

# SNS炎上報知器

— 個人情報流出・炎上防止見守りシステム —

自由部門

登録番号 20014

# はじめに —社会背景—

SNSによる炎上が増加している！

- 現在、SNSの利用者は**5000万人**を超えている
- SNS上の不適切な発言が発見・拡散され、取り返しのつかない事態へ発展する事案が**多発**

## 過去3年間の炎上件数

- 個人的な炎上 90件
  - ・ 個人情報晒された事例 60件
  - ・ 晒されるまでに至らなかった事例 30件
- 職務や企業が関係する炎上 21件

## 実際にあった事例

- ・ コンビニ店員がアイス冷凍庫に横たわる写真を投稿して炎上  
結果：フランチャイズ**契約解消**
- ・ スポーツ店店員が来店したサッカー選手の同伴者を中傷  
結果：店員の個人情報が晒され店は**謝罪**
- ・ 看護学生が患者の臓器写真を投稿して炎上  
結果：学生は**退学処分**

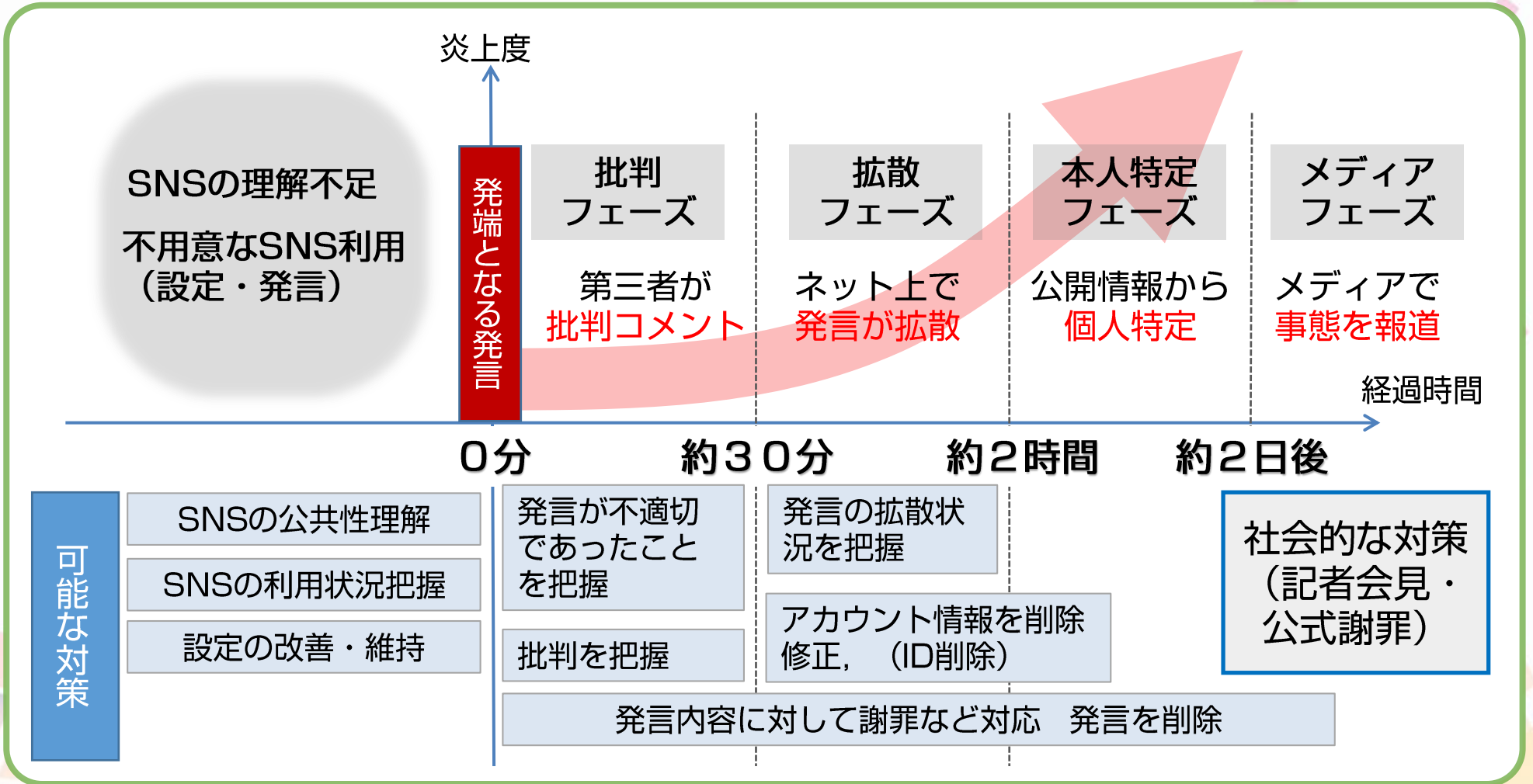
## 原因

- ・ SNSの公共性の**認識不足**からの不適切な発言
- ・ 個人情報が公開される**設定を放置**しSNSを利用
- ・ 炎上把握と対策が**短時間では困難**

## 炎上として社会問題化

（「炎上」急増 予防に躍起、日経新聞2013.7.6）

# 典型的な炎上のプロセス



炎上被害は  
**2時間以内**の対策が必須  
→注意するだけでは対応困難

事前対策, 炎上の危険性察知と対応を促す

**SNS炎上報知器** を開発!!

# SNS炎上報知器のシステム構成

## 炎上防止機能

## 個人情報の管理

### 公開設定のチェック

定期的に、プロフィール等の公開設定を確認

設定の不備に対する「気づき」をユーザへ与え  
炎上を未然に防止します

### NGワードの検出

NGワードが投稿にないかリアルタイムに確認

ユーザにアラートし、  
削除・謝罪を促します

### 炎上感知

リツイート、リプライ、シェア等の数の  
異常な増加をチェック

