

ユーザの関連性を用いた情報伝播モデルの評価実験

竹内 亨[†] 鎌原 淳三[‡] 下條 真司[◇] 宮原 秀夫[†]

[†] 大阪大学 大学院基礎工学研究科 情報数理系専攻

〒560-8531 大阪府豊中市待兼山町1-3

Tel: 06-6850-6588 Fax: 06-6850-6589

{stakeuti, miyahara}@ics.es.osaka-u.ac.jp

[‡] 神戸商船大学 情報処理センター [◇] 大阪大学 サイバーメディアセンター

kamahara@cc.kshosen.ac.jp

shimojo@cmc.osaka-u.ac.jp

概要

インターネット上の情報流通における情報フィルタリング手法がいくつか提案されているが、情報の価値を判断する際の基準として、情報の発信者とそれを伝えてきた中継者に対する信頼性が考えられる。そこで我々は、隣接ユーザとの関連性に基づいて情報を連続的に転送する情報伝播モデルの提案を行っている。このモデルでは、情報転送の可否を転送を行うユーザが判断するため、そのユーザに対する信頼性を情報に対する優先度として評価することによって、情報に対するフィルタリングが可能となる。これをヒューマンネットフィルタリング (Human Network based Filtering) と呼ぶ。本モデルに対する評価実験の結果から、ユーザ間の関連性に一定のフィルタリング効果があることが分かった。

キーワード: 情報伝播モデル, ヒューマンネットフィルタリング, フィルタリング, 情報流通, P2P 型通信モデル

Evaluation of the Information-Spread Model based on Users' Relevance

Susumu Takeuchi Junzou Kamahara Shinji Shimojo Hideo Miyahara

Abstract

There are some methods for filtering the distributed information in the Internet, but no method based on the authority of information or transmitter. We consider that the significance of information is determined by user's reliability toward the author or the transmitter. Therefore, we proposed the information-spread model in which user re-transmits information to his neighbors or discard based on the relations to the neighbor. Each user does such process one by one. In this model, since the neighbors judge whether the information has a worth to read, the reliability of information can be estimated by the confidence to them. Those reliabilities can be used as filtering. We named this "Human Network based Filtering." The result of the experiment in this article shows that the relevance between users takes filtering effects.

key words: Information-Spread Model, Human Network based Filtering, Filtering, Distribution of Information, P2P Communication Model

の情報を転送する。情報を転送されたユーザ C がその情報に対して興味を持たなかった場合、その情報は隣接ユーザへ転送されなくなる。そのため C の隣接ユーザ D には情報が伝達されないことになる。しかし F も D と同様に C の隣接ユーザではあるが、他の隣接ユーザである E が情報に対して興味を持つことによって、情報が伝達されてくる。

このように本モデルは、情報の確実な伝達は保証されないが、隣接ユーザが興味を持った情報だけが伝わってくるという、いわゆる「ロコミ」をモデル化したものであるといえる。そのモデル化のために、多数のユーザがそれぞれ隣接ユーザと関連性を持って接続しあうという P2P 型通信モデルを採用しており、これによって多数のユーザで情報を伝播し共有することができる。本研究では、このような情報伝播モデルをシステムとして実現することによって、ユーザが必要とする情報を十分に得ることができることを目的とする。

2.2 ヒューマンネットフィルタリング

情報伝播モデルにおいては、ユーザ同士が特定の関連性を持ちながら情報を提供しあうことが基本となっている。これは特定のユーザには隣接ユーザが興味を持った情報だけが伝わることを意味し、何らかの関連性を持ってつながっている人間が情報の選別を行うということになる。さらに多数のユーザが網目状に接続しあい、この転送するべきかどうかの判断を自立的に分散協調処理することによって、全体として効率よく情報をフィルタリングできることになる。これを人のネットワークによって情報をフィルタリングするという意味で、ヒューマンネットフィルタリング (Human Network based Filtering) と呼ぶことにする。

以下では、このヒューマンネットフィルタリングの特徴を説明する。

2.2.1 伝達経路による情報の信頼性

我々は日常生活において、情報の重要性や信頼度などの情報に対する価値を、その情報の内容だけでなく発信者や直接知らされた人物や機関などの信頼度で測る。情報伝播モデルにおいては、発信者からの伝達経路がユーザに分かるようになっており、その伝達経路によって重要なユーザを経由している情報は、通常より重要であると判断することができる (図 2)。

