利用者選好の半順序性に着目したWeb探索とナビゲーションの個別化

山口　雅史[†]　　大島　裕明[††]　　小山　　聡[††]　　田中　克己[††]

† 京都大学工学部情報学科 〒606-8501 京都府京都市左京区吉田本町
†† 京都大学大学院情報学研究科社会情報学専攻 〒606-8501 京都府京都市左京区吉田本町
E-mail: †{yamaguti,ohshima,oyama,tanaka}@dl.kuis.kyoto-u.ac.jp

あらまし　ユーザの嗜好に基づいたウェブページを推薦し提示することは，情報探索においてユーザの負担を大きく軽減させることが可能であり，それにより探索効率が向上すると考えられる．本論文では，ユーザの主観的な選好を定量的に観測することが困難であり，しばしば半順序的に与えられるものであるということに着目し，半順序グラフによりユーザの選好を表現した上で，閲覧中ページに含まれるリンク集合をランキングする手法について述べる．加えてランキングにおいて上位のリンクアンカーに対して，強調表示や自動ナビゲーションを行いユーザの情報探索を支援するウェブブラウザを試作した．ユーザプロファイルはユーザの Web ページ閲覧行為からリアルタイムに自動生成し，ユーザに明示的な選好を入力させることなく，暗黙的なパーソナライゼーションを可能にした．本システムにより，ユーザは通常通りのブラウジング操作を行うだけで，選好を自然に表現するユーザプリファレンスグラフを生成することができ，これにより選好に合致したウェブページを容易に発見することが可能となった．

キーワード　プロファイル 知識処理 パーソナライゼーション

Personalization of Web Search and Navigation by Partially-ordered User Preferences

Masashi YAMAGUCHI[†], Hiroaki OHSHIMA[††], Satoshi OYAMA[††], and Katsumi TANAKA[††]

† School of Informatics, Kyoto University Yoshidahonmati, Sakyou-ku, Kyoto,606-8501 Japan
†† Department of Social Informatics, Graduate School of Informatics, Kyoto University Yoshidahonmati, Sakyou-ku, Kyoto 606-8501 Japan
E-mail: †{yamaguti,ohshima,oyama,tanaka}@dl.kuis.kyoto-u.ac.jp

Abstract　To recommend the web page that is relevant to the user's interest can reduce the user's load. This paper describes a recommendation system that assists information seeking for the World Wide Web. It analyzes the page that the user is browsing, and extract hyperlinks from it then display list of hyperlinks ordered by the user's preference. The best ranked page is suggested in another web browser automatically. The user's preference is often given in partial order, not by a quantitative value. In this paper, user preference is presented by " User Preference Graph" based on partially-ordered graph. By using this, the proposal system suggests some web pages that might be useful to the user. " User Preference Graph" is generated automatically by observing the user's web page browsing activity. Neither the keyword about the user's interest nor the feedbacks to the recommendation are demanded. It is an implicit personalization method. By using this system, the user can generate " User Preference Graph" that can express the user's preference. The user can obtain information easily that is suitable for him without extra cost.

Key words　User Profile, Knowledge Processing, Personalization

1. はじめに

近年，WWW の発展は目まぐるしくその様相は変化に富んでいる．WWW 上には膨大な情報が存在し，様々な情報を得ることが可能であるが，ユーザは自分にとって有用な情報を得るまでの過程で常に情報を取捨選択しなければならない．WWW においては，現在ブラウザによる情報探索が主流であるが，これはユーザの能動的操作が必要不可欠である．

また，情報を得る上で現在 Google(注1)等の検索エンジンが頻繁に利用されている．検索エンジンにおいてはユーザの問い合わせに忠実な結果を返す反面，絞込みが不足しているなど，クエリの選定が不十分であれば，ユーザは大量の検索結果の中から個々のウェブページを順次閲覧し，求める情報を探さなければならない．

本研究ではユーザの選好に基づいたウェブページを推薦し提示することで，情報探索活動においてユーザの負担を軽減させる手法について述べる．ユーザが閲覧しているページに含まれているリンクアンカーとリンク先ページを解析した上で，ユーザの選好に基づき順位付けを行い，その中でも特に有用であるであると判断されるページを別ウィンドウに自動的に提示するウェブブラウザを試作し，システムの考察を行った．

リンクアンカー群の並べ替えを行うことで，以下のような利点がある．

- ユーザは素早く選好に合致したウェブページを発見することが可能．
- 上位にランクされたページを次々とユーザに提示することでユーザの自発的な操作を必要としない自動ナビゲーションを行うことが可能．
- 本来，検索結果の絞込みが重要となる検索エンジン利用時において，比較的少ない検索語でも目的の情報を発見する事が可能．

