

Text Analytics and Data Access as Services

*A case study in transforming a legacy client-server text analytics workbench and
framework to SOA*

E. Michael Maximilien, Ph.D.

<http://maximilien.org>

IBM Almaden Research Center
San Jose, CA

v0.5.0

About me

- Research Staff Member
 - ▶ Almaden Services Research group
 - ▶ Silicon valley, CA
- 13 years experience in software research, architecture, design, and implementation
- Research interests in SOA, Web APIs, services, mashups, OOD/P, SE, and Agile methods
- Ph.D. in computer science from NC State University, Raleigh



Agenda

- Motivation
- Current solution
- Requirements
- Solution and benefits
- Architecture
- Demo
- Directions
- Acks and references

Motivation

- Information increasingly key asset
- Collaborative data and tooling
 - ▶ Wiki
 - ▶ Web 2.0 community generated data
 - ▶ Ratings and reputation data
- Need to perform text analytics on data set
- Need to incorporate text analytics functions into various business processes

Current solution

- Advantages
 - ▶ First of class algorithms
 - ▶ Tons of successful deployments
 - ▶ Comprehensive
- Disadvantages
 - ▶ Thick-client and monolithic
 - ▶ Single user
 - ▶ Difficult to use (designed for experts)

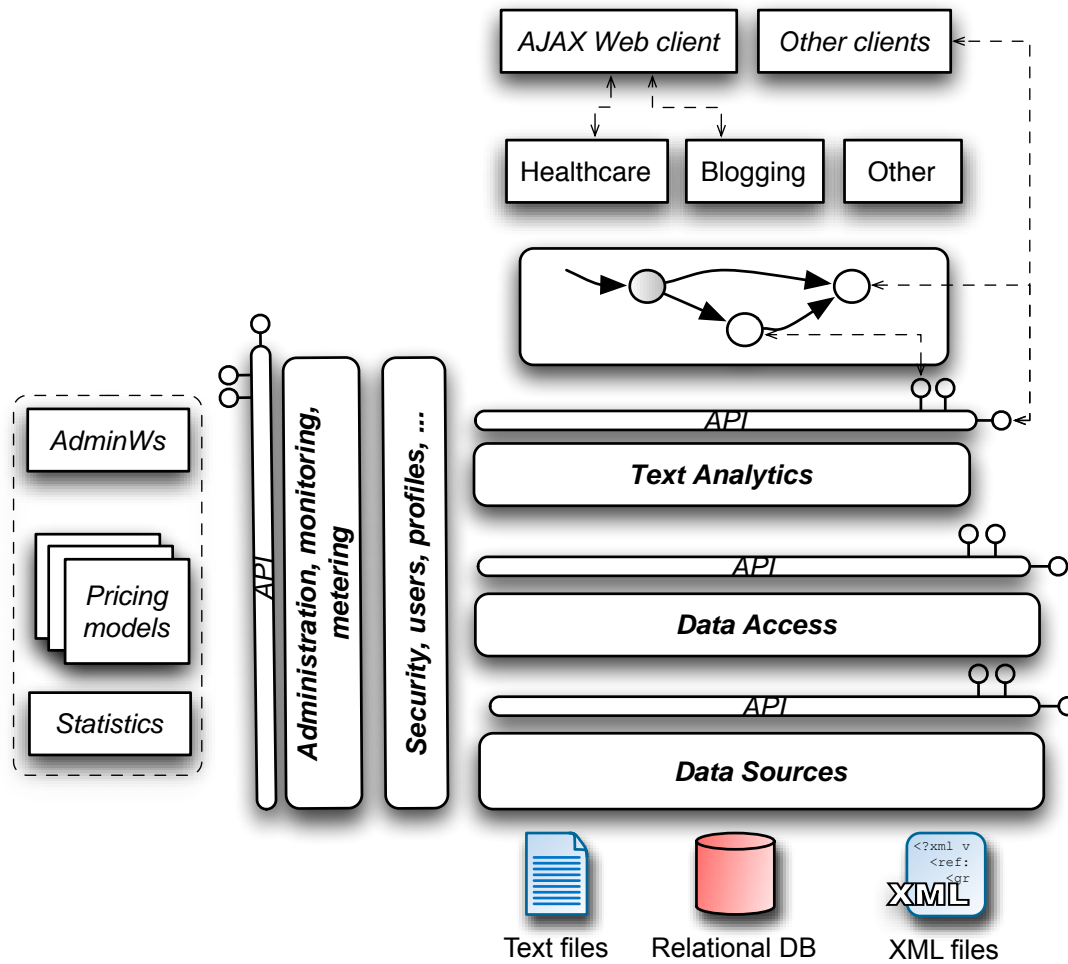
Requirements

- Scalability
- Multi-user
- Fine- and coarse-grained services
- Flexible deployments
- Reusable components
- Leverage old code

