

IPRO 327

A Geographically Searchable News Aggregator

SPRING 2009

Team

- Ori Rawlings (CS)
- Dan Copeland (PS)
- Evan Estola (CS)
- Jay Mundrawala (CPE)
- Daniel Price (CPE/CS)
- Daniel Sirotzke (CS)
- Max Kaim (Psych/CS)
- Laura Rodriguez (CPE)
- Steven Peterson (ITM)

Objective

- Create a news search engine that understands the content of articles
 - Allows search by “meaning” versus by “string matching”
 - Example: What happened in Chicago in 1996?

Why News Articles?

- Contains reasonably factual content
- Articles printed regularly
- Articles have relatively standard format



Your Search

January 1, 1996

Search

Today

Past 7 Days

Past 30 Days

Past 12 Months

All Results Since 1851

All Results | Articles | Multimedia

1-10 of 10,000+ Results

Sponsored Links

Sponsored Links

[January 1996](#) January 1996 Newspapers In The World's Largest Newspaper Archive!
www.NewspaperArchive.com/1990's

[\\$1402. Nikon 1996](#) 80-400mm f/4.5-5.6D Zoom Nikkor Eligible Units Ship Free Tax Varies
amazon.com

[I-96](#)

Up to the Minute Detroit Traffic Reports to Your Email & Cell Phone
www.Traffic.com/Detroit

[July 1 2006](#)

Find Info On Hazardous Substances Restrictions at Business.Com
www.business.com

[January 1st](#)

Get the Facts About January 1st.
WiseTo.com/January+1st

[Review Listings - Theater](#)

Spoke Man (March 4, 1996). State Fair (March 28, 1996). The Three Lives Of Lucie Chabrol (August 9, 1996). True Crimes (January 1, 1996) ...

[CALENDAR:When to Park Where in 1996](#)

Published: **January 1, 1996**. Listed below are the days on which parking regulations will be suspended this year in New York City. Street cleaning regulations ...

January 1, 1996

[Review Listings - Theater](#)

Animations (Play Trilogy) (January 30, 1996) ... Dancing On Her Knees (March 1, 1996). Dark Rapture (May 24, 1996). Dark Ride (January 20, 1996) ...



Your Search

January 1, 1996

Search

Today

Past 7 Days

Past 30 Days

Past 12 Months

All Results Since 1851

All Results | Articles | Multimedia

1-10 of 10,000+ Results

Sponsored Links

Sponsored Links

[January 1996](#) January 1996 Newspapers In The World's Largest Newspaper Archive!
www.NewspaperArchive.com/1990's

[\\$1402. Nikon 1996](#) 80-400mm f/4.5-5.6D Zoom Nikkor Eligible Units Ship Free Tax Varies
amazon.com

[I-96](#)

Up to the Minute Detroit Traffic Reports to Your Email & Cell Phone
www.Traffic.com/Detroit

[July 1 2006](#)

Find Info On Hazardous Substances Restrictions at Business.Com
www.business.com

[January 1st](#)

Get the Facts About January 1st.
WiseTo.com/January+1st

[Review Listings - Theater](#)

Spoke Man (March 4, 1996). State Fair (March 28, 1996). The Three Lives Of Lucie Chabrol (August 9, 1996). True Crimes (January 1, 1996) ...

[CALENDAR:When to Park Where in 1996](#)

Published: **January 1, 1996**. Listed below are the days on which parking regulations will be suspended this year in New York City. Street cleaning regulations ...

January 1, 1996

[Review Listings - Theater](#)

Animations (Play Trilogy) (January 30, 1996) ... Dancing On Her Knees (March 1, 1996). Dark Rapture (May 24, 1996). Dark Ride (January 20, 1996) ...



