

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5192475号
(P5192475)

(45) 発行日 平成25年5月8日(2013.5.8)

(24) 登録日 平成25年2月8日(2013.2.8)

(51) Int.Cl. F1
G06F 17/30 (2006.01) G06F 17/30 210D

請求項の数 9 外国語出願 (全 16 頁)

(21) 出願番号	特願2009-243546 (P2009-243546)	(73) 特許権者	505418870 エヌイーシー (チャイナ) カンパニー, リミテッド NEC (China) Co., Ltd. 中華人民共和国 100191 베이ジン ハイディエンディストリクト, シュエユ エンロード 35, シーニンタワー 20 階
(22) 出願日	平成21年10月22日 (2009.10.22)	(74) 代理人	100093595 弁理士 松本 正夫
(65) 公開番号	特開2010-170529 (P2010-170529A)	(72) 発明者	リー ジェンチャン 中華人民共和国 100084 베이ジン チンフア サイエンス パーク, イ ノベーション プラザ, ビルディング エー, 11エフ
(43) 公開日	平成22年8月5日 (2010.8.5)		
審査請求日	平成21年10月22日 (2009.10.22)		
(31) 優先権主張番号	200810173612.5		
(32) 優先日	平成20年10月30日 (2008.10.30)		
(33) 優先権主張国	中国 (CN)		

最終頁に続く

(54) 【発明の名称】 オブジェクト分類方法およびオブジェクト分類システム

(57) 【特許請求の範囲】

【請求項1】

オブジェクト分類システムを構成するコンピュータによるオブジェクト分類方法であって、

オブジェクト取得手段が、オブジェクト集合を取得するオブジェクト取得ステップを実行し、

クエリログベース分類手段が、クエリログに基づいてオブジェクトを分類し、第1の分類結果をクエリリストの形式で生成するクエリログベース分類ステップを実行し、

オントロジ情報ベース分類手段が、オントロジ情報に基づいてオブジェクトを分類し、第2の分類結果を概念リストの形式で生成するオントロジ情報ベース分類ステップを実行し、

意味的融合手段が、前記第1と第2の分類結果を意味的に融合し、最終的な分類結果を生成する意味的融合ステップを実行し、

前記意味的融合手段が備える第1の調整手段が、前記第2の分類結果に従って前記第1の分類結果を調整する第1の調整ステップを実行し、

前記意味的融合手段が備える第2の調整手段が、前記第1の分類結果に従って前記第2の分類結果を調整する第2の調整ステップを実行し、

前記意味的融合手段が備える統合手段が、調整後の前記第1の分類結果及び調整後の前記第2の分類結果を統合して最終的な分類結果を生成する統合ステップを実行し、

前記第1の調整ステップで、前記第2の分類結果に形成された意味的な構造を前記第1

10

20

の分類結果に追加し、

前記第2の調整ステップで、前記第1の分類結果においてクリック頻度の高いクエリを疑似概念として前記第2の分類結果に追加し、

前記意味的な構造が、オントロジに基づく分類方法によって生成されることを特徴とするオブジェクト分類方法。

【請求項2】

前記オブジェクト取得手段が備える入力手段が、ターゲットクエリを入力する入力ステップを実行し、

前記オブジェクト取得手段が備える検索手段が、前記ターゲットクエリに従ってオブジェクトデータベースを検索し、検索結果を分類すべきオブジェクト集合として出力する検索ステップを実行する

ことを特徴とする請求項1に記載のオブジェクト分類方法。

【請求項3】

前記オブジェクトデータベースの検索が、キーワードに基づく情報検索方法を利用することを特徴とする請求項2に記載のオブジェクト分類方法。

【請求項4】

前記クエリログベース分類手段が備えるクエリログ取得手段が、クエリログを取得するクエリログ取得ステップを実行し、

前記クエリログベース分類手段が備える関連クエリ抽出手段が、前記クエリログからターゲットクエリに関する全てのクエリを抽出する関連クエリ抽出ステップを実行し、

前記クエリログベース分類手段が備えるクラスターベースカテゴリ学習手段が、抽出された関連クエリをクラスタリングし、各クラスターのクラスター中心をオブジェクトカテゴリとして決定するクラスターベースカテゴリ学習ステップを実行し、

前記クエリログベース分類手段が備える分類手段が、全ての検索結果を決定したカテゴリへ分類する分類ステップを実行する

ことを特徴とする請求項2に記載のオブジェクト分類方法。

【請求項5】

前記オントロジ情報ベース分類手段が備えるオントロジ情報取得手段が、前記オントロジ情報を取得するオントロジ情報取得ステップを実行し、

前記オントロジ情報ベース分類手段が備えるオブジェクト注釈手段が、前記オントロジ情報を参照して検索結果である全ての前記オブジェクトに注釈を付するオブジェクト注釈ステップを実行し、

前記オントロジ情報ベース分類手段が備えるクエリ注釈手段が、前記ターゲットクエリに注釈を付するクエリ注釈ステップを実行し、

前記オントロジ情報ベース分類手段が備えるカテゴリ生成手段が、前記オントロジ情報における、注釈を付されたターゲットクエリの意味的な関連概念に従ってオブジェクトカテゴリを生成するカテゴリ生成ステップを実行し、

前記オントロジ情報ベース分類手段が備える分類手段が、全ての検索結果を、生成されたカテゴリへ分類する分類ステップを実行する

ことを特徴とする請求項2に記載のオブジェクト分類方法。

【請求項6】

コンピュータ上で構成されるオブジェクト分類システムであって、

オブジェクト集合を取得するオブジェクト取得手段と、

クエリログに基づいてオブジェクトを分類し、第1の分類結果をクエリリストの形式で生成するクエリログベース分類手段と、

オントロジ情報に基づいてオブジェクトを分類し、第2の分類結果を概念リストの形式で生成するオントロジ情報ベース分類手段と、

前記第1と第2の分類結果を意味的に融合し、最終的な分類結果を生成する意味的融合手段とを備え、

前記意味的融合手段が、

10

20

30

40

50

前記第 2 の分類結果に従って前記第 1 の分類結果を調整する第 1 の調整手段と、
前記第 1 の分類結果に従って前記第 2 の分類結果を調整する第 2 の調整手段と、
調整後の前記第 1 の分類結果及び調整後の前記第 2 の分類結果を統合して最終的な分類
結果を生成する統合手段とを含み、

