

【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22年9月9日(2010.9.9)

【公表番号】特表2009-545809(P2009-545809A)

【公表日】平成21年12月24日(2009.12.24)

【年通号数】公開・登録公報2009-051

【出願番号】特願2009-522778(P2009-522778)

【国際特許分類】

G 06 F 17/30 (2006.01)

【F I】

G 06 F 17/30 3 4 0 B

G 06 F 17/30 1 7 0 A

G 06 F 17/30 3 5 0 C

【手続補正書】

【提出日】平成22年7月20日(2010.7.20)

【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】

コンピュータ実行可能命令を記録したコンピュータ可読記録媒体であって、該命令はコンピュータによって実行されると、

ユーザによって入力された、1または2以上の検索クエリ用語を含む検索クエリに応答して、使用データに基づいて文書をランク付けするためのインクリメンタルに更新可能なクエリ分類器を使用して、ネットワーク上の文書をランク付けすることと、

事前に計算され、検索クエリ用語のそれぞれについて記憶済みの値($w_i, Asset$)および $\log[(w_i, Asset)]$ と、事前に計算され、文書のそれぞれについて記憶済みの値#($Asset$), $\log[\#(Asset)]$ および#($w_i, Asset$)を含む使用データに基づいてクエリ分類器によってユーザランク付けに対する文書を表示することと、

閲覧するためにユーザが文書を選択することに応答して、

カウント値#($Asset$), ($w_i, Asset$)および#($w_i, Asset$)を更新すること、

値 $\log[(Asset)]$ および $\log[(w_i, Asset)]$ を算出すること、および

事前に計算された値を置き換えて更新されたユーザデータを記憶することにより、使用データを更新することとをコンピュータに実行させることを含み、

#($Asset$)は、ネットワーク上の所与の文書が閲覧のために選択された回数を表し、

$\log[(Asset)]$ は、($Asset$)のログを表し、

($w_i, Asset$)は、ネットワーク上の所与の文書および検索クエリの検索クエリ用語 w_i がマッチした回数を表し、

$\log[(w_i, Asset)]$ は、($w_i, Asset$)のログを表し、

#($w_i, Asset$)は、ネットワーク上の所与の文書および検索クエリ用語 w_i がマッチした回数の合計を表す、ことを特徴とするコンピュータ可読記録媒体。

【請求項2】

前記ネットワーク上の各文書は、前記検索クエリに対するその関連度に基づいてランク付けされ、以下の公式（I）を用いて文書関連度スコアが提供され、

【数1】

$$\log[P(Asset | Query)] = \log[\#(Asset)] - \log[\#(T)] + \sum_{i=1}^{N_Q} \left(\log[\#(w_i, Asset)] - \log[\#(Asset)] + \frac{\#(w_i, Asset)}{\#(Asset)} \right) - \sum_{i=1}^V \frac{\#(w_i, Asset)}{\#(Asset)} \quad (I)$$

上式で、

$P(Asset | Query)$ は、ユーザによって入力された特定の検索クエリ $Query$ が与えられた場合に所与の文書 $Asset$ を返す確率を表し、

N_Q は前記検索クエリ中の用語の数であり、

V は前記ネットワークの語彙のサイズであり、

$\#(T)$ は任意のユーザによって処理された検索クエリの総数であることを特徴とする請求項1に記載のコンピュータ可読記録媒体。

【請求項3】

前記インクリメンタルに更新可能なクエリ分類器モデルは各期間の終わりに更新され、前記期間の長さは24時間以下であることを特徴とする請求項1または2に記載のコンピュータ可読記録媒体。

【請求項4】

カウント値 $\#(Asset)$ 、 $\#(w_i, Asset)$ 、および $\#(w_i, Asset)$ を更新することは、所与の期間中に収集された新しいデータがあればそれを、前に記憶されたカウント値 $\#(Asset)$ (old)、 $\#(w_i, Asset)$ (old)、および $\#(w_i, Asset)$ (old) に加えることを含むことを特徴とする請求項1から3のいずれかに記載のコンピュータ可読記録媒体。

