

# 子どもの保護を目的とした Web アクセス監視支援システム

上田 達 巳<sup>†1</sup> 高井 昌 彰<sup>†2</sup>

近年、インターネットは子どもたちにとって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。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保護者が監視することが望まれるが、一方で子どもたち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 上に実装される Web アクセス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つねに行い、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およびアクセス傾向分析によるシステムの評価について述べる。

## Web Access Watchdog Systems for Children Protection

TATSUMI UEDA<sup>†1</sup> and YOSHIAKI TAKAI<sup>†2</sup>

For today's children, the Internet is one of the most familiar and useful information media. However, there is harmful information somewhat in the Internet. Parents 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 a Web access watchdog system which operates on 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 downloaded text of Web pages. The collected words are immediately compared with the systems' dictionaries of different categories, and then children's Internet access tendency is analyzed. A summary of the access tendency analysis is periodically reported to the parents via e-mail. We have implemented the watchdog system on a Windows platform, and evaluated the overhead for packet capture and access tendency analysis.

### 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たどっている<sup>1)</sup>。また、小中高校における情報機器の配備および、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さかんに行われている。このように、現代の子どもたち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ものとなっている<sup>2),3)</sup>。

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ことによる有害情報の

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、子どもがリンクをたどっていくことで意図せずに有害情報に接触したり、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ったりする可能性がある。また、子どもがインターネットを利用する際には、保護者がそばにいて監視することが推奨されているが、実際にはつね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 PC 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が多い。

これらの問題から子どもたちを保護するために、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、取り締まることの難しさや“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

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

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

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められている<sup>4)</sup>．

## 2. 既存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者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、学校などの公的機関や子どものいる家庭への導入が推奨されている<sup>5)</sup>．しかしながら、フィルタリングソフトウェアが取得したアクセス情報をすべて管理者である保護者に伝えることにより、子どものプライバシーが犠牲となったり、ログデータの管理コストが非常に大きくなってしまったりするという問題点がある．

既存のフィルタリングソフトウェアとしては、CYBERsitter<sup>6)</sup>、i-フィルター4<sup>7)</sup>、InterSafe Personal<sup>8)</sup>などがある．また類似研究として、井ノ上らのフィルタリングソフトの開発<sup>9)</sup>や、西埜らのレーティング、フィルタリング技術の開発<sup>10)</sup>が行われている．これらと本稿で提案する Web アクセス監視支援システム(以下、本システムという)との特徴の比較を表 1 に示す．

既存のフィルタリングソフトウェアは主に URI に基づいた監視を行っており、アクセス先に関する情報が登録されたデータベースから、スコアやレートを取得し判定を行っている．しかしながら、Web サイトが頻繁に更新される場合や作成されて間もない場合では、データベース内に情報が存在しないことや、存在していても実際のドキュメントの内容と食い違っていることがある．そのため、有効なデータベースの作成と保守には、非常に大きな労力とコストを要することとなる．また、使用できるブラウザの種類やバージョンに制限があったり、追加のプロキシ設定が必要だったりするソフトウェアも多く、運用に支障を来す場合もある．

本システムでは、子どもが実際に取得を行っている Web ドキュメントの内容から直接情報を抽出し、異なるカテゴリごとの辞書と比較して一致した単語数をスコアとしてカウントすることによりアクセス傾向を解析した結果を、保護者に対して定期的に報告する．保護者が子どもの動向を把握し、監視を支援することが目的のため、アクセスのフィルタリングは行っていない．ホストが送受信しているパケットを内部でキャプチャして通信内容を取得するため、暗号化されていない通信は、ブラウザの種類にかかわらず判定可能

表 1 既存のフィルタリングソフトウェアとの比較  
Table 1 Comparison with the existing filtering software.

	監視		プライバシーの保護	ブラウザ非依存	Proxy 設定不要	カテゴリ数
	URI	内容				
CYBERsitter5	○	○	×	△	△	33
i-フィルター4	○	○	×	△	△	33
InterSafe Personal	○	×	×	○	○	72
フィルタリングソフト(井ノ上ら)	○	○	×	○	×	2
フィルタリングソフト(西埜ら)	○	×	△	△	○	4
本論文で提案するシステム	△	○	○	○	○	7

ある．また、1度判定を行った URI に対するスコアをキャッシュすることで、ドキュメントが更新されていない場合には辞書との比較処理を省くことができるため高速にスコアリングを行うことも可能である．

