

Info-Plaza : 進化的クラスタリング機能を備えた WWW マルチブラウザ

平岩 真一 神田 陽治

(株) 富士通研究所 情報社会科学研究所

最近、急速に普及しているWWW上の情報にアクセスするためのツール、WWWマルチブラウザ Info-Plaza を開発した。Info-Plaza は、URL-Map と Plaza-Server から構成される。ユーザは、複数の URL 間の関係を URL-Map 上に視覚的に分類・整理できる。URL-Map の情報を Plaza-Server に送信することで、自分と興味の似通った人と URL 情報の交換がなされ、ユーザには新鮮で役立つ情報が盛り込まれた URL-Map が返される。この過程を繰り返すことで、URL-Map は徐々にユーザの興味を表す認知マップへと進化していく。

Info-Plaza : WWW Multi Browser with Dynamic Clustering Capability

Shinichi Hiraiwa Youji Kohda

Institute for Social Information Science
FUJITSU LABORATORIES LTD.

E-mail: hiraiwa@iias.flab.fujitsu.co.jp, kohda@iias.flab.fujitsu.co.jp

We have prototyped a multi-browser for WWW, Info-Plaza, which consists of URL-Maps on the user side and Plaza-Server on the server side. Users edit the relations between URLs visually, by putting their URLs on the URL-Map. Plaza-Server returns a newly created URL-Map when users send it their URL-Maps, by crossing over similar URL-Maps. Thus users who share similar interests could interchange their interesting URLs with the help of Plaza-Server and URL-Maps would gradually grow up as user's mental-maps of interests.

1 はじめに

World Wide Web(以下、WWW)[1]の普及によって、最近、インターネット上で提供される情報が、急速に増加している。WWWでは、情報のアクセス方法、位置を表すものとして、インターネット上の資源に対する統一的な名前付けの手法であるURL(Universal Resource Locator)を用いる。毎日、何件ものURLが紹介され、その数はますます増加している。WWWにより、情報の提供が容易に行えるようになった。しかし、必要な情報をいかに見つけるかということが大きな問題となった。

通常、WWWの情報にアクセスするときは、MosaicやNetscapeなどのWWWブラウザを用いる。こうしたブラウザでは、すでに知っているURLを直接指定するか、リンクをたどっていくという方法しか情報にアクセスする手段がない。情報を見つけるには、NetNews、雑誌といった他の媒体に流れている情報を利用したり、知り合いに聞くなどが主な手段である。

WWWを用いて情報を探す手段を提供するために、検索可能なメタインデックスを用意しているサーバがいくつか存在している[2][3][4]。この種のサーバを利用しキーワードによる検索を行って、情報を見つけることができるが、キーワード検索では、必要としない情報が入ってきたり、必要な情報が得られなかったりすることが起こり得る。

今回提案するWWWマルチブラウザInfo-Plazaは、自分と同じ興味を持つ人と情報を交換しあうことで、自分にとって新鮮で役立つ情報を発見するという仕組みを提供する。

Info-Plazaは、グループウェアとして見たとき、“不特定多数”が集まる“情報の広場”に集約される情報を、“特定多数”が見られる“広場の地図”としてユーザに見せる、そして、ユーザがそれを評価していくという基本的な仕組みを持つ。情報共有のあり方についての新しい提案である。

Info-Plazaは、複数のURLの関係を、二次元マップとして視覚的に表し、複数のWWWサー

バに同時にアクセスし情報を表示する。二次元マップには、Plaza-Serverとの通信によって、興味の似通った人とURL情報が交換されて、興味をそそると思われるURLが追加されていく。Info-Plazaは、生きた情報の広場であり、ユーザは、Info-Plazaを用いることで、徐々に進化していく二次元マップを通して、新しい情報を効率よく発見することが可能になる。

2 Info-Plazaの基本原理

同じ興味を持つ人同士が、有用な情報を交換し合う場があれば、お互いに役に立つ情報を手に入れることができる。情報を提供し合う人は、たとえ知らない人同士であっても別に構わない。Give and Takeで、お互いに知っている情報同士を交換し合うのは、情報を集めるのには非常に有効な手段である。

では、興味が同じということ、どうやって判断したらよいのだろうか。

通常、人は、収集した情報を話題ごとに分類して整理する。その整理の仕方は、人それぞれであるが、ある話題について考えた場合、その話題のカテゴリー名が関連したものなら、同じような情報をそのカテゴリーに分類する人がいると考えられる。また、たとえカテゴリー名が違ったとしても、分類された情報の内容が似通っていれば、共通の話題についての情報を集めていると推察することが可能である。つまり、カテゴリー名自体が関連している、または、そのカテゴリーの中に同じものが含まれてるというケースについて、そのカテゴリー同士の興味が近いと考えた。

