コンピュータと教育 59－6
（２００１．３．２３）

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

西埜 覚*1　　西塚 一樹*2　　苗村憲司**
通信・放送機構 横浜コンテントリサーチセンター*1,*2
慶應義塾大学 大学院 政策・メディア研究科**

あらまし　WWW ページのコンテントをテキスト処理で情報の数値化を行い、更にリンク先ページの情報も組み合わせて用いて、学校教育における有害な情報を効率的に検出する研究開発を実行した。有害な情報の検出のために収集したページを中心にレイティングデータを作成し、それを適用させる PICS Rules 準拠のフィルタリングソフト（HCB-P フィルタ）を開発した。その HCB-P フィルタを実際に学校で試用して評価を行った。この一連の技術開発の状況と成果について報告する。

キーワード　レイティング、フィルタリング、有害な情報、PICS

A method of detecting harmful contents on the Internet and rating/filtering technology for schools

Satoru Nishino*1　Kazuki Nishiduka*2　Kenji Naemura**
Telecommunications Advancement Organization of Japan*1,*2
The Graduate School of Media and Governance, Keio University**

Abstract　A method for detecting harmful contents on WWW for school uses is developed. The method is constructed in two parts, one to analyze the text structure, the other to apply the information of forward-linked contents. A filtering program(HCB-P Filter) that conforms to PICS Rules is developed, and rating data made from the method mentioned above is applied to HCB-P Filter. The HCB-P Filter is evaluated through the use by pupils in schools. This paper reports the process and result of our study and development.

Key words　Rating, Filtering, Harmful content, PICS

1.はじめに

インターネットは、高度情報通信社会を支えるネットワークとして期待されており、世界の World Wide Web (WWW) のページ（以降、ページという）数は急激に増加し、利用者も増加している。インターネットには教育に役立つ有益な情報が多く公開されており、小中学校等でページの内容を授業等で積極的に活用するようになった。しかし、一方で違法又は有害な情報も数多く公開されており、学校では情報（コンテント）を選んで閲覧する必要がある。

違法又は有害な情報への対処策としては、①発信者に対する規律、②プロバイダ（通信事業者）に対する規律、③受信者による選択的受信（コンテントのフィルタリング）が考えられる。しかし、①、②は、1996 年米国通信品位法[1]が違憲判決を受けたように「表現の自由」等から慎重な検討が必要である。③の策が憲法の基本的人権に抵触することがない方策として有望視されており、郵政省の研究会報告[2]で、各地域の歴史的、文化的事情に合わせて、情報を取捨選択する方法が提言された。

コンテントのフィルタリングを行なうには、有害な情報を検出し、コンテントのレイティングを行なわなければならない。これらの作業は、ページを目視するなど人手に頼らざろうを得ないため、多大な労力が必要である。また、レイティングデータは、それを実施する組織の考え方に基づいて作成され、組織別に分散されて保存されている。その分散しているレイティングデータを連携させれば、データを有効に活用することが可能になる。

筆者らは、学校教育でインターネットを利用することを前提にして、

- ページの記述内容、ページのリンク関係を解析してコンテントを数値化し、4 カテゴリ（アダルト、暴力、差別、悪い情報）・5 レベル（0,1,2,3,4）の多段階レイティングで有害な情報の候補を効率的に検出すること
- その数値化の結果をもとにページを目視してレイティングデータを作成し、WWW コンソーシアム (W3C) の PICS(Platform for Internet Content Selection) 仕様[3]準拠のフィルタリング機能を実現すること
- そのフィルタリング機能の実験を行なって評価、課題を抽出すること
- また、他の組織のレイティングデータをフィルタリング機能に取込むこと