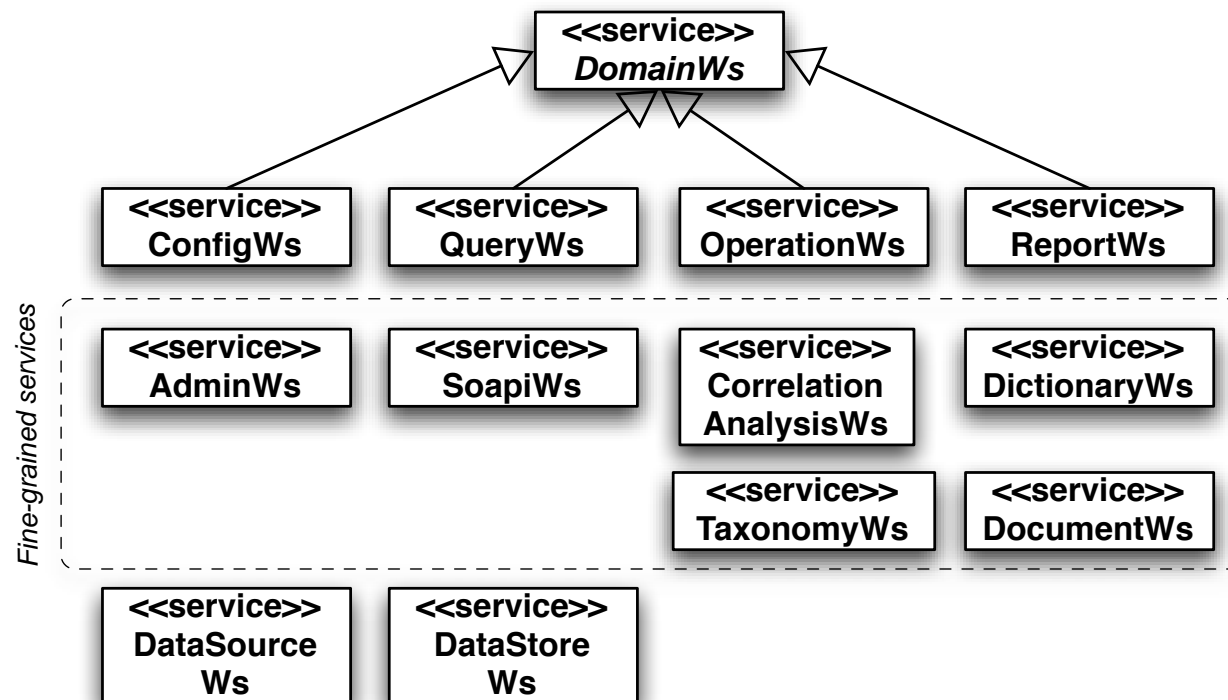
Solution

- Layered architecture
 - ▶ Common layer for non-functional properties, e.g., metering
 - ▶ Data Access Services (DAS) layer
 - ▶ Text Analytics Services (TAS) layer
 - ▶ Domain-specific layers
- Reuse existing code
- Agile “spike” to implement a few complete solutions

