

(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2018-538603  
(P2018-538603A)

(43) 公表日 平成30年12月27日(2018.12.27)

(51) Int.Cl.	F I	テーマコード (参考)
<b>G06F 17/30 (2006.01)</b>	G06F 17/30	340B
	G06F 17/30	380E
	G06F 17/30	220Z
	G06F 17/30	120A

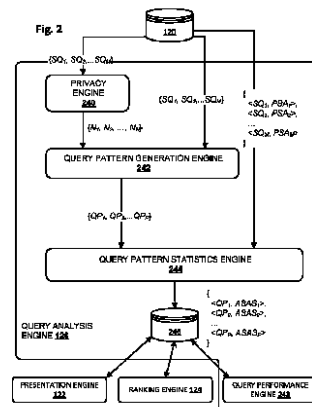
審査請求 有 予備審査請求 未請求 (全 22 頁)

(21) 出願番号	特願2018-521028 (P2018-521028)	(71) 出願人	502208397 グーグル エルエルシー アメリカ合衆国 カリフォルニア州 94043 マウンテン ビュー アンフィシ アター パークウェイ 1600
(86) (22) 出願日	平成28年11月21日 (2016.11.21)	(74) 代理人	100108453 弁理士 村山 靖彦
(85) 翻訳文提出日	平成30年6月7日 (2018.6.7)	(74) 代理人	100110364 弁理士 実広 信哉
(86) 国際出願番号	PCT/US2016/063175	(74) 代理人	100133400 弁理士 阿部 達彦
(87) 国際公開番号	W02017/091519	(72) 発明者	マイク・ベンダースキー アメリカ合衆国・カリフォルニア・94043・マウンテン・ビュー・アンフィシ アター・パークウェイ・1600
(87) 国際公開日	平成29年6月1日 (2017.6.1)		最終頁に続く
(31) 優先権主張番号	14/950,052		
(32) 優先日	平成27年11月24日 (2015.11.24)		
(33) 優先権主張国	米国 (US)		

(54) 【発明の名称】 検索クエリ間におけるクエリパターンおよび関連する総統計の特定

(57) 【要約】

様々な実施形態において、本明細書において説明する特徴は、検索クエリを処理する際に情報の機密性を維持するのに寄与し、計算コストを比較的安く抑えながらN-gramをプライベートまたは非プライベートとして分類するのを可能にする。プライベート検索クエリの検索ログ内に存在する複数の非プライベートN-gramに基づいて複数のクエリパターンが生成されてもよい。各クエリパターンに関連する総検索活動統計(「ASAS」)が検索ログから判定されてもよい。各クエリパターンに関連するASASは、クエリパターンと一致する検索ログ内の複数のプライベート検索クエリに関連する検索後活動を示してもよい。所与のクエリパターンに関するASASが成績基準を満たしているとの判定にตอบสนองして、この所与のクエリパターンと一致する検索クエリにตอบสนองして提示されるデータを生成するための方法が、この所与のクエリパターンに関連するASASに基づいて変更されてもよい。



## 【特許請求の範囲】

## 【請求項 1】

コンピュータによって実施される方法であって、

プライベート検索クエリおよび対応する検索後活動の検索ログ内において、前記プライベート検索クエリに含まれプライバシー基準を満たす複数の非プライベートN-gramを特定するステップと、

前記複数の非プライベートN-gramに基づいて前記検索ログから複数のクエリパターンを生成するステップと、

前記複数のクエリパターンの各々に関連する総検索活動統計を前記検索ログから判定するステップであって、各クエリパターンに関連する総検索活動統計が、前記クエリパターンと一致する前記検索ログ内の複数のプライベート検索クエリに関連する検索活動を示す、ステップと、

所与のクエリパターンに関する総検索活動統計が成績基準を満たしているとの判定に回答して、前記所与のクエリパターンに関連する総検索活動統計に少なくとも部分的に基づいて前記所与のクエリパターンと一致する検索クエリに回答して提示されるクエリサジェスションまたは検索結果をランク付けするステップとを含む、コンピュータによって実施される方法。

10

## 【請求項 2】

前記プライバシー基準は、頻度しきい値を満たす頻度で前記検索ログ内に出現するN-gramによって満たされる、請求項1に記載のコンピュータによって実施される方法。

20

## 【請求項 3】

前記プライバシー基準は、少なくとも最小回数だけ前記検索ログ内に出現するN-gramによって満たされる、請求項1に記載のコンピュータによって実施される方法。

## 【請求項 4】

前記生成するステップは、前記プライベート検索クエリ間において前記非プライベートN-gramがどれだけの頻度で出現するかに基づいて前記非プライベートN-gramのヒストグラムを生成するステップを含む、請求項1に記載のコンピュータによって実施される方法。

## 【請求項 5】

前記複数のクエリパターンの各々に関連する前記総検索活動統計は、前記クエリパターンと一致する前記検索ログ内のすべての検索クエリを含む、請求項1に記載のコンピュータによって実施される方法。

30

## 【請求項 6】

前記複数のクエリパターンの各々に関連する前記総検索活動統計は、前記クエリパターンと一致する前記検索ログ内のいくつかのそれぞれに異なる検索クエリを含む、請求項1に記載のコンピュータによって実施される方法。

## 【請求項 7】

前記所与のクエリパターンに関連する前記総検索活動統計は、前記所与のクエリパターンと一致する前記検索ログ内の複数のプライベート検索クエリに関連する検索後活動に関する総統計を含む、請求項1に記載のコンピュータによって実施される方法。

## 【請求項 8】

前記所与のクエリパターンに関連する前記総検索後活動統計は、前記所与のクエリパターンに関連するクリック位置統計を含む、請求項7に記載のコンピュータによって実施される方法。

40

## 【請求項 9】

前記所与のクエリパターンに関連する前記総検索後活動統計は、前記所与のクエリパターンに関連するクリック率を含む、請求項7に記載のコンピュータによって実施される方法。

## 【請求項 10】

前記所与のクエリパターンに関連する前記総検索後活動統計は、前記所与のクエリパターンに関連する離脱率を含む、請求項7に記載のコンピュータによって実施される方法。

50

## 【請求項 1 1】

前記成績基準は、前記離脱率が最小しきい値を満たす場合に満たされる、請求項10に記載のコンピュータによって実施される方法。

## 【請求項 1 2】

メモリと、前記メモリ内に記憶された命令を実行するように動作可能な1つまたは複数のプロセッサとを含むシステムであって、

プライベート検索クエリおよび対応する検索後活動の検索ログ内において、前記プライベート検索クエリに含まれプライバシー基準を満たす複数の非プライベートN-gramを特定するための命令と、

前記複数の非プライベートN-gramに基づいて前記検索ログから複数のクエリパターンを生成するための命令と、

前記複数のクエリパターンの各々に関連する総検索活動統計を前記検索ログから判定するための命令であって、各クエリパターンに関連する総検索活動統計が、前記クエリパターンと一致する前記検索ログ内の複数のプライベート検索クエリに関連する検索活動を示す、命令と、

所与のクエリパターンに関する総検索活動統計が成績基準を満たしているとの判定に回答して、前記所与のクエリパターンに関連する総検索活動統計に少なくとも部分的に基づいて前記所与のクエリパターンと一致する検索クエリに回答して提示されるクエリサジェスションまたは検索結果をランク付けするための命令とを含む、システム。

## 【請求項 1 3】

前記プライバシー基準は、頻度しきい値を満たす頻度で前記検索ログ内に出現するN-gram、または少なくとも最小回数だけ前記検索ログ内に出現するN-gramによって満たされる、請求項12に記載のシステム。

## 【請求項 1 4】

前記生成するための命令は、前記プライベート検索クエリ間において前記非プライベートN-gramがどれだけの頻度で出現するかに基づいて前記非プライベートN-gramのヒストグラムを生成するための命令を含む、請求項12に記載のシステム。

## 【請求項 1 5】

命令を含む少なくとも1つの非一時的コンピュータ可読記憶媒体であって、コンピューティングシステムによる前記命令の実行に回答して、前記コンピューティングシステムに

プライベート検索クエリおよび対応する検索後活動の検索ログ内において、前記プライベート検索クエリに含まれプライバシー基準を満たす非プライベートコンテンツを特定する動作と、

前記非プライベートコンテンツに基づいて前記検索ログから複数のクエリパターンを生成する動作と、

前記複数のクエリパターンの各々に関連する総検索活動統計を前記検索ログから判定する動作であって、各クエリパターンに関連する総検索活動統計が、前記クエリパターンと一致する前記検索ログ内の複数のプライベート検索クエリに関連する検索活動を示す、動作と、

所与のクエリパターンに関する総検索活動統計が成績基準を満たしているとの判定に回答して、前記所与のクエリパターンに関連する総検索活動統計に基づいて前記所与のクエリパターンと一致する検索クエリに回答して提示されるデータを生成する方法を変更する動作とを実行させる、非一時的コンピュータ可読記憶媒体。

