

連想マルチメディアリンクに基づく検索エンジンにおける概念ベース連携方式についての提案

3V-3

渡邊 拓二 西永 誠司

NTT ソフトウェア研究所

The proposal of enhanced DREAM(DiREctory for Associative Multimedia link) using concept base

Takuji Watanabe Seiji Nishinaga

NTT Software Laboratories

1.背景とねらい

近年のWWWの発展により、数多くの画像等のマルチメディア情報がWWW上を流通しており、その中からの確に情報を収集し、流通させるための検索技術が求められている。

この要求に応えるため、画像とキーワードによる複合的な検索環境を実現する、連想マルチメディアリンクに基づく検索エンジン(DREAM)を開発した。DREAMでは、画像間の類似度を定量化して評価するORDB(Object Relational Data Base)の機能を利用して、画像自体を検索キーとして用いることを可能にしている。また、画像が存在するコンテンツページよりHTMLのタグ情報を基に抽出した名詞を、画像に関連するキーワードリストとして登録しておき、入力した単語と一致するキーワードの個数により、キーワードと画像の類似度の重み付けを行うことで、総合的な類似度を算出している。

2.従来方式の問題点と提案方式

しかしこのキーワード評価方式では、登録されたキーワードリストと、入力された単語が一致する時のみキーワードとして有効なため、表記の違いによってキーワードが一致しないことがある。そこでキーワードと画像の類似度の正確性を高める為には、類義語なども含めて大量のキーワードを登録しておく必要があり、登録の更新などのキーワード処理時間が非常に大きくなってしまいう問題点があった。

そこでこの問題を解決するために、単語間の類似度を定量的に計算できる概念ベースをキーワードの

評価に利用することにした。新しい方式では、画像のタイトルをキーワードとして登録し、検索キーとして入力された単語との類似度を概念ベースで提供されているライブラリ関数で計算する。そしてORDBにより算出した画像間の類似度と概念ベースより算出された類似度を正規化し、一定の係数をかけて加算したものを類似度とする。これによって大量のキーワードを管理することなく、キーワードと画像の類似度を定量的に評価することが可能になり、検索処理速度と検索精度の向上が期待される。

従来方式				概念ベース連携方式		
総合的類似度	類似度	キーワード類似度	キーワード類似度	総合的類似度	類似度	キーワード類似度
10.0=	1.10	2.00=	○一致	10.0=	1.10	1.00
18.0=	2.00	1.00=	○	18.0=	2.00	0.80
25.0=	3.00	1.00=	×不一致	25.0=	3.00	0.10
50.0=	10.00	0=	×	50.0=	10.00	0.01

図1. 従来方式と今回の方式の比較

3.提案方式の評価

概念ベースを利用した実験システムを試作し、①検索処理速度、②検索精度、という観点から適用評価を行った。モデルとして、登録されている画像の一部分を拡大した画像をキー画像として、元画像を検索する方法を使用した。

評価対象値として、検索処理速度は、クライアントからの検索要求を受け取ってから結果を返すまでの時間を測定した。検索精度は、利用者が一度に検索結果をみることができ満足する基準を、検索結果が類似度の上位10位以内に入ることと仮定して、検索結果が実際に10位以内に入る割合を使用した。

入力する単語は、10人の被験者に元画像を見せ、その画像から連想される語を答えてもらった。検索処理時間は10枚の画像に対して、5回ずつ測定した結果の平均を示し、検索結果は被験者それぞれについて10位以内に入った画像の割合の平均化を示した。

測定結果より、処理速度が約3.4倍、検索精度が約1.5倍向上したことが確認できた。

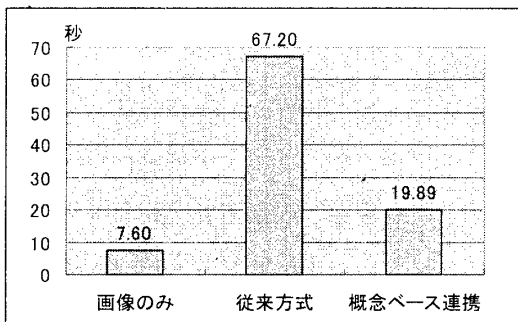


図2. 検索処理時間平均

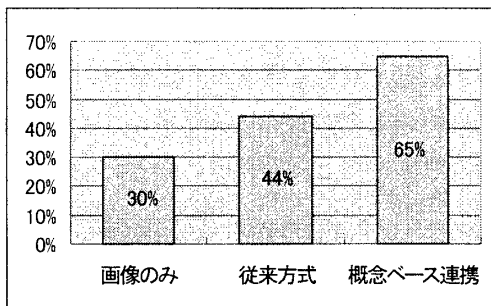


図3. 結果10位以内率

4. 考察

画像をキーとして用いる時、自ら画像を作り出すことは、実際に利用する立場から考えると手間がかかる上、検索精度的にもよくない。そこで今回は手元にキー画像が存在してそれを元に類似画像を検索するという条件を想定し、3項で述べたモデルを使用した。

この評価実験によって得られた検索結果は、図4のように、ほとんどの画像とキーワードの組み合わせで、従来方式より順位が上がっており、概念ベースの利用による効果が認められる。

しかしキーワードによっては、従来方式よりも順位が下がる場合も見受けられた。それらのキーワードは概して抽象的な語(たくさんの語と類似する語)の場合で、キーワードによる効果が表れにくかった

ためと思われる。この問題は、キーワードとして抽象的な語を入力しないよに注意を促すガイダンスなどによって対処可能と考えられる。

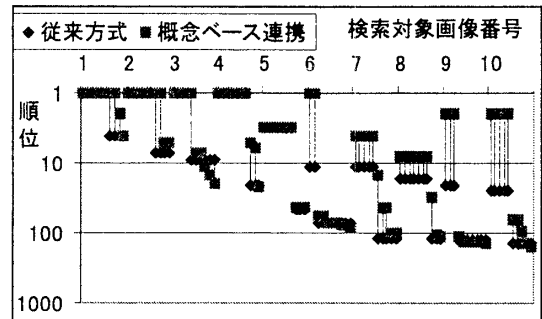


図4. 検索結果順位詳細データ

5. まとめと今後の課題

今回は DREAM と概念ベースを連携させたシステムを試作し、その有効性を確認できた。今後、より柔軟で精度の高い検索を実現するためには、複数キーワードによる絞り込みなどの機能拡充を行う必要があると考える。

またキーワードと画像の重み付け、入力した単語の情報などの画像個々への蓄積や、WWW上の新しい流れであるXML情報の利用、音声などの他のマルチメディア情報への対応など、利用者の立場に立ったより柔軟な検索環境の構築方法について検討を進める予定である。

〈参考文献〉

[1]吉田勝彦, 山本修一郎:「インターネット上のマルチメディアデータの柔軟な検索方式の提案」, 知能ソフトウェア工学研究会, 1997年11月
 [2]飯田敏幸, 松澤和光, 池上哲彦, 石野福弥, 今井賢一:「想起型情報検索システムについて」, システムソフトウェアとオペレーティング・システム.77-4, マルチメディア通信と分散処理.87-4, 1998年2月
 [3] 吉田勝彦, 山本修一郎, 赤間浩樹:「WebBASEによるイメージ画像検索システム:DREAM」, NTT技術ジャーナル, 1998年5月
 [4]INFORMIX:「Image DataBlade Guide」, 1994年6月