等の研究開発を行なってきた。

以下ではこれまで行なった研究開発の内容と成果について報告する。

2.レイティングとフィルタリング

レイティングとフィルタリングは組み合わされてページの閲覧を制御する。

①公開されるページは、コンテントの提供者等によってレイティングがなされる。

②受信者は、ページの閲覧を制御するためのレイティングレベルをフィルタリングソフトに指定する。

③受信したページのコンテントのレイティングレベルが、指定したレベルより低ければ閲覧でき、高ければ閲覧をブロックする。

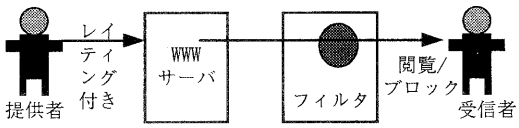


図1 レイティングとフィルタリング

PICS はレイティング、フィルタリングを実現するための標準仕様のひとつである。レイティングを行なった PICS ラベルの配布とフィルタリング機能実現のための形式、通信プロトコルが定められている。また、PICS ラベルをフィルタリング機能に適用させる方法や、ページの閲覧を受信者の考え方で制御するための PICS Rules も仕様として定めている。これらの PICS 仕様は、W3C のホームページで公開されており、これに基づいてレイティング、フィルタリングのシステムを開発することが可能である。PICS 仕様に基づくレイティング方式の一つである RSACi(Recreational Software Advisory Council on the Internet)[4]では、4 カテゴリ・5 レベルの多段階レイティングが可能である。マイクロソフト社(MS 社)の Internet Explorer は、PICS 仕様に準拠しており、このブラウザを利用すれば RSACi のラベルが記述されたページの閲覧を制御することが出来る。しかし、全ての

ページが RSACi を採用されている訳ではないため、受信側で有害な情報を閲覧ブロックする多種多様なフィルタリングソフトが利用されている。

3.有害な情報の検出

インターネットで公開されるコンテントを学校教育で利用することを前提に

・「有益な情報」とは、児童生徒に閲覧させることが授業目的の達成に役立つ情報
・「有害な情報」とは、児童生徒に閲覧させることが授業目的の達成を妨げる情報

と定義した。学校では有益な情報は積極的に閲覧させたいし、有害な情報は閲覧を事前に防ぎたい。

また、学校教育においては、情報の「有害」と「有益」は単純な対称的関係にあるのではなく

・「あるページの一部が有害」なら、そのページは有害と判断するべき
・しかし、「あるページの一部が有益」でも、必ずしも有益と判断すべきでない(例えば、交通安全の説明で車の制限速度の注意を書いた後で、スピード違反を逃れる方法が書いてあるページは有益ではない)

と考えられる。

これらの考え方をもとに、ページのコンテントを数値化し、ページのリンク関係を加えて、有害な情報の候補のページを検出する。それらページの数値等を利用者に提示し、目視によってレイティングを決定することで、人手による負荷の軽減を図ることを目標にしている。ただし、数値化の結果を自動的にレイティングデータとすることは目標にしていない。

(1) 個別ページのコンテントの数値化

ページに記述されたテキストを解析して、語句の抽出や語句の構成などについて以下の①から⑤を実行する。その結果を4カテゴリ別で重みを加算し、有害な情報として 5 レベルに数値化した。

①キー単語の出現頻度
- 有害な情報の特徴に関わる単語（キー単語）を、本文と HTML 構文の中（タグで囲まれた部分）の両方から抽出する。
- 本文、HTML 構文によって異なる重みを与える。
- 対象とする HTML 構文は、「Title タグ」、「META タグの Description、KeyWords」、「 <h1>、<h2>タグ」の 5 つ。タグによって与える重みは異なる。
- 重みは、キー単語に与えられた重みと出現回数によって加算して算出する。

②2単語の組み合わせの出現頻度
- 単独ではキー単語でなくとも、組み合わせることで有害な情報の示す場合がある。
- 文章の「。」、または HTML 構文の
、<p>などの「区切り」迄を一文章とみなし、この中に「指定の2単語が一定距離（バイト数）以内」にある時に単語の組み合わせの重みを与える。
- 例えば、[自殺、方法、距離＝4、重さ=x]の定義に対して、「・・自殺の方法には・・。」の文章は、距離=2 なので重み=x を与える
- 異なる単語の組み合わせについて全文章を調べ、加算して重みを算出する。

③文節（短い文章）の出現頻度
- 単語の組み合わせと同様に、一文章の中に「規定の文節と一致する部分」があれば重みを与える。
- 異なる文節について全文章を調べ、加算して重みを算出する。
- 例えば、[覚醒剤にはダイエット効果]（重み＝a)、[LSD の密輸ビジネス]（重み=b）が規定の文節なら重み=a+b を与える。

④自己ページ URL（Uniform Resource Locator）中の語句の出現頻度
- 自己の URL の中に「規定の語句」があれば重みを与える。
- 例えば、[bomb]（重み＝c)、[weapon]（重み＝d）が規定の語句なら、URL が「--.co.jp/bomb/weapon/--」の場合、加算して重み=c+d を与える。但し、同一語句の繰り返しは加算しない。