## 【発明の詳細な説明】

## 【技術分野】

## 【0001】

検索クエリ間におけるクエリパターンおよび関連する総統計の特定に関する。

## 【背景技術】

## 【0002】

10

20

30

40

50

検索クエリの検索ログは、検索クエリ間において共通のクエリパターンを判定するうえで有用である場合がある。クエリパターンは、代替クエリサジェスションの生成および/またはランク付け、検索結果のランク付けなど、様々な目的に有用である場合がある。しかし、いくつかの検索クエリはプライベートユーザ情報を含む場合がある。このことは、ユーザおよび/またはユーザが関連するエンティティ(たとえば、ユーザの雇用者)に対してプライベートと見なされるシステムに検索クエリがサブMITされる状況に特に当てはまる。たとえば、eメールシステム、ファイルシステム、ドキュメント管理システムなどの多くのパーソナルシステムまたはエンティティ内システムが、ユーザによって、名前、連絡先情報、アカウント番号などのプライベート情報を含む可能性が比較的高い検索クエリを使用して検索される場合がある。

10

**【発明の概要】****【課題を解決するための手段】****【0003】**

本開示は主として、検索クエリを処理する際に情報の機密性を維持することに関する。本開示は概して、場合によってはプライベートな情報を人間のユーザに公開することなく(多くがプライベートでありおよび/または場合によってはプライベートな情報を含み得る)検索クエリおよび対応する検索後活動(Post-search Activity)の検索ログを分析して1つまたは複数のクエリパターンを生成するための方法、装置、およびコンピュータ可読媒体(一時的および非一時的コンピュータ可読媒体)を対象とする。1つまたは複数の生成されたクエリパターンについて、1つまたは複数のクエリパターンと一致する検索クエリに関連する総検索活動統計(Aggregate Search Activity Statistics)(検索後活動に関する統計を含む)が判定されてもよい。次いで、それらの1つまたは複数のクエリパターンおよび対応する総検索活動統計が様々な目的に使用されてもよく、それらの目的のいくつかについて以下に説明する。たとえば、所与のクエリパターンと一致する検索クエリをサブMITするユーザに提示されるデータを生成するのに使用される方法を変更するために、所与のクエリパターンに関連する総検索活動統計が使用されてもよい。

20

**【0004】**

いくつかの実装形態では、コンピュータによって実施される方法であって、プライベート検索クエリおよび対応する検索後活動の検索ログ内において、プライベート検索クエリに含まれプライバシー基準を満たす複数の非プライベートN-gramを特定するステップと、複数の非プライベートN-gramに基づいて検索ログから複数のクエリパターンを生成するステップと、複数のクエリパターンの各々に関連する総検索活動統計を検索ログから判定するステップであって、各クエリパターンに関連する総検索活動統計が、クエリパターンと一致する検索ログ内の複数のプライベート検索クエリに関連する検索活動を示す、ステップと、所与のクエリパターンに関する総検索活動統計が成績基準を満たしているとの判定に回答して、所与のクエリパターンに関連する総検索活動統計に基づいて所与のクエリパターンと一致する検索クエリに回答して提示されるデータを生成するための方法を変更するステップとを含む、コンピュータによって実施される方法が提供される場合がある。

30

**【0005】**

この方法および本明細書において開示する技術の他の実装形態の各々は、場合によっては以下の特徴のうちの1つまたは複数を含んでもよい。

40

**【0006】**

いくつかの実装形態では、この方法は、所与のクエリパターンと一致する検索クエリに回答して提示されるクエリサジェスションが、この所与のクエリパターンに関連する総検索活動統計に少なくとも部分的に基づいてランク付けされるように変更されてよい。

**【0007】**

このように、クエリサジェスションが機密情報を含む可能性の調節を実現することができる。

**【0008】**

いくつかの実装形態では、この方法は、所与のパターンと一致する検索クエリに回答し

50

て提示される検索結果が、この所与のクエリパターンに関連する総検索活動統計に少なくとも部分的に基づいてランク付けされるように変更されてもよい。

【0009】

このように、検索結果が機密情報を含む可能性の調節を実現することができる。

【0010】

いくつかの実装形態では、プライバシー基準は、頻度しきい値を満たす頻度で検索ログ内に出現するN-gramによって満たされてもよい。いくつかの実装形態では、プライバシー基準は、少なくとも最小回数だけ検索ログ内に出現するN-gramによって満たされてもよい。いくつかの実装形態では、生成するステップは、プライベート検索クエリ間において非プライベートN-gramがどれだけの頻繁で出現するかに基づいて非プライベートN-gramのヒストグラムを生成するステップを含んでもよい。

10

【0011】

これらの特徴は、計算コストを比較的安く抑えながらプライベートN-gramと非プライベートN-gramを分類するのを可能にする。

【0012】

いくつかの実装形態では、複数のクエリパターンの各々に関連する総検索活動統計は、このクエリパターンと一致する検索ログ内のすべての検索クエリを含んでもよい。いくつかの実装形態では、複数のクエリパターンの各々に関連する総検索活動統計は、このクエリパターンと一致する検索ログ内のいくつかのそれぞれに異なる検索クエリを含む。

【0013】

いくつかの実装形態では、所与のクエリパターンに関連する総検索活動統計は、この所与のクエリパターンと一致する検索ログ内の複数のプライベート検索クエリに関連する検索後活動に関する総統計を含んでもよい。いくつかの実装形態では、所与のクエリパターンに関連する総検索後活動統計は、この所与のクエリパターンに関連するクリック位置統計を含んでもよい。いくつかの実装形態では、所与のクエリパターンに関連する総検索後活動統計は、この所与のクエリパターンに関連するクリック率を含んでもよい。いくつかの実装形態では、所与のクエリパターンに関連する総検索後活動統計は、この所与のクエリパターンに関連する離脱率を含んでもよい。いくつかの実装形態では、成績基準は、離脱率が最小しきい値を満たす場合に満たされる。

20

【0014】

他の実装形態は、上述の方法のうちの1つまたは複数などの方法を実行するためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含んでもよい。さらに別の実装形態は、メモリと、上述の方法のうちの1つまたは複数などの方法を実行するための、メモリに記憶された命令を実行するように動作可能な1つまたは複数のプロセッサとを含むシステムを含んでもよい。

30

【0015】

前述の概念および本明細書においてより詳細に説明する他の概念のすべての組合せが、本明細書において開示される主題の一部と見なされることを諒解されたい。たとえば、本開示の最後に記載された請求された主題のすべての組合せが、本明細書において開示される主題の一部と見なされる。

40

【図面の簡単な説明】

【0016】

【図1】様々な実装形態による、クエリパターンおよび対応する総検索活動統計を生成するために本開示の様々な構成要素によって検索クエリの1つまたは複数のログが分析される場合がある環境を示す図である。

【図2】様々な実装形態による、クエリパターンおよび関連する総検索活動統計を取得するにはプライベート検索クエリおよび対応する検索後活動をどのように処理したらよいかを示す図である。

【図3】様々な実装形態による、クエリパターンおよび対応する総検索活動統計を生成し、その情報を様々な目的に適用する例示的な方法を示すフローチャートを示す図である。

50

【図4】コンピュータシステムの例示的なアーキテクチャを概略的に示す図である。

【発明を実施するための形態】

【0017】

図1は、プライベートと見なされる場合がある様々なシステム上でユーザが検索を実行する場合がある例示的な環境を示す。この例示的な環境は、クライアントデバイス102と、ドキュメントストレージシステム104と、eメールシステム106とを含む。ドキュメントストレージシステム104および/またはeメールシステム106は、たとえばネットワーク(図示せず)を介して通信する1つまたは複数のコンピュータに実装されてもよい。ドキュメントストレージシステム104および/またはeメールシステム106は、本明細書において説明する技術が実施される場合があり、ならびに/または本明細書において説明するシステム、構成要素、および技法が相互作用する場合があるシステムの例である。

10

【0018】

ユーザはクライアントデバイス102を介してドキュメントストレージシステム104および/またはeメールシステム106と対話してもよい。ドキュメントストレージシステム104および/またはeメールシステム106は、クライアントデバイス102から検索クエリ108を受信し、検索クエリ108に回答して検索結果110を返してもよい。本明細書において使用する「検索クエリ」108は、検索エンジンなどの検索インターフェースにサブMITTされる場合がある情報を求める要求を指すことがある。多くの場合、検索クエリは、情報を求める要求を集合的に形成する1つまたは複数のN-gramを含んでもよい。しかし、これは限定を目的としたものではない。検索クエリ108は、画像、オーディオなどの他の形式であってもよい。いくつかの実装形態では、他のクライアントデバイス、ならびに/あるいはドキュメントストレージシステム104および/またはeメールシステム106のプロバイダと提携しているウェブサイト用のサービスを実施する1つまたは複数のサーバなどの、他のコンピュータデバイスが、ドキュメントストレージシステム104および/またはeメールシステム106に検索クエリをサブMITTしてもよい。しかし、説明を簡潔にするために、各例についてクライアントデバイス102の文脈において説明する。

