

子どもを対象とした Web アクセス監視支援システムの構築と評価 Implementation and Evaluation of Web Access Watchdog System for Child Protection

上田 達巳†
Tatsumi Ueda

高井 昌彰‡
Yoshiaki Takai

1. はじめに

近年、インターネットは非常に身近なものになってきており、子どもたちがインターネットから情報を取得する機会が増えてきている[1]。インターネットでは有益な情報を得ることが可能である反面、誰でも自由に情報を公開することができ、また情報取得も容易であることから、情報の受信者が意図せずに有害な情報に接触してしまう危険性ははらんでいる。この危険を回避するために、インターネット上に存在する有害なコンテンツを排除しようという動きがあるが、現段階では法的に情報発信を規制するには限界がある[2]。

この問題に対応するために、フィルタリングソフトウェアが用いられている[3]。フィルタリングソフトウェアはネットワーク上のルータや使用者が操作を行うホストで通信に含まれる情報が有害であるか否かの検出を行う。しかしながら、フィルタリングソフトウェアには子どもが取得している情報を詳細に渡って監視、記録し、保護者に対して通知することにより、子どものプライバシーが犠牲になってしまうという問題点がある。また、コンピュータの取り扱いにあまり詳しくない保護者にとって、フィルタリングソフトウェアのログの取り扱いは煩雑なものとなっている。

そこで、本稿では子どもが取得している情報を、具体的な内容を含まない、カテゴリに基づく傾向分析の形式で保護者に通知する、Web アクセス監視支援システムの Windows プラットフォーム上での構築と評価実験について述べる。

2. システムの概要

2.1 全体構成

本システムの概略図を図1に示す。本システムは子どもが直接利用するホスト上に設置し、インターネットとホスト間で送受信されるパケットに含まれる Web ドキュメントを監視対象とする。Web ドキュメント内に有害な情報が含まれていないかどうか検知し、電子メールを用いて分析結果の要旨を保護者へと通知する。なお、システムの運用に当たっては、子どもと保護者の間での事前了解を取ることが望ましい。

2.2 端末内でのパケットの取得

インターネットを介して使用されるアプリケーションの多様性を考慮し、汎用的に通信の監視を行うことのできるパケットキャプチャによって通信内容の取得をするために、

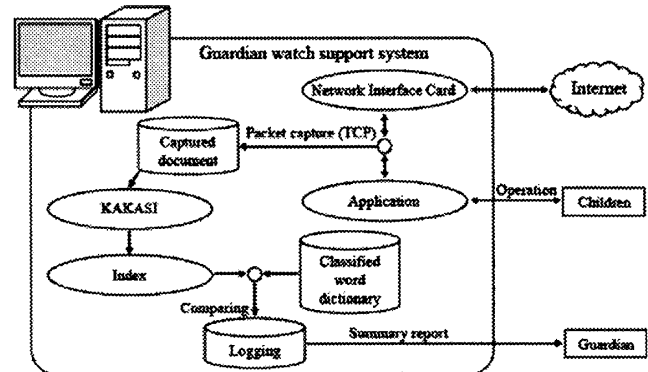


図1 監視支援システムの概要
Fig.1 Architecture of watchdog system.

パケットドライバ WinPcap[4]を用いてシステムを実装した[5]。キャプチャされたパケットはデータリンク層のフレームとして得られ、ヘッダを取り除いた IP データグラムを取り出す。同様にしてネットワーク層、トランスポート層のヘッダを参照して TCP/IP の通信のみを選択し、各ヘッダを取り除いた TCP データを得る。

また、これらのヘッダを用いる事で、TCP/IP およびアプリケーション層のプロトコルの挙動を模擬する。インターネットとアプリケーションの間の通信を再現する事により、通信に含まれる画像、音声などのバイナリデータを取り除き、テキストデータを得ることができる。

本稿では、これらのうち暗号化されていない通信において送受信される Web ドキュメントに含まれる日本語のテキストデータを取得する。

2.3 カテゴリ別単語の辞書化

情報が有害であるかどうかの判断は、受信者によって大きく異なる。例えば、子どもにとって過激な性表現は有害な情報となるが、そのような情報に興味を持つ成人にとっては有害な情報とはならない。

本稿における有害な情報とは、子どもにとって有害な情報を意味するものとし、具体的には性、暴力、犯罪、麻薬、人権侵害等を示す。これらの情報を表現するためによく用いられる単語を集めて辞書化する。表1に辞書に含まれる単語の一例を挙げる。また、子どもが興味を持っている分野を調べるために、様々な分野の単語を集めた辞書を作成する。

2.4 単語の検索

日本語の Web ドキュメントは、Shift-JIS, EUC-JP, UTF-8 などの文字コードを用いて記述されているため、一律に取り扱うために、nkf[6]を用いて文字コード変換を行った後、ドキュメントに含まれるテキストデータに対し KAKASI[7]により分かち書きを行い単語ごとに区切る。分

† 北海道大学大学院情報科学研究科, Graduate school of Information Science and Technology, Hokkaido University

‡ 北海道大学情報基盤センター, Information Initiative Center, Hokkaido University

かち書きによって得られた単語は辞書と比較を行い、一致した数がカウントされ、検出を行った時間と共にログへと書き出される。子どもがインターネットから取得している情報の傾向は、各カテゴリにおいて一致した単語の割合により求められる。