前記第 1 の調整手段が、
前記第 2 の分類結果に形成された意味的な構造を前記第 1 の分類結果に追加し、
前記第 2 の調整手段が、
前記第 1 の分類結果においてクリック頻度の高いクエリを疑似概念として前記第 2 の分
類結果に追加し、

前記意味的な構造が、オントロジに基づく分類方法によって生成される
ことを特徴とするオブジェクト分類システム。

10

【請求項 7】

オブジェクトを記憶するオブジェクトデータベースを備え、
前記オブジェクト取得手段が、
ターゲットクエリを入力する入力手段と、
前記ターゲットクエリに従って前記オブジェクトデータベースを検索し、検索結果を分
類すべきオブジェクト集合として出力する検索手段とを含むことを特徴とする請求項 6 に
記載のオブジェクト分類システム。

【請求項 8】

前記クエリログベース分類手段が、
前記クエリログを取得するクエリログ取得手段と、
前記クエリログから前記ターゲットクエリに関する全てのクエリを抽出する関連クエリ
抽出手段と、

20

抽出された関連クエリをクラスタリングし、各クラスタのクラスタ中心をオブジェクト
カテゴリとして決定するクラスターベースカテゴリ学習手段と、

全ての検索結果を決定したカテゴリへ分類する分類手段とを含むことを特徴とする請求
項 7 に記載のオブジェクト分類システム。

【請求項 9】

前記オントロジ情報ベース分類手段が、
前記オントロジ情報を取得するオントロジ情報取得手段と、
前記オントロジ情報を参照して検索結果である全ての前記オブジェクトに注釈を付する
オブジェクト注釈手段と、

30

前記ターゲットクエリに注釈を付するクエリ注釈手段と、
前記オントロジ情報における、注釈を付されたターゲットクエリの意味的な関連概念に
従ってオブジェクトカテゴリを生成するカテゴリ生成手段と、

全ての検索結果を、生成されたカテゴリへ分類する分類手段とを含むことを特徴とする
請求項 7 に記載のオブジェクト分類システム。

【発明の詳細な説明】

【技術分野】

【0001】

40

本発明は、一般に情報検索およびオブジェクト（例えば、ドキュメント、検索結果等）
 分類に関し、特に、オブジェクト分類と組織化のためにクエリログに基づいた分類結果お
 よびオントロジ情報に基づいた分類結果を相互に利用する自動オブジェクト分類のための
 方法およびシステムに関する。

【背景技術】

【0002】

インターネットの普及を背景とした電子情報の爆発的増加に伴い、大量で多様な情報が
 ウェブ上に集積され、今なお信じがたいほどの速度で増加を続けている。ネット市民にと
 って、この膨大な情報の集積から有用な情報を見つけ出すことは非常に困難な作業である

50

【0003】

情報検索（IR）とは、オブジェクト集合（例えば、ドキュメント）内の情報を検索するための技術であり、さらに（１）ドキュメントに含まれる特定情報の検索、（２）ドキュメントそのものの検索、（３）ドキュメントを説明するメタデータの検索、（４）独立型リレーショナルデータベースや、インターネット、イントラネットのようなハイパーテキストでネットワーク化されたデータベース内のテキスト、音声、画像、データの検索に分類することができる。長い歴史を持つこの検索分野から出現したウェブ検索エンジン（例：Google、百度（Baidu））は、ウェブ上で情報を見つけるためのドキュメント検索システムであり、特定の基準を満たすコンテンツ（典型的には、与えられた単語または語句を含むコンテンツ）を検索し、基準に一致する項目リストを取り込むことができる。

10

【0004】

オブジェクト分類は、予め定義された集合から主題カテゴリを有するオブジェクト（例えば、ドキュメントや自然言語テキスト）をラベル付けする作業であり、例えば、語義曖昧性解消、ドキュメント編成、テキストフィルタリングおよびウェブページ検索等のIRおよびテキストデータマイニングの多くの利用形態に適用することが可能である。オブジェクトクラスタリングは、異なるグループへのオブジェクトの分類である。正確には、各部分集合のドキュメントがある共通の特徴を共有するように、部分集合（クラスター）内へドキュメント集合等のオブジェクト集合を区分化することである。

【0005】

これらの一般の検索エンジンからの返却結果が大量であることを考慮すると、ウェブ利用者が本当に望むものを探し出すことはなお一層困難である。オブジェクトクラスタリング/分類の手法は、検索結果を整理して利用者に関連ドキュメントに素早く導く効果的な方法を可能するという大きな可能性を提供する。

20

【0006】

上述したように、電子メディアコンテンツの急成長に伴って、検索エンジン（ウェブページあるいはデスクトップ・ドキュメントのために）は、人々が有益な情報を見つけ出すのを支援するにあたり重大な役目を果たす。しかしながら、大量の返却結果（それらは多くの場合話題とジャンルにおいて種々雑多である）は、関心のある情報を見つけ出す利用者にとって大きな重荷になるであろう。

30

【0007】

関連技術として、多くの既存の自動情報分類アルゴリズムが存在する。

例えば、論文、シュエンホイ ワーン、チュヨンシアーン ジャイ（XuanHui Wang, ChengXiang Zhai）による、「Learn from Web Search Logs to Organize Search Results（検索結果を整理するためのWeb検索ログからの学習）」、SIGIR2007, pp.87-94（非特許文献1）においては、検索結果を検索エンジンログから学習した特徴によって整理するという、検索結果分類方法が提供されている。他の例としては、特開2005-182280号公報（特許文献1）が、検索結果を整理する他の方法を提供している。この方法は、まず、あらかじめ記憶されたオントロジ情報に基づいてオブジェクトカテゴリを抽出し、次に、抽出されたカテゴリに従って検索結果を整理する。

40

【先行技術文献】

【特許文献】

【0008】

【特許文献1】特開2005-182280号公報

【非特許文献】

【0009】

【非特許文献1】シュエンホイ ワーン、チュヨンシアーン ジャイ（XuanHui Wang, ChengXiang Zhai）「Learn from Web Search Logs to Organize Search Results（検索結果を整理するためのWeb検索ログからの学習）」、SIGIR2007, pp.87-94

【発明の概要】

50

【発明が解決しようとする課題】**【0010】**

クエリログに基づいたオブジェクト分類方法は、カテゴリ選択において背景知識（すなわち、オントロジ）を考慮に入れていない。このため、分類精度が十分に良いとは言えない。さらに、解決方法が履歴情報にとっても依存するので、発見されたカテゴリ情報が利用者にとって分かりにくい可能性ある。このため、分類結果がユーザにとって使い勝手の良いものとは言えない。