⑤リンク先ページ URL 中の語句の出現頻度
- ④と同様の重みの算出をリンク先ページの URL についても行う。

公開されている 1,218 ページについて、上述した数値化で得た結果と目視した結果を有害な情報の判定で比較した（図2）。

結果が一致したものが 1,028 ページ(84.4%) と大半を占めた。だが、目視の結果

が有害な情報であるにも関わらず甘い評価(有害な情報ではないの判定) になったものが 57 ページ (2.2%) あった。

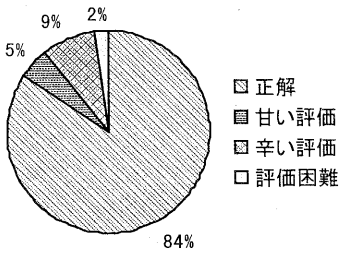


図 2 有害な情報の数値化と目視の結果

甘い評価になったものには、コンテントが画像のみや、テキストがほとんど書かれていないページが含まれていた。

(2) リンク先ページのコンテントの利用

WWW ページは、図 3 のようなリンクで作られており、一つ先のページを容易に閲覧できる構造 (ページ A から B、ページ B から C の閲覧) を持つ。

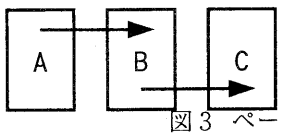


図 3 ページのリンク関係

ページ A と B は、以下のような関係が考えられる。

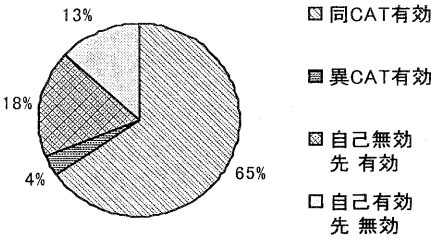
① ページ A の作成者は、ページ B のコンテントを理解している。

② ページ B の作成者は、ページ A のコンテントを知らない。

①の関係から「事前にページ B が有害な情報を持つと評価された場合は、ページ A の閲覧時にページ B のコンテントを利用して (以降、リンク先ページの情報先取りという)、ページ A は有害な情報の候補として扱う (例えば、ページ B が暴力的な情報なら、ページ A も暴力的な情報を持つケースが多い) ことが可能」の仮説が考えられる。

この仮説が正しいと確認出来れば、図 3 においてページ A の数値化が実行出来なくとも、事前にページ B が有害な情報と判断される場合には、ページ A も有害な情報を持つ候補とすることが出来る。このリンク先ページの情報先取りを用いれば、図 2 で示したページの「甘い評価」を改善することが可能な場合がある。

異なるホスト (ページ作成者が異なる場合が多い) にリンク先ページを持つ 3,854 ページで有害な情報の関連性を調べた結果を図 4 に示す。同じカテゴリで有害が 2,514 ページ (65.2%)、異なるカテゴリで有害が 141 ページ (3.7%)、リンク先ページで有害が 700 ページ (18.2%)、リンク先ページで有害でないが 499 ページ(12.9%)であった。高い割合でリンク先ページの情報先取りが可能である。



「CAT」は、カテゴリの略語として用いた
「CAT 有効」等で用いる“有効”とは、数値化の結果が有害な情報であること

図 4 リンク先ページのコンテントの関連性

(3) 有害な情報の候補の検出

上述したように、ページのテキストにコンテントを特徴づける記述が存在すれば、それを解析してコンテントの数値化を行なって有害な情報の候補を検出することが可能である。また、画像のみ (例えば、バナー広告のみのページ等) や、テキストがほとんど書かれていない FRAME タグや META タグの Refresh 等を用いた書き換えのみのが目的のページについても、リンク先ページの有害な情報を先取りして利用することが高い割合で可能である [5], [6]。

個別ページのコンテントの数値化とリンク先ページの情報先取りを組み合わせることで有害な情報を効率的に検出出来ることを確認した。

4.フィルタリング機能の開発

(1) レイティングデータの作成

有害な情報の候補になったページの URL と自己、リンク先ページの数値をユーザに提示して、ページを閲覧させて、目視で確認してレイティングレベルを設定するツールを作成した。この目視で決定したレベルが該当ページのレイティングデータとして保存される。

レイティングデータは、ページの URL とカテゴリ別のレベル、PICS Rules 用の情報で構成する。レイティングデータは、特定のフィルタリングソフトのみを対象にしておらず、簡単なツールで市販のソフトにも適用出来る構造を採っている。

