

## 子どもの保護を目的としたインターネット利用監視支援システムの開発

上田 達巳<sup>†</sup> 高井 昌彰<sup>††</sup>

<sup>†</sup> 北海道大学大学院情報科学研究科 〒060-0814 北海道札幌市北区北14条西9丁目

<sup>††</sup> 北海道大学情報基盤センター 〒060-0811 北海道札幌市北区北11条西5丁目

E-mail: †{ueda,takai}@iic.hokudai.ac.jp

あらまし 近年、インターネットは子ども達にとって非常に身近なものになってきており、子ども達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。インターネットでは有益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ある。この危険を回避するために、子どものインターネット利用状況を監視しなければならないが、一方で子ども達のプライバシーにも配慮する必要がある。本稿ではインターネットに接続している子どものPC上で全ての処理を行う、インターネット利用監視支援システムについて述べる。イーサネットのパケットキャプチャによってインターネットアクセスの監視を常に行い、Webページなどから取得された単語を抽出し、異なるカテゴリーごとの辞書と比較することにより、アクセス傾向を解析した結果を、保護者に対して電子メールで定期的に報告する。本稿ではシステムのWindowsプラットフォーム上での実装と、パケットキャプチャ及びアクセス傾向分析による負荷の評価について述べる。

キーワード 監視支援, 子どもの保護, 有害情報, Webドキュメント

## Development of Internet Watchdog Systems for Children Protection

Tatsumi UEDA<sup>†</sup> and Yoshiaki TAKAI<sup>††</sup>

<sup>†</sup> Graduate school of Information Science and Technology, Hokkaido University Kita 14, Nishi 9, Kita-ku, Sapporo, 060-0814 Japan

<sup>††</sup> Information Initiative Center, Hokkaido University Kita 11, Nishi 5, Kita-ku, Sapporo, 060-0811 Japan

E-mail: †{ueda,takai}@iic.hokudai.ac.jp

**Abstract** For today's children, the Internet is one of the most familiar and useful information media. However, the Internet sometimes can provide information harmful to the children. Guardians need to watch their children's use of the Internet to avoid potential risks of contact with harmful information. On the other hand, we have to consider the children's privacy in communication. In this paper, we propose an Internet watchdog system which operates in every child's PC connected to the Internet. The watchdog system always checks the Internet access by Ethernet packet capturing, and samples the words from the downloaded texts such as Web pages. The collected words are immediately compared with the system's dictionaries of different categories, and then children's Internet access tendency is analyzed. This summary of analysis is periodically reported to the guardians via e-mail. We have implemented the Internet watchdog system on the Windows platform, and evaluated the overhead for packet capturing and access tendency analysis.

**Key words** Watch support, Child protection, Harmful information, Web document

### 1. はじめに

近年、急速なインターネット接続環境の整備により、日本国内におけるインターネット利用率は年々増加の一途を辿っている [1]。また、小中高校における情報機器の配備及び、インターネット接続環境の整備も進んでおり、情報教育やインターネッ

トを利用した教育も盛んに行われている。このように、現代の子ども達にとってインターネットは家庭・学校の双方において非常に身近な物となっている [2] [3] [4]。

インターネットでは有益な情報を入手できる反面、誰でも自由に情報を発信可能な事による有害情報の存在や、インターネットを利用した犯罪の増加が問題視されてきており、子ども

がリンクを辿っていくことで意図せずに有害情報に接触することや、操作ミスによる個人情報の誤発信、掲示板などで名誉毀損的な発言をしてしまい、收拾がつかなくなるといった被害に遭ってしまう可能性がある。また、子どもがインターネットに接続する際には、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、実際には常に監視し続けることは難しく、また保護者よりも子どもの方がパーソナルコンピュータの扱いに長けている場合もあり、子どもがインターネット上でどのような行動をしているのかを、保護者が十分に把握しきれていない場合もある。

これらの問題から子ども達を保護するために、有害な情報を含むコンテンツの発信を制限する法規制や、インターネットサービスプロバイダによる自主規制が行われているが、取り締まる事の難しさや“表現の自由”との関連もあり、現段階では情報を発信側から制御する方法には限界がある。このため子どもから有害な情報を遠ざけ、さらに有害な情報を取得していないかどうかを監視するために、現在フィルタリングソフトウェアの普及が進んでいる。

