【報道発表資料】

株式会社 KDDI 研究所
2011 年 10 月 25 日

インターネット上の有害情報を高精度に自動識別するシステムを開発

有害情報の「見た目」と「内容」を分析、識別精度 95%を達成

株式会社 KDDI 研究所（本社：埼玉県ふじみ野市、代表取締役所長：中島 康之）は、インターネット上にある Web ページや掲示板などの書き込みに含まれる有害情報を柔軟かつ高精度に自動識別する『有害コンテンツ高度識別システム』を開発しました。本システムは、①有害情報の「見た目」と「内容」の両方を分析して高精度に識別できる高い識別精度、および、②適用先ごとの有害判定基準の違いに配慮し、有害としたい情報を 72 カテゴリから柔軟に設定できる高い柔軟性を特長とします。

なお今回の技術は、独立行政法人情報通信研究機構からの委託研究である「インターネット上の違法・有害情報検出技術の研究開発」の研究成果です。

【背景】

インターネット上の Web ページ数の増大に伴い、出会いや犯罪予告などを目的とした有害ページも急増しており、現在、ブラックリスト※1 方式や、コンテンツフィルタリング方式※2 を代表とした有害情報フィルタが注目されています。特に単語に基づき有害情報を識別するコンテンツフィルタリング方式は、新たな URL であっても識別可能であることが利点であり、近年利用が増加していますが、単語を追加し過ぎると識別誤りも増えるため、必要なレベルまで識別精度が高くならないという課題がありました。また、有害情報フィルタは青少年向けだけでなく、企業の Web 閲覧制限用途にも利用されていますが、この場合、娯楽関連カテゴリも制限対象となる場合もあり、このような判定基準の柔軟な設定が必要とされていました。

【今回の成果】

このたび、こうした課題を解決するために、有害情報の柔軟かつ高精度な自動識別機能を簡単に利用できる『有害コンテンツ高度識別システム』を開発しました。利用者はまず、判定システムを利用する前に、「誹謗中傷」や「アダルト」「オンラインゲーム」「掲示板」などの 72 カテゴリの中から有害としたいカテゴリを自由に組み合わせることで、フィルタリング対象を簡単に設定できます。識別精度については、コンテンツに含まれる有害な単語検出やその係り受け※3 関係抽出などの精密な「言語的コンテンツ分析」と、判定済みコンテンツとの内容の類似判定や、有害の可能性が高い背景色やリンク先の検出などの高速な「外形的コンテンツ分析」を組み合わせた総合判定により、高精度化を両立することに成功しました。識別精度※4 はおよそ 95%を実現しています。また、本システムは Web インタフェースを搭載しており、Web ブラウザ経由で簡単に利用することができます。

【今後の展望】

現在、数社と協力して、開発したシステムの実証実験を実施しており、2012 年度のサービス開始を予定しています。また本システムを 2011 年 10 月 26～28 日に幕張メッセ（千葉市美浜区）で開催される第 1 回情報セキュリティ EXPO【秋】の KDDI 研究所ブース内に出展致します。

※実証実験に参加して頂ける事業者は随時募集中です。(2011 年 12 月末終了予定)

以上

本件に関するお問合せ先
株式会社ＫＤＤＩ研究所 営業企画グループ
TEL:049-278-7545　E-mail：inquiry@kddilabs.jp

【別紙】

用語説明

※1　ブラックリスト方式

有害ページの URL を記載したリストをあらかじめ作成しておき、利用者が Web ページにアクセスしようとすると、その Web ページの URL がブラックリストに記載された URL と一致した場合に、有害と判定するフィルタリング方式。

※2　コンテンツフィルタリング方式

Web ページや掲示板の書き込みなどの文章を分析し、有害かどうかを判定する方式。あらかじめ有害と分かっているデータを用いて、有害サイトに特徴的に出現する文章や単語を抽出しておき、それに基づいて有害かどうかを判定する。

※3　係り受け

名詞 A を修飾する形容詞 B、などの、文法的な単語間の関係。

※4　識別精度

システムが有害と判定したコンテンツのうち、目視確認の結果、実際に有害であった割合。

システムの使い方

①システムにログイン

②判定対象データのアップロード

③判定結果ファイルのダウンロード

想定出力ファイル(1000件).tsv

	A	B	C
1	ファイル名	有害/無害	
2	U1 N1 6977320,28	0	
3	U1 N1 9534290,89	0	
4	U1 N1 8495090,89	0	
5	U1 N1 7200270,89	1	
6	U1 N23995930,89	0	
7	U1 N1 7073840,89	0	
8	U1 N20845880,89	0	
9	U1 N1 8573180,89	1	
10	U1 N1 5764780,89	0	
11	U1 N24063350,89	0	
12	U1 N27354250,89	1	

④結果確認

システムの機能構成

入力：HTML(Web ページ、ブログサイト)、またはテキスト(記事、コメント)
出力：総合判定結果

高精度・高速化：見た目判定＋言葉判定
柔軟なポリシー設定：判定基準
(効率化：URL フィルタ(判定対象をあらかじめ絞り込む))