ユーザ個人に特化した情報を推薦する，いわゆるパーソナライゼーションに関する手法については，これまで多くの研究がなされてきた．従来の研究では，

（1） ユーザの明示的な嗜好入力

（2） 嗜好の絶対的な評価

を必要とするものが主である．

前者の特色としては，ユーザが登録した嗜好カテゴリに関するキーワードをプロファイルに利用するものや，ある対象に対してユーザが評価を行い，システムにフィードバックするといったものである．ユーザはシステムに対する何らかの入力と引き換えにシステムからの推薦を受ける．

後者の絶対的な評価として現在多用されているものに SD 法がある，個々の対象に 5 段階ないしは 7 段階評価などによって嗜好の強さを表現する手法であるが，このような絶対的評価は一個人の嗜好を表現するのには不適切である．

逆に「あれよりはこれ」,「この 3 つの中ではこれ」という順序による評価が現実では頻繁に行われる．つまり，ある対象の嗜好の度合いは単独で測定することが困難であり，他の対象との比較で表現する手法の方がより適切で，自然であると考えられる．

本研究では，上記のようにユーザの嗜好が定量的に与えられるのものではなく，しばしば半順序的に与えられるものであるということに着目し，ユーザプロファイルを半順序グラフを用いた「ユーザプリファレンスグラフ」で学習するシステムを提案する．

ユーザプリファレンスグラフはユーザのウェブページ閲覧行為の監視によりリアルタイムに自動生成し，ユーザは明示的に嗜好を入力する必要は無い．本システムにおいてユーザは普段のブラウジング操作を行うだけで選好に合致した Web ページを閲覧することができる．

2. 関連研究

2.1 Web Watcher

Web Watcher [1] は Carnegie Mellon University の Armstrong らによって開発された，ユーザの探索支援を行うソフトウェアエージェントである．システムは一種のプロキシサーバであり，探索の初めに興味や探索の対象についての入力を行う．ユーザが WebWatcher に対してウェブページをリクエストすると，WebWatcher は WWW に対してそのウェブページをリクエストし，以下の変更を行う．

（1） 操作メニューの挿入

（2） WebWatcher に引数として与えるための URL を置換

（3） ユーザに対する推薦

ユーザは探索を終了する時に挿入された操作メニューにより WebWatcher に満足した結果を得られたかどうかをフィードバックする．

類似の研究を以下に述べる．Pazzani ら [2] の Syskill&Webert はユーザによるウェブページの評価を基にして，推薦及び LYCOS(注2)への検索クエリを生成する．Balabanović ら [3] はユーザによる評価を基にしてヒューリスティクスを更新するエージェントについて述べた．いずれもユーザの嗜好入力及びフィードバックの必要性において本研究と異なる．

2.2 Letizia

Letizia [4] は Massachusetts Institute of Technology Media Laboratory の Lieberman らによって研究されているソフトウェアエージェントで，個人に特化したウェブページを推薦するものである．Netscape などの Web ブラウザと同時に動作し，ユーザのブックマーク及び閲覧履歴から tf/idf を用いたキーワードベースのユーザプロファイルを生成する．ユーザが現在閲覧しているウェブページのリンク先ページを別ウインドウのウェブブラウザ上で推薦する．

Letizia はユーザが操作中のウィンドウには介入せず，操作中でないウィンドウにおいて Letizia がそのユーザの嗜好に合致したページを推薦している．そのためユーザはエージェントに影響されずに自分の判断によるブラウジングを続行することも可能であるし，次にどのページを閲覧すればよいのか分からなくなった時など，必要になった時には Letizia の推薦ページを参照することができる．類似の研究には Armstrong ら [5] の Personal Web Watcher がある．いずれも，本研究とはユーザプロファイル及びその生成手法が順序に基づいたものではないという点で異なる．

(注1)：http://www.google.co.jp/

(注2)：http://www.lycos.com/

3. 選好の半順序性とその抽出方法

本章ではまず，本論文の提案手法に関して概要を記述する．ユーザがウェブページを閲覧した際，通常複数のハイパーリンクが存在するが，それらの選択状況を監視することで，本手法のユーザプロファイルとなる「ユーザプリファレンスグラフ」を生成する.「ユーザプリファレンスグラフ」はリンクアンカーが選択される度に更新され，それを用いてシステムはユーザが新たなページを閲覧した際にリンクアンカーの推薦を行う．