【0011】

一方、オントロジ情報に基づくオブジェクト分類方法については、予め格納されているオントロジ情報に限定されるため、オントロジに基づく分類方法のカテゴリ集合による検索結果は柔軟性がなく、ユーザの関心の変化を反映することができない。

10

【課題を解決するための手段】**【0012】**

本発明は、関連技術における上述の課題を解消するためになされたものである。本発明のオブジェクト分類方法は、オブジェクト（例えば、ドキュメントまたは検索結果）分類の質を向上させるために、オントロジ情報によって提供される背景知識をクエリログに含まれる履歴情報と組み合わせる。

【0013】

本発明によるオブジェクト分類方法は、オブジェクト集合を取得するステップと、クエリログに基づいてオブジェクトを分類し、第1の分類結果を生成するステップと、オントロジ情報に基づいてオブジェクトを分類し、第2の分類結果を生成するステップと、第1と第2の分類結果を意味的に融合し、最終的な分類結果を生成するステップとを有する。

20

【0014】

本発明によるオブジェクト分類システムは、オブジェクト集合を取得するオブジェクト取得手段と、クエリログに基づいてオブジェクトを分類し、第1の分類結果を生成するクエリログベース分類手段と、オントロジ情報に基づいてオブジェクトを分類し、第2の分類結果を生成するオントロジ情報ベース分類手段と、第1と第2の分類結果を意味的に融合し、最終的な分類結果を生成する意味的融合手段とを含む。

【0015】

本発明が提供するオブジェクト分類方法は、主に、（1）クエリログに基づいたオブジェクト分類、（2）オントロジに基づいたオブジェクト分類、（3）上記の2つの結果の意味的な組合せ、の3つのステップを含む。

30

【0016】

第1に、クエリログに基づいたオブジェクト分類方法においては、参考文献1に述べられるように、検索エンジンクエリログが検索結果の潜在的な特徴（カテゴリ集合）を表わす関連クエリを記憶しているので、この方法は、検索エンジンログから学習された特徴によって検索結果を整理する。まず、利用者はクエリログから関連クエリを抽出する。次に、それらの関連クエリをクラスタリングし、クラスタ中心を潜在的な特徴として見なす。最後に、全ての検索結果を対応するカテゴリに分類する。

【0017】

40

第2に、オントロジに基づくオブジェクト分類方法に関しては、特許文献1に述べられるように、オントロジ情報に反映された背景知識は、利用者にとって非常に分かり易いので、この方法は、オントロジから抽出された特徴に検索結果を分類する。まず、オントロジ情報に従って、利用者は、オブジェクト（例えば、ドキュメント）および入力クエリすべてに注釈を付ける。次に、特徴（カテゴリ集合）を意味的な結合性解析に基づいて生成する。最後に、全ての検索結果を対応するカテゴリに分類する。

【0018】

最後に、意味的な融合ステップは、以下の3つのケースを含む。
1) オントロジに基づく方法の結果によって、クエリログに基づく方法の結果を調整するか、あるいは改善する。

50

2) クエリログに基づく方法の結果によって、オントロジに基づく方法の結果を調整するか、あるいは改善する。

3) 上記2つの結果の特徴を組み合わせることで最終的なオブジェクト分類結果を取得する。

【発明の効果】

【0019】

本発明によれば、分類精度は向上することができる。また、ユーザにとって使い勝手の良い分類結果の表示を提供することができる。

【0020】

適切なオブジェクトカテゴリ集合がない状態での、カテゴリに対する分類精度は、ある程度において無意味である。クエリログに基づいた方法によって生成された精度の安定しないカテゴリ集合に、オントロジに基づく分類方法によって生成される意味的な構造 (framework) を追加することにより、本発明は、オントロジ知識によって修正され、ユーザのクエリ/検索履歴を反映するオブジェクトカテゴリ集合を動的に生成し、それによって分類精度を向上させる。

【0021】

さらに、利用者がオントロジ情報の背景知識を十分に理解できるため、本発明は、ユーザにとって使い勝手の良い検索結果の表示を提供する。

【0022】

オントロジに基づく方法とクエリログに基づく方法からの結果間の意味的な整合は、クラスタリング結果が利用者の関心の変化を反映することを可能にし、それにより、カテゴリの柔軟性が向上する。

【0023】

本発明の前述した特徴と他の特徴及び効果は、添付図面と共に以下の説明を参照することによりよく理解されるであろう。本発明の範囲が、ここに説明された例あるいは特定の実施の形態に限定されないことは言うまでもない。

【図面の簡単な説明】

【0024】

本発明の前述した特徴及び他の特徴は、以下の説明と添付図面とを参照することにより、さらに明らかになるであろう。

【図1】本発明の実施の形態によるオブジェクト分類システム100の内部構成を示すブロック図である。

【図2】図1に示すシステム100の動作手順の例を説明するフローチャートである。

【図3】図1に示されるシステムにおけるクエリログベース分類手段102のより詳しい内部構成を示すブロック図である。

【図4】クエリログに基づいたオブジェクト分類手順を説明するための概略図である。

【図5】図1に示されるシステムにおけるオントロジ情報ベース分類手段103のより詳しい内部構成を示すブロック図である。

【図6】オントロジ情報に基づくオブジェクト分類手順を説明するための概略図である。

【図7】クエリログに基づく方法の分類結果をオントロジに基づく分類方法の結果を参照して調節する際の第1の意味的な融合手順を示すフローチャートである。

【図8】図7に示される第1の意味的な融合手順と結果をより詳細に説明するための概略図である。

【図9】図7に示される第1の意味的な融合手順と結果をより詳細に説明するための概略図である。

【図10】クエリログに基づく分類方法の結果を参照してオントロジに基づく方法の分類結果を調整する際の第2の意味的な融合手順を示すフローチャートである。

【図11】図10に示される第2の意味的な融合手順と結果をより詳細に説明するための概略図である。

【図12】図10に示される第2の意味的な融合手順と結果をより詳細に説明するための概略図である。

10

20

30

40

50

【図13】最終的な意味的な融合結果を示すための概略図である。

【発明を実施するための形態】

【0025】

以下、本発明の実施の形態について図面を参照して説明する。以下に述べる実施の形態は例として説明するものであることは理解されるであろう。ただし、本発明の範囲は、以下に説明するどの特定の実施の形態にも限定されない。