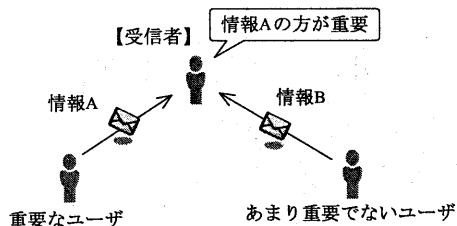


図 2: 伝達経路による情報の信頼性

既存の放送型の情報配信の場合、その発信者が明確であり、その発信者からユーザは直接情報を受け取るためにこの価値判断は行いやすいが、既にあるネームバリューの影響を受けやすい。それに対してヒューマンネットフィルタリングでは、情報を受け取ったユーザは発信者だけでなく中継者も分かるため、発信者に信頼が置けない場合であっても、その経路のユーザを調べることによって信頼度を補完することができる。一方で一対一型の通信モデルによって知らされた情報に対して、その伝えてきたユーザに信頼を置けない場合についても、同様に発信者や途中の中継者を見ることによって信頼度を判断することができる。

2.2.2 伝達距離による情報の信頼性

2.2.1 節のように伝達経路によって情報の信頼性が判断できることが本モデルの特徴であるが、情報を受信したユーザがあらかじめ重要だと判断できないユーザから情報を伝えられた場合、ユーザにとってその情報がどのような価値を持つのか容易に判断できないものとなる。

しかし本モデルでは、情報を受信したユーザに到達するまでに経たユーザ数が多いほど価値の高い情報であると考えることができる。これは、本モデルにおいて多数のユーザを経たということは多数のユーザがその情報に対して興味を持ったことを示しているためであり、この経たユーザ数である伝達距離が大きいほど重要な情報であると考えることができる (図 3)。

したがってこの指標を用いることで、情報を受信したユーザがあらかじめ重要だと判断できないユーザから情報が送られてきた場合であっても、その情報の重要性を判断することが可能となる。さらにあらかじめ重要なユーザからの情報であると分かっていたとしても、2.2.1 節の指標とこの指標

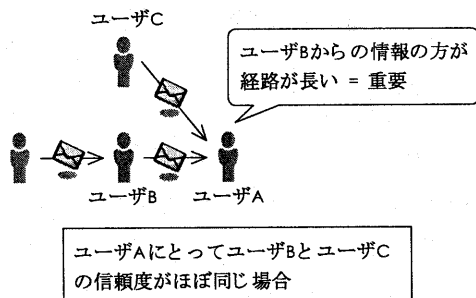


図 3: 伝達距離による情報の信頼性

を組み合わせることによって、より情報の重要度をユーザに明確に示すことができるようになる。

2.2.3 柔軟な情報の取捨選択

既存のコンテンツベースフィルタリング [4] においても、情報の取捨選択を自動的に行うことはできた。しかしこれは、ユーザプロファイルのような体系づけられたものに対して情報を伝えるべきかどうかをシステムが選択していたに過ぎない。これに対してヒューマンネットフィルタリングでは、人が情報を選別することから、より柔軟に情報を選別することができる。

例えば興味がある分野の情報であっても、人が見れば明らかに無駄な情報であることがある。そういった場合にコンテンツベースフィルタリングを適用すると、優先的に表示してしまうという問題が発生する。ヒューマンネットフィルタリングではこの選別を人間が行うことから、人が見れば明らかに不必要な情報は確実に遮断することができるようになる。

2.2.4 意外性の高い情報の迅速な伝達

特定のユーザに隣接したユーザは、何らかの関連性を持ってつながっている。そのため基本的にはその関連性に基づいた情報がやりとりされることになるが、実際にはその関連性自体とは直接関係のない情報をやりとりすることもユーザの判断で可能なため、意外性の高い情報も同じ機構で入手することができる (図 4)。

これは協調フィルタリング [5] においても可能であるが、嗜好が似たユーザの数が重視されるため、多数のユーザが興味を持たない限り推薦されることはなく、迅速に意外性の高い情報が伝わってくる

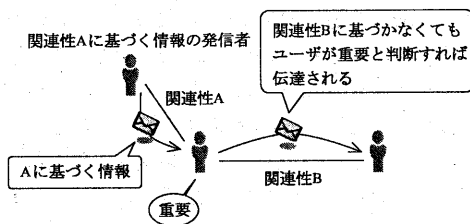


図 4: 意外性の高い情報の伝達

とはいえない。しかしヒューマンネットフィルタリングでは、隣接するユーザさえ重要だと思えばその情報が伝わってくるため、協調フィルタリングに比べて迅速な伝達が可能であるといえる。

