

研究指導 中澤真 准教授

# Web上の児童被害防止におけるフィルタリングのありかた

大賀 さとみ

## 1 はじめに

インターネットの普及に伴い、インターネットを利用する児童<sup>1</sup>も増加してきた。しかし、同時に児童誘引や個人の中傷、出会い系サイトなど、有害情報<sup>2</sup>に触れる機会が増えてしまい、インターネット上での児童の被害も増加を続けている。そこで、安全なサイトのみを閲覧でき、有害なサイトを遮断できるフィルタリングが近年注目されている。

現在のフィルタリングソフトは、サイトの特徴ごとにカテゴリ分けをし、児童に見せたいカテゴリと見せたくないカテゴリを保護者が選択できるタイプのものが多い。これらはURLフィルタリングとよばれる方式を採用しており、ソフトウェアベンダーが人手によって各サイトのURLをカテゴリごとに登録し、このリスト上のURLに該当するサイトをブロック<sup>3</sup>するというものである。しかし、人手による分類作業では、インターネット上の全サイトを適切なカテゴリに振り分けるのは難しく、未分類の扱いとなるサイトが数多く残ってしまうという問題がある。これに対応する一つの方法がキーワードフィルタリングであり、サイト内に含まれるキーワードをチェックすることによって、各サイトをフィルタリングしている。しかし、既存のキーワードフィルタリングには、安全なサイトを過剰ブロックしてしまう、カテゴリ別に設定を反映できないなどの問題がある。この原因として、有害なキーワードのみをリスト登録しており、カテゴリ別のリストを用いていないことや、1つのキーワードの含有の有無で判別していることが考えられる。

そこで本研究では、HTMLのdescriptionタグ内の文章から共起頻度の高いキーワードを抽出することにより、各カテゴリを特徴付けるキーワードを複数選び出すことで問題の解決を図る。そして、既存のキーワードフィルタリングとのブロック漏れ率と過剰ブロック率の比較・検証を行い、フィルタリング精度の観点から提案システムの有効性を示す。

## 2 情報インフラ整備による児童被害の現状とフィルタリングの必要性

### 2.1 児童を取り巻くネット環境とフィルタリング

情報インフラが整備されたことにより、自宅はもちろん学校やネットカフェなど、私たちはいつでもどこでもインターネットの利用が可能となった。これに伴い、近年では児童もインターネットを利用する時代へと変化した。現在では、小学生が家庭でインターネットを利用している割合は約7割にも及んでいる[1]。今やインターネットは、学習時の情報源として、また友人とのコ

ミュニケーションツールとしても、児童にとって不可欠なものになりつつある。しかし、インターネットの利用が児童に浸透したことで、違法・有害情報に児童が触れる機会も増えてしまった。さらに、掲示板やブログ、SNSなどのコミュニケーション色の強いサイトの普及により、これらのサイトを通しての児童誘引が多発するなどの新たな問題も生じている。

このような違法・有害情報から児童を守る対応策の一つにフィルタリングがある。フィルタリングとは、広義には大量の情報の中からユーザにとって必要な情報を取り出し、不要な情報を除外する処理を行う技術のことである。この技術が組み込まれたシステムを用いて児童に利用させるべきではない情報やサイトの取捨選択をさせることにより、児童が有害情報に触れてしまうのを防ぐことが可能となる。

以下では、インターネットにまつわる児童の被害状況から、フィルタリングの必要性について述べる。

### 2.2 インターネットに起因する児童の被害とフィルタリングの必要性

まず、警察庁による出会い系サイトに関係した事件等の被害状況と、総務省によるフィルタリングの普及率をまとめたグラフを図1に示す[2][3]。この図から明らかに、フィルタリングの普及率は、青少年ネット規制法[4]の「児童の携帯フィルタリング義務化」により近年伸びていることが読み取れる。それにもかかわらず性的犯罪などの児童の被害数は増加しており、現在のフィルタリングの普及率では、依然として児童被害の歯止めには至っていない。

