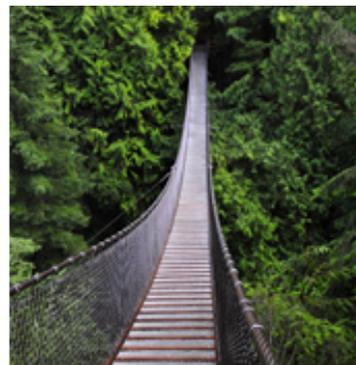
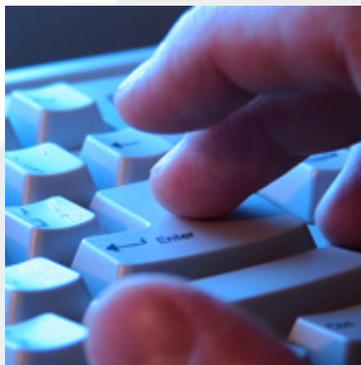


*Add value to your audience*



## **Scan & Target**



**Surveillance temps réel des contenus textes  
web et mobile**

L'usage des moyens de  
communication  
numériques connaît une  
croissance exponentielle



# Genèse du projet

**GARY HAYES' SOCIAL MEDIA**

In the last 71 seconds **there have been...** social media mobile games

**now** +1 day +1 week +1 month +1 year

- 745 **new blog posts**
- 498 **members added on Facebook**
- \$ 12487 **spent on virtual goods globally**
- 3315 **tweets sent on Twitter**
- 828702 **videos watched on YouTube**
- 4192 **iPhone apps downloaded**
- \$ 158 **spent on Facebook gifts**
- 25 **hours uploaded onto YouTube**
- 39 **new Twitter users**
- 89500 **text messages between avatars in Second Life**
- \$ 51538 **made from global mobile messaging and data**
- 35911 **user hours spent on Facebook**
- 13812 **voice hours between users in Second Life**
- 54 **events created on Facebook**
- 1657408 **searches made on Google**
- 174027777 **emails sent globally**
- \$ 1135 **made in advertising revenue on Facebook**
- 13622527 **sms's sent worldwide**
- 111 **videos uploaded to Facebook**
- 2485 **images uploaded to Flickr**
- 469 **new internet users globally**

made by Gary Hayes twitter @GaryPHayes blog <http://personalizemedia.com>

- Un langage écrit numérique n'est plus standard aujourd'hui, on rencontre un grand nombre de variantes / altérations liées aux:
  - Contraintes du terminal (SMS, IM...)
  - Communautés (race/religion/sexe/âge...)
  - Instantanéité
  - Habitudes anti filtrage

### Cas réels de messages bloqués pour un même joueur en moins de 10 s:

- c'est l'arnaque
- c lar nakeeeee
- ya de larnaque laaaaaa
- larnake ce jeu loll
- c l.a.r.n.a.k.e
- c larnakeee

## Genèse du projet

- Les services de police et de sécurité souhaitent de plus en plus **surveiller les communications digitales** notamment écrites (projet Hérisson/loi Loppsi2 en France)
- Les éditeurs proposent de plus en plus souvent des **espaces participatifs** à leurs usagers mais:
  - Risque juridique (LCEN en France notamment)
  - Risque de détournement de trafic
- Les outils traditionnels de Text Mining ne sont pas adaptés pour traiter ces **données non structurées**:
  - Forte volumétrie / Temps réel
  - Utilisation de langages naturels et non structurés
- **Besoin de nouveaux outils** d'analyse et d'aide à la décision

## Qui sommes nous?

Un **éditeur de logiciels créé en décembre 2007** spécialisé dans **l'analyse temps réel des contenus textes web et mobiles** (SMS, e-mails, blogs, forums, chats...)