また、既存のフィルタリングソフトウェアにおけるログは詳細にわたっており、各レコードはアクセス日時、URI、検出されたカテゴリや単語などの具体的な項目からなり、アクセス回数分のレコードが生成されるため俯瞰性に乏しい．これに対し、本システムでは起動時間と通信データ量のほか、通信内容の傾向のみが 1 時間ごとにログに記録される．具体例として一般的なユーザのインターネット利用状況を考慮し<sup>11)</sup>、1カ月あたりのインターネット利用時間が 20 時間で、閲覧ページ数が 2,200 ページであるような環境を仮定すると、i-フィルター 4 ではアクセス 1 回あたり約 250 Bytes 程度のレコードが生成されるため、総計約 550 KBytes のログサイズとなる．一方、本システムの生成するログは 1 時間あたり約 80 Bytes 程度であるため、20 時間分では約 1.6 KBytes となりきわめてコンパクトである．

## 3. Web アクセス監視支援システム

### 3.1 想定環境

本システムは一般家庭において、子どもが PC からインターネットに接続する環境を想定している<sup>12)</sup>．具体的には、Microsoft Windows を搭載した PC を利用し、インターネットで Web サイトの閲覧や掲示板への書き込みを行うという状況である．

### 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とっており、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 ドキュメントから抽出された単語がシステム内のカテゴリ別の辞書に含まれるかどうかを調べることで有害な情報が含まれていないかどうか

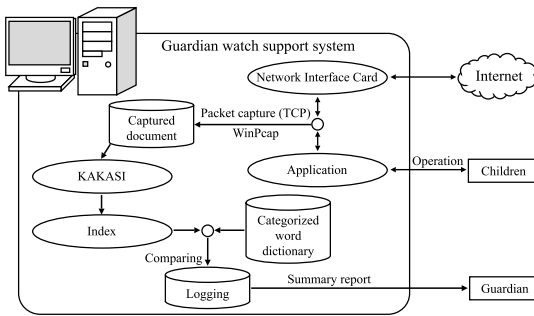


図 1 Web アクセス監視支援システムの概略図

Fig. 1 Architecture of Web access watchdog system.

を判定し、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。また、一連の処理はスタンドアロンで行われる。

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と拡張性を考慮し、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<sup>13)</sup>を使用した。WinPcap はインターネットと Web ブラウザなどのアプリケーションの間でやりとりされるパケットのコピーを生成し、アプリケーションの通信内容をシステム内部で取得するものである。

通常のネットワークアクセスにおいては、送受信される各パケットに対して階層化された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ソケットを用いたデータアクセスを容易に行うことが可能となっている。WinPcap で取得されるパケットには PC 上で動作しているすべ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いるため、パケットの IP および TCP ヘッダに記載されている IP アドレスとポート番号を基準に、Web アクセスに使用されているコネクションの再構築処理を行い、Web トラフィックを取得した(図 2)。

3.2.3 テキストドキュメントの取得

取得した Web トラフィック中には、検出の対象となる日本語で記述された Web ドキュメントのほか、圧縮された Web ドキュメントや、画像・音声といったバイナリデータなどの一部が含まれている。圧縮された Web ドキュメントは適切な解凍処理を行うことで検出が可能である一方、バイナリデータは不要な処理コストが発生するため、Web トラフィック中から Web ドキュメントだけを抽出することが必要となる。そこで Web トラフィックに対し、HTTP のプロトコルを

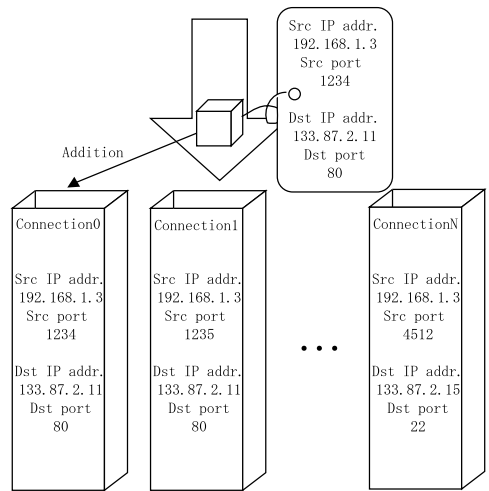


図 2 TCP ストリームの再構築

Fig. 2 Reconstruction of TCP streams.