Your Search

January 1, 1996

Search

Today

Past 7 Days

Past 30 Days

Past 12 Months

All Results Since 1851

All Results

Articles

Multimedia

1-10 of 10,000+ Results

Sponsored Links

Sponsored Links

[January 1996](#) January 1996 Newspapers In The World's Largest Newspaper Archive!
www.NewspaperArchive.com/1990's

[\\$1402. Nikon 1996](#) 80-400mm f/4.5-5.6D Zoom Nikkor Eligible Units Ship Free Tax Varies
amazon.com

[I-96](#)

Up to the Minute Detroit Traffic Reports to Your Email & Cell Phone
www.Traffic.com/Detroit

[July 1 2006](#)

Find Info On Hazardous Substances Restrictions at Business.Com
www.business.com

[January 1st](#)

Get the Facts About January 1st.
WiseTo.com/January+1st

[Review Listings - Theater](#)

Spoke Man (March 4, 1996). State Fair (March 28, 1996). The Three Lives Of Lucie Chabrol (August 9, 1996). True Crimes (January 1, 1996) ...

[CALENDAR:When to Park Where in 1996](#)

Published: **January 1, 1996**. Listed below are the days on which parking regulations will be suspended this year in New York City. Street cleaning regulations ...

January 1, 1996

[Review Listings - Theater](#)

Animations (Play Trilogy) (January 30, 1996) ... Dancing On Her Knees (March 1, 1996). Dark Rapture (May 24, 1996). Dark Ride (January 20, 1996) ...

[Review Listings - Theater](#)

Animations (Play Trilogy) (**January 30, 1996**) ... Dancing On Her Knees (March **1, 1996**). Dark Rapture (May 24, **1996**). Dark Ride (**January 20, 1996**) ...

Entity

- Real World Object
 - Person/Organization
 - Place
 - Date

Old Swyne Interface

SwyneProject

It is munchin' all your news articles...



Latitude Longitude Radius Start End
(mi) Date Date

[Tutorial](#) | [IPRO 327](#)

Old Swyne Interface

SwyneProject

It is munchin' all your news articles...



Latitude

Longitude

Radius
(mi)

Start
Date

End
Date

Submit

[Tutorial](#) | [IPRO 327](#)

Old Swyne Interface

SwyneProject

It is munchin' all your news articles...



Latitude

Longitude

Radius
(mi)

Start
Date

End
Date

Submit

[Tutorial](#) | [IPRO 327](#)

Intended Users

- Researchers
- Journalists
- Intelligence agents
- Real estate agents
- Marketers

Group Goals

- Experience and Leadership
- Research and Communication
- Open-ended Problem Solving

Group Goals Approach

- Research of relevant scientific articles
- Written critiques of scientific articles
- Formal presentation of research results
- 360 degree feedback
- Wiki-based project planning and time management

System Goals

- Support multiple news sources
- Improve entity identification
- Improve user experience

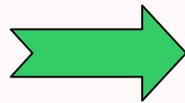
Approach

- Divide system goals into three groups
 - Assign a team to each group of goals
 - Crawling
 - Server
 - Usability

Teams

- Crawling
 - Supply system with news content
- Server
 - Organize and manage data
- Usability
 - Present content to user

