iAD – information Access Disruptions

Dag Johansen
University of Tromsø and Fast Search & Transfer

Geneva, 10th October 2007

Outline

(1) Background

- Enterprise Search.

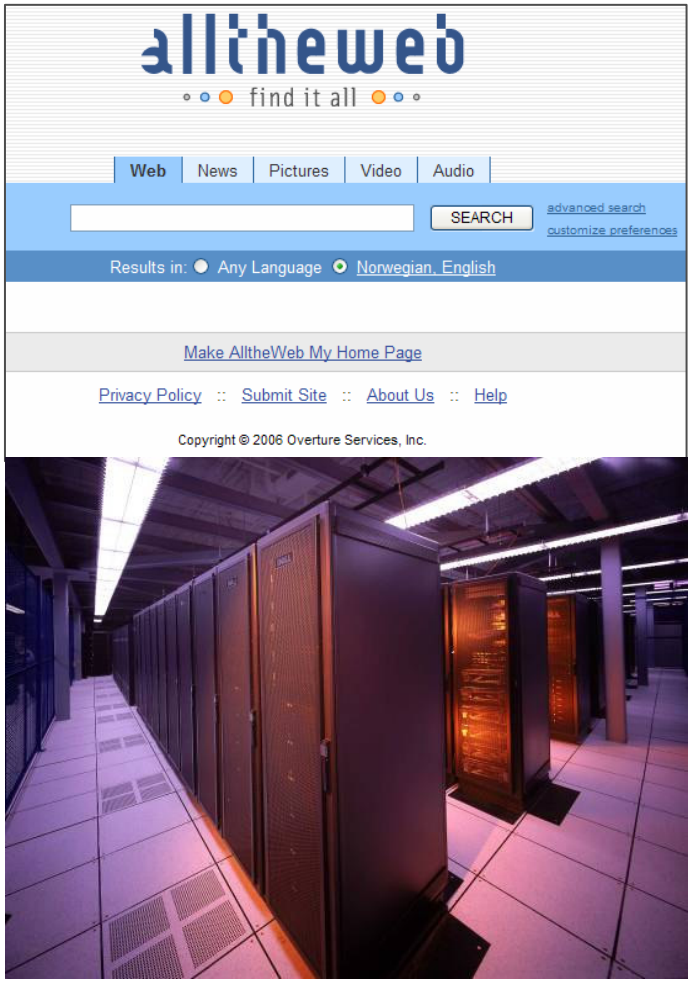
(2) Research Problems.

(3) The iAD project.

(1) Enterprise Search

– We used to provide an Internet search service:
  - More than x0M QPD.
  - Average: below 300ms.
  - 10+G index.

– From hosted service to enterprise *software*.



Applicability

- Search as interface to enterprise data:
  - Internal search service.
  - External search service.
  - Real-time filtering mechanism.
  - Database replacement.
  - Federation of multiple data sources.
  - Layer between data (stores) and analysis applications.

Beyond Search

- Enterprise search is beyond document retrieval.
- Embedded in a lot of business-critical applications:
  - Business intelligence.
  - e-commerce.
  - Surveillance and security.
  - Fraud detection.
  - Quality management.
  - …….

Example: Rakuten

“eBay of Japan”

楽天 RAKUTEN ICHIBA

Handling Extreme Query Volumes While Keeping the Catalog Up-to-date

- ✓ Handling 500 updates per second.
- ✓ Extreme performance requirements with 2,000 QPS during normal traffic, peak level of 3,000 QPS in real-time scenario.
- ✓ Generates $12M of transactions per day through search.

fast

Example: Comcast

The leading broadband provider of cable TV, Internet, and VoIP in the U.S.