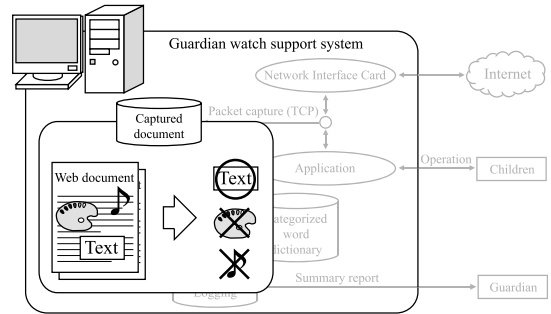


図 3 テキストドキュメントの取得

Fig. 3 Acquisition of text documents.

解釈し、Web ブラウザと同様の動作をエミュレートすることで、Web ドキュメントの取得を行った。

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、テキスト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。また、Web トラフィックから画像、音声などのバイナリデータ部分を取りのぞき、圧縮された Web ドキュメントの解凍処理を行う(図 3)。

取得後のドキュメントは、漢字フィルタ nkf<sup>14)</sup>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。その後、表示書式を記述するタグ情報部分の削除を行う。

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別するために、ドキュメントに含まれる単語をカテゴリ化する。本システムで使用する

表 2 辞書に含まれる単語数と単語の一例

Table 2 Examples of words included in the dictionaries.

カテゴリ	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
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

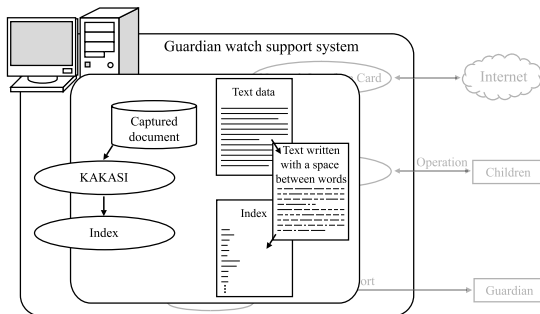


図 4 テキストのわかち書き

Fig. 4 Leaving a space between words of a sentence.

単語辞書のカテゴリと、それぞれの辞書に含まれる単語数および単語の一例を表 2 に示す。単語辞書はインターネット上で公開されている分野別辞書サイトから索引項目を抽出した後に、一般的に用いられる単語と複数カテゴリに分類されると思われる単語を除去して作成した。また、隠語的表現として用いられる単語を他カテゴリの誤検出とならない範囲で追加している。新しい単語に対応し、傾向の分析をより確かなものにするためには、カテゴリ別に分けられている単語辞書に対象となる新たな単語を追加して辞書を拡充すればよい。

はじめに、パケットから取得されたドキュメントに対して、kakasi<sup>15)</sup>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と各単語の出現回数を取得する(図 4)。ドキュメント内から得られた単語は、各カテゴリの辞書に含まれるかどうかそれぞれ検索され、辞書にヒットした単語の出現回数の総和をそのカテゴリにおけるスコアとする。カテゴリ別のスコアの集計値は、1 時間ごとにログへと書き出される(図 5)。

また、サーバから取得したドキュメントの更新時刻、ハッシュ値および URI とともにスコアをキャッシュに保存し、同じドキュメントが取得された場合には、キャッシュ中のスコアが参照され、ドキュメントに含まれる単語を検索する処理を省くことにより高速化を

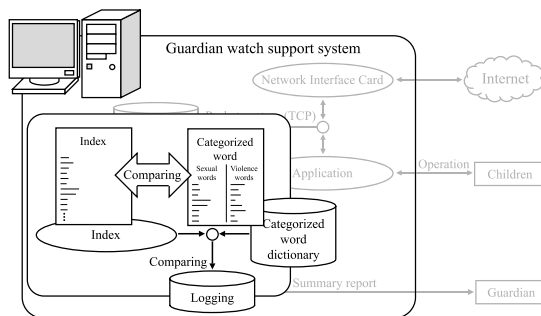


図 5 辞書との比較

Fig. 5 Comparison with words of the dictionaries.

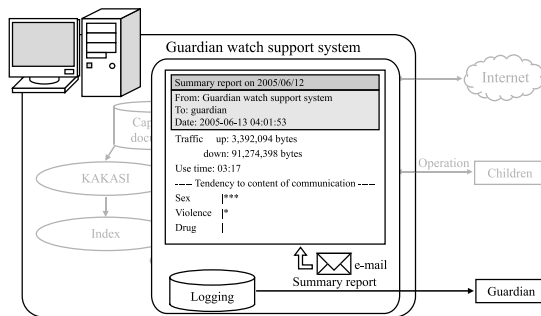


図 6 保護者への通知

Fig. 6 Notice to the guardian.