3 情報伝播モデルの実装

3.1 実験システムの概要

今回の実験システムにおいては、評価情報収集のため、ユーザ間の一対一接続を図 5 のように一般的なメールサーバを通すメールクライアントアプリケーションとして実装することとした。

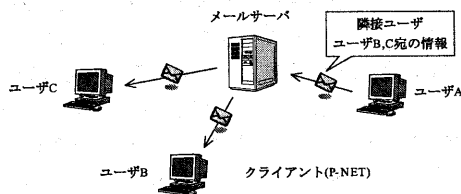


図 5: システムの概要

これによって、送受される情報は全てメールとしてクライアント間を転送されることになる。この際、情報は特定のフォーマットにしたがっていると拡張性や汎用性が増すため、ここでは XML [6, 7] によって記述することとした。さらに Java [8] を用いてクライアントを記述することで、よりシステム全体が拡張性や汎用性に富むように配慮した。

個々のシステムの流れは、3.2 節で決定した仕様に基づき、図 6 のようになっている。まず隣接ユーザから送信された情報がメールとしてシステムに受信される。受信された情報はシステムが全てユーザに示し、ユーザが面白いと思ったときはさらに次の隣接ユーザ全員へ自動的に転送される。逆につまらないとユーザが思ったときは、次のユーザに転送されずに破棄される。当然の事ながら、各ユーザ毎

に異なる隣接ユーザのリストを所持しているため、全員が全く同じ情報を共有することにはならない。

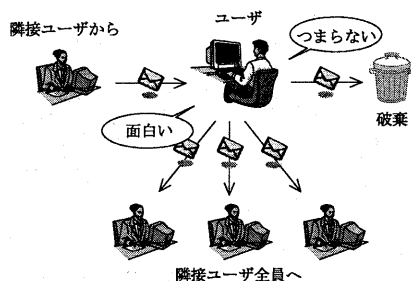


図 6: システムの流れ

3.2 実験システムの枠組み

2章で示した情報伝播モデルを実際に運用し、ヒューマンネットフィルタリングの評価を行うため、次のようなシステムの設計を行った。なお実験のため、いくつかの条件を仮定している。

3.2.1 情報の発信

情報伝播モデルではどのユーザも情報の発信者となることができるため、ユーザが新しい情報を発信する場合は、その情報を記入するだけで全隣接ユーザに自動的に送信されるようにした。

3.2.2 情報の受信

特定のユーザに対しては全隣接ユーザから情報が送られてくるため、同じ情報が異なる経路から届くことがある。こういった場合は先に届いた方が迅速性があるということで、二度目以降に届いた情報は全て無視されるようにした。

3.2.3 情報の転送とコメントの追加

隣接ユーザから届けられた情報に対してユーザが興味を持った場合は、3.2.1 節と同様に全隣接ユーザにその情報が転送される。また転送時にそのユーザ独自のコメントを追加することで、より有意義な情報を次の隣接ユーザに送り届けることができるようにした。

3.2.4 隣接ユーザの設定

ユーザ間ネットワークを構成する際、接続するユーザ同士はお互いが登録していることを認識する必要がある。そのため、隣接ユーザとなる双方のユーザが認証した場合にのみ接続するようにした。

4 情報伝播モデルの実験と評価

3章で実装したシステムを用いて実験を行い、ヒューマンネットフィルタリングの評価を行った。

4.1 実験および評価の方法

4.1.1 実験方法

前回行った実験 [2] と同様、実装したシステムによる実験は筆者の所属する研究室の学生など合計 22 人で行った。まず実験の元となるユーザ間ネットワークを構成する必要があるが、このために各ユーザは 1~6 人の範囲で隣接ユーザを登録し (平均 3.23 人)、その結果構成されたネットワークの接続グラフは次の図 7 のようになった。なお以下の評価においては、途中から参加または辞退したユーザは 0.5 人として換算している。またこの図においては、それらのユーザは描かれていない。