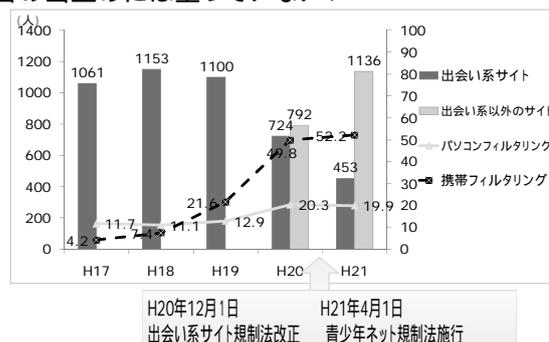


図1 インターネット上での児童被害数とフィルタリング普及率

一方、出会い系サイトと非出会い系サイト<sup>4</sup>に着目すると、出会い系サイト規正法改正[5]などの効果により出会い系サイトにおける児童の被害数は減少している。しかし非出会い系サイトによる被害は急激に増加しており、法の網の目をくぐった犯罪被害が増えていることがわかる。また、警察庁の調査により[6]、非出会い系サイトにおいて被害を受けた児童の99%がフィルタリ

<sup>1</sup> 18歳未満の者。

<sup>2</sup> 有害情報とは、ポルノ、出会い系、暴力的な画像、反社会的行為の助長、自殺の肯定などである。

<sup>3</sup> サイトを遮断し、閲覧できないようにすること。

<sup>4</sup> 主に、SNSなどのコミュニティサイトを指す。

ングを導入していないことが明らかとなった。これらのデータは、児童被害を減少させるためには法整備だけでなく、フィルタリングの導入率を高めることが必要不可欠であることを示している。

インターネットには児童誘引のような直接的な被害だけでなく、個人情報の公開や自殺の助長などの違法・有害情報を児童が目にしてしまう危険性もある。東京都教育庁の調査によると、学校非公式サイト<sup>5</sup>における書込みのうち、76%が不適切な書込みであった。その中には、飲酒や喫煙などの違法行為や、誹謗中傷など問題となる情報が多く見られた[7]。このような有害情報から児童を守るためにもフィルタリングは有効であり、この普及率を高める必要がある。

そこで、本研究ではフィルタリングの普及を妨げている技術的な要因を探り、フィルタリングシステムを改善することで普及率の向上を目指す。

### 3 既存のフィルタリングソフトの調査

現在、携帯電話とパソコンではフィルタリングの提供方法が異なっている。携帯電話では通信キャリアがそれぞれフィルタリングサービスを提供しており、利用者は申し込み手続きをするだけでフィルタリングサービスを利用できる。パソコンの場合はフィルタリングソフトをインストールする方法や、各プロバイダが提供しているフィルタリングサービスを利用する方法などがある。本研究では、フィルタリングの普及率向上を技術的な視点から取り組むために、携帯電話のフィルタリングと比較して普及率が極端に低いパソコンフィルタリングにターゲットを絞る。特に、無料で利用できるものも多数あり、プロバイダに左右されずに、それぞれの家庭のルールにカスタマイズできるという点を考慮し、パソコンインストール型のフィルタリングソフトを対象とする。

今回の調査ではフィルタリングの機能と精度について、5社のフィルタリングソフト<sup>6</sup>を実際に使用し、調査・検証を行った。調査したソフト名と、料金、制限カテゴリ数<sup>7</sup>については表1に示した。なお、以下本稿ではyahooあんしんねつを「あんしんねつ」、Windows ファミリーセーフティを「WFS」、Norton Online Familyを「NOF」と略す。

表1 フィルタリングソフトの料金と制限カテゴリ数

	あんしん ねつ	Net nanny	親子ネット	WFS	NOF
制限 カテゴリ数	6種類	31種類	17種類	13種類	49種類
料金	無料	3780円	月額300円	無料	無料

