SocialWall: Make Your Semantic Government-related Poster on iPad

Haofen Wang, Xinruo Sun, Xing Niu, Lin Zhou, Zhongyang Fu, Shu Rong, and Yong Yu

Shanghai Jiao Tong University
{whfcarter, xrsun, xingniu, gorden, zhongyang.fu, rongshu, yyu}@apex.sjtu.edu.cn

Abstract. The novel mobile application SocialWall enables its users to discover and share useful information by interactively designing meaningful posters. We think posters are of a good form to present information and discover interesting facts. Thus, we let users to choose what data they wanted to see on the poster and how they shall be visually presented. This kind of user need is represented in a SPARQL query along with a visual template. The integration engine processes the query by fetching and integrating related data from web APIs on-the-fly. Resulting RDF documents are transferred back to mobile end and corresponding poster images are generated, which could be projected on the map, timeline or a poster wall. Users could share their queries and posters on their favorite social sites. Over time, these user-contributed posters will provide interesting views on the mentioned entities. Our demonstration focuses on web APIs about the politics of the United States.

1 Introduction

Recently, governments are making more and more data public accessible. For example, `http://data.gov` publishes over 250,000 data sets from the U.S. government and `http://data.gov.uk` contains more than 2,500 data sets from various UK government departments.

Since data published by governments are raw data containing details derived from too many data sets for a user to understand, some developers make efforts to extract useful data and re-publish it in a user friendly way [7][4][6][3]. Some has build widgets [5] and cool visulizations [8] to help users understand and consume information.

Innovative applications have been developed in competitions [2][1] that aims to seek accessible and comprehensible ways to government data. Some have developed web APIs that could access these data from application servers or even directly from web browsers. Such initiatives include Sunlight Labs[1], OpenCongress.org[2], etc. They share one same goal: to make raw data user-friendly.

[1] http://sunlightlabs.com

[2] http://www.opencongress.org

Experts reorganizes the data the create diagrams out of the data sets that citizens are more familiar with.

We are working on the same direction but in a completely new framework called SocialWall. SocialWall emphasize three aspects:

1. **utilize semantic web technologies**
   We built the application by mashing up open data and APIs. Unlike other sites, we try to reuse data from existing web APIs instead of maintaining a database. User needs are fulfilled by on-the-fly semantic integrating of heterogeneous data sources and SPARQL querying over these data sources.
2. **visualize query and results**
   Instead of providing a few hand-craft interfaces that each presents limited area of data, we let users choose what data aspects they want and how they are visually presented. This kind of user need is represented in a visual query. The engine fetches related data and composes visually attractive posters. Users could interactively modify the layouts or visualization methods to find their best shots. Users can also view the posters as a whole on the map, timeline or even compare two posters.
3. **enable a social environment**
   Users could share their queries and posters on their favorite social sites such as Blogs, Twitter and Facebook. Over time, these user-contributed posters will provide interesting views on the mentioned entities.

We are working on a demonstration that currently focuses on the politics of the United States. There are over 20 APIs available from 7 API providers. The demonstration website(`http://linkedus.org`) builds on Sencha Touch[3] so it works well on mobile devices(especially on iPad!). On PC, Safari or Chrome works fine.

2 Framework

This section will introduce the main objects in Fig. 1.

A visual query is created using Query Builder Interface (Section 4). The interface lets users choose what facets they want to see on the poster. A facet is a unit of data such as an organization's name, a politician's top contributors and total obliged amount of a state in the past ten years, etc. This interface can also generate a corresponding SPARQL query. In the future advanced users may also write SPARQL queries to fulfill complicated data needs.

Given a query, Integration Engine (Section 5) executes the query and returns a list of RDF documents as result. These RDF documents could be given permanent URIs so that users can share or comment on the individual document/poster.