20

【0019】

様々な実装形態では、クライアントデバイス102は、ローカルエリアネットワーク(LAN)またはインターネットなどのワイドエリアネットワーク(WAN)などの1つまたは複数のネットワーク(図示せず)を介してドキュメントストレージシステム104および/またはeメールシステム106に通信可能に結合されたコンピュータであってもよい。クライアントデバイス102は、たとえば、デスクトップコンピューティングデバイス、ラップトップコンピューティングデバイス、タブレットコンピューティングデバイス、スマートフォンコンピューティングデバイス、ユーザの車両のコンピューティングデバイス(たとえば、車載通信システム、車載エンターテインメントシステム、車載ナビゲーションシステム)、またはコンピューティングデバイスを含むユーザのウェアラブル装置(たとえば、コンピューティングデバイスを有するユーザの時計、コンピューティングデバイスを有するユーザの眼鏡)であってもよい。追加および/または代替のクライアントデバイスが設けられてもよい。クライアントデバイス102は一般に、検索クエリのサブMITTおよびネットワークを介したデータの送受信を容易にするための1つまたは複数のアプリケーションを含む。たとえば、クライアントデバイス102は、ユーザが検索クエリを作成してドキュメントストレージシステム104および/またはeメールシステム106にサブMITTするのを可能にするブラウザ112、eメールクライアント114、および/またはドキュメントクライアント116などの1つまたは複数のアプリケーションを実行してもよい。

30

40

【0020】

図1では、ドキュメントストレージシステム104は、提示エンジン122<sub>A</sub>と、ランク付けエンジン124<sub>A</sub>と、クエリ分析エンジン126<sub>A</sub>とを含む。同様に、eメールシステム106は、それ自体の提示エンジン122<sub>B</sub>と、ランク付けエンジン124<sub>B</sub>と、クエリ分析エンジン126<sub>B</sub>とを含む。しかし、これは限定を目的としたものではない。様々な実装形態では、ドキュメントストレージシステム104とeメールシステムが、たとえば、同じエンティティによって動作

50

させられることにより、および/またはユーザが認証情報の単一のセットを使用してアクセスできるようにすることによって相互に関係付けられてもよい。そのような実装形態では、ドキュメントストレージシステム104とeメールシステム106は、共通の提示エンジン122、ランク付けエンジン124、および/またはクエリ分析エンジン126を共有してもよい。説明を簡潔にするために、ドキュメントストレージシステム104またはeメールシステム106の特定の構成要素が参照されていない限り、これらの構成要素は総称的に参照符号122、124、および126によって参照される。いくつかの実装形態では、エンジン122、124、および/または126のうちの1つまたは複数省略されてもよく、組み合わされてもよく、ならびに/あるいはドキュメントストレージシステム104および/またはeメールシステム106から分離された構成要素に実装されてもよい。いくつかの実装形態では、エンジン122、124、および/または126のうちの1つまたは複数、あるいはそれらの任意の動作可能部分、ならびに/あるいはドキュメントインデックス付けエンジン128および/またはeメールインデックス付けエンジン132(後述)は、クライアントデバイス102によって実行される構成要素に実装されてもよい。

10

20

30

40

50

#### 【0021】

ドキュメントストレージシステム104は、ユーザがたとえば、無料でまたは料金を支払ってドキュメントを記憶するためのオンラインストレージ空間を構成してもよい。いくつかの実装形態では、ドキュメントストレージシステム104は、「クラウドストレージ」と呼ばれることが多い態様でユーザ用のドキュメントを受け入れるファイルストレージおよび同期サービスであってもよい。ユーザは、クラウドストレージに関してユーザによって動作させられるローカルデバイス(たとえば、クライアントデバイス102)のファイルシステム上の位置を指定してもよい。クライアントデバイス102のクラウドストレージ位置に記憶された任意のドキュメントを、ユーザによって動作させられる他のデバイスの同様に指定された位置に伝播させてもよく/このような位置と同期させてもよい。本明細書において参照される「ドキュメント」には、たとえば、ウェブページ、文書処理ドキュメント、ポータブルドキュメントフォーマット(PDF)ドキュメント、画像、eメール、映像、およびスプレッドシート、プレゼンテーション、マルチメディアファイル、写真、およびオーディオファイルを含めてもよい。各ドキュメントは、たとえば、テキスト、画像、映像、音声、埋込み情報(たとえば、メタ情報および/またはハイパーリンク)、ならびに/あるいは埋込み命令(たとえば、JavaScriptなどのECMAScript実装形態)などのコンテンツを含んでもよい。

#### 【0022】

ドキュメントストレージシステム104は、インデックス130をドキュメントストレージシステム104によって使用できるように維持するように構成されたドキュメントインデックス付けエンジン128を含んでもよい。ドキュメントインデックス付けエンジン128は、ドキュメントを検索可能にするために、従来のインデックス付け技法および/またはその他のインデックス付け技法を使用してドキュメントを処理しインデックス130におけるインデックスエントリを更新してもよい。様々な実装形態では、ユーザは、ユーザが指定されたクラウドストレージに記憶しているドキュメントを検索するためにドキュメントストレージシステム104に検索クエリ108をサブミットすることが可能であってもよい。たとえば、ユーザは、ドキュメントストレージシステム104上に記憶されたドキュメントにアクセスするためのインターフェースとして動作するウェブサイトログインするようにブラウザ112を動作させてもよい。ウェブサイトは、ユーザが、たとえば1つまたは複数のN-gramの検索クエリ108をサブミットして特定のドキュメントを見つけるのを可能にする検索バーまたはその他の同様のインターフェースを構成してもよい。検索クエリ108がドキュメントストレージシステム104にアップロードされた後、ドキュメントインデックス付けエンジン128は、検索クエリと一致するドキュメントを検索するための様々な従来の技法またはその他の技法を実行してもよい。追加または代替として、ユーザは、ユーザがクラウドストレージに記憶したドキュメントをインデックス130において検索するために、クライアントデバイス102のオペレーティングシステムによって実現されるファイルマネージャ

に類似しているかまたは場合によってはこのファイルマネージャに組み込まれる場合があるドキュメントクライアント116を動作させてもよい。

【0023】

eメールシステム106は、eメール、およびeメールに関連して記憶されることが多い添付ファイル、連絡先、メモ、予定表項目などのeメールに関連するその他の情報を維持するように特定の構成されることを除いて、多くの点でドキュメントストレージシステム104と同様に動作してもよい。ドキュメントストレージシステム104の場合と同様に、eメールシステム106は、eメールおよび/または関連情報を検索可能にするために、従来のインデックス付け技法および/またはその他のインデックス付け技法を使用してeメールおよび関連情報を処理しインデックス134におけるインデックス項目を更新するeメールインデックス付けエンジン132を含む。したがって、たとえば、ユーザは、eメール、添付ファイル、予定表項目、メモ、タスク、リマインダー、連絡先などのコンテンツを検索するようにeメールクライアント114の検索インターフェースを動作させてもよい。あるいは、ドキュメントストレージシステム104と同様に、ユーザはeメールシステム106用のウェブインターフェースにログインするようにブラウザ112を動作させてもよい。ウェブインターフェースは、ユーザがeメール関連コンテンツを検索するのを可能にする検索バーまたは同様の機能を実現してもよい。

10

【0024】

様々な実装形態では、ドキュメントストレージシステム104および/またはeメールシステム106は検索ログ120を維持してもよい。図1では、ドキュメントストレージシステム104およびeメールシステム106用にそれぞれ、別個の検索ログ120<sub>A</sub>および120<sub>B</sub>が設けられるが、他の実装形態では、検索ログ120<sub>A</sub>および120<sub>B</sub>を単一の検索ログとして統一することができる。検索ログ120は、ドキュメントストレージシステム104および/またはeメールシステム106にサブMITされる検索クエリの様々な属性を記憶してもよい。たとえば、いくつかの実装形態では、検索ログ120は、サブMITされた検索クエリごとに、クエリをサブMITしたユーザに関連するユーザ識別子、タイムスタンプ、検索クエリの全テキスト(すなわち、クエリのすべてのN-gram)、および検索クエリに関連するいわゆる検索後活動を記憶してもよい。特定の検索クエリに関連する「検索後活動(Post-search Activity)」は、限定はしないが、検索クエリがサブMITされ検索結果が返された後でユーザによって行われる活動を含んでもよい。たとえば、いくつかの実装形態では、検索後活動には、検索クエリがサブMITされた後でユーザによってクリックされるドキュメント/検索結果/グラフィカル要素のランク付け、クリックが行われなかった(たとえば、ユーザが検索クエリを放棄し、その代わりに、新しい検索クエリをサブMITすることなどの他の活動を行った)ことを示す表示、ユーザが、検索クエリをサブMITする間またはその後クリックした他の何らかのグラフィカル要素の表示(たとえば、クエリサジェスション、eメールコンテキスト内の連絡先など)などを含めてもよい。

20

30

【0025】

本明細書では、「データベース」および「インデックス」という用語は、データの任意の集合を指すために広く使用される。データベースおよび/またはインデックスのデータを特定の方法で構造化する必要はなく、これらのデータは1つまたは複数の地理的位置にあるストレージデバイスに記憶することができる。したがって、たとえば、インデックス120、130、および/または134は、データの複数の集合を含んでもよく、各集合が異なるように構成されアクセスされてもよい。

