

LOUIS: ラベルに基づく情報流通システム

横田 健彦、梅木 秀雄、田中 利一

(株) 東芝 研究開発センター
ヒューマンインターフェースラボラトリー*

概要: 興味ある Web ページの URL を保存し、再利用するための有効な手段として、Web ブラウザに付属のブックマーク機能が広く用いられている。LOUIS は通常のブックマークに比べてより柔軟な情報整理が可能なラベルつけ(ラベリング)を採用し、各ユーザが行ったラベリングの情報を共有することで、情報収集と関連情報検索を統合した Web ページ推薦システムである。LOUIS ではラベル名による検索を行わないため、他人のラベリング情報を直接公開することなく、自分と類似した興味をもつ他人から未知の有用そうな情報を獲得することができる。また、ラベリングという仕組みを利用することで、ラベル公開による協調的な情報収集や Web ページに対する意見収集が可能となる。

LOUIS: A Labeling-based Recommendation System for Adaptive Information Sharing on the Web

Takehiko Yokota, Hideo Umeki and Toshikazu Tanaka

Human Interface Laboratory
R&D Center, Toshiba Corporation

Abstract. Bookmarking facilities in Web browsers are widely used to store and retrieve interesting Web pages. LOUIS is a recommendation system that allows a user to label, or categorize, interesting Web pages. Individual users' labeling information is then used by LOUIS to determine other relevant Web pages of interest. LOUIS incorporates a relevance propagation scheme that, from a database of labeled URLs, can establish a recommendation list of ranked Web pages.

1. はじめに

現在、インターネット上では、WWW (World Wide Web) を介して膨大な情報が入手可能である。WWW で公開されている情報から自分の必要とする情報を見つけ出す方法としては、Infoseek¹や goo²等の検索エンジンがあり、見つけた Web ページの URL (Uniform Resource Locator) は、再利用のために Netscape Navigator³等の Web ブラウザに付属の「ブックマーク」

や「お気に入り」に保存することが一般的である。

ブックマークによる情報整理では、フォルダを作成することによって階層的に URL を管理することができる。しかし、フォルダを用いた階層的な情報整理にはいくつかの問題点がある[1]。例えば、ある URL を保存するのに適当と思われるフォルダが複数存在する場合に、どのフォルダに保存すべきか決定するのが困難である。また、保存している URL の個数が増大するにつれて、目的の URL がブックマーク中のどこに存在するかが分りにくくなり、必要な情報を再利用することが困難になってくる。

つぎに、情報収集に目を向けてみる。情報収集手段として幅広く用いられている WWW 上

*連絡先: 〒210-8501 川崎市幸区柳町 70 番地

¹ <http://japan.infoseek.com/>

² <http://www.goo.ne.jp/>

³ Copyright (C) 1994-1998 Netscape

Communications Corporation, All rights reserved.

の検索サイトでは、基本的に検索キーは単語であり、検索キーとして指定した単語が含まれている Web ページのリストが検索結果として得られる。このとき、自分が必要とする情報を表す的確な単語あるいは単語の組がはっきりしていれば、適切な情報を得やすいが、その反面、指定した単語を含んではいるが内容的に有用でない情報も多くリストアップされるといった欠点もある。テキスト解析だけでは得られない有用性や関連性については、協調フィルタリング技術を用いた情報検索が重要である。

協調フィルタリング手法を応用したシステムとしては、Firefly や GroupLens などがよく知られているが、映画や音楽といったドメインが限定される上、システムから提示されるアイテム(映画名やアーティスト名など)に対して、ユーザが個々に得点付け(rating)をしていく必要がある。

同様の協調フィルタリングを一般の Web ページ推薦システムに適用した例としては、SiteSeer[2]がある。SiteSeer はブックマークの共有に基づく Web ページの推薦システムである。このシステムはそれぞれのユーザのブックマークを収集し、ブックマーク中のフォルダ同士を、フォルダに含まれる URL の重なる度合いに基づいて比較し、他人のブックマークのフォルダのうち、ユーザが指定したフォルダに類似するフォルダを見つけ出し、そこに含まれる URL を類似情報としてユーザに提示する。

本稿では、ユーザが任意に作成したラベルを URL に貼り付けることによって、ブックマークよりも効率良く情報整理・再利用を支援し、各ユーザのラベリング情報を共有することで、共有したラベリング情報を用いて効率的かつ柔軟な Web ページの推薦を行なう(類似情報検索)システム LOUIS (Label Outlook system for Information Search, 「るい」と発音)[3]を紹介する。