2.5 保護者への通知

パケットに含まれる単語のカテゴリ分析の結果は、図2に示されるように可視化され、あらかじめ登録された保護者の電子メールアドレスに通知される。これには、通信時間、転送量、パケット内の情報の傾向が示されている。また、インターネットとホスト間で送受信されたテキストデータが一切含まれていないため、子どものプライバシーを守る事にもつながる。単語の辞書を有害なものだけでなく、幅広いカテゴリについて用意する事により、保護者が子どもの興味の変遷を知る事もできる。通知の間隔は保護者が任意に設定することができる。

3. システムの評価実験

Windows PC 上に実装を行ったシステムに対し、インストールを行ったホストに対する負荷および、Web サイトの傾向分析の測定を行った。システムの実装には Visual C++ .NET を用いており、プログラムサイズは 148KBytes、辞書のサイズは 8 カテゴリ 4179 語 (59.1KBytes) である。測定時のホストとして、CPU: PentiumM 1.2GHz、Memory: 1.24GBytes、OS: Windows XP を搭載したマシンを使用した。

ローカルネットワーク上に設置した HTTP サーバから、28KBytes の Html ドキュメントを 1 秒間隔で転送した際のシステムによる平均 CPU 使用率および平均通信速度低下率を調べたところ、平均 CPU 使用率は 13.6%、平均通信速度は未使用時 247Kbps、使用時 239Kbps であり低下率は 3.24% となった。本システムの動作時にホストに与える負荷が軽微であることがわかる。

Web サイトの傾向分析として、スポーツ関連のニュースサイトである NIKKEI NET スポーツ総合、IT 関連の話題が交わされているスラッシュドットジャパン、アダルト関連の情報サイトである Oh-Yeah! すずきの情報リンクに掲載されているコンテンツに含まれる情報の傾向分析を行った結果を表2に示す。それぞれの Web サイトにおいて、メインとなっている情報のカテゴリ取得を行うことができていくことがわかる。

このことから、本システムは低負荷でありながらも、高い精度を持った実時間の Web アクセス監視機能を実現している。

4. まとめ

Web アクセス監視支援システムの Windows プラットフォーム上への構築と、システム使用時におけるホストに対する負荷および、Web サイトの傾向分析の測定を行った。今後、辞書の拡充および長期的なログの利用により、カテゴリ判定精度のさらなる向上を図ると共に、取得された情報を基にしたカテゴリ辞書の適応的な自己修正機能の実装を目指す。

表1 各カテゴリの辞書に含まれる単語の一例
Table 1 Example of words included in dictionary.

カテゴリ	単語例
アダルト	アダルトビデオ, イメクラ, ボルノ, 痴漢
ドラッグ	コカイン, シンナー, ヘロイン, 大麻
人権侵害	(放送禁止用語などから抜粋)
IT	イーサネット, エンコード, スパイウェア, 集積回路
料理	油通し, グラッセ, シチュー, 割した

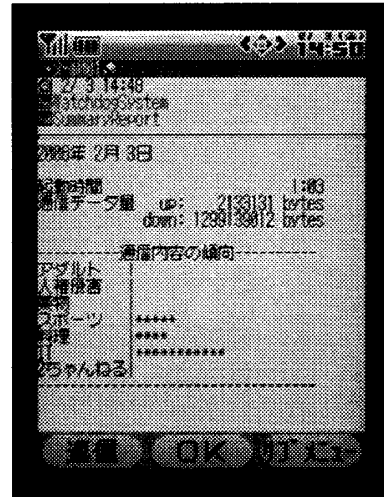


図2 保護者への通知メール
Fig.2 Notification mail to guardian.

表2 Web サイトの傾向分析の結果
Table 2 Trend analysis of Web site access.

カテゴリ \ サイト	NIKKEI NET スポーツ総合	スラッシュドット ジャパン	Oh-Yeah! すずきの情報リンク
アダルト	2 (0.9%)	0 (0%)	201 (88.5%)
人権侵害	0 (0%)	0 (0%)	0 (0%)
薬物	0 (0%)	0 (0%)	0 (0%)
ギャンブル	0 (0%)	0 (0%)	0 (0%)
料理	1 (0.4%)	0 (0%)	0 (0%)
スポーツ	226 (97%)	0 (0%)	1 (0.5%)
IT	4 (1.7%)	48 (100%)	25 (11%)
単語種数	1022	1357	1569
延べ単語数	3496	2793	6998

参考文献

- [1] 平成 16 年版 情報通信白書, 総務省
- [2] 五味崇, 安田裕樹, “インターネットコンテンツに対する法規制の実現可能性に関する検討”, UFJ Institute REPORT 2002.12 Vol.8 No.1 (2002)
- [3] “フィルタリング情報ページ”, 財団法人インターネット協会, <http://www.iajapan.org/rating/>
- [4] WinPcap, <http://winpcap.polito.it/>
- [5] 上田, 高井, “子供の Web アクセスにおける保護者監視支援システム”, 電気情報通信学会研究報告 Vol.105 No.193, pp.119-124, 2005
- [6] nkf, <http://www.ic.u-ryukyu.ac.jp/~kono/nkf/>
- [7] KAKASI, <http://kakasi.namazu.org/>