## 2. 従来のフィルタリングソフトウェアの問題点

フィルタリングソフトウェアでは受信側で受け取る情報を選択または制限する手法が用いられており、公的機関への導入推進および子どものいる家庭への導入の推奨が行われている。しかしながら、フィルタリングソフトウェアには取得した情報をすべて記録し管理者である保護者に対して伝えることにより、子どものプライバシーが犠牲となったり、ログデータの量が膨大になることで管理コストが非常に大きくなってしまいう問題がある。

## 3. 監視システムの開発

本稿では子どものプライバシーを保護しながらも、有害な情報を取得していないかどうか、どのような情報に興味を示しているかを保護者に報告するインターネット監視支援システムの開発について述べる。

### 3.1 想定環境

本システムは一般家庭において、子どもがインターネットに接続する環境を想定している [5]。具体的には、OSはMicrosoft WindowsのPCを利用し、インターネットでWebサイト閲覧、電子メールの送受信、チャットを行うという状況である。また、一般家庭におけるPCの利用目的を考慮に入れ、ローカルネットワーク上にサーバは存在しないと仮定する。

### 3.2 動作概要

#### 3.2.1 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PCにインストールして使用するホスト型の設置形式を取っており、インターネットとPC間で送受信されるパケットに含まれるWebドキュメントなどを監視対象とする。検出はコンテンツチェック方式を採用し、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を調べ、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。また、一連の処理

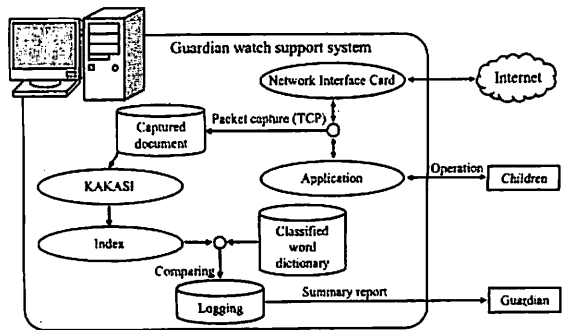


図 1: システムの概略図

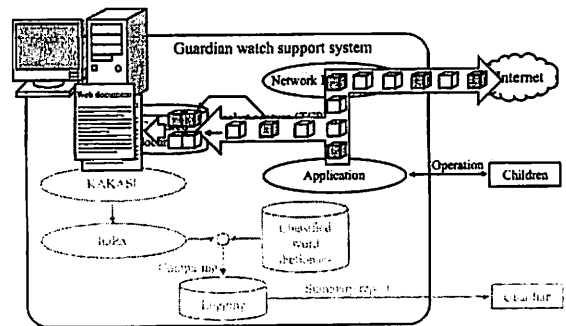


図 2: パケットの取得

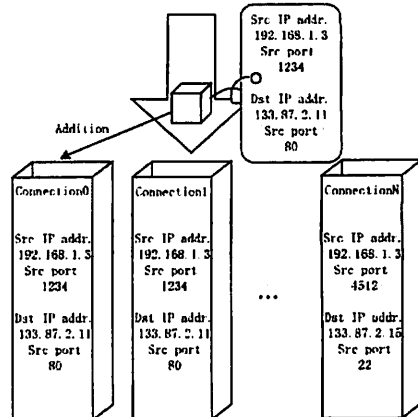


図 3: TCP ストリームの再構築

はスタンドアロンで行われる。

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事が望ましい。

#### 3.2.2 パケットの取得とストリームの再構築

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行うために、パケットドライバ WinPcap [6] を使用した (図 2)。

通常のネットワークアクセスにおいては、送受信される各

表 1: 辞書に含まれる単語数と単語の一例

カテゴリー	単語数	単語例
アダルト	182	アダルトビデオ, イメクラ, ボルノ, 痴漢
ドラッグ	77	コカイン, シンナー, ヘロイン, 大麻
人権侵害	38	(放送禁止用語などから抜粋)
ギャンブル	75	ガセ演出, 換金率, 等価交換
2ちゃんねる	893	きばんぬ, ぬるぼ, もうだめば, インスコ, オマエモナー
IT	1962	イーサネット, エンコード, スパイウェア, 集積回路
料理	402	油通し, グラッセ, シチュー, 割りした
スポーツ	550	アーチェリー, フリースローライン, 準備運動