システムの流れを図 1 に示す．

なお，図中のリンクエレメントに関しては第 3.2 節に述べる．

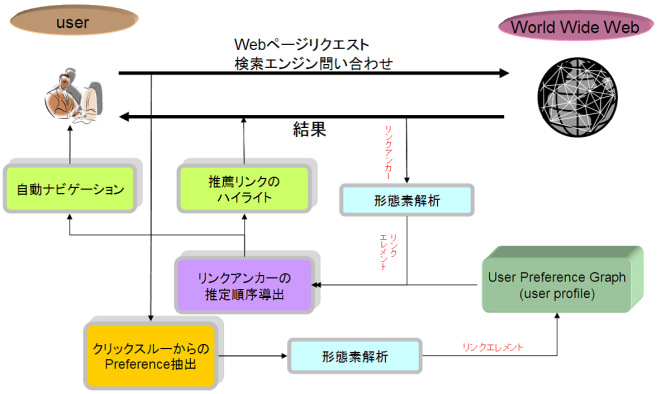

図 1　システムの流れ

3.1 選好 (Preference) について

3.1.1 選好の半順序性

人間の好みを表現する手法について考察する．第一に好みの対象となるものに対して，ある評価値を与えることによって，その対象の好みの強度を表すことが考えられる．たとえば A を 95，B を 50，C を 15 とした場合，A はその人の嗜好の尺度においては高く評価されており，逆に C は低く，B は中間程度といった具合である．この様な定量的観測法は一見有効であるように見える．しかし評価の対象が比較的多量になるとになると，ユーザ本来の嗜好に関しての大小関係と矛盾する評価値を与えてしまう場合がしばしば生じる．この様な現象はユーザの評価の基準が変動しやすく，再度たとえ同一の対象物を評価したとしても以前とは異なった評価値を与えてしまう可能性があることが原因である．

印象評価を数値化する代表的な方法に Semantic Differential(SD 法) がある．印象を，対極にある形容詞の組によって測定するものである．神嶌ら [6] は個人の嗜好を表現する手段として SD 法が不適であることについて述べた．

日常生活に於いて，複数の対象物から好みものを選択する機会は往々にしてあるが，人間の選好は，半順序性を持つと考えられる．日常生活でも，

- あれよりはこれが好み
- これら 3 つの中ではこれが好み
- これらに順序をつけるならば〜

というようにしばしば順序で与えられる事からも分かる．この様な評価手法は，対象に絶対尺度で評価する手法に比べ，容易かつ自然であり，上で述べた矛盾の発生を抑えることが可能である．

3.1.2 半順序 (partial order)

集合 X が半順序であるとは X に二項関係 $\preceq$ が定義されていて，次を満たすことを言う．

反射律 (reflexivity)　すべての x に対し $x \preceq x$

推移律 (transitivity)　$x \preceq y$ かつ $y \preceq z$ ならば $x \preceq z$

反対称律 (anti-symmetry)　$x \preceq y$ かつ $y \preceq x$ ならば $x = y$

3.1.3 事 前 実 験

20 代から 40 代の男女 20 名の被験者に 16 個の対象集合からランダムに 2 つの対象を提示し，選好に基づく優劣を測定した．提示した対象の対はのべ 480 組である．選好について推移律，すなわち「$x \preceq y$ かつ $y \preceq z$ ならば $x \preceq z$」が成立することを調査した．のべ 160 の対象集合 $\{x, y, z\}$ において，x と y，y と z，z と x の順序対 3 組の推移性に関してまとめたものが表 1 である．これにより推移律に基づいた推薦は有効である可能性が示された．尚，成立しなかった 2 組に関しては 5 段階評価で全て同程度の値となった組であった．当該被験者において優劣の付けがたい対象組であったと言える．

表 1　選好の推移性

	対象集合数	割合
推移律成立	158	98.7%
推移律不成立	2	1.3%
合計	160	100%

3.2 リンクエレメント

Web ページには通常複数のリンクアンカーが存在し，各リンクアンカーをクリックすることで逐次情報を獲得することができる．リンクアンカーは大抵，参照先の文書を端的に表した文字列である．ここで以下の要素をリンクアンカー関連文字列とする．(図 2)

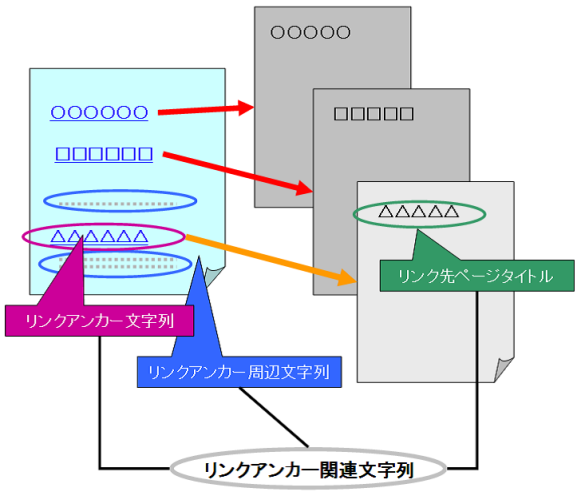

図 2　リンクエレメント

（1） リンクアンカー文字列
(但し，リンクアンカーが画像であった場合は alt タグの文字列で代用する)
（2） リンクアンカーの周辺文字列
（3） リンク先ページのタイトル

そしてリンクアンカー関連文字列に含まれている名詞集合を「リンクエレメント」と定義する．これはリンクアンカーを表現する名詞集合と呼ぶことができる．以後「リンクエレメント」とは，これらの名詞集合を指す．尚，ウェブページの名詞抽出処理には茶筌 (ChaSen) [7] を用いている．

