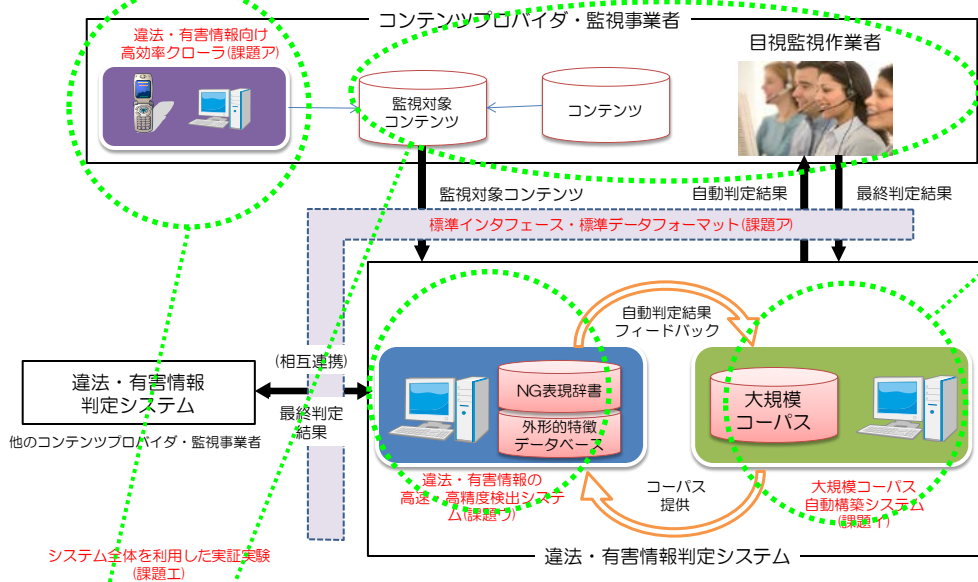


インターネット上の違法・有害情報検出技術の主な成果

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

課題ア: モバイル対応型情報収集・解析結果提供技術

○違法・有害情報に関するPCサイト・携帯サイトのクローリング作業の効率化技術に関する研究

- 携帯専用サイトを自動識別する機能(精度91.2%)を開発
- 有害Webページを効率的に(適合率90%、再現率70%)収集可能な有害クローラを開発

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

○実証実験の実施と商用化に向けた施策の実施

- 成果の積極的な広報活動(展示会8回・プレスリリース4回)を実施。その結果、実験参加5社(問い合わせは10社以上)、取材6回・報道10回を獲得。
- CP2社、Webサイト管理1社、監視事業者1社の合計4社と実証実験を実施。うちCP2社と商用利用に向けた検討を開始。

課題イ: 違法・有害情報の大規模コーパス自動構築技術

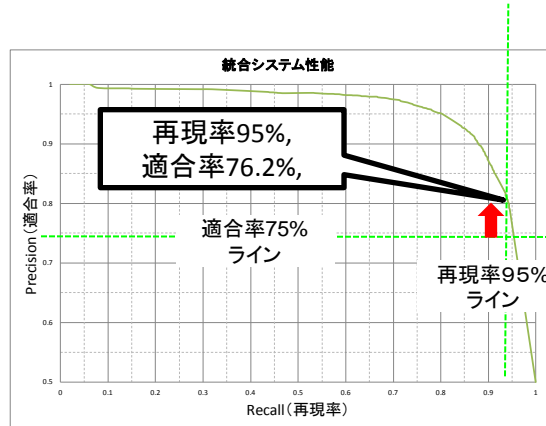
○違法・有害情報の高精度な監視ツール構築のベースとなる、3億記事の大規模コーパスを構築する技術に関する研究

- 機械学習向けの人手によるラベル付きの3種類のコーパスを構築
 - ・カテゴリラベル付与コーパス3220万記事(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)
- 機械学習システムの段階的適用によりコーパスを自動構築するシステムを開発し、これを用いて3億記事のコーパスを構築(目標:3億記事)
- 上記コーパスによるネット言語対応の形態素解析辞書を構築

課題ウ: 違法・有害情報の高速高精度検出技術

○違法・有害情報の監視業務を効率化するため、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

- 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
- 伏せ字やギャル文字といったネット特有の言語を高精度に解析するための手法を世界で初めて開発
- 最終目標判定精度である再現率95%以上、適合率75%以上、処理時間1.2秒以内、をすべて達成



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」

ギャル文字自動翻訳
【日経産業新聞第一面
2009年12月3日】他、
NHK「おはよう日本」、
テレビ東京「WBS(トレ
たまコーナー)」

4. これまで得られた成果(特許出願や論文発表等) ※成果数は累計件数と()内の当該年度件数です。

国内出願	外国出願	研究論文	その他研究発表	プレスリリース	展示会	標準化提案	取材・報道
21 (5)	0	7 (1)	14 (1)	4 (1)	6 (2)	0	12 (2)

5. 成果の対外アピール、実証実験の成果について

(1) 外部への積極的な成果アピールを実施し、多大な反響を得た

・成果の積極的な広報活動(展示会6回・プレスリリース4回)を実施し、取材・報道12回を獲得した。
TV報道: NHK「おはよう日本」に2回、テレビ東京「WBS(トレたまコーナー)」
新聞掲載: 日経産業新聞(第一面)、毎日新聞、読売新聞、東京新聞、その他Webニュース多数掲載。
その他問い合わせ: 警察庁(有害クローラ)、東京都庁(取り組み全般)、など、十数社からの問い合わせあり。

(2) 4社と連携して実証実験を実施、うち1社と商用利用に向けて進展

・監視事業者1社、カード決済サービス提供企業1社、地図サービス運用企業1社、スマートフォンアプリ提供サイト管理企業1社、の合計4社と実証実験を実施した。
・このうち地図サービス運用企業、およびカード決済サービス提供企業と、システム利用に向けた検討を継続して進めている。
・商用利用のモデルケースとして、協力会社と連携して商用システムの運用計画(環境構築方法、運用形態、運用費の算出など)を進めている。

6. 今後の研究開発計画

この成果により、今後、どのような研究を行うのかを例示を上げながら、具体的、かつ簡潔に記載して下さい。

(1) 商用展開・普及活動

・商用時の運用形態の検討、他社連携も含めた運用体制の検討など、商用展開に向けた施策を進める。
・商用展開について現在進捗中の1社に加えて、実証実験参加社およびこれまでに問い合わせを受けた企業を中心に、商用展開を進める。
・商用展開可能なソフト(ドキュメントの整備を含む)の完成後、プレスリリースにより広報活動を進める。

(2) 研究活動

・くだけた表現の自動補正技術に関し、今回対象とした表現(伏字、組み文字など)の範囲を広げるための研究開発を新たな枠組みで進める必要がある。
・判定性能の維持・向上のため、新たに出現した違法・有害情報の検出方式の研究開発を新たな枠組みで進める必要がある。