パケットに対して階層化された各プロトコルスタックがパケットヘッダに基づいた適切な処理を行うことにより、アプリケーション側からはソケットを用いての容易なデータアクセスを行うことが可能となっているが、WinPcapにより取得されるパケットはPC上で動作している全てのネットワークアプリケーションが送受信するものが混在しており、直接取り扱うことは困難であり、IP、TCPヘッダを用いてコネクションの処理を行った(図3)。

### 3.2.3 テキストドキュメントの取得

3.2.2において処理が行われ渡されたデータには、テキストデータと画像・音声などのバイナリデータが混在してしまっている。Webドキュメントでは日本語の文字コードとして、SHIFT-JIS、EUC-JP、UTF-8、UTF-16、ISO-2022-JPなどが用いられており、情報を取得するためには文字コード変換を行う必要があるが、この際にバイナリデータの存在が問題となるため取り除く必要がある。また、場合によってはデータが圧縮されている場合もあるため、適切な解凍の処理を行った。

取得後のドキュメントは、漢字フィルタ nkfを用いて日本語版 Windowsのネイティブな文字コードである、SHIFT-JISへと文字コード変換を行う[7]。その後、表示書式を記述するタグ情報部分の削除を行った。

### 3.2.4 傾向分析

子どもがインターネットからどのような情報を取得しているのかを判断するために、ドキュメントに含まれる単語をカテゴライズする処理を行った。本システムで使用する単語辞書のカテゴリーと、それぞれの辞書に含まれる単語数と単語の一例を表1に示す。

本システムではパケットから取得されたドキュメントに対して、kakasi[8]を用いてわかち書きを行い、ドキュメントに含まれる単語の取得を行った(図5)。

わかち書きによって得られた単語は、それぞれのカテゴリーの情報を表現する際に用いられる単語が記述されている単語辞書と比較され一致した数がスコアとしてカウントされる。このスコアは、検出を行った時間と共にログへと書き出される(図6)。

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、一定期間に取得されたドキュメントのスコアの合計値に対する、各カテゴリーごとのスコアの割合により求められる。

