

【公報種別】特許法第 17 条の 2 の規定による補正の掲載
 【部門区分】第 6 部門第 3 区分
 【発行日】平成20年9月11日(2008.9.11)

【公表番号】特表2008-519328(P2008-519328A)
 【公表日】平成20年6月5日(2008.6.5)
 【年通号数】公開・登録公報2008-022
 【出願番号】特願2007-539077(P2007-539077)
 【国際特許分類】

G 0 6 F 13/00 (2006.01)

G 0 6 F 17/30 (2006.01)

【 F I 】

G 0 6 F 13/00 5 4 0 E

G 0 6 F 17/30 3 5 0 C

G 0 6 F 17/30 3 4 0 A

G 0 6 F 17/30 1 8 0 Z

【手続補正書】

【提出日】平成20年7月17日(2008.7.17)

【手続補正 1】

【補正対象書類名】特許請求の範囲

【補正対象項目名】全文

【補正方法】変更

【補正の内容】

【特許請求の範囲】

【請求項 1】

検索結果セットにおいて検索ヒットをランク付けする、コンピュータによって実現される方法であって、

ユーザからクエリを受取るステップと、

クエリに関連するヒットのリストを生成するステップとを備え、

ヒットの各々はクエリに関連し、

少なくとも 1 つのヒットは、ブーストドキュメントにおけるリンクによって指し示され

、

ブーストドキュメントにおけるリンクは、少なくとも 1 つのヒットの、クエリとの関連性を人為的に高め、前記方法はさらに、

前記少なくとも 1 つのヒットにメトリックを関連付けるステップを備え、メトリックは、前記少なくとも 1 つのヒットの、クエリとの関連性を人為的に高める前記少なくとも 1 つのヒットへのリンクを含むブーストドキュメントの数を表わし、前記方法はさらに、

メトリックを閾値と比較するステップと、

一部比較に基づいて、修正されたリストを形成するためにヒットのリストを処理するステップとを備え、

前記少なくとも 1 つのヒットは、前記修正されたリストから除外されるか、または前記ヒットのリストにおける前記少なくとも 1 つのヒットに起因していた関連性よりも関連性が低い状態で前記修正されたリストの中で提示されるかのいずれかであり、前記方法はさらに、

修正されたリストをユーザに伝送するステップを備える、方法。

【請求項 2】

少なくとも 1 つのヒットにメトリックを前記関連付けるステップは、前記メトリックを形成するステップを備え、前記メトリックを形成するステップは、

前記少なくとも 1 つのヒットについて第 1 の尺度を形成するステップを備え、第 1 の尺

度は、前記少なくとも1つのヒットのリンクポピュラリティを表わし、前記メトリックを前記形成するステップはさらに、

前記少なくとも1つのヒットについて第2の尺度を形成するステップを備え、第2の尺度は、前記少なくとも1つのヒットが評判のよいドキュメントである可能性を示し、前記メトリックを前記形成するステップはさらに、

第1の尺度および第2の尺度を備える組合せを形成するステップを備え、組合せは、第1の尺度と第2の尺度との間の差を表わす、請求項1に記載の方法。

【請求項3】

前記メトリックを前記形成するステップは、クエリを前記受取るステップの前に実行される、請求項2に記載の方法。

【請求項4】

第2の尺度を前記形成するステップは、

評判のよいドキュメントのシードセットを形成するステップを備え、評判のよいドキュメントのシードセットは他のドキュメントへのリンクを備え、第2の尺度を前記形成するステップはさらに、

シードセットにおけるドキュメントの各々に信頼値を割当てるステップと、

シードセットにおけるドキュメントのうちの少なくとも1つによって指し示される複数のドキュメントの各々に信頼値を伝播させるステップと、

比例配分された信頼値を、シードセットにおけるドキュメントのうちの少なくとも1つによって指し示される複数のドキュメントの各々に割当てるステップとを備える、請求項2に記載の方法。

【請求項5】

シードセットを前記形成するステップは、

複数のドキュメントの各々ごとに、複数のドキュメントの各々に含まれるアウトリンクの数を表わすアウトリンクメトリックをそれぞれ求めるステップと、

アウトリンクメトリックを使用して複数のドキュメントをランク付けするステップと、複数のドキュメントにおいて最も高くランク付けされたドキュメントのセットを識別するステップと、

最も高くランク付けされたドキュメントのセットの各々の質を評価するステップと、

最も高くランク付けされたドキュメントのセットから不適切であると考えられるドキュメントを除去することによって、最も高くランク付けされたドキュメントの修正されたセットを形成するステップと、

最も高くランク付けされたドキュメントの修正されたセットを使用してシードセットを形成するステップとを備える、請求項4に記載の方法。

【請求項6】

検索結果セットにおいて検索ヒットをランク付けするための命令を格納する、コンピュータによって実現されるコンピュータ可読記憶媒体であって、命令は、

ユーザからクエリを受取るステップと、

クエリに関連するヒットのリストを生成するステップとを実行するための命令を含み、

ヒットの各々はクエリに関連し、

少なくとも1つのヒットは、ブーストドキュメントにおけるリンクによって指し示され

、

ブーストドキュメントにおけるリンクは、少なくとも1つのヒットの、クエリとの関連性を人為的に高め、前記命令はさらに、

前記少なくとも1つのヒットにメトリックを関連付けるステップを実行するための命令を含み、メトリックは、前記少なくとも1つのヒットの、クエリとの関連性を人為的に高める前記少なくとも1つのヒットへのリンクを含むブーストドキュメントの数を表わし、前記命令はさらに、

