Google Fusion Tables
Web Centric Data Management and Collaboration


Aravinth Bheemaraj
aravinth@cise.ufl.edu
What is the paper about?

- Introduction
- Design Foundations
- Data management with Fusion Tables
- Fusion Tables API
- Related Work
- Conclusion
Overview

- Data Management functionality extension.
  - For broader class of users
  - Wider range of applications.
- How do we achieve the following:
  - Seamless web integration
  - Ease of use
  - Data sharing
- Support for data acquisition, collaboration, visualization and web publishing
Need for new functionality

- How to design data management functionality for today’s connected world?
  - Support collaboration
  - Appealing to less technically skilled users
  - Seamless integration with web
    - Data collection, Presentation and Visualization should be web-compatible
So, what is Google Fusion Tables?

- Cloud-based data management and integration service.
- Focus on both organization and users
  - Organization: Help to make data available online
  - User: Help to collaborate on data management across multiple enterprises.
- Supports tabular data files up to 100mb.
- Data can be geographical objects!
Google Fusion Tables Cont.

- Provide Data Visualization
  - Charts, Maps, Timelines
- Export feature - KML (Google Earth view)
- Supports Filters & Aggregates
- Data Integration
  - Joins across different users tables.
- Sharing – public/restricted sharing/private
KML Mappings
What’s new?

- Discussion feature
  - Collaborators can post/respond to comments

- Interaction with data
  - Web Interface/API based
Sample Applications

- Ecologists collecting specimen of animal and plant life.
- Publishing data about availability and usage of water resources. (Visualization – Intensity based)
- Data distribution about Coffee Exports and Imports.
- Exploring & publishing biking trials on a Map.
- Data based collaboration within dairy farms.
Principles

1. Seamless Integration with the web

- Need?
  - Easy access to internet
  - Entry point into data management
  - Venues for publishing & visualizing data.
- Fusion tables allows users to publish their visualization on the web.
- Make data available as public dataset.
- Integrate seamlessly with established model for documents/spreadsheets.
Principles
2. Emphasize ease of use

- Need?
  - Reach broader class of potential users, who are not techie!
  - Need reduced initial investment.

- Advantages of Fusion tables
  - cloud-based service: requires no initial installation
  - Automatic detection of column types
Principles
3. Incentives for sharing data

- Need?
  - Afraid of misuse/corruption of their data.
  - Not finding the required data easily.

- How can Google Fusion Tables help?
  - Specifying access (Private/Public/Restricted)
  - Make public dataset crawlable.
Principles
4. Facilitate Collaboration

- Need?
  - Require collaboration among multiple parties.
  - Valuable insight from multiple sources.

- How can Google Fusion tables help?
  - Facilitate joining of data.
  - Enable collaborators to discuss/comment on data.
Data Management with Fusion Tables

1. Data Acquisition
2. Data Sharing and collaboration
3. Data Manipulation and Visualization
Data Management

1. Data Acquisition

- Enables uploading files (structured data)
  - Supports CSV, Excel, Open Office, KML
- System detects the header row automatically
  - User just verifies!
- Responsive Import process
  - Background processes are transparent!
- System can guess data types
  - But users can specify if they wish!
Data Management

2. Data Sharing & Collaboration

• Issues:
  • Loss of control over their data once uploaded
  • Losing credit for creating the data
  • Possibility of others using the data incorrectly.

• Solutions:
  • Users can finely control what they share
  • Specify attribution for data.
  • Users can restrict the ability of other users to export outside Fusion Tables.
Data Management

2. Data Sharing & Collaboration (Cont.)

- Search
  - Make public data discoverable by Search engines.
  - Advanced search for tables within Fusion Tables.

- Sharing and Integration
  - Typical model for document sharing in the cloud.
  - Supports Merge operation (key-based Joins)

- Discussions
  - Supports in-depth collaborations
  - Identify outliers, incorrect data to improve the quality
  - Based on particular view of data.
Data Management

3. Data Manipulation & Visualization

- Data exploration using combination of visualization and SQL-like querying
- System suggests visualization type
  - Geographic locations: Maps
  - Date/Time: Timelines/Motion chart
- Create HTML snippet for the visualization and this can be embedded in other sites.
- Rendering of large geographic datasets.
Fusion Tables API

- Extends the functionality of the site using the Fusion Tables API.
- Allows external developers to write apps that use Fusion Tables as a dataset.
- API supports querying of data through SQL like statements.
Related work

- Inspired by ManyEyes (similar data visualization site)
  - Notable visualizations: Word Cloud, Network Diagram
- Online databases (ex: DabbleDB) is close enough
  - Misses collaboration aspect
  - Paid service 😞
- Google public data explorer provides view of government data and visualizations.
  - Excellent Timeline animation
Conclusions

- Enable larger class of users to manage their data.
- Make data management process easier.
- Provide query capabilities.
- Adequate performance for large datasets.
- Priority: Map based visualization and API.
Big questions??

- In what way, Google Fusion Tables differ from previously discussed data management tools?
- Possible extension for unstructured data?
- What other visualization techniques can be more interesting than Maps?
Thank you!