

Outsmart traffic. Go511.com **511**
Free traffic updates and transit planning

Username
Password
 Remember Me

PoynterOnline.

Search Poynter Online



NEWS & TIPS TRAINING CHATS

Romenesko Reporting Ethics & Leadership Visual Online & TV & Journalism
Latest News & Writing Diversity & Business Journalism Technology Radio Education

Top Story



Ana Marie Cox, Fark.com's Drew Curtis Speak at Poynter's Newseum Event

E-Media Tidbits

Home > Online & Technology > E-Media Tidbits

Tools: Text Size or , Print , RSS , Subscribe via e-mail



Matt Baume

A group weblog about the intersection of news & technology

TIDBITS CONTRIBUTORS

- Alan Abbey
- Dorian Benkoil
- Paul Bradshaw
- Matthew Buckland
- Juan C. Camus
- Michelle Ferrier
- Amy Gahrn
- A. Adam Glenn
- Rich Gordon
- Tish Grier
- Etan Horowitz
- David Herrold
- Barb Iverson
- David Johnson
- Damon Kiesow
- Steve Klein
- Maryn McKenna
- Michelle Minkoff
- Steve Outing
- Kim Pearson
- Ernst Poulsen
- Ken Sands
- Maureen Skowran
- Will Sullivan
- Megan Taylor
- Fons Tuinstra
- Peter M. Zollman

How the Semantic Web Can Connect News and Make Stories More Accessible

Posted by [Matt Baume](#) at 3:23 PM on Jun. 25, 2010

Tom Tague isn't content to let an article just be an article. "How do I take a chunk of text," he asked, "and turn it into a chunk of data?"

He was speaking Thursday night at a [panel discussion hosted by Hacks/Hackers](#), a San Francisco-based group that bridges the worlds of journalism and engineering. Coinciding with the [2010 Semantic Technology Conference](#), Thursday's presentation dealt with the Web's evolution from a tangle of text to a database capable of understanding its own content.

[Tague](#), vice president for platform strategy with Thompson Reuters, was joined by New York Times Semantic Technologist [Evan Sandhaus](#), allVoices CEO [Amra Tareen](#), and Read It Later creator [Nate Weiner](#). The semantic Web is already here, they explained; and it's getting smarter.

Make news worth more

Simply put, the semantic Web is a strategy for enabling communication between independent databases on the Web.

For example, Sandhaus said, there's a wealth of priceless data in databases at Amazon, the Environmental Protection Agency, the Census Bureau, Twitter and Wikipedia. "But they don't know anything about one another," he said, so there's no way to answer questions like, "What is the impact of pollution on population?" or "What do people tweet about on smoggy days?" (Sandhaus said he did not do his presentation as a representative of the Times.)

This is a particular problem for news publishers, said Tague. Publishers need to monetize content, engage with users and launch new products; since news articles lie in a "sweet spot" between fleeting tweets and durable scientific journals, they have the most potential to grab and retain readers.

In other words, it's possible for publishers to improve the value and shelf life of news. All that's required is rich metadata.

Metadata, Tague said, improves reader engagement by linking together related media. For readers, that means more context on each story and a more personalized experience. And for advertisers, it means better demographic data than ever before.

But there's a problem: Currently, the economics of online

Most Recent Articles

1. Chicago Reader publisher explains editor's firing
5:01 PM Jun. 29, 2010
2. Ana Marie Cox, Fark.com's Drew Curtis Speak at Poynter's Newseum Event
4:36 PM Jun. 29, 2010
3. Law firm considering class action suit against Apple
3:01 PM Jun. 29, 2010
4. Finding Opportunities in Sports Journalism
12:01 AM Jun. 29, 2010
5. Survey Finds UV Protection Not Always a Priority for People Buying Sunglasses
12:01 AM Jun. 29, 2010

More Recent Articles

- Most E-mailed
- Recent Comments
- Recent Tags
- Community Activity

Poynter Training.



Use our content for free.
NEWS21.com

Poynter Careers.

