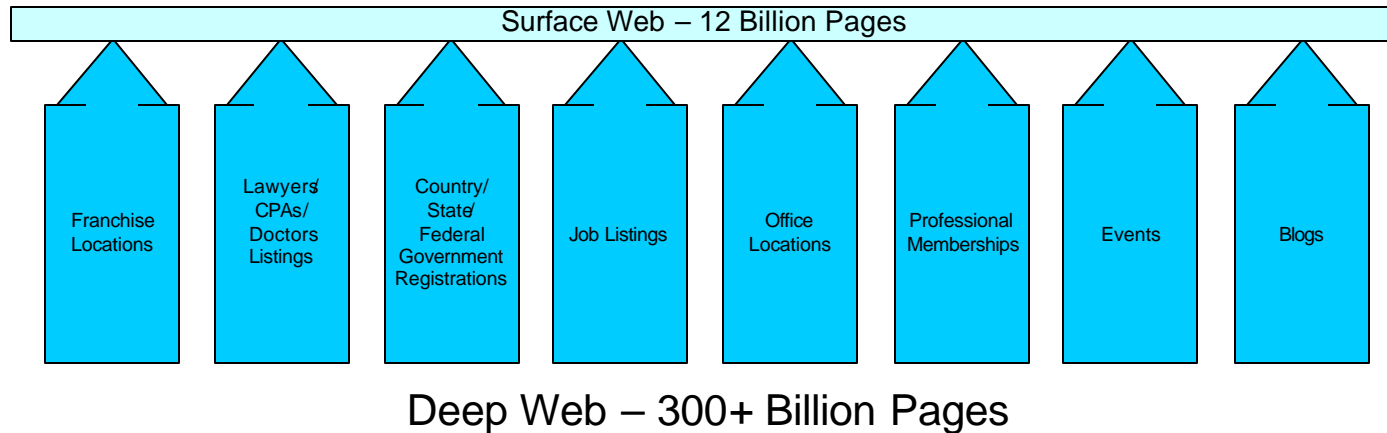


# Deep Web Search

**Julia Komissarchik**  
**Co-Founder & VP Products**  
**Glenbrook Networks**

# Deep Web vs. Surface Web



Surface Web - This is what search engines (e.g., Google) sift through

## Deep Web

- Pages revealed only after actions on prior page(s)
- Rich in content
- Pages often are created on the fly, in response to a specific inquiry
- Content is often perishable
- More and more of the desired content is stored in Deep Web

Challenge - To harvest desired data from the Surface Web and determine intelligently when and how to go to the Deep Web to collect more

# Deep Web Examples

---

- Travel
  - [United Airlines](#)
  - [Expedia](#)
- Job Search
  - [URS](#)
  - [Express Script](#)
- Store Locator
  - [Wendy's](#)
  - [H&R Block](#)
- Katrina
  - [Katrina Survivor – Connector List](#) from “Gulf Coast News”

# Problem

---

- Size - 10-30 times larger than the existing Surface web covered by Google, Yahoo, MSN and others
- Pages are dynamically generated in response to a question entered into DHTML forms – no passwords just appropriate questions
- Pages don't have urls - non-addressable space, no classic search engine link-based page ranking
- Designed for humans not machines
- Long tail – millions of web sites with DHTML forms of arbitrary structure

# Solution

---

- Vertical focus –
  - Manageable size
  - Restricted Semantics
  - Pragmatic knowledge
- Sophisticated AI techniques – penetration through forms
- Fact-based page ranking vs. link-based page ranking

# Prediction - 3-5 year out impact

---

- Web
  - Deep Web will become pervasive
  - Most businesses will have an extensive online presence
  - No Standardization in view for Facts Representation
- Factual Search will become prevailing
  - Expectations of answers to a question, not references to a bunch of documents to be read
  - Ability to ask questions like
    - Find a restaurant in 5 mile radius that serves chicken enchilada under \$5 and is open right now
    - Find a local plumber that has been in business for more than 10 years, has experience with sprinkle systems and accepts Visa
    - Find a business that is within 5 minutes walk of a bus stop for a line that is within 10 minutes of a given location
    - Find locations that have the highest increase of job openings in IT industry
    - Who are the top 10 RBI players in American league over last month
    - What was the apartment rentals trend in San Francisco over the last 6 months

# So – A Lot Of Business Opportunities

---

- Successful vertical applications
  - Business Info/Local Search
  - Events/Entertainment
  - Travel
  - Sports
  - Health
  - Job Search
- New Google in breeding – synthesis of multiple factual vertical Deep Web search engines into a massive horizontal search

# Glenbrook Networks

## **Data Collection and Fact Extraction Via Intelligent Web Trawling**



# Glenbrook Networks


---

- Glenbrook delivers data using intelligent patented search technology capable of extracting targeted information from the web with high precision, efficiency, and speed
- What makes us unique:
  - Glenbrook trawls the Deep Web, not just crawls the Surface Web
  - Glenbrook's output is a structured factual data feed rather than a simple list of links to pages of potential interest
  - It is not template based (so it is highly scalable)

