

擬似 Push 型 P2P 情報配信をするニュース記事閲覧システム

News article browsing system based on quasi-PUSH type P2P information delivery

安達 真¹⁾
Shin Adachi

1) 早稲田大学理工学部情報学科 2 年 (E-mail: shn@suou.waseda.jp)

ABSTRACT. The aim of this project is to develop a new Internet news browser that gathers latest news articles by pushing among users using P2P networking. The P2P networking functionality enables us to easily obtain appropriate articles and to enjoy automatic clustering of articles according to user's preferences, and consequently the possibility of ushing mutually-popular articles will increase. Additionally, the use of RSS (RDF Site Summary) approach as a common basis of content description helps this system easily adapt to coming standardizations on WWW content sharing technology.

1 背景と目的

インターネット及び WWW が一般家庭に普及してから久しく、現在 WWW 上には膨大な量の情報が存在するようになった。そしてこの膨大な情報を効率的に扱う手法が求められており、文献のような更新頻度の低い (Static な) コンテンツにたいしては Google のようなインデックス検索サービスが代表される。しかしこのような Web クロールを元にした検索エンジンは、ニュースや日記 (blog も含む) のような更新頻度の高い (Dynamic な) コンテンツには向いておらず、これらのコンテンツに適応もしくは特化した手法の開発が求められている。

そしてこれらのコンテンツの特徴として更新頻度が高いことのほかに、各々のコンテンツがトピック毎に小さくまとまっており、さらに各々コンテンツの間に多くの相関が存在する事が上げられる。例えばコンテンツ間の因果関係や、多数のサイトが同じトピックを扱った場合などが含まれる。現在このようなダイナミックなコンテンツを簡便にブラウズするために、WWWC^[5] やはてなアンテナ^[4] に代表される Web アンテナのような更新チェック、airWeb^[6] のような自動巡回エージェントが存在し、評価を得ているが、これらのサービスは前述のようなコンテンツの相関については考慮していない。

このコンテンツの相関を意識し、ユーザによりよいブラウジング環境を提供するために、本年度では第 1 段階として P2P ネットワークを利用した擬似 Push 型情報配信をするニュース記事閲覧システムを開発した。インターネットにおける Push 型情報配信は、ユーザ数に応じたサーバ設備を必要とするなど運営コストの問題、配信されるニュースが限定されているなどのコンテンツプロバイダーとしての問題が存在し、一般の利用者に普及する事が難しく、現在でもそれほど広く利用されているとはいえない。本開発では、情報配信をするための基盤として P2P ネットワークを利用し、低コストで運用する事に成功した。さらにネットワークを動的に再構成し、趣向によってクラスタリングする事によって、コンテンツの相関を利用したニュースブラウジングが可能となった。

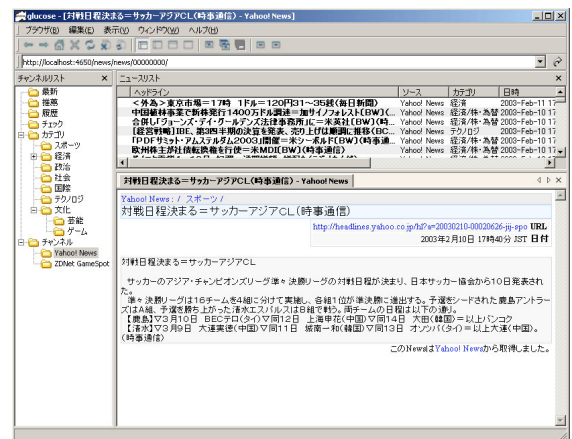


図 1: glucose

2 ソフトウェア構成

本システムは、クライアントとなる glucose と、P2P ゲートサーバとなる budou により構成される。

(1) glucose

glucose は

- エージェントによるニュースの自動収集
- ニュースのブラウジング
- P2P によるニュースの Push 型配信と協調フィルタリング

の機能を持つ。この三つの機能はシステムの構造としても三つのものに分担し実現されており、これらが総合されて効果を発揮するように工夫されている。まず、エージェントは各ニュースサイトを定期的に巡回して最新の情報を収集するモジュールで、その実現には Python スクリプト言語を用いる事で、開発コストを下げると共に、強力なカスタマイズ機能を持たせた。ブラウジングには、glucose が HTTP サーバを内包する形を取り、既存のブラウザでも違和感なくニュースブラウジングが出来るようにした。

(2) budou

図 2: 電光掲示板風の表示

budou サーバは P2P ネットワーク上のノードを管理するサーバである。ユーザが新しく P2P ネットワークに参加する場合、ハイブリッド P2P では中央サーバに接続すれば良かったが、ピア P2P ではその中央サーバがないためユーザはどこに接続すれば良いのか判らないという状況になる。これはピア P2P ネットワークの根本的な問題であるが、いままでの多くの P2P サーバがこの問題を極めて軽く扱っていた。Winny^[8] では、ノードのリストをボランティアで運営される WWW サイトなどよりユーザ自身が入手し、それぞれに対して接続を試行することでネットワークに参加する。gnuttela^[9] 互換サーバでは、FreePeer というサーバがノードリストを提供していたが、FreePeer はノード情報の質、鮮度を保つことに無関心であったため、接続開始まで何回も存在しないノードに対して接続を試行をする、といったことをしなければならなかった。ファイル交換といった、魅力のあるサービスならば、その接続開始までの時間や手間は苦にならないかもしれないが、ニュースを読むといった日常的なサービスを提供するのに、ネットワークに接続するまでの時間が遅れるのは問題である。このため、本プロジェクトでは、ノードが迅速にネットワークに接続出来るようにするための、鮮度の高いノード情報を提供する budou サーバを開発した。