(2) PICS Rules 準拠の HCB-P フィルタの開発

市販のフィルタリングソフトは多数存在するが、有料等の理由から広く普及しているとは言い難い。多くのユーザが利用しているパソコン用ブラウザの MS 社 Internet Explorer バージョン 5.0(IE5.0)は、PICS Rules に基づくフィルタリング機能が組み込まれている。上述のレイティングデータをこの機能で利用する HCB-P(Harmful Content Block-Prototype)フィルタを開発した。その仕組みを図 5 に示す。

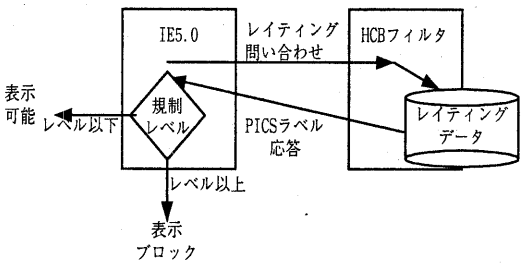


図 5 IE5.0 と HCB-P フィルタのインタフェイス

フィルタリング機能の実現の概略は以下の通りである。

①事前に、受信者は表示を制御する規制レベルを IE5.0 に設定する。

②受信者が閲覧を望んだページのレイティングレベルを、IE5.0 が HCB-P フィルタに問い合わせる。

③HCB-P フィルタはレイティングデータのデータベース(R-DB)を検索し、データがあれば“PICS ラベル”を応答する。データが無ければ、“ラベル無し”を応答する。

④IE5.0 は、HCB-P フィルタからの PICS ラベルと設定された規制レベルを比較し、PICS ラベルが規制レベル以下なら表示する。規制レベル以上なら表示をブロックする。

⑤HCB-P フィルタからの“ラベル無し”の応答に対しては、IE5.0 の設定によって「表示する/表示をブロック」のいずれかを選択することが出来る。

(3) レイティングデータの有効活用

HCB-P フィルタは R-DB に登録されたレイティングデータに基づいて PICS ラベルを応答する。サーチロボットを用いて有害な情報を中心にページを収集してレイティングデータを作成した。仮に与えたデータを含めて、約 8 万 5 千ページ分が登録されている。しかし、このページ数は小中学校の利用に限っても少ない数である。

PICS Rules では PICS ラベルの応答の実現方法は自由である。例えば、“「go.jp/」を URL に含むページは全てのカテゴリのレベルをゼロである”や、“「xxx.com/(xxx はある語句を与える)」のページはあるカテゴリのレベルが 3 である”といったことが可能である。PICS Rules の実現方法で、対象ページを拡大することが出来るが、ページ「xxx.com/」のレベルが固定化される等の課題が残る。他の組織が公開するレイティングデータを利用出来れば柔軟性と適用範囲が拡大するという利点がある。

(財)ニューメディア開発協会は SafetyOnline[7]の名称で独自のレイティング、フィルタリング機能を提供している。PICS Rules に基づく利用インタフェイスが公開されており、URL の問い合わせに PICS ラベル(5 カテゴリ・5 レベル)を応答する。HCB-P フィルタの R-DB に PICS ラベルが未登録なら、SafetyOnline から PICS ラベルを取得し、HCB-P フィルタの 4 カテゴリ・5 レベルに変換して活用するプログラムを開発した。これによって、フィルタリングの対象となるページを拡大することが出来た。

また、作成したレイティングデータを他の組織が利用することも可能である。

①PICS ラベル形式：PICS Rules に適用。
②URL リスト形式：規制レベルを指定して、該当するレベルを持つ URL のリストを生成。NOT リストとして、市販のフィルタリングソフトに適用。

5.小中学校での試用評価、実験

本研究開発には、研究フィールドとして 18 校の小中学校が参加した。これらの学校の先生、生徒の協力を得て、HCB-P フィルタの機能や、レイティングに関わる評価の確認を行なった。

この評価では、前述した約 8 万 5 千ページ分のレイティングデータを用いた。

(1) レイティングレベルの評価

暫定的に定めたレイティングデータのカテゴリ、レベルを表 1 に示す。RSACi 等を参考にして定めており、アダルトは RSACi のヌードとセックスを合わせたもの。暴力は RSACi と同一。差別と悪い情報は独自に設けたカテゴリである。