# Application Examples


## Glendor Local Search

**Metro Mobile Communications**

**Logo**  
Company's Logo  **MetroMobile Communications**

**CompanyInfo**  
Company URL [www.metromobile.com](http://www.metromobile.com)

**Address**  
Company Address 3549 Haven Ave, Suite A Menlo Park, CA 94025-1070  
Zip Code 94025  
Company Phone Number (650) 367-1992  
Company Toll Free Number (800) 383-2929  
Company Fax Number Fax: (650) 367-1995

**Executive**  
Executive's Name John Singer  
Executive's Title President & CEO  
Executive's Picture 

**Executive**  
Executive's Name Paul Einhorn  
Executive's Title Vice President

Map of the San Francisco Bay Area showing locations like San Francisco, Oakland, Hayward, and San Jose. The map is powered by Google and includes data from Tele Atlas.

# Application Examples

## Glendor Local Search

http://glendor.com - Glendor Local - Milk Pail Market - Microsoft Internet Explorer

**Milk Pail Market**

**Logo**

Company's Logo




photo by Oliver Duvellier

**CompanyInfo**

Company URL [www.milkpail.com](http://www.milkpail.com)

**Address**

Company Address 2585 California Street  
Mountain View, CA 94040


Zip Code 94040

Company Phone Number (650) 941-2505

Company Fax Number fax: (510) 471-1645

**OfficeHours**

Hours of Operation Open daily 8:30 am - 8 pm (Sat 7 pm / Sun 6 pm)



Map Satellite Hybrid

San Leandro Hayward Union City Fremont Newark Milpitas Sunnyvale San Jose

Map data © 2005 File Atlas - Terms of Use

Done Internet

# Application Examples

## Glendor Deep Web Job Search

**GLENDOR SHOWCASE**

Home About Company Technology Products Contact us Help Glendor Blog

Position:  Description:

City or ZipCode in the Bay Area:  Within N miles: 10 Harvest Time Preference: Within the last 7 days Source Preference: All

LIST JOBS 20 MAP JOBS 20 Jobs 1-20 of 100+

Map Satellite Hybrid

- Human Resources Business Partner  
Company: CREDECE SYSTEMS US - CA, Militas (Headquarters)  
Sources: CREDECE SYSTEMS
- Entry Level Engineering  
Positions: ADEPT TECHNOLOGY Livermore, CA  
Sources: ADEPT TECHNOLOGY
- Senior Technical Writer  
Company: ARIBA Sunnyvale, CA  
Sources: ARIBA
- Materials Manager 3  
Company: CREDECE SYSTEMS US - CA, Sunnyvale  
Sources: CREDECE SYSTEMS
- SOFTWARE ENGINEER  
Company: ACTUATE South San Francisco, CA  
Sources: ACTUATE
- PROGRAM COORDINATOR 2  
Company: CREDECE SYSTEMS US - CA, Militas (Headquarters)

