

階層構造を持つタグを利用した探索的検索のための ユーザインタフェース

佐藤 進也[†]

日本工業大学[†]

1 はじめに

検索要求自体を明確化しながら必要とする情報への到達を目指すプロセスは探索的検索と呼ばれている [1]. クエリを修正しながら検索を繰り返す中で検索要求に関わるキーワードを発見し、それを手掛かりに得たかった情報に辿り着く、といった一連の検索行為もこの範疇に属する.

検索要求の明確化には検索対象への理解が必要である. 探索的検索の支援においても, 検索対象の理解を助けることがその機能として求められる [2]. 一般に, 対象をよりよく理解するためには, (i) 個々の事柄の詳細を把握する (ii) 関連する事柄の相互関係を把握する (iii) 把握する対象を広げる, といった行為が有効であると考えられる. 本研究では (ii), (iii) に焦点をあて, これらの支援を狙った探索的検索のためのユーザインタフェースを提案する. 提案手法では, 検索対象にタグが付与されており, そのタグの集合が階層構造を有していることを前提とし, その構造を支援に利用する. 具体例として, 提案手法を Wikipedia に適用した結果を示す.

2 関連研究

検索対象となる情報相互の関係理解を支援する方法の一つに, 検索結果を構造化して提示することが挙げられる. 構造化手法としては, クラスターリング [3], カテゴリ分類 [4], ファセットの利用 [5] など様々なアプローチが提案されている. Mirylenka らは検索結果の各文書を Wikipedia のカテゴリに関連付けし, そのカテゴリ名の集合から要約を生成してユーザの理解を支援している [6]. 要約の生成には Wikipedia カテゴリの階層構造を利用している.

我々の提案手法も, タグが付与されている検索結果を, そのタグどうしが持つ階層関係を利用して構造化している. この点で Mirylenka らの手法と類似しているが構造化の目的は異なるため得られる構造も異なったものとなる. 提案手法における構造化の狙いは要約を示すことではなく, 検索結果の全体と部分 (に対応するトピック) の関係を示すこと, そして, そのトピックを情報探索活動の範囲を広げる手段として提供することにある.

3 提案手法

3.1 検索結果の構造化

本節では, 図 1 に示した例を用いて検索結果の構造化手法について述べる. 図 1 左の木構造はタグの相互関係を示したものである. A, B などそれぞれタグであり, タグ間のエッジは上位/下位 (例えば A は B の上位) の関係を表している. いま, あるクエリ q で検索した結果, 文書集合 $R(q)$ が得られたとする. 図 1 左に示したタグの中で白抜き文字で表しているものがあるが, これは $R(q)$ のいずれかの要素に付与されたタグであること, すなわち, 検索結果に付与されたタグの分布を示している. この検索結果に付与されたタグを R タグと呼ぶことにする. この R タグの分布に基づき, 検索結果をテーブル形式 (図 1 右) に構造化する.

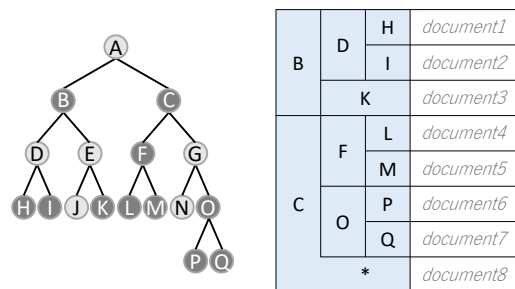


図 1 タグの階層関係を利用した検索結果の構造化

まず, 以下の手順により, タグの相互関係をテ

A User Interface for Exploratory Search Using Tags with Hierarchical Structure

[†] Shin-ya Sato, Nippon Institute of Technology

ブルヘッダとして表す. (1) 各 R タグ t を起点として下位方向に h ホップの範囲にあるタグの集合 $T(t)$ を作成する (例では $h=3$). (2) タグ集合のうち要素を共有するものを併合する (例では $T(C)$ と $T(O)$ を併合する). (3) 各タグ集合において R タグ以外を除去する. ただし, 複数の R タグの上位に位置するタグ (例えば D) は残す. (4) 得られたタグ集合に属する要素の階層関係を表形式で表わす.