3.3 リンクアンカーの選択状況に基づく選好抽出

WWW のページを巡回し，情報探索を行う場合，複数のリンクアンカーから一つを選択し，リンク先ページを表示すると行為を繰り返し行う．リンクアンカーの選択は，情報探索の目的によるばらつきは存在するものの個人の選好による一種の傾向がみられる．

本研究においてはユーザのリンクアンカーの選択状況を監視しユーザプロファイルを自動生成する．従って，ユーザは明示的に興味カテゴリ等の情報をシステムに入力する必要は無い．

リンクアンカーの選択状況から選好を抽出する手法を述べる．ここでは簡単に，あるウェブページに 5 個のリンクアンカーが存在する場合を想定する．(図 3)

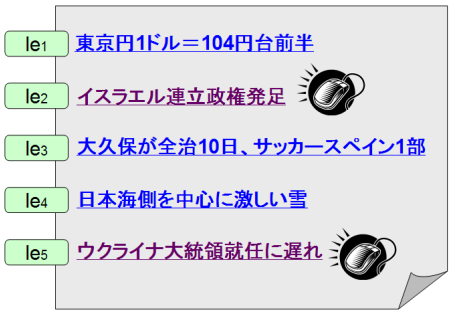

図 3 リンクアンカーの例

ここであるユーザが le_2 及び le_5 を選択したとすると，そのユーザの選好は

$$le_2 \prec_r le_1 \;\; le_2 \prec_r le_3 \;\; le_2 \prec_r le_4$$

$$le_5 \prec_r le_1 \;\; le_5 \prec_r le_3 \;\; le_5 \prec_r le_4$$

であると考えられる．但し r はユーザの選好を表現する半順序関係である．
一般化すると以下のようになる．

C:ユーザが選択したリンクエレメントの集合
$le_1, le_2, le_3, \cdots$:ウェブページに含まれているリンクエレメント
r:ユーザの選好を表現する半順序関係
とした時，
$1 \leqq j < i, i \in C, j \notin C$ を満たす全ての i, j について
$le_i \prec_r le_j$ を得る．

実際のウェブページはさらに多くのリンクアンカーが存在し，様々な目的で配置されている．

例としては，

- 他のウェブページを参照する外部リンク
- back， next といったような内部リンク
- バナー広告などの画像，動画

などがある．

上で述べた手法をページ全体に適用すると，異種の目的で記述されたリンクアンカーを選好抽出の対象に含めてしまう可能性が極めて高い．本手法ではこれを利用して選択されたリンクアンカーが table 要素内にあり，かつ 2 つ以上のリンクアンカーが存在する場合，同一 table 内でのみ上記の抽出を行い，ノイズを極力抑えた．これは同種の目的で配置されたリンクアンカー群がウェブページの配置の問題や視認性の良さから，table 要素に包括されていることが多い事を利用している．

3.4 対象と問題

4. 章および 5. 章では，リンクエレメント間の複数の順序関係が与えられた場合に，未知のリンクエレメント集合の順序を導出する手法について述べる．この問題は一般化すると以下のように表すことができる．

入力
- 既知の名詞集合の集合 S 上の半順序関係 r
- 未知の名詞集合の集合 T

出力
- T 上の半順序関係 r'

入力の関係 r と出力の関係 r' は明示的に区別しているが，以後は特に必要なければ $\prec, \succ, \approx$ で表す．

4. ユーザプロファイルの生成

4.1 リンクエレメントの半順序関係からの名詞間半順序関係の生成

リンクエレメントは 3.2 節で述べたように，名詞集合である．クリックスルーによって得られるリンクエレメントの半順序より，リンクエレメント中の名詞の半順序を以下のように得る．

$$le_i \prec_r le_j, i \neq j, \forall t_i \in le_i, \forall t_j \in le_j, t_i \neq t_j$$

において，

$$t_i \prec_{r*} t_j$$

得られた名詞の半順序は次節に述べるユーザプリファレンスグラフを用いて表現する．

4.2 ユーザプリファレンスグラフ

ユーザプリファレンスグラフとは有向グラフによるユーザプロファイルモデルである．各節点は原則として個々の名詞に対応し，有向枝により各名詞の半順序関係を表現する．ユーザプリファレンスグラフの一例を図 4 に示す．

節点 A から節点 B へ有向枝が存在することを $linkTo(A, B)$

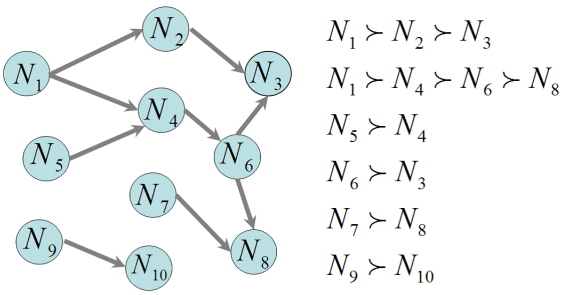

図 4　ユーザプリファレンスグラフ

と表すと，

$$linkTo(A, B) \Rightarrow B \prec A$$

である．

ユーザプリファレンスグラフはあるユーザにとっての選好をあらわすものであるが，同一のユーザとはいえ複数の半順序関係を統合したものであるため，全体としての半順序関係が必ずしも保持されるわけではない．例を挙げると，図 4 において，$N_4 \prec N_3$ が追加されると，N_3, N_4, N_6 は一つの閉路を形成し，推移律が崩壊する．結果，全体として半順序の要件を満たさなくなるため，ここでは閉路が出来た場合 N_3, N_4, N_6 を併合し一つの節点とすることで，推移律を保持する．(図 5) この場合，併合された節点は単一の名詞ではなく名詞集合となる．これに

