

収集履歴と編纂履歴に基づく近傍検索方式

石井大輔[†] 浜田伸一郎[†] 石谷康人[†]

(株)東芝 研究開発センター 知識メディアラボラトリー[†]

1 はじめに

我々は、インターネットとデスクトップ上の文書に対し、タブレット PC 上の簡便なペン操作により検索、閲覧、スクラップをシームレスに実施し、効率よく情報を収集することが可能なシステムを開発した[1]。このシステムでは、検索クエリ、文書 URL、スクラップといった情報収集の履歴が蓄積され、ユーザはこれらの履歴へアクセスすることにより、情報収集過程における任意の時点へ立ち戻ること(近傍検索[2])ができる。さらに我々は、収集したスクラップに対し、ドラッグ&ドロップによる構成、表示スタイルの変換、アノテーションといった情報編纂を施し、新たな2次成果物の作成が可能なシステムを開発した[3]。

しかし現状では、上述した情報収集と情報編纂を異なるシステム上で行うため、2次成果物の編纂中にスクラップの収集過程へ立ち戻ったり、収集過程で参照した文書を2次成果物へ引用するといったことが困難であった。本研究ではこのような問題点を解決するために、情報収集および情報編纂の履歴をともに蓄積し、各作業結果へ横断的にアクセスすることを可能とする新しい近傍検索方式を提案する。これにより、情報収集と情報編纂の作業結果を効率よく再利用し、さらには循環的に新しい作業を実施することが可能となる。

以下、本論文では提案方式の基本コンセプト、実装のシステム構成、ユースケースについて述べる。

2 基本コンセプト

本研究で提案する情報活用は情報収集、情報編纂、収集履歴と編纂履歴に基づく近傍検索で構成されている(図1)。以下に各々の概要について説明する。

情報収集[1]: 文献[4]の技術を利用し、閲覧している文書上で任意の語句に線が引かれると、当該語句とその近傍の語句の意味に基づいて関連文書を効率よく的確に検索することができる。このようにしてアクセスした文書の部分をペンで囲むことにより、当該箇所をスクラップすることができる(図1(a))。そして、これらの情報検索、閲覧、スクラップの操作を収集履歴として蓄積することができる。

情報編纂[3]: 収集したスクラップに対して、ドラッグ&ドロップによるコピーや移動、削除、並べ替え、スタイル変換、編集、アノテーションなどの編纂作業を行い、複数のスクラップで構成される2次成果物を作成することができる(図1(b))。さらに、各編纂作業に対して時間、作業種別、IDのメタデータを付与し、編纂履歴として蓄積することができる。

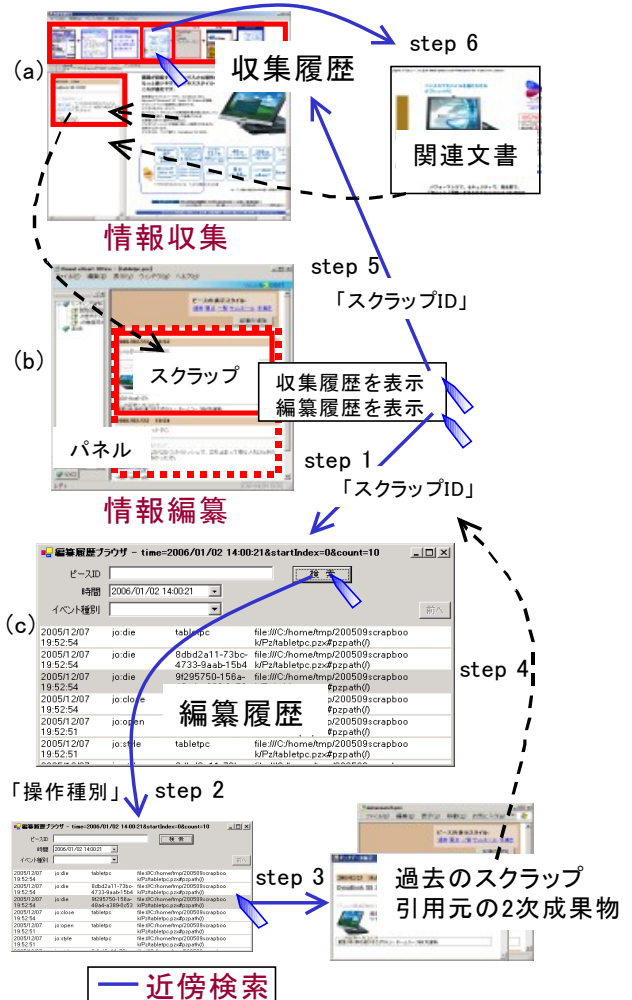


図1:収集履歴と編纂履歴に基づく近傍検索方式による、情報収集と情報編纂の連携

近傍検索: 時間、作業種別、IDのメタデータに基づいてスクラップの編纂履歴を検索することができる(図1(c))。このようにして検索した編纂履歴の1つを選択することにより、任意のスクラップにアクセスすることができる。このスクラップに限定して編纂履歴をさらに検索することにより、引用元の2次成果物や編集前のスクラップなどにアクセスすることができる。また、スクラップの収集履歴を辿ることによりスクラップ元の文書にアクセスすることができる。

このような情報活用環境を実現することにより、(1) 任意のスクラップの編纂履歴から任意の時点の編纂作業に戻る、(2) 再現された2次成果物の任意のスクラップから収集履歴に基づいてスクラップ元の文書を検索する、(3) スクラップ元の文書から再び情報検索を行い新しい情報にアクセスする、(4) 新しい情報の一部をスクラップする、(5) 新しいスクラップを(2)で再現した2次成果物