2. LOUIS のシステム構成

LOUIS のシステム構成図を図 1 に示す。LOUIS はクライアント・サーバ型のシステム構成をとっている。サーバ部にはユーザに関する情報、ユーザが作成したラベルに関する情報、ユーザがラベルをつけた情報、URL に関する情報等が RDB (Relational Database) に保持管理されている。また他人のラベリングに対する類似情報検索もサーバ部のデータベース操作

によって行なわれる。一方、各ユーザはクライアント部にあるユーザインターフェイスを介してラベル付けや情報検索を行なう。

現在、サーバマシンとしては UNIX ワークステーションを使用している。また、クライアントマシンとしては、Windows95/98/NT の動作する PC を想定しており、ユーザインターフェイスプログラムは Windows95/98/NT 上で動作する。

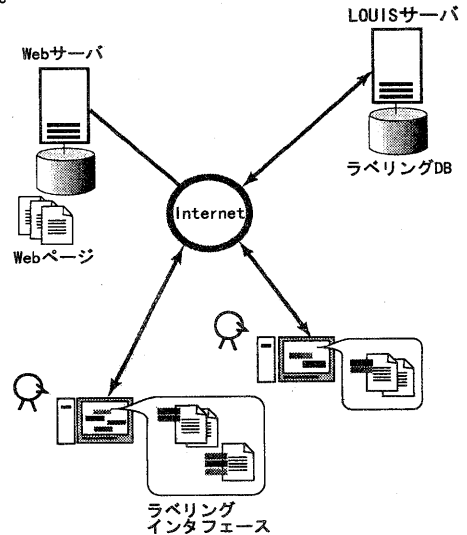


図1 LOUIS のシステム構成

3. LOUIS のユーザインターフェイス

LOUIS では、「パレット」と呼ばれる、ラベルを管理するためのユーザインターフェイスを用意している(図2)。ユーザは、任意の名前のラベルを作成し、パレットのウィンドウ上に自由に配置することができる。また、「場所」フィールドには、操作対象となる URL が表示されている。「場所」フィールドには、Netscape Navigator が現在表示している Web ページの URL が自動的に入力される他、キーボードからユーザが自由に URL を入力することもできる。

LOUIS では、ラベルを URL に貼り付けることによって URL の保存及び整理を行なう。ラベルを URL に貼り付けるためには、パレットのウィンドウ上のラベルを 1 つまたは複数個選択し、貼り付けボタンを押せばよい。どのラベルをどの URL に貼り付けたか、という情報(ラベリング情報)は、LOUIS サーバが持つ

データベースに蓄積される。

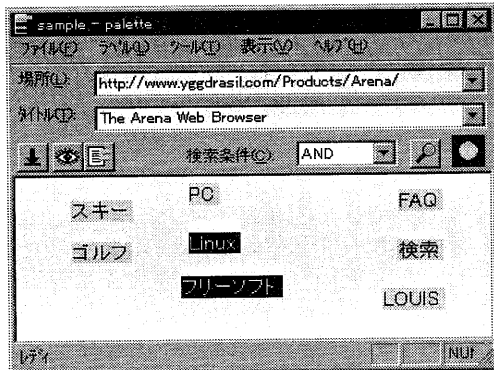


図2 パレット

保存した URL を再利用するために、ラベルをキーとした検索を行なうことができる。ユーザはパレットのウィンドウ上のラベルを1つまたは複数個選択し、検索ボタンを押す。この操作により、選択したラベルを検索キーとして URL を検索することができる。検索結果は「ファインダ」と呼ばれるユーザインターフェイスに表示される。図3は、「フリーソフト」と「Linux」という2つのラベルが貼り付けられている URL を検索した結果である。

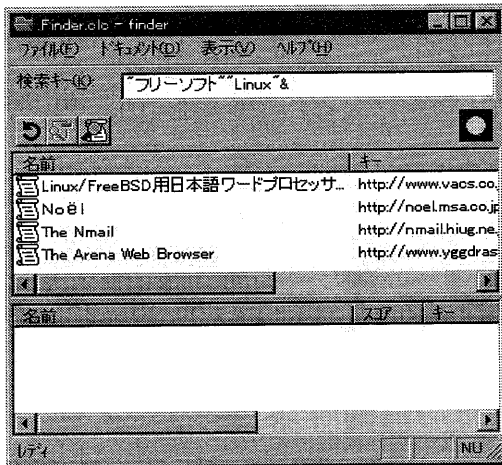


図3 ファインダ

LOUIS では、他人がラベル付けした URL のうち、現在ファインダに表示されている URL 群と関連が深いと考えられる URL を見つけ出し、ユーザに推薦する機能を提供している（類