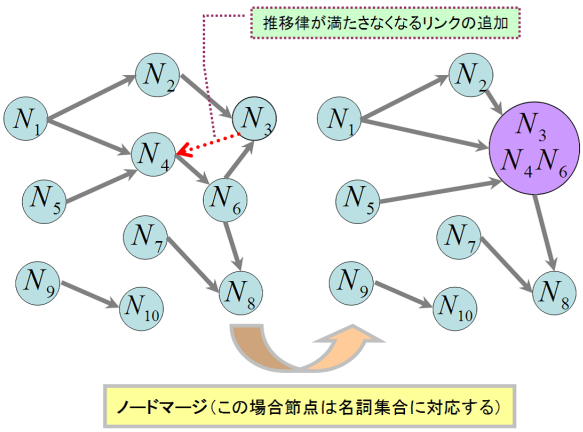

図 5　推移律の保持

よりユーザプリファレンスグラフは無閉路有向グラフ (directed acyclic graph) となる．

4.3　可到達行列

4.2 節で述べたユーザプリファレンスグラフは具体的には可到達行列で表す．可到達行列は常に特定の節点から有向枝によって辿り着くことが可能となる節点を表現している．

図 6 に図 4 に対応するものを示した．

可到達行列は有向枝が一箇所追加されるごとに更新される．$N_v \prec N_u$ が与えられたときの更新の手続きを Algorithm1 に示す．$M(u, v)$ は u 行 v 列の要素 (bool 値) であり，D は可到達行列の次元数である．

なお，N_u または N_v が未知の単語であった場合，可到達行

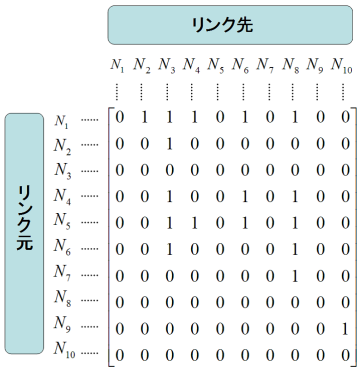

図 6　可到達行列

Algorithm 1 可到達行列の更新

```
1: if M(u,v) = true then
2:     exit
3: end if
4: M(u,v) ⟵ true
5: for i = 0 to D − 1 do
6:     M(u,i) ⟵ M(v,i) || M(u,i)
7: end for
8: for i = 0 to D − 1 do
9:     if M(i,u) = true then
10:        for j = 0 to D − 1 do
11:            M(i,j) ⟵ M(i,j) || M(u,j)
12:        end for
13:    end if
14: end for
```

列の次元を 1 ないしは 2 増加し，あらかじめ対応する名詞の行及び列を追加しておく．行 5-7 においては第 u 行に第 v 行との OR 演算の結果を代入，行 8-14 においては第 u 列が true となる全ての行に第 u 行との OR 演算の結果を代入している．$N_3 \succ N_4$ を図 4 に追加した場合の更新手続きを図 7 に示した．この手続きにより，

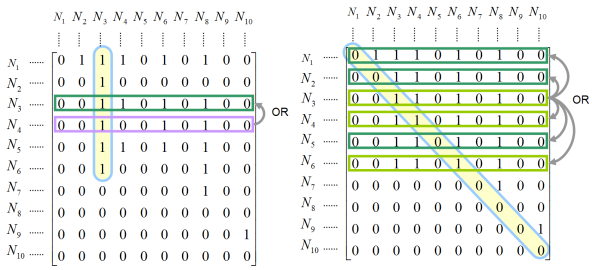

図 7　可到達行列の更新

$$\begin{cases} N_u \succ N_v & (M(u,v) = true \quad and \quad M(v,u) = false) \\ N_u \prec N_v & (M(u,v) = false \quad and \quad M(v,u) = true) \\ N_u \approx N_v & (M(u,v) = true \quad and \quad M(v,u) = true) \\ undefined & (M(u,v) = false \quad and \quad M(v,u) = false) \end{cases}$$

が言えることは明らかであり，また，

$$M(i, i) = true \Leftrightarrow (N_i \text{ は閉路を構成している節点})$$

$$M(i,j)\ and\ M(j,i) = true \Leftrightarrow (N_i \text{ と } N_j \text{ は同一閉路上に存在})$$

