GeoTemporal Reasoning in a Web 3.0 World

Jans Aasman, Ph.D.

Franz Inc.
2201 Broadway, Suite 715
Oakland, CA 94612
ja@franz.com

Abstract

We demonstrate a Web 3.0 application that organizes social events for friends in the larger San Francisco Bay Area using real data from available web sources. We use a collection of techniques that will be at the heart of the next wave of applications on the web; entity extraction, querying federated databases, efficient spatial reasoning, temporal reasoning, practical RDFS++ reasoning, reasoning over preferences, social network analytics, activity planning and activity recognition.

Overview

Tim Berners-Lee, James Hendler and Ora Lassila's Scientific American article [1] provides a compelling vision of the Semantic Web. The body of the article covered interesting use cases that showed what the Semantic Web will offer individual users.

Nearly all of the use cases are about helping people to organize their lives better, for example planning a visit to a doctor within the constraints of your working life and family life. In their use cases it is never about one person, but always about the person in a social context of family, circle of friends, work or service providers (like a doctor and his office). All use cases deal with events that happen at some point in time and at some particular place, where the agent must take into consideration that you have to travel between locations and be aware of traffic constraints. Finally, in many use cases it is clear that you have to deal with preferences of people.

We decided to test if the data on the web, and the currently available Semantic Web tools were sufficient to implement a typical use case, as mentioned above, on real scenarios using data available today.

The example we chose was an event-planning-agent for the San Francisco Bay area. The use case in the example (and our demo) is to organize an evening with your friends where you leave it up to the agent to determine a movie and restaurant, and then invite the right people and deal with the fact that they all have to come from different locations and will have different traveling times. Obviously the agent will also have to deal with the preferences that people have about food and movies.

Collected Data

• **Restaurant Data.** We collected restaurant data from Yelp (www.yelp.com) since it is popular and fairly complete for restaurants in the Bay Area. We also used the Yelp classification scheme for type of food, price range, rating, and with the help of Google we can also accurately get the geospatial coordinates.

• **Movie Data.** We scraped part of the Internet Movie Database (www.imdb.com) to get movie titles, actors playing in movies, the rating for movies, the types of movies, reviews, etc.

• **Theater Data.** We used Yahoo's movie service to find all theaters in the Bay Area plus the current releases and their running times. We used the Google API again to find the geospatial coordinates

• **Transit Data.** We contacted the regional transit authority (www.mtc.ca.gov) and obtained transit data for all transportation modes in the greater Bay Area, in this case BART, MUNI, ferries, and local trains.

Simulated Data

Due to privacy issues and data availability challenges we opted to simulate a large social network based on US Census Data (2000). We created a network of 5000 people; including names, frequency statistics, gender, age, work/study, a set of preferences for movies and type of food, a history for restaurants and movies visited, and rela-

tionships between people (school/lab/club mates, family, in love, etc).

Demonstration Scenarios

1. There is a new major cinematic release this weekend and I want to see it with someone. Tell me which of my friends may be interested in this movie, the most convenient theater for our group, and determine a casual inexpensive restaurant location for after the movie. I need to be home by midnight.

2. I want to schedule a date with my girlfriend, but I don't want to drive a car. Tell me the most romantic and recent movie release she may like (if there are several choices, include those I may enjoy), include a quaint but not too expensive restaurant in the plans. The theater must be near the restaurant. I prefer that the transit fee for my girlfriend should be cheaper than mine, she has no wait time, and she can be home by her 11PM curfew.

3. Suddenly today's lecture is cancelled. Several of my classmates want to go see a movie but can't be late for their next lecture. Find the nearest theater available with movies starting soon (but we can be on time) and we can return back to our current location before the next lecture.

If there are several choices, choose one the group will prefer on average.

4. I'd like to plan a dinner with some old school friends.

I'd like to watch a movie with the largest number that have similar movie interests (not specific) before/after dinner, please suggest members, films, and theater nearest to the suggested restaurant. The location should be mutually inconvenient for everyone.

Summary of required reasoning capabilities of the Agent.

It falls outside the scope of this short demo paper to explain in detail the inner workings of the agent. However, here are some things that the agent uses:

1. Geospatial computations. Both within the trip planner and for finding the right pairs of movie theaters and restaurants.

2. Temporal reasoning for all events. In the trip planner, finding the times between having dinner and going to the theater, etc.

3. Social network analysis. The agent needs to compute interest groups based on preferences and types of relationships between people

4. RDFS++ reasoning to reason over types of restaurants, types of movies, etc. RDF Schema is a W3C standard that adds semantics to RDF, essentially providing basic elements for the description of ontologies. The most important construct being that of classes and subclasses.

RDFS++ refers to the fact that we also added more advanced description logic features. We refer to [3] for an extensive description of RDFS++.

5. Federated querying. The restaurant data, the movie data, the theater data, the social networks and the transit data all reside in their own RDF triple stores. All the queries are done over a federated store that encompasses all these triple stores.

6. Trip planning. In this case is a well understood cost based planner that works directly on the RDF store (a type of A* search [2]).

Presenter's Bio

Jans Aasman started his career as an experimental and cognitive psychologist, earning his Ph.D in cognitive science with a detailed model of car driver behavior using Common Lisp and Soar. He has spent most of his professional life in telecommunications research, specializing in intelligent user interfaces and applied artificial intelligence projects. From 1995 to 2004 he was also a part-time professor in the Industrial Design department of the Technical University of Delft. Dr. Aasman is currently the CEO of Franz Inc., the leading supplier of commercial, persistent and scalable RDF database products that provide the storage layer for powerful reasoning and ontology modeling capabilities for Semantic Web applications.

References

[1]   1st Scientific American article on the Semantic Web, May 2001.   http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21&ref=sciam

[2]   Dechter, Rina; Judea Pearl (1985). "Generalized best-first search strategies and the optimality of A*". Journal of the ACM 32 (3): pp. 505536

[3]   Allemang, Dean; Hendler, James (2008). "Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL"