

BlogINT: Weblogs as a Source of Intelligence

John Griffith

781-271-7736 • griffith@mitre.org

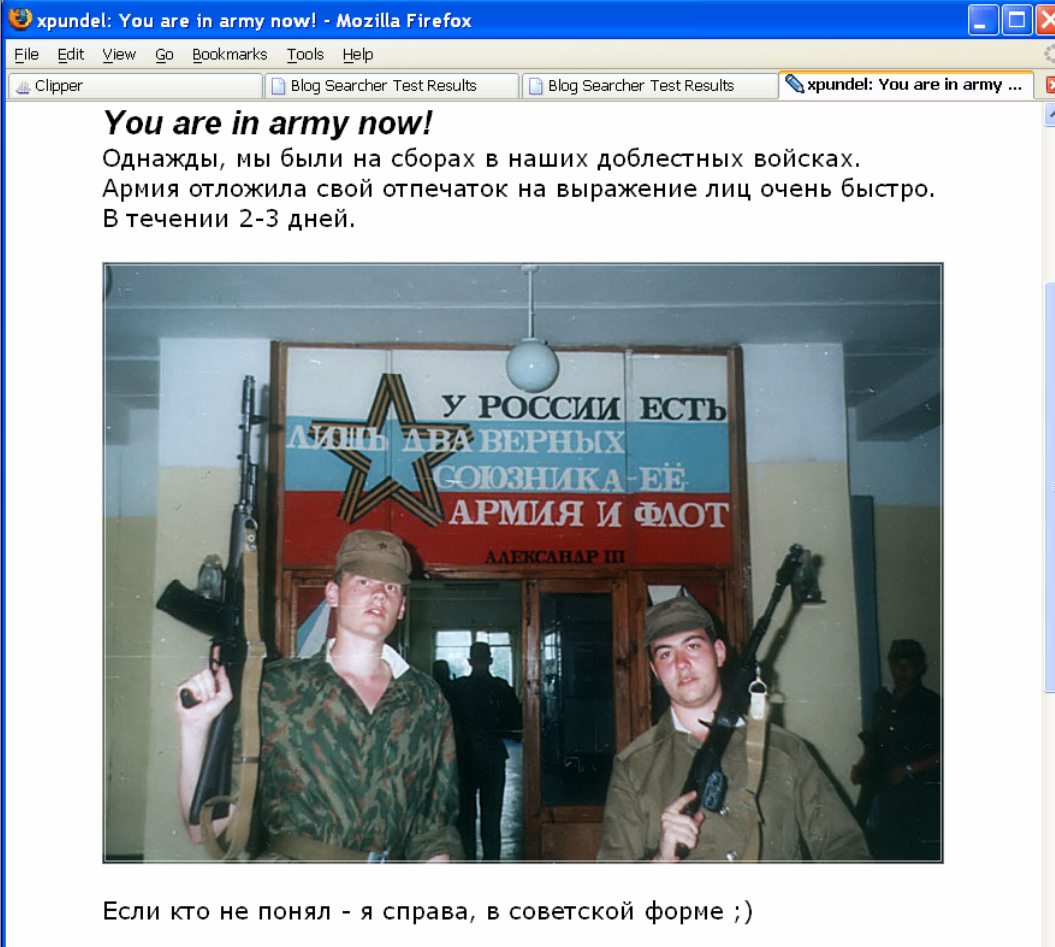
MITRE Sponsored Research



Problem

- Web logs, *blogs*, are an emerging source of information.
 - Contain personal reports and images
 - Often authored on mobile devices and published in seconds
- Their value as a source of intelligence is unknown.
 - Must be sifted for pertinent information
 - New, relevant events must be detected

Background



The screenshot shows a Mozilla Firefox browser window with the title "xpundel: You are in army now! - Mozilla Firefox". The address bar contains "xpundel: You are in army ...". The page content includes the following text:

You are in army now!
Однажды, мы были на сборах в наших доблестных войсках.
Армия отложила свой отпечаток на выражение лиц очень быстро.
В течении 2-3 дней.

The central image shows two young men in military uniforms standing in front of a sign. The sign features a large star and the Russian text: "У РОССИИ ЕСТЬ ЛИШЬ ДВА ВЕРНЫХ СОЮЗНИКА - ЕЁ АРМИЯ И ФЛОТ" and "АЛЕКСАНДР III". The man on the left is wearing a camouflage uniform and a cap, holding a rifle. The man on the right is wearing a brown uniform and a cap, also holding a rifle.

Если кто не понял - я справа, в советской форме ;)

There are millions of active blogs containing multilingual and multimedia information available as XML documents.

