1ZB-05

# 児童被害防止のための SNS の有害コメント収集プラットフォームの開発

千葉　翔也†　　角田　裕†

†東北工業大学工学部情報通信工学科

## 1. はじめに

SNS における未成年の犯罪被害への対策が必要となっている．その一環として，援助交際を誘引するコメントなどの児童に有害なコメント（以下，有害コメント）の削除を目的としたボランティアによるサイバーパトロール活動が実施されている［1］．しかし，有害コメントの数の多さに，通報が追い付いておらず，活動を効率化する仕組みが必要である．これまでに，機械学習による有害コメントの抽出［2］や個々人の活動の支援用アプリの開発［3］などの試みはあるが，グループでの協調活動を前提とした効率化の試みは進んでいない．本研究では，有害コメントの収集と複数の参加者による情報共有のプラットフォームを開発し，収集を効率化する仕組みを構築した．

## 2. コメント収集プラットフォームの提案

筆者らの研究室は宮城県警のサイバーパトロール活動に参画し，Twitter 上の有害コメントを Twitter 社の通報窓口に通報している．活動参加者は以下の手順で通報を行う．

**(C-1)** コメントをキーワード検索
**(C-2)** 発見したコメントが有害コメントかどうか目視で確認
**(C-3)** 有害コメントの URL を CSV ファイルにコピー&ペーストし通報理由とともに蓄積
**(C-4)** CSV ファイルをもとに機械的に通報

上述の手順では，((C-1))コメントの検索と((C-3)) URL のファイルへの集約をすべて手動で行っているため，検索やウィンドウの切り替え操作などに時間を取られている．加えて，複数の参加者が独立して検索・通報を行っているため，(A)同一語句による重複検索，(B)同一有害コメントの重複通報という非効率性がある．

A development of the platform for collecting harmful SNS comments to protect children
Shoya Chiba†, Hiroshi Tsunoda†
†Tohoku Institute of Technology

そこで，以下の機能を有する有害コメントの収集と判定のためのプラットフォーム（図１）を構築した．

- 単語・ハッシュタグでコメントを自動検索
- 検索で得られたコメントの参加者間で共有
- 各コメントが有害か否かの記録の保持

プラットフォームの実現には以下のサービスを用いた．

- Queryfeed［4］
- Slack［5］とその RSS アプリ［6］

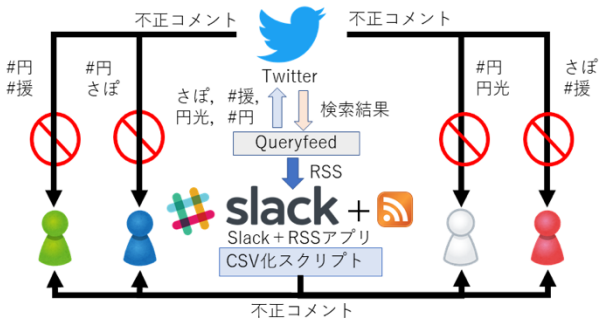


**図 1　プラットフォームの概念図**

本プラットフォームは，図１のように各参加者の代わりに，特に数の多い援助交際誘引目的の対象に，複数のハッシュタグや単語によりコメントを一括検索し，その結果をチャットシステムを介して参加者間で共有できるようにする．参加者は各コメントを閲覧し有害コメントと判断したものの URL だけを，スクリプトによってチャットから CSV ファイルとしてダウンロードし通報する．

Queryfeed は，SNS などを定期的に単語などで検索した結果の RSS フィードを生成するサービスであり，自動検索の実現のためにこれを用いた．検索結果は，ビジネスチャット Slack 上で RSS アプリ用いて RSS フィードを購読することで参加者内で共有した．チャット上に共有されたコメントの例を図 2 に示す．図 2 中の“⑱”の絵文字は，コメントを閲覧した参加者が有害コメントと判断した印であり，Slack の機能を利用して付

