

利用者間に構築したネットワークを利用する 分散型 URL フィルタの提案

A proposal of distributed URL filter

大井 恵太† 林 憲亨† 向垣内 岳弥†
Keita Ooi Noriyuki Hayashi Takeya Mukaigaito

1. はじめに

Web を用いた情報提供が広範に利用されるにつれて、利用者が得られる情報が増え、情報過多ともいえる状態になっている。

もともとは性善説に基づいて運営されていたインターネットではあるが、有害な情報を提供するサイトや、ウイルスを含んだサイト、フィッシング詐欺を行う偽のサイトなど、必ずしも利用者が望まないサイトも多く現れるようになった。

ポータルサイトやニュースサイトでも、利用者の意思に関わらず、広告情報を閲覧させることが多い。企業や学校などでは、閲覧者が不要な情報の閲覧を強いられることによって、生産効率を下げている場合もあり、URL フィルタと呼ばれる情報を遮断する機構が導入されつつある。

これらの URL フィルタについては、フィルタ情報の管理の点で課題がある。本稿では、閲覧傾向の似た利用者同士とフィルタ情報を共有管理する URL フィルタシステムを提案する。

2. 従来システムとその課題

通常使われている URL フィルタとしては、以下のような方式がある。

1. 第三者あるいは利用者が作成した URL リストによるフィルタ(adblock[1]など)
2. URL 中に現れる文字列によるフィルタ
3. サイト運営側が第三者機関へ申告した分類情報を用いたフィルタ(PICS[2]など)

これらを単独で使用する方式に加えて、3 つを組み合わせたフィルタ情報を用いて、利用者の URL へのアクセスの可否を判断する方式がある。

上記の従来方式の有効性は全て、フィルタ情報に大きく依存している。そのため、精度を高めるためには大量のフィルタ情報を保持する必要があり、また、日々変化するインターネットでの情報提供に応じて、フィルタ情報も更新を行う必要がある、などフィルタ情報の管理が重要な課題である。

フィルタ情報の管理方式としては、以下の方式がある(図 1)。

- A) 利用者全員で同一のフィルタ情報を共有する方式
- B) 各利用者がフィルタ情報を管理し、情報は共有しない方式

† 日本電信電話株式会社 NTT 情報流通プラットフォーム研究所

NTT Information Sharing Platform Laboratories, NTT Corporation

方式 A) は、利用者がフィルタ情報の管理を行う必要がない利点がある反面、以下のような問題がある。

- a1. 利用者の嗜好を反映していないため、遮断されるべき情報が遮断されず、遮断してほしくない情報が遮断される場合がある。
 - 利用者にとっては必要の無い情報もフィルタ情報として保持されているため、処理に時間を要する要因ともなる場合がある。
 - 速度向上のために、URL 中の文字列のみで識別を行うと、精度低下の要因となる。
- a2. 管理者が莫大なフィルタ情報を扱う必要があることから情報の精査が困難である。
- a3. フィルタ情報が多くの利用者に影響を及ぼすため、高い精度の情報が求められるが、そのためにフィルタ情報の精査が必要であり、更にフィルタ情報の更新に時間がかかる。

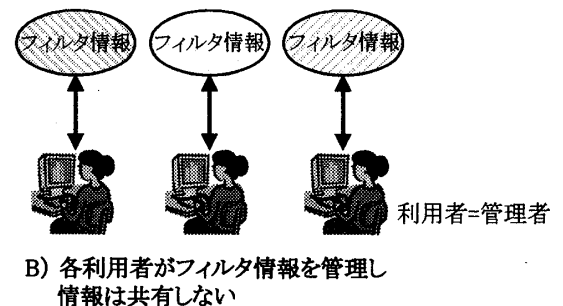
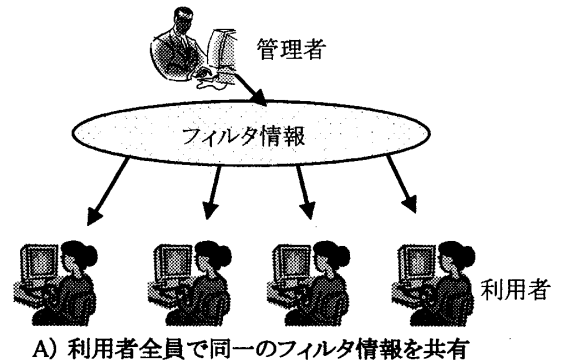


図 1: 既存のフィルタ情報管理方式

方式 B) は、方式 A) とは異なり、利用者は自身の嗜好を完全に反映した情報を作成可能であるという利点がある。しかし、全てのフィルタ情報の管理が各端末の利用者に委ねられており、利用者は、全てのフィルタ情報を自身で管理する必要がある。このため、方式 B) には、以下のような問題がある。