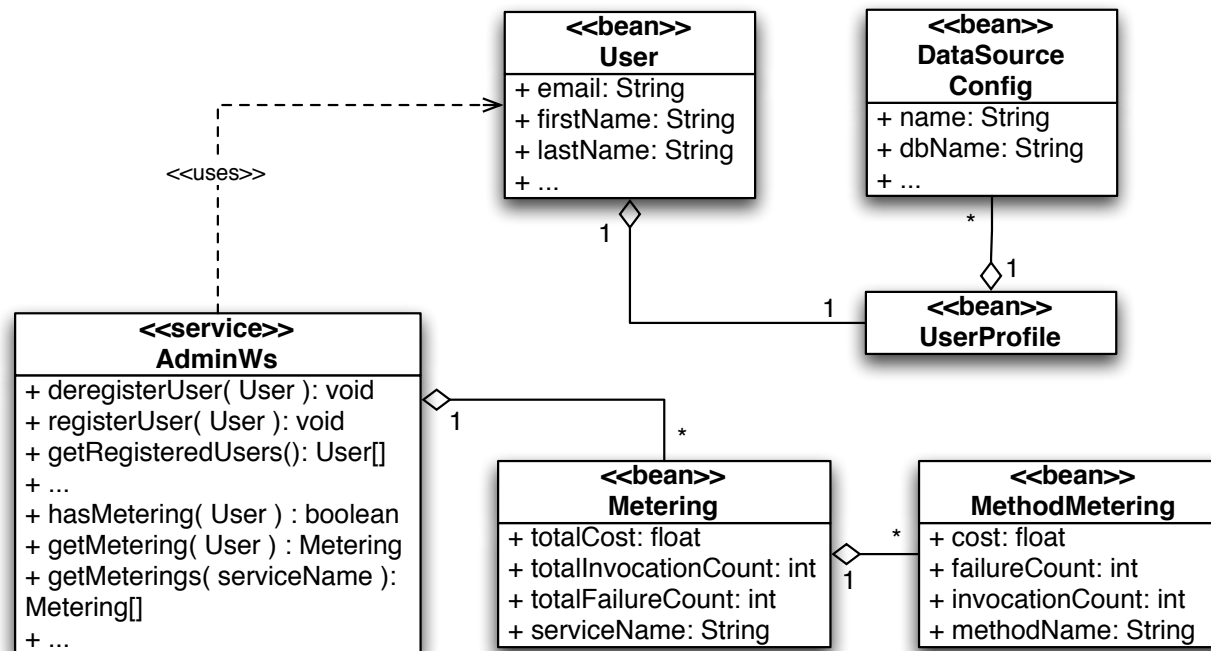
Architecture



Architecture - TAS



Architecture - Metering



Demo

The screenshot shows a web browser window titled "Serenity - Mozilla Firefox". The browser's address bar and menu bar are visible. The page content includes a navigation menu on the left with sections for "Data sources", "Solutions", and "Steps". The "Steps" section contains a list of actions: "Query", "Edit categories", "Select categories", "Co-occurrence table", and "Generate report". The main content area displays the title "BISON" and the subtitle "Business Information Services On a Network". Below this, there is a section titled "Select the step for the RCA solution" with a numbered list of the same five actions. At the bottom of the main content area, there is a footer that reads "Brought to you by the Service Oriented Technology team at Almaden Services Research" and a link to "E-mail us your feedback".

Serenity - Mozilla Firefox

File Edit View Go Bookmarks Yahoo! Tools Help

Serenity_{beta} Welcome back, test@test.org
[Log Out](#) [About](#)

Data sources

Solutions

Steps

- [Query](#)
- [Edit categories](#)
- [Select categories](#)
- [Co-occurrence table](#)
- [Generate report](#)

Info [Query](#) [Edit](#) [Select](#) [Analysis](#) [Report](#)

BISON

Business Information Services On a Network

Select the step for the RCA solution

1. Query
2. Edit categories
3. Select categories
4. Co-occurrence table
5. Generate report

Brought to you by the [Service Oriented Technology](#) team at [Almaden Services Research](#)

[E-mail us your feedback](#)

Reports

Demo (cont.)

Serenity - Mozilla Firefox

File Edit View Go Bookmarks Yahoo! Tools Help

Serenity *beta* Welcome back, test@test.org
[Log Out](#) [About](#)

Data sources

Solutions

Steps

- Query
- [Edit categories](#)
- [Select categories](#)
- [Co-occurrence table](#)
- [Generate report](#)

[Info](#) [Query](#) [Edit](#) [Select](#) [Analysis](#) [Report](#)

Edit categories step number 2 of 5

Name	Size	Remove	Merge
tech_verify	6	<input checked="" type="checkbox"/> Remove	<input type="checkbox"/> Merge
registers	4	<input checked="" type="checkbox"/> Remove	<input type="checkbox"/> Merge
power_ser	5	<input checked="" type="checkbox"/> Remove	<input type="checkbox"/> Merge
bkq	6	<input type="checkbox"/> Remove	<input type="checkbox"/> Merge
ser_tib	6	<input type="checkbox"/> Remove	<input type="checkbox"/> Merge
ends_fire	8	<input type="checkbox"/> Remove	<input type="checkbox"/> Merge
pose	7	<input type="checkbox"/> Remove	<input type="checkbox"/> Merge

1 - 7 of 40 [next >](#)

Name: power_ser

Browse documents: [◀](#) [▶](#)

text.cad: FN8QDDP

text.comments: 09/06 1441 PIX FIRE WALL IS DEAD WILL NOT POWER ON_SER#06009943_SEE TIB_*

Reports

Benefits

- **Agile** and **lean** development
- Evolutionary approach to service design
- Ability to iterate over service APIs and implementations
- Significant reuse of code base
- Multi-user at top levels of stack and slowly down

Directions

- More domain solutions
- User studies to improve ease of use
- Integrate services into business processes
- Experiment with various costing models
- Other service types, e.g., REST RSS, and Atom
- Mashups

Acks

- **Development**

- ➡ Scott Spangler

- ➡ James Rhodes

- ➡ Ana Lelescu

- ➡ Amit Behal

- **Management**

- ➡ Ying Chen

- ➡ Jeff Kreulen

- ➡ Ruoyi Zhou

धन्यवाद
Hindi

多謝
Traditional Chinese

ขอบคุณ
Thai

Спасибо
Russian

Gracias
Spanish

شكراً
Arabic

Thank You
English

Obrigado
Brazilian Portuguese

Grazie
Italian

多谢
Simplified Chinese

Danke
German

Merci
French

நன்றி
Tamil

ありがとうございました
Japanese