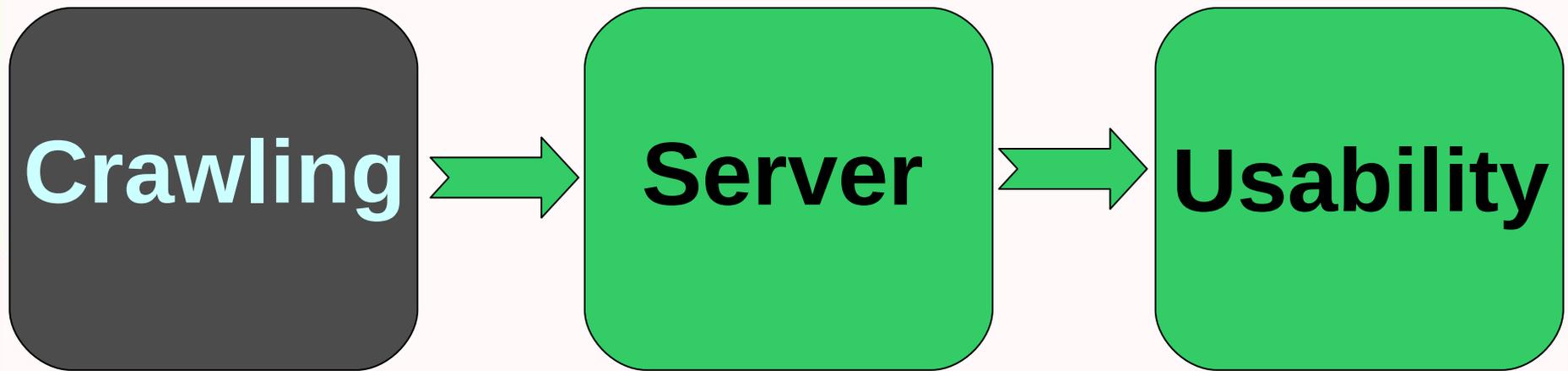
Crawling



Server



Usability



Crawler

- Goals
 - Up-to-date news in system
 - Automatic discovery of new articles
 - Obtain articles from many locations

Approach

- Utilize RSS and Atom Feeds



Site Services

- » Help
- » Contact Us
- » RSS
- » News in PDF
- » Archives

News/Opinion

- California | Local
- National
- World
- Business
- Sports
- Washington
- Science
- Environment
- Opinion

Arts/Entertainment

- Entertainment
- The Guide
- Company Town
- Arts & Culture
- Calendar
- The Envelope
- TV Listings
- Movie Showtimes

Living

- Travel

RSS

What is RSS?

The acronym RSS stands for Really Simple Syndication or Rich Site Summary. An RSS feed is a syndicated news feed in an XML format. Through a news feed reader, you can subscribe to various RSS feeds to get customizable news feeds delivered directly to your desktop. Certain feeds are also available via [Twitter](#).

[Click here for more information about RSS feeds.](#)

NEWS
Top News
Most E-mailed Stories
California Local News
California Politics
Los Angeles
Inland Empire
Orange County
San Fernando Valley
California's War Dead
Data Desk
National
Campaign '08
Business
Personal Finance
Education
Environment
Religion
Science

ARTS & ENTERTAINMENT
Art Reviews
Entertainment News
The Envelope
Movie News
Movie Reviews
Music News
Music Reviews
Restaurant Reviews
Theater & Dance Reviews
TV News
TV Reviews

SPORTS
Sports
Auto Racing
Baseball
Angels
Baseball

LIVING
Autos
Books
Food
Health
Home & Garden
ICU: Los Angeles Connections
Image
Living Green
Real Estate
Travel
Europe
Hawaii
Las Vegas
Mexico

BLOGS
All The Rage

Show: **10 new items** - [all items](#)

[show details](#)

☆ **Scolding and a stiff sentence for Carona** »

12:03 AM (13 hours ago)

A federal judge lectures the former Orange County sheriff about honesty and hands down a 5 1/2-year prison sentence for witness tampering.

Mike Carona's fall from "America's Sheriff" to convicted felon reached bottom Monday as a federal judge gave Orange County's former top law enforcement officer a half-hour lecture about honesty before sentencing him to 5 1/2 years in prison for attempting to obstruct a grand jury investigation.

◆ [Email this](#) ◆ [Technorati Links](#) ◆ [Add to del.icio.us](#)

☆ [Add star](#)  [Share](#)  [Share with note](#)  [Email](#) [Keep unread](#)  [Add tags](#)

☆ **Attacks intensify in L.A. city attorney race** »

Apr 27, 2009 (14 hours ago)

New ads by Councilman Jack Weiss accuse attorney Carmen Trutanich of defending environmental polluters. They cost Weiss the backing of L.A. County Sheriff Lee Baca.