Latest jobs (more)

- Regional Editor South Associated Press
- Looking for News Researcher? News Research
- Night News Editor The Day Publishing Company
- Managing Editor The Cleveland Jewish News
- James H. Ottaway Sr. Endowed Professorship SUNY New Paltz

- [See All Jobs](#)
- [Add Your Resume](#)
- [Post Your Job](#)
- [Become a Member](#)

More media jobs

M.A. in Journalism - Regent University - Converged Media Program-- Apply Here!
Online/On Campus.



Global Expert Finder

Analysis on demand

A free resource of Global Experts



Al Tompkins, Broadcast and
Online Group Leader

Poynter Seminars

Small, in-person training
experiences.

1. Narrative Writing On
Deadline
Apply NOW
2. Pants on Fire: Sorting
Out the Facts in
Florida's 2010
Elections, July 12
workshop
Apply NOW
3. New Tools (and They're
Free!) for Improving
Your Government
Coverage
Apply NOW

All Poynter Seminars

News University

Today's most popular
courses on NewsU, Poynter's
e-learning site for
journalists.

1. The Be A Reporter
Game
2. Best of
Photojournalism: What
Makes a Winner
3. The Lead Lab

All NewsU Courses

Webinars

Our online classroom is just
a click away. Learn more.

All Webinars

news doesn't support the manual creation of metadata.

Let algorithms curate

Tague's solution to the Internet's overwhelming volume of news is [OpenCalais](#), a Thomson Reuters tool that can examine any news article, understand what it's about, and connect it to related media.

This is more than a simple keyword search. OpenCalais extracts "named entities," analyzing sentence structure to determine the topic of the article. It is able to understand facts and events. For example, when fed [a short article](#) about a hurricane forming near Mexico, [an OpenCalais demo tool](#) recognized locations like Acapulco, facilities like The National Hurricane Center and an even occupations like "hurricane specialist." It also understood facts, synthesizing a subject-verb-object phrase to express that a hurricane center had predicted a hurricane.

OpenCalais has already been put to work at a wide range of news organizations, including The Nation, The New Republic, Slate, and Aljazeera. Each site's implementation is unique; for example, [DailyMe](#) uses semantic data to monitor each user's reading habits, presenting the user with personalized reading suggestions.

Both The Nation and The New Republic saw immediate benefits to the use of OpenCalais, Tague said; the tool coincided with significant gains in time-on-site, and it automatically generates pages dedicated to a single topic, which had been a labor-intensive process for editors.

Overcome overwhelming content

As OpenCalais frees editors from the minutiae of searching for complementary stories, Nate Weiner's software facilitates the gathering of reading material. [Read It Later](#) integrates with browsers and RSS readers; when users see something that they want to read later, they simply flag the page and the application gathers it for later consumption.

Unfortunately, users can sometimes wind up with an overwhelming, disorganized collection of articles. So Weiner decided to teach the application how to group similar items, making them easier to skim and select.

Initial experiments with manual tagging didn't work out, since users weren't interested in taking the time to add tags to every article they collected. So Weiner turned to semantic applications that could automatically analyze each article and organize related topics. His tool of choice: OpenCalais, which turned Read It Later's "Digest" view from an unwieldy list into a magazine-like layout.

Organize the organizing

Sandhaus described the alchemy of the semantic Web as "graphs of triples," which drew furrowed brows from his audience. But it turned out not to be as complicated as it sounds; the "triples" are just simple subject-verb-object sentences, chained together. For example, if a tool detects "Barack Obama" in an article, it will scan nearby words to create a relationship like "Barack Obama is the President." Then it can build on its knowledge of "the President" to branch further out: "The President lives in the White House," "The White House was burned in 1814," and so on.

These relationships are derived from massive databases that grow larger and larger by the day. For example, [DBpedia](#) has turned Wikipedia into a database of 2.6

Quick
reactions on
breaking news

Diverse
voices
on complex
issues

Office Depot

Search Amazon.com:

All Products

Keywords:

Search New and Used
amazon.com.

million entities; [Freebase](#) is a database of databases with 11 million topics; [GeoNames](#) tracks 8 million place names, and [MusicBrainz](#) can recognize 9 million songs.