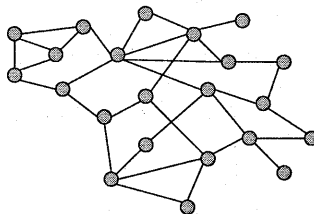


図 7: ユーザ間ネットワークの接続グラフ

このように構成されたユーザ間ネットワーク上で、評価情報を得るための情報を各ユーザが適宜発信することとした。しかし全てのユーザが均一に情報を発信することは不可能であり、ユーザ毎に情報発信数のばらつきが見られた。なお、流通した全情報量は 51 件となった。

4.1.2 評価方法

前回行った実験と同様に、4.1.1 節のような実験環境の上で実際に情報を流通させ、ユーザによる情報の評価結果を集計した。さらに実験終了後、各ユーザにユーザ間ネットワーク上に流通した全ての情報に対して、それぞれ興味を持てるかどうかについてアンケートを行った。

前回の実験においては、これらの評価情報を元に次のような評価を行った。

- 精度・再現率の評価
- 不要な情報のフィルタリング率

精度・再現率の評価においては、隣接ユーザ数が増加すると精度も再現率も増加する傾向が見られた。一般的に隣接ユーザ数が増加すると情報量が増加するため、隣接ユーザ数が増加すると精度は減少し再現率は増加するものと考えられたが、評価結果は異なった。そこでさらに、不要な情報量がどれだけフィルタリングされていたのかを評価した。その結果、隣接ユーザ数が増加すると確かに精度・再現率は増加していたが、不要な情報のフィルタリング率が悪化しているという傾向が見られた。

これらの前回行った評価においては、隣接ユーザ数が増加すると情報量が増加するという前提で行っていた。しかし実際には、情報源となるユーザに近いかどうか情報量に与える影響が強かったため、正しい評価となっているとはいえなかった。そこでここでは、次のような評価を行った。

- 精度・再現率の傾向
- 精度・再現率にみるフィルタリング効果
- 伝達距離と情報棄却率の相関関係

以下ではこれらについて詳しく述べることにする。

4.2 ヒューマンネットフィルタリングの評価

4.1.2節で示した評価方法に基づいて評価した結果を、以下に示す。

4.2.1 精度・再現率の傾向

前回の実験から、隣接ユーザ数が増加すると精度・再現率共に上昇するという傾向が見られた。同様の評価を今回も行ったところ、図8のように同様の傾向を見ることができた。なお、一つの点は一人のユーザを表している。

しかし隣接ユーザが増加すると、一般的にはユーザが得られる情報量が増えるため、精度は減少し再現率は上昇するものと考えられる。そこで今回は情報のフィルタリング率を用いて評価を行ったが、今回は隣接ユーザではなくユーザが受信した情報量と精度・再現率の関係を評価した(図9)。その結果、情報量が増加すると精度は減少し、再現率が増加するという傾向を得ることができた。

この結果から、確かに隣接ユーザが増加すると情報量が増加する傾向にはあるが、実際には情報源に近いユーザも大量の情報を得ているため、実

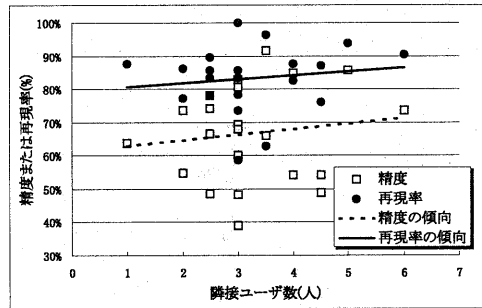


図 8: 隣接ユーザ数と精度・再現率の関係

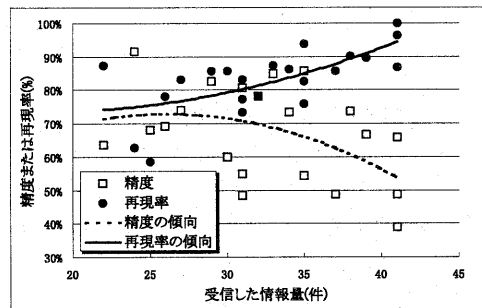


図 9: 情報量と精度・再現率の関係

際に得る情報量が精度・再現率に影響を与えることが分かる。

4.2.2 精度・再現率にみるフィルタリング効果

4.2.1節のように一般的な結果が得られたため、ユーザが受信する情報量が増加した場合にどの程度精度に影響を与えているか、また逆に減少した場合にどの程度再現率に影響を与えているのかという、フィルタリング効果について評価を行った。