【0026】

本発明は自動オブジェクト分類に関する。説明を簡単にするために、検索結果としてドキュメントを例に用いて、本発明による方法とシステムを詳述する。もちろん、本発明はこの例に限定されず、他のオブジェクト分類の関連分野に、より広範囲に適用できることは、当業者にとって容易に理解できるであろう。

【0027】

図1は、本発明の実施の形態によるオブジェクト分類システム100の内部構成を示すブロック図である。図示のように、オブジェクト分類システム100は、主として、オブジェクト取得手段101、クエリログベース分類手段102、オントロジ情報ベース分類手段103および意味的融合手段104を含む。クエリログベース分類手段102は、クエリログ記憶装置106に記憶された検索エンジンのクエリログを参照することにより、クエリログに基づいたオブジェクト分類を実行する。また、オントロジ情報ベース分類手段103は、オントロジ情報記憶装置107からオントロジ情報を参照することにより、オントロジ情報に基づいたオブジェクト分類を実行する。例えば、前述の非特許文献1および特許文献1は、クエリログに基づく分類方法とオントロジに基づく分類方法の2つの例をそれぞれ提供する。もちろん、本発明が、非特許文献1および特許文献1に記載された分類方法にのみ限定されるものでなく、当業界において知られている他のクエリログに基づく分類方法やオントロジに基づく分類方法のいずれにも適用できることは当業者であれば理解できるであろう。

【0028】

図1の例において、利用者に使用可能な全てのオブジェクト（例えば、ドキュメント）はオブジェクトデータベース105に記憶されている。利用者は、入力ユニット1011を介してターゲットクエリ(target query)を入力する。次に、検索ユニット1012は、オブジェクトデータベース105において従来のキーワードに基づく情報検索を実行し、分類すべきオブジェクト集合として、逆索引による検索結果を出力する。

【0029】

クエリログベース分類手段102による分類結果およびオントロジ情報ベース分類手段103による分類結果は、クエリリストの形式および概念リスト(Concept List)の形式でそれぞれ出力され、意味的な組合せのために意味融合手段104に提供される。最後に、意味融合手段104は、クエリリストおよび概念リスト(Concept List)を調整し、最終的な分類結果を出力する。

【0030】

図2は、図1に示すシステム100の動作手順の具体例を示すフローチャートである。まず、ステップ201において、利用者が入力ユニット1011を介してターゲットクエリを入力する。ステップ202において、検索ユニット1012は、オブジェクトデータベース105を検索し、分類するためのオブジェクト集合を取得する。ステップ203において、クエリログベース分類手段102は、クエリログに基づいて入力されたオブジェクト集合（例えば、ドキュメント集合）を分類し、第1の分類結果（すなわち、クエリリスト）を生成する。ステップ204において、オントロジ情報ベース分類手段103は、オントロジ情報に基づいてドキュメントを分類し、第2の分類結果（すなわち、概念リスト）を生成する。最後に、ステップ205において、意味融合手段104は、第1と第2分類結果を意味的に融合させ、最終的な分類結果を生成する。その後、手順200が終了する。

【0031】

以下に、クエリログに基づいたオブジェクト分類手順について、図3および図4を参照して説明する。

【0032】

図3に示すように、クエリログベース分類手段102は、クエリログ取得ユニット301、関連クエリ抽出ユニット302、クラスターベースカテゴリ学習ユニット303および分類ユニット304を含む。クエリログ記憶装置106は、検索エンジンのクエリ履歴、すなわち、利用者によって入力された全てのクエリと利用者のクリックスルー情報(click through information)(以下、疑似ドキュメントと称する)を記憶する。各クエリは、利用者によってクリックされたクエリから導かれた検索結果と利用者のクリック時間などのような他の情報を記録する1つの疑似ドキュメントに対応している。

10

【0033】

まず、クエリログ取得ユニット301は、クエリログ記憶装置106に記憶されたクエリログを取得する。次に、関連クエリ抽出ユニット302は、利用者によって入力されたターゲットクエリとクエリログ取得ユニット301によって取得されたクエリログ内の疑似ドキュメントとの間の類似度に従って関連クエリを抽出する。次に、クラスターベースカテゴリ学習ユニット303は、全ての関連クエリをクラスタリングし、各クラスターの中心をカテゴリとして出力する。これらのオブジェクトカテゴリは、入力されたターゲットクエリで示される利用者の関心に対応しているはずである。例えば、図4に示すように、ターゲットクエリとして「War Room」を入力した例において、クラスタリング結果は、例えば、図の左側に示すように生成される。また、図の右側に示すクエリリストにおいて、左側のクラスタリング方式から生成されるクラスターの中心となる、War Room Plan、War Room Spec、Information Management、Tag、Ontology等がオブジェクトカテゴリとして使用される。

20

【0034】

ターゲットクエリに関して、1つの関連クエリだけ、すなわちターゲットクエリそれ自体が抽出される他のケースがある。この場合、各オブジェクトカテゴリの名前は、返された結果を単に統計的に分析することにより生成することが可能である。例えば、ターゲットクエリ「War Room」に対して返された合計100のドキュメントがあると仮定する。100のドキュメントについてのタイトル分析と頻出単語の統計値によって、3つのカテゴリ名、すなわち、「Desktop System」、「Ontology」、「Automatic Office System」が取得されるかもしれない。この3つのカテゴリは以下のオブジェクト分類のために使用することができる。もちろん、クエリに基づくオブジェクトカテゴリ生成方法は、クエリリストあるいは単一のクエリに基づく上述した例に限定されない。当業者にとって周知である他の関連技術を、本発明に適用することも可能である。

30

【0035】

分類ユニット304は、異なるカテゴリにオブジェクト(例えば、検索結果)を分類するために既存の分類方法を利用することが可能である。例えば、分類ユニット304は、検索結果のTF-IDFベクトルと、あるカテゴリにおけるドキュメントの全ベクトルの平均との間のコサイン類似度に従って異なるカテゴリに検索結果を分類することができる(すなわち、重心に基づく方法)。

40

【0036】

図5と図6は、オントロジ情報に基づくオブジェクト分類方法およびその動作原理をより詳細に示している。図5に示すように、オントロジ情報ベース分類手段103は、オントロジ情報取得ユニット501、オブジェクト注釈ユニット502、クエリ注釈ユニット503、カテゴリ生成ユニット504および分類ユニット505を含む。オントロジ情報記憶装置107は、2つの部分、すなわち背景知識記憶データベース1071と、メタデータ記憶データベース1072に分けられる。図5に示す例において、まず、オントロジ情報取得ユニット501は、背景知識記憶データベース1071から背景知識を取得し、それを、オブジェクト(例えば、ドキュメント)の意味的注釈(semantic annotation)のためにオブジェクト注釈ユニット502に提供する。オブジェクト注釈ユニット502