40

【0026】

提示エンジン122(122<sub>A</sub>または122<sub>B</sub>を含む)は、クライアントデバイス102のユーザに提示される検索結果110を生成してもよい。たとえば、提示エンジン122は、検索クエリ108に回答して、検索結果および/またはその他のユーザインターフェース要素のリストをクライアントデバイス102に与えてもよい。いくつかの実装形態では、提示エンジン122は、関連性、タイムリー性などに関係する様々な基準を満たすいくつかの検索結果を他の検索結果よりも視覚的に強調させるかまたは他の検索結果よりも前に提示させてもよい。たとえ

50



ば、いくつかの結果は、フォント、フォントサイズ、色、動画、境界、マージン、またはその他の視覚的証印を使用して他の結果よりも視覚的に強調されてもよい。他の場合には、いくつかの結果が最初に提示され、その後関連性のより低い検索結果が提示されてもよい。

【0027】

ランク付けエンジン124(124<sub>A</sub>または124<sub>B</sub>を含む)は、たとえば、従来の情報検索技法および/またはその他の情報検索技法を使用し、それぞれのインデックス(たとえば、130および/または134)を使用して検索クエリ108に回答してドキュメント、eメール関連コンテンツ、および/またはその他の情報を特定してもよい。いくつかの実装形態では、ランク付けエンジン124は、たとえば、1つまたは複数のランク付け信号を使用して、検索クエリ108に回答して特定されたドキュメント、eメール関連コンテンツ、およびその他の情報に関するスコアを算出してもよい。各ランク付け信号は、ドキュメント、eメール関連コンテンツに関する情報、またはその他の情報自体、ドキュメント、eメール関連コンテンツ、または情報と検索クエリ108との間の関係、および/あるいはドキュメント、eメール関連コンテンツ、または情報と検索を実行するユーザとの間の関係を提示してもよい。

10

【0028】

クエリ分析エンジン126(126<sub>A</sub>または126<sub>B</sub>)は、プライベート検索クエリおよび対応する検索後活動の検索ログ120(図1における120<sub>A</sub>または120<sub>B</sub>)内において、プライベート検索クエリに含まれプライバシー基準を満たす複数の非プライベートN-gramなどの非プライベートコンテンツを特定するように構成されてもよい。いわゆる「プライバシー基準」は様々な形態であってもよい。たとえば、いくつかの実装形態では、頻度しきい値を満たす頻度で検索ログ120内に出現するN-gramによってプライバシー基準が満たされる。いくつかの実装形態では、少なくとも最小回数だけ検索ログ120内に出現するN-gramによってプライバシー基準が満たされる。いくつかの実装形態では、非プライベートN-gramおよびそれらに関連する頻度、すなわち、検索ログ120内に記憶された検索クエリに非プライベートN-gramが存在する頻度が、ヒストグラムとして構成されてもよい。より頻度の低いN-gramは、プライベート情報を有する可能性がより高く、ヒストグラムから破棄または排除されてもよい。これらの特徴は、計算コストを比較的安く抑えながらN-gramをプライベートまたは非プライベートとして分類することを可能にする。

20

【0029】

クエリ分析エンジン126は、プライバシー基準を満たした非プライベートコンテンツに基づいて、検索ログ120における別のパスを利用して複数のいわゆるクエリパターンを生成してもよい。「クエリパターン」は、検索ログ120内の1つまたは複数の検索クエリと一致する場合がある。クエリパターンは、限定はしないが、正規表現(たとえば、「resume\*」、「\* design doc」、「white paper\*」など)を含む様々な形態であってもよい。「プライベート」N-gram(すなわち、上述のプライバシー基準を満たさなかったN-gram)は、非プライベートN-gramのいずれとも一致せず、たとえばワイルドカードによって置き換えられてもよい。したがって、検索ログ120内に記憶された検索クエリに含まれていた可能性があるプライベート情報はクエリパターンから除去されてもよい。

30

【0030】

クエリ分析エンジン126はまた、複数のクエリパターンの各々に関連する総検索活動統計を検索ログ120から判定するように構成されてもよい。様々な実装形態では、各クエリパターンに関連する「総検索活動統計(Aggregate Search Activity Statistics)」は、クエリパターンと一致する検索ログ120内の複数のプライベート検索クエリに関連する検索活動および検索後活動を示してよい。

40

【0031】

様々な実装形態では、所与のクエリパターンに関連する総検索活動統計は、たとえば、この所与のクエリパターンと一致する検索ログ120内のすべての検索クエリおよび/または一意の数の検索クエリ、この所与のクエリパターンと一致する検索クエリをサブミットしたすべてのユーザおよび/または一意の数のユーザ、この所与のクエリパターンと一致す

50

る検索クエリ間において見られるセマンティックタイプのN-gram、プレフィックス/サフィックス照合に使用されるN-gram IDなどを含んでもよい。

【0032】

様々な実装形態では、所与のクエリパターンに関連する総検索後活動統計(Aggregate Post-search Activity Statistics)は、各検索クエリに関連して検索ログ120内に記憶されるデータとして上記において説明した検索後活動を表す集計データを含んでもよい。たとえば、総検索後活動統計は、所与のクエリパターンと一致する検索クエリをユーザがサブミットした後に観測されるいわゆるクリック位置統計(たとえば、ユーザがウェブ結果、クエリサジェスション、連絡先結果をクリックしたかどうか)、ユーザが所与のクエリパターンと一致する検索クエリをサブミットした後の離脱率(Abandonment Rate)(クリックが記録されない)、所与のクエリパターンと一致する検索クエリ全体にわたって検索された結果の平均数などを含んでもよい。

10

【0033】

図2は、ドキュメントストレージシステム104および/またはeメールシステム106の様々な構成要素、ならびに特にクエリ分析エンジン126の様々な(必ずしもそうであるとは限らないが、一般にはソフトウェア)構成要素が、検索クエリの検索ログ120を分析して1つまたは複数のクエリパターンおよび関連する総検索活動統計を生成し得るにはどうすればよいかの例を示す。図2は、生成されたクエリパターンおよび対応する統計が使用され得るにはどうすればよいかのいくつかの例も示す。

【0034】

まず、図2の1番上において、検索ログ120における複数の検索クエリ $\{SQ_1, SQ_2, \dots, SQ_M\}$ がプライバシーエンジン240に与えられ、および/またはプライバシーエンジン240によって取得されてもよい。プライバシーエンジン240は、たとえば、プライベート検索クエリに含まれプライバシー基準を満たす複数の非プライベートN-gram $\{N_1, N_2, \dots, N_N\}$ を特定してもよい。たとえば、いくつかの実装形態では、毎日少なくともx人の一意のユーザによってサブミットされる検索クエリに存在すること、毎週少なくともy人の一意のユーザによってサブミットされる検索クエリなどに存在することなどの、頻度しきい値を満たすN-gramのみが、非プライベートN-gramと見なされてもよい。上述のように、非プライベートN-gramは、いくつかの実装形態では、非プライベートN-gramおよび対応する頻度のヒストグラムとして構成されてもよい。

20

30

【0035】

様々な実装形態では、プライバシーエンジン240は他の技法を使用して複数の非プライベートN-gramを特定してもよい。たとえば、いくつかの実装形態では、プライバシーエンジン240は、N-gramが出現する複数の検索クエリにわたって、たとえばi18nなどの国際標準に従ってN-gramを分析し、これらの検索クエリにわたる母国語の分布を判定してもよい。次いで、これらの分布を使用してN-gramが非プライベートと見なされるかどうかを判定してもよい。プライバシーエンジンの他の実装形態は、様々なトークナイゼーション、字句解析、正規化、および/またはステミング技法を適用してもよい。たとえば、検索クエリに含まれる単語および語句は、分解され、さらに、たとえば構文解析、テキストマイニングなどを含む場合があるさらなる処理を受けてもよい。

40

【0036】

複数の検索クエリ $\{SQ_1, SQ_2, \dots, SQ_M\}$ は、クエリパターン生成エンジン242に与えられてもよく、またはクエリパターン生成エンジン242によって取得されてもよい。クエリパターン生成エンジン242は、プライバシーエンジン240によって特定された複数の非プライベートN-gram $\{N_1, N_2, \dots, N_N\}$ を使用して、複数の検索クエリから複数Pのクエリパターン $\{QP_1, QP_2, \dots, QP_P\}$ を生成してもよい。クエリパターン生成エンジン242は、様々な技法を使用してクエリパターンを特定してもよい。いくつかの実装形態では、検索クエリは、共有されるプレフィックス、サフィックス、および/またはインフィックスによってクエリパターンへとグループ化されてもよい。たとえば、検索クエリ「resume Bob Smith」、「resume Jane Doe」、および「Resume Arnold」はすべてプレフィックス「resu