Jobs 1-20 of 100+

# Trend Analysis of Job Postings

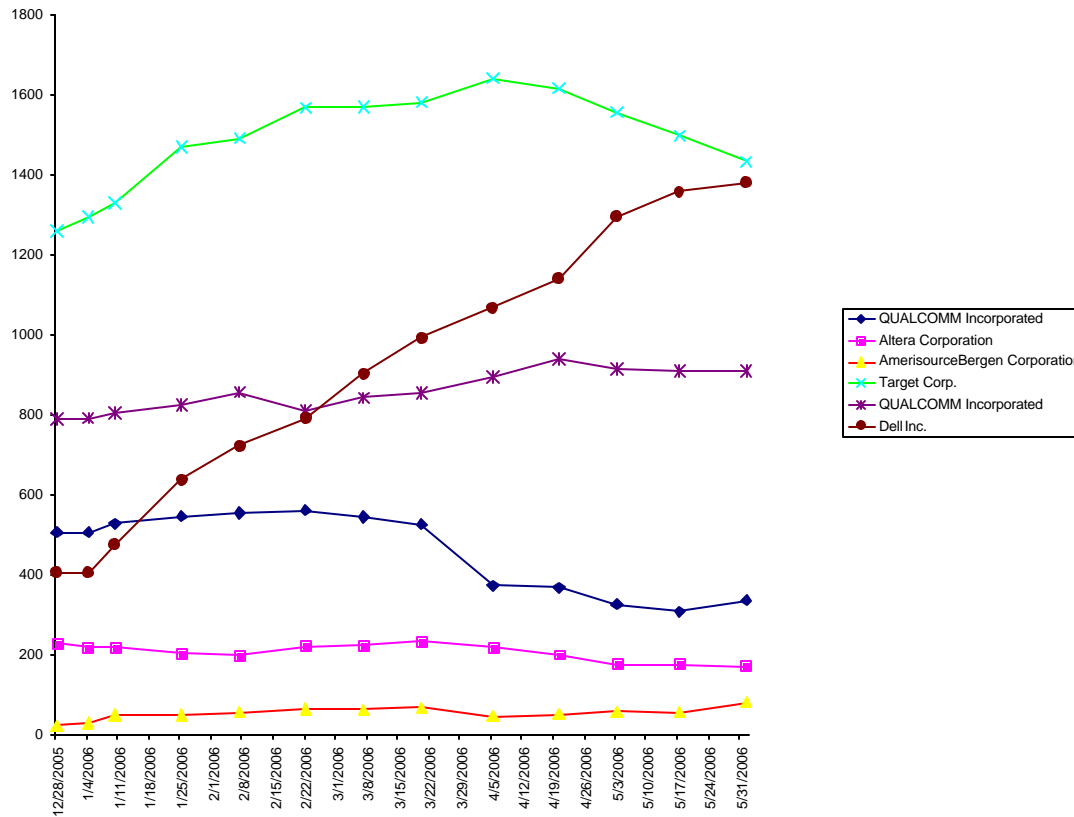
## Case Study

---

- A large Wall Street financial institution has approached Glenbrook to help collecting in-depth information about public companies, in particular job postings
- Using Glenbrook Deep Web trawler and Fact Extractor data was collected biweekly directly from public companies websites
- The data feed was used by analysts to perform trend analysis that influenced their recommendations for stock market

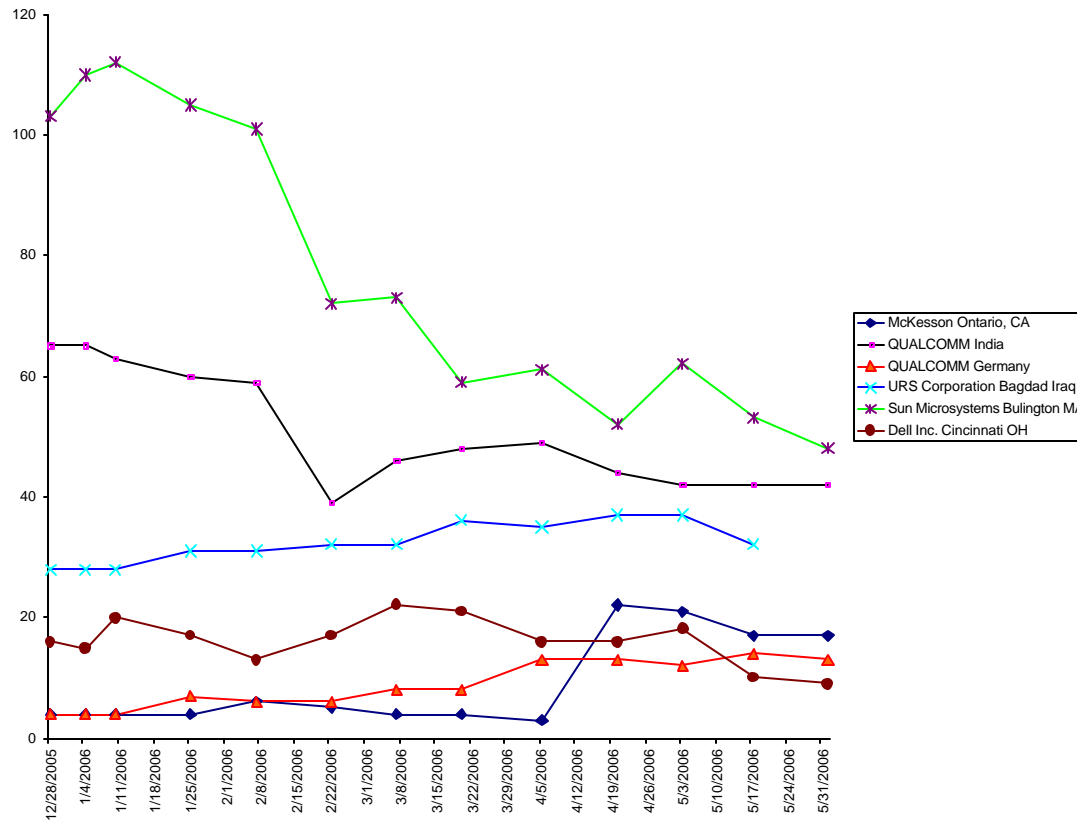
# Public Companies Job Postings

## Trend Analysis (by Company)



# Public Companies Job Postings

## Trend Analysis (by Company and by City)



Thank you

---

Julia Komissarchik

[julia@glenbrooknet.com](mailto:julia@glenbrooknet.com)

650 759 3959

