングされることができる(同じアイテム内で使用される「blog」及び「blogs」などは同じことを意味する可能性が高い)。少しの手動介入を用いて、「weblog」が任意の所与のアイテムの定義キーワードで「blog」又は「blogs」と共に何度も使用される場合、「weblog」を同じ意味に関連付けることができる。同様に、異なる自然言語にまたがるキーワードが同じ意味に関連付けられることができる。より包括的な方法として、定義キーワード及びそれらの対応するアイテムの行列を取り、我々が共起、及び2次、3次、又はそれより高次の共起を含む語形間の相関を調べて「related-To」関係及び定義概念並びにそれらのキーワードの割り当てを生成することを可能にするLSIなどの相関解析を実行することができる。これは、IRにおけるその他のそのような試みとは大きく異なる。本発明のメカニズムによって生成される定義キーワードは、実世界のアイテムに関する単語の意味のコミュニティの解釈であり、実際の使用における意味を示す。これは、意味メタデータ及びそれらの関係を導出するためにLSIなどのパターン認識方法及びその他の相関方法が使用されることができる質の高いデータ集合を与える。カテゴリコンテキストも、意味メタデータを導出するためにそのような解析のために利用されることができる同様の質の高いデータ集合を表す。そのようなメタデータは、2つの特許出願によってそれらの機能のために要求される形態に変換されることができ、次に本発明の精度を向上するために使用される

30

40

50

ことができる。

【0132】

上記と同様に、これは、アノテータが「band = Beatles」などのキーワードの形態で関係を指定することを可能にし、定義キーワードを決定するのと同様の方法を使用して所与のアイテム / 概念の属性名の発見を可能にし、それを使用して自動的又は半自動的な態様でオントロジを生成することによってセマンティックウェブによって定義されるようなより豊富なオントロジを生成することに拡張されることができる。

【0133】

本発明の好適な実施形態は、組織内の処理機能を増強するための新しい方法として使用されることができる。例として、組織内で自身の机を移動したい人は、単に「移動要求」などのコンテキストにメッセージを発行することができ、移動要求を処理することに関わるすべての関係者がこのトピックにサブスクライブし、同時に通知されることができる。これらの関係者は、要求、設備、技術などのオーソライザを含むことができる。これらの関係者のそれぞれは、元の要求のコンテキストへのメッセージの連鎖を可能にするためにアイテムを元のアイテムIDと共に1つにして発行することができる。これらの発行されたアイテムは、要求を処理するための組織のワークフロー内のステップの完了に対応する可能性がある。例として、そのようなステップは、要求の認可、要求のその他のコンテキストへの通知、要求の終了、要求の別のコンテキストへの割り当てなどを含むことができる。これは、「移動要求」意味メタデータを定義し、それをアイテムの種類に割り当てるなど、発行されたアイテムを意味メタデータを用いて厳密に分類することによって強化されることができる。より広範な処理機能を容易にするために、このメカニズムは既存のBPMシステムとインターフェースを取られることができる。処理の自動化に関する重要な問題のうちの一つは、例外的な状況を扱うことが難しいことである。本発明のパラダイムなどの多対多コミュニケーションパラダイムは、そのようなインタラクションに対してより有機的で適応性のある構造をもたらす助けとなることができる。

【0134】

幅広く説明された本発明の範囲又は精神を逸脱することなく、特定の実施形態に示された本発明に対して多くの変更及び / 又は修正がなされ得ることが当業者によって理解されるであろう。したがって、本実施形態はすべての点において例示的とみなされるべきであり、限定的とみなされるべきではない。