50

me」を共有し、したがって、すべてがパターン「resume [A-z]+」と一致し得る。いくつかの実装形態では、検索クエリ同士が、構文に加えてまたは構文の代わりにセマンティクスに基づいて照合され得る。第4の検索クエリ「CV Allison Jones」が検索ログに記憶されていると仮定する。「CV」は「resume」と構文的には異なるが、この2つの単語は類似の意味を有する。したがって、すべての4つの検索クエリを取り込むために、「(resume|CV) [A-z]+」などの、よりロバストなクエリパターンが生成され得る。

【0037】

いくつかの実装形態では、クエリパターン生成エンジン242は、クエリパターンを次のように特定してもよい。所与の検索クエリQを、プライバシーエンジン240によって特定された複数の非プライベートN-gram  $\{N_1, N_2, \dots, N_N\}$ の各非プライベートN-gram  $N_i$ と照合することが試みられてもよい。「 $N^*$ 」、「 $*N$ 」、および/または「 $N^*N$ 」のいずれかがQと一致するかどうか判定されてもよい(ここで、「 $*$ 」はワイルドカードである)。3つの変数のいずれもQと一致しない場合、Qは次の非プライベートN-gram  $N_{i+1}$ と比較されてもよい。3つの変数のうちの1つまたは複数がQと一致する場合、クエリパターン生成エンジン242は、変数/Qの組合せをパターンとして保存し、次の非プライベートN-gram  $N_{i+1}$ に進んでもよい。クエリパターン生成エンジン242は、Qをすべての特定された非プライベートN-gramと比較することによって、Qおよび非プライベートN-gramを使用して作成することのできる最長のN-gramを特定してもよい。たとえば、プライバシーエンジン240が以下の2つの非プライベートN-gram「CV」および「CV DRAFT」を特定すると仮定する。クエリパターン生成エンジン242は最終的に「 $CV^*$ 」よりもパターン「 $CV DRAFT^*$ 」を優先して特定してもよい。その理由は「 $CV DRAFT^*$ 」の方が長いからである。

【0038】

いくつかの実装形態では、クエリパターン生成エンジン242は、それが生成するクエリパターンにセマンティック情報を組み込んでもよい。たとえば、クエリパターン生成エンジン242によって、たとえばプライベートN-gramが破棄される前にこれらのプライベートN-gramからセマンティック情報が抽出されてもよい。この抽出されたセマンティック情報は、より微細に調整されたワイルドカード、たとえば関連するセマンティックタイプを有するワイルドカードを作成するために使用されてもよい。たとえば、上記の例を使用すると、resume/CVに関する複数の検索クエリにおける関心対象の人物を特定するプライベートN-gramが、これらの検索クエリをサブミットしたユーザの連絡先(たとえば、eメール連絡先)を構成することがわかったときに、クエリパターン生成エンジン242によってクエリパターン「(resume|CV) <contact\_name>」が生成されてもよい。

【0039】

クエリパターン生成エンジン242の出力は、複数Pのクエリパターン、すなわち $\{QP_1, QP_2, \dots, QP_P\}$ であってもよく、検索ログ120に記憶された検索クエリおよび検索後活動 $\{<SQ_1, PSA_1>, <SQ_2, PSA_2>, \dots, <SQ_M, PSA_M>\}$ とともにクエリパターン統計エンジン244に与えられてもよい。クエリパターン統計エンジン244は、所与のクエリパターンQPと一致する検索クエリに関連する総検索活動統計を判定してもよい。これらの総統計には、総検索後活動統計およびその他の総検索活動統計を含めることができる。これらの統計については上記において説明した。

【0040】

クエリパターン統計エンジン244の出力は、クエリパターンおよび対応する総検索活動統計 $\{<QP_1, ASAS_1>, <QP_2, ASAS_2>, \dots, <QP_P, ASAS_P>\}$ であってもよい。このデータは、(検索ログ120の一部であってもよい)インデックス246に記憶されてもよく、それによって、限定はしないが提示エンジン122、ランク付けエンジン124、および/またはクエリ成績エンジン248を含む様々なダウンストリーム構成要素によってアクセス可能になる。これらの構成要素は、データを様々な方法で使用する場合がある。たとえば、提示エンジン122、ランク付けエンジン124、および/またはクエリ成績エンジン248のうちの1つまたは複数は、クエリを受信したことに応答してリアルタイムにデータを分析するかまたはその他のときに(たとえば、周期的に)データを分析し、所与のクエリパターンに関連する総検

索活動統計に基づいてこの所与のクエリパターンと一致する検索クエリに応答して提示されるデータを生成する方法を変更してもよい。

【0041】

様々な実装形態では、所与のクエリパターンに関連付けられインデックス246に記憶された総検索活動統計は、いわゆる「クリック位置」統計を含んでもよい。クリック位置統計は、所与のクエリパターンと一致する検索クエリをサブミットした後にユーザがクリックしたかまたはその他の方法で選択したものに関する統計(たとえば、網膜モニタは、選択したものを、ユーザがそれを見ていることに基づいて判定することができる)を指すことがある。ユーザは検索結果のリストのうちの特定の検索結果(たとえば、第1の位置、第2の位置、第10の位置など)をクリックすることがある。ユーザは、検索結果とともに(たとえば、検索結果の前にまたは検索結果の横に)返される広告をクリックすることがある。ユーザは、ハイパーリンクに加えてまたはハイパーリンクの代わりに純粋な情報を含むことがあるいわゆる「情報結果」をクリックすることがある。たとえば、検索クエリ「Ronald Reagan」をサブミットするユーザには、従来の検索結果に加えて、この元大統領について概略的に説明する抜粋が提示されることがある。

10

【0042】

クリック位置統計は、コンテキストに応じて、ユーザが検索結果をサブミットした後にクリックする必要がある他の項目と関係してもよい。特定のeメールを検索する検索クエリ108を図1におけるeメールシステム106にサブミットするユーザには、たとえば、この検索クエリと一致するeメールのリストに加えて、結果eメールに関連するeメール連絡先、および/またはこの検索クエリと一致するユーザに/からeメールを送った/受信した人のeメール連絡先など、ユーザがクリックすることができる他の項目が提示されてもよい。特定のドキュメントを検索する検索クエリ108を図1におけるドキュメントストレージシステム104にサブミットするユーザには、たとえば、この検索クエリと一致するドキュメントのリストに加えて、ドキュメントが属するプロジェクト、ドキュメントを作成した/編集した/「所有」している従業員の連絡先、ドキュメントが属する組織などの、他のクリック項目が提示されてもよい。ユーザは、これらの様々な構成要素をクリックするかまたは他の方法で選択してクリック位置統計を生成してもよい。

20

【0043】

様々な実装形態では、提示エンジン122または別のダウンストリーム構成要素は、クリック位置統計を分析して、クエリパターンと一致する検索クエリをサブミットするユーザに様々な情報を提示するのに使用される方法を変更してもよい。たとえば、比較的高いクリック率を実現するクリック位置に配置された各構成要素については、状況に応じて浮かだたせおよび/または強調を抑えてもよい。eメールシステム106にサブミットされた、特定のクエリパターンと一致する検索クエリ108の60%において、ユーザがeメール連絡先をクリックすると仮定する。提示エンジン122<sub>B</sub>は、将来同じクエリパターンと一致する検索クエリが受信されるときにこのことを考慮して、たとえば、eメール連絡先をより目立つように提示してもよい(またはユーザがeメール連絡先をそれほど頻繁にクリックしないことを望む場合には目立たなくしてもよい)。

30

【0044】

様々な実装形態では、提示エンジン122または別の構成要素は、クリック位置統計に加えてまたはクリック位置統計の代わりに、所与のクエリパターンに関連するクリック率を分析して、この所与のパターンと一致する検索クエリに対して結果をどのように提示するかを決定してもよい。特定のクエリパターンのクリック率が非常に低く、このことが、離脱率が高いことと合致し、および/または「低成績の」クエリパターンと見なされる場合があると仮定する。このことは、所与のクエリパターンと一致する検索クエリに、たとえば、それらがよくあるスペルミスを含むことに起因して欠陥があることを示す場合がある。提示エンジン122または別の構成要素は、将来同じクエリパターンと一致する検索クエリが受信されるときにこのことを考慮して、たとえば、「より高成績の」クエリパターンに関連するクエリサジェスチョンを(たとえば、検索バー上のドロップダウンメニューと

40

50

して、または別個の検索結果として)提示してもよい。たとえば、低成績のクエリパターンがずっと高成績のクエリパターンにかなり類似している(たとえば、1つのN-gramが変更されている)と仮定する。低成績のクエリパターンと一致する検索クエリが受信されるとき、この検索クエリは、クエリサジェスションをユーザに提案する場合により高成績のクエリパターンとともに使用されてもよい。このクエリサジェスションが元の検索クエリにおけるスペルミスを訂正してもよい。

**【0045】**

この技法は、一般にスペル訂正ソフトウェアによって使用される辞書に含まれているにもかかわらず誤って入力された用語または語句を検索クエリが含む例において、従来のスペル訂正技法に勝る利点を有する場合がある。たとえば、ユーザが「no reply」という語句を検索するつもりであったが誤って検索クエリ「no replay」をサブミットしたと仮定する。「no reply」と「no replay」がどちらもスペル訂正辞書に含まれるN-gramを含むことがあるので、従来のスペル訂正ソフトウェアはこのエラーを取り込まない場合がある。しかし、本明細書における技法では、eメールシステム106にサブミットされた、「no replay」という語句を含む検索クエリは非常に低成績であるが、「no reply」という用語を含む検索クエリは非常に高成績であることが示される場合がある。そのような場合、提示エンジン122または別の構成要素は、たとえば提案されるスペル訂正として「no reply」をユーザに提示してもよい。

