

【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1月31日(2008.1.31)

【公開番号】特開2005-222532(P2005-222532A)

【公開日】平成17年8月18日(2005.8.18)

【年通号数】公開・登録公報2005-032

【出願番号】特願2005-4310(P2005-4310)

【国際特許分類】

G 06 N 3/08 (2006.01)

G 06 F 17/30 (2006.01)

【F I】

G 06 N 3/08 Z

G 06 F 17/30 1 7 0 A

G 06 F 17/30 2 2 0 Z

G 06 F 17/30 3 4 0 A

【手続補正書】

【提出日】平成19年12月10日(2007.12.10)

【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】

コンピュータで機械学習アプローチを実施してドキュメントの関連性を決定するシステムであって、

肯定的テストケースとして使用すべき、人または機械が選択した1組のアイテムを受け取るストレージコンポーネントと、

肯定的テストケースとしての前記人または機械が選択したアイテム、および否定的テストケースとしての1つまたは複数の他のアイテムを用いて少なくとも1つの分類機構トレーニングしてクエリから独立したモデルを提供するトレーニングコンポーネントとを備え、

前記トレーニング済みの分類機構が、統計ベース又は確率ベースの探索から得られたドキュメントをフィルタするために使用されることを特徴とするシステム。

【請求項2】

前記否定的テストケースは、統計的探索によって選択されることを特徴とする請求項1に記載のシステム。

【請求項3】

前記トレーニング済みの分類機構は、新しい肯定的ケースを選択する際に個人を助けるために使用されることを特徴とする請求項1に記載のシステム。

【請求項4】

前記フィルタの出力は、肯定的ケースが否定的ケースの前に格付けされるように格付けされることを特徴とする請求項1に記載のシステム。

【請求項5】

前記出力は、それらが肯定的ケースであるという確率に従って格付けされることを特徴とする請求項1に記載のシステム。

【請求項6】

前記ストレージコンポーネントは、ユーザにとって関心のある関連サイト、ドキュメン

ト、またはデータアイテムのログを含むことを特徴とする請求項1に記載のシステム。

【請求項7】

前記ストレージコンポーネントは、集中ストアについての、または、ウェブサイト、ドキュメントのコレクション、百科事典、ローカルデータソース、リモートデータソースなど分散ソースからの情報を含むことを特徴とする請求項6に記載のシステム。

【請求項8】

前記分類機構は、前記ストレージコンポーネント中のデータを自動的に分析して、ユーザインターフェースと対話することができる1つまたは複数のツールを支援するために使用されることを特徴とする請求項1に記載のシステム。

【請求項9】

前記ツールは、管理ツール、編集ツール、および格付けツールのうちの少なくとも1つを含むことを特徴とする請求項8に記載のシステム。

【請求項10】

前記ツールは、オンライン方式およびオフライン方式のうちの少なくとも一方で使用されることを特徴とする請求項8に記載のシステム。

【請求項11】

前記分類機構を肯定的テストデータおよび否定的テストデータに従ってトレーニングして、ドキュメント、または有用な情報の他のサイトを示唆するリンクなどからアイテムの関連性を決定することを特徴とする請求項1に記載のシステム。

【請求項12】

機械学習分類機構をトレーニングするための1組の手動で選択されるドキュメントまたはアイテムをさらに含むことを特徴とする請求項11に記載のシステム。

【請求項13】

前記分類機構を新しい用語に適用して最善策または関連ドキュメントを識別することを特徴とする請求項11に記載のシステム。

【請求項14】

様々なトレーニングの反復にわたって新しいモデルをポートストラップし、より正確な情報検索活動のために使用される学習表現の成長モデルを実施するステップをさらに含むことを特徴とする請求項11に記載のシステム。

【請求項15】

編集者によって手動で選択される最善策をさらに含むことを特徴とする請求項14に記載のシステム。

【請求項16】

ユーザが閲覧または検索するのに十分に関心があると考える可能性が高いタイプのドキュメントまたはアイテムを表示する可能性を最大化するコンポーネントをさらに含むことを特徴とする請求項15に記載のシステム。

【請求項17】

前記分類機構は、以下の学習技法、すなわちサポートベクトルマシン(SVM)、ナイーブベイズ、ベイズネット、判定ツリー、類似ベース、ベクトルベース、隠しマルコフモデル、または他の学習技法のうちの少なくとも1つを含むことを特徴とする請求項1に記載のシステム。

【請求項18】

情報の後処理を実施して、ユーザまたは管理者に対するドキュメントまたはサイトの関連性を決定するコンポーネントをさらに含むことを特徴とする請求項1に記載のシステム。

【請求項19】

前記後処理は、所定の確率しきい値による格付けを含み、関連性についてのより高い確率を有するアイテムは、より低い確率を有するアイテムの前に提示されることを特徴とする請求項18に記載のシステム。

【請求項20】

表示アイテムに追加されてドキュメントまたはサイトの関連性または重要度を示す明示的な注釈をさらに含むことを特徴とする請求項1_8に記載のシステム。

【請求項 2 1】

請求項 1 に記載の前記トレーニングコンポーネントおよび前記ストレージコンポーネントを実施するためのコンピュータ読み取り可能命令が記憶されていることを特徴とするコンピュータ読み取り可能媒体。

【請求項 2 2】

データ用語についてのトレーニングセットを決定する手段と、
前記トレーニングセットを自動的に分類する手段と、
前記分類済みのトレーニングセットから新しいアイテムを決定する手段と、
情報検索要求に従って前記新しいアイテムを提示する手段と
を備えることを特徴とするコンピュータベースの情報検索システム。

【請求項 2 3】

前記分類されたトレーニングセットをテストする手段をさらに備えることを特徴とする請求項2_2に記載のシステム。

【請求項 2 4】

コンピュータで自動化された情報検索を実施する方法であって、
 n を整数として、データログから n 個のクエリを処理するステップと、
前記 n 個のクエリから関連候補を識別するステップと、
分類機構をトレーニングして後続の探索活動のための他の関連候補を識別するステップと
を含むことを特徴とする方法。

【請求項 2 5】

1つの情報が、所与のクエリまたはトピックについて提示することが望ましいか否かを判定する分析を編集者に転送するステップをさらに含むことを特徴とする請求項2_4に記載の方法。

【請求項 2 6】

可能性のあるドキュメントまたはサイトのリストから関連候補を抽出するステップと、最善策を他の統計的格付けの前に自動的に配置するステップとをさらに含むことを特徴とする請求項2_4に記載の方法。

【請求項 2 7】

あるドキュメントが関連する確率によって結果を再格付けするステップをさらに含み、各ドキュメントは、ダウンロードされ、用語は、前記ドキュメント中に現れる用語について抽出され、ルックアップされることを特徴とする請求項2_4に記載の方法。

【請求項 2 8】

分類すべき少なくとも 1 つのカテゴリを決定するステップをさらに含むことを特徴とする請求項2_4に記載の方法。

【請求項 2 9】

トレーニングデータセットのサブセットを使用して前記分類済みのカテゴリをテストするステップをさらに含むことを特徴とする請求項2_8に記載の方法。

【請求項 3 0】

関連性カテゴリについてのトレーニングデータセットに関連した第 1 のデータフィールドと、

前記関連性カテゴリに属する新しい 1 組のデータアイテムに関連した第 2 のデータフィールドと、

前記新しい 1 組のデータアイテムについての確率格付けに関連した第 3 のデータフィールドと

を含む、データ構造が記憶されていることを特徴とするコンピュータ読み取り可能媒体。