

検索結果の再構成によるナビゲーション支援*

2AC-2

巖寺 俊哲 鷲崎 誠司 菊井 玄一郎 林 良彦†

NTT 情報通信研究所‡

e-mail: {iwadera,suzaki,kikui}@isl.ntt.co.jp, hayashi@nttnly.isl.ntt.co.jp

1 はじめに

近年、WWW上の検索エンジンが、盛んに開発されている。このような検索エンジンは、通常、検索結果を検索要求と結果の類似度あるいは関連度に基づいて順位付けし出力する。本稿では、この検索結果を様々な観点（ビュー）から再構成することにより、ナビゲーションを支援する機能について報告する。

2 ナビゲーション支援

人間が、検索エンジン等を利用して、能動的な情報収集活動を行なう場合、ある程度明確化された情報要求を持っている場合が多い。この情報要求を基に、特定の情報に到達し（一次的検索）、さらにそこを起点として関連する情報を収集する（ブラウジング）という過程をとることが多いと考えられる。情報空間のナビゲーション過程とは、上記の2つの基本操作を繰り返す過程と考えられる[1]。

我々は、現在までにWWW検索システムTITANを開発してきた[1]。その中で、ナビゲーションを支援するために、多言語情報検索支援[2]、タイトル翻訳[2]、検索結果への注釈記述付与[1]、あるいは、ネットワークポロジ表示[3]によるブラウジング支援、GUIによる検索条件の調整[4]等の機能を提案してきた。多言語情報検索支援は、検索要求の記述言語と同一言語で記述された情報資源だけでなく、異なる言語で記述されたものをも検索可能とする。たとえば、日本語での検索要求を英語に変換することにより日英両言語に渡る検索を行なうことができる。タイトル翻訳は、検索結果のWWWページのタイトルを日本語へ変換する。これにより検索結果のタイトルを日本語で一覧することが可能となる。検索結果への注釈記述付与では、注釈情報をアイコン化しタイトルと共に表示する。付与する注釈は、各ページの発信元ドメイン（国）、テキスト情報、画像情報等の分類を示すメディアタイプ、固有情報を含む程度、リンク情報を含む程度を示す情報型/案内型分類である。これらの情報を基に、例えば、「アメリカから発信された画像情報にアクセスする」といったブラウジングの意志決定支援ができる。GUIによる検索条件の調整では、検索条件をベン図を用いて表示し、これをGUIで直接操作する。これにより適切な検索結果数を得られるように条件を直感的に調整できる。

次に検索結果を再構成することによりナビゲーションを支援する機能について述べる。

3 検索結果の再構成

3.1 機能概要

検索エンジンは、通常、検索結果を検索要求と結果の類似度あるいは関連度に基づいて順位付けし出力する。しかし、必要とするページが、高順位にランクされている保証はない。結局、検索結果すべてをタイトル等を参考に調べていく必要がある。この問題を解決する1

つの方法として、検索結果をある条件にしたがって整理して提示することが考えられる。

例えば、「桜の花見」に関する検索をした場合を想定する。検索結果は、これに関して何らかの関係があるページであり、タイトルのリストから次に閲覧するページを選択するための手がかりとなる情報が少ない。しかし、この結果を「見頃の時期」や「地域」などの条件によって整理して提示することにより、選択がより容易となる。

我々が提案する検索結果の再構成機能は、予め設定されている観点（ビュー）に従って検索結果を再構成することにより、手がかりとなる情報を提示し、ナビゲーションを支援する。複数のビューが設定可能であり、同一の検索結果を様々なビューから再構成することができる。また、再構成に使用する知識は、サンプルデータを利用することにより自動的に獲得することができる。このため容易にビューを追加することが可能である。

3.2 機能構成

検索結果の再構成機能は、(1)再構成知識獲得処理、と(2)検索結果再構成処理とにより実現されている。機能構成の概観を図1に示す。

あるビューからの検索結果の再構成は、そのビューに対応する分類体系に従って検索結果を分類するタスクと考えることができる。例えば、「桜の花見」に関する検索結果を「見頃」ビューで再構成することは、「3月下旬」、「4月上旬」といった項目別に分類することに対応する。分類タスクとみなすと上記、処理(1)は、分類知識の学習に、(2)は、分類処理に対応する。以下、各処理について述べる。

