

子供の保護のためのインターネット利用監視支援システム

上田達巳[†]
北海道大学大学院
情報科学研究科[†]

高井昌彰[‡]
北海道大学
情報基盤センター[‡]

1. はじめに

近年インターネットの利用は極めて一般的なものとなっており、子供達もまたインターネットから様々な情報を取得する機会が増えている[1]. インターネットでは、有用な情報を得ることができる反面、意図せずに有害な情報に接触してしまう危険性が存在している. 有害な情報から子供達を保護するために、インターネットから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].

このため、受信者側で受け取る情報を選択または制限する方法、すなわちフィルタリングが用いられている[3]. しかし、フィルタリングには過度に情報を制限してしまい、本来有用となる情報を得られないという問題がある.

また、子供がインターネットに接続する場合は、操作をしている間、保護者がそばにいて監視することが推奨されているが、常に監視し続けることは難しく、子供がインターネットで何をしているかを十分に把握しきれていない場合もある.

本稿では、有害なインターネットの利用から子供たちを保護するための監視支援システムについて述べる. また、Windows でのプロトタイプの実装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

本システムは子供が直接利用する PC 内に設置し、Web を監視対象とする. Web 通信に有害な情報が含まれていないかどうか監視を行い、電子メールを用いて結果の要旨を保護者へと通知する.

2.2 端末の通信内容の取得

Web access watching support system for child's protection

[†]Tatsumi Ueda, Graduate school of Information Science and Technology, Hokkaido University

[‡]Yoshiaki Takai, Information Initiative Center, Hokkaido University

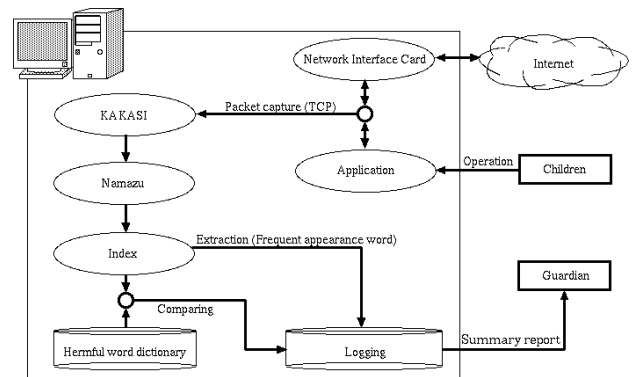


Fig. 1 監視支援システムの概要

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、パケットドライバ WinPcap を用いたツールを作成した[4].

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、ヘッダを取り除いて 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て TCP データを得る.

また、これらのヘッダを用いることで、TCP/IP およびアプリケーション層のプロトコルを模擬する. インターネットとアプリケーション間の通信を再現することにより、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、html やメールなどの暗号化されていないテキストデータを得る. 本稿では、これらの暗号化されていないテキストデータを用いる.

2.3 有害単語およびカテゴリ別単語の辞書化

情報が有害であるかどうかは、受信者によって大きく異なる. 子供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ならない.

本稿における有害な情報は、子供にとって有害な情報を意味するとし、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。また、子供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

キャプチャされたパケットに含まれるテキストデータに対しわかち書きを行い、単語ごとに区切る。Namazu[5]でテキストデータに含まれる単語のインデックスを作成する。インデックスと有害単語辞書、カテゴリ別単語辞書を比較し、送受信されたパケットにこれらの単語が含まれているかどうかを調べる。また、パケットに高い頻度で含まれる単語も調べて記録する。これにより、子供が有害な情報に接触していないかどうか、どのような情報に興味を持っているかを知ることができる。

2.5 保護者への通知

単語を検索した結果の要旨を保護者の電子メールアドレスに通知する。この要旨には、通信時間、転送量、パケット内の単語の傾向が示される。また、生のテキストデータが含まれていないため、子供のプライバシーを守ることもつながる。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ことで、保護者が子供の興味の変遷を知ることが支援される。

3. システムの評価実験

システムを評価するために、Impress Watch, Yahoo! Japan NEWS, 2ちゃんねる News カテゴリ, アダルトサイトをそれぞれ1時間閲覧し、検出された有害単語の分布を比較する。

Impress Watch と Yahoo! Japan News では IT に関連する単語、アダルトサイトでは性に関連する単語が多く検出された。また2ちゃんねるでは複数のカテゴリに渡って有害単語が少量ずつ含まれていた。

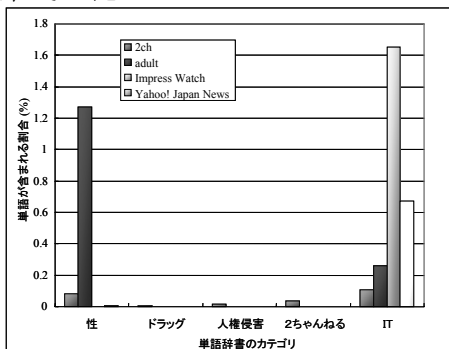


Fig. 2 辞書との比較結果

4. まとめ

インターネット利用監視システムの Windows プラットフォーム上への実装と有害単語検出の評価実験について述べた。

今後、パソコンにあまり詳しくない保護者に配慮し、簡単に操作できる GUI の実装を目指すとともに、パケットを検索して有害な情報と判別されたものを保護者が確認して妥当性を判断し、フィードバックすることで有害判別の精度の向上を図る。

性	: アダルトビデオ, イメクラ, ポルノ, 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)
2ちゃんねる	: きぼんぬ, ぬるぼ, もうだめぼ, インスコ, オマエモナー
IT	: イーサネット, エンコード, スパイウェア, 集積回路

Table. 1 有害およびカテゴリ辞書に含まれる単語の一例



Fig. 3 保護者への通知メール

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] Namazu, <http://www.namazu.org/>