50

は、外部からオブジェクト集合を取得し、オントロジ情報取得ユニット501によって入力された背景知識を参照して、受信したオブジェクトに意味的な注釈を付け、意味的な注釈結果として生成されたメタデータを、後で利用するためにメタデータ記憶データベース1072に格納する。ドキュメントなどのようなオブジェクト集合についてオブジェクト注釈ユニット502によって実行された意味的な注釈手順は、バックエンドで実行され、背景知識で定義されたエンティティとドキュメントに現れるエンティティとの間のリンクを記録する。

【0037】

図6のフローチャートを参照すると、クエリ注釈ユニット503は、ステップ601でターゲットクエリで提示されたキーワード（例えば、「War Room」）を取得し、そのキーワードに意味的に注釈を付け、背景知識のエンティティとターゲットクエリにおける関心のあるエンティティとの間の関連を特定する（ステップ602）。例えば、クエリ注釈ユニット503は、エンティティ「War Room」を背景知識における概念「project」の例であると認定する。次に、ステップ603において、カテゴリ生成ユニット504は、オントロジの注釈を付けられたターゲットクエリの意味的な関連概念に基づいて、背景知識記憶データベース1071に記憶された背景知識から適切なオブジェクトカテゴリ集合を抽出する。例えば、背景知識によって、カテゴリ生成ユニット504は、概念「project」が概念「Time」、「People」、「Team」等と関連すると認定し、それにより、可能性のあるオブジェクトカテゴリとしてこれらの概念を抽出する。このようにして、カテゴリ生成ユニット504は、オブジェクト分類のために図6に示すようなConcept List 1（概念リスト）を生成する。最後に、ステップ604において、分類ユニット505は、背景知識を媒体として利用して、オブジェクト注釈ユニット502によってメタデータ記憶データベース1072に記録されたそれぞれのオブジェクトに関するメタデータを参照して意味的な結合性解析（connectivity analysis）を行い、オブジェクト分類のために適切なカテゴリを見つけ出す。

【0038】

クエリログに基づくオブジェクト分類方法およびオントロジに基づくオブジェクト分類方法の例について、図3、4および図5、6を参照して詳細に説明した。上述したように、クエリログに基づく分類方法は背景知識を考慮に入れないため、分類精度が十分に良いとは言えない。また、利用者にとっての可読性も悪い。一方、オントロジに基づく分類方法によって生成されるカテゴリ集合は、柔軟性がなく、利用者の関心の変化を反映することができない。

従って、オブジェクト分類の精度と利用者のfriendnessを改善するために、2つの方法を相互に組み合わせる統合方法を提供する必要がある。

【0039】

本発明の実施の形態によれば、図1に示すシステム100における意味融合手段104の動作手順は、主に以下の3つの特徴を含んでいる。

- 1) オントロジに基づく方法の結果に従ってクエリログに基づく方法の結果を調整すること。
- 2) クエリログに基づく方法の結果に従ってオントロジに基づく方法の結果を調整すること。
- 3) 前述した調整した2つの結果を結合して最終的なオブジェクト分類結果を生成すること。

【0040】

以下、意味融合手段104の意味的な融合処理の動作原理について、図7から図3を参照して詳細に説明する。

【0041】

図7は、第1の意味的な融合処理を示している。この処理において、クエリログに基づいた方法の分類結果（すなわち、クエリリスト）は、オントロジに基づいた方法の分類結果（すなわち、概念リスト）に従って調整される。図8および図9は、図7の第1の意味

10

20

30

40

50

的な融合処理を示す概略図である。

【0042】

図7に示すように、意味融合手段104は、まず、クエリログベース分類手段102およびオントロジ情報ベース分類手段103から、第1の分類結果(すなわち、クエリリスト)と第2の分類結果(すなわち、概念リスト)を受け取る。例えば、クエリリストと概念リストの構造は、図8に示すような構造である。次に、ステップ701において、まず、クエリリストの各クエリと概念リストの各概念との間の類似度を計算する。例えば、図8の例において、クエリリストのクエリ「WarRoom Spec」に対応するドキュメントがそれぞれ取得され、利用者はそのメタデータ(Time、People、Team、・・・)を分析する。例えば、カテゴリ「WarRoom Spec」の90%のドキュメントが単語「Meng Xin」を含むこと、あるいは「Meng Xin」によって作成されている。このような方法で、クエリ「WarRoom Spec」と概念リストの概念「Meng Xin」との間の類似度が、90%であることが判定される。同様に、クエリ「WarRoom Spec」と概念「Time」の下の副概念「Jan」、「Feb」、「Mar」との間の類似度は、それぞれ、15%、80%および5%と決定される。次に、ステップ702において、計算された類似度が所定の閾値を越えているかどうかを判定する。超えていれば、第1の意味的な融合がステップ703において実行される。例えば、クエリ「WarRoom Spec」と概念「Meng Xin」は、概念「Meng Xin」の後に括弧を追加し、その括弧にクエリ「WarRoom Spec」を追加することにより融合させることができる。ステップ704において、括弧に含まれたクエリは、ステップ701において計算された類似度に従ってランク付けされる。その後、手順700が終了する。

10

20

【0043】

図9は、第1の意味的な融合処理の結果を示す。第1の意味的な融合の後、オントロジに基づく分類処理によって生成された概念リストの意味的な構造は、クエリログに基づく分類処理に生成されたクエリリストに追加され、その結果として、図9に示される、融合リスト1が得られる。この処理によって、オントロジ情報における関連概念が、クエリログに基づく方法の精度の安定しないカテゴリ集合に追加され、より精度の高いカテゴリ集合を生成でき、カテゴリへの分類精度を意味あるものに行うことができる。さらに、利用者がオントロジ情報の背景知識をよりよく理解することができるので、第1の意味的な融合を行えば、ユーザにとって分かりやすい分類結果の提示を提供することができる。