MITRE

© 2005, The MITRE Corporation

Objective

- **Characterize the benefits, costs, risks, opportunities, and challenges associated with exploiting blogs as potential sources of intelligence**
- **Compare blogs to traditional Web-based news and broadcast television in terms of content, accuracy, and timeliness**

Activities

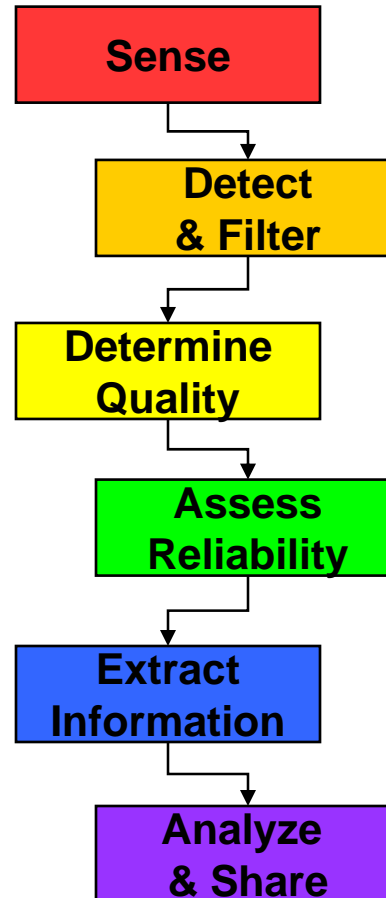
- **Collecting blogs, TV and Web news**
- **Developing a blog-specific search engine**
- **Comparing event coverage in blogs versus traditional media**
- **Developing techniques for blog classification and novelty detection**
- **Assessing annotation tools for multimedia**

Highlight

- Volume
- Temporal behavior
- Language, modality
- Camera type

- Use higher quality items with comparable information

- Event detection & extraction
- Identify direct observation



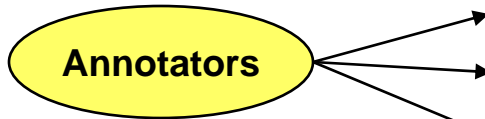
- Detect novelty
- Filter junk
- Search
- Co-cluster with news

- Reputation v. new authors
- Rumors with confidence
- Fact checking sites

- Propagate evidence
- Situational awareness
- Mobile coordination

Demonstration

Training a document classifier



Reconciling Document 19573933 - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

Annotation History (reconciliation in bold)

Tagset: MilitaryTask		
cdoran	Yes	2005-03-11 10:50:52
jlndrsn	Yes	2005-03-11 11:25:13
griffith	No	2005-03-11 14:22:11
cleo	Yes	2005-03-11 16:12:47
griffith	Yes	2005-03-12 11:25:55
jtsmith	No	2005-03-14 09:56:40
jtsmith	No	2005-03-14 09:56:55

Reconciling Document 19573933

Blog: 104/100

Tag: Yes No

Comment:

Tagset: **MilitaryTask**

Annotator: griffith

Bundle: bundle0001

The US officially observes Martin Luther King Day as a national holiday for the first time

The space shuttle Challenger explodes moments after lift off, killing 6 astronauts and a teacher

A major nuclear disaster occurs at the Chernobyl nuclear power plant in the Soviet Union

Japanese video game maker Nintendo introduces its games to America

US warplanes bomb Libyan headquarters in retaliation for terrorist attacks

The Soviet Union launches the Mir space station

IBM unveils the PC Convertible, the first laptop computer

Done

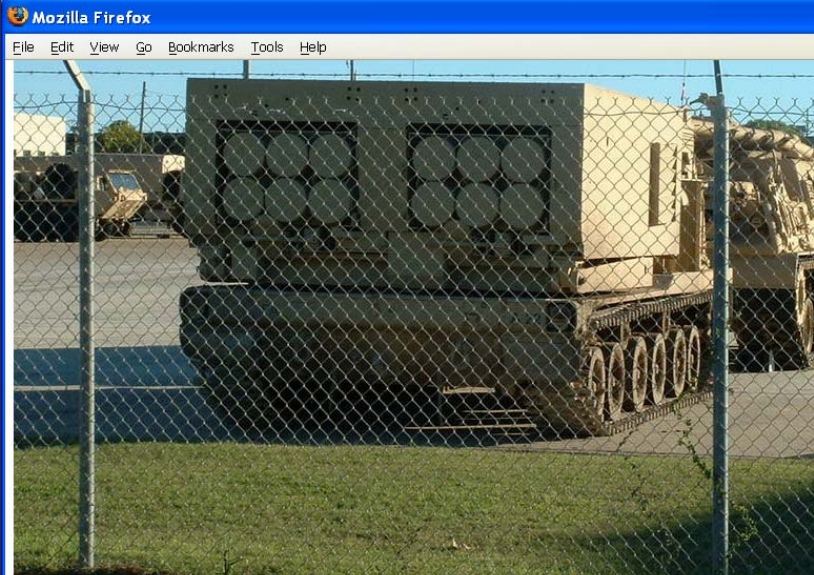
Impacts

- **Characterization of intelligence value of blogs**
- **A large, longitudinal data set**
- **Real-time blog collection and indexing system**
- **Tools and techniques for retrospective analysis of events in blogs**

Future Plans

Open our lab doors to analyst-centered pilot studies

Utilize found metadata as well as automatically extracted metadata to further characterize the items in our collection



http://flora.mitre.org/blogint/blogint2_newlocation/results/images/img_photobuc

Aperture Stop	F8.2
Brightness	333.14foot-lambert
CompressedBitsPerPixel	2
Copyright	
Date	2004:11:06 17:19:32
DateTimeDigitized	2004:11:06 17:19:32
DateTimeOriginal	2004:11:06 17:19:32
Exposure comp.	0.00EV
Flash	Did not fire
Focal Length	14.2mm
FocalPlaneXResolution	2442
FocalPlaneYResolution	2442
Make	FUJIFILM
MaxApertureValue	3
Model	FinePix2800ZOOM

MITRE

© 2005, The MITRE Corporation