

汎用 Web サービス型スパムフィルタリングサービスの事業化

－無料ウェブサービス上のスパム行為の検出－

1. 背景

近年、ブログや SNS などの CGM(Consumer Generated Media)が、インターネットユーザの情報発信・情報共有の場として注目されている。また、企業がマーケティング調査やプロモーションを行うクチコミ情報のメディアとしても注目されている。

一方で、検索エンジンやユーザーランキング、コメントやトラックバックなどの CGM の機能を巧みに利用して、さまざまな商品・サービスの販売や詐欺行為を目的とするスパムが数多く存在する。

ブログにおいては、全体の 4 割以上がスパムとされ、CGM ポータルサイトを運営する企業や CGM の検索サービスを提供する企業、CGM からクチコミ情報を解析する企業、CGM 上でクチコミを生成する企業などの、CGM 関連の各事業者はスパムへの対応を求められている。


2. 目的

CGM のスパムは内容や手法が多岐に渡り、キーワード検出などの従来の手法で検出することは困難である。

スパム行為の特長を突いた独自のスパム検出技術を用いて、CGM 関連事業者に対して、SaaS 型のスパム検出サービスを提供する。

3. 開発の内容

開発したシステムは、クローリング機能、スパム検出機能、インターフェイス機能に分けることができる。



【クローリング機能】

スパム検出の対象となる CGM にユーザから新着記事が投稿された場合に、クローリングによって、その記事の本文部分を抽出し、収集する。

【スパム検出機能】



クローリング機能により収集した記事から、他人の文章のコピーで生成された記事や、ソフトにより自動的に生成された記事をスパムとして検出する。

【インターフェイス機能】



スパムの URL や記事本文を一覧で確認し、スパム判定を確認するためのインターフェイスを提供する。
このインターフェイスは、Microsoft Internet Explorer で動作するため、PC にソフトウェアをインストールする必要がない。

4. 従来技術(または機能)との相違

CGM のスパムは、ウェブサイトの内容や金銭獲得の手段が多岐に渡り、一般の人が行うアフィリエイト等との区別が難しいため、検出が困難である。

従来技術のうち、代表的なものとして、キーワードによる検出やリンクによる検出が挙げられる。

- キーワードによる検出は、成人向けの内容や金儲けに関する内容など、特定の内容に対して有効であるが、アフィリエイトなどにより、一般的な商品を扱うスパム行為を検出することは困難である。
- リンクによる検出は、被リンク数などからスパムサイトを見つけ、それにリンクする多くのサイトを見つける点で優れているが、コンテンツマッチ広告などのアフィリエイト行為を検出することは非常に難しい。また、リンクによる場合にも、ウェブページの転送やフラッシュの利用など、スパムサイトの特定を回避する各種の手法も存在する。

本製品で用いられている独自の検出技術は、文章の一部がコピーされていることを検出するため、成人向けや金儲けの情報を扱ったスパムに限らず、様々な内容のスパムを検出することができる。また、大量の CGM を運用して行われるアフィリエイトについても、スパムサイトとして検出することができる。

5. 期待される効果

【高い検出精度】

独自の検出技術により、従来は検出が困難であったスパムを検出することができる。

【導入の容易さ】

SaaS 型のサービスにより、顧客企業は、自社システムの構成を変更することなく、スパム対策を行うことができる。

【検出精度の維持】

SaaS 型のサービスなので、日々増え続け、手法も変化するスパムに対する最新の対策手法を取り入れて提供することができる。

6. 普及(または活用)の見通し

現在、約 10 社が導入に向けて検討中である。

7. 開発者名(所属)

坂根 毅、児濱 大 (株式会社ユーテラス)

(参考)開発者URL

<http://www.u-tellus.com/>