例を示す。Aさんが天文学に興味を持っていて、天文学に関する本をたくさん読んでいる。同様にBさんも天文学に興味を持っていて、多くの本を読んでいる。また、Cさんは、物理学に興味を持っているとしよう。ここで、AさんとBさんは、天文学というカテゴリー名が同じということで、興味が近い。また、もし、Aさんの持っている本とCさんの持っている本が似ていれば、興味が近いととらえる。

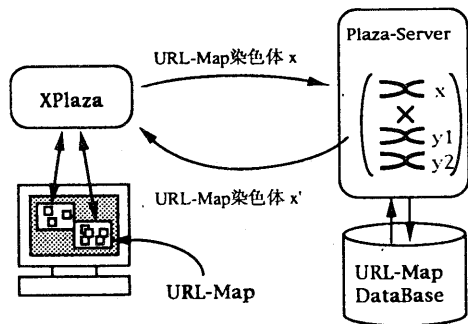


図 1: Info-Plaza の構成

この例での情報交換とは、本のタイトルの情報を交換し合うことである。

3 Info-Plaza の仕組み

Info-Plaza の仕組みを説明する。Info-Plaza において、収集される情報は URL である。Info-Plaza のシステム構成を図 1 に示す。

Info-Plaza は、クライアントである XPlaza とサーバである Plaza-Server から構成される。XPlaza は、視覚的なユーザインタフェースで URL を分類・整理することを支援する。分類・整理された URL は、Plaza-Server に集約される。ユーザは、URL の情報を Plaza-Server に送り (Give)、Plaza-Server は、データベース中の URL の情報から、興味が近いと思われるものを見つけだし、新鮮で役に立つ URL をユーザに送り返す (Take)。

Info-Plaza では、この仕組みに、遺伝アルゴリズム [5](進化的クラスタリング) 的な考え方を採り入れた。以下に、Info-Plaza の各機能について詳しく説明する。

3.1 興味を反映する URL-Map

Mosaic などの WWW ブラウザには、興味をもった URL をあとでもう一度参照するため、Hotlist や Bookmark といった機能が備わっている。XPlaza では、この機能を視覚化し、興味をもった URL をマウスの簡単な操作で二次元マッ

プに取り込み登録する。この二次元マップを URL-Map と呼ぶ。

XPlaza を用いて作成した URL-Map の例を図 2 に示す。URL-Map 上では、文字、イメージがそのまま縮小され、ミニチュア化した Mosaic のようなミニブラウザとして URL が表示される。もちろんスクロールもできる。もともと大きな文字や、画像は、縮小しても判読が容易であり、一目見ただけでどの URL かを思い出しやすい。このミニブラウザをマウスで自由に配置し、自分なりの URL-Map を作成する。URL のコピー、削除、Map-Map 間の移動なども可能であり、URL-Map は、Hotlist のビジュアルエディタとともとらえることができる。また、ミニブラウザをクリックするだけで、その URL の情報の内容が本来の大きさの WWW ブラウザで表示される。ユーザは、複数の URL-Map を作成することができ、各 URL-Map には興味に応じたラベルを付ける。XPlaza では、URL-Map に登録された複数の URL を、マルチウィンドウで同時にブラウザすることが可能である。

こうして作成された URL-Map は、ユーザの視点から URL の関係の明示し興味を整理したものであり、ユーザの興味を表すメンタルマップと捉えられる。

URL-Map から、それと同等の情報をもつ URL-Map 染色体 x を生成する。ユーザは、URL-Map の情報交換を行うために、URL-Map 染色体を Plaza-Server に対して送信する。

3.2 Plaza-Server

Plaza-Server は、送られてくる URL-Map 染色体を URL-Map データベースにプールする。プールする際、ユーザ名と URL-Map に付けられたラベルもいっしょに記録する。Plaza-Server には、ユーザが登録した数だけの URL-Map 染色体が保存される。

Plaza-Server は、ユーザから送られてきた染色体 x と、プールされた URL-Map を交叉することで、ユーザが興味を持ちそうな新しい URL-Map を生成しそれをユーザに提供する。データ

