

フロー情報収集・活用のための知的検索システム Fit

2 T-8

(1) コンセプト*

上田隆也 大谷紀子 伊藤史朗 柴田昇吾 池田裕治

キヤノン(株) 情報メディア研究所

1 はじめに

インターネットの普及により、NetNewsやWWWなどネットワーク経由でアクセス可能な情報が広く利用されるようになってきている。これらのうちには、新聞記事のように常に新しい情報が提供されるものも多い。こうした情報をフロー情報と呼ぶ[1]。

フロー情報の利用においては、従来の「必要が生じた時に文書データベースから検索する」という形態とは異なり、「継続的に新しい文書を収集する(収集フェーズ)とともに、必要な文書は保存しておいて後に活用する(活用フェーズ)」という形態が広く見られる。これにともない、検索システムも収集フェーズと活用フェーズを総合的に支援するものが必要である。

継続的な情報収集を考えた場合、1) 情報収集の目的が最初から明確であるとは限らない、2) 収集の過程で情報に対する見方が変化する、という問題がある。我々は、情報に対する視点に着目し、視点に基づいてフロー情報の収集・活用を支援する知的検索システム Fit(Flexible Information Tracking system)を開発している。本稿ではFitのコンセプトについて述べる。

2 フロー情報収集・活用における視点

2.1 視点

フロー情報の収集において、通常ユーザは複数の興味の対象に関して平行して文書を収集し、自分の見方に沿って文書を整理する。各々の興味の対象について、情報に対する見方が最初から明確に定まっているとは限らず、収集の過程で徐々に明確になっていくことが多い。新しい情報の発生によって新しい見方が生じる場合もある。一方、活用の際には、個々の文書でなく、ある興味の対象に関する文書をまとめて利用することが多い。

「興味の対象」や「情報の見方」をここでは視点として扱う。そして、収集フェーズではユーザが必要とする文書に視点を付与して保存し、活用フェーズでは視点を中心にして文書を利用するというモデルを考える。収集フェーズにおいては、新たな視点の発生や文書に付与した視点の変更がありうる。このように視点は時間

の経過にもよって変化するので、視点の表現には動的な枠組が必要である。

2.2 フロー情報の収集・活用に対する支援

視点を考慮すると、フロー情報の収集・活用の支援には次の三つの要件がある。

- 視点の変化に対応できる
- 収集時に視点の付与がユーザの負担にならない
- 活用時に視点を用いて精度の高い検索ができる

これらの点について、収集時と活用時の視点の変化を模式的に表した図1によって考察する。

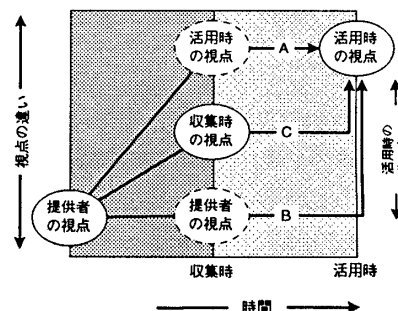


図1: 視点の変化とユーザの負荷

パスAで示すように活用時の視点を収集時に付与することができれば、活用の際に目的の文書を取り出すのは容易である。しかし、視点は収集時に定まっているとは限らず、また時間の経過とともに変化するのが普通なので、これは期待できない。

一方、パスBに示すように、提供者の視点(提供者が付与したカテゴリなど)をそのまま付与すれば、収集時の負荷は小さい。しかし、提供者の付与するカテゴリは情報源によって異なるため、複数の情報源からの関連する文書をまとめることができない。また、現状の技術レベルでは活用時に、視点に合致する文書を過不足なく検索することは難しい。

そこで、我々はパスCのように、収集時にはその時点での視点を付与し、活用時には視点間のギャップを埋めるという方法をとる。多くの場合、収集時と活用時で視点のギャップはそれほど大きくないので、収集時の視点を用いた検索結果から、活用時の視点に合致した文書に到達するのは容易であると考えられる。

*Fit, an Intelligent Retrieval System to Collect and Reuse Flowing Information (1) - Concept -

UEDA Takaya, OTANI Noriko, ITOH Fumiaki, SHIBATA Shogo and IKEDA Yuji (Media Technology Laboratory, Canon Inc.)