似情報検索）。見つけ出された URL は、関連が深いと思われる順にファインダの下部（候補表示部）に表示される。図4は、「フリーソフト」と「Linux」というラベルが貼り付けられている URL を検索キーとして、類似情報検索を行なった結果である。関連の深さ（スコア）の計算方法については後述する。ユーザは、候補表示部に表示された URL について、必要であればラベルを貼り付けることによってその URL を保存することができる。候補表示部に表示された URL に対して簡単にラベル付けを行なうために、ファインダではドラッグ&ドロップ操作によるラベル付けが可能になっている。すなわち、候補表示部からファインダ上部に URL をドラッグ&ドロップすることで、検索キーとして指定された全てのラベルが自動的に URL に貼り付けられる。

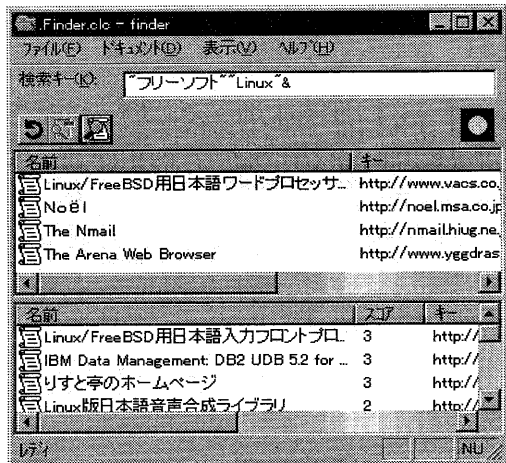


図4 類似情報検索の結果

4. 類似情報検索の基本アルゴリズム

類似情報検索の基本アルゴリズムを簡単に説明する。LOUIS では、各ユーザは図5のように URL にラベルを貼り付けている。ここで、ラベルと URL を結ぶ線がラベル付けの関係を示している。このようなラベル付け情報を「ラベリング情報」もしくは単に「ラベリング」と呼ぶ。

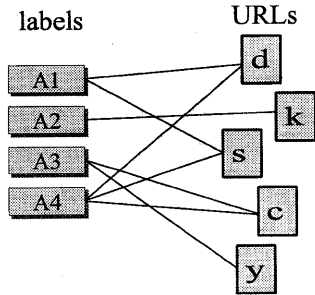


図5 ラベリング

類似情報検索では、他人のラベリング情報を用いて、自分の興味に近いと思われる URL を見つけ出す。検索のキーとなるのは、URL の集合である。図 6 は、ユーザ A について、ラベル A1 かつ A3 が付けられた URL 集合（ここでは a と e）を検索キーとして、他人のラベリング（ここでは B と C）からこれらの URL に関連した候補 URL を取得する仕組みを示している。

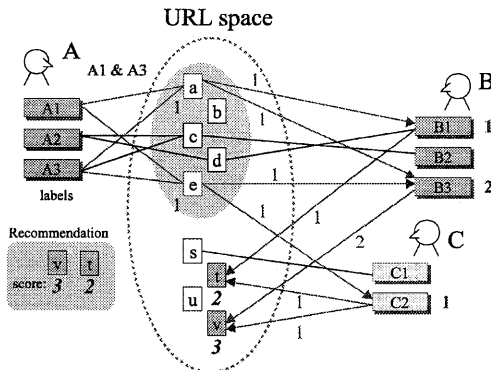


図6 類似情報検索

候補 URL と、検索キーとの関連の深さ（スコア）を得るまでの流れは以下の通りである。

- (1) ラベリング B 及び C 内において、URL a と e が付けられているラベルを検索し、ラベル付け関係ごとに各ラベルに対して重み 1 を与える。例えば、ラベル B1 は URL a と e のうち a だけに貼り付けられているので、重みは 1 となる。一方、ラベル B3 は a と e が貼り付けられているため、重みは 2 となる。この重みを「ラベルスコア」と呼ぶ。
- (2) 検索キー以外の URL のうち、ラベルスコアが 0 でないラベルが貼り付けられてい