With nursery rhyme characters, giant inflatable rats and "high noon" debate challenges, it's getting hard to quibble with those who complain that the tone of the Los Angeles city attorney's race has gotten downright sophomoric.

◆ [Email this](#) ◆ [Technorati Links](#) ◆ [Add to del.icio.us](#)

☆ [Add star](#)  [Share](#)  [Share with note](#)  [Email](#) [Keep unread](#)  [Add tags](#)

☆ **Tough work, if you can get it** »

Apr 27, 2009 (14 hours ago)

The pay is lower and the jobs are fewer, but the pool of laborers looking for work at sanctioned locations grows larger.

Challenges

- Telling system to listen to additional feeds
- Extracting article content from web pages

Accomplishments

- Fully functioning crawler
- Administrative Interface
- Currently Supports:
 - Boston Globe
 - Los Angeles Times
 - Houston Chronicle
 - New York Times
- 1449 Articles Indexed in 3 days



Server

- Goals:

- Enhance identification of entities in news articles
- Improve browsing of relationships between entities

Entity Identification Example

- Disambiguation
 - A name might refer to two different entities
 - George Bush
 - George W. Bush
 - George H.W. Bush

Entity Browsing

- Link news articles by common entities
- Allow users to follow links

Approach

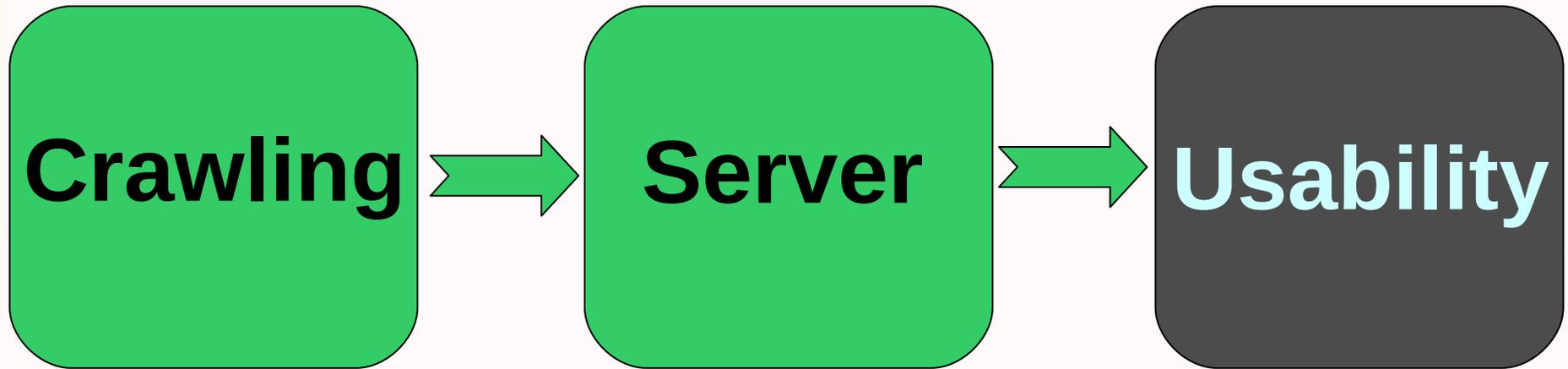
- Entity Identification
 - Extract semantic knowledge from Wikipedia
- Entity Browsing
 - Develop custom interface between website and semantic database

Challenges

- Assigning unique identifiers to entities that refer to the same real world object
- Storing and accessing entity data effectively

Accomplishments

- Entity extraction and disambiguation
- Entity search
- Entity browsing interface
- Over 14,000 entities in database



Usability

Goals:

- Streamline last semester's interface
- Redesign the new interface to be more:
 - User-friendly
 - Simple
 - Accessible
- Create a team brand

Approach

- Research other search engine layouts
- Perform regular usability testing
- Revised prototype based on recommendations

Challenges

- Learning a new technology
- Striking a balance between functionality and simplicity
- Choosing the appropriate test subjects