次に, 検索結果 $R(q)$ の各要素 (図 1 の *document1* など) をそれに付与されたタグの行に配置する. 図では, *document8* が C の欄につながっている * で示した行に配置されているが, これはこの検索結果にタグ C のみが付与されている (言い換えれば, C ではあるが, F や G ではない) 状況を示している. 一方, 一つの検索結果に上位/下位関係にある複数のタグが付与されていて場合には, そのうちより下位のタグの行に配置する.

3.2 多様な情報へのアクセス支援

提案手法では, 構造化された検索結果を利用して, より多様な情報へのアクセスを支援する. 以下, その情報探索プロセスについて述べる. 図 2 は, 説明のために検索対象となる情報 (文書) の分布を図式的に示したものである. 簡単のため, 各文書にはタグ $\text{Tag}_1 \sim \text{Tag}_4$ のうち 1 つが付与されており, そのため, 検索対象が 4 つの部分に分けられているとする.

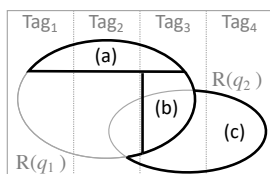


図 2 情報探索支援のシナリオ

この情報探索プロセスは, ユーザがクエリ q_1 で検索を行うことから始まる. ユーザに検索結果として示されるのは, q_1 に適合する文書のうち上位にランク付けされたものである (図で (a) として示した部分). この検索結果に対して提案手法による構造化を行いユーザに提示する. 提示されたタグの中からユーザが興味の対象となるタグ (この例では Tag_3) を選ぶことで, このトピックに関する情報 (図の (b)) に効率的にアクセスできる. さらに, q_1 と Tag_3 に関連する新たなクエリ候補 q_2 がユーザに提示される. ユーザは q_2 を手掛かりに関連情報 (図の (c))

にたどり着くことができる. なお, q_1 と Tag_3 から q_2 を導き出す方法については, 別稿に譲る.

4 Wikipedia への適用例

我々は, 提案手法を用いて Wikipedia 記事検索の実験システムを作成し, 有効性を検証している. 本実験ではタグとして Wikipedia のカテゴリを利用している. 図 3 は「うどん」で検索した結果の一部である. Web 検索エンジンの UI と同様に, 検索結果を選択 (クリック) することで Wikipedia の当該記事にアクセスできる. さらに, テーブルヘッダに表示されているタグを選択することでそのトピックに焦点をあてた情報探索を進めていくことができる.

日本の食文化	ご当地グルメ	高槻うどんギョーザ
	食文化に関する日本の組織	日本の外食事業者 ウエスト (飲食店)
		日本の食品メーカー 白雪食品
		* 日本うどん学会
	香川県の食文化	恐るべきさぬまうどん
日本の郷土料理	味噌煮込みうどん	

図 3 Wikipedia 記事検索への適用例

参考文献

- [1] R. W. White and R. A. Roth. *Exploratory Search: Beyond the Query-Response Paradigm*. Morgan & Claypool, 2009.
- [2] R. W. White, B. Kules, S. M. Drucker, and M. C. Schraefel. Supporting Exploratory Search — Introduction. *Comm. of ACM*, Vol. 49, No. 4, pp. 36–39, 2006.
- [3] D. Cutting, D. R. Karger, J. O. Pedersen, and J. W. Tukey. Scatter/Gather: A Cluster-based Approach to Browsing Large Document Collections. In *Proc. of ACM SIGIR*, pp. 318–329, 1992.
- [4] H. Chen and S. Dumais. Bringing Order to the Web: Automatically Categorizing Search Results. In *Proc. of ACM CHI*, pp. 145–152, 2000.
- [5] M. A. Hearst. Clustering versus Faceted Categories for Information Exploration. *Comm. of ACM*, Vol. 49, No. 4, pp. 59–61, 2006.
- [6] D. Mirylenka and A. Passerini. Navigating the Topical Structure of Academic Search Results via the Wikipedia Category Network. In *Proc. of ACM CIKM*, pp. 891–896, 2013.