Semantic Poster is an image that visualizes a RDF document. Render Engine (Section 6) is responsible for generating SVG images from the documents

[3] http://www.sencha.com/products/touch/

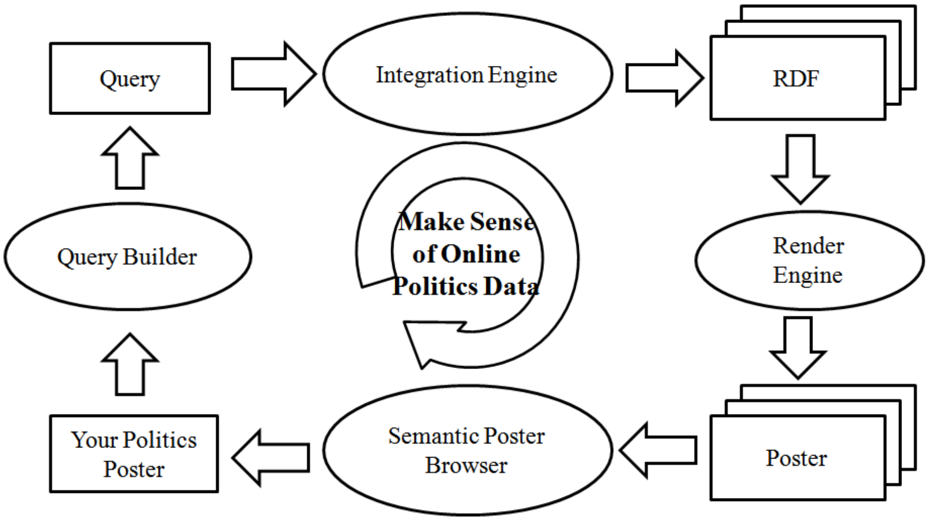


Fig. 1. SocialWall Framework

according to the user's design. The Semantic Poster Browser is used to view the generated posters in different views.

This process is repeated as the user refines his poster.

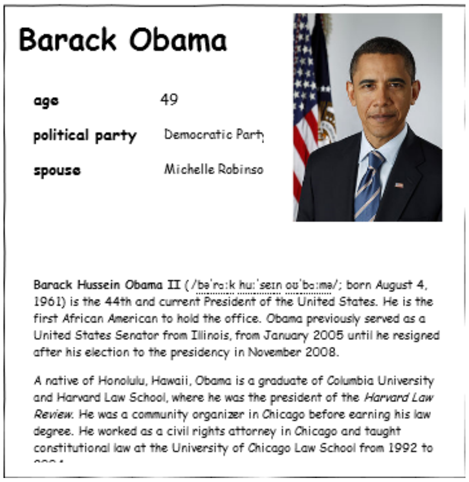
3 Running Examples

Example 1: Show me the President of the United States and his bio.

- The system shows a list of facets
- The user drag the facet 'Democratic candidates' onto the canvas
- The system searches through the Web and generates the posters
- The user scrolls through the list, expands the list into a wall of posters
- The user views them in the map and timeline view
- The user decides to focus on the President
- The user selects the information he is concerned with such as what policies the President is advocating
- The system then generates a poster (like Fig. 2 showing these information
- The user publishes the poster
- His friend sees the poster on Facebook or Flipboard and likes it and writes a comment
- Later, his friend decides to create a similar poster for his favorite politician also on SocialWall

Example 2: Show me recipient candidates of campaign donations whose donor come from the health sector.

Example 3: Show me the total amount of donation coming from each sector to Obama, and compare with McCain.



Barack Obama

age	49
political party	Democratic Party
spouse	Michelle Robinso

Barack Hussein Obama II (/bəˈrɑːk huːˈseɪn oʊˈbɑːmə/; born August 4, 1961) is the 44th and current President of the United States. He is the first African American to hold the office. Obama previously served as a United States Senator from Illinois, from January 2005 until he resigned after his election to the presidency in November 2008.