Accomplishments

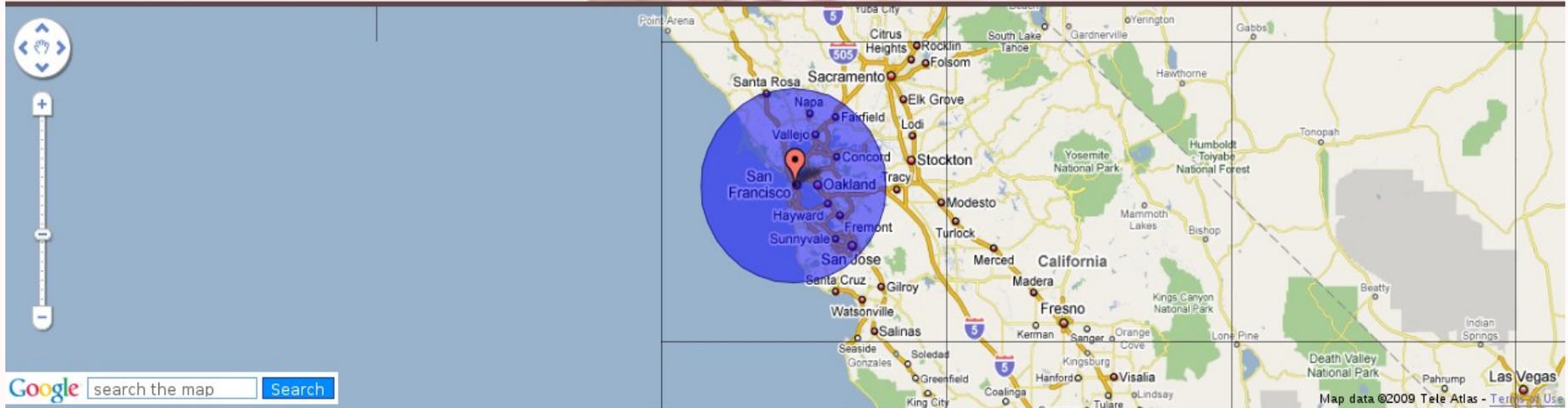
- Unified Look-and-Feel of entire project
 - Implemented color scheme and layout
- Executed 5 test/revision cycles
 - 14 unique test subjects
- Added User Interface Functionality
- Gathered valuable feedback for future work

Group Accomplishments

- Read and prepared presentations on over 20 articles
- Almost 1,000 man hours logged

System Accomplishments

SwyneProject



[Tutorial](#)

Start Date End Date Radius

Governor

copyright © IPRO 327

Interface Inputs

Start Date

End Date

Radius

Entity Disambiguation

SwyneProject

37.75681843152008 -122.442626953125 50 04/01/2009 04/30/2009 Governor

Results for: Governor

Governor

A governor is a governing official, usually the executive (at least nominally, to different degrees also politically and administratively) of a non-sovereign level of government, ranking under the head of state.

Eliot Spitzer

Eliot Laurence Spitzer (born June 10, 1959) is an American lawyer and former politician of the Democratic Party.

David Paterson

David Alexander Paterson (born May 20, 1954) is an American politician and the current Governor of New York.

Sarah Palin

Sarah Louise Heath Palin (; born February 11, 1964) is the Governor of Alaska and the Republican Party's vice-presidential nominee for the presidential election of 2008 in the United States of America.

Arnold Schwarzenegger

Arnold Alois Schwarzenegger (German ; born July 30, 1947) is an Austrian-American bodybuilder, actor, businessman, and politician, currently serving as the 38th Governor of the state of California.

Rick Perry

James Richard "Rick" Perry (born March 4, 1950) is a Republican politician and the current Governor of Texas.

Entity Disambiguation

SwyneProject

37.75681843152008 -122.442626953125 50 04/01/2009 04/30/2009 Governor

Results for: Governor

Governor

A governor is a governing official, usually the executive (at least nominally, to different degrees also politically and administratively) of a non-sovereign level of government, ranking under the head of state.