加する．これにより，参加者同士での判断済み情報の共有と，有害コメントの URL の機械的な抽出を可能にした．


図 2　絵文字による有害コメント分類

有害コメントを表す“🔞”付きコメントの情報はスクリプトによって CSV ファイルとして取得・ダウンロードできるようになっている．また，重複した通報を防止するために，このスクリプトは取得済みの有害コメントに対して“✔”の絵文字を付け（図 2 参照），この絵文字が付いたコメントは以後取得しないようになっている．

これらの機能により，プラットフォームがコメントの検索を一括して実行するため，参加者各自が検索をする必要がなく，同一語句による重複検索の問題を解消できた．また，スクリプトにより CSV ファイルが作成されるため，URL をファイルへコピー&ペーストする手間がなくなる．

プラットフォーム使用時の通報活動参加者は，以下の手順で通報活動を実施する．

**(P-1)** Slack 上でコメントを確認
**(P-2)** 有害コメントであれば絵文字で印を付与
**(P-3)** 有害コメントの情報を CSV ファイルとしてダウンロード
**(P-4)** CSV 形式のリストをもとに通報

## 3. プラットフォームの有効性評価

### 3.1 有効性の評価方法

有効性を評価するため，プラットフォーム使用時通報手順と，既存手順の比較を行う．通報にかかる時間は，プラットフォーム，既存手順で変わらない．そこで，既存手順の (C-1) ～ (C-3) にかかる所要時間と，プラットフォーム使用時の (P-1) ～ (P-3) にかかる所要時間を測定する．以下の場合についてそれぞれ 5 回ずつ測定し，その平均の値を比較することで評価する．

- URL 3 件の記録に要する時間
- URL 5 件の記録に要する時間

このプラットフォームは 3 時間で 20 件程度のコメントを収集し，そのうち通報対象である有害コメントは 10 件に満たないため，記録する URL は 3 件及び 5 件とした．

### 3.2 有効性評価の結果

所要時間の測定結果の平均値を表 1 に示す．

表 1　所要時間の測定結果

| | *従来* | *プラットフォーム利用* |
|---|---|---|
| *3 件* | 294 秒 | 137 秒 |
| *5 件* | 569 秒 | 283 秒 |

この結果から，開発したプラットフォームを使用することで URL の記録に要する時間が半分以下に短縮されており，通報作業の効率化が実現されたことが分かる．これは，プラットフォームと CSV 化スクリプトが，検索とファイルの作成を代行した効果だと考えられる．

## 4. まとめ

本研究では，児童被害防止のための有害コメント収集プラットフォームを構築し，その有効性の評価を行った．評価の結果，本プラットフォームは，検索及びリスト作成の自動化により，有害コメントの通報活動にかかる時間を短縮することで，効率化に貢献できることがわかった．

一方，有効性評価の過程で，いたずら目的のコメントや凍結済みコメントが多く見受けられたことから，収集精度の向上，凍結済み有害コメントの除外などの必要性が確認された．また，本プラットフォームの機能は既存サービスに依存しているため，今後は，その依存の解消についても取り組んでいきたい．

## 参考文献

[1] “取組紹介 - 宮城県警察サイバー犯罪対策課,” 宮城県警察，[オンライン]. Available: http://www.police.pref.miyagi.jp/hp/cyber/torikumi.html.

[2] 住田淳，亮隆弘，菱田隆彰，“児童被害を抑止するための SNS 上の不正コメント抽出方法,” *情報処理学会第 80 回全国大会講演論文集*, pp. 117-118, 2018.

[3] 亮隆弘，住田淳，菱田隆彰，“サイバーパトロール活動支援アプリケーションの開発とその有効性,” *情報処理学会第 80 回全国大会講演論文集*, pp. 125-126, 2018.

[4] Queryfeed, “Queryfeed | Twitter, Instagram, Google Plus and Facebook on RSS,” [オンライン]. Available: https://queryfeed.net/.

[5] Slac, “よりシームレスなチームワークを実現する、ビジネスコラボレーションハブ | Slack,” [オンライン]. Available: https://slack.com/intl/ja-jp/.

[6] Slack, “RSS | Slack App ディレクトリ,” [オンライン]. Available: https://slack.com/apps/A0F81R7U7-rss.