



URI免疫化：参加型システムに おけるスパム避けの一手法

国立情報学研究所

北本 朝展



研究の経緯

- 不特定多数参加型システム「台風への眼」「台風前線」を2004年から運用。
- 最近トラックバックスパムに悩まされる。
- トラックバックスパムを避けるための一つのやり方を思いついた。
- いろいろ考え、URI免疫化と名づけた。
- ひょっとすると、インターネット免疫学という新しい体系が作れるのでは？



本発表が対象とする問題

- ブログのトラックバックやコメントなどに対する低コストで簡便なスパム防御。
- ユーザ管理をする必要がないサイトにおけるスパム防御。
- トラックバック承認プロセスを省きたいサイトにおけるスパム防御。
- スパムの行動を追跡しながら対策を動的に適応させるスパム防御。

本発表が対象としない問題

- 人間が自らの手で行うスパムの攻撃の防御（荒らしや宣伝など）。
- スパムの攻撃に巨大な価値があるサイトの防御（あらゆる手段が使われる）
- 閉鎖的参加型システムにおける防御（パスワードによるユーザ管理等）。
- リンクスパムなど参加型システムとは関係のないウェブスパムへの対策。
- 「これだけで何でも解決する」という方法。



デジタル台風

<http://www.digital-typhoon.org/>

- 台風情報に特化したポータルサイト。
- 台風画像の大規模・網羅的なデータベース化という点がユニーク。
- 過去の膨大な観測データから、台風解析や予測に有用な情報を発見する。
- ウェブサイトのページビューは、トップページ約1100万、サイト全体約4160万など、3年間で多くの利用者を得た。

デジタル台風

デジタル台風: 台風画像と台風情報 - Mozilla Firefox

ファイル(E) 編集(E) 表示(V) 移動(G) ブックマーク(B) ツール(T) ヘルプ(H)

http://agora.ex.nii.ac.jp/digital-typhoon/

北本 朝展 @ 国立情報学研究所 (NII) 検索 サイトマップ

デジタル台風: 台風画像と台風情報

agora ホーム デジタル台風 English

:: お知らせ :: ケータイ版 :: PDA版 :: Google Earth版 :: iPod版 :: 壁纸 :: 動画 :: RSS / Atom :: キッズ :: 台風への眼 / 台風前線 :: MTSATプロダクツ ::

リアルタイム台風情報

更新日時: 2006年 11月 7日 5時 8分 (JST)

台風200619号



MTS106110600
200619 (WNP)
(N14.0, E114.0)
1002 hPa / 0 kt

台風発生数

現在 = 19個 (台風経路図)
平年 = 23.5個 (1951-2005)

最新台風ニュース

- 台風ニュース・ウェブログ
2006年台風19号(シマロン | CIMARON) --
2006年11月03日
- 台風ニュース・トピックス
- 地域情報ポータル
- 災害情報ポータル
: 気象災害データベース :: 台風被害データベース :
- その他のニュース
ご意見・お問い合わせ 開設(2006.11.06)

台風データベース	静止気象衛星画像
----------	----------

1 気象データによる検索
完了



台風への眼

<http://eye.tc/>

- 不特定多数の現地の人から発信された情報を集約・共有する「参加型メディア」。
- ウェブログ(ブログ)からのトラックバックを集約してリアルタイムの情報収集。
- トラックバックURL(情報の送り先)の構成法を工夫することで、情報を地域ごとに集約して整理。

トラックバックによる情報集約

「台風への眼」の使い方

トラックバックURL	http://eye.tc/trackback/ping/200604/pRZ
トラックバックを見る	http://eye.tc/trackback/view/200604

まずは「どれ」(WHICH)に関するトラックバックかを設定します。現在、テーマは **台風200604号** に設定されていますが、6桁の台風番号を変更することにより、他の台風に関するトラックバックを送ることもできます。

次に「どこ」(WHERE)に関するトラックバックかを設定します。都道府県ごとのトラックバックURLを以下に示します。なお都道府県名をクリックすると、地方自治体ごと・郵便番号ごとのトラックバックURLを順次表示します。また[検索](#)も可能です。

カタカナ読み : [昇順](#) | [降順](#) :: トラックバックURL : [昇順](#) | [降順](#)

01 **北海道** [[トラックバックのリストを見る](#)]

トラックバックURL : <http://eye.tc/trackback/ping/200604/01/pRZ>

02 **青森県** [[トラックバックのリストを見る](#)]

トラックバックURL : <http://eye.tc/trackback/ping/200604/02/pRZ>

03 **岩手県** [[トラックバックのリストを見る](#)]