異常をユーザにアラートし、  
炎上を警告します

### 個人情報の検出

投稿に個人情報が無いかリアルタイムに確認

ユーザにアラートし、  
削除・訂正を促します

### 統計処理・特徴抽出

一定期間内のTwitterとFacebookの投稿  
内容から個人を特定できる可能性のチェック

特定される可能性を数値化して  
ユーザに警告します

# 実現方法 —炎上防止機能(投稿直後30分以内)—



NGワードを発見したらすぐにユーザに知らせます

ユーザ

スマートフォン

登録したアカウントの投稿内容を常に監視します

SNS

アラート

投稿の監視

ユーザはアラートを受け、このアプリから素早く投稿の削除をすることができます

Twitter, Facebookの投稿内容をアプリが監視し、NGワード辞書に含まれる単語があった場合、利用者にアラートを届けて知らせます

NGワード辞書

チェック

# 実現方法 —炎上防止機能(投稿後2時間以内)—

リツイート・リプライ・コメント等を検知

反応の数による処理

リツイート, リプライ, シェア数を  
定期的に取り得し, 増加度を計算

増加度が異常である場合  
ユーザへアラート

反応の内容による処理

リプライ, コメントを取得し,  
NGワードを検索する

NGワードが見つかった場合  
ユーザへアラート

アラートを受け, その投稿を  
ユーザが確認し, 対処する

# 実現方法 —個人情報の管理—

## リアルタイム処理

TwitterとFacebookを  
常時見守り、  
個人情報に繋がるツイートや投稿を  
発見・通知し、訂正を促します

## 統計処理・特徴抽出

TwitterとFacebookの  
プロフィール・過去の投稿を  
遡って比較し、  
互いの類似性が高い場合は  
「個人が特定される恐れあり」  
として投稿の修正を促します

## 定期チェック

プロフィールの公開設定の確認、  
Facebookはそれに加え、  
投稿の共有範囲のデフォルト設定を  
確認し、広範囲に拡散する可能性が  
ある場合は、警告します。