30

【0044】

まず、図10に示すように、意味融合手段104は、クエリログベース分類手段102およびオントロジ情報ベース分類手段103から、第1の分類結果(すなわち、クエリリスト)と第2の分類結果(すなわち、概念リスト)を受け取る。次に、ステップ1001で、ある期間内におけるターゲットクエリ(例えば、「WarRoom」)の概念に関連する全てのクエリが、オントロジ情報記憶装置107から収集される。ステップ1002で、これらのクエリに関する全ての疑似ドキュメントが収集される。ステップ1003で、それぞれのクエリに対応する疑似ドキュメントに関して、クエリリストの各カテゴリ(すなわち、クエリ)における検索結果のクリック回数を計算する。

例えば、図11に示すように、クエリリストにおけるクエリ「WarRoom Plan」、「WarRoom Spec」に対応する検索結果のクリック回数が30回、100回であると計算され、クエリリスト中に記録される。

40

次に、ステップ1004で、各クエリのクリック頻度が所定の閾値を越えてあるかどうかを判定する。超えていれば、ステップ1005で、すなわち、クリック頻度のより高いクエリカテゴリを概念リストへ追加する第2の意味的な融合が実行される。

例えば、図11の例において、クエリ「Tag」および「Information Management」のクリック頻度が比較的高いので、「WarRoom」を検索する場合、これらのクエリは疑似概念としてみなされ、概念リストに追加される。

次に、ステップ1006で、追加された疑似概念はそれらのクリック頻度に従ってランク付けされる。

50

【 0 0 4 5 】

図 1 2 は、第 2 の意味的な融合処理の結果を示す。

上述したように、クエリログに基づく分類方法によって生成されたクエリリストを分析することにより、クリック頻度の比較的高い、クエリ「 T a g 」および「 Information Management 」は、オントロジに基づく分類方法によって生成された概念リストに疑似概念として追加され、その結果、図 1 2 に示すような融合リスト 2 が得られる。多くの場合、オントロジ情報の記述はそれほど完全でないかもしれないし、利用者の関心の変化を反映できない可能性がある。第 2 の意味的な融合処理を利用することにより、クエリ履歴に基づいて生成された疑似概念が概念リストに追加することができる。従って、分類結果はより柔軟性のあるものになる。これにより、オントロジの柔軟性の無さを解決し、動的に利用者の関心の変化を反映することが可能となる。

10

【 0 0 4 6 】

最後に、前述の 1 番および第 2 の意味的な融合処理による結果は、最終的な分類結果を生成するために、さらに相互に組み合わせられる。図 1 3 は、最終的な意味的な融合処理の概略図を示す。最終結果は、例えば図 1 3 の融合 3 となる。2 つの意味的な融合結果の組合せを通して、オントロジに基づく分類結果とクエリログに基づく分類結果の調整および改善が実現される。

【 0 0 4 7 】

本発明によれば、オブジェクト（例えば、ドキュメントまたは検索結果）分類の精度およびユーザにとっての使い勝手（分かりやす）が改善される。

20

【 0 0 4 8 】

まず、ユーザに対する分かりやすさの観点から、オントロジ情報に基づいて生成された概念リスト（ C o n c e p t L i s t ）の意味的な構造を精度の安定しないクエリリストへ追加することにより、利用者は、関連するクエリの意味的な属性をより迅速に理解することが可能となる。関連する疑似概念を追加する手法は、オントロジの柔軟性のなさを改善し、利用者にとって最も関心が高くかつ関連性の高いクエリ結果を見つけ出すことが簡単となる。

【 0 0 4 9 】

さらに、精度の観点から、クエリログに基づく方法の精度が十分とはいえないカテゴリセットに、意味的な構造（ framework ）（オントロジにおける関連概念）を追加することは、よりよいカテゴリセットを生成し、カテゴリへの分類精度を意味あるものにするのを可能にする。さらに、どのような検索エンジンも最初はクエリログを有していないし、異なるドメインのクエリログは全く異なる。このため、本発明によるエンジンに直接利用することはできない。従って、オントロジに基づく分類方法は、最初の段階でのクエリログの欠如を補うことができ、それによって、オブジェクト分類の精度を改善できる。さらに、クリック頻度に従って各クラスのランクを調整することが、利用者の関心の変化を反映する。

30

【 0 0 5 0 】

上記では本発明を特定の実施例を参照して説明したが、本発明は図面に示した特定の構成や処理に限定されるものではない。上述した実施の形態において、幾つかの特定のステップを例として示しかつ説明している。しかしながら、本発明の方法の工程は、これらの特定のステップに限定されない。当業者であれば、これらのステップを変更し、修正し、補完することができること、そしていくつかのステップの順番を、本発明の精神および本質的な特徴から外れずに変更することができることを十分に理解するであろう。

40

【 0 0 5 1 】

本発明に要素は、ハードウェア、ソフトウェア、ファームウェアあるいはその組み合わせにおいて実現され、そのシステム、サブシステム、構成部品あるいはサブコンポーネントにおいて利用することができる。ソフトウェア中で実現された時、本発明に要素は、必要なタスクを実行するためのプログラム、あるいはコードセグメントである。プログラムまたはコードセグメントは、コンピュータ読み取り可能な媒体に格納するか、あるいは

50

は伝送ケーブルか通信リンク上の搬送波に包含されたデータ信号によって送信することが可能である。コンピュータ読み取り可能な媒体には、情報を格納するか転送することが可能であるすべての媒体を含む。コンピュータ読み取り可能な媒の具体例は、電子回路、半導体記憶装置、ROM、フラッシュ・メモリー、消去可能ROM (EROM)、フロッピー・ディスク、CD-ROM光ディスク、ハードディスク、光ファイバー媒体、無線周波数(RF)リンクなどを含む。コードセグメントは、インターネット、イントラネットなどのようなコンピュータネットワークを経由してダウンロードすることも可能である。

【0052】

上記では本発明を特定の実施例を参照して説明したが、本発明は上記の特定の実施例や、図面に示した特定の構成に限定されるものではない。例えば、図示した一部のコンポーネントは、互いに組み合わせて1つのコンポーネントとしたり、1つのコンポーネントを複数のサブコンポーネントに分割したり、他の既知のコンポーネントを追加することも可能である。また、動作プロセスも同様に、例に示されたものに限定されない。本発明はその精神と主要な特徴から逸脱することなく他の様々な形態で実装できることは、当業者によって理解されるであろう。従って、現在の実施の形態は、全ての点において例示でありかつ限定的でないとして考慮されるべきである。本発明の範囲は、前述の説明によってではなく添付された特許請求の範囲に示される。よって、請求項と同等の意味および範囲の内で生ずる変更は全て本発明の範囲に包含される。