But the real magic happens when the databases come together, such as when the BBC wanted to create [a comprehensive resource for information about bands](#). By merging its own information with entries from Wikipedia and MusicBrainz, the BBC created a website that seems to know everything about music.

Trust algorithms, but trust humans more

As smart as the semantic Web can be, it's still not as smart as a human editor. "Our algorithms can never be perfect," said allVoices CEO Amra Tareen. Her company provides citizen journalists with their own news platform, incentivizing high-quality reporting with payments based on page views.

Since its launch in 2008, allVoices has scanned articles to generate what Tareen called a "bag of words" that connects each story to complementary reporting. Depending on a reporter's algorithmically calculated reputation and users' engagement with the story, the story can work its way up from a local section to national or even global focus on the site.

Tareen estimates that the curating of news on the site is about 20 percent human and 80 percent algorithmic.

Expect to see more semantic Web tools

Expect to see more semantic Web technology -- lots more, and soon. "There's growing momentum in this space," said Sandhaus, gesturing to a slide showing exponential growth of connected databases. "The more that you put yourself out there and people point back to you, the easier you are to find."

Fortunately for journalists, the semantic Web will work for humans, not the other way around. "We don't want to get in the way of the journalistic process," said OpenCalais' Tague. That's welcome news to any reporter who has been frustrated by a clunky content management system, a labyrinthine tagging and categorization system or manual photo management.

Semantic Web developers' goal, Tague said, is to free journalists to report, rather than sentencing them to generate endless metadata for the sake of SEO. "I hate the idea of journalists writing for searchability," he said. "That's a problem we should solve on the tech side."

Weiner of Read It Later agreed. Speaking on behalf of developers, he advised journalists, "Keep doing what you're doing. We'll try to adapt."

RELATED

NewsU Webinar Replay

[Discover Your Missing Links: The Semantic Web](#)

Tags: [Online databases](#), [Online traffic and metrics](#)

Tools:

[Comment](#) [e-mail](#) [Permalink](#) [Share](#) [BOOKMARK](#) [f](#) [t](#) [e](#) ...

Poynter. Aug. 23-26 | Oct. 25-28
Critical Tools for the Non-Traditional Journalist **Apply Today**
With Poynter's Kelly McBride

Latest Poynter Blogs (See All Blogs)

Romenesko

- Chicago Reader publisher explains editor's firing

AI's Morning Meeting

- Survey Finds UV Protection Not Always a Priority for People Buying Sunglasses

E-Media Tidbits

- How the Semantic Web Can Connect News and Make Stories More Accessible

Mobile Media

- Law firm considering class action suit against Apple

The Biz Blog

- MediaNews, Journalism Online Set to Start Charging for Content

NewsPay

- Why McClatchy, Others Use PlaceLocal to Help Advertisers Reach Small Audiences

Transformation Tracker

- Tracking the Future of Nonprofit Journalism

Ask the Recruiter

- Finding Opportunities in Sports Journalism

SuperVision

- What Great Bosses Know about the 5 Daily Realities They Must Face

Shop

The Kennedys: America's Front Page Family

50 years of historic newspapers with captions
\$16.99



About Poynter

Who We Are & What We Do

History and mission

Where is Poynter?

The Institute's location

Faculty & Staff Listings

Contact information

Poynter on the Record

Faculty in the news

Resource Center

Tips & Bibliographies

Give to Poynter

Invest in Journalism

Your gifts support Poynter's teaching and provide scholarships.

Advertise

You aim, we deliver

Reach thousands of journalists with your message on Poynter Online.

[RSS](#) | [Podcasts](#) | [Mobile](#) | [Twitter](#) | [Facebook](#) | [Contact](#) | [FAQ](#)
[Guidelines](#) | [Corrections](#) | [Privacy](#) | [Site Map](#) | [Press](#) | [Advertise](#)

© 1995-2010 The Poynter Institute

801 Third Street South | St. Petersburg, FL 33701

Phone (888) 769-6837 | Fax (727) 553-4680