# 独創的な点

Android端末だけですべて実現

Android端末にインストールするだけで、**全ての機能**を利用可能



スタンドアロンで解析・通知

Android端末内で解析・通知をすることで炎上情報や個人情報特定が危険性を**手軽に確認**することができ、危険なつぶやき・投稿を抑制できます

炎上が起こるような発言をした時も素早く対応

Android端末内でリアルタイムで監視することで問題のある発言の発見やアラートをすぐに出すことが可能になり、**早期に対策を行う**ことができます



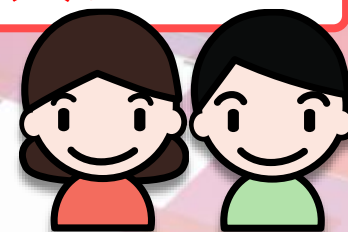
個人の特定を予防

普段、気にすることの少ない公開設定を定期的に管理することで、個人の特定を**事前に予防**できます



# 対象者


Twitter・Facebookを利用している**人すべて**





# 類似サービスとの相違点

## 主要な類似サービス調査結果

e-miningアラート目視サービス（有料，企業向け）  e-mining

→ 事前設定した**リスクキーワード**をSNS上でコンピュータが検知し、**人が目視で異常を判断する**。異常を確認した場合、メールで通知する。企業向けの炎上検知サービス。教育サービス。

評Ban（有料，企業向け）  評Ban

→ 特定の製品名（観測キーワード）がTwitterで**ポジティブに紹介されているか**、**ネガティブに扱われているか**を分析する。個人向けのサービスではない。観測キーワードのつぶやき頻度が上がった場合に通知。

Social Media Visualizer（有料，企業向け）  SM VIZ

→ 特定の製品（観測キーワード）の**評判や客層の調査を行うWebサービス**。**Twitterのみ**を監視し、観測キーワードのつぶやき頻度が上がった場合に通知。製品の市場調査が主目的。

## 調査したその他の類似サービス

whotwi ,つぶやきマネージャー ,コミュニティパトロール ,見える化エンジン ,  
ホットアラート , Omusubi ,WEB監視サービス,Cikappo ,  
ソーシャルメディアWebモニタリングサービス (Deloitte トーマツ) …等

早期に炎上を検知し、**注意を促す個人向けサービス・アプリはほとんどない。**

**社員向け教育サービスとマーケティングサービスがほとんど**

# 将来のビジネスプラン

## ビジネス化へのステップ

個人向けに完全無料で普及を図る



一定のユーザー数や評価を獲得



有料の企業向け製品の開発



企業向けに販売

## 想定される法人向けサービス

### SNS監視

社員や所属タレントのSNS投稿をサーバで一括してフィルタリング・自動監視し、危険な投稿内容、情報漏洩、炎上の検出を行い、企業のダメージを未然に防止するシステム

### ブログ監視

所属する社員等のブログやコメントをサーバーで常に監視し、投稿・コメント内容、コメント数などから、ブログを発端とする炎上を未然に防止・警告するサービス

想定利用法人：

芸能プロダクション、  
野球・サッカー等のプロチーム

# 開発環境

## 開発環境

Windows 7 / Windows 8 /  
Java / Android SDK

## 実行環境

端末：Android搭載  
スマートフォン  
(Android OS)

# 開発計画

タスク名	担当者	2014年06月			2014年07月			2014年08月			2014年09月			2014年10月		
		上旬	中旬	下旬	上旬	中旬	下旬	上旬	中旬	下旬	上旬	中旬	下旬	上旬	中旬	下旬
Androidアプリの設計	全員	■														
API関連	全員	■														
NGワード辞書作成	全員	■														
個人情報辞書作成	全員	■														
過去情報から個人情報特定	B	■														
NGワード分類器製作	C	■														
リアルタイム個人情報監視	A							■								
リアルタイムNGワード監視	AC							■								
投稿の公開範囲アラート	E							■								
拡散アラート	E							■								
初期設定監視機能	E							■								
類似性調査	AB							■								
アプリUI・デザイン	E	■														
クラス内公開・フィードバック	全員							■								
機能のAndroidへの実装	全員							■								
GooglePlayに公開	全員													■		
アンケートサイトでフィードバック	全員													■		
調整	全員													■		
プロコン														■		

9月15日  
GooglePlayへ公開  
フィードバック収集