### 3.2.5 報告メールの作成と送信

辞書との比較を行った結果を基に、報告メールの作成を行っ

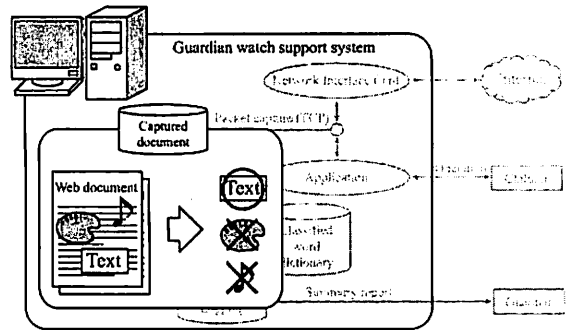


図 4: テキストドキュメントの取得

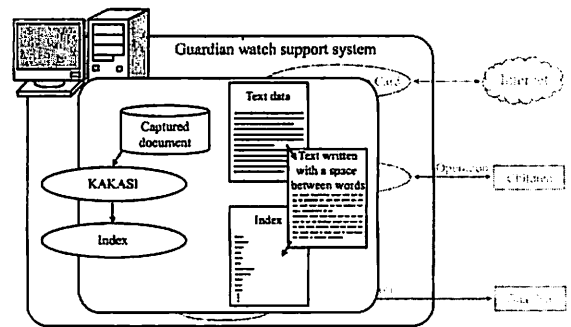


図 5: テキストのわかち書き

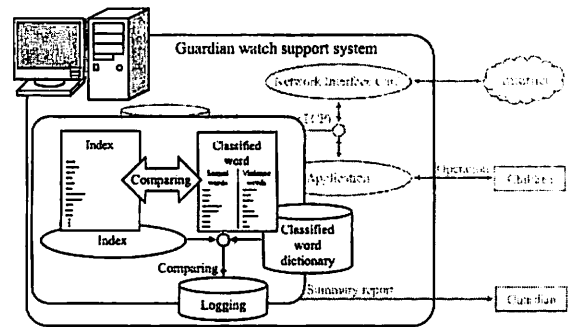


図 6: 辞書との比較

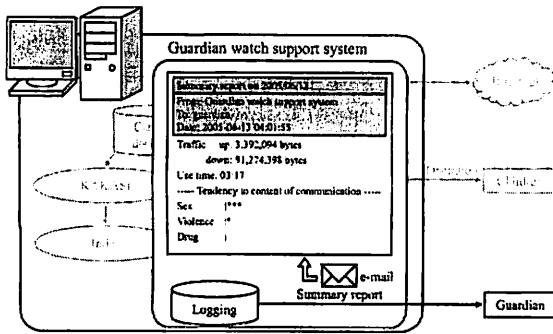


図 7: 保護者への通知

た。報告メールにはパソコンの利用時間、インターネットと送受信した通信データ量、Web ドキュメントに含まれる情報の傾向が記述されている。

ログから過去一週間のパソコンの利用時間、インターネットとの通信量および通信に含まれる情報の傾向をテキストデータに出力する。通信に含まれる情報の傾向は、全カテゴリーのスコアの総計に対する各カテゴリーごとのスコアの割合によって求められ、この割合から横棒グラフを作成する。

作成された報告メールは、あらかじめ登録しておいた保護者の電子メールアドレスに対して送信される。送信されたメールを図 8 に示す。この報告メールには、インターネットと送受信した Web ドキュメントに含まれる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りながらも、子どもがどのような情報を取得しているかを確認する事ができる。

### 3.3 システムの実装

Windows プラットフォーム上で、Microsoft Visual C++ .NET 2003 を用い実装を行った。本システムはバックグラウンドで常にパケットキャプチャを行いながら、取得が終わった Web ドキュメントへの自然言語処理を並行して行うため、マルチスレッドプログラミングによる実装を行っている。プログラムサイズは 148Kbytes、辞書のサイズは 8 カテゴリー 4179 語 (59.1Kbytes) である。ライブラリとして winpcap, nkf, kakasi, safestr, zlib を使用している。

## 4. システムの評価実験

### 4.1 通信速度、CPU 負荷および処理速度への影響

本システムを用いた場合、監視対象のパーソナルコンピュータにおいて通信速度および CPU 負荷にどの程度の影響を与えるか測定を行った。実験に用いたサーバ、PC およびネットワークについて表 2、3 に示す。

サーバとホスト間のネットワークについて図 9 に示す。サーバとホストは、それぞれギガビットイーサネット接続されており、ギガビットのスイッチングハブを間に挟むような形になっている。

#### 4.1.1 TCP レイヤにおける影響

TTCP を用いてサーバ・ホスト間で TCP レイヤでの通信を行い、通信速度およびホストでの CPU 使用率への影響を測定

した。実験環境は以下の通りである。

- 測定内容
  - － サーバ・クライアント間の通信速度
  - － クライアントの CPU 使用率
- 測定に用いるプログラム・データの概要
  - － 使用アプリケーション：TTCP
  - － 送信データ長：81920Bytes
  - － 送信データ数：8192 個
  - － 使用ポート番号：5001
  - － 通信内容：ASCII 文字の羅列
- 比較条件
  - － システム未使用時
  - － システム使用時 (パケットキャプチャ、TCP ストリーム再構築)

システム未使用時、システムでパケットキャプチャのみを行った場合およびパケットキャプチャと TCP ストリーム再構築を行った場合の比較を行う。

TTCP による計測を行った結果について、図 10-11 に示す。システム未使用時の TCP レイヤにおける平均速度は 734Mbps となった。システムにおいて WinPcap により TTCP のパケットをキャプチャし、そのまま破棄した場合の平均速度は 645Mbps であり 12.2% の速度低下となった。また、キャプチャ後に TCP ストリームの再構築まで行った場合は、606Mbps となり 17.6% 程度の速度低下が見られた。

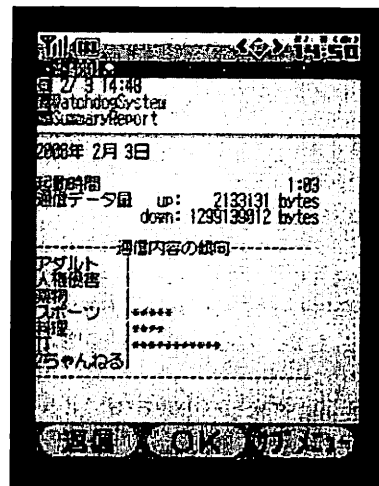


図 8: 保護者への報告メール

表 2: 通信速度、CPU 負荷への影響測定時の実験環境 (サーバ)