【符号の説明】

【0053】

- 100：オブジェクト分類システム
- 101：オブジェクト取得手段
- 102：クエリログベース分類手段
- 103：オントロジ情報ベース分類手段
- 104：意味的融合手段
- 105：オブジェクトデータベース
- 106：クエリログ記憶装置
- 107：オントロジ情報記憶装置
- 1011：入力ユニット
- 1012：検索ユニット
- 301：クエリログ取得ユニット
- 302：関連クエリ抽出ユニット
- 303：クラスターベースカテゴリ学習ユニット
- 304：分類ユニット
- 501：オントロジ情報取得ユニット
- 502：オブジェクト注釈ユニット
- 503：クエリ注釈ユニット
- 504：カテゴリ生成ユニット
- 505：分類ユニット
- 1071：背景知識記憶データベース
- 1072：メタデータ記憶データベース

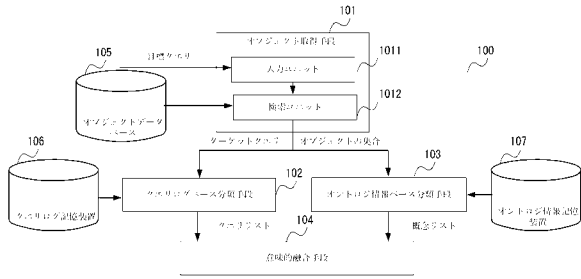
10

20

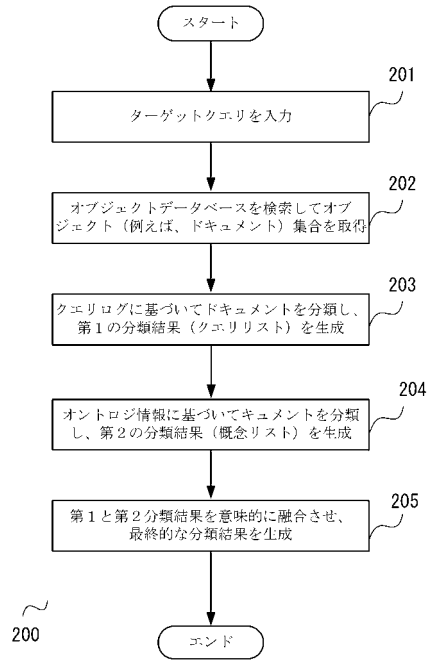
30

40

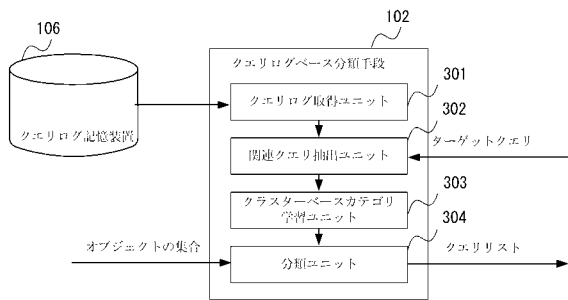
【図1】



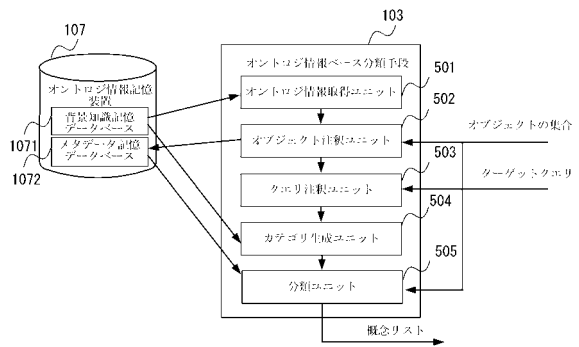
【図2】



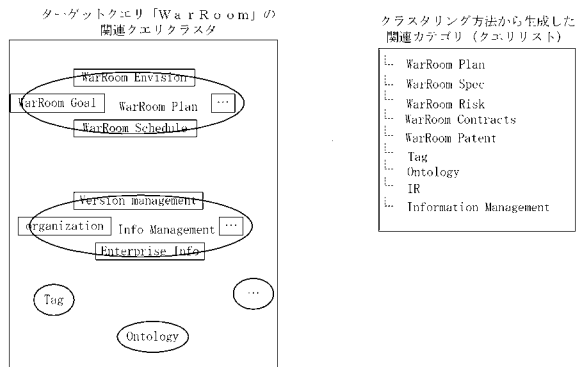
【図3】



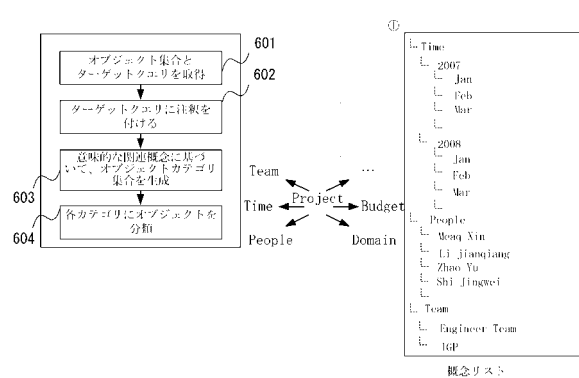
【図5】



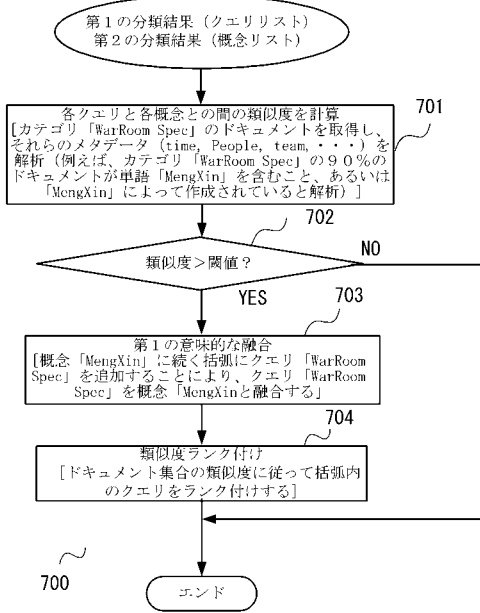
【図4】



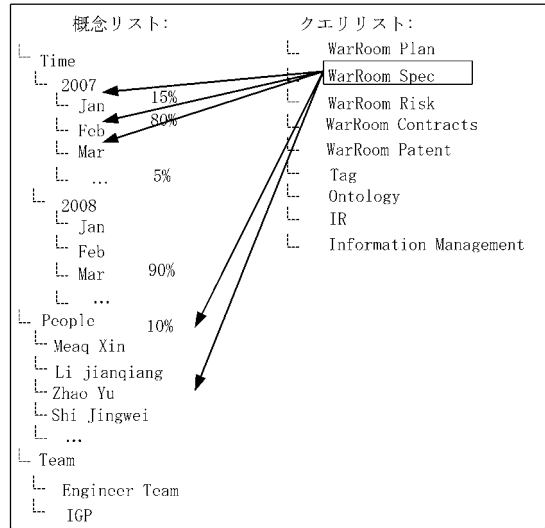
【図6】



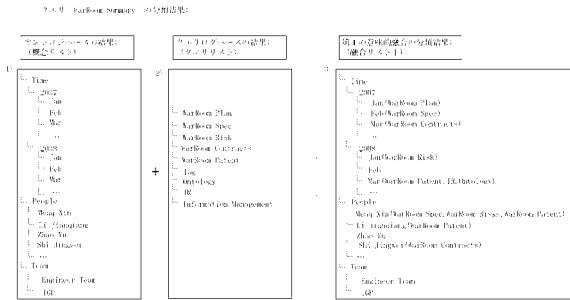
【図7】



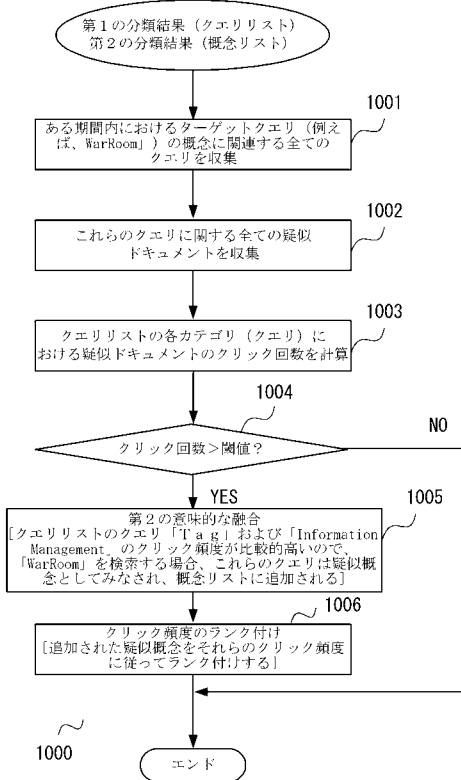
【図8】



