

誹謗中傷を含む情報のフィルタリング結果を表示するインタフェース

石丸隆司

Ryuji Ishimaru

1 前回での課題

インタフェースについての考えが不十分であり、工夫が足りなかった。自分のテーマに対する考えが弱かったので今一度テーマを見直して、考え直した。

前回での課題を解決するために今回私は、立命館大学2013年度卒の岩佐さんの論文を読むことで誹謗中傷の情報に対するフィルタリングについて調べてきました。また、フィルタリング後に表示させるインタフェースについて考えてきました。

2 今回の課題

2.1 誹謗中傷の情報に対するフィルタリング

誹謗中傷を含む情報をフィルタリングするうえで、誹謗中傷ワードの判定が重要となってくる。

岩佐さんの論文『電子掲示板において投稿を避けるべき文章の自動判断システム』より、誹謗中傷ワード(以後バッドワード)を大きく2種類に分けていた。

一つ目は、どのテーマや内容にも共通して誹謗中傷と捉えることができる共通バッドワードである。岩佐さんの研究において以下のものが共通バッドワードとして扱われていた。

ブス、デリヘル、エロゲ、ホモゲ、クソ、糞、自殺、アホ、アフォ、阿呆、バカ、馬鹿、ボケ、クズ、屑、カス、キチガイ、マジキチ、基地外、キモ、キモイ、ウザ、ウザイ、老害、マンコ、チンポ、インポ、包茎、パクリ、パク、チョン、リョナ、フルボッコ、グロ、無能、DQN、ブサイク、不細工、ブサ、駄作、愚作、ババア、ババァ、ブチギレ

二つ目は、スレッドのテーマに応じて誹謗中傷と捉えることができるスレッドバッドワードである。例えば、「学校」というテーマに対して「小学生みたい」と書かれても不思議ではないが、ほかのテーマに対して「小学生みたい」とテーマに関係ない形で書かれたら、相手からバカにされていると感ずることができる。このような単語をスレッドバッドワードと判別させるために岩佐さんは、投稿文章の中から記事重要単語、共通バッドワード、既存のスレッドバッドワードと同時に使われることが多ければその単語をデータベースに格納し、同時に使われることが少なくなればデータベースより削除して随時スレッドバッドワードを更新するという方法を用いられていた。

本実験でのバッドワードの判定は、この方法より行い

たいと思う。

2.2 インタフェースについて

結果を表示させるインタフェースだが表示させ方に様々な問題が生じる。子供やインターネットの経験が浅い方たちだとバッドワードは悪影響であるが、インターネットに慣れている方からしたら2chなどでのバッドワードは当たり前であり、左右されることもないと思う。また、バッドワードのあとに否定の言葉が来る場合に対して文全体やその人の書き込みすべて表示させないというのは閲覧側の情報量の減少につながると思う。一方で、「株式会社 詐欺」という書き込みがあったとする。この会社が本当に詐欺の会社であれば閲覧側にとって有益な情報であるがこの会社が詐欺の会社でなければ閲覧側と会社側どちらに対してもいい情報ではない。ネットの情報だけで真偽の判別がつかないからこそこのような誹謗中傷に対しては、信じさせすぎるのもいけないし、また、逆のこともいけないと思う。

以上のことよりインタフェースについての課題は

- 経験や年齢層によってバッドワードの受け取り方が違う。
- 文に否定がつく場合もあるので単語や文、書き込みすべてを消去すべきではない。
- ネットの情報なので真偽があいまいであるため情報は簡単に削除できない。

となるので、これらのことを考えると

- 経験や年齢層に応じてフィルタリング後の表示を変える。
- 単語や文、書き込み全体を非表示にする必要はない。

という機能がインタフェースにほしいと感じインタフェースを考えた。

私たちのような若者やインターネット経験が浅くない方には、共通バッドワードとスレッドバッドワードを消す必要はないと考えた。しかし、見ていていい気分ではないと思ううえに、上記にも書いたように「詐欺」という単語が仮にスレッドバッドワードになったとする。この場合、「詐欺」という単語の真偽がわからずバッドワードと判別しづらいものとなる。それを踏まえて裏面に元のスレッド画像 (Fig1) とフィルタリング後のイメージ画

像 (Fig2) を表示した。共通パスワードに対しては上から線を引き、スレッドパスワードや真偽がわからないものに対しては、その文に下線で色付けしたいと思う。共通パスワードリストの中の単語は、いかなる時でもパスワードととらえることができるものばかりなので表示させないようにする。スレッドパスワードは、必ずしもパスワードと捉えられるわけではないが注意すべき言葉であるため警告させる。

子供やインターネット経験が浅い方たちに向けたインタフェースは Fig3 である。共通パスワードがある場合は上から完全に塗りつぶします。仮にその情報が重要であり、必要だったとしても子供に共通パスワードを見せる必要はないと考えたからである。また、緑色というそんなに目立たない色で消すことで消した部分を強調させないようにする。あとの表示方法は Fig2 と同じである。



Fig.1 元のスレッド画像



Fig.2 若者、経験者向けのインタフェース

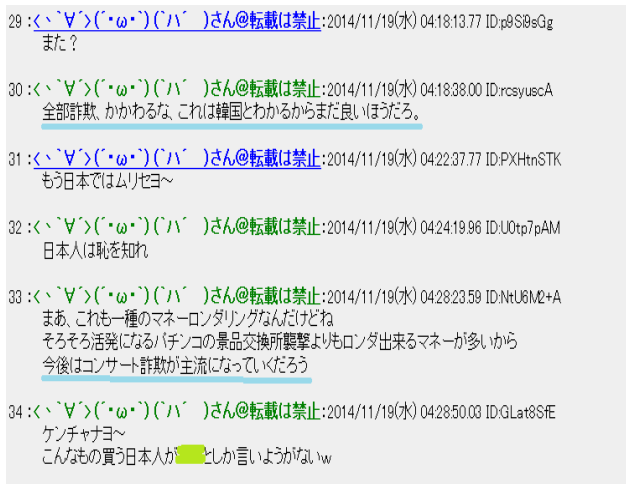


Fig.3 子供、不慣れな方向向けのインタフェース

3 次回の発表までにすること

- インタフェースについてのさらに工夫する。
- Fig3 向けの人たちの年齢や経験歴についての具体的な数字を考える。
- 否定された文への対応を考える。

4 参考文献

- 立命館大学 2013 年卒 岩佐一樹『電子掲示板において投稿を避けるべき文章の自動判断システム』
- 2ch 【芸能】なんと怪しい韓流ライブの舞台裏、「T-ARA」チケット詐欺疑惑の真相... 主催する日本のイベント会社が“雲隠れ” [11/19] (<http://yomogi.2ch.net/test/read.cgi/news4plus/1416333960/>)