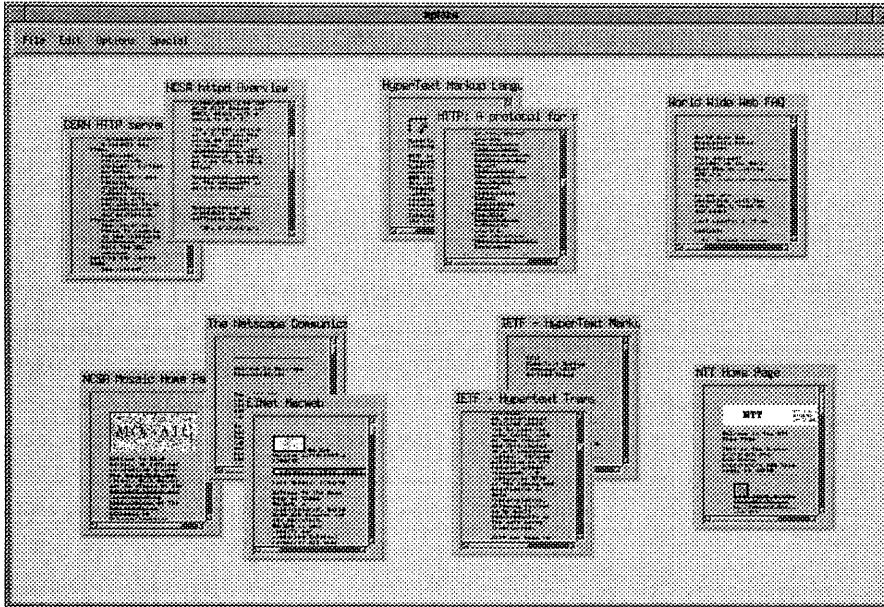


図 2: URL-Map: 利用者の興味に応じて配置された URL の地図

ベースにプールされた URL-Map 染色体全体の集合を M とすると、新しく生成される URL-Map 染色体 x' は、その生成関数を f とした場合、

$$x' = f(x, M)$$

で表される。

3.3 URL-Map の生成

生成関数 f について説明する。

遺伝アルゴリズムでは、染色体をランダムに交叉させる。しかし、単にランダムに URL-Map 染色体の交換をしてしまうと、興味を同じくする人との情報交換という目的が達成できない。このため、URL-Map の内容の近さと、URL-Map についてのラベルの関連度から、興味の近さを計測する。

いま、URL 染色体 x に対して、プールされた染色体の集合 M の中の一つ m_i を比較した場合、その類似度 s_i は、関数 g によって、

$$s_i = g(x, m_i)$$

で表記される。この類似度 s_i が一定値を越えたものを選択する。選択されたものを y_i と書く。また、これと同時に URL-Map に付けられたラベルの比較を行い、関連があるものを類似度とは関係なしに選出する。これも y_i と表記する。

こうして、データベースにプールされた染色体集合 M から、興味が近いと思われる染色体 y_i ($i = 1, 2, \dots, k$) が選択されたことになる。

次は、実際に URL 情報の交換を行う。この場合、選択確率を興味の近さを表す類似度 s_i に比例した確率で、染色体 y_i を選ぶ。選択したら、 x と y_i の間で URL の交換を行う。これを一定回数繰り返して、 x' を生成する。

URL-Map の生成は、類似度が低い URL-Map との情報交換を行ったり、交換する URL の数を多くしたりすることで、「このサーバにアクセスすると、Map の変化が大きい、いろいろ発見ができる」というように、サーバごとに特徴を持たせることができる。また、Plaza-Server には、ユーザからさまざまな URL-Map が集まる。どのようなユーザが Plaza-Server にアクセ

スするかによって、Plaza-Server に集まる URL-Map も変わる。当然、URL-Map 全体の集合 M が変われば、 x' も変化する。このため、複数の Plaza-Server が立ち上げればそれぞれ特徴のあるサーバになることは想像がつくだろう。ユーザは、好みの Plaza-Server を選ぶことができる。

3.4 ユーザによる URL-Map の評価

新しく生成された URL-Map 染色体 x' は、 x を送ってきたユーザに返される。XPlaza がこれを受け取ると、該当する URL-Map に、 x' の結果が反映されて新しい URL が表示される。

ユーザは、この新しい URL-Map をそのままにしてもよいが、おそらくこれに修正を加えるだろう。

ユーザは、新しく追加された URL の内容を見て、必要があると感じたら自分の好みの位置に URL を再配置し、興味を持たなければその URL を削除する。追加された URL はもちろんであるが、削除された URL についても、URL-Map の情報として保存され、次回以降の情報交換の時に利用される。

染色体の内容自体に変更を加えるという点では、この修正は、ユーザの感覚を評価関数 E とした遺伝アルゴリズムにおける染色体の評価にあたると言えるだろう。複数の候補染色体から高い評価値の染色体を選び出すかわりに、新しく生成された染色体をユーザによって高い評価値になるまで修正してもらうのである。ユーザは、URL-Map が自分の思い通りに変更できたら、再び、Plaza-Server に URL-Map を送信してもよい。こうした操作が繰り返されることで、ユーザの興味は URL-Map に蓄積されていく。

4 Info-Plaza の拡張

ここでは、Info-Plaza のその他の機能について説明する。

4.1 URL-Map マップの共有

前記の方法で生成されていく URL-Map は、いわば不特定多数の匿名グループによる情報共

有である。URL-Map は、特定多数による情報共有も可能とする。

URL-Map は、ちょうど UNIX のファイルと同じように、グループと、オーナー以外のすべての人に対して、それぞれ read, write の Permission を設定できる。read が許可されている他人の URL-Map を、自分の URL-Map の一つとして利用することができるし、また、write が許可されている URL-Map には、URL の登録、削除を行うことができる。