【図面の簡単な説明】

【0135】

【図1】典型的なアノテーションの分布を示す図である。

【図2】べき乗則の分布を示すグラフである。

【図3】本発明による基本的なシステム設計を示す図である。

【図4】例示的なコンピューティング環境の図である。

【図5】ブラウザベースのアノテーションイベントジェネレータのためのユーザインターフェースの図である。

【図6】ファイルシステムベースのアノテーションイベントジェネレータのためのユーザインターフェースの図である。

【図7】電子メールソフトウェアベースのアノテーションイベントジェネレータのためのユーザインターフェースの図である。

【図8】例示的な記憶ダイアログの図である。

【図9】例示的な発行ダイアログの図である。

【図10】アノテーションブラウザのためのユーザインターフェースの図である。

【図11】アノテーションの収集のプロセスを示す流れ図である。

【図12】検索及びパーソナライズのプロセスを示す流れ図である。

【図13】発行及びサブスクライブのプロセスを示す流れ図である。

【図14】例示的な記憶アノテーションイベントの図である。

【図15】サブスクリプションクリックストリームアノテーションイベントの図である。

【 図 1 6 】 例示的な発行イベントの図である。

【 図 1 】

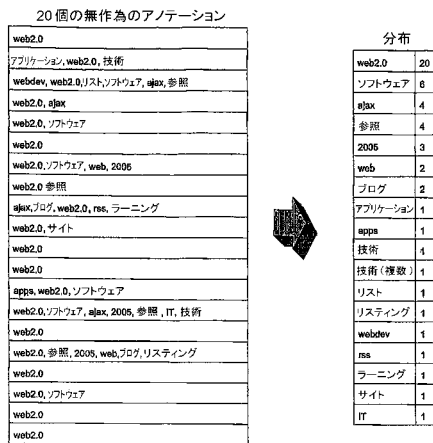


Fig 1: del.icio.us における http://web2.wsj2.com に関するアノテーション

【 図 2 】