Eliot Spitzer

Eliot Laurence Spitzer (born June 10, 1959) is an American lawyer and former politician of the Democratic Party.

David Paterson

David Alexander Paterson (born May 20, 1954) is an American politician and the current Governor of New York.

Sarah Palin

Sarah Louise Heath Palin (; born February 11, 1964) is the Governor of Alaska and the Republican Party's vice-presidential nominee for the presidential election of 2008 in the United States of America.

Arnold Schwarzenegger

Arnold Alois Schwarzenegger (German ; born July 30, 1947) is an Austrian-American bodybuilder, actor, businessman, and politician, currently serving as the 38th Governor of the state of California.

Rick Perry

James Richard "Rick" Perry (born March 4, 1950) is a Republican politician and the current Governor of Texas.

News Results



SwyneProject

37.75681843152008 -122.442626953125 50 04/01/2009 04/30/2009

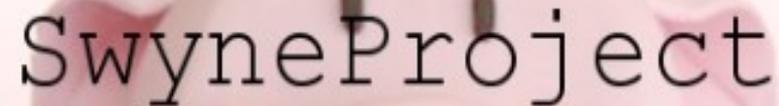
Results

[Newsom officially enters governor's race -- and Twitters the announcement](#)

LATimes 2009-04-22

San Francisco, California, Arnold Schwarzenegger, Jerry Brown, Meg Whitman

News Results



SwyneProject

37.75681843152008 -122.442626953125 50 04/01/2009 04/30/2009

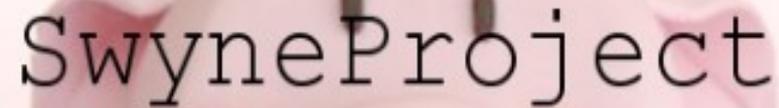
Results

Newsom officially enters governor's race -- and Twitters the announcement

LATimes 2009-04-22

San Francisco, California, Arnold Schwarzenegger, Jerry Brown, Meg Whitman

News Results



SwyneProject

37.75681843152008 -122.442626953125 50 04/01/2009 04/30/2009

Results

Newsom officially enters governor's race -- and Twitters the announcement

LATimes 2009-04-22

San Francisco, California, Arnold Schwarzenegger, Jerry Brown, Meg Whitman

Article Page

SwyneProject

Newsom officially enters governor's race -- and Twitters the announcement

LATimes 2009-04-22

San Francisco Mayor Gavin Newsom made official Tuesday what Californians who have watched him traverse the state for months already suspected -- he is running for governor in 2010.

As might be expected for a campaign that has tried to maximize its presence on social networking websites, Newsom made the announcement electronically: "Today, I am announcing, via Tweet, my candidacy for governor," the Democrat's website declared.

"It's official -- running for Gov of CA. Wanted you to be the first to know. Need your help," he said on his Twitter feed, the real-time microblog that limits missives to 140 characters.

The San Francisco mayor's entrance into the race came as the only other official Democratic candidate, Lt. Gov. John Garamendi, prepared to switch to a race for Congress. According to sources close to him, Garamendi is expected to announce shortly that he will seek the seat to be vacated by Rep. Ellen O. Tauscher (D-Alamo) after she takes a State Department post later this year. His race for governor has been crippled by fundraising difficulties.

Entities found:

Barack Obama

Democratic Party (United States)

Mayor

United States Republican Party

San Francisco, California

Twitter

John Garamendi

Arnold Schwarzenegger

Article Page

SwyneProject

Newsom officially enters governor's race -- and Twitters the announcement

LATimes 2009-04-22

San Francisco Mayor Gavin Newsom made official Tuesday what Californians who have watched him traverse the state for months already suspected -- he is running for governor in 2010.

As might be expected for a campaign that has tried to maximize its presence on social networking websites, Newsom made the announcement electronically: "Today, I am announcing, via Tweet, my candidacy for governor," the Democrat's website declared.