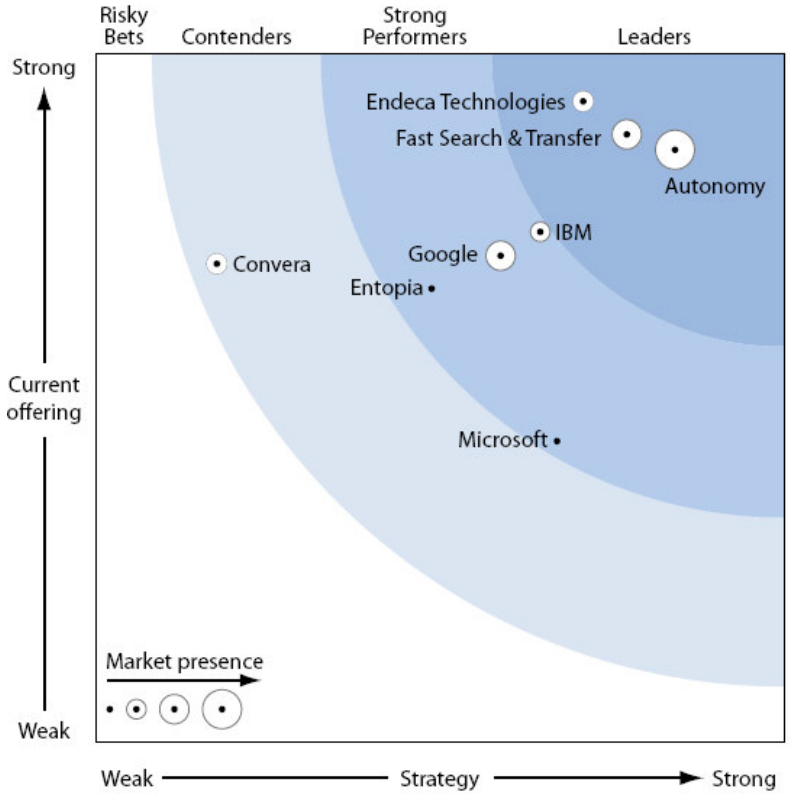
Providing information to more than 27 million customers

- ✓ The sixth busiest search engine on the Internet.
- ✓ Traffic is more than 3 million queries per day.
- ✓ Indexing 14 million Yellow Page records.
- ✓ Federating entertainment content of popular TV shows.

Many Industrial Players

FORRESTER

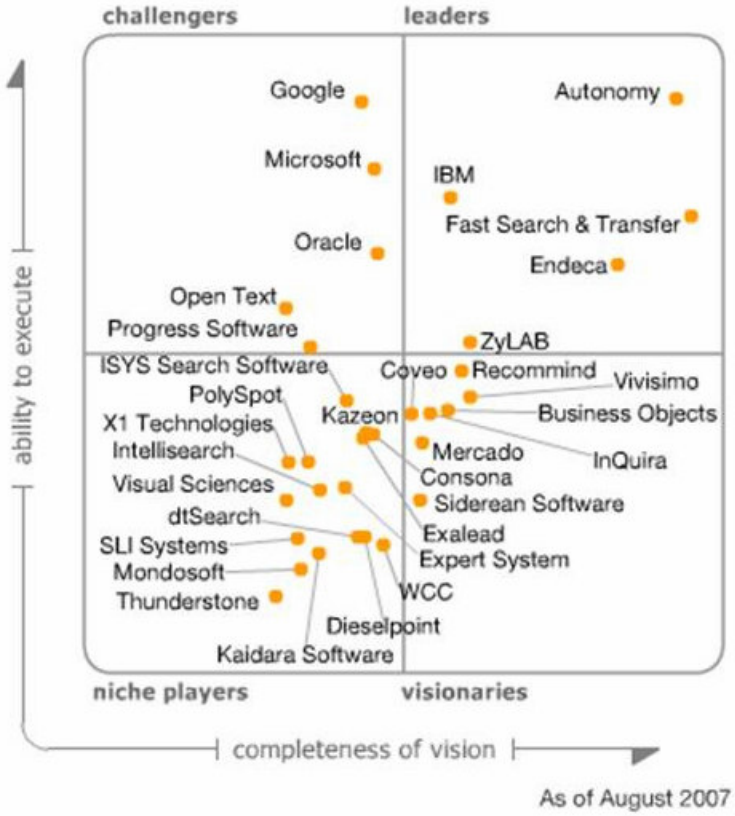
Forrester Wave™: Enterprise Search Platforms, Q2 '06



Source: Forrester Research Inc.