が言える．

($\because M(i,j)\ and\ M(j,i) = true \Rightarrow (1 < k < D-1$ において $M(i,k) = M(j,k))$

であるが，逆は偽である．)

図 7 右において N_3, N_4, N_6 において対角成分が $true$ であり，これら 3 つの節点は閉路を成している．

また

$$M(3,4)\ and\ M(4,3) = true$$

$$M(4,6)\ and\ M(6,4) = true$$

$$M(6,3)\ and\ M(3,6) = true$$

よりこれらは同一閉路上の節点であることがわかる．

可到達行列を用いることで，任意の節点間の順序を $O(1)$ で求めることができる．これはリアルタイムにウェブページを推薦する本システムにおいて有意義である．

5. ユーザプロファイルの利用

ウェブページ内に含まれるリンクエレメントを順序付けし提示するために，複数のリンクエレメント間の直列順序を導出する必要がある．

この章では，まず，2 つのリンクエレメントの比較に関して，次に全リンクエレメントの直列順序導出について記述する．

5.1 リンクエレメントの比較

ここで未知のリンクエレメントの集合が与えられた時，ユーザプリファレンスグラフからそれらの最適な順序を求める手法を考察する．ここで Lagrue [8] の極小元による近似的推定手法を用いる．与えられたそれぞれのリンクエレメントに対して，極小元を求める．

$$Min(\Sigma) = \{N_i \in \Sigma |\ \nexists N_j \in \Sigma, N_j \prec N_i\}$$

各リンクエレメントの比較は求めた極小元の比較により行う．二つのリンクエレメント X，Y において，それぞれの極小元，$Min(X)$ と $Min(Y)$ に関して以下が成立するとき「X は Y に先行する」とする．

$$X \prec Y \Longleftrightarrow \forall N_j \in Min(Y), \exists N_i \in Min(X)\ such\ that\ N_i \prec N_j$$

例を以下に示す．

$$X = \{N_2, N_3, N_6, N_{10}\},\quad Y = \{N_1, N_4, N_5, N_9\}$$

であるとき，図 4 のユーザプリファレンスグラフに基づいて極小元を算出すると以下のようになる．

$$Min(X) = \{N_2, N_6, N_{10}\},\quad Min(Y) = \{N_1, N_5, N_9\}$$

この場合

$$N_2 \prec N_1, N_6 \prec N_5, N_{10} \prec N_9$$

より，$X \prec Y$ が推定される．

5.2 全リンクエレメントの直列順序

5.1 節で述べた手法により得られる，リンクエレメント間の関係を用いて，全リンクエレメントの順序を導出する．得られる関係は半順序関係であり，それを表現する有向グラフは無閉路になるため，トポロジカルソートを適用することができる．

トポロジカルソートは，与えられた半順序関係から，それに矛盾しない直列順序を求める手法である．アルゴリズムを簡略に述べると以下のようになる．

（1） 全ての節点 i に対して，$t_in(i) \leftarrow -1, t_out(i) \leftarrow -1$

（2） 基点となる節点 n を選択

（3） $visit(n)$

　（a） $t_in(n)$ にタイムスタンプ設定

　（b） 全ての n の接続先 n' に対して

　　i. $t_out(n') < t_in(n')$ ならば閉路が存在するため停止．

　　ii. そうでないなら $visit(n')$ を実行．

　（c） $t_out(n)$ にタイムスタンプ設定

（4） $t_in(m) = -1$ となる節点 m が存在するなら，m を基点として選択し 3 へ．

（5） m が存在しないなら終了．

タイムスタンプは時間的前後関係を特定出来ればよいので，呼び出される度にインクリメントされるカウンタで設定すればよい．

5 が終了した段階で全ての節点 i に対して $t_out(i)$ が正の値をとり，その順序に節点を整列したものが全節点の直列順序の一つである．

6. プロトタイプ

6.1 開 発 環 境

システムのプロトタイプを実装した．以下に環境を示す．

- OS: WindowsXP Professional Edition
- CPU: Xeon 3.20GHz
- RAM: 2.00GB
- 開発環境: Microsoft Visual Studio.NET(C#)

