

平成22年度「インターネット上の違法・有害情報検出技術の研究開発」の開発成果について

1. 施策の目標

ISPやコンテンツ監視事業者が共通で利用可能な「監視業務支援システム」を研究開発し、監視業務のコストを削減すること

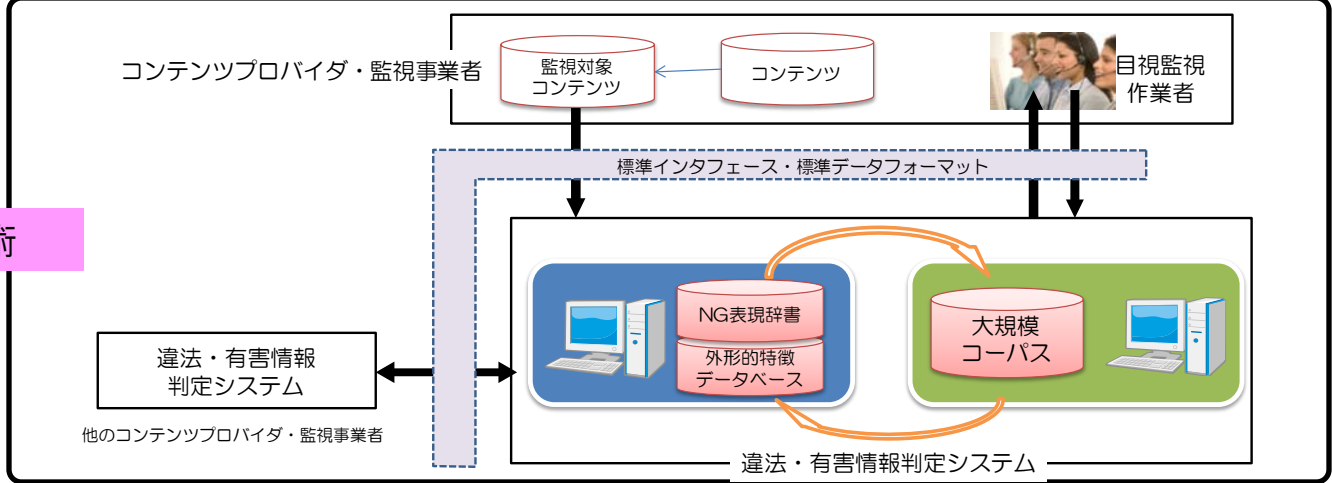
2. 研究開発の背景

インターネットが国民生活に必要不可欠となっている一方で、違法・有害情報による犯罪の助長や自殺の誘因が社会的に大きな影響を与えており、近年では、ISPやコンテンツ監視事業者などによる監視作業が進められている。しかしながら、今後も爆発的に情報が増大すること、およびその内容が多岐に渡ることを勘案すると、それらの監視対象に対して現在実用化されている技術だけでは限界が生じることは推測に難くない。そのため、システムが監視業務を支援する割合をこれまで以上に増やし、人手による業務を軽減するための新たな技術が必要である。

3. 研究開発の概要と期待される効果

以下の3つの要件を満たす監視業務を支援する違法・有害情報検出技術を研究開発する。これにより、コンテンツ事業者や監視事業者が低コストで違法・有害監視作業を実施できるようになり、ISPの自治の促進や安価なASP型違法・有害判定サービスの実現が可能となる。

- ・各事業者が共通で利用できること(共通ポリシー)
- ・各事業者の監視ポリシーを反映できること(事業者個別ポリシー)
- ・監視コストを大幅に下げることができること(コスト:人件費、設備費、維持費など)

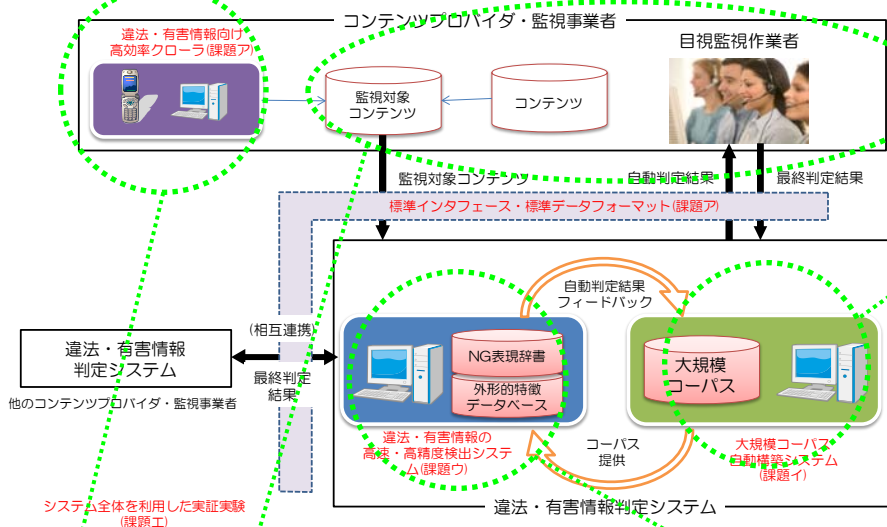


4. 研究開発の期間及び体制

平成21年度～平成23年度(3年間)
NICT委託研究(株式会社KDDI研究所)