トラックバックURL : <http://eye.tc/trackback/ping/200604/03/pRZ>



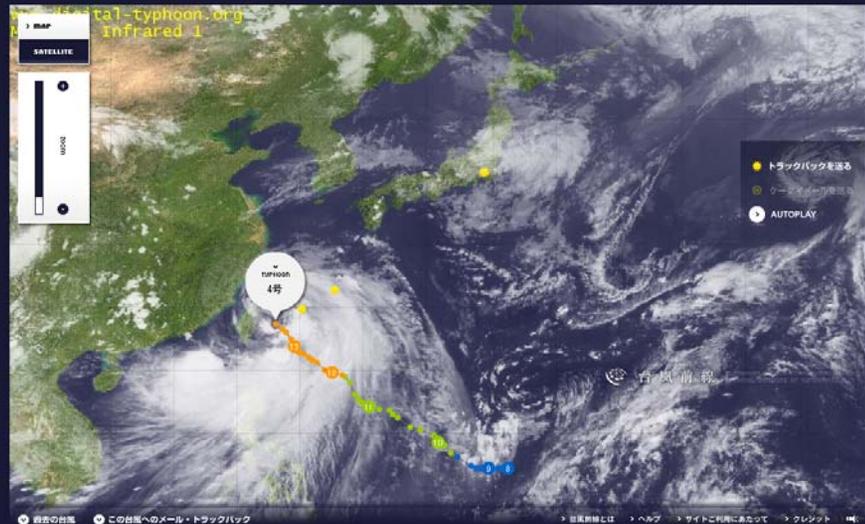
「台風への眼」の経過

- 2004年6月10日に、台風200404号を対象にして実験を開始。
- トラックバックを送信して下さったブログは218(654件)。日本を縦断した台風200514号には74件のトラックバック。
- そのうち7-8割には位置情報が付与されており、地図上に表示可能。
- **時間情報の把握が困難という問題。**

「台風前線」

<http://front.eye.tc/>

台風の動きと、各地の
発信情報との、時間的
な関連が一目瞭然。



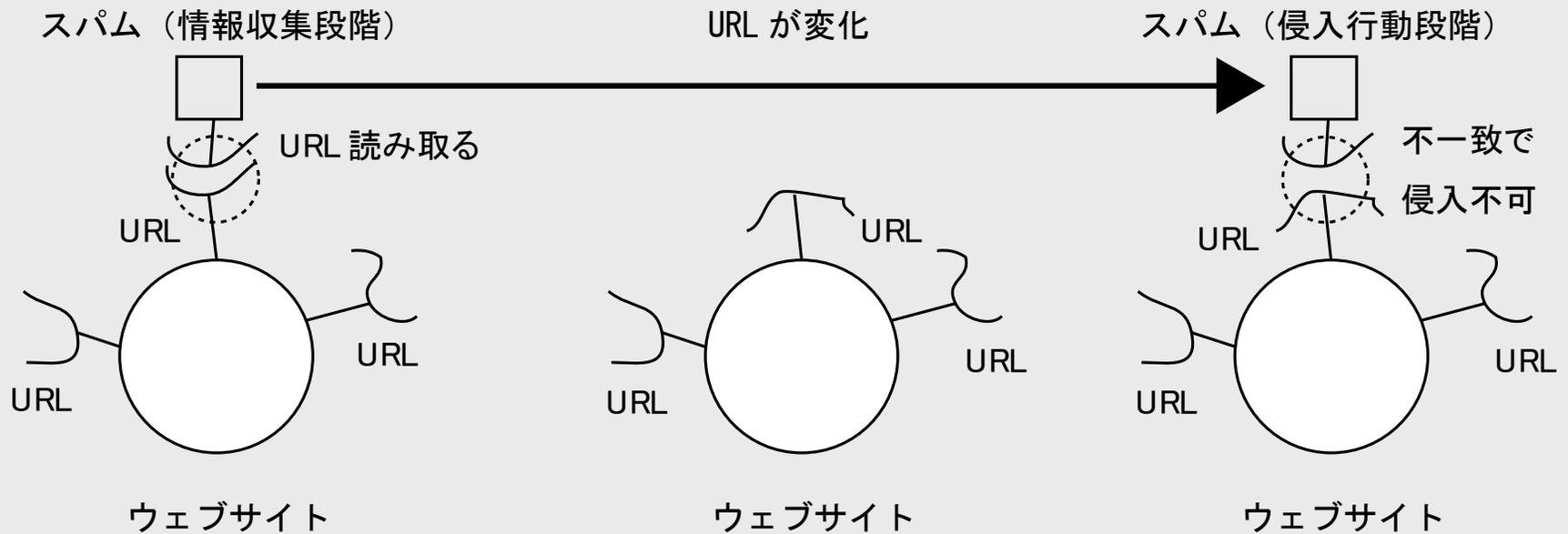
参加型台風情報の特殊性

以下のよくあるスパム対策は使いたくない。

1. 事前登録を要求する。
2. 承認プロセスを組み込む。
3. 言及リンクを要求する。
4. 特定の文字コードを要求する。
5. トラックバック内容の特徴を学習する。