"It's official -- running for Gov of CA. Wanted you to be the first to know. Need your help," he said on his Twitter feed, the real-time microblog that limits missives to 140 characters.

The San Francisco mayor's entrance into the race came as the only other official Democratic candidate, Lt. Gov. John Garamendi, prepared to switch to a race for Congress. According to sources close to him, Garamendi is expected to announce shortly that he will seek the seat to be vacated by Rep. Ellen O. Tauscher (D-Alamo) after she takes a State Department post later this year. His race for governor has been crippled by fundraising difficulties.

Entities

found:

Barack
Obama

Democratic
Party (United
States)

Mayor

United States
Republican
Party

San Francisco,
California

Twitter

John
Garamendi

Arnold
Schwarzenegger

Article Page

SwyneProject

Newsom officially enters governor's race -- and Twitters the announcement

LATimes 2009-04-22

San Francisco Mayor Gavin Newsom made official Tuesday what Californians who have watched him traverse the state for months already suspected -- he is running for governor in 2010.

As might be expected for a campaign that has tried to maximize its presence on social networking websites, Newsom made the announcement electronically: "Today, I am announcing, via Tweet, my candidacy for governor," the Democrat's website declared.

"It's official -- running for Gov of CA. Wanted you to be the first to know. Need your help," he said on his Twitter feed, the real-time microblog that limits missives to 140 characters.

The San Francisco mayor's entrance into the race came as the only other official Democratic candidate, Lt. Gov. John Garamendi, prepared to switch to a race for Congress. According to sources close to him, Garamendi is expected to announce shortly that he will seek the seat to be vacated by Rep. Ellen O. Tauscher (D-Alamo) after she takes a State Department post later this year. His race for governor has been crippled by fundraising difficulties.

Entities found:

Barack Obama

Democratic Party (United States)

Mayor

United States Republican Party

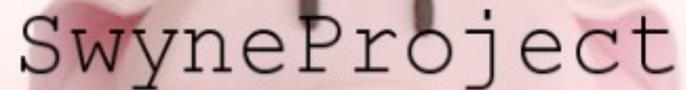
San Francisco, California

Twitter

John Garamendi

Arnold Schwarzenegger

Entity Browsing



SwyneProject

Twitter

website

- [Lawyers Enter Twitter Tempest](#)
- [Newsom officially enters governor's race -- and Twitters the announcement](#)
- [Our Towns: Atwitter About a Mom in Scarsdale](#)

Neighbor Entities found:

as of 2007

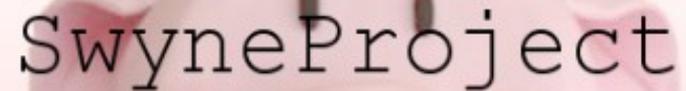
Sarah Palin

Best response

Restaurateur

Daily News
(New York)

Entity Browsing



SwyneProject

Twitter

website

- Lawyers Enter Twitter Tempest
- Newsom officially enters governor's race -- and Twitters the announcement
- Our Towns: Atwitter About a Mom in Scarsdale

Neighbor Entities found:

as of 2007

Sarah Palin

Best response

Restaurateur

Daily News
(New York)

Societal Impact

- Improve access to information
- Inspire future search engines to use semantic data to provide users with specific answers

Management Insights

- Strategic Planning
 - Scheduling
 - Division of Labor
 - Integration of Work
- Teamwork
 - Interpersonal Communication
 - Conflict Resolution

Ethical Considerations (Personal)

- Individual Work Ethic
- Accurate Reporting of Effort
- Respect and Communication

Ethical Considerations (Technical)

- Copyright Infringement
- Fair Use of Software
- User Privacy

Future Work

- Crawler that adapts to new sources
- Optimization for speed
- Expanded usability testing

Questions?

?

Contact Information

Wai Gen Yee: yee@iit.edu