An Information Hopping Method Based on History of Information Collecting and Remixing
Daisuke ISHII[†], Shinichiro HAMADA[†], Yasuto ISHITANI[†]
[†]TOSHIBA Corporate Research and Development Center

にコピーして新しい2次成果物を作る、というシームレスな情報活用が可能となる。

3 システム構成

提案する近傍検索方式を図2のような構成により、Microsoft .NET 1.1環境上に実装した。

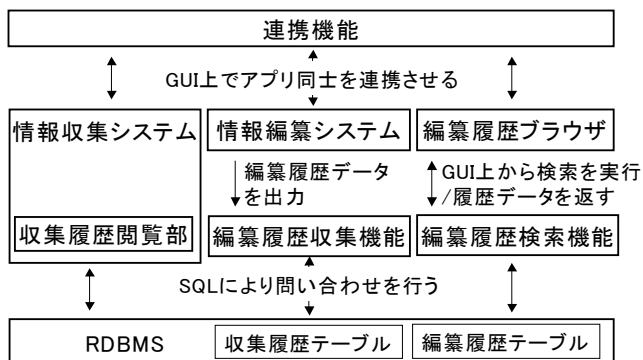


図2: システム構成

システムを構成する各機能について説明する。情報収集システムには文献[1]の実装を利用した。情報収集システムは内部に収集履歴閲覧部を持ち、収集履歴を収集するとともに、属性に応じた収集履歴へのアクセス機能を実現している。情報編纂システムには文献[3]のVisual eXcartを利用した。情報編纂システムは編纂履歴データを出力する機能を持つ。編纂履歴データを構成するおもな要素を以下に示す。

操作時刻: 操作が行われた時間を表す。

IPアドレス: クライアント識別のための情報。

操作種別: 編纂操作の種別を表す。

主体ID: 操作対象のスクラップ識別情報。

所属ID1、所属ID2: スクラップが所属するページ識別情報。引用操作の場合、引用元と引用先を出力する。

ダンプデータ: 操作時のページデータを保持する。

編纂履歴収集機能は出力された履歴データを受け取りRDBMSに格納する。編纂履歴検索機能は後述の編纂履歴ブラウザからの検索クエリを受けてSQL文を作成し、RDBMSへの問い合わせを行う。RDBMSにはSQLite[5]を利用した。編纂履歴ブラウザは編纂履歴データを図1(c)に示した時系列のリスト表示により可視化する。クエリ入力部に編纂履歴データの要素に基づくクエリを入力し、履歴検索を実行できる。情報収集システム、情報編纂システム、編纂履歴ブラウザの間にMicrosoft Windowsのシェルおよびウィンドウメッセージを介したGUI上での連携機能を実装した。

次に、上記システム構成における処理の流れを説明する。情報収集システム上で操作が行われると、収集履歴閲覧部が収集履歴を収集し、RDBMSに格納する。情報編纂システム上で操作が行われると、スクラップ部品毎の事前設定に基づき編纂履歴データが出力され、これを編纂履歴収集機能が収集し、RDBMSに格納する。

ユーザが情報収集システム上で編纂履歴検索を行うと、編纂履歴ブラウザのプロセスヘクエリが送信される。編纂履歴ブラウザは編纂履歴検索機能を介し、RDBMSから履歴データが得て表示を行う。情報収集システム上で収集履歴検索が行われた場合には、情報収集システムのプロセスヘクエリが送信される。

4 ユースケース

開発システムによる有用なユースケースとして、タブレットPCの製品調査を行う例を図1に沿って説明する。ここではユーザが図1(a)の情報収集システムにより製品調査を行い、図1(b)の情報編纂システム上に2次成果物である製品調査結果が読み込まれているとする。

- step 1.** 製品調査結果に含まれるスクラップをペンで選択し、メニュー「編纂履歴を表示」を選択すると、図1(c)の編纂履歴ブラウザが開く。ブラウザには選択スクラップの編纂履歴が表示される。
- step 2.** 履歴ブラウザのクエリ入力欄に操作種別「テキスト編集」を設定し検索を行うと、当該操作のみの履歴表示に絞り込まれる。
- step 3.** 絞り込まれた履歴表示中の1つの操作表示をペンでタップすると、ウィンドウが開き変更前のスクラップデータが再表示される。
- step 4.** ユーザは再表示されたスクラップデータを引用し、さらなる編纂を行うことができる。
- step 5.** 2次成果物中のスクラップ上でメニューをポップアップし、メニュー「収集履歴を表示」を選択すると、図1(a)の情報収集システムがスクラップの由来を表示した状態で開く。
- step 6.** スクラップ元文書の収集履歴を活用し、新たな検索を行うことができる。

5 まとめ

本論文で提案した近傍検索方式により、情報収集および情報編纂をシームレスに行うことが可能となり、知的生産業務を効率よく実施できる。また、複数ユーザの操作履歴に基づき近傍検索を行うことにより、たとえばオフィス内での生産性向上が期待できる。

参考文献

- [1] 石井大輔, 鈴木優, 石谷康人: ペン操作型情報収集とイベント型情報再利用に基づく情報活用システム. FIT2005, 2005.
- [2] 増井俊之, 塚田浩二, 高林哲: 近傍関係にもとづく情報検索システム. 増井俊之(編), インタラクティブシステムとソフトウェア XI, 2003.
- [3] 浜田伸一郎, 石谷康人: Webコンテンツの安全かつ簡便な再利用を可能とする編集技術 Visual eXcart. FIT2005, 2005.
- [4] 鈴木優, 布目光生, 石谷康人: ユーザの思考を妨げないペン操作によるインタラクティブな情報検索～意味解析と意図推定に基づく連鎖情報検索～. インタラクティブ 2005, 2005.
- [5] SQLite home page, <http://www.sqlite.org>.