3 ニュース記事取得

glucose では、ニュース記事の取得・管理に RSS^[1] を用いている。RSS(RDF Site Summary) は XML/RDF を用いてウェブサイトの見出しや要約を記述するフォーマットで、BBC や NY Times などのニュースサイトや多くの Wiki, blog で採用されている。又、bulknews^[2] や rss.jp^[3] のように RSS を提供していないサイトから RSS を作成し提供しているサービスも利用できる。

しかし、RSS を提供していないサイトは数多く存在し、無視できない存在である。そのため glucose では、サイト毎にセンサーと呼ぶ HTML → RSS コンバータを用意する事により、そのようなサイトも対応可能にした。この方法の欠点である、サイトに新規対応するためのコスト及び、デザイン変更によるメンテナンスのコストをできる限り低減するために、センサーの記述言語に Python を採用する事にした。2003 年 3 月時点までに 11 サイトのセンサーを作成し、いずれも 50~200 行程度の Python スクリプトで記述されている。

4 ニュースのブラウジング

glucose では、ニュースの一覧表示とニュースがリンクする WWW へのブラウジングの両方の要求を満たすために、MUA とタブブラウザを組み合わせたユーザインターフェースを採用した。ニュースはセンサーによりカテゴリ化され、さらに Push 頻度による重要度などを総合して一覧表示される。ニュースは、WWW との親和性を保つために HTML によって表示され、その表示には昨今評価の高いタブブラウザ形式を用いた。さらに、ユーザが自動的に新着や Push されたニュースの情報を得るために、電光掲示板風に表示する機能も搭載している。

5 P2P 及びクラスタリング

(1) NAT/FireWall 内クライアントへの対応

NAT や Firewall などにより外部から接続する事の出来

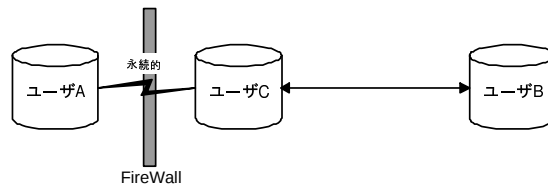


図 3: NAT/FireWall 内クライアントとの通信

ないコンピュータは非常に多数存在するため、P2P フレームワークとして対応する必要があった。

FireWall 内ユーザ A は、内部から FW 外のユーザへは接続できるが、逆に外部から接続をする事が出来ない。その場合、ユーザ A は外部のユーザ C に対して永続的なコネクションを確立させ、これをトンネルとして用いる。ユーザ A と通信したいユーザ B は、A の代わりに C と通信を行い、C は通信データを A に転送する。

(2) ゲートサーバ budou

前述のとおり、ユーザが迅速かつ手軽に P2P ネットワークに接続できるように、最初に接続する際のゲートとなる budou サーバを作成した。さらに budou は、(1) のような NAT 内クライアントにトンネル先となるユーザを紹介する機能も備え、NAT 内のユーザも特に意識することなく P2P を利用する事が可能である。さらに認証及び、暗号化機能を加える事により、セキュリティが求められる用途にも対応可能である。

ユーザがログインすると、budou はユーザの初期クラスタとなる 10 個程度のノード情報をユーザに配布する。ユーザは最初これらのノードと通信を行い、ニュースの閲覧状況によりクラスタを再構成する。

(3) Push 配信とクラスタリング

現在のバージョンでは、ユーザが読んだニュースは自動的に回りのユーザに Push される。さらにこのニュースを他のユーザに伝送することで、多くのユーザにニュースを Push する事が可能になる。自分の好みのニュースを配信するユーザとの距離を近く、そうではないニュースを配信するユーザとの距離を遠くするようにノード間距離を調節する事で、自分の好みのニュースが交換されるクラスタに配置されるようにする。

6 開発成果

本開発では、Python 及び RSS を基盤としたニュース記事収集システム、NAT 内ノードをサポートし、迅速な接続を行う P2P フレームワーク、P2P ネットワークを利用しニュースの Push 型配信を行うソフトウェアを開発し、現在ベータテスト中である。

開発成果は、下記の URL で公開中である。

<http://glucose.dip.jp/>

今後はさらに、自然言語処理によるニュースの自動カテゴリ化及び検索、blog へのサポートの強化などを行う予定である。

7 参加企業及び機関

なし

参考文献

- [1] RSS(RDF Site Summary/Rich Site Summary)
<http://web.resource.org/rss/1.0/>
- [2] <http://www.bulknews.net/>

- [3] <http://rss.s7.xrea.com/x/>
- [4] <http://a.hatena.ne.jp/>
- [5] WWWC <http://www.nakka.com/wwwc/>
- [6] airWeb <http://www.airclub.org/>
- [7] Python <http://www.python.org/>
- [8] Winny <http://www.geocities.co.jp/SiliconValley/2949/>

- [9] Gnutella <http://www.gnutella.com/>