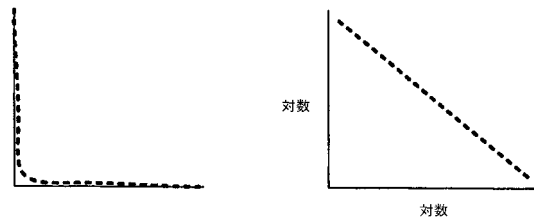


Fig 2: べき乗則の分布

【 図 3 】

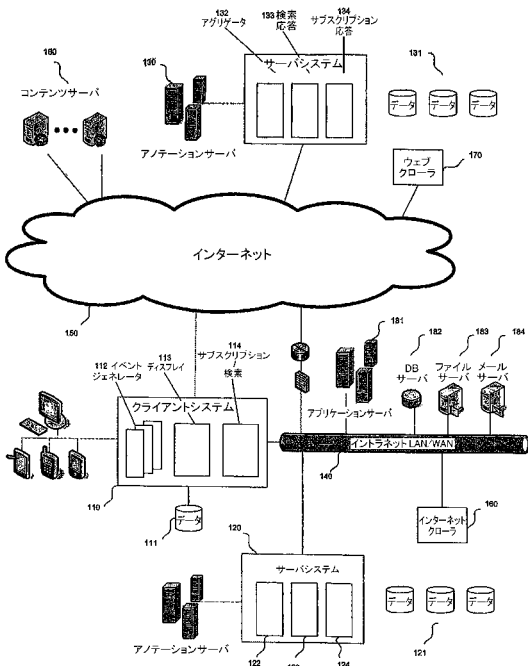


Fig3 基本的なシステムの図

【 図 4 】

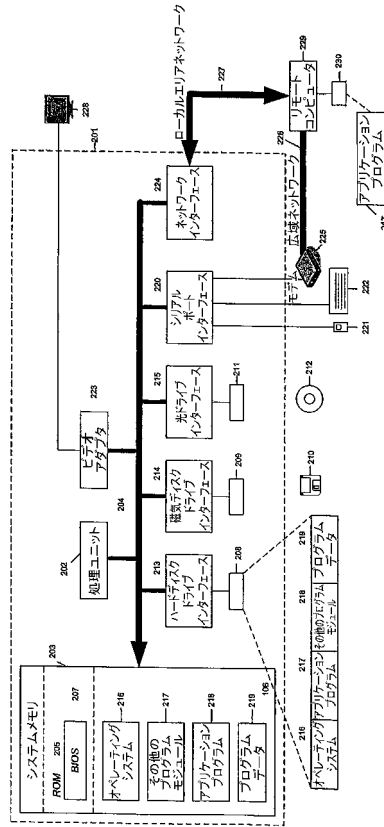


Figure 4 典型的なコンピューティング環境

【 図 5 】

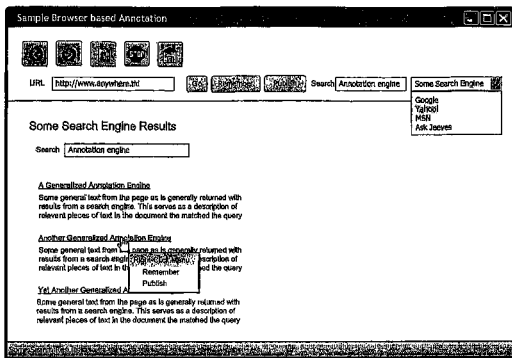


Fig5: ブラウザベースのアノテーションイベントジェネレータの例

【 図 6 】

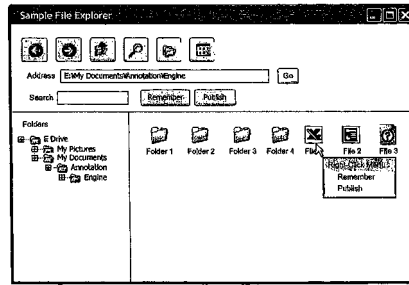


Fig6. ファイルシステムベースのアノテーションイベントジェネレータの例

【 図 7 】

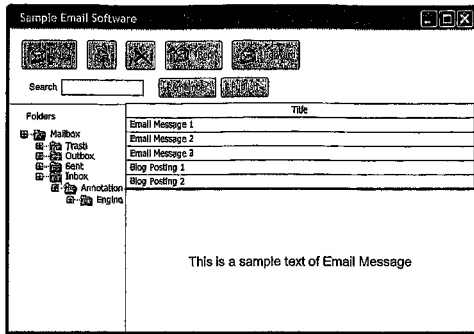


Fig7. 電子メールソフトウェアベースのアノテーションイベントジェネレータの例

【 図 8 】

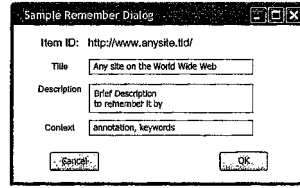


Fig8. 記憶ダイアログの例

【 図 9 】

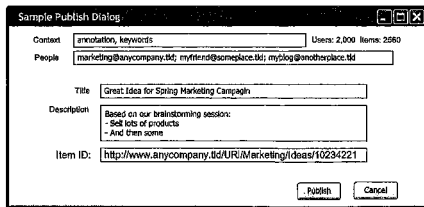


Fig9. 発行ダイアログの例

【 図 10 】

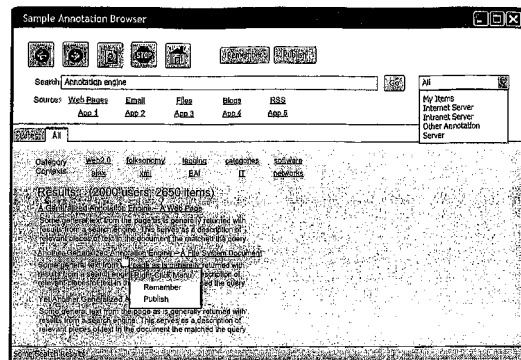
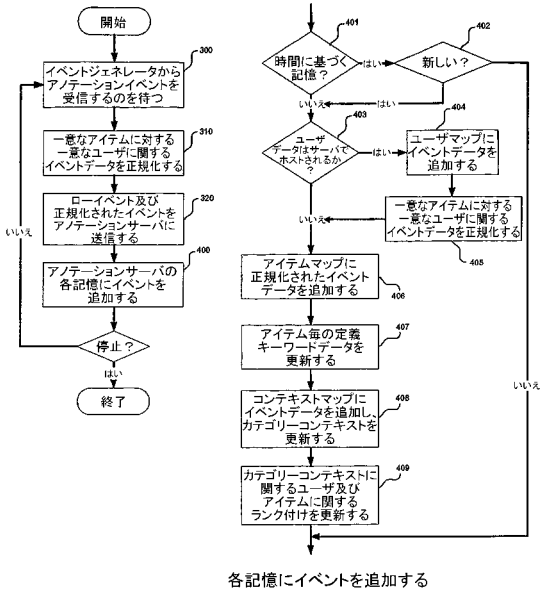