る URL でユーザ A が持っていないもの全てについて、貼り付けられているラベルのラベルスコアを加算し、URL のスコアを計算する。例えば、URL v にはラベル B3 と C2 が貼り付けられており、ラベル B3 と C2 のラベルスコアはそれぞれ 2、1 なので、URL v のスコアは $2+1=3$ となる。

- (3) 適合度がある値以上になっている URL を候補 URL としてそのスコアでソートされた URL リストをユーザ A に提示する。

この検索アルゴリズムにおいて特筆すべきなのは、検索にはラベルと URL の対応に関する情報だけを用いており、ラベルの文字列の内容や URL の内容そのものに関する情報は一切用いていないという点である。このため、キーワード検索と比較し、次のような利点がある。

- キーワードを含むだけで内容的に無関係な URL を提示することがない。
- キーワードを含んでいないが内容的には有用な URL を検索からもらすことがない。
- キーワード毎にシソーラスを準備する必要がない。

一方、課題としては、類似情報検索における検索精度が挙げられる。類似情報検索は、ユーザの主観的なラベリング情報に基づいているため、検索を行ったユーザにとって、候補 URL として提示された Web ページ情報が必ずしも有用であるとは言えない。また、少数のユーザではラベリングした URL の重なりが乏しいため十分な類似情報検索結果が得られないことが考えられる。しかし、その場合でも、何らかの興味を共有する組織内における情報共有や、ある共通のサイト内の有用情報検索という目的であれば、有効な情報共有手段となりうると考えられる。

5. ラベルの流通

LOUIS の基本的なコンセプトは、ラベル名による検索ではなく、ラベリング情報に着目した情報検索機構である。しかし、ラベリングという形態を別の目的に利用することも可能である。それが、ラベルの公開による協調的な情報収集と Web ページに対する意見収集の機能である。

LOUIS では通常、誰がどのようなラベルを持っているかという情報は得ることができないが、ラベル名を意識的に公開することが可能である。この公開ラベルを Web ページに貼ることによって、別のユーザがその Web ページにアクセスすると、LOUIS のパレット上に公開ラベルが見え、公開ラベルを自分のラベルとしてインポートし、利用することができる。これにより、公開ラベルを介して、協調的な情報収集を行うことができ、自分だけでなく、他人が貼りつけた情報を直接検索することができる。

また、意見収集の場合も、ある Web ページに対する意見を集めるための特殊な公開ラベルを利用することで、自分が作成したページだけでなく、一般の Web ページに対する意見を収集することができる (図 7 を参照)。

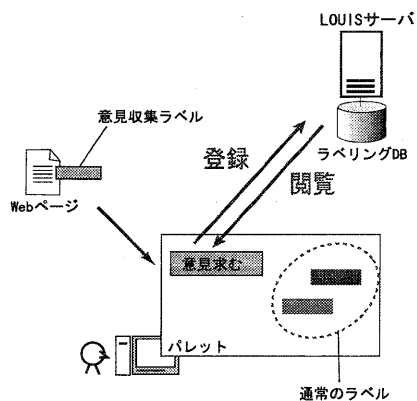


図 7 意見収集ラベル

6. まとめ

ラベリングに基づく URL 情報の整理及び収集システム LOUIS を紹介した。LOUIS では、ラベル付けによって URL を整理・保存することにより、ブックマークのような階層的な情報整理の抱える「情報の分類・再利用のしにくさ」といった問題点の 1 つの解決策を与えている。

また、LOUIS では、共有したラベリング情報を用いて、ユーザの興味に近い情報を見つけ出す機能 (類似情報検索) を提供している。類似情報検索では、ラベルの文字列や URL の内容に関する情報を用いず、ラベルと URL の対応付けに関する情報を用いることで、既に有用な情報を収集済みであるユーザから有用な情報を引き出すことが可能になっている。これに

より、キーワード検索では発見できなかった情報をも発見できるようになる。

参考文献

- [1] R.M.Keller *et al.*, Comput. Netw. ISDN Syst., Vol.29, 1103-1114, 1998.
- [2] J.Rucker and M.J.Polanco, Commu. of the ACM, Vol.40, No.3, 73-75, 1997.
- [3] H.Umeki and T.Yokota, Proc. the 8th WWW Conf., Toronto, 1999, in press.