6.2 システムのインターフェース

プロトタイプのスクリーンショットを図 8，9 に示す．

図 8 インターフェース 1

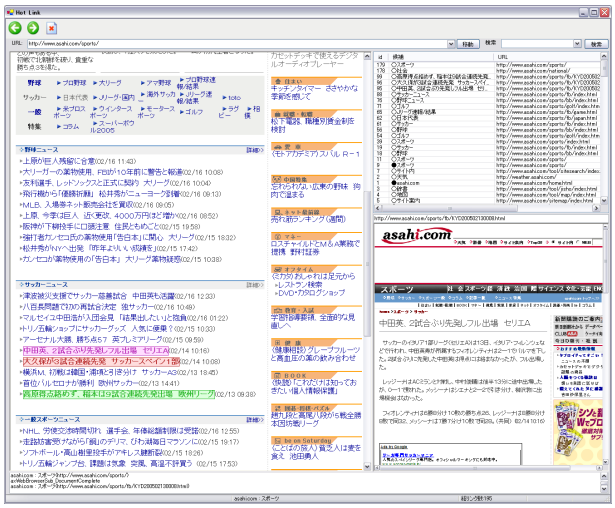

図 9　インターフェース 2

ウィンドウ内には 3 つのペインがある．それぞれの用途は以下の通り．

左側ペイン

ユーザが操作するウェブブラウザ．システムはこのペインでのユーザの行動を監視する．ランキングで上位 5 つのリンクアンカーは強調表示を行い，視認性を高めている．

右上リスト

左側閲覧中のウェブページに含まれているリンクアンカー群．ユーザプリファレンスに基づいてランキングされている．

右下ペイン

システムが制御するウェブブラウザ．右上リストの中で最も評価が高いページを提示する．

システムが制御している右側のペインはユーザが自由にサイズ変更ができ，非表示にすることも可能である．また，右上のリストにおいてはユーザが項目をクリックすることで，そのページを左側ペインに表示できる．

ウィンドウ右上には検索ボックスがあり，クエリを入力することで容易にパーソナライズされた検索結果を取得できる．

ユーザはまず左側ペインにおいて通常のブラウザと同様にウェブページを閲覧する．システムは閲覧中のリンクアンカーを抽出し，右側のリストとペインでプロファイルに基づいた推薦を行う．左側ペインで強調表示されているリンクアンカーはシステムの推薦リンクである．ユーザがリンクアンカーのクリックを行うとそこから選好が抽出されプロファイルに反映され，新しいページを表示するとそのページにおける推薦を行う．以上を繰り返し，ユーザの情報探索を支援する．

システムはユーザの閲覧を妨げる動作を行わない．システムはウェブページを提案するだけであり，それを採用するかはユーザ次第である．

6.3　システムの評価

本システムにより，ユーザプリファレンスグラフの節点数が 3000 程度になるまで適当な回数のウェブページを閲覧した後，様々なウェブページについて評価を行った．中でも特に有効であったと考えられるものを以下に示す．

WWW 検索エンジン

システムの右上に配置されている検索ボックス(注3)にクエリを入力した．検索エンジンから得られる検索結果はクエリを含むウェブページを評価の高い順に表示したものである．本システムにより行われた再ランキングではクエリに含まれていない語に関しての評価の差が反映されていた．具体的には「飲食店」というクエリの場合,「関西の飲食店サイト」や,「京都料理飲食店ガイド」などが他の地域の飲食店のウェブページよりも上位に表示された．この場合，関西，京都などはユーザプリファレンスグラフにおいて上位にある単語である．

オンライン店舗

WWW 上には多種多様のオンライン店舗が存在し，最近ではネットオークションも盛んである．オンライン店舗は商品の陳列に物理的拘束が無いという特性で，大量のカテゴリにわたり，膨大な数の商品が販売されている．本システムにおいては選好に合致したリンクアンカーを素早く発見することができるため，このように大量のリンクアンカーがあるウェブページにおいては特に有効であった．

地域情報

天気予報や店舗案内などの地域情報は，主に利用者の居住地域など特定の地域に関して参照される．このような局所的情報はユーザの傾向が表れやすく適していた．

特に検索エンジン利用時においては，少ない検索語であってもユーザの選好に合致したウェブページが上位にくるため，利便性が大きく向上した．

ユーザプロファイルを単語ベースの半順序グラフにすることで，ユーザプロファイルのデータ量を減らしながらも，少ない計算時間でリンクエレメント間の近似的な推定順序を与えることができた．

7.　考　　察

本章では提案手法に関する考察を行う．本手法の問題点として以下の 2 点が挙げられる．

- 節点数の収束
- 分解による情報損失

それぞれについての考察を以下に述べる．

7.1　節点数の収束

ループが形成された場合に推移律を保持する為，節点合成を行う．従って，節点数の変化は新たに生成される節点の増加分と合成による減少分の兼ね合いとなる．1 万語という制限の元でランダムに順序対を与えた場合の節点数の推移を図 10 に示した．

凡例の項目についてはそれぞれ

- 節点数 (左軸).. 合成しない場合の節点数
- 合成後の節点数 (左軸).. 合成した場合の節点数

(注3)：使用した検索エンジンは Google(http://www.google.co.jp/) である．

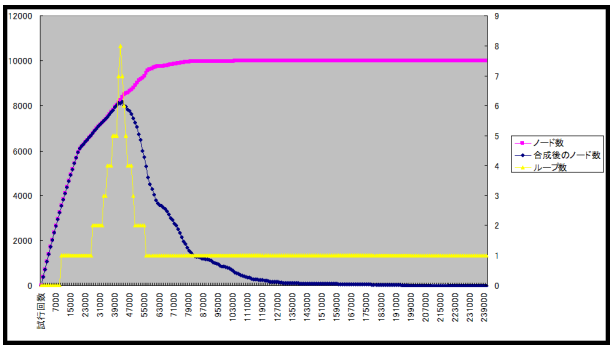

図 10　節点数の推移

- ループ数 (右軸).. ループ数

を表す．図 10 より合成を行うと節点数が 1 に収束することが分かる．この場合ユーザプリファレンスグラフは意味を成さない．しかし，以下の 3 点を踏まえると，ユーザプリファレンスグラフが短期間で意味を成さなくなる危険性はほとんど考慮しなくてよいと思われる．