A native of Honolulu, Hawaii, Obama is a graduate of Columbia University and Harvard Law School, where he was the president of the *Harvard Law Review*. He was a community organizer in Chicago before earning his law degree. He worked as a civil rights attorney in Chicago and taught constitutional law at the University of Chicago Law School from 1992 to

Fig. 2. Sample Poster

4 Query Builder

Query Builder interface (Fig. 3) is composed of three parts: a poster design canvas, a list of available facets and a panel showing result posters.

User designs his poster by dragging the facets that he is interested in onto the canvas. The position of each facet is recorded which is later used to generate the posters from data.

4.1 Semantic Poster Browser

After the posters are generated, the user can view them in the Semantic Poster Browser which has four major views:

- Wall view tiles posters on the screen, like Fig. 4.
- Map view shows posters according to a geographic facet if available
- Timeline view shows posters acccording to a temporal facet if avaiable
- Compare view is a special view that allows the user to compare the information in two posters.

4.2 Poster Viewer

The user can tap on a poster to open the Poster Viewer. This viewer has the following functions in different pages:

- it shows the poster in fullscreen size
- it guides the user to publish the poster to the social web and to the semantic web
- it allows the user to select this poster for later comparing with another poster
- it shows related posters published by others that contains the same entities

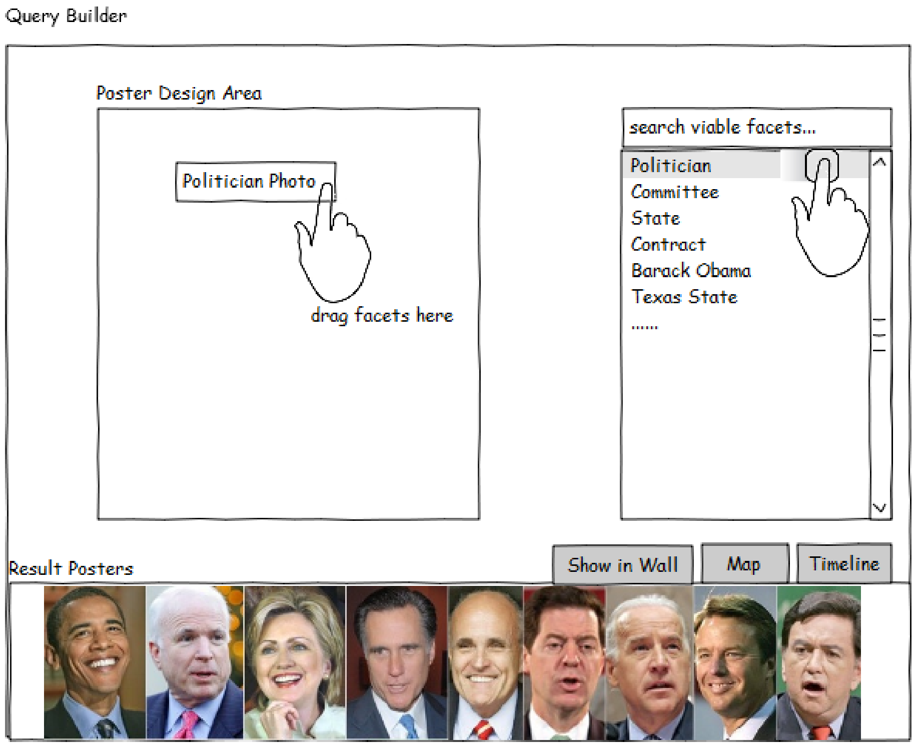


Fig. 3. Query Builder Interface

5 Integration Engine

Integration Engine integrates heterogeneous data sources and tries to answer users' queries on the fly.

The visual query that the engine works on can be represented as a set of facets. The engine's goal is to find RDF documents that satisfy these facets.

