

【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6月19日(2008.6.19)

【公開番号】特開2005-322244(P2005-322244A)
 【公開日】平成17年11月17日(2005.11.17)
 【年通号数】公開・登録公報2005-045
 【出願番号】特願2005-134488(P2005-134488)
 【国際特許分類】

G 0 6 F 17/30 (2006.01)

【F I】

G 0 6 F 17/30 3 5 0 C

G 0 6 F 17/30 3 7 0 Z

【手続補正書】

【提出日】平成20年4月30日(2008.4.30)

【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】

コンピュータシステムにおいて、検索結果のドキュメントを順位付ける方法であって、
 前記検索結果の各ドキュメントについて、前記ドキュメントの情報の豊富さにより類似度の順位を初期化するステップであって、ドキュメントの前記情報の豊富さは、前記ドキュメントがその話題に関係する情報をどの程度含むのかを示す尺度であり、ドキュメントの前記類似度の順位は、1つのドキュメントの前記情報が、別のドキュメントの前記情報に、どの程度含まれているかを示す尺度に基づいているステップと、

類似するドキュメントの各グループについて、最も高い類似度の順位を除く前記類似度の順位が、前記グループ内のドキュメントの最も高い類似度の順位と比較して低いよう、前記グループ内のドキュメントの前記類似度の順位を調整するステップと
 を備えたことを特徴とする方法。

【請求項 2】

前記グループ内のドキュメントの前記類似度の順位を調整する前記ステップは、前記グループ内の最も高い類似度の順位を有する前記ドキュメントの前記類似度の順位を除いて、前記グループ内の各ドキュメントの前記類似度の順位を減少させるステップを含むことを特徴とする請求項 1 に記載の方法。

【請求項 3】

最も高い類似度の順位を有する前記ドキュメントに、より類似するドキュメントは、最も高い類似度の順位を有する前記ドキュメントに、より類似しないドキュメントよりも非常に減少させた類似度の順位を有することを特徴とする請求項 2 に記載の方法。

【請求項 4】

前記グループ内のドキュメントの前記類似度の順位を調整する前記ステップは、
 前記グループから最も高い類似度の順位を有する前記ドキュメントを除去するステップと、
 前記グループの残りのドキュメントの前記類似度の順位を減少させるステップと
 を含み、前記ドキュメントを除去する順序が前記検索結果の前記ドキュメントの前記順位を表すことを特徴とする請求項 1 に記載の方法。

【請求項 5】

各ドキュメントについて、前記調整された類似度の順位および検索ベースの適合度により、前記ドキュメントの適合度を計算するステップを含むことを特徴とする請求項 1 に記載の方法。

【請求項 6】

コンピュータシステムにおいて、検索結果のドキュメントの順序付けを行い、高い順序のドキュメントの話題の多様性を増大させる方法であって、

検索要求を受信するステップと、

前記検索要求を満たすドキュメントを検索するステップと、

前記検索要求を満たすいくつかの前記ドキュメントを前記検索結果として指定するステップと、

前記検索結果が指定された後に、1つのドキュメントの前記情報が、別のドキュメントの前記情報に、どの程度含まれているかを示す、ドキュメントの複数の組の間の類似度、ドキュメントの組の間の前記類似度を示す類似度の図式に基づいて、前記検索結果に類似するドキュメントのグループを識別するステップと、

前記識別されたグループの各々から1つのドキュメントを選択するステップと、

前記検索結果の他のドキュメントの上位に前記選択されたドキュメントを順位付けるステップと

を備えたことを特徴とする方法。

【請求項 7】

各ドキュメントが初期の順位を有し、および前記順位は、より高い初期の順位を有する別のドキュメントより高く、選択したドキュメントを順位付けるステップを含むことを特徴とする請求項 6 に記載の方法。

【請求項 8】

各ドキュメントが初期の順位を有し、および前記識別されたグループの各々から前記選択されたドキュメントは、最も高い初期の順位を有する前記ドキュメントであることを特徴とする請求項 6 に記載の方法。

【請求項 9】

グループの選択されないドキュメントを、前記グループの前記選択されたドキュメントの類似性により、再度順位付けるステップを含むことを特徴とする請求項 6 に記載の方法。

【請求項 10】

前記再度順位付けるステップは、グループの前記選択されたドキュメントに最も類似する、前記グループの前記選択されないドキュメントに、前記グループの前記ドキュメントの順位において最も大きい減少を与えることを特徴とする請求項 9 に記載の方法。

【請求項 11】

前記グループの前記選択されないドキュメントは、再度順位付けるステップに従って順位付けされることを特徴とする請求項 10 に記載の方法。

【請求項 12】

再度順位付けるステップの後に前記識別されたグループの各々から1つのドキュメントを選択するステップと、

前記ドキュメントを、まだ選択されていない他のドキュメントの上位に順位付けるステップと

を含むことを特徴とする請求項 10 に記載の方法。

【請求項 13】

前記再度順位付けるステップは、類似性のペナルティを適用することを特徴とする請求項 9 に記載の方法。

【請求項 14】

前記グループの各々から前記選択されたドキュメントは、前記グループ内の前記ドキュメントで最も高い情報の豊富さを有することを特徴とする請求項 6 に記載の方法。

【請求項 15】

前記グループは、類似度の図式を使用して識別されることを特徴とする請求項 6 に記載の方法。

【請求項 16】

コンピュータシステムにおいて、ドキュメントの集まり内のドキュメントの情報の豊富さを計算する方法であって、

前記集まり内の各ドキュメントが有する、前記ドキュメントの類似度を識別するステップであって、前記類似度は、1つのドキュメントの前記情報が、別のドキュメントの前記情報に、どの程度含まれているかを示す尺度であるステップと、