Notre métier est de fournir des solutions aux **services de renseignement / Police / éditeurs de services** leur permettant d'**analyser, modérer, monétiser** tout **contenu électronique texte**

### Nos solutions fonctionnent:

- ✓ En **temps réel**
- ✓ Quelle que soit la **volumétrie**
- ✓ Sur tout type de contenus (internet, mobile ou TVI), **structurés ou non**
- ✓ Pour toute langue utilisant des **caractères latins, arabe et japonais**
- ✓ Implémentation **flexible** (mode SaaS / SaaS on premises / appliance)

# Exemples de références clients



## Comment ça marche (1/2)

- Contrairement aux solutions basées sur des mots clés simples ou la sémantique, notre technologie peut traiter les différentes **variantes et altérations** des expressions pour analyser le contenu:
  - Utilisation lettres majuscules / minuscules
  - Répétition de lettres (vvviiiagrerra par exemple)
  - Variantes orthographiques (vi@gra, vIagra, v1@gra, v149r4)
  - Elision de lettres dans certains cas (v|agra, v agra...)
  - Utilisation de symboles non alpha numériques (v.i.a.g.r.a, v\_i °ag#r:a, v-iagra, viagr"a...)
  - Altérations phonétiques
  - Formes plurielles et conjuguées
  - Langages SMS et chat

## Comment ça marche (2/2)

---

- Notre solution utilise des thésaurus thématiques. Les filtres sont catégorisés pour permettre aux clients d'affiner leur analyse
- Les **algorithmes de détection** peuvent être **personnalisés** en fonction de vos besoins spécifiques
- La solution Scan & Target repose sur un système de notation non pas du seul mot mais de la totalité de contenu lors de son passage dans le moteur de filtrage. Les **mots** sont replacés **dans leur contexte**.
- Le système de notation prend en compte l'objectif de l'analyse, le media mais aussi les critères spécifiques du service
- **Systeme d'analyse conditionnelle** réduisant le nombre de faux positifs (aujourd'hui inférieur à 0,005% en production)

## Exemples surveillance terrorisme

- Pour détecter des messages potentiellement terroriste dans un flux (SMS, e-mail, chat...)
  - ✓ Vocabulaire des thesaurus actes terroristes (violence / explosifs...)
  - ✓ Vocabulaire des thesaurus organisations terroristes (personnes / organisations...)
  - ✓ Meeting pattern (heure / lieu...)

*“J’ai le C4, rdv à la mosquée al-Rahman à 2h demain matin” ou “3andi C4, khaliina nelte’i fi-djami3 Ar-rahman fis-saa3a at-taniyé Fil-layl »*

remonteront une alerte, contrairement à :

*« Vendredi on se retrouve à la mosquée Al-Rahman à 14h » ou « J’ai essayé la dernière Citroën C4 »*

## Avantages concurrentiels

---

1. Capacité à travailler en temps réel des fortes volumétries sur Internet ou sur Mobile, etc.
2. Pertinence des algorithmes de détection (prise en compte langage naturel et altérations, analyse conditionnelle) avec très peu de faux positifs
3. Solution multilingue et rapide à implémenter (quelques semaines à quelques mois)