再構成知識獲得処理では、再構成に使用する知識をサンプルデータを用いて獲得する。例えば、「桜の花見」の「見頃」ビューに関する再構成知識は、予め「3月下旬」、「4月上旬」といった項目別に分類されているサンプルデータを用いて獲得される。学習は、 χ^2 検定の考え方を利用した手法（例えば、[5, 6]）を利用して行なう。学習結果は、サンプルデータ中のすべての異なり単語毎に各分類項目に対する特徴付けのスコアを表すテーブルとなる。従って、分類項目数が m 個、異なり単語数が n 語のとき、 $m \times n$ のテーブルとなる。

検索結果再構成処理では、前述した再構成知識獲得処理により作成されたスコア・テーブルを用いて検索結果を再構成する。再構成は、検索結果の各ページ毎に単語を抽出し、スコア・テーブルを用いて単語毎に項目別のスコアを付与する。さらに、このスコアを項目毎にページ中から抽出したすべての単語について合計し、最高スコアを持つ項目へそのページを分類する。

ここで用いているスコア・テーブルを変更することによって、同一の検索結果を様々なビューから再構成することが可能となる。

3.3 実行例

「桜の花見」に関する検索結果を「見頃」ビューで再構成した結果を図2に示す。再構成結果では、まず、各項目毎の件数が棒グラフで示される。さらに、その下に各項目別に検索結果のタイトルが表示される。また、棒グラフ中でクリックすることで項目を指定することにより、その項目のタイトルの表示位置へ直接移動することができる。