【請求項5】

前記インクリメンタルに更新可能なクエリ分類器モデルはさらに時間減衰コンポーネントを含み、最近の検索クエリおよびユーザ応答には過去の検索クエリおよびユーザ応答よりも大きい重みが与えられることを特徴とする請求項1から4のいずれかに記載のコンピュータ可読記録媒体。

【請求項6】

【数2】

$$\#(Asset) = \sum_{t=0}^{\infty} \lambda^t [\#(Asset)(t)];$$

$$\#(w_i, Asset) = \sum_{t=0}^{\infty} \lambda^t [\#(w_i, Asset)(t)]; \text{ および}$$

$$\Sigma \#(w_i, Asset) = \sum_{t=0}^{\infty} \lambda^t [\Sigma \#(w_i, Asset)(t)];$$

上式で、

λ は1.0未満の値を有する重み付け乗数であり、

t はカウント値コンポーネントの古さを表す整数であることを特徴とする請求項1から5のいずれかに記載のコンピュータ可読記録媒体。

【請求項 7】

カウント値 $\#(\text{Asset})$ 、 $\#(w_i, \text{Asset})$ 、および $\#(w_i, \text{Asset})$ を更新することは、

$\#(\text{Asset})(\text{new}) = \#(\text{Asset})(0) + [\#(\text{Asset})(\text{old})]$ 、

$\#(w_i, \text{Asset})(\text{new}) = \#(w_i, \text{Asset})(0) + \#(w_i, \text{Asset})(\text{old})$ 、および

$\#(w_i, \text{Asset})(\text{new}) = \#(w_i, \text{Asset})(0) + [\#(w_i, \text{Asset})(\text{old})]$

のように $\#(\text{Asset})$ 、 $\#(w_i, \text{Asset})$ 、および $\#(w_i, \text{Asset})$ を再計算することを含み、

上式で、

$\#(\text{Asset})(\text{new})$ 、 $\#(w_i, \text{Asset})(\text{new})$ 、および $\#(w_i, \text{Asset})(\text{new})$ は各々独立して、カウント値 $\#(\text{Asset})$ 、 $\#(w_i, \text{Asset})$ 、および $\#(w_i, \text{Asset})$ のインクリメンタルに更新された値をそれぞれ表し、

$\#(\text{Asset})(0)$ 、 $\#(w_i, \text{Asset})(0)$ 、および $\#(w_i, \text{Asset})(0)$ は各々独立して、最後の期間内の発生数をそれぞれ表し、

$\#(\text{Asset})(\text{old})$ 、 $\#(w_i, \text{Asset})(\text{old})$ 、および $\#(w_i, \text{Asset})(\text{old})$ は各々独立して、前記最後の期間よりも前の累積カウント値をそれぞれ表すことを特徴とする請求項 6 に記載のコンピュータ可読記録媒体。

【請求項 8】

ユーザによって入力された前記検索照会を受け入れ、前記ネットワーク上の前記文書の検索を実施して複数の文書を含む検索結果を生成し、前記インクリメンタルに更新可能なクエリ分類器モデルを使用して前記検索結果の前記複数の文書をランク付けしてランク付き検索結果を生成し、前記ランク付き検索結果を前記ユーザに対して表示するためのコンピュータ実行可能命令をさらに含むことを特徴とする請求項 1 から 7 のいずれかに記載のコンピュータ可読記録媒体。

【請求項 9】

検索エンジン中のランク付け関数コンポーネントとして使用するのに適したクエリ分類器モデルをインクリメンタルに更新する方法であって、

ユーザによって入力された、1または2以上の検索クエリ用語を含む検索クエリに応答して、使用データに基づいて文書をランク付けするためのインクリメンタルに更新可能なクエリ分類器を使用して、ネットワーク上の文書をランク付けすることと、

事前に計算され、検索クエリ用語のそれについて記憶済みの値 (w_i, Asset) および $\log[(w_i, \text{Asset})]$ と、事前に計算され、文書のそれについて記憶済みの値 $\#(\text{Asset}), \log[\#(\text{Asset})]$ および $\#(w_i, \text{Asset})$ を含む使用データに基づいてクエリ分類器によってユーザランク付けに対する文書を表示することと、

閲覧するためにユーザが文書を選択することに応答して、

カウント値 $\#(\text{Asset}), (w_i, \text{Asset})$ および $\#(w_i, \text{Asset})$ を更新すること、

