

Information is Nothing without Retrieval



Web Technology & Information Systems [www]
Bauhaus University Weimar

Wellenbrecher für die Kommentarflut im Web

- Martin Potthast
- Benno Stein

Wellenbrecher für die Kommentarflut im Web



Anti-Terrorist Data Mining Doesn't Work Very Well

Journal written by [Presto Vivace \(882157\)](#) and posted by [kdawson](#) on Tuesday October 07, @03:05PM from the [could-have-told-you-and-did](#) dept.

Presto Vivace and others sent us this CNet report on a just-released NRC report coming to the conclusion, which will surprise no one here, that [data mining doesn't work very well](#). It's all those darn false positives. The submitter adds, "Any chance we could go back to [probable cause](#)?"

"A report scheduled to be released on Tuesday by the National Research Council, which has been years in the making, concludes that automated identification of terrorists through data mining or any other mechanism 'is neither feasible as an objective nor desirable as a goal of technology development efforts.' Inevitable false positives will result in 'ordinary, law-abiding citizens and businesses' being incorrectly flagged as suspects. The whopping 352-page report, called 'Protecting Individual Privacy in the Struggle Against Terrorists,' amounts to [be] at least a partial repudiation of the Defense Department's controversial data-mining program called Total Information Awareness, which was limited by Congress in 2003."

Wellenbrecher für die Kommentarflut im Web

The screenshot shows a Slashdot page with a search bar and navigation links. The main content is a comment thread. On the left side, there is a vertical text overlay that reads: "Anti-Terrorist Journal writer by kdawson on from the could- Presto Vivace ar on a just-release conclusion, which data mining doe darn false positi chance we could "A report sche by the Nationa been years in automated ide data mining or feasible as an of technology will result in 'o being incorrec 352-page repo Struggle Again partial repudia controversial Awareness, wh".

Slashdot IT IS WHAT IT IS. Search

► Log In | Create Account | Help | Subscribe | Why Log In? | Why Subscribe?

The mere concept of a list of names of people who are too "dangerous" to let fly ... but not dangerous enough to track ... that just fucking stupid.

Think about how many people could be killed in the airport terminal itself WITHOUT getting on a plane ... say during the Thanksgiving or Christmas rushes there.

What idiot would let the people on that list (if they were really a threat) into a terminal? Wouldn't you expect them to STOP them BEFORE they get into a position to do that kind of damage?

Reply to This Parent

Re:'d run on that platform. (Score:5, Insightful)

by Fulcrum of Evil (560260) on Tuesday October 07, @03:31PM (#25290757)

The no fly list doesn't identify people, just names, and it's very exact, so changing charles to chuck will defeat it. The upshot is that it's utterly useless for stopping bad guys, so you can't even identify who's on there - John Smith is on the list, but there are 10,000 of them.

Reply to This Parent

Re:'d run on that platform. (Score:5, Informative)

by Geoffrey Landis (926948) on Tuesday October 07, @04:16PM (#25291263) Homepage

The no fly list doesn't identify people, just names, and it's very exact, so changing charles to chuck will defeat it.

No, actually it won't. The newspapers are full of stories of people who were detained or forbidden from flying because their name was *similar* to a name on the list, or a *nickname* of a name on the list, or a possible alternative spelling of a name on the list, or names that had once been used as an *alias* of names on the list.

for example, the name "T. Kennedy" was on the list. Senator *Edward Kennedy* (whose name does not begin with "T", but who is nicknamed "Teddy") was stopped: from [Wikipedia](#) [wikipedia.org]

"In August 2004, Senator Ted Kennedy (D-MA) told a Senate Judiciary Committee discussing the No Fly List that he had appeared on the list and had been repeatedly delayed at airports. He said it had taken him three weeks of appeals directly to Homeland Security Secretary Tom Ridge to have him removed from the list. Kennedy said he was eventually told that the name "T Kennedy" was added to the list because it was once used as an alias of a suspected terrorist. There are an estimated 7,000 American men whose legal names correspond to "T Kennedy". (Senator Kennedy, whose first name is Edward and for whom "Ted" is only a nickname, would not be one of them.)"

Reply to This Parent

Re:'d run on that platform. (Score:4, Insightful)

by low (121541) on Tuesday October 07, @04:36PM (#25291499) Journal

Right, but flying under your middle name does work. As does claiming that you lost your ID (but if you refuse to show it on principle, you can't fly). As does using one boarding pass with matching ID at security, and a different boarding pass with matching ID at the gate.

The really sad thing is, the people who the government feels are a real threat based on strong intelligence are *not* on the no-fly list! The government doesn't want to reveal to the real suspects that their being watched.

Reply to This Parent

Re: (Score:3, Interesting) by kalrion (728907) What would happen if terrorists got nicknames after all major U.S. and U.K. political figures.

Try "chuck" (Score:3, Funny) by EmbeddedJanitor (597931) It probably uses Google search engine or similar. Tell TSA-guy you're German and the "s" are part of your name. At \$8/hr TSA-guy isn't paid to think.

Re:'d run on that platform. (Score:5, Informative)

by Free the Cowards (1280296) on Tuesday October 07, @05:48PM (#25292473)

It doesn't matter, because the only place where you *have* to get your ID checked is at the TSA checkpoint, and they don't check it against any databases.

Wellenbrecher für die Kommentarflut im Web

Forschungsfragen

- ❑ Was sind geeignete Retrieval-Modelle?
- ❑ Wie misst man Relevanz? (nicht über Textähnlichkeit!)
- ❑ Wie extrahiert man Stimmungen?
- ❑ Wie erkennt man Polarisierungen oder Desktruktivität?
- ❑ Wann und wie helfen User-Profile?
- ❑ ...
- ❑ Wie visualisiert man die Analyse?

Wellenbrecher für die Kommentarflut im Web

YouTube-Beispiel [Becker/Loose/Potthast/webis]



Wellenbrecher für die Kommentarflut im Web

YouTube-Beispiel: Technologie

- Filtern der Kommentare d mit $|d| > 30$.
- Parsen von d und Vergleich mit Stimmungs-Dictionary D_S .
- Test bzgl. Modifikatoren (not, very, ...)
- Visualisierung (Impact, Stimmung, etc.)

Herausforderung ist die Konstruktion von D_S :

- Ausgangspunkt ist ein von Menschen erstelltes Seed-Dictionary.
- POS-Analyse vieler YouTube-Kommentare zur Filterung von Adjektiven.
- Co-occurrence-Analyse auf Basis des Seed-Dictionaries und Vergabe von Stimmungswerten

Projektschwerpunkte

Modelle und Algorithmen

IR, IE: Sentiment Analysis,
Polarization Analysis,
Text Summarization,
maschinelle Lernverfahren

Voraussetzung: Interesse an Statistik, analytisches Denken

SWT, Web, Soziologie

Konzeptionell: Architektur von Web 2.0 Anwendungen verstehen,

Interface: Integration in Web-Anwendungen

Evaluation: Korpora erstellen
Entwicklung von Belastungstests

Voraussetzung: Kreativität, Java, Web-Sprachen,
gutes bis sehr gutes Verständnis für Web-Technologie