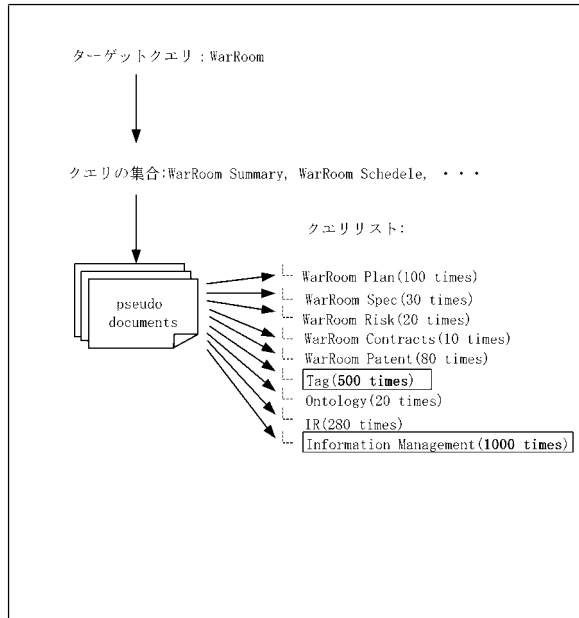
【図9】



【図10】

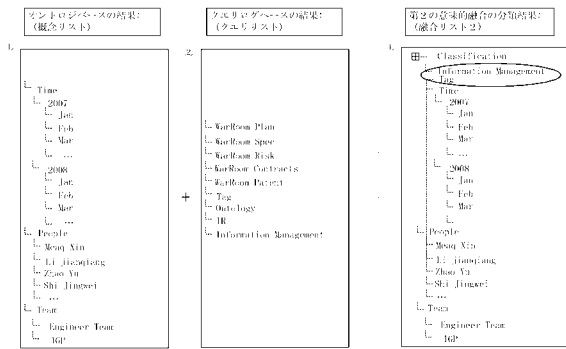


【図11】



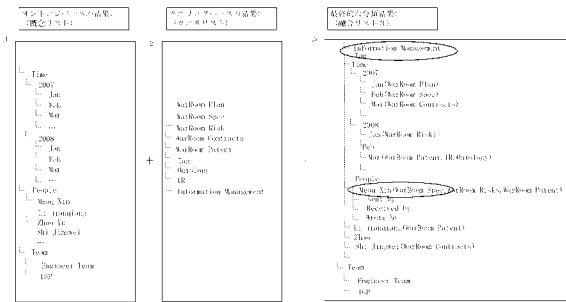
【 図 1 2 】

クエリ「WorkItem Summary」の分類結果:



【 図 1 3 】

クエリ「WorkItem Summary」の分類結果:



フロントページの続き

(72)発明者 メン シン

中華人民共和国 100084 ベイジン, チンファ サイエンス パーク, イノベーション
プラザ, ビルディング エー, 11エフ

(72)発明者 ジャオ ユウ

中華人民共和国 100084 ベイジン, チンファ サイエンス パーク, イノベーション
プラザ, ビルディング エー, 11エフ

(72)発明者 シー ジンウェイ

中華人民共和国 100084 ベイジン, チンファ サイエンス パーク, イノベーション
プラザ, ビルディング エー, 11エフ

審査官 高瀬 勤

(56)参考文献 特開2005-182280(JP,A)

特開2005-148843(JP,A)

市瀬 龍太郎, 分類をみつめなおす, 情報の科学と技術, 日本, 社団法人情報科学技術協会, 2
008年 2月 1日, 第58巻, 第2号, 第78-83ページXUANHUI Wang, CHENGXIANG Zhai, Learn from Web Search Logs to Organize Search Results,
[online], SIGIR2007, 2007年, P.87-94, Internet, <URL:http://dl.acm.org/citation.cfm?id=1277759>

(58)調査した分野(Int.Cl., DB名)

G06F 17/30

JSTPlus(JDreamII)