Fig10: アノテーションブラウザの例

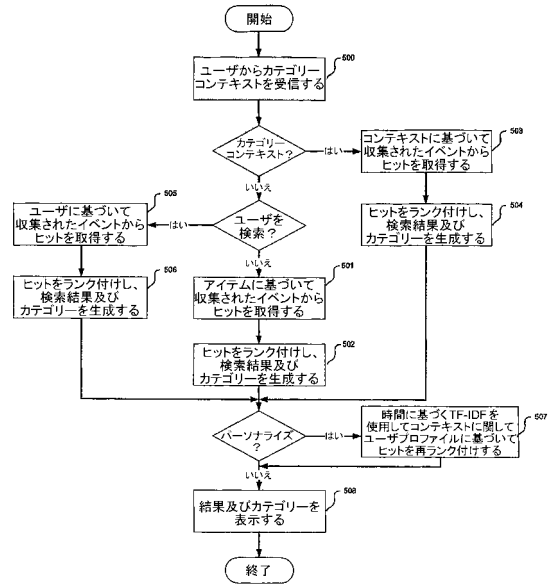
【 図 1 1 】

Figure 11 アノテーションの収集



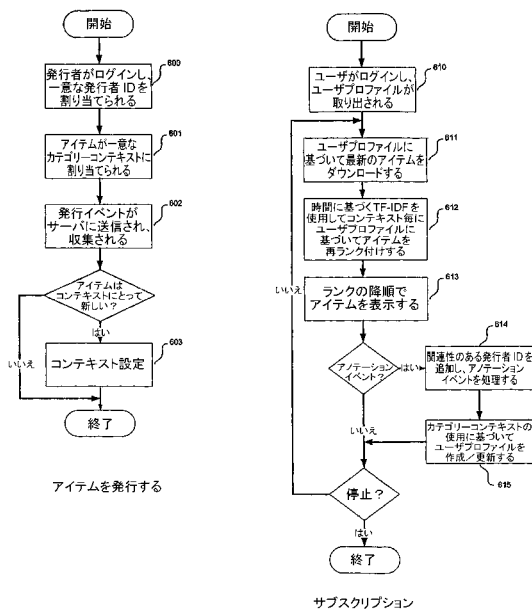
【 図 1 2 】

Figure 12: 検索及びパーソナライズ



【 図 1 3 】

Figure 13: 発行及びサブスクリプション



【 図 1 4 】

```

<AnnotationEvent>
  <EventSourceID>123456</EventSourceID>
  <EventSourceLabel>RememberDialog</EventSourceLabel>
  <UserID>http://annotationserver/UserID/JohnDoe</UserID>
  <Item>
    <ItemID>http://www.anysite.tid</ItemID>
    <Title>Any site on the World Wide Web</Title>
    <Description>Brief description to remember it by</Description>
  </Item>
  <OriginalPublisherID>
  <PublisherID>
  <Keyword>annotation</Keyword>
  <Keyword>keywords</Keyword>
</AnnotationEvent>

```

Fig 14: 記憶アノテーションイベントの例

【 図 15 】

```

<AnnotationEvent>
  <EventSourceID>234567</EventSourceID>
  <EventSourceLabel>ClickStream</EventSourceLabel>
  <UserID>http://annotationserver/UserID/JohnDoe</UserID>
  <Item>
    <ItemID>http://www.anysite.tld</ItemID>
    <Title/>
    <Description/>
  </Item>
  <OriginalPublisherID>Original</OriginalPublisherID>
  <PublisherID>Known</PublisherID>
  <Keyword>annotation</Keyword>
  <Keyword>keywords</Keyword>
</AnnotationEvent>

```

Fig15: サブスクリプションクリックストリームアノテーションイベントの例

【 図 16 】

```

<PublishingEvent>
  <EventSourceID>345678</EventSourceID>
  <EventSourceLabel>PublishDialog</EventSourceLabel>
  <UserID>http://annotationserver/UserID/JohnDoe</UserID>
  <Item>
    <ItemID>http://www.anycompany.tld/URI/Marketing/Ideas/10234221</ItemID>
    <Title>Great Ideas for Spring Marketing Campaign</Title>
    <Description>Based on our brainstorming session
      - Sell lots of products
      - And then some</Description>
  </Item>
  <OriginalPublisherID>JohnDoePublisherID</OriginalPublisherID>
  <PublisherID>JohnDoePublisherID</PublisherID>
  <Keyword>annotation</Keyword>
  <Keyword>keywords</Keyword>
</PublishingEvent>