2.3 視点の表現方法

視点を表現する手段としてはフォルダを用いる。一つの視点に一つのフォルダが対応する。収集時には文書をフォルダ別に保存し、活用時にはフォルダ単位で文書を利用する。視点 V に対応するフォルダ f は次の三つ組で表す。

1. 視点 V を表現するラベル
2. 収集された文書の集合
3. 文書が視点 V に合致するか否かの判定条件

視点の変化はフォルダの変化として表される。フォルダの分割・統合や新規フォルダの生成だけでなく、フォルダへの文書の追加・削除も視点の変化を反映していると考えられる。

なお、フォルダに階層を設けると、階層を越えた変更が困難になり、視点の変化に対応しづらくなるので、フォルダには階層を設けない。フォルダ数が増えた場合に一覧性が低下するという問題に対しては、類似したフォルダを検索する機能と関連するフォルダをグループ化する機能を設けて対応する。

3 Fit の支援機能

3.1 収集フェーズの機能

効率的な情報収集を支援するとともに、視点付与の負荷を軽減するため、以下の機能を用意した。

視点別の文書提示機能

ユーザの視点に合致する文書をフロー情報の中から選択し、視点ごとに分けて提示する。一つの文書に複数の視点が付与される場合もある。提示された視点が正しい場合は、視点付与のために余分な操作を行わずによい。また、複数情報源からの関連した文書をまとめて見ることができるので、効率的な収集が行なえる。

収集フェーズの途中でユーザの視点が変化した場合にそれはそれを反映する。

保存候補のリストアップ機能

文書の保存先として適当なフォルダをリストアップする。Fit が提示したのと別の保存先を探す場合に用いる。多数のフォルダから望みのフォルダを探すのは困難なので、この機能は有効である。なお、一覧できる数の候補の中に正解があれば十分であり、現状の検索技術の精度でも十分に実現できる。

3.2 活用フェーズの機能

活用時の視点に合致する文書の検索を支援するとともに、情報を有効に活用できるように、以下の機能を用意した。

フォルダ単位の検索機能

フォルダの集合の中から活用時の視点に近いフォルダを検索する。検索機能には次の二つがある。

条件検索機能 ユーザが検索条件 (検索語とそのブール演算式) を与えて検索を行なう。検索条件への合致度の高いフォルダを検索する。[2]

類似検索機能 ユーザがフォルダや文書の一つを与えて、それに内容が類似しているフォルダを検索する。活用時の視点に近いフォルダ見つけた後、さらにそれを類似したフォルダを検索できる。[3]

フォルダ内の文書の共通部分が視点を反映していると考えられるので、フォルダ全体として条件を満足したり類似している方を優先することで、活用時の視点により近いフォルダを得ることができる。

また、文書の検索についても、フォルダを検索の対象とすることで、視点に合致する文書を容易に検索することができる。フォルダ内には、単独では条件を満足したり類似していると判定されなくても、視点に合致した文書が含まれる可能性が高いからである。

視点に応じた情報提示機能

活用時には、個々の文書を見るだけでなく、複数の文書を関連づけて見ることが多い。この際、どのような形で関連づけて見たいかは視点によって異なる。例えば、一連の事象について時系列順に並べて見たい場合や、共通する項目について比較対照したい場合がある。

そこで、時系列表示・表形式表示などいくつかの情報提示方法を用意し、視点ごとに設定できるようにした。着目する情報は文書中から情報抽出によって取り出す。これによって、視点に応じた情報の有効活用を支援する。

4 おわりに

フロー情報の収集・活用を支援する知的検索システム Fit のコンセプトについて述べた。今後は、「視点間の関係を扱う機能」の作成と「視点に応じた情報の提示機能」の拡充によって、フロー情報の収集・活用支援に必要な機能を充実させる。

参考文献

- [1] 上田他：フロー情報を対象にした情報検索システム (1) 概要, 情報処理学会第 50 回全国大会 4F-7, 1995.
- [2] 伊藤他：フロー情報収集・活用のための知的検索システム Fit(2) 処理方式, 本大会予稿 2T-9, 1996.
- [3] 大谷他：フロー情報収集・活用のための知的検索システム Fit(3) 類似度判定, 本大会予稿 2T-10, 1996.