#### 3.1 カテゴリ数による問題

フィルタリングソフトでは保護者が児童に対する関

<sup>5</sup> 学校の公式サイトとは別に、同じ学校に通う生徒間での交流や情報交換を目的とする非公式なサイト。

<sup>6</sup> あんしんねつ, <http://anshin.yahoo.co.jp/>

Net nanny, <http://oyako119.jp/>

親子ネット, <http://www.oyakonet.com/>

WFS, <http://explore.live.com/windows-live-family-safety?>

NOF, <https://onlinefamily.norton.com/familysafety/loginStart.fs>

<sup>7</sup> アダルト、ゲームなどフィルタリングを適用できるカテゴリの数。

覧可否設定をカテゴリごとに行うように設計されているが、どのようなカテゴリにサイトを分類するか、またいくつかのカテゴリを設けるかは、ソフトウェアによって異なる。実際、表1に示した調査結果からも現在のフィルタリングソフトのカテゴリ数にはばらつきがあることが分かる。

フィルタリングソフトにおけるカテゴリ数の少なさは保護者の意図した通りの細かい設定ができず、利便性の低下を招いてしまう恐れがあるため問題となる可能性が高い。例えば、あんしんねつのカテゴリは不適切、掲示板、ショッピング、チャット、娯楽、メールの6種類と少ないため、表2に示した「mixi」などのSNSサイトや、10代の女性を中心とした情報交換サイトの「ふみコミュ」が全て出会い系サイトやアダルトサイトと同じ「不適切」カテゴリとして扱われてしまう。そのため、表2のサイトを児童に見せたい場合には、保護者は1サイトずつホワイトリスト<sup>8</sup>に登録するなどの手間がかかり、ユーザビリティ<sup>9</sup>の低下を招いてしまうことになる。

表2 yahoo あんしんねつにおける不適切カテゴリとして分類されるサイト例

不適切なサイト	URL
mixi	<a href="http://mixi.jp/">http://mixi.jp/</a>
GREE	<a href="http://gree.jp/?action=login">http://gree.jp/?action=login</a>
モバゲー	<a href="http://yahoo-mbga.jp/">http://yahoo-mbga.jp/</a>
ふみコミュ	<a href="http://www.fumi23.com/">http://www.fumi23.com/</a>
twitter	<a href="http://twitter.com/">http://twitter.com/</a>

以上述べたように、フィルタリングソフトでは保護者の規制プランに沿った制限ができるように、カテゴリ数が多いことが望ましい。加えて、各サイトが正しいカテゴリに振り分けられていることも求められる。次節では、このカテゴリ分けの精度によって生じる問題について論じる。

#### 3.2 カテゴリ分けの精度の問題

現在、ほとんどのフィルタリングソフトはURLフィルタリング方式を採用している。これはあらかじめ人手によって集めたサイトのURLをカテゴリ別に分類し、このカテゴリ単位でサイトの閲覧の可否を判別する方法である。手作業によるカテゴリ分けの場合、人間自身がサイト内容を把握した上で分類するため精度は高いというメリットがあるが、2011年時点で2億7千万[8]を超えるサイトのみならず、随時増え続ける新しいサイトまでも分類することは不可能である。そこで、NOFでは未登録のサイトを全て「未分類」というカテゴリに割り当て、これらのサイトをすべて閲覧不可と設定できる機能を設けている。しかし、このカテゴリには児童にとって無害なサイトと有害なサイトが混在している可能性がある。この場合、保護者が未分類カテゴリの閲覧を許可する設定にすると、有害なサイトも児童の目に触れる危険があり、逆にこのカテゴリの閲覧をブロックする設定にすると、安全なサイトまでも過剰ブロックされてしまうという問題が生じる。そこで、NOFにおける「未分類」カテゴリに、無害サイトと有害サイトがどの程度の割合で

<sup>8</sup> 児童に見せてもいいと設定したサイトのリスト。

<sup>9</sup> 利用者にとっての使いやすさ。

含まれているのかを明らかにするための調査を実施した。本調査では、小中学生を対象としたポータルサイト「Yahooきっず<sup>10</sup>」で「本」と検索した場合と、検索エンジン「Google<sup>11</sup>」において「出会い系サイト」と検索した場合の上位 20 件のサイトを調査対象とし、その内、未分類カテゴリに属しているサイトの数をカウントした。この結果を表 3 に示す。