この際の指標として、再現率は情報母集団の影響を受けにくいものの、精度は情報母集団の影響を大きく受けてしまうため、情報母集団に対して興味をもてる情報の割合を用い、精度に対するフィルタリング効果率 f_{prec} を次のように定義した。

$$f_{prec}(u) = \frac{\frac{\alpha(u)}{\alpha(u)+\beta(u)}}{\frac{A(u)}{A(u)+B(u)}}$$

ただし、情報母集団に対してユーザ u が興味があると思った情報量を $A(u)$ 、不要だと思った情報量 $B(u)$ とし、さらにユーザが得ることができた情報中で興味があると思った情報量を $\alpha(u)$ 、不要だと思った情報量を $\beta(u)$ とする。

この f_{prec} 値が 100% の場合、母集団から無作為に情報が抽出された状態を表し、100% 以上の値をとる場合は何らかのフィルタ効果が得られていると考えることができる。またこの指標は、母集団に対して興味を持てる情報が少ないユーザに対して改善が見られた場合はより大きな値をとるため、ユーザの満足度が反映され易いものとなっている (図 10)。

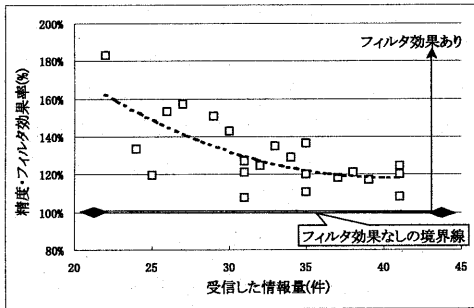


図 10: 精度にみるフィルタリング効果

この結果から、ユーザが受信する情報量が増加すると次第にフィルタリング効果が減少していくことが分かる。しかし全てのユーザの f_{prec} 値は 100% を越えており、情報量が増加したとしても、本モデルではフィルタリング効果を得られ続けていることが分かった。

続いて、再現率に対するフィルタリング効果について評価を行った。仮に母集団から無作為に情報を抽出してユーザに示した場合、再現率は $\frac{a+b}{A+B}$ という線形的な割合で増加していくと考えられる。このため、フィルタリング効果が得られている場合はこれより大きな値をとるものと考えることができる (図 11)。

この結果から、ユーザが受信する情報量がどのような状態であっても、高いフィルタリング効果が得られていることが分かる。さらに情報量が減少する割合に対して再現率が減少する割合が小さいことから、ユーザが興味を持てるような情報は他のユーザによって遮断されにくいという本モデルの有効性を示すことができた。

以上の評価結果から、ユーザにとって有用な情報を得るための手段として、本モデルは有効であることが示されたといえる。

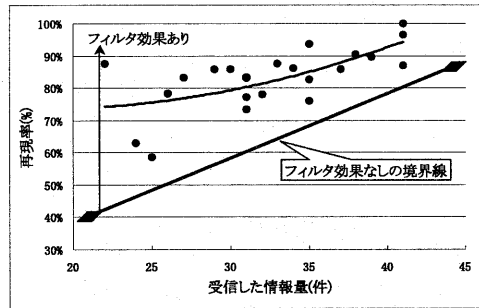


図 11: 再現率にみるフィルタリング効果

4.2.3 伝達距離と情報棄却率の相関関係

2.2.2 節のとおり、ユーザを多数経て伝わってきた情報の価値は、一般的に大きいものと考えることができる。そこでここでは、特定の情報に対してユーザが行った判断を元に、その情報の伝達距離 (ユーザが受信するまでに経由したユーザ数) と情報棄却率 (その情報をユーザが「つまらない」と判断する確率) の関係について、全ユーザの情報を元に評価を行った (図 12)。

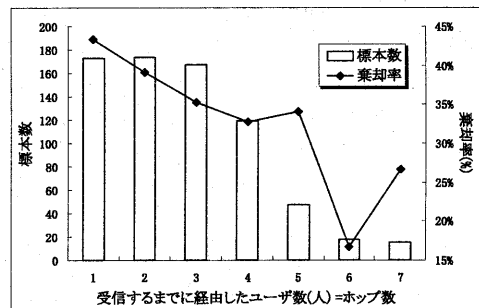


図 12: 情報棄却率の推移

その結果、伝達距離が増加するにつれ、情報棄却率が減少するという傾向が見られた。しかしこの評価においては、伝達距離が増加すると急激に標本数が減少するため、誤差が急激に増大してしまうという傾向が見られた。そこで、標本数が多量にある 4 ホップ目までを抽出した (図 13)。