値 $\log[(\text{Asset})]$ および $\log[(w_i, \text{Asset})]$ を算出すること、および

事前に計算された値を置き換えて更新されたユーザデータを記憶することにより、使用データを更新することとを含み、

$\#(\text{Asset})$ は、ネットワーク上の所与の文書が閲覧のために選択された回数を表し、

$\log[(\text{Asset})]$ は、 (Asset) のログを表し、

(w_i, Asset) は、ネットワーク上の所与の文書および検索クエリの検索クエリ

リ用語 w_i がマッチした回数を表し、

$\log[(w_i, Asset)]$ は、 $(w_i, Asset)$ のログを表し、
 $\#(w_i, Asset)$ は、ネットワーク上の所与の文書および検索クエリ用語 w_i がマッチした回数の合計を表す、ことを特徴とする方法。

【請求項 10】

前記期間の長さは 24 時間以下であることを特徴とする請求項 9 に記載の方法。

【請求項 11】

記憶されたカウント値 $\#(Asset)$ 、 $\#(w_i, Asset)$ 、および $\#(w_i, Asset)$ を更新することは、

$\#(Asset)(new) = \#(Asset)(0) + [\#(Asset)(old)]$ 、

$\#(w_i, Asset)(new) = \#(w_i, Asset)(0) + \#(w_i, Asset)(old)$ 、および

$\#(w_i, Asset)(new) = \#(w_i, Asset)(0) + [\#(w_i, Asset)(old)]$

のように $\#(Asset)$ 、 $\#(w_i, Asset)$ 、および $\#(w_i, Asset)$ を再計算することを含み、

上式で、

$\#(Asset)(new)$ 、 $\#(w_i, Asset)(new)$ 、および $\#(w_i, Asset)(new)$ は、カウント値 $\#(Asset)$ 、 $\#(w_i, Asset)$ 、および $\#(w_i, Asset)$ のインクリメンタルに更新された値をそれぞれ表し、

$\#(Asset)(0)$ 、 $\#(w_i, Asset)(0)$ 、および $\#(w_i, Asset)(0)$ は、最後の期間内の発生数をそれぞれ表し、

$\#(Asset)(old)$ 、 $\#(w_i, Asset)(old)$ 、および $\#(w_i, Asset)(old)$ は、前記最後の期間よりも前の累積カウント値をそれぞれ表し、
 は重み付け乗数であることを特徴とする請求項 9 または 10 に記載の方法。

【請求項 12】

は 1.0 未満の値を有することを特徴とする請求項 11 に記載の方法。

【請求項 13】

ユーザによって入力された検索クエリに対するネットワーク上の文書の文書関連度スコアを決定する方法であって、請求項 9 から 12 のいずれかに記載の方法を含むことを特徴とする方法。

【請求項 14】

前記文書関連度スコアは、以下の公式 (I)

【数 3】

$$\log[P(Asset | Query)] = \log[\#(Asset)] - \log[\#(T)] + \sum_{i=1}^{N_Q} \left(\log[\#(w_i, Asset)] - \log[\#(Asset)] + \frac{\#(w_i, Asset)}{\#(Asset)} \right) - \sum_{i=1}^V \frac{\#(w_i, Asset)}{\#(Asset)} \quad (I)$$

を使用して決定され、

上式で、

$P(Asset | Query)$ は、ユーザによって入力された特定の検索クエリ $Query$ が与えられた場合に所与の文書 $Asset$ を返す確率を表し、

N_Q は前記検索クエリ中の用語の数であり、

V は前記ネットワークの語彙のサイズであり、

$\#(T)$ は任意のユーザによって処理された検索クエリの総数であることを特徴とする請求項 13 に記載の方法。

【請求項 15】

検索クエリの検索結果をランク付けする方法であって、
請求項 1 3 または 1 4 に記載の方法を使用してネットワークの各文書の文書関連度スコアを決定するステップと、

各文書の前記文書関連度スコアに基づいて前記文書を降順でランク付けするステップとを含むことを特徴とする方法。

【請求項 1 6】

請求項 1 0 から 1 5 のいずれかに記載の方法を実施するための手段を備えることを特徴とするコンピューティングシステム。