Gartner

Magic Quadrant for Information Access Technology, 2007



As of August 2007

Source: Gartner (August 2007)

Disclaimer:
The Magic Quadrant is copyrighted 6 October 2007 by Gartner, Inc. and is reused with permission. The Magic Quadrant is a graphical representation of a marketplace at and for a specific time period. It depicts Gartner's analysis of how certain vendors measure against criteria for that marketplace, as defined by Garter. Gartner does not endorse any vendor, product or service depicted in the Magic Quadrant, and does not advise technology users to select only those vendors placed in the "Leaders" quadrant. The Magic Quadrant is intended solely as a research tool, and is not meant to be a specific guide to action. Gartner disclaims all warranties, expressed or implied, with respect to this research, including any warranties of mechantability or fitness for a particular purpose.

(2) (Some) Research Problems

- Scale and performance (in-house service).
- Data freshness and *time consistency* (time to market vital).
- Extreme precision (frequent changes, context).
- Data evolution (track temporal or spatial changes).
- Trends (zipf behaviour, related trends, group conformity).
- New data formats and types (semantic and automatic annotation).
- Security (user integrity and privacy).
- Mediation (deduce conclusions, syndication and aggregation).
- Proactivity and recommendation (push web).
- Autonomic computing (initial assembly, dynamic adaption).

(3) iAD - information Access Disruptions

– Norwegian cluster of academics and companies focussing on search:
  - NTNU (Trondheim) and University of Tromsø.
  - Fast Search & Transfer.
  - Yahoo.
  - Google.
  - Sesam (Schibsted).
  - ……

Innovation Key

– Innovation and research key to existence.

– Applied for Norwegian Centre of Excellence status with FAST as host

→ Large-scale Norwegian Research Centre for *information access*.

Who is FAST?

Key Facts

Founded in 1997.

Headquarters: Oslo & Boston.

Employees: 730.

Revenues 2006: USD 162M.

INFORMATION ACCESS DISRUPTIONS

http://www.iad-centre.no

sfi Centre for Research-based Innovation

Established by the Research Council of Norway

iAD – Goals

- Core research for next generation precision, analytics and scale in information access.
- Build international networks to identify and execute on global disruption opportunities.
- Educate 25 PhDs.
- **Timeline:** Jan 2007, duration: up to 8 years.
- **Budget:** €30m.

iAD – partners

iAD mobilizes…

Partner network

- FAST (host institution).
- Accenture as global partner to define next generation information management solutions.
- Schibsted to validate iAD results.

Customer experience

- Access to customer requirements of global Fortune 500 companies.
- Collaborative live laboratories.

Academic network

- Cornell University.
- AIC Dublin (DCU, UCD).
- Norwegian universities:
  - NTNU Trondheim.
  - U Tromsø.
  - U Oslo.
- Norwegian School of Management.

iAD: Customer Requirements

…and many more FAST customers.

INFORMATION ACCESS DISRUPTIONS

http://www.iad-centre.no

sfi Centre for Research-based Innovation

Established by the Research Council of Norway

iAD – Sub Projects

Service

Technology

Collaboration

Schema agnostic indexing services (NTNU)

- Schema-agnostic end2end design.
- Consolidation of query model.

Processing high-speed data streams (Cornell)

- Capturing & extracting knowledge from data streams.
- Pervasive sensor networks, RFID readers, multimedia feeds, …

Scalable infrastructure for push and pull based computing (U Tromsø)

- Robust principles and services for next generation infrastructure for distributed information access.

Extreme precision and recommendation in multimedia access (Dublin)

- Extreme precision solutions for access to multimedia content.
- Social networks with recommender functions.

Understanding and managing the disruptive potential of iAD (Norw. School of Management)

- Analyze business and societal impact.
- Assess disruptive potential.

INFORMATION ACCESS DISRUPTIONS

http://www.iad-centre.no

Summary

- Enterprise search is becoming a critical factor within the IT infrastructure of modern enterprises.
- Extremely complex software with hard research problems:
  - Performance and scaling requirements.
  - Security requirements.
  - Fault-tolerance requirements.
  - Deployment and maintenance requirements.
- Search is killer application with high impact on society:
  → *Economic, ethical, social, cultural, and political*.

Thank you – questions?