前記集まり内の他のドキュメントが有する、前記ドキュメントの前記類似度により、前記ドキュメントの情報の豊富さを決定するステップであって、ドキュメントの前記情報の豊富さは、前記ドキュメントがその話題に関係する情報をどの程度含むのかを示す尺度であるステップと

を備えたことを特徴とする方法。

【請求項 17】

前記各ドキュメントの類似度を識別するステップは、類似度の図式を生成するステップを含むことを特徴とする請求項 16 に記載の方法。

【請求項 18】

前記類似度は、1つのドキュメントの情報内容が別のドキュメントに含まれる程度の尺度であることを特徴とする請求項 16 に記載の方法。

【請求項 19】

前記類似度は、

【数 1】

$$\text{aff}(d_i, d_j) = \frac{\vec{d}_i \cdot \vec{d}_j}{\|\vec{d}_i\|}$$

として定義されることを特徴とする請求項 16 に記載の方法。

【請求項 20】

前記情報の豊富さは、1つのドキュメントの情報内容が、他のドキュメントの情報内容を含む程度の尺度となることを特徴とする請求項 16 に記載の方法。

【請求項 21】

前記情報の豊富さは、

【数 2】

$$\text{InfoRich}(d_i) = \sum_{\text{all } j \neq i} \text{InfoRich}(d_j) \cdot \tilde{M}_{ji}$$

として定義されることを特徴とする請求項 16 に記載の方法。

【請求項 22】

方法によりコンピュータシステムにドキュメントの順位付けを行わせる命令を含む、コンピュータ可読媒体であって、前記方法は、

各ドキュメントについて、前記ドキュメントの情報の豊富さにより、類似度の順位を初期化するステップであって、ドキュメントの前記情報の豊富さは、前記ドキュメントがその話題に関係する情報をどの程度含むのかを示す尺度であり、ドキュメントの前記類似度の順位は、1つのドキュメントの前記情報が、別のドキュメントの前記情報に、どの程度含まれているかを示す尺度に基づいているステップと、

ドキュメントが高い類似度の順位を有する場合、関連するドキュメントの前記類似度の順位を減少させるステップと

を備え、前記類似度の順位が前記ドキュメントの前記順位付けを表すことを特徴とするコンピュータ可読媒体。

【請求項 23】

前記ドキュメントの情報の豊富さは、ドキュメントの組の類似度により計算されること

を特徴とする請求項 2 2 に記載のコンピュータ可読媒体。

【請求項 2 4】

前記情報の豊富さは、

【数 3】

$$\text{InfoRich}(d_i) = \sum_{\text{all } j \neq i} \text{InfoRich}(d_j) \cdot \tilde{M}_{ji}$$

として定義されることを特徴とする請求項 2 3 に記載のコンピュータ可読媒体。

【請求項 2 5】

前記類似度は、

【数 4】

$$\text{aff}(d_i, d_j) = \frac{\bar{d}_i \cdot \bar{d}_j}{\|\bar{d}_i\|}$$

として定義されることを特徴とする請求項 2 3 に記載のコンピュータ可読媒体。

【請求項 2 6】

前記高い類似度の順位を有するドキュメントに、より類似する関連ドキュメントは、前記高い類似度の順位を有する前記ドキュメントに、より類似しないドキュメントよりも非常に減少させた類似度の順位を有することを特徴とする請求項 2 2 に記載のコンピュータ可読媒体。

【請求項 2 7】

各ドキュメントについて、前記ドキュメントの類似度の順位および前記ドキュメントの検索ベースの適合度により、前記ドキュメントの適合度を計算することを含むことを特徴とする請求項 2 2 に記載のコンピュータ可読媒体。

【請求項 2 8】

ドキュメントの集まり内のドキュメントの情報の豊富さを計算するためのコンピュータシステムであって、

前記ドキュメントに対する、前記集まり内の各ドキュメントが有する類似度を識別するコンポーネントと、

前記ドキュメントに対する、前記集まり内の他のドキュメントが有する前記類似度により、前記ドキュメントの情報の豊富さを決定するコンポーネントと

を備えたことを特徴とするコンピュータシステム。

【請求項 2 9】

前記識別するコンポーネントは、類似度の図式を生成することを特徴とする請求項 2 8 に記載のコンピュータシステム。

【請求項 3 0】

前記類似度は、1つのドキュメントの情報内容が別のドキュメントに含まれる程度の尺度であることを特徴とする請求項 2 8 に記載のコンピュータシステム。

【請求項 3 1】

前記情報の豊富さは、1つのドキュメントの情報内容が他のドキュメントの情報内容を含む程度の尺度となることを特徴とする請求項 2 8 に記載のコンピュータシステム。