図っている。

### 3.2.5 報告メールの作成と送信

ログから過去 1 週間の PC の利用時間、インターネットとの通信量および Web ドキュメントに含まれる情報の傾向をテキストデータに出力する(図 6)。アクセス情報の傾向は、全カテゴリのスコアの総計に対する各カテゴリごとのスコアの割合によって求められ、この割合からグラフを作成する。このグラフは、子どもがアクセスした Web ドキュメントに、どのカテゴリに属する単語が多く含まれているかの大きな分布を示している。

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。システムから送信された報告メールの一例を図 7 に示す。この報告メールには、送受信された Web ドキュメント内のテキストデータがいっさい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ことができる。

### 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

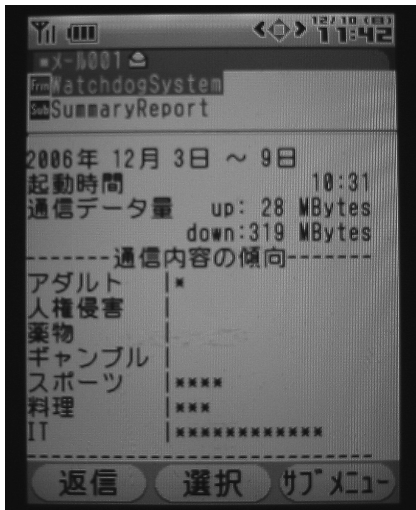


図 7 保護者への報告メール

Fig. 7 Report email to the guardian.

バックグラウンドでつね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 148 Kbytes、辞書のサイズは 7 カテゴリ 3,294 語 (48.5 Kbytes) 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 通信速度および CPU 負荷への影響

本システムを用いた場合、監視対象の PC 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。表 3 に実験で用いた Web サーバと PC のスペックを示す。また、サーバと PC 間のネットワーク構成を図 8 に示す。

##### 4.1.1 TCP ストリーム再構築まで行った場合における影響

TTCP<sup>\*1</sup>を用いて TCP レイヤでの通信を行い、通信速度および PC での CPU 使用率への影響を測定した。

送信データ長 81.92 KBytes, 送信データ数 8,129 個, 使用ポート番号 5001 の条件で TTCP による計測を行った。一般的な Web トラフィックにおいては、Web ドキュメントのリンク選択の操作時間や、Web ブラウザ内における HTTP の解釈と、描画などの処

\*1 2 台のパソコン上で TCP もしくは UDP を利用してコネクションを張り、データが含まれるパケットを 1 方向に送信して、ネットワークのデータ転送のパフォーマンスを測定するコマンドラインツール。

表 3 通信速度および CPU 負荷の評価実験環境

Table 3 Conditions of the evaluation experiments.

Device	Component	Specification
Web Server	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)
PC	CPU	Intel PentiumM processor 1.2GHz
	Memory	DDR2 SDRAM 1280MBytes
	Ethernet Card	Marvell Yukon 88E8053 PCI-E GE Controller
	OS	Windows XP Professional SP2

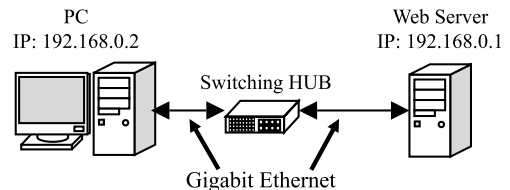


図 8 Web サーバと PC 間のネットワーク図

Fig. 8 Network of Web server and PC.

理遅延による影響があり、データの転送には待機時間が発生する。ここでの実験パラメータは Web ブラウザ操作時間や HTTP の処理遅延によるオーバヘッドを考慮に入れず、Web コンテンツに含まれる JPEG 画像 1 個分程度のデータを待機時間なしで大量に転送し、ネットワーク帯域幅を限界近くまで利用している状況に等しい。計測の結果、システム未使用時の平均速度は 734 Mbps であるのに対し、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 645 Mbps であった。

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、606 Mbps となり 17.6%程度速度低下が見られた。