表 1 HCB-P フィルタのレイティングレベル

●アダルト：A	
0	なし
1	露出的な服装や情熱的なキス
2	部分的な露出や着衣での性的接触
3	全裸や性的接触の描写
4	煽情的な裸や性行為の描写

●暴力：V
なし
人や動物の傷害
殺人や動物の殺害
攻撃的な暴力や流血を伴う殺人
残忍で過激な暴力

●差別：D	
0	なし
1	穏やかな差別的な表現
2	差別的な表現
3	不快感を与える差別
4	排他的な差別

●悪い情報：B
なし
いかがわしい情報
不快な情報
破壊的な情報
有害な情報

この基準で目視を行ない、レイティングデータを作成した。しかし、カテゴリやレベルは、必ずしも学校教育での利用に基づく結果や先生の意見を反映しているわけではない。そこで、明らかに有害な情報であると判断出来るページを除き、学校の授業や課外活動でページ検索の方法によっては生徒が閲覧する可能性がある 6 ページ(ページの特徴を表 2 に示す)を選び、18 校の先生にページを閲覧してもらい、レイティングレベルと閲覧が許される年齢などについて、意見・感想を得た。その結果を表 3 に示す。表 3 のマスクされた部分が筆者によるレイティングレベルで、その下段が 18 校の先生によるレベルを平均した値である。

表 2 評価に用いたページ

URL	ページの特徴
URL#1	悪い情報へのリンク
URL#2	ドラッグの情報
URL#3	ドラッグ防止の情報（有益な情報）
URL#4	中国人の戦争被害に関する情報
URL#5	個別のビジネスへの意見 (悪いビジネスとして紹介)
URL#6	個人の日記風ページ

表 3 ページ別のレイティング評価

URL番号	レベル：下段が評価結果 A	V	D	B	閲覧許可の年齢 ES	JH	HS	AT	OT
URL#1	2	0	0	2					
	2.18	0.24	0.24	2.24	0	2	3	11	1
URL#2	0	0	0	2					
	0	0	0	1.94	1	5	3	8	1
URL#3	0	0	0	1					
	0	0	0	0.44	7	8	3	0	0
URL#4	0	1	1	0					
	0	0.67	0.61	0	5	9	1	3	0
URL#5	0	0	1	1					
	0	0	0.71	1.06	2	4	3	8	0
URL#6	0	1	1	1					
	0.39	0.83	0.72	1.06	3	3	4	7	0

年齢欄の ES は小学高学年、JH は中学生、HS は高校生、AT は成人、OT はその他を示す。

- レイティングレベルの平均値は、定めた値に近いが、大半のページで先生毎のバラ付きが大きい。
- URL#1 はリンク先ページの内容から成人が妥当。
- URL#3 は薬物の教育の実施状況で評価に差が出た。
- URL#4 は歴史的な事実とはいえ、記述内容の判断によって評価が分かれた。
- URL#5, #6 はページを見る立場の違いが評価の差になった。

この評価から、有害な情報の捉え方は利用する立場（学年や科目）や先生（個人）の考え方による差が大きいと考えられる。今後は、学校でのインターネット利用の目的、ページ閲覧の方針とレイティングレベルをリンクした検討、議論が求められるだろう。

(2) フィルタリング機能の評価

HCB-P フィルタの機能の確認を小学校と中学校で実施した。その構成を図6に示す。HCB-P フィルタは研究協力校と離れたセンターに設置されたサーバで稼動しており、学校とは INS64 で接続する。学校のパソコンからページを閲覧する時には必ず PICS ラベルの取得が行われる。その他の評価条件は以下の通りである。

①学校が利用するパソコンは14台。
②全カテゴリでレベルがゼロのページなら閲覧を許可。
③レイティングデータが未登録なら閲覧を許可。
④評価の時間は、小学校では授業の約45分間、中学校では課外活動の約90分間。
⑤生徒は、検索エンジンを利用して閲覧したいページを探す。