*Re-formulating retrieval results for navigating in information space

†Toshiaki IWADERA, Seiji SUSAKI Gen'ichiro KIKUI and Yoshihiko HAYASHI

‡NTT Information and Communication Systems Laboratories

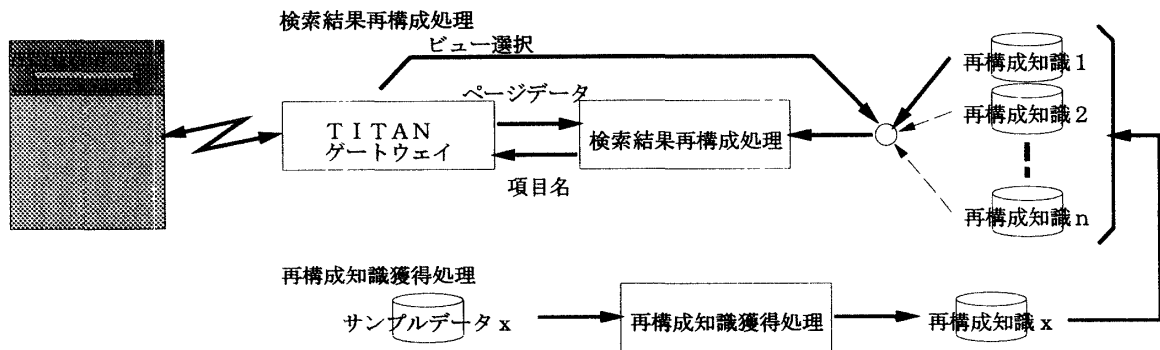


図1: 検索結果の再構成機能

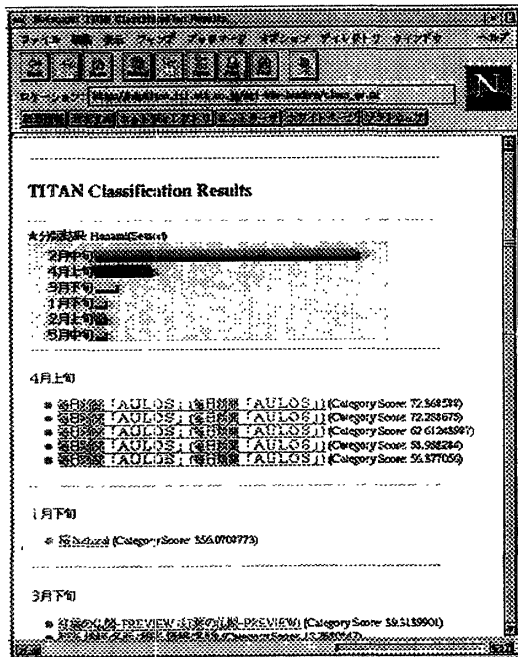


図2: 再構成結果

4 考察

検索結果を再構成する研究例として、Scatter/Gather[7]やSenseMaker[8]がある。前者は、検索結果を各ページを内容に従ってクラスタリングし、各クラスタに内容に合致したラベルを付与する。利用者は、必要なクラスタを選択する。システムは、選択されたクラスタをさらにクラスタリングする。これを繰り返すことにより検索結果を絞り込むことができる。後者は、例えば、同一のURLや類似のタイトルといった属性情報を用いて検索結果を整理し提示する。

クラスタリング手法では、クラスタリングの算法中で用いる予め決められた距離の定義に従ってクラスタが構成され、検索結果が整理される。従って距離の定義が固定的であり、同一の検索結果を様々な観点から整理できない。また、各クラスタを構成する基準が不明確であり、クラスタ間の相違がわかりづらい。これに対して、本機能では、同一の検索結果に対して複数のビューから再構成することが可能である。また、各分類項目は指定したビューに従って分離されており、項目間の相違がわかり易い。例えば、「桜の花見」の例では、同一検索結果を「見頃」や「地域」など複数のビューで整理することができる。また、「地域」ビューで再構成した場合は、各項目間の相違は「地域」に基づいていることが容易にわかる。

属性情報を用いる手法では、各属性毎にその属性値によって検索結果がまとめられる。このため、項目間の相違は明確であり、また、属性毎に整理することができる。しかし、この手法では、属性を自動的に追加する

機能が定義されておらず、追加が困難である。これに対して、本機能では、サンプルデータを利用することにより上記属性に対応するビューを容易に追加することができる。例えば、「桜の木の種類」ビューを追加する場合は、「桜の木の種類」毎に予め分類したサンプルデータを利用することにより容易にビューを追加できる。

5 おわりに

検索結果を再構成することによりナビゲーションを支援する機能について報告した。本機能により、同一検索結果を様々なビューから再構成することができる。また、再構成に用いる知識をサンプルデータから獲得することができ、容易にビューを追加することが可能である。今後は、再構成に必要な知識を予め学習するのではなく、検索結果から自動的に抽出する手法を検討する。

参考文献

- [1] 林良彦, 菊井玄一郎, 鷲崎誠司, 砂場 倫太郎. WWW 情報空間における resource discovery と navigation 支援. 電子情報通信学会技術研究報告 AI95-31, 1995.
- [2] Genichiro Kikui, Yoshihiko Hayashi, and Seiji Susaki. Cross-lingual information retrieval on the WWW. In *Proceedings of Multilinguality in Software Engineering: AI Contribution (MULSAIC-96)*, 1996.
- [3] Seiji Susaki, Yoshihiko Hayashi, and Gen'ichiro Kikui. Navigation interface in cross-lingual WWW search engine, TITAN. In *Proceedings of AUUG 96 and Asia Pacific World Wide Web*, 1996.
- [4] 田中智博, 菊井玄一郎, 奥雅博. 情報探索における GUI に関する一考察. 情報処理学会第51回全国大会, 1995.
- [5] 田村淳, 渡辺道枝, 原良憲, 笠原裕. 統計的手法による文書自動分類. pp. 1305-1306. 情報処理学会, March 1988. 第36回全国大会.
- [6] 河合敦夫. 意味属性の学習結果にもとづく文書自動分類方式. 情報処理学会論文誌, Vol. 33, No. 9, pp. 1114-1122, 1992.
- [7] Cutting Douglass R., Karger David R., Pedersen Jan O., and tukey John W. Scatter/Gather: A cluster-based approach to browsing large document collection. In *Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1992.
- [8] Michelle Q Wang Baldonado and Terry Winograd. Sensemaker: An information-exploration interface supporting the contextual evolution of a user's interests. In *Proceedings of CHI'97*, 1997.