**表 3 未分類サイトの割合に関する調査結果**

	未分類数/検索数
yahooきっず検索:本	7/20件
google検索:出会い系サイト	10/20件

「Yahooきっず」はあらかじめ登録された小中学生向けのサイトのみを検索結果として返すため、今回の検索結果のサイトもすべて健全なものであった。一方、「Google」で出会い系サイトと検索した場合の上位 20 件のサイトは、そのすべてが児童にとって有害なものであった。調査の結果、どちらも約半数が未分類カテゴリに属していた。このため、未分類の閲覧許可設定を可としても不可としても、半数のサイトが保護者の意図に反した形で処理されてしまうことになり、結果として児童の利便性が低下する原因となってしまう。

### 3.3 既存のキーワードフィルタリングの問題

前節で述べた問題点を解決する一つの手段がキーワードフィルタリング方式である。キーワードフィルタリングとは、あらかじめ指定してある有害なキーワードと、サイト内の文を照らし合わせ、キーワードを含むページをブロックするというものである。キーワードフィルタリングは自動分類によって表示するサイトを随時判断しているため、新しく作成されたサイトに対してもフィルタリングを行うことができ、未分類サイトにも対応できるが、過剰ブロックやブロック漏れとなる可能性がある。

そこで、この問題を明らかにするために、キーワードフィルタリング機能を実装している「あんしんねっと」を用いて検証した。今回は、検索エンジン「Google」において「ゲーム」というキーワードで検索した際の、上位 100 件中、内容的に「ゲーム」カテゴリに分類される 57 件のサイトに対するブロック率を検証した。その結果、URL フィルタリングでは 9 件のサイトをブロックすることができなかった。そこでキーワードフィルタリング機能を有効にした時、このブロック漏れの問題を解消できるか検証したが、キーワードリストに「ゲーム」が登録されておらず 9 件全てブロック漏れとなってしまった。

一方、このキーワードフィルタリングには過剰ブロックについての課題もある。例えば、「あんしんねっと」のキーワードフィルタリングを有効にした状態では、検索サイトにおいて「出会い」というキーワードで検索すると、出会い系サイトをブロックするために検索結果自体が表示されないような強い規制が働く。しかし、「出会い」というキーワードを含む出会い系サイトでない安全なサイトでさえも表示されず、過剰ブロックになってしま

う。

そこで次節では、これまで述べたキーワードフィルタリングの問題を解決するための新しいフィルタリング手法について論じる。

## 4 共起関係がある 2 語を利用したキーワードフィルタリングの提案

### 4.1 description タグ内の文を利用したカテゴリ分け

まず、カテゴリごとのフィルタリングを可能にするために、それぞれのカテゴリ特有のキーワードを抽出する。この方法として、それぞれのカテゴリに属するすべてのサイト内の文章から頻度の高いキーワードを選び出すことにした。このとき考慮することは、キーワードをサイト内のどの部分から抽出するかということである。HTMLソース全体に対して頻出キーワードを抽出した場合、他サイトへのリンクなどコンテンツ本来の内容から外れたキーワードも多く含まれるため、そのサイトが属するカテゴリの概念となるキーワードを抽出できない可能性がある。そこで、今回着目したのはdescription タグである。このdescription タグとは、そのページの概要を記述する部分であり、検索エンジンサイトにおけるスニペット<sup>12</sup>にも使用されることが多いものである。そのため、サイト本来の内容だけが記述されており、サイトが属するカテゴリのキーワードを含んでいる可能性が高い。さらに、頻度が高いキーワードを使用することで、その可能性はより高まると考え、description タグの文章を用いたキーワード抽出が適切であると考えた。