このような方法で、URL-Map を特定のグループ内で共有することができる。URL を共有を容易に行えらると、WWW をベースとする情報システムを構築すれば、研究活動などにおいて必要となる情報源へのアクセス手段としても Info-Plaza を利用できる。また、URL-Map を電子メールを利用して配布すれば、より多くのユーザに Map を配ることも可能となる。

4.2 URL へのアクセス履歴

URL-Map には、各 URL へのアクセスに関する情報も記録される。これを利用すれば、ユーザが、最近よくアクセスする URL だけを表示したり、1 カ月アクセスにっていない URL だけを表示したりすることもできる。

URL-Map を作成しても、常にその Map へアクセスに行くとは限らないし、人間の興味も徐々に変化していく。

この機能は、古い情報を捨てて新しい情報を得ようと、最近、興味を持っていることだけを選択した URL-Map を元に情報交換を行ったり、忘れていた情報を確認するといったことも可能にする。

4.3 URL-Map の自動生成

Plaza-Server が集める URL-Map は、ユーザからの情報提供に頼っている。ユーザから URL-Map が送信されてこなければ、Plaza-Server は、URL の存在を知る機会がない。この点を改良するため、NetNews などからの情報を集めて、自動的に URL-Map を生成する機能を Plaza-Server

に持たせることを検討している。

例えば、WWWに関係したニュースグループの記事で、SubjectをURL-Mapのラベルとして、記事の中で紹介されているURLをURL-Mapの要素として自動的にPlaza-Serverにプールしていく。紹介されているURLは1つであっても複数であっても構わない。この機能により、新しいURLが自動的にPlaza-Serverに登録されれば、Plaza-Serverにアクセスするユーザは、偶然新しい情報に出会うという機会に恵まれることになる。

4.4 XPlazaのデータキャッシュ

URL-Map上に複数のURLを登録した場合、登録したURLすべてのデータの取得を行うため、通信量が増大してしまう。そのため、現在のシステムでは、WWWサーバのキャッシュ機構をXPlazaにそのまま利用しているが、問題は多く残っている。

例えば、URLが指す情報の内容が更新された場合は、実際のWWWサーバにアクセスしなければならない。情報更新の頻繁なURLばかりを登録したURL-Mapを作ったときには、多くのWWWサーバにアクセスに行くことになり、表示までに非常に時間がかかる。

また、URL-Mapでは、URLがミニブラウザとして表示される。これは、ユーザが一目見るだけで、どのURLかを見つられるといった効果も大きい。情報が更新されたからといって、急にイメージが変化してしまうと、せっかく登録した情報を見つけるのに手間がかかってしまう。

このために、URL-Map上では、情報の更新があった場合、更新があったURLの枠の色だけを変えて、実際に情報を読むときに、更新された情報を新たに取得するとなどの機構をここに採り入れたいと考えている。

5 おわりに

WWWは、情報提供を容易にしたが、その反面で情報をいかに見つけるかが大きな問題となった。Info-Plazaでは、こうした問題に対して、

従来のキーワードによる検索という手法とは全く違った、さまざまなユーザの情報を交換し合うという方法を用いることを提案した。

現在、クライアントであるXPlazaのシステム作りを進め、Plazaサーバの詳細パラメータについての検討を進めている。今後は、システムの運用を通じて、評価・検討を行って行く予定である。また、WWWという枠組だけでなく、他の情報システムとの関係も可能にしていきたいと考えている。

この研究は、現在我々が進めているソーシャルウェア研究 [6][7] の一環として、検討およびシステムの開発を進めている。

参考文献

- [1] <http://www.w3.org/>
- [2] <http://info.cern.ch/hypertext/DataSources/bySubject/Overview.html>
- [3] C. Mic Bowman, et al, "The Harvest Information Discovery and Access System", Proc. of the 2nd International WWW Conference, Oct 1994.
<http://harvest.cs.colorado.edu/harvest/demobrokers.html>
- [4] Oliver A. McBryan, "GENVL and WWW: Tools for Taming the Web", Proc. of the 1st International WWW Conference, May 1994.
<http://www.cs.colorado.edu/home/mcbryan/WWW.html>
- [5] 北野 宏明 編, "遺伝的アルゴリズム", 産業図書, 1993.
- [6] 増井, 園部, "SocialwareLand: WWW上の視聴率・モニター調査機能の実現", 情処研報, 95-IM-20-5, 1995.
- [7] 神田, 園部, "Socialware: "架想"世界とデータ世界を現実社会で取り結ぶ情報メディアアーキテクチャの提案", 情処研報, 95-IM-20-6, 1995.