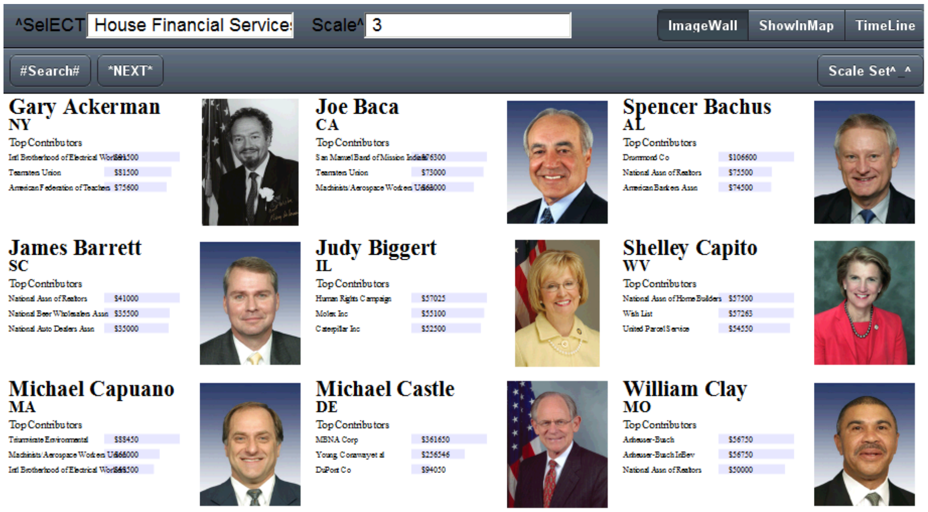
5.1 Facet Graph

We call an interface provided by a data source an API. Each API is represented as a few triples: <API name, input facet, output facet>. For example a RESTful endpoint to get a politician's basic information given his name can be written as: <API1, name, politican_info>.

Facet graph is represented using RDF Schema. The result RDF documents should follow this schema.

5.2 Execution Stage

Four major steps of the engine are:



^SelECT House Financial Service:

Scale^ 3

Gary Ackerman
NY
Top Contributors

Intl Brotherhood of Electrical Wor	$91500
Teamsters Union	$81500
American Federation of Teachers	$75600

Joe Baca
CA
Top Contributors

San Manuel Band of Mission Ind	$76300
Teamsters Union	$73000
Machinists/Aerospace Workers U	$63000

Spencer Bachus
AL
Top Contributors

Drummond Co	$106600
National Assn of Realtors	$75500
American Bankers Assn	$74500

James Barrett
SC
Top Contributors

National Assn of Realtors	$41000
National Beer Wholesalers Assn	$35500
National Auto Dealers Assn	$35000

Judy Biggert
IL
Top Contributors

Human Rights Campaign	$57025
Molex Inc	$55100
Caterpillar Inc	$52500

Shelley Capito
WV
Top Contributors

National Assn of Home Builders	$57500
Wish List	$57263
United Parcel Service	$54550

Michael Capuano
MA
Top Contributors

Triumvirate Environmental	$88450
Machinists/Aerospace Workers U	$66000
Intl Brotherhood of Electrical Wor	$58500

Michael Castle
DE
Top Contributors

MBNA Corp	$361650
Young Conawayet al	$256546
DuPont Co	$94050

William Clay
MO
Top Contributors

Anheuser-Busch	$56750
Anheuser-Busch InBev	$56750
National Assn of Realtors	$50000

Fig. 4. Result View: Poster Wall

1. Query planning and routing
2. Convert the query plan to API requests
3. submit API requests
4. result combination and data reconciliation

For each API, we write adapter programs to fetch its data and convert into an internal data format.

5.3 Data Sources

In this demonstration, the following data sources are used.

- OpenSecrets (CRP)
- Transparency Data
- USASpending
- FollowTheMoney.org
- OpenCongress.org
- Congressional Biographical Directory
- GovTrack

6 Render Engine

Render Engine renders RDF documents into SVG images according to user's poster design. The engine knows the visualization method for each kind of facet, thus given all facets in one document the engine is able to compose a poster by putting each facet on the desired position.

The possible visualization methods are:

- **text** to present simple data types like number, string
- **image** to present facets that have image properties
- **bar chart** to present facets that represent number series
- **pie chart** to present facets that represent distribution

More visualization methods can be easily added into the system by implementing plugins.