**【0046】**

いくつかの実装形態では、提示エンジン122は、所与のクエリパターンと一致する検索クエリに回答してユーザに提示されるユーザインターフェース要素の属性、およびそのような要素とのユーザの対話にどのような傾向があるかを考慮してもよい。たとえば、上記においてRonald Reaganに関して説明したような情報結果は、検索結果ウェブページの強調表示された領域などの別個のユーザインターフェース要素またははっきりと異なるユーザインターフェース要素、あるいはポップアップウィンドウにおいてユーザに提示されてもよい。所与のクエリパターンと一致する検索クエリをサブミットするユーザが検索をサブミットした後に特定のユーザインターフェース要素をクリックする傾向がある場合、そのユーザインターフェース要素を将来より目立つように提示してもよく(たとえば、「浮きだたせる」)、または少なくとも引き続き提示してもよい。一方、ユーザがそのユーザインターフェース要素をクリックする傾向がない場合、そのユーザインターフェース要素はより目立たないように(たとえば、別の形式で)提示されてもよく、またはまったく提示されなくてもよい。あるいは、ユーザが比較的使用頻度が低いユーザインターフェース要素をより頻繁にクリックすることが望まれる場合、クリック頻度を高めるようにそのユーザインターフェース要素を浮きだたせるかまたは他の方法でより目だたさせてもよい。

**【0047】**

提示エンジン122に加えてまたは提示エンジン122の代わりに、他の構成要素が、インデックス246に記憶されたデータを利用して、データをどのようにユーザに提示させるかを変更してもよい。たとえば、所与のクエリパターンに関連する総検索活動統計が、ユーザは、検索から返された、このパターンと一致する検索結果において、非常に関連性の高い検索結果よりも非常に新しい検索結果の方をクリックする可能性がより高いことを示していると仮定する。そのような場合、ランク付けエンジン124は、検索クエリに回答する検索結果をランク付けする際にそれらの検索結果の新しさに対して、他の場合よりも強い重み付けを施してもよい。ランク付けエンジン124は、様々な実装形態において、そのランク付けを示すデータを提示エンジン122に与えてもよく、それによって、提示エンジン122は、最終的にクライアントデバイス102のユーザに提示されるデータを生成することができる。

**【0048】**

別の例として、成績に関係する様々な目的のために、クエリパターンおよび関連する総検索活動統計、特にクリック率、離脱率、クリック位置などを分析するためにクエリ成績エンジン248が使用されてもよい。クエリパターンの成績は、限定はしないがクリック率

10

20

30

40

50

、平均逆順位(「MRR」)、離脱率などを含む様々な評価基準に基づいて評価されてもよい。これらの評価基準は様々な用途に使用されてもよい。

【0049】

ユーザの特定のサブセット(たとえばランダムに選択されたIPアドレスを有するユーザ、特定の地域のユーザ、ベータユーザなど)に検索結果を与えるために、たとえばドキュメントストレージシステム104またはeメールシステム106上に実験的アルゴリズムが実装されると仮定する。さらに、他の「対照」ユーザがレガシーアルゴリズムを使用して引き続き検索結果を受信すると仮定する。この実験的アルゴリズムをレガシーアルゴリズムと比較する場合、本明細書の技法を使用して、ユーザのサブセットによってサブミットされた検索クエリに関連する検索後活動を分析し、クエリパターンおよび対応する総検索後活動統計を特定してもよい。次いで、これらの総検索後活動統計を、たとえばクエリ成績エンジン248によって、レガシーアルゴリズムに関して生成されたクエリパターンに関連する総検索後活動統計と比較して、検索後活動が実験的アルゴリズムとレガシーアルゴリズムとの間でどのように変化したかを判定してもよい。特定のクエリパターンがレガシーアルゴリズムの場合は高成績であるが実験的アルゴリズムの場合は低成績である場合、実験的アルゴリズムが、バグなどの1つまたは複数の問題を有することを示すことがある。

10

【0050】

クエリ成績エンジン248は、他の目的のためにクエリパターンを分析してもよい。いくつかの実装形態では、クエリ成績エンジン248は、所与のクエリパターンと一致する検索クエリに関連するクエリ活動統計を、時期、時刻、主題/テーマなどの様々な属性に基づくグループとして構成してもよい。たとえば、税金関連のクエリは確定申告の時期に急増しやすい。このことはクエリ成績エンジン248によって確認されおよび/または(たとえば、インデックス246に)記録されてもよい。確定申告の時期には、次いで、提示エンジン122またはランク付けエンジン124などの他の構成要素が、確定申告以外の時期とは異なるように、上記の確認情報を使用して検索結果および/またはその他のデータ/ユーザインターフェース要素をランク付けするかまたはその他の方法でユーザに提示してもよい。

20

【0051】

さらに他の実装形態では、クエリ成績エンジン248は、クエリパターンに関連する総検索後活動統計の変化を経時的に監視してもよい。所与のクエリパターンと一致する検索クエリに回答して提示されるインターフェースとのユーザの対話の状態が急激におよび/または大幅に変化した場合、クエリ成績エンジン248は、適切な管理要員(たとえば、ドキュメントストレージシステム104またはeメールシステム106の管理者)に自動的に通知することなど、様々な処置を講じてもよい。

30

【0052】

次に図3を参照すると、検索ログ内の検索クエリからクエリパターンを生成する例示的な方法300が示されている。説明の都合上、フローチャートの動作については、これらの動作を実行するシステムを参照しながら説明する。このシステムは、様々なコンピュータシステムの様々な構成要素を含んでもよい。さらに、方法300の動作は特定の順序で示されているが、これは限定を目的としたものではない。1つまたは複数の動作の順序変更、省略、または追加が行われてもよい。

40

【0053】

ブロック302において、システムは、図2のプライバシーエンジン240に関して上記において説明したように検索クエリのログに記憶された検索クエリに含まれる非プライベートコンテンツを特定してもよい。たとえば、プライバシー基準は、頻度しきい値を満たす頻度で検索ログ内に出現するN-gramなどのコンテンツによって満たされてもよく、または検索ログ内に少なくとも最小回数だけ出現するN-gramなどのコンテンツによって満たされてもよい。本明細書ではN-gramは様々な例について説明するために使用されるが、これは限定を目的としたものではない。様々な実装形態では、コンテンツが他の種類の単位またはトークンへとグループ化されおよび/または構文解析されてもよい。

【0054】

50

ブロック304において、システムは、ブロック302において特定された非プライベートコンテンツに基づいて複数のクエリパターン(たとえば、上記において説明し図2に示した $\{QP_1, QP_2, \dots, QP_N\}$ )を生成してもよい。たとえば、非プライベートコンテンツが各検索クエリと比較されてもよい。ブロック302において特定されたどの非プライベートコンテンツとも一致しない検索クエリのコンテンツはいずれも、たとえばプライベートと見なされる場合があるので無視または破棄されてもよい。クエリパターンを特定し得るにはどうすればよいかの他の例については、クエリパターン生成エンジン242および図2に関して上記において説明した。

【0055】

ブロック306において、システムは、ブロック304において生成された複数のクエリパターンの各々に関連する総検索活動を、たとえば検索ログ120から判定してもよい。上述のように、このことには、図2に示すクエリパターン統計エンジン244などの構成要素が、クエリパターンと一致する検索クエリから、一致する検索クエリの数、一意の一致する検索クエリの数、一致する検索クエリが最も頻繁にサブミットされる(または通常のようにサブミットされない)タイムフレームなどの、全般的な総検索活動統計を特定することを含めてもよい。このことにはまた、クエリパターン統計エンジン244などの構成要素が、クリック率、クリック位置、提示された/対話したユーザインターフェース要素、離脱率、MRRなどの、所与のクエリパターンと一致する検索クエリに関連する検索後活動から総検索後活動統計を特定することを含めてもよい。

【0056】

ブロック308において、システムは、所与のクエリパターンに関連する総検索活動統計(たとえば、図2における $\{ASAS_1, ASAS_2, \dots, ASAS_p\}$ )が成績基準(Performance Criterion)を満たしていると判定してもよい。たとえば、提示エンジン122は、たとえば検索クエリを受信したことに応答して、受信された検索クエリと一致するクエリパターンが、離脱しきい値を満たす総検索活動統計に関連していると判定してもよい。離脱しきい値は、たとえば、クリック率が、10%、5%などよりも低いことなど(これらの割合に限定されない)、ある割合より低いことによって満たされてもよい。別の例として、クエリ成績エンジン248は、ベータユーザに検索結果を提供するために実験的検索アルゴリズムが使用される場合、特定のクエリパターンの成績が著しく変化すると判定してもよい。