```

Fig16: 発行イベントの例

【 国際調査報告 】

INTERNATIONAL SEARCH REPORT

International application No
PCT/SG2006/000280

A. CLASSIFICATION OF SUBJECT MATTER INV. G06F17/24 G06F17/30 G06Q10/00		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) G06Q G06F		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the International search (name of data base and, where practical, search terms used) EPO-Internal, WPI Data, INSPEC, IBM-TDB, COMPENDEX		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2005/081159 A1 (GUPTA ANOOP [US] ET AL) 14 April 2005 (2005-04-14)	1-18, 28-51, 64,65
Y	paragraphs [0034], [0051], [0052], [0089], [0124], [0134]; figures 3,4,6,7,15,16	19-27, 52-63
X	US 2003/179222 A1 (NOMA TSUNETAKE [JP] ET AL) 25 September 2003 (2003-09-25)	1-18, 28-51, 64,65
Y	paragraphs [0011], [0044], [0075], [0076]; figures 3,13	19-27, 52-63
X	US 2002/019827 A1 (SHIMAN LEON G [US] ET AL) 14 February 2002 (2002-02-14)	1-18, 28-51, 64,65
Y	paragraph [0110] - paragraph [0113]; figure 5	19-27, 52-63
	----- -/-	
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.		
* Special categories of cited documents : *A* document defining the general state of the art which is not considered to be of particular relevance *E* earlier document but published on or after the international filing date *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) *O* document referring to an oral disclosure, use, exhibition or other means *P* document published prior to the international filing date but later than the priority date claimed *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art. *Z* document member of the same patent family		
Date of the actual completion of the international search 27 November 2006		Date of mailing of the international search report 12/12/2006
Name and mailing address of the ISA/ European Patent Office, P.B. 6818 Patentlaan 2 NL - 2260 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016		Authorized officer Deane, Inigo

INTERNATIONAL SEARCH REPORT

International application No

PCT/SG2006/000280

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2004/267798 A1 (CHATTERJEE AVIJIT [US] ET AL) 30 December 2004 (2004-12-30) paragraph [0122] - paragraph [0124]	1, 64, 65
Y	JOON HO LEE ED - FOX E A ET AL ASSOCIATION FOR COMPUTING MACHINERY: "Combining Multiple Evidence from Differnet Properties of Weighting Schemes" SIGIR '95. PROCEEDINGS OF THE 18TH. ANNUAL INTERNATIONAL ACM SIGIR CONFERENCE ON RESEARCH AND DEVELOPMENT IN INFORMATION RETRIEVAL. SEATTLE, JUL. 9 - 13, 1995, SIGIR. PROCEEDINGS OF THE ANNUAL INTERNATIONAL ACM SIGIR CONFERENCE ON RESEARCH AND DEVELO, 9 July 1995 (1995-07-09), pages 180-188, XP002159143 ISBN: 0-89791-714-6 abstract	19-21
Y	DATABASE INSPEC THE INSTITUTION OF ELECTRICAL ENGINEERS, STEVENAGE, GB; Inspec No. AN8651126, 2005, DONATE, D. ET AL.: "Stability and similarity of link analysis ranking algorithms" XP002409186 abstract & PROC. AUTOMATA, LANGUAGES AND PROGRAMMING, 32ND INTERNATIONAL CONF., 11 July 2005 (2005-07-11), - 15 July 2005 (2005-07-15) pages 717-729, Lisbon, Portugal	22-27
Y	GOLDBERG D: "USING COLLABORATIVE FILTERING TO WEAVE AN INFORMATION TAPESTRY" COMMUNICATIONS OF THE ASSOCIATION FOR COMPUTING MACHINERY, ACM, NEW YORK, NY, US, vol. 35, no. 12, 1 December 1992 (1992-12-01), pages 61-70, XP000334368 ISSN: 0001-0782 page 63, left-hand column, line 48 - right-hand column, line 3	52-63

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/SG2006/000280

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2005081159	A1	US 6956593 B1	18-10-2005
US 2003179222	A1	NONE	
US 2002019827	A1	NONE	
US 2004267798	A1	NONE	

フロントページの続き

(81)指定国 AP(BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), EA(AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), EP(AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OA(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG), AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW

(特許庁注：以下のものは登録商標)

- 1 . L i n u x
- 2 . J A V A
- 3 . U N I X

(72)発明者 サーカー , デヴァジョティ
シンガポール , シンガポール 4 6 8 7 0 2 , ナンバー 0 3 - 0 1 ランドベイ コンドミニ
アム , ジャラン ハジジャ 7

Fターム(参考) 5B075 KK02 ND20 PR06 UU40