CPU	Intel Pentium4 processor 2.4GHz
Memory	DDR SDRAM 512MBytes
Ethernet Card	Intel PRO/1000 GT Desktop Adapter
OS	Gentoo Linux 2005.1-r1 (Linux Kernel 2.6.14)

表 3: 通信速度, CPU 負荷への影響測定時の実験環境 (クライアント)

CPU	Intel PentiumM processor 1.2GHz
Memory	DDR2 SDRAM 1280MBytes
Ethernet Card	Marvell Yukon 88E8053 PCI-E Gigabit Ethernet Controller
OS	Windows XP Professional SP2

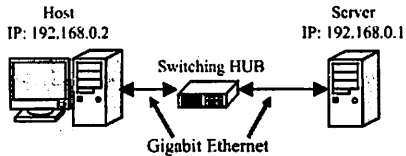


図 9: サーバとホスト間のネットワーク図

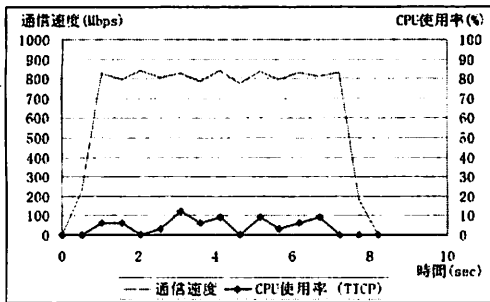


図 10: TCP レイヤにおける影響: システム未使用時

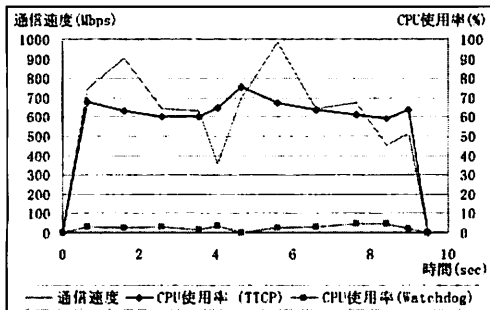


図 11: TCP レイヤにおける影響: システム使用時

#### 4.1.2 HTTP レイヤにおける影響

サーバ上で動作している Apache2 から Wget を用いてドキュメントのダウンロードを行い, 通信速度およびホストでの CPU 使用率への影響を測定した。実験環境を以下に示す。

- 測定内容
  - サーバ・クライアント間の通信速度
  - クライアントの CPU 使用率
- 測定に用いるプログラム・データの概要
  - HTTP サーバ: Apache 2.0.55-r1
  - HTTP クライアント: Wget 1.9
  - HTTP バージョン: 1.0
  - 転送に用いたファイル: <http://madeira.cc.hokudai.ac.jp/>