次に、抽出するキーワード数について検討する。前節で示した例のように、「出会い」といった単一のキーワードでフィルタリングを行うと、カテゴリを正しく識別できず、過剰ブロックを誘発する可能性がある。これを防ぐためには、複数の単語によってフィルタリングを行う必要がある。本研究では、この複数の単語を選び出すため、キーワードの共起関係に着目した。共起とは文章において、ある文字列とある文字列が同時に出現することである。つまり、共起頻度が高いということは、その単語同士の関連性が強いことを意味する。そのため、単一キーワードではできなかったカテゴリの識別を、同時に出現したもう一方のキーワードによって行う事が可能であると判断した。

キーワード抽出の具体的な方法について実例を挙げながら説明する。ここでは表 4 に示した 4 つのカテゴリに対するキーワードを抽出するために、各カテゴリに属する全てのサイトに対して、共起頻度の高いキーワード上位 5 件を抽出した。表 4 の結果から、単独ではカテゴリの特性を必ずしも表してはいないが、2 語となることで所属カテゴリとの関連性があるキーワードを抽出できたことが分かる。

本システムでは、保護者がブロック設定をしたカテゴリに対し、先に述べた方法で抽出したキーワードを用いてキーワードフィルタリングを実行する。具体的には、共起頻度の高い上位 5 件のいずれかのキーワー

<sup>10</sup> <http://kids.yahoo.co.jp/>

<sup>11</sup> <http://www.google.co.jp/>

<sup>12</sup> 検索エンジンで検索をした際に、検索結果のタイトルの下に記述されたページの説明文のこと。

ドの組み合わせが、チェック対象のサイトの description タグの文章に使用されていた場合、そのサイトをブロックする。なお、上位いずれかのキーワードによって判断する理由としては、1位の共起キーワードは含まれていないが、2位以下のキーワードは含まれているサイトが存在するからである。また、キーワードの照合をサイトの description タグ内に限定した理由は、キーワード抽出時に description タグ内に限定した理由と同じである。次節では、既存のキーワードフィルタリングとの精度の比較を行った結果について述べる。

**表4 各カテゴリに頻繁に出現する共起の組合せ<sup>13</sup>**  
(括弧内は40サイト中共起キーワードが登場したサイト数)

順位	カテゴリ名	出会い系	アダルト	ショッピング	ゲーム
1		出会い系(32)	アダルト(24)	通販・ショッピング(17)	ゲーム・無料(23)
2		出会い 無料(17)	アダルト・動画(15)	通販・サイト(16)	ゲーム・オンライン(12)
3		サイト・無料(15)	動画・サイト(14)	サイト・ショッピング(15)	無料・オンライン(10)
4		出会い 紹介(12)	アダルト・比較(10)	通販・商品(13)	ゲーム・遊べる(8)
5		サイト 紹介(10)		商品・ショッピング(12)	ゲーム・サイト(8)

#### 4.2 あんしんねっとのキーワードフィルタリングとの比較に基づいた検証

提案したフィルタリング方式の有効性を示すために、本システムと、「あんしんねっと」のキーワードフィルタリングの精度を比較・検証した。有効性の検証方法として、検索エンジン Google において「出会い」、「アダルト」、「ゲーム」、「ショッピング」と検索した際のそれぞれ上位100件中「あんしんねっと」のURLフィルタリングでブロック漏れした有害サイトに対するのブロック確率、また、同じくURLフィルタリングで、無害なサイトと判断されたサイトに対するの過剰ブロック確率を検証した。そこで、それらのサイトに対するのブロック漏れ率と過剰ブロック率を検証した。

**表5 有害なサイトに対するブロック確率**

検索ワード	あんしんねっと	本研究
出会い	53.3%	80.0%
アダルト	50.0%	75.0%
ゲーム	0.0%	57.1%
ショッピング	0.0%	62.5%

**表6 無害なサイトに対する過剰ブロック率**

カテゴリ名	あんしんねっと	本研究
出会い	100.0%	33.3%
アダルト	100.0%	14.0%
ゲーム	0.0%	20.0%
ショッピング	0.0%	14.7%

表5、表6に示したように、本研究のシステムでは既存のキーワードフィルタリングよりも、すべてのカテゴリにおいて高いブロック率となり、カテゴリごとのキーワードフィルタリングが有効に機能することが明らかとな