メトリックを閾値と比較するステップと、

一部比較に基づいて、修正されたリストを形成するためにヒットのリストを処理するス

テップとを実行するための命令を含み、

前記少なくとも1つのヒットは、前記修正されたリストから除外されるか、または前記ヒットのリストにおける前記少なくとも1つのヒットに起因していた関連性よりも関連性が低い状態で前記修正されたリストの中で提示されるかのいずれかであり、前記命令はさらに、

修正されたリストをユーザに伝送するステップを実行するための命令を含む、コンピュータ可読記憶媒体。

【請求項7】

少なくとも1つのヒットにメトリックを前記関連付けるステップは、前記メトリックを形成するステップを備え、前記メトリックを形成するステップは、

前記少なくとも1つのヒットについて第1の尺度を形成するステップを備え、第1の尺度は、前記少なくとも1つのヒットのリンクポピュラリティを表わし、前記メトリックを前記形成するステップはさらに、

前記少なくとも1つのヒットについて第2の尺度を形成するステップを備え、第2の尺度は、前記少なくとも1つのヒットが評判のよいドキュメントである可能性を示し、前記メトリックを前記形成するステップはさらに、

第1の尺度および第2の尺度を備える組合せを形成するステップを備え、組合せは、第1の尺度と第2の尺度との間の差を表わす、請求項6に記載のコンピュータ可読記憶媒体。

【請求項8】

前記メトリックを前記形成するステップは、クエリを前記受取るステップの前に実行される、請求項7に記載のコンピュータ可読記憶媒体。

【請求項9】

第2の尺度を前記形成するステップは、

評判のよいドキュメントのシードセットを形成するステップを備え、評判のよいドキュメントのシードセットは他のドキュメントへのリンクを備え、第2の尺度を前記形成するステップはさらに、

シードセットにおけるドキュメントの各々に信頼値を割当てするステップと、

シードセットにおけるドキュメントのうちの少なくとも1つによって指し示される複数のドキュメントの各々に信頼値を伝播させるステップと、

比例配分された信頼値を、シードセットにおけるドキュメントのうちの少なくとも1つによって指し示される複数のドキュメントの各々に割当てするステップとを備える、請求項7に記載のコンピュータ可読記憶媒体。

【請求項10】

シードセットを前記形成するステップは、

複数のドキュメントの各々ごとに、複数のドキュメントの各々に含まれるアウトリンクの数を表わすアウトリンクメトリックをそれぞれ求めるステップと、

アウトリンクメトリックを使用して複数のドキュメントをランク付けするステップと、

複数のドキュメントにおいて最も高くランク付けされたドキュメントのセットを識別するステップと、

最も高くランク付けされたドキュメントのセットの各々の質を評価するステップと、

最も高くランク付けされたドキュメントのセットから不適切であると考えられるドキュメントを除去することによって、最も高くランク付けされたドキュメントの修正されたセットを形成するステップと、

最も高くランク付けされたドキュメントの修正されたセットを使用してシードセットを形成するステップとを備える、請求項9に記載のコンピュータ可読記憶媒体。

【手続補正2】

【補正対象書類名】明細書

【補正対象項目名】0011

【補正方法】変更

【補正の内容】

【0011】

別の局面では、第2の尺度は評判のよいドキュメントのシードセット (seed set) を形成することによって生成され、評判のよいドキュメントのシードセットはリンクを張るドキュメントであり、上記第2の尺度はさらに、シードセットにおけるドキュメントの各々に信頼値を割当てることと、リンクを張るドキュメントによって指し示されるリンクされるドキュメントの各々にその信頼値を伝播させることと、比例配分された信頼値を、リンクされるドキュメントの各々に割当てることによって生成される。

【手続補正3】

【補正対象書類名】明細書

【補正対象項目名】0039

【補正方法】変更

【補正の内容】

【0039】

リンクベースのスパム検出

図2に示されるように、検索サーバシステム160は、ウェブスパムページが格下げされるかまたはリストから除去された場合に修正された検索リストを返すリンクベースのスパム検出器180と組合せられて機能し、検索サーバシステム160の出力(結果、提案、媒体コンテンツなど)をリンクベースのスパム検出器180に与える。検索サーバシステム160は、この発明の実施例による検索エンジンを動作させるよう構成される。検索エンジンは、3つの部分、すなわち1つ以上のスパイダ162と、データベース163と、ツール/アプリケーション167とから成る。情報を集めるスパイダ162はインターネット中をゆっくり進み、データベース163にはスパイダが集めた情報および他の情報が入っており、ツール/アプリケーション167は、データベースを検索するためにユーザによって使用される検索ツール166などのアプリケーションを含む。データベース163には、検索ツールによって使用されるページ索引170が入っている。さらに、この発明の実施例による検索エンジンは、スパム検出器180を含む。スパム検出器180は、以下に記載されるさまざまなアルゴリズムを実行し、ページについてのウェブスパムメトリック181をページ索引170に格納する。上で説明されたように、この発明の実施例によるスパム検出器180は、ヒットの有効質量に一致するメトリックを推定し、検索ツール166およびページ索引170と組合せられて機能し、有効質量がリンクベースのスパムによってヒットを人為的にブーストする可能性を高くするそのヒットを格下げする。所与のウェブドキュメントについての有効質量を求めることは、所与のウェブドキュメントのリンクベースのポピュラリティ(たとえば、ページランク)と、所与のウェブドキュメントの信頼性(たとえば、トラストランク)との間の相違を部分的に査定する技術の組合せに依拠する。一実施例では、ウェブスパム検出器180は、ページ索引170におけるすべてのページを処理して、索引におけるページについてウェブスパムメトリック181を計算し、ウェブスパムメトリック181をデータベース163に格納する。メトリック181は、ドキュメントを検索結果に含ませる検索クエリから独立している。