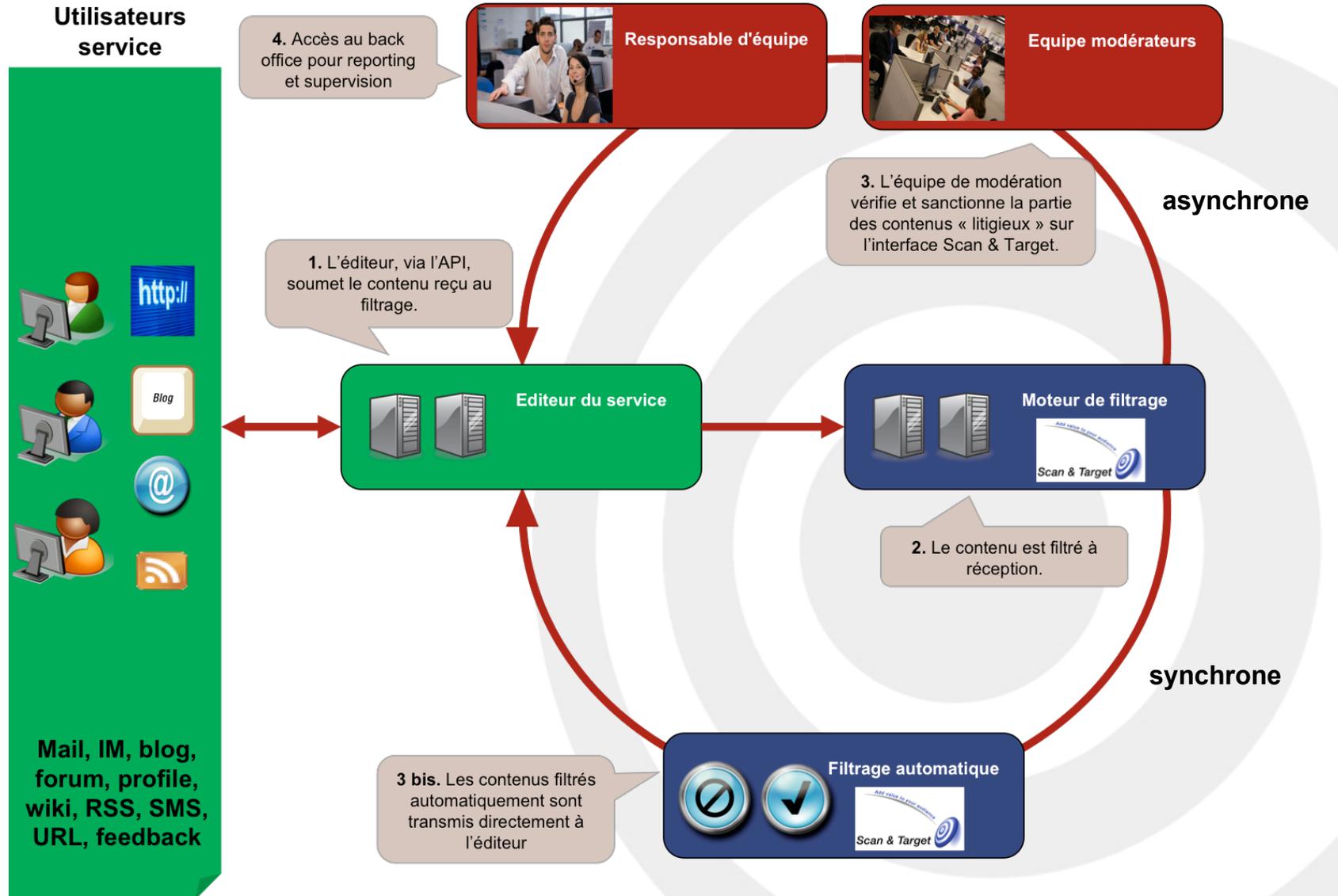
## Notre solution d'analyse

- Notre solution peut **analyser en temps réel tout contenu** web ou mobile texte (emails, SMS, chat, forums...)
- Le moteur va en temps réel catégoriser 100% du trafic et **identifier les messages** potentiellement intéressants par rapport à vos critères d'analyse
- Sur l'application modération, les messages non conformes sont marqués comme à bloquer **avant publication, et sans perturber** l'expérience utilisateur
- Les messages intéressants sont remontés dans une **console de surveillance** dans laquelle les agents peuvent les **analyser ou agir** sur ces contenus ou sur leurs rédacteurs

*Nous n'utilisons **aucune donnée personnelle** pour l'analyse*



# Workflow modération



## Scan & Target SAS

80 rue des Haies  
75020 Paris  
France

Bastien Hillen  
CEO

Phone: +33 (0)6 11 25 53 80  
Mail: [b.hillen@scanandtarget.com](mailto:b.hillen@scanandtarget.com)