- 節点数を最大 1 万個に制限しているため，収束速度は現実問題と比較して速くなっている．
- ユーザの選好に基づいたものではなく，ランダムな順序対を与えた．
- 1 万個の節点に対して，ループ数は高々10 個程度である．

さらに，個人の選好が長期間に渡り変化しないことは考えにくく，一定期間で既存のプロファイルを破棄し，新たなユーザプロファイルを作成する必要性があるため，この問題点はとりわけ大きな弊害にはならないと考えられる．

別手法として，有向枝に 2 つの単語間の選好コスト $Pref(a,b) \in [0,1]$ を定め,a よりも b を好む頻度を表すことも考えられ，この場合は推移律を成立させる為の節点合成は不必要である.（この場合極小元による推定の際は，閾値を設定し条件を緩和する必要がある.）しかしこの場合は 2 つの節点の大小関係を，それらの中間に位置する節点との関係から推移律を用いて導くことが不可能になるという欠点がある．

7.2　分解による情報損失

リンクエレメントは名詞の集合であるが，ユーザプリファレンスグラフで表現する際に個々の名詞の分解を行った．データ量を減らすと同時に一般のリンクエレメントの評価が容易にできるという利点があるが，語の共起に関する情報が失われてしまう．共起の情報を利用することができれば，推定順序の改善につながると思われる．

共起に関する情報を得る手段としては，相関ルールの確信度を用いる方法がある．これにより，ユーザプリファレンスグラフから得られたリンクエレメント間の順序を補正することで，共起による影響を近似的に考慮することができると思われる．

また別に，リンクエレメントを名詞に分解するのではなく集合の順序として扱い，共起による影響が無いような一般的とみなすことのできる単語を除き，データ量を削減するという手法が考えられる．

8.　おわりに

本提案手法により得られたユーザプリファレンスグラフは，順序関係をベースとしたユーザプロファイルであるため，人間の選好を自然に表現することができた．また，複数の順序関係を組み合わせることで，比較の幅を広げた．

ユーザプリファレンスグラフはウェブページの推薦だけでなく，名詞集合の順序を推定する問題であれば広範囲にわたって利用できると思われる．

今後の課題としては，抽出した選好の情報損失を極力抑えつつ，コンパクトなユーザプリファレンスグラフを生成する手法について，さらに考察を深めて行きたい．

謝　　辞

本研究の一部は，文部科学省科学技術振興費プロジェクト「異メディア・アーカイブの横断的検索・統合ソフトウェア開発」（代表：田中克己）および，平成 16 年度科研費特定領域研究 (2) 「Web の意味構造発見に基づく新しい Web 検索サービス方式に関する研究」（課題番号：16016247，代表：田中克己）および，21 世紀 COE プログラム「知識社会基盤構築のための情報学拠点形成」によるものです．ここに記して謝意を表すものとします．

文　　献

[1] Armstrong, R., Freitag, D., Joachims, T. and Mitchell, T.: Web Watcher: A Learning Apprentice for the World Wide Web, *Proceedings of IJCAI97* (1997).

[2] Pazzani and M Muramatsu, j. B. D.: Syskill&Webert:Identifying interesting web sites., *AAAI Spring Symposium on Machine Learning in Infomation Access, Stanford* (1996).

[3] Balabanović, Shoham, M. and Y: Learning Information Retrieval Agents:Experiments with Automated Web Browsing, *AAAI 1995 Spring Symposium on Information Gathering from Heterogeneous,Distributed Environments,Stanford* (1995).

[4] Lieberman, H.: Letizia:an agent that assists web browsing., *International Joint Conference on Artificial Intelligence, Montreal* (1995).

[5] Mladenic, D.: Personal WebWatcher: Implementation and Design., *Technical Report IJSDP-7472, Department of Intelligent Systems, J.Stefan Institute, Slovenia* (1996).

[6] 神嶌敏弘: なんとなく協調フィルタリング-順序応答に基づく推薦, *The 17th Annual Conference of the Japanese Society for Artificial Intelligence* (2003).

[7] 奈良先端科学技術大学情報科学研究科松本研究室: 茶筌ホームページ (http://chasen.naist.jp/hiki/ChaSen/).

[8] Lagrue, S., Devillers, R. and Besqueut, J.-Y.: Partially ordered preferences applied to the site location problem in urban planning, *Database and Expert Systems Applications*, pp. 151–160 (2004).

[9] Joachims, T.: Optimizing search engines using clickthrough data., *SIGKDD 02* (2002).