【0057】

ブロック308における判定に応答して、ブロック310において、システムは、所与のクエリパターンと一致する検索クエリに応答して提示されるデータ(たとえば、検索結果ウェブページに関するHTMLデータ、eメールクライアント114またはドキュメントクライアント116において提示すべきデータ)を生成するための方法を変更してもよい。この変更は、所与のクエリパターンに関連する総検索活動統計に少なくとも部分的に基づいて行われてもよい。たとえば、ブロック308において、ユーザがしきい値(たとえば、ある最小割合)を満たす特定の頻度で所与のクエリパターンと一致する検索クエリに応答して提示される特定のユーザインターフェース要素をクリックすると判定されると仮定する。これに応答して、システムは、たとえば、ユーザインターフェース要素をより目立つように提示するか、浮きだたせるか、または場合によってはまったく新しい方法で(たとえば、ポップアップとして)提示することによって、ユーザインターフェース要素を提示するために使用する方法を変更してもよい。同様に、ユーザインターフェース要素は、クリックされた回数が最小しきい値未満である場合、より目立たないように提示されてもよく、または場合によってはもはや提示されなくてもよい。

【0058】

図4は、例示的なコンピュータシステム410のブロック図である。コンピュータシステム410は一般に、バスサブシステム412を介していくつかの周辺デバイスと通信する少なくとも1つのプロセッサ414を含む。これらの周辺デバイスは、たとえばメモリサブシステム425およびファイルストレージサブシステム426を含むストレージサブシステム424と、ユーザインターフェース出力デバイス420と、ユーザインターフェース入力デバイス422と、ネ

10

20

30

40

50

ットワークインターフェースサブシステム416とを含んでもよい。入力デバイスおよび出力デバイスは、ユーザとコンピュータシステム410との対話を可能にする。ネットワークインターフェースサブシステム416は、外部ネットワークとのインターフェースを構成し、他のコンピュータシステムにおける対応するインターフェースデバイスに結合される。  
【0059】

ユーザインターフェース入力デバイス422には、キーボード、マウス、トラックボール、タッチパッド、もしくはグラフィックタブレットなどのポインティングデバイス、スキャナ、ディスプレイに組み込まれたタッチスクリーン、音声認識システムなどのオーディオ入力デバイス、マイクロフォン、および/または他の種類の入力デバイスを含めてもよい。一般に、「入力デバイス」という用語を使用する場合、コンピュータシステム410内または通信ネットワーク上に情報を入力するためのすべての考えられる種類のデバイスおよび方法を含むことが意図される。

10

【0060】

ユーザインターフェース出力デバイス420には、ディスプレイサブシステム、プリンタ、ファックス、またはオーディオ出力デバイスなどの非視覚的ディスプレイを含めてもよい。ディスプレイサブシステムには、陰極線管(CRT)、液晶ディスプレイ(LCD)などのフラットパネルデバイス、投影デバイス、または視覚的画像を作成するための何らかの他の機構を含めてもよい。ディスプレイサブシステムはまた、オーディオ出力デバイスを介することなどによって非視覚的ディスプレイを構成してもよい。一般に、「出力デバイス」という用語を使用する場合、コンピュータシステム410からユーザまたは別の機器もしくはコンピュータシステムに情報を出力するためのすべての考えられる種類のデバイスおよび方法を含むことが意図される。

20

【0061】

ストレージサブシステム424は、本明細書に記載されたモジュールのうちのいくつかまたはすべての機能を実現するプログラミング構造およびデータ構造を記憶する。たとえば、ストレージサブシステム424は、方法300の選択された態様を実行し、ならびに/あるいは提示エンジン122、ランク付けエンジン124、クエリ分析エンジン126の1つもしくは複数の構成要素(たとえば、240、242、244、246、248)、ドキュメントインデックス付けエンジン128、および/またはeメールインデックス付けエンジン132のうちの1つまたは複数を実装するための論理を含んでもよい。

30

【0062】

これらのソフトウェアモジュールは一般に、プロセッサ414のみによって実行されるかまたは他のプロセッサと組み合わせて実行される。ストレージサブシステム424において使用されるメモリ425には、プログラム実行時に命令およびデータを記憶するためのメインランダムアクセスメモリ(RAM)430、ならびに一定の命令が記憶される読取り専用メモリ(ROM)432を含むいくつかのメモリを含めることができる。ファイルストレージサブシステム426は、プログラムおよびデータファイル用の永続的なストレージを構成することができ、ハードディスクドライブ、関連する取外し可能媒体を伴うフロッピーディスクドライブ、CD-ROMドライブ、光学ドライブ、または取外し可能なメディアカートリッジを含んでもよい。いくつかの実装形態の機能を実現するモジュールは、ファイルストレージサブシステム426によって、ストレージサブシステム424に記憶されてもよく、またはプロセッサ414によってアクセス可能な他のマシン内に記憶されてもよい。

40

【0063】

バスサブシステム412は、コンピュータシステム410の様々な構成要素およびサブシステムを意図された通りに互いに通信させるための機構を構成する。バスサブシステム412は単一のバスとして概略的に示されているが、バスサブシステムの代替実装形態は複数のバスを使用してもよい。

【0064】

コンピュータシステム410は、ワークステーション、サーバ、コンピューティングクラスタ、ブレードサーバ、サーバファーム、または任意の他のデータ処理システムもしくは

50



コンピューティングデバイスを含む様々な種類のシステムであってもよい。コンピュータおよびネットワークは絶え間なく変化するので、図4に示すコンピュータシステム410の説明は、いくつかの実装形態を例示することを目的とした特定の例を対象としたものに過ぎない。コンピュータシステム410の他の多くの構成は、図4に示すコンピュータシステムよりもより多い構成要素またはより少ない構成要素を有することが可能である。

【0065】

本明細書において説明するシステムがユーザに関する個人情報収集するかまたは個人情報を利用する場合がある状況では、プログラムまたは機能がユーザ情報(たとえば、ユーザのソーシャルネットワーク、社会活動、職業、ユーザの好み、またはユーザの現在の地理的位置に関する情報)を収集するかどうかを調節するか、あるいはユーザとの関連性がより高い場合があるコンテンツをコンテンツサーバから受信するかどうかおよび/またはどのように受信するかを調節する機会がユーザに与えられ得る。さらに、特定のデータは、個人を特定可能な情報が除去されるように、記憶または使用される前に1つまたは複数の方法で処理されてもよい。たとえば、ユーザに関して個人を特定可能な情報を判定することができないようにユーザの識別子が処理され得るか、あるいはユーザの特定の地理的位置を判定できないように地理的位置情報が取得されるユーザの地理的位置が(市レベル、郵便番号レベル、または州レベルなどまで)一般化されるようにユーザの識別子が処理されてもよい。したがって、ユーザは、ユーザに関する情報がどのように収集されおよび/または使用されるかを調節してもよい。

【0066】

本明細書においていくつかの実装形態について説明し図示したが、本明細書において説明した機能を実行しならびに/あるいは本明細書において説明した結果および/または利点のうち1つもしくは複数を取得するための他の様々な手段および/または構造が利用されてもよく、そのような変形実施形態および/または修正実施形態の各々が、本明細書において説明する実装形態の範囲内であると見なされる。より一般的には、本明細書において説明するすべてのパラメータ、寸法、材料、および構成は例示的なものであり、実際のパラメータ、寸法、材料、および/または構成は、本教示が使用される特定の適用例によって決まる。当業者には、本明細書において説明する特定の实装形態の多くの均等物が認識されるか、または当業者はそれらを日常的な実験のみを使用して確認することができよう。したがって、前述の実装形態は一例としてのみ提示されており、添付の特許請求の範囲およびその均等物の範囲内で、明示的に記載され請求された実装形態以外の実装形態が実施されてもよいことを理解されたい。本開示の実装形態は、本明細書において説明した各々の個々の特徴、システム、物品、材料、キット、および/または方法を対象とする。さらに、そのような特徴、システム、物品、材料、キット、および/または方法が互いに矛盾しない場合、2つ以上のそのような特徴、システム、物品、材料、キット、および/または方法の任意の組合せが本開示の範囲内に含まれる。

【符号の説明】

【0067】

- 102 クライアントデバイス
- 104 ドキュメントストレージシステム
- 106 eメールシステム
- 108 検索クエリ
- 110 検索結果
- 112 ブラウザ
- 114 eメールクライアント
- 116 ドキュメントクライアント
- 120<sub>A</sub> 検索ログ
- 120<sub>B</sub> 検索ログ
- 122<sub>A</sub> 提示エンジン
- 122<sub>B</sub> 提示エンジン