った。また、表6の過剰ブロック率では、「あんしんねっと」において「アダルト」と「出会い」はキーワードフィルタリングが適用され、検索結果自体が表示されず過剰ブロックとなってしまったが、本研究のシステムでは過剰ブロック率をかなり抑えることに成功した。

#### 5 むすび

本研究では、既存システムで問題となっていた、カテゴリ別に設定を反映できない、過剰ブロックが発生するというキーワードフィルタリングの問題を改善する手法を提案した。この手法ではHTMLのdescriptionタグ内の文章から共起頻度の高いキーワードを抽出することにより、各カテゴリを特徴付けるキーワードを複数抽出することを可能とした。さらに、既存システムで生じた1キーワードの含有の有無のみの判断による過剰ブロックを防ぐために、共起頻度の高いキーワードを用いることにした。その結果、既存システムよりも精度の高いフィルタリングを実現することができた。

しかし、本システムには、僅かだが過剰ブロックの問題が残っている。その原因として、全カテゴリにおいて共起頻度が高いキーワードが抽出されてしまうケースがあるかである。今後、これらのキーワードの処理方法を検討することにより、より精度の高いフィルタリングが可能になるであろう。

#### 参考文献

- [1] アスキー総合研究所, 小学生のPCでのインターネット利用経験, <http://asciimw.jp/info/release/pdf/20100614.pdf>
- [2] 警察庁, 平成22年上半年期の出会い系サイトに関係した事件等の検挙状況について, サイバー犯罪対策, <http://www.npa.go.jp/cyber/statics/h22/pdf02-1.pdf>
- [3] 総務省, 平成21年通信利用動向調査, 情報通信統計データベース, [http://www.soumu.go.jp/johotsusintokei/statistics/data/100427\\_1.pdf](http://www.soumu.go.jp/johotsusintokei/statistics/data/100427_1.pdf)
- [4] 衆議院, 青少年が安全に安心してインターネットを利用できる環境の整備等に関する法律, 制定法律第79号, [http://www.shugiin.go.jp/itdb\\_gian.nsf/html/gian/honbun/houan/g16901030.htm](http://www.shugiin.go.jp/itdb_gian.nsf/html/gian/honbun/houan/g16901030.htm)
- [5] 衆議院, インターネット異性紹介事業を利用して児童を誘引する行為の規制等に関する法律の一部を改正する法律, 制定法律第52号, <http://www.shugiin.go.jp/index.nsf/html/index.housei.htm>
- [6] 警察庁, 非出会い系サイトに起因する児童被害の事犯に係る調査分析について, サイバー犯罪対策, <http://www.npa.go.jp/cyber/statics/h22/H22deai-bunseki.pdf>
- [7] 教育庁, 学校非公式サイト等の監視結果について(平成22年10月), 東京都教育会 [http://www.kyoiku.metro.tokyo.jp/pickup/seisaku/kanshi\\_1010.pdf](http://www.kyoiku.metro.tokyo.jp/pickup/seisaku/kanshi_1010.pdf)
- [8] Net craft, January2011WebServerSurvey, <http://news.netcraft.com/archives/2011/01/12/january-2011-web-server-survey-4.html>
- [9] 松元英樹, “フィルタリングソフトで有害サイトを遮断,” 日経パソコン, no.492, pp.30-33, 2005/10.
- [10] 洪井哲也, “青少年ネット規制法とフィルタリングのあり方-“有害情報”とフィルタリング-,” 電子情報通信学会技術研究報告, vol.108, no.168, pp.13-16, 2008.
- [11] 井出明, “社会システムとしてのフィルタリング問題,” 情報の科学と技術, 56巻, 10号, pp.446-451, 2006.
- [12] 新保史生, “フィルタリングと法,” 情報の科学と技術, 56巻, 10号, pp.475-481, 2006.

<sup>13</sup> アダルトの場合5位以下は登場サイト数が並んでいたため、4位までの抽出とした。