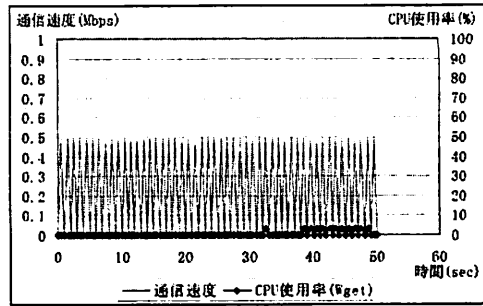


図 12: HTTP レイヤにおける影響: システム未使用時

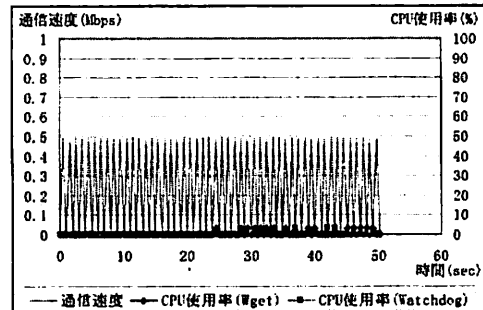


図 13: HTTP レイヤにおける影響: システム使用時

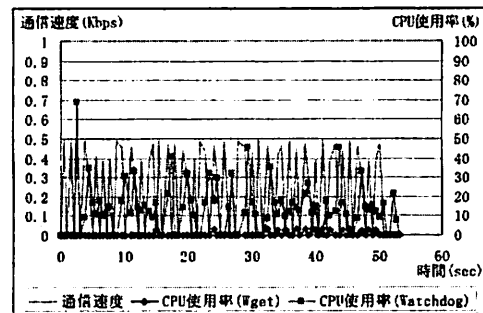


図 14: 自然言語処理による影響: KAKASI

のトップページのミラー

- 転送ファイル形式: HTML 形式
- 転送ファイルサイズ: 28KBytes
- 転送回数および頻度: 1 秒おきに 50 回
- 比較条件
  - システム未使用時
  - システム使用時 (パケットキャプチャ, TCP ストリーム再構築, Web ドキュメント取得)

計測は <http://madeira.cc.hokudai.ac.jp/> のトップページのミラーをサーバ上に作成し, 1 秒間隔で転送を行い, システム未使用時とシステムで Web ドキュメント取得までの処理をした場合の比較を行っている。なお, この実験ではシステムは自然言語処理は行わない。

表 4: 傾向測定の結果

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!すすきの 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	228 (97.0%)	0 (0%)	1 (0.5%)
IT	4 (1.7%)	48 (100%)	25 (11.0%)
単語種数	1022	1357	1569
延べ単語数	3496	1312	1331

Apache から Wget でドキュメントをダウンロードして計測を行った結果について、図 12-13 に示す。通信速度および CPU 使用率への影響は皆無であった。

#### 4.1.3 自然言語処理による影響

HTTP の処理を行い取得したドキュメントに対して自然言語に関する処理を行った場合の影響の測定を 4.1.2 と同様の HTTP サーバ・クライアント、ドキュメント、転送回数・頻度の条件の下に行った。

取得した Web ドキュメントに対して辞書との比較を行って計測した結果について、図 14 に示す。若干、CPU への負荷の増加が見られるが、ほぼリアルタイムで処理を行うことができていることがわかる。

#### 4.2 Web サイトを用いたの検出実験

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! すす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 4 に示す。

辞書と比較する事により検出できた全単語のうち、NIKKEI NET スポーツ総合においてはスポーツに関連する単語が 97%、スラッシュドットジャパンでは IT に関連する単語が 100%、Oh-Yeah!すすきの情報リンクではアダルトに関連する単語が 88.5%を占めており、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ることがわかる。また、傾向の測定に用いられた単語は取得された単語のうち 1.1%~6.6%程度となっている。

## 5. まとめ

本稿では子どものプライバシーを守りながらも、子どもがインターネットから取得したドキュメントに含まれる情報の傾向を判別し報告する事で、子どもが有害な情報に接触していないかどうか、どのようなことに興味を示しているかを保護者に対して伝え、保護者が子どもに対してさらなる理解を深めることを手助けすることができるホスト型のインターネット利用監視支援システムの開発について述べた。

しかしながら、本システムで用いている単語辞書は 8 カテゴリーであるため、取得可能な情報は限られている。インターネット上に存在する多くの情報を検出するためには、さらなる

単語辞書の拡充および整理が必要である。また、SMTP、POP といった電子メールの他、チャットのプロトコルに対応することにより、より広範囲に渡って子どもの保護のための監視を行うことを目指す。

#### 参考文献

- [1] 平成 16 年度版 情報通信白書、総務省
- [2] “学校における情報教育の実態等に関する調査結果”、文部科学省、平成 15 年度
- [3] “有害サイトブロックに関するアンケート”、パナソニックネットワークサービス株式会社、トレンドマイクロ株式会社、<http://www.trendmicro.com/jp/about/news/pr/archive/2004/news040705.htm>
- [4] “インターネット活用のための情報モラル指導実例集”、財団法人コンピュータ教育開発センター
- [5] “家庭教育におけるテレビメディア調査/青少年とインターネット等に関する調査”、(社)日本 PTA 全国協議会、平成 16 年度
- [6] WinPcap、<http://winpcap.org/>
- [7] nkf のページ、<http://www.io.u-ryukyu.ac.jp/~kono/nkf/>
- [8] KAKASI、<http://kakasi.namazu.org/>