10

20

30

40

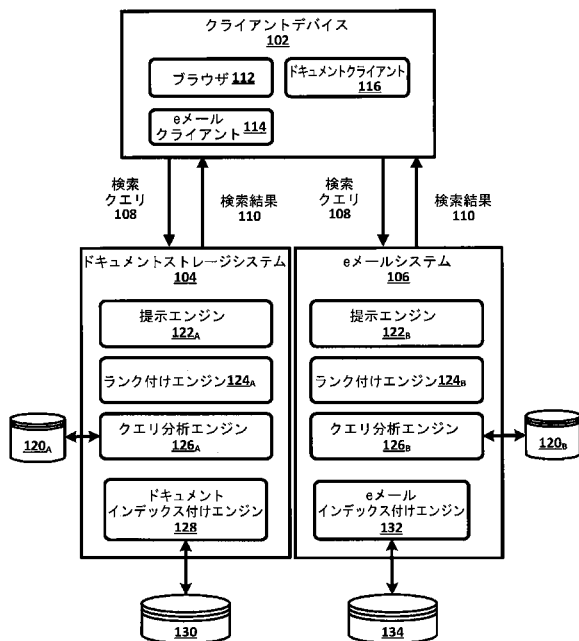
50

- 124<sub>A</sub> ランク付けエンジン
- 124<sub>B</sub> ランク付けエンジン
- 126<sub>A</sub> クエリ分析エンジン
- 126<sub>B</sub> クエリ分析エンジン
- 128 ドキュメントインデックス付けエンジン
- 130 インデックス
- 132 eメールインデックス付けエンジン
- 134 インデックス
- 240 プライバシーエンジン
- 242 クエリパターン生成エンジン
- 244 クエリパターン統計エンジン
- 246 インデックス
- 248 クエリ成績エンジン
- 410 コンピュータシステム
- 412 バスサブシステム
- 414 プロセッサ
- 416 ネットワークインターフェースサブシステム
- 420 ユーザインターフェース出力デバイス
- 422 ユーザインターフェース入力デバイス
- 424 ストレージサブシステム
- 425 メモリサブシステム
- 426 ファイルストレージサブシステム
- 430 メインランダムアクセスメモリ (RAM)
- 432 読取り専用メモリ (ROM)

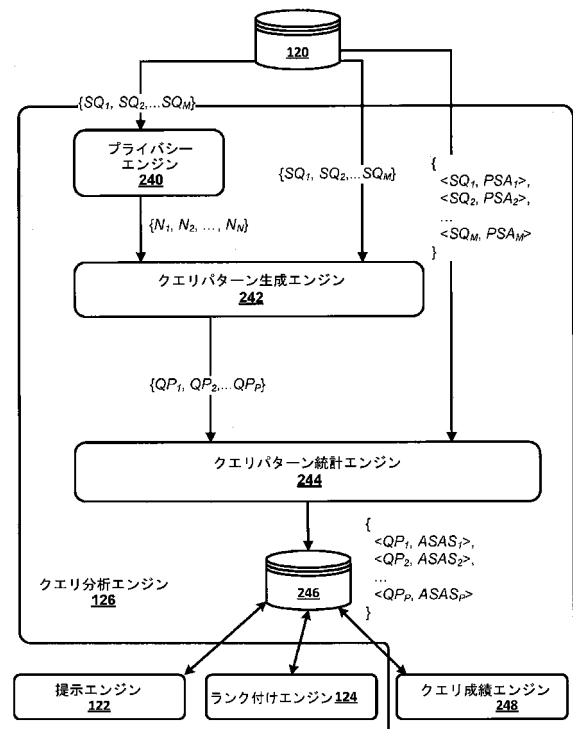
10

20

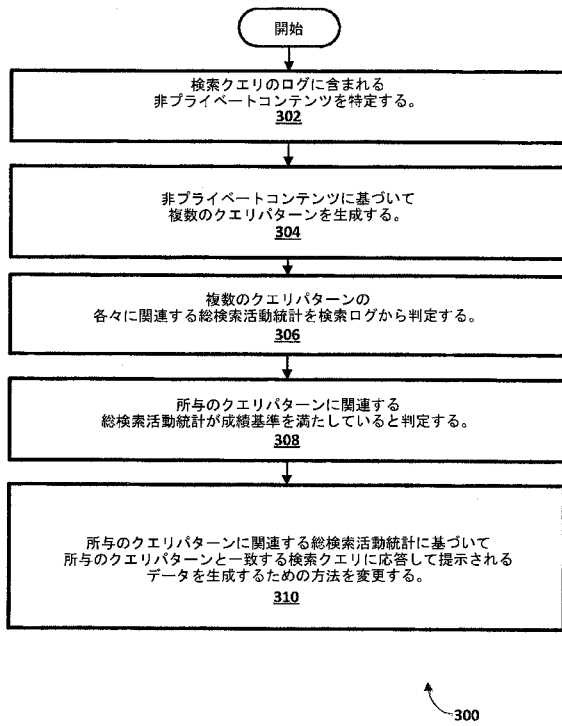
【 図 1 】



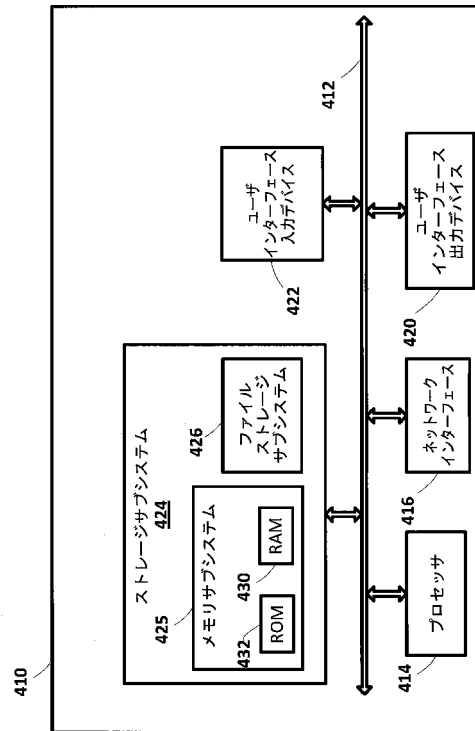
【 図 2 】



【 図 3 】



【 図 4 】



## 【 国際調査報告 】

## INTERNATIONAL SEARCH REPORT

International application No PCT/US2016/063175
---

<b>A. CLASSIFICATION OF SUBJECT MATTER</b> INV. G06F21/62 G06F17/30 ADD.		
According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b> Minimum documentation searched (classification system followed by classification symbols) 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 practicable, search terms used) EPO-Internal, WPI Data		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2014/089285 A1 (GIBBS KEVIN A [US]) 27 March 2014 (2014-03-27) paragraph [0046] - paragraph [0082]; figures 1-14 -----	1-15
A	US 8 364 709 B1 (DAS ABHINANDAN S [US] ET AL) 29 January 2013 (2013-01-29) column 5, line 26 - column 7, line 47; figures 2-7 -----	1-15
A	US 2008/294630 A1 (YAN WEIPENG [US] ET AL) 27 November 2008 (2008-11-27) paragraph [0033] - paragraph [0056]; figures 2-4 -----	1-15
<input 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 application or patent 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 "&" document member of the same patent family
Date of the actual completion of the international search		Date of mailing of the international search report
9 February 2017		20/02/2017
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016		Authorized officer  Savvides, George

**INTERNATIONAL SEARCH REPORT**

Information on patent family members

International application No

PCT/US2016/063175

Patent document cited in search report	Publication date	Patent family member(s)	Publication date	
US 2014089285	A1	27-03-2014	CN 101194256 A	04-06-2008
			JP 5459958 B2	02-04-2014
			JP 5557862 B2	23-07-2014
			JP 2008520037 A	12-06-2008
			JP 2012108959 A	07-06-2012
			KR 20070086055 A	27-08-2007
			US 2006106769 A1	18-05-2006
			US 2014089285 A1	27-03-2014
			US 2015046422 A1	12-02-2015
			WO 2006055120 A2	26-05-2006
			US 8364709	B1
US 8930399 B1	06-01-2015			
US 9239888 B1	19-01-2016			
US 2008294630	A1	27-11-2008	AU 2008254644 A1	27-11-2008
			CN 101796515 A	04-08-2010
			EP 2171621 A1	07-04-2010
			ES 2428546 T3	08-11-2013
			JP 5255055 B2	07-08-2013
			JP 2010530566 A	09-09-2010
			US 2008294630 A1	27-11-2008
			WO 2008144732 A1	27-11-2008

## フロントページの続き

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

(特許庁注：以下のものは登録商標)

1. J A V A S C R I P T

- (72)発明者 ドナルド・メッツラー  
アメリカ合衆国・カリフォルニア・94043・マウンテン・ビュー・アンフィシアター・パーク  
ウェイ・1600
- (72)発明者 マーク・アレクサンダー・ナジョーク  
アメリカ合衆国・カリフォルニア・94043・マウンテン・ビュー・アンフィシアター・パーク  
ウェイ・1600
- (72)発明者 ドル・ナヴェ  
アメリカ合衆国・カリフォルニア・94043・マウンテン・ビュー・アンフィシアター・パーク  
ウェイ・1600
- (72)発明者 ブラッド・パナイト  
アメリカ合衆国・カリフォルニア・94043・マウンテン・ビュー・アンフィシアター・パーク  
ウェイ・1600
- (72)発明者 シュアンファイ・ワン  
アメリカ合衆国・カリフォルニア・94043・マウンテン・ビュー・アンフィシアター・パーク  
ウェイ・1600