- a1. 情報の投入、更新を利用者自身で行う必要がある。

b2. 複数端末を利用する際にフィルタ情報の同期をとる必要がある。

これらの問題によって、方式 B) では、利用者の情報管理コストが増大する。

上記のような問題から URL フィルタの利用は、現状では、一部に限られている。

3. 提案システムの構成

上述の問題に対して、利用者に属するフィルタ情報を共有し、利用者向けに修正されたフィルタを作成・利用するシステムを提案する。図2に概念図を示す。

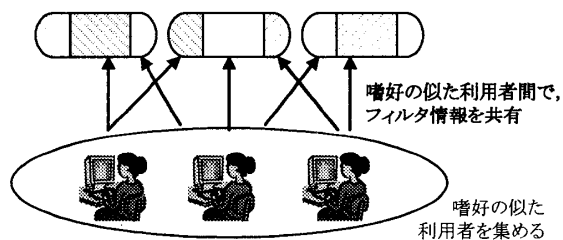


図2: 提案システムの概念図

本システムでは、利用者間でフィルタ情報を共有するために利用者をノードとするネットワークを作成する。利用者の閲覧履歴の一部を周囲の利用者に配信し、各利用者は周囲の利用者の閲覧履歴に基づき、ネットワークを再構成する。この再構成によって、閲覧傾向の似た利用者ノードが近くに配されたネットワークとなることが期待できる。

このネットワークに基づき、以下のルールに従って利用者の登録したフィルタ情報を共有する。

1. 利用者はフィルタ情報登録時に自身のフィルタ情報に追加すると同時に周囲の利用者にフィルタ情報を配信する
2. 同種のフィルタ情報を一定数以上受信した場合に、有効なフィルタ情報として利用する
3. フィルタ情報を利用した際に周囲の利用者に該当するフィルタ情報を配信する

このルールに基づき、嗜好の似た、近くの利用者が利用中のフィルタ情報と、利用者自身によって登録されたフィルタ情報から、自身利用するフィルタ情報を作成する。

これにより、利用者のフィルタ情報を分散管理するだけでなく、利用者ノードで適切な接続先を選択することから、利用者の閲覧履歴に基づくネットワークの再構成に関しても分散環境下で実現可能である。

提案方式は、方式 A) に比べ、利用者の嗜好に合致した過不足の無いフィルタ情報の作成が可能であるため a1. の問題が解決可能であり、全員のフィルタ情報を保持する必要はないため、a2. の問題が生じない。フィルタ情報の精度については、自身で投入したフィルタ情報と周囲の利用者が実際に用いた情報を活用するため、利用者にとって精度の高いフィルタ情報が随時作成され、a3. の問題を解決できる。

フィルタ情報の投入については、複数の利用者に情報投入のコストが分散されるため、各利用者の情報作成・管理の手間は、嗜好の似た利用者の数に反比例し、方式 B) における b1. の投入コストの問題も解決できる。同一利用者

による複数端末の利用時の b2. の問題も、同一利用者はネットワーク上で、近くに配置されるため、特別な作業を要せずにフィルタ情報の同期が実現できる。

更に、利用回数の多いフィルタ情報は、2. 3.によって、最初にフィルタ情報の登録を行った利用者から、広い範囲に伝播することが期待できる。

4. 評価

提案システムを用いて、のべ 21 名の利用者と約 3000 名の仮想ユーザを対象に実験を行った。この結果、以下の効果が確認できた。

- 各利用者は一部のフィルタ情報を入力するだけで、閲覧頻度の高いサイトの閲覧時に望まれるフィルタ情報が利用できる
- 利用したフィルタ情報を周囲の利用者に配信する機構によって、利用率の高い情報が広い範囲に伝播する
- 複数の端末を利用する者についても、情報のコピーなどの作業を行わずに、必要なフィルタ情報を獲得できる

5. まとめ

本稿では、閲覧履歴に基づいて閲覧傾向の似た利用者を近くに配置するネットワークを構成し、嗜好の近い利用者間でフィルタ情報を共有するフィルタ情報管理方式を提案した。

提案した URL フィルタシステムでは、利用者はフィルタ情報の管理コストを増大させることなく、閲覧を望まない情報を高い精度で遮断することが出来る。更に、第三者が利用者閲覧させたくない情報を遮断するといった目的に対しては、旧来のフィルタ情報を一元管理する方式との連携が必要であると考えられる。

一方、本システムのサービスが広範に利用された場合、情報を遮断したい利用者と情報を提供したい事業者との間で利害関係の対立が発生することも予想できる。この対立については、利用者に適した情報を選択的に提示するシステムの導入も進んでいるため、利用者情報提供者の両者が合意できるサービスの開発によって回避することができる。加えて、このような利害関係からは、故意に流される誤情報に対する防御機構も検討すべきであり、例えば情報や情報発信者の評価に基づいて利用の可否を判断する手法などが有効であろう。

今後は、本稿で提案した URL フィルタに限らず、利用者間の仮想ネットワークを通じて情報を共有・利用を行う、多種のサービスへの適用を検討していく。

参考文献:

- [1] Adblock, <http://adblock.mozdev.org/>
- [2] Platform for Internet Content Selection (PICS), <http://www.w3.org/PICS/>