7 Publishing

Users could publish their posters to the social web and to the semantic web at the same time.

When published, a poster is assigned a permanent ID. The poster content is fixed as it is published, which means it would not change when the underlying data sources have updated data.

Two URLs based on the permanent ID are generated at the same time. One is for the poster image, which can then be used to publish to blog, Twitter, Facebook, etc. The other is for the RDF document.

Over time, these user-contributed posters will provide interesting views on the mentioned entities.

8 Conclusion

In this paper, we presented a novel mobile application called SocialWall. It enables its users to discover and consume information by interactively designing meaningful posters. When shared to the social web and the semantic web, these will provide interesting views on mentioned entities.

9 Appendix

9.1 Minimal Requirements

The application has to be an end-user application, i.e. an application that provides a practical value to general Web users or, if this is not the case, at least to domain experts. YES. We designed a web application focusing on the politics domain. Any user knowing politics terminology could design their own politics poster.

The information sources used should be under diverse ownership or control. YES. We integrated more than 7 data sources from different providers, see Section 5.3.

The information sources used should be heterogeneous (syntactically, structurally, and semantically). YES. We used RESTful APIs and dumps of these data sources which all have different format, data instance format.

The information sources used should contain substantial quantities of real world data (i.e. not toy examples). YES.

The meaning of data has to play a central role. YES.

Meaning must be represented using Semantic Web technologies. YES. We provide a RDF document for each result posters. Internally, we use an ontology to represent the integration of concepts and data sources.

Data must be manipulated/processed in interesting ways to derive useful information. YES. Not only are the result posters visually attractive, but users could also view them in different views such as map and timeline.

This semantic information processing has to play a central role in achieving things that alternative technologies cannot do as well, or at all. YES. Data source integration using ontology. A federated query engine decompose a semantic query and route to remote APIs.

9.2 Additional Desirable Features

The application provides an attractive and functional Web interface (for human users). YES. Go to the website and see for yourself!

The application should be scalable (in terms of the amount of data used and in terms of distributed components working together). Ideally, the application should use all data that is currently published on the Semantic Web. Not quite scalable

Rigorous evaluations have taken place that demonstrate the benefits of semantic technologies, or validate the results obtained. None at the moment

Novelty, in applying semantic technology to a domain or task that have not been considered before. YES. To the politics domain and we are made it possible for normal users (not programmers) to create semantic posters.

Functionality is different from or goes beyond pure information retrieval. YES. It is search+visual+social

The application has clear commercial potential and/or large existing user base. Maybe. As our application is easy to use for normal users, we expect the user base can be established in a short period of time. We haven't thought about it's business model yet.

Contextual information is used for ratings or rankings. YES. The most common case is that the mobile platform could provide user's location which could be an input to the system.

Multimedia documents are used in some way. YES. The search results themselves are images.

There is a use of dynamic data (e.g. workflows), perhaps in combination with static information. NO

The results should be as accurate as possible (e.g. use a ranking of results according to context). Yes and No. The results have accruate attributes which are utilized to show on different views such as map and timeline. Since normal users could hardly write SPARQL queries, we provide an intuitive inferface to help them create the queries, with some loss of accuracy.

There is support for multiple languages and accessibility on a range of devices. YES. Since we develop it as a web application, it could run on multiple platforms including mobile devices and desktop browsers.

References

1. Apps for californians. http://www.ca.gov/appsforcalifornians/.
2. Design for america. http://sunlightlabs.com/contests/designforamerica/.
3. Follow the money. http://www.followthemoney.org/.
4. OpenCongress - track bills, votes, senators, and representatives in the U.S. congress. http://www.opencongress.org/.
5. Politiwidgets. http://politiwidgets.com/.
6. Search state and federal campaign contribution data - TransparencyData.com. http://transparencydata.com/.
7. USAspending.gov. http://www.usaspending.gov/.
8. Where does my money go? http://www.wheredoesmymoneygo.org/.