違法・有害情報検出技術の主な成果

違法・有害情報検出技術



- 課題ア: モバイル対応型情報収集・解析結果提供技術
- 課題イ: 違法・有害情報の大規模コーパス自動構築技術
- 課題ウ: 違法・有害情報の高速高精度検出技術
- 課題エ: ISPやコンテンツ監視事業者等の実運用環境における実証実験

課題イ: 違法・有害情報の大規模コーパス自動構築技術

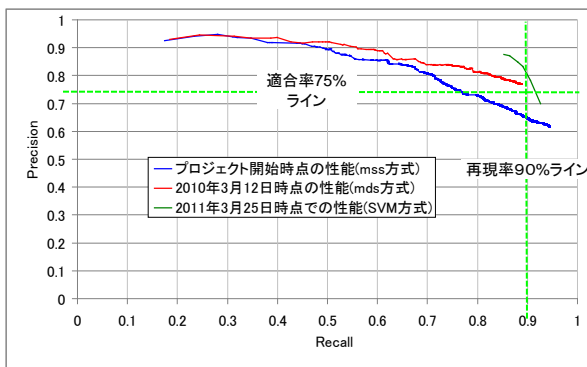
[目的] 違法・有害情報の高精度な監視ツール構築のベースとなる3億記事の大規模コーパスを構築する技術に関する研究

- [成果] 4種類のコーパスを構築し、H22年度目標を達成
- ・カテゴリラベル付与コーパス2665万記事(目標: 2000万記事)
 - ・品詞ラベル付与コーパス50万記事(目標: 50万記事)
 - ・品詞・格構造ラベル付与コーパス10万記事(目標: 10万記事)
 - ・有害・無害ラベル付与コーパス1.3億記事(目標: 1億記事)

課題ウ: 違法・有害情報の高速高精度検出技術

[目的] 違法・有害情報の監視業務を効率化するための、違法・有害情報の頑健かつ高速高精度な自動検出技術に関する研究

[成果] Webサイトの背景色や構成など有害サイトに特有な外形的特徴に基づいて有害サイトを高速かつ高精度に検出する技術を開発。また、高速SVMを用いた大規模データによる学習の結果、再現率を3%以上向上。さらに、年度目標である再現率90%、適合率75%を達成



有害サイト高速に検出
毎日新聞、読売新聞、東京新聞、NHK「おはよう日本」にて紹介

課題ア: モバイル対応型情報収集・解析結果提供技術

[目的] Webサイトの監視作業効率化のための、携帯専用サイト識別技術、有害サイトクローラ技術に関する研究

[成果] 携帯専用サイト識別技術の評価実験を行い、91.4%の精度を達成。また、有害サイトクローラの評価実験を行い、一般のクローラと比較して3.5倍の有害サイト収集を達成

課題エ: ISPやコンテンツ監視事業者等の実運用環境における実証実験

[目的] 本開発の実証実験の実施に向けた環境構築

[成果] H23年度の実証実験に用いる判定エンジンを構築。また、H23年度の実証実験に参加する事業者・CPの募集を開始

1. これまで得られた研究成果(特許出願や論文発表等)

H21年9月～H23年3月

	国内出願	外国出願	研究論文	その他研究発表	報道発表	展示会	標準化提案
インターネット上の違法・有害情報検出技術の研究開発	16	0	5	14	7	6	0

2. 特筆すべき成果

- ・課題ウ: 外形的特徴に基づく有害サイト検出手法について報道発表により研究成果を幅広くアピール
 - ①一般誌4誌に掲載
読売新聞(9月14日、8面)、毎日新聞(9月14日、27面)、東京新聞(9月14日、8面)、フジサンケイビジネスアイ(9月16日、7面)
 - ②NHK番組「おはよう日本」で紹介(10月17日)
- ・課題ア: 有害サイトクローラ技術について国際会議発表(10月18日)および報道発表(2月1日)を行い、Webニュースを中心に50サイトに掲載されるなど、研究成果を幅広くアピール

3. 表彰

- ・船井ベストペーパー賞を受賞(2010年9月8日)
外形的特徴に基づく有害サイト検出手法について、2010年9月に開催された情報科学技術フォーラム(FIT2010)に投稿した論文が受賞。(700件以上の論文のうちわずか3件のみ)
- ・大会奨励賞を受賞(2011年3月3日)
係り受け解析を用いた高精度有害サイト検出手法について、情報処理学会全国大会で発表した研究が受賞。