PC の性能限界に近い通信速度が必要となる場合、本システムによる速度低下が悪影響を及ぼすことも考えられるが、一般家庭におけるインターネット利用には十分耐える通信速度を確保しているといえる。実際、家庭からの Web アクセスにおいては、プロバイダとの接続帯域および遅延がボトルネックとなり、Web サーバとの通信速度が大きく低下することが多く、事実上問題とはならない。

##### 4.1.2 HTTP のエミュレートまで行った場合における影響

PC 上のアプリケーションとして Wget<sup>\*2</sup>を用いて、サーバ上で動作している Apache2 から Web ドキュメントのダウンロードを行い、通信速度および PC での

\*2 HTTP や FTP のプロトコルに従って、サーバからファイル取得を行うコマンドラインツール。

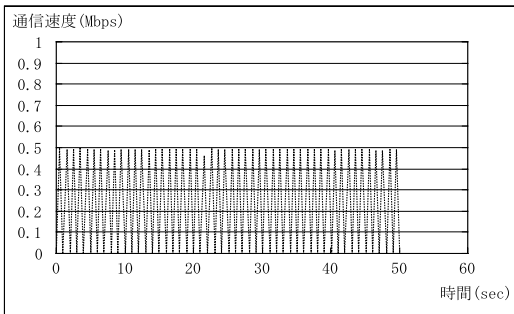


図 9 システムによる影響：システム未使用時  
Fig.9 Influence of the system (system nonuse).

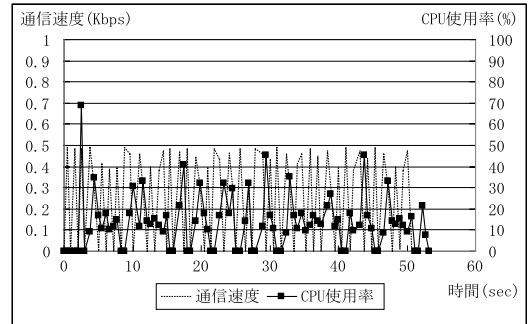


図 11 システムによる影響：システム使用時（システム全体）  
Fig.11 Influence of the system (whole system in use).

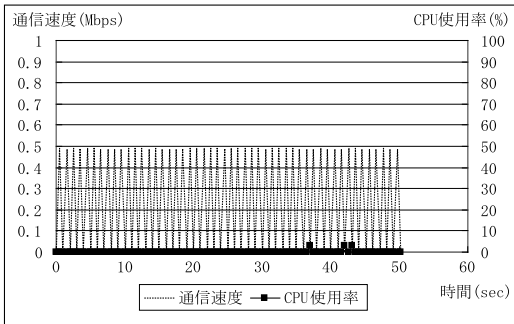


図 10 システムによる影響：システム使用時（HTTP のエミュレートまでの処理）  
Fig.10 Influence of the system (HTTP emulation).

CPU 使用率への影響を測定した。Wget は Web サーバ上のトップページ（サイズ 28 KBytes）を 1 秒間隔で 50 回ダウンロードする。一方、Web アクセス監視支援システムでは辞書を用いた分析処理は行わないよう機能制限を施し、Web ドキュメント取得までの処理のみを行う。

システム未使用時の通信速度を図 9 に、システム使用時の通信速度と CPU 使用率を図 10 に示す。ここでの通信速度とは Wget が Web ドキュメントをダウンロードする際のデータ転送の速度を意味し、CPU 使用率は Web アクセス監視支援システムのプロセスが CPU に与える負荷を表す（Wget のプロセスは含まれていない）。Wget は 1 秒間隔で Web ドキュメントにアクセスするため、システム未使用時の通信速度は 0.5 Mbps をピークに、周期的に変化している。システムを稼働させ TCP ストリームの再構築および Web ドキュメントの取得を行うと、その CPU 使用率は平均 0.09% であり、通信速度には平均 0.15% の低下しか見られなかった。

本システムが HTTP プロトコルをエミュレートし、Web ドキュメントを取得するまでの負荷は非常に小

さいといえる。

#### 4.1.3 システム全体での影響

取得した Web ドキュメントに対する分析処理を含む監視支援システム全体での影響を 4.1.2 項と同様の条件の下で測定した。キャッシュは無効にし、Web ドキュメントの内容に対して、毎回辞書との比較処理を行っている。

図 11 にシステム全体を稼働させた際の通信速度と CPU 使用率を示す。図 9、10 の結果と比較すると、システム全体では CPU 使用率は平均 13.59% まで増加し、通信速度は 3.5% 低下している。しかしながら、1 回あたりの Web ドキュメントのダウンロードに余分にかかった時間は約 0.06 秒であり、Web アプリケーションの利用者が体感するのが難しい程度の時間であるため、ほぼリアルタイムに監視支援処理を実行できているといえる。