いずれも場当たりのな解決策であり、台風情報を対象とする場合に問題がある。

動く標的URI



スパムが読み取ったURLを変えると、URL
が完全一致せずスパムは侵入できない！

時間差が重要

1. クローラーを巡回させる、あるいは検索エンジンを利用して、トラックバックURIを収集しデータベース化。
2. データベースからターゲットなるURIを読み出し、スクリプトを用いて自動攻撃。

行動1と行動2の間には時間差が生じる。



その間にURIを変えてしまえばよい！

CAPTCHAによる情報提示

- URIが時間変化するため、**最新のURIを不特定多数に提示する必要がある。**
- Trackback Auto-discovery等、スパムが簡単に読み取れる形式では時間差が生じない。
- ゆえに**機械「難読」な形式で、最新URIを提示する必要がある。**
- **CAPTCHA**を用いて、人間とスパムの能力差を時間差に変換する。



URI免疫化とは

- URIを時間変化させる(動く標的URI)。
- 最新URIを機械難読な形式で提示する(CAPTCHAによる情報提示)。
- スпамが情報を収集して、攻撃を開始するまでの時間差に、上限を課すことと同様の効果をもつ。
- これをなぜ「免疫化」とよぶのか？そのところは？その利点は？

免疫系における抗原と抗体

- 体内の異物を認識して除去するには？
- 異物として認識される部分を抗原という。抗原の3次元構造にぴったりとはまる抗体があれば、異物の認識ができる。
- 抗原の形状は無限に近い多様性をもつので、予測不能であるにもかかわらず、それと形状が(特異的に)一致する抗体を事前に用意しておかねばならない。



利根川博士の大発見

- 抗体はどのように多様性を生み出しているのかという問題に取り組んだ。
- 遺伝子が組換えや突然変異により変化して、多様性を生み出すことを発見。
- それまで遺伝子是不変であると考えられていたので、常識を覆す発見だった。
- この研究によって、利根川博士はノーベル賞を受賞した。

可変性

- **生物の免疫系**では、「変化する」というメカニズムにより、無限の多様性をもつ外敵の形状と(たまたま)**一致する**形状の抗体を生成する。
- **URI免疫化**では、「変化する」というメカニズムにより、無限の多様性をもつスパムが提示するURLと**一致しない**URLを生成する。



URIの3分類

- iURI (Import URI) 外部から内部に情報を取り入れるURI
- eURI (Export URI) 内部から外部に情報を吐き出すURI
- sURI (Symport URI) ImportとExportの両方をおこなうURI