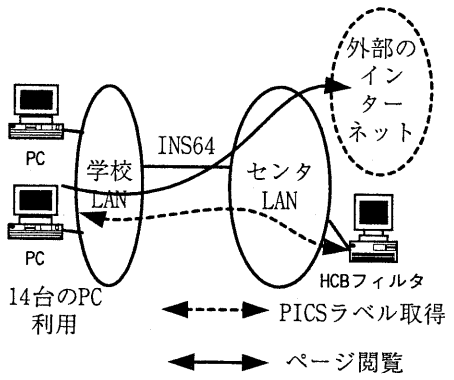


図6 フィルタリング機能評価の構成

評価時のアクセス件数とレイティングレベルがゼロ以外で表示がブロックされたページ数の結果を表4に示す。

表4 アクセス件数とブロック件数

	ページアクセス数	表示ブロック数
小学校	204	0
中学校	321	2

・授業などでの利用では、有害な情報へのアクセスはほとんど無い。
・中学校でブロックされた2ページは、戦争、兵器に関わるコンテントを含んでいた。
・評価したパソコンの台数、所要時間に比べてアクセス数は少ない。この理由は、①検索エンジンのアクセス数は除く、②検索エンジンの利用法に不慣れ、③印刷待ち（小学校）、④メールの送受信は除く（中学校で3台のパソコン）等による。
・PICS ラベルを INS64 で接続して取得することが利用上のアクセス低下にはならなかった。

このフィルタリング機能の評価で用いたレイティングデータは、有害な情報を検出する目的で収集したページをもとに作られている。小学校では表示ブロックの発生は無かった。中学校では、平和や戦争のページを閲覧対象にしたので、表示ブロックが2件あった。また、レイティングデータが登録されていないために有害な情報（戦争による死体）が2件表示された。この評価の結果から、先生が立ち会っている授業や課外活動などでは、生徒が有害な情報にアクセスする割合は少ないと考えられる。

また、アクセス件数はパソコンの操作の慣れ、検索エンジンの利用方法や閲覧したページを印刷、情報を保存する機器等の条件に影響を受ける。学校でのインターネット利用ではこれらの充実が求められると考えられる。

6.終わりに

インターネットで公開されているページのテキストを解析してコンテントを数値化し、さらにリンク先ページのコンテントも加えて、有害な情報の候補を効率的に検出する方式を開発した。また、有害な情報の候補のページを目視で確認してレイティングデータを作成し、PICS Rules に準拠する HCB-P フィルタで閲覧するページを制御すること、また、他の組織・団体が公開するレイティングデータを活用することを実現した。

さらに、研究フィールドとして参加した学校の協力によって、HCB-P フィルタを学校の授業や課外活動で利用し、今後の学校教育で

インターネットを利用するときの課題の一部が明らかになった。

学校教育でインターネットを利用する場合、授業教科にそった有益な情報にアクセス出来るよう、先生が事前にリンク集等を準備して活用する方法が中心になるであろう。一方、休み時間や放課後などで生徒が自由にインターネットを使用する場合、検索エンジン等で有害な情報の閲覧につながることも考えられる。これへの対応策が求められる。

また、欧米では、青少年が有害な情報を閲覧することに関して、家庭を第一の場と考え、更に公共の図書館等での閲覧についても議論が進んでいる。日本では、インターネットの利用について、家庭を前提にした議論が進まず、学校に閲覧するコンテントを任してことも多い。

青少年が学校や家庭でインターネットを利用する機会は増加し続ける。両方の場を前提にしたインターネットの利用方法、利用するコンテントの検討、議論が進められるだろう。

同時に、青少年の保護とインターネットの発展のために、表現の自由を保証しつつ、受信者が閲覧するコンテントを適切に選択する仕組みが強く求められる。このようなニーズに応え、従来のレイティング、フィルタリングシステムが抱える課題を解決するために、ICRA(Internet Content Rating Association)[4]によって新しい方式の国際標準化作業が進められている。既存のレイティングデータの相互利用や市販フィルタリングソフトへの適用に加えて、インターネットのグローバル性を考慮に入れたICRAの取り組みなどへの積極的な対応も求められる。

謝辞

HCB-Pフィルタの試用評価、実験は、横浜市教育委員会、市内の小中学校18校のご協力を得て実施した。ここに感謝の意を表します。

参考文献

[1]米国通信品位法(1996年) http://www.fcc.gov/telecom.html
[2]郵政省の「電気通信における利用環境整備に関する研究会」報告(平成8年12月) http://www.mpt.go.jp/policyreports/japanese/group/denki/61226y01.html
[3]PICS仕様 http://www.w3.org/PICS/
[4]RSACiレイティング、ICRA http://www.icra.org/
[5]西塚、他："WWWの有害情報検出に関わるある種のリンク先機能についての一考察"，情処学会2000年前期全国大会,5U-08(2000)
[6]西埜、他："WWWの有害な情報検出におけるリンク先情報の先取りの可能性の検証"，情処学会2000年前期全国大会,5U-09(2000)
[7] SafetyOnline http://www.nmda.or.jp/enc/rating/index.html