このことから、本システムは監視対象の PC に過大な CPU 負荷をかけることなく動作し、Web ブラウザ利用者にアクセス遅延を感じさせないことが分かる。

#### 4.2 Web サイトを用いての検出実験

アクセス傾向の検出能力を評価するため、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<sup>16)</sup>、YOMIURI ONLINE スポーツ<sup>17)</sup>、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<sup>18)</sup>、ITmedia News<sup>19)</sup>、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<sup>20)</sup>、動宝<sup>21)</sup>、子ども向けサイトであるキッズ goo<sup>22)</sup>、Yahoo!きっず<sup>23)</sup> 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。

辞書と比較すること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 98.68%、YOMIURI ONLINE スポーツにおいては 91.75%、スラッシュドットジャパンでは IT に関連する単語が 95.48%、ITmedia News で

表 4 傾向測定の結果

Table 4 Results of tendency measurement.

	NIKKEI NET スポーツ総合	YOMIURI ONLINE スポーツ	スラッシュドット ジャパン	ITmedia News	Oh-Yeah! すずきの情報リンク	動宝	キッズgoo	Yahoo!きっず
アダルト	0 (0%)	0 (0%)	2 (0.48%)	9 (0.54%)	625 (82.89%)	1218 (86.38%)	0 (0%)	0 (0%)
人権侵害	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	2 (0.14%)
薬物	2 (0.05%)	0 (0.00%)	1 (0.24%)	1 (0.06%)	0 (0%)	7 (0%)	0 (0%)	1 (0.07%)
ギャンブル	36 (0.90%)	3 (0.10%)	0 (0%)	2 (0%)	0 (0%)	0 (0%)	0 (0%)	2 (0.14%)
料理	2 (0.05%)	0 (0.00%)	0 (0%)	1 (0%)	2 (0.27%)	25 (1.77%)	31 (9.37%)	65 (4.65%)
スポーツ	3950 (98.68%)	2737 (91.75%)	16 (3.81%)	42 (2.54%)	0 (0%)	41 (3%)	52 (15.71%)	251 (17.95%)
IT	13 (0.32%)	243 (8.15%)	401 (95.48%)	1598 (96.67%)	127 (16.84%)	119 (8.44%)	248 (74.92%)	1077 (77.04%)
合計	4003	2983	420	1653	754	1410	331	1398

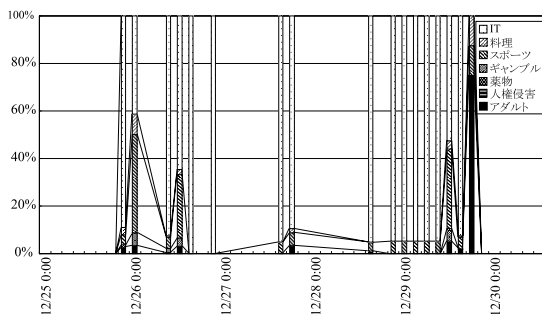


図 12 Web アクセスの嗜好変遷の一例

Fig. 12 An example of preference changes in Web access.

は 96.67%，Oh-Yeah!すずきの情報リンクではアダルトに関連する単語が 82.89%，動宝では 86.38%を占めており，これらの Web サイトにおいて，主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分かる．また，キッズ goo と Yahoo!きっずで検出された単語は料理，スポーツ，IT といった一般的なカテゴリのものであったが，Yahoo!きっずでは，無害な単語をシステムが複数にわかち書きすることにより，有害な人権侵害の単語として認識してしまう検出ミスが確認された．傾向の測定に用いられた単語は，ドキュメントから取得された全単語のうちの 0.4%～3.7%程度となっている．

#### 4.3 Web アクセスの嗜好変遷

本 Web アクセス監視支援システムによる，Web アクセスの時間推移ともなう傾向分析の様子を確認するため，著者の所属する研究室の大学院学生 1 名を被験者として，2006/12/25～2006/12/30 の期間で，研究室からインターネットへの Web アクセスの監視を行った．3 時間ごとにまとめられた Web アクセスの傾向分析結果の変遷を図 12 に示す．結果は検出されたスコアの合計値で正規化してある．期間中に取得された情報の 85.1%が IT カテゴリのものであったが，スポーツ，料理，ギャンブル，アダルトといった情報も取得されており，Web アクセスにおける嗜好の変遷がカテゴリの割合の時間変化として現れている．