スパム防御であれば、ウェブサイトのiURI(とsURI)のみに対策すればよい。



免疫的URIの構成

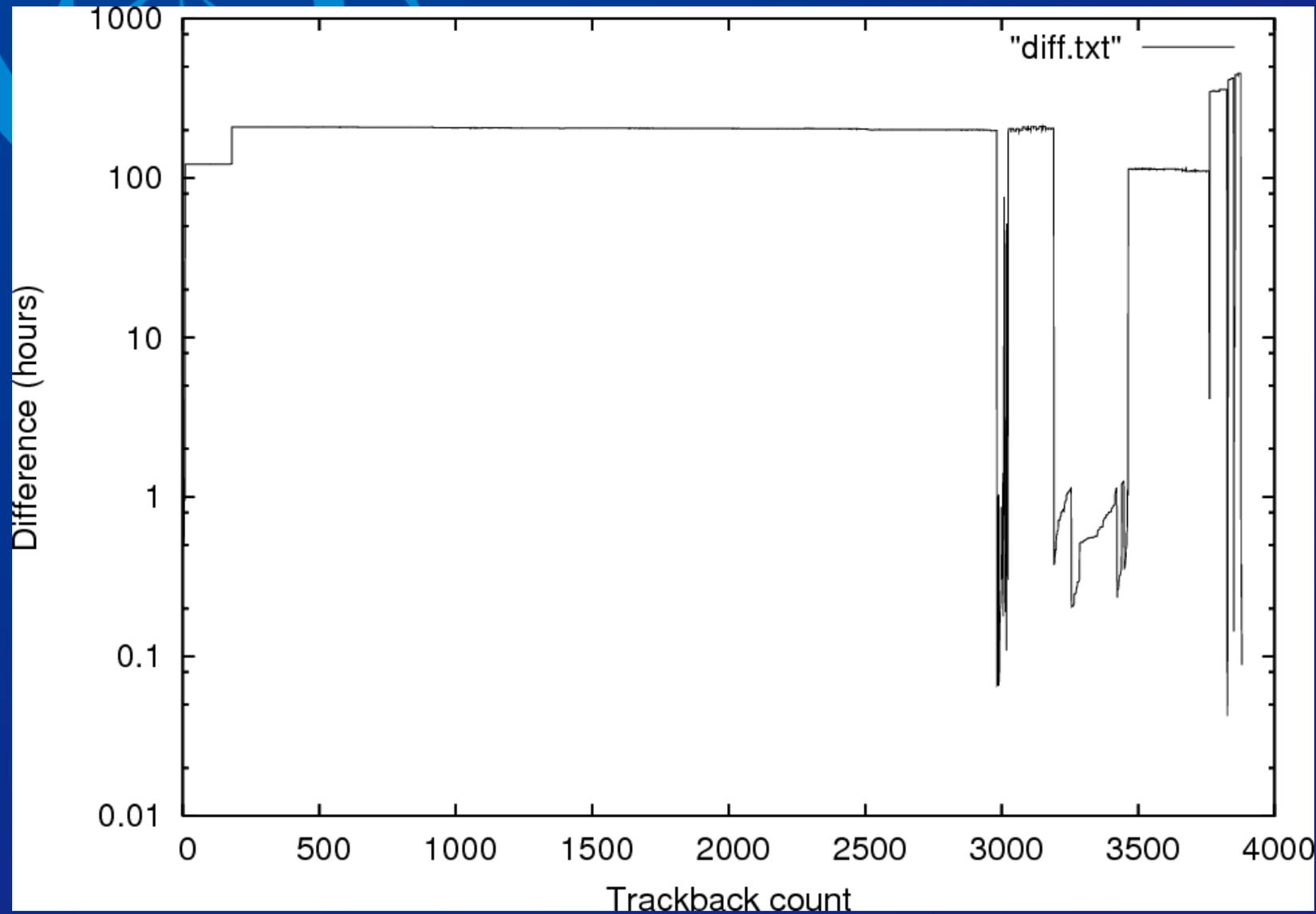
- 免疫的URI = 定常領域 + 可変領域
- 定常領域はリソースを表す不変な記号列（従来の意味でのURI）。
- 可変領域はリソースと無関係な記号列。「ランダム」である必要はなく、スパムが容易に予測できなければよい。
- 可変領域を囲URLと考えると、履歴からスパムが情報収集した時間を推測可能。

解析結果(1)

POSTリクエスト13360件の解析結果

受理アクセス	196
非受理アクセス	3882
時刻差10分未満	14 (0.36%)
時刻差30分未満	81 (2.1%)
時刻差1時間未満	183 (4.7%)
時刻差10時間未満	55 (1.4%)
時刻差10時間以上	3549 (91.4%)
可変領域なし	1278
不正なURL	8004

解析結果(2)



RESTアーキテクチャとの関係

- 「すべてのリソースは一意に識別できるURIをもつ」というウェブの原則に矛盾する？
- PermalinkではなくEphemelinkである。



- 実際に可変URIにすべきなのはiURIのみ。こちらは元々Permalinkである必要はなく、関連するeURIから情報入手できればよい。
- 「すべてのリソースはURI集合で表される互いに素なアドレス集合をもつ」と拡張する。