この結果から、伝達距離と情報棄却率は指数分布にしたがっていることが分かった。この結果を用いることで、あらかじめ重要かどうか判断できないユーザから届いた情報であっても、その伝達距離数を元にした情報の評価を行うことができる。また、

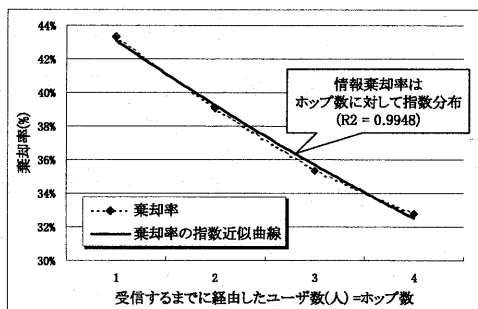


図 13: 情報棄却率の推移 (4 ホップまで)

あらかじめユーザが重要だと判断できるユーザから届いた情報であっても、さらにその重要度を増大させる指標として用いることができる。

5 まとめ

本稿では、多数のユーザが何らかの関連性のもとで網目状に接続しあった P2P 型ネットワークを構成し、そのネットワーク上で隣接するユーザとの関連性に基づいた情報をお互いに転送しあうことで、ネットワーク上に情報を波及的に伝播する「情報伝播モデル」という個人間コミュニケーション手段の提案を行った。この情報伝播モデルにおいては、それぞれのユーザが独立して情報の選別を行うことにより、人間のネットワークによって情報のフィルタリングを行っていると考えることができる。これをヒューマンネットフィルタリングと名付けた。そしてこのヒューマンネットフィルタリングを評価するため、実際にシステム化して実験を行った。その結果、既存の情報フィルタリング手法よりも柔軟に情報が伝達されてくることが確認できた。

またその評価を分析した結果、ユーザに伝わる情報量が増加した場合は精度が減少するが、その減少幅はフィルタリング効果が失われるほどではないことが分かった。同様に情報量が減少した場合は再現率も減少するが、その減少傾向は緩やかであり、フィルタリング効果が得られていることが分かった。また、伝達距離が増加すると指数関数的に情報棄却率が減少することも分かった。このことから全体として、各ユーザは独立に情報に対してフィルタリングを行っているものの、各ユーザに対してはフィルタリング効果があることから、ユーザ間の関連性がフィルタリングにおいて一定の役割を果

たしていると言うことができる。

このようなことから今後の課題として、ユーザが得られる情報量を適正に保つためにどの様に隣接ユーザを組み替えるのかということや、伝達経路や距離を用いてどのように情報に対する優先付けを行っていくのかなどについて、取り組んでいく必要がある。そして最終的に、各ユーザが必要十分な量の情報を受け取ることができるようなシステムの実現を目指していきたい。

参考文献

- [1] S. Ramakrishnan and V. Dayal, "The Point-cast Network," in *ACM SIGMOD International Conference on Management of Data*, p. 520, June 1998.
- [2] 竹内 亨, 鎌原 淳三, 下條 真司, 宮原 秀夫, "ユーザの関連性に基づいた情報伝播モデル," in 第 12 回データ工学ワークショップ, Mar. 2001.
- [3] 有限会社 グラム・デザイン, "Boogie." available at <http://www.gram.co.jp/>.
- [4] D. Billsus and M. Pazzani, "A Hybrid User Model for News Classification," in *Proceedings of the Seventh International Conference on User Modeling*, pp. 99-108, June 1999.
- [5] U. Shardanand and P. Maes, "Social information filtering: Algorithms for automating "word of mouth"," in *Proceedings of CHI'95*, pp. 210-217, 1995.
- [6] T. Bray, J. Paoli, and C. SperbergMcQueen, "Extensible Markup Language (XML) 1.0 (Second Edition)." World Wide Web Consortium (W3C), Oct. 2000. available at <http://www.w3.org/TR/2000/REC-xml-20001006>.
- [7] XML/SGML サロン, 標準 XML 完全解説. 技術評論社, 1998.
- [8] Sun Microsystems, Inc., "Java 2 SDK Standard Edition Documentation." Version 1.3. available at <http://java.sun.com/j2se/1.3/docs/>.