このことから，本システムを用いることにより，Web アクセスにおける子どもの興味対象の移り変わりを保護者が知る手助けが可能であることが分かる．

## 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ことで，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 Web アクセス監視支援システムの開発について述べた．

本システムは低負荷で動作するため利用者にストレスを与えることがなく，子どものプライバシーに配慮した形で，保護者に子どもの Web アクセスの状況を知ることができる．また，時間とともに推移する子どもの興味対象を提示することで，保護者が子どもの挙動に対する理解を深める手助けが可能である．

しかしながら，本システムで用いている単語辞書は 7 カテゴリであるため，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が，辞書の整備にはコストを要するため，Web ドキュメントから取得した情報からシステムが辞書を自律的に構築することが必要となる．また，SMTP，POP といった電子メールのほか，チャットのプロトコルに対応することにより，さらに広範囲の監視支援を行うことを目指す．

## 参 考 文 献

- 1) 平成 18 年度版情報通信白書，総務省．
- 2) 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>
- 3) 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
- 4) 国文, 清水: インターネットにおけるコンテンツ・レイティングとフィルタリング, 情報処理, Vol.40, No.1, pp.57-63 (1999).
  - 5) 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/>
  - 6) CYBERSitter.  
<http://www.iqs-j.com/filtering/cyber5/>
  - 7) i-フィルター 4. <http://www.daj.jp/cs/>
  - 8) InterSafe Personal. [http://www.alsi.co.jp/security/isp/isp\\_01\\_gaiyou.html](http://www.alsi.co.jp/security/isp/isp_01_gaiyou.html)
  - 9) 井ノ上, 帆足, 橋本: 文書自動分類手法を用いた有害情報フィルタリングソフトの開発, 電子情報通信学会論文誌, Vol.J84-DII, No.6, pp.1158-1166 (2001).
  - 10) 西埜, 西塚, 苗村: 学校教育における有害な情報の検出とレイティング, フィルタリング技術の開発, 情報処理学会研究報告, Vol.2001, No.34, pp.39-46 (2001).
  - 11) comScore Publishes the First Comprehensive Review of Asia-Pacific Internet Usage, comScore Inc. <http://www.comscore.com/press/release.asp?press=1520>
  - 12) 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.
  - 13) WinPcap. <http://winpcap.org/>
  - 14) nkf のページ. <http://www.ie.u-ryukyu.ac.jp/kono/nkf/>
  - 15) KAKASI. <http://kakasi.namazu.org/>
  - 16) NIKKEI NET スポーツ総合.  
<http://sports.nikkei.co.jp/>
  - 17) YOMIURI ONLINE スポーツ.  
<http://www.yomiuri.co.jp/sports/>
  - 18) スラッシュドットジャパン. <http://slashdot.jp/>
  - 19) ITmedia News.

- <http://www.itmedia.co.jp/news/>
- 20) Oh-Yeah! すすきの情報リンク.  
<http://www.oh-yeah.jp/>
  - 21) 動宝. <http://www.douhou.com/>
  - 22) キッズ goo. <http://kids.goo.ne.jp/>
  - 23) Yahoo! きっず. <http://kids.yahoo.co.jp/>

(平成 19 年 6 月 12 日受付)

(平成 19 年 12 月 4 日採録)



上田 達巳 (学生会員)

平成 16 年北海道大学工学部システム工学科卒業. 平成 18 年同大学大学院情報科学研究科修士課程修了. 現在, 同大学院博士後期課程在学中. 主として, コンピュータネットワークに関する研究に従事. 電子情報通信学会学生会員.



高井 昌彰 (正会員)

昭和 58 年東北大学工学部電子工学科卒業. 昭和 63 年同大学大学院工学研究科博士課程修了. 工学博士. 同年東京大学理学部助手. 平成元年北海道大学工学部講師. 平成 4 年同助教授. 平成 7 年同大学大型計算機センター助教授. 平成 15 年同大学情報基盤センター教授. 平成 16 年同大学情報基盤センター副センター長. 平成 18 年同大学 CIO 補佐官. 現在に至る. 超並列・分散処理システム, コンピュータグラフィックス, コンピュータネットワークの研究に従事. 電子情報通信学会, IEEE, 国際 CIO 学会各会員.