eURIの免疫化

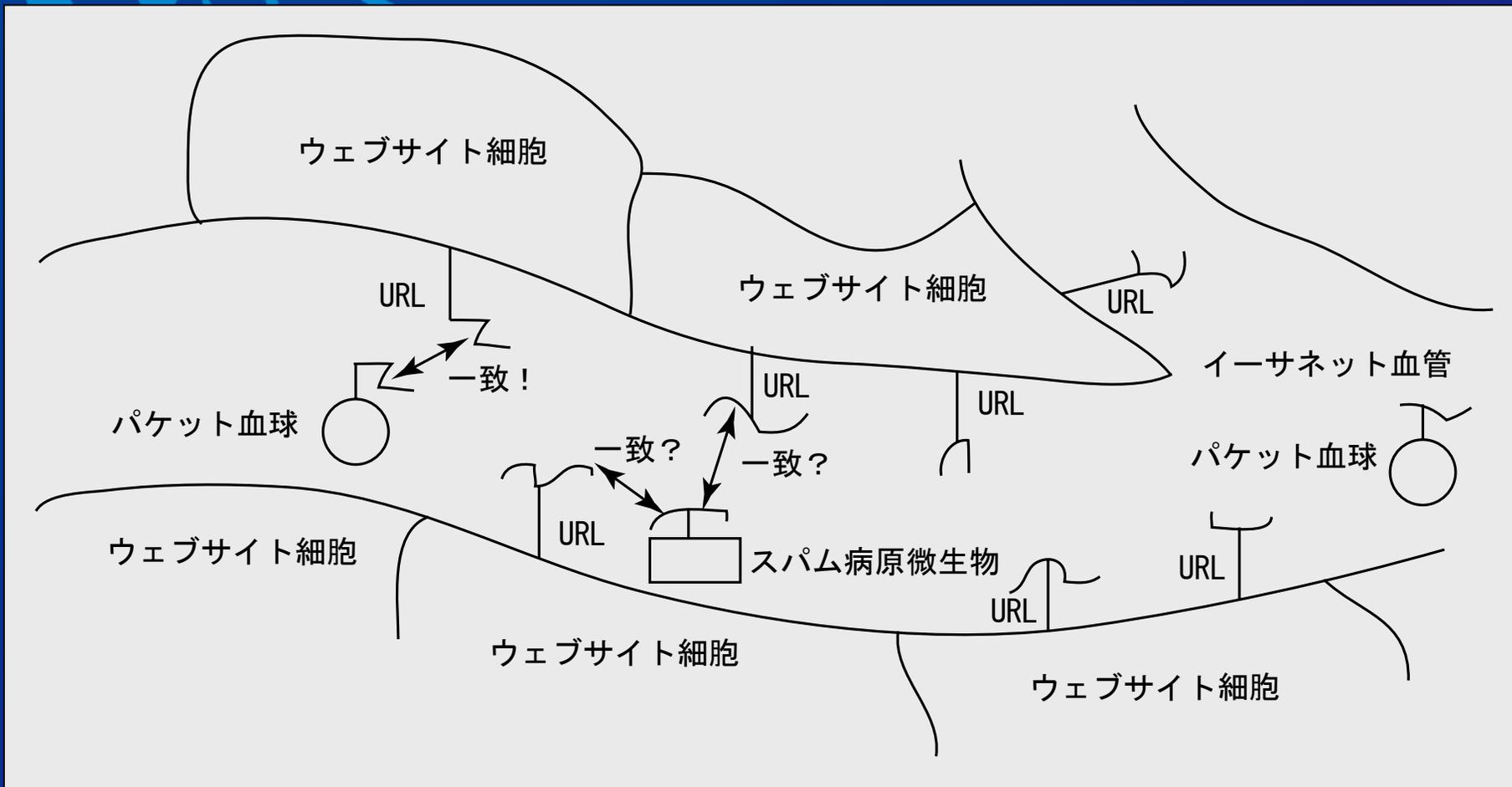
- eURIを可変にすると、外部から(永続)リンク不可能なウェブページが作れる。
- サイト内部で可変記号列を共有すれば、内部での(短時間)リンクは問題ない。
- アーキテクチャレベルで、リンク可能なURIとリンク不可能なURIを指定可能。
- 免疫学的アイデアは、スパム対策に止まらない応用へと広がる可能性あり。



インターネット免疫学

- 生物が外敵の侵入に対抗して進化させてきた免疫系から学び、そのエッセンスをインターネットのアーキテクチャに組み込むことを目的とする。
- アーキテクチャレベルで対抗できるので、スパム防御がより確実になるかも。
- インターネットの制約(レガシーなプロトコル等)を反映した免疫系の構築。

逆免疫システムの全体像





まとめ

- 参加型システムにおけるトラックバックスパムの防御という問題に対して、URI免疫化という方法を提案した。
- URIを可変にすることで、ほぼ100%のスパムを防御し特性を評価できた。
- ノウハウ自体は新しいものではないが、免疫学とのアナロジーから理解することで、見通しがよくなるという利点がある。



呼びかけ

世の中にあるブログツールへの
プラグインを作ってみませんか。
簡単なのでたぶん半日もあれば
作れると思いますよ。もしたく
さんプラグインが生まれれば、
まとめサイトを作るかも。。



謝辞・連絡先

- 参加型システムでは多くのブロガーの方々の自発的な協力を頂いています。
- その他、プロジェクトでは多くのデータを利用しています。
- メールアドレス：kitamoto@nii.ac.jp
- ウェブサイト：
<http://agora.ex.nii.ac.jp/~kitamoto/>