Trustworthy Search

Silvija Seres
VP Strategic Market Development
Fast Search & Transfer

Abelia Innovasjons Fagnettverk for Informasjonssikkerhet
September 2006

Contents

1. Corporate Overview
2. Trustworthy Search
3. Use Cases
4. Conclusions

FAST Corporate Overview

Silvija Seres, FAST

Abelia Innovasjons

Fagnettverk for Informasjonssikkerhet

Vision

*fast is transforming search into an **IT infrastructure necessity** and a **key enabler** for **mission-critical applications***

Corporate Overview

FAST's enterprise search solutions are changing entire industries and enabling governments to perform more effectively.

Tackling the most demanding problems

Rock-Solid Stability

- Founded in 1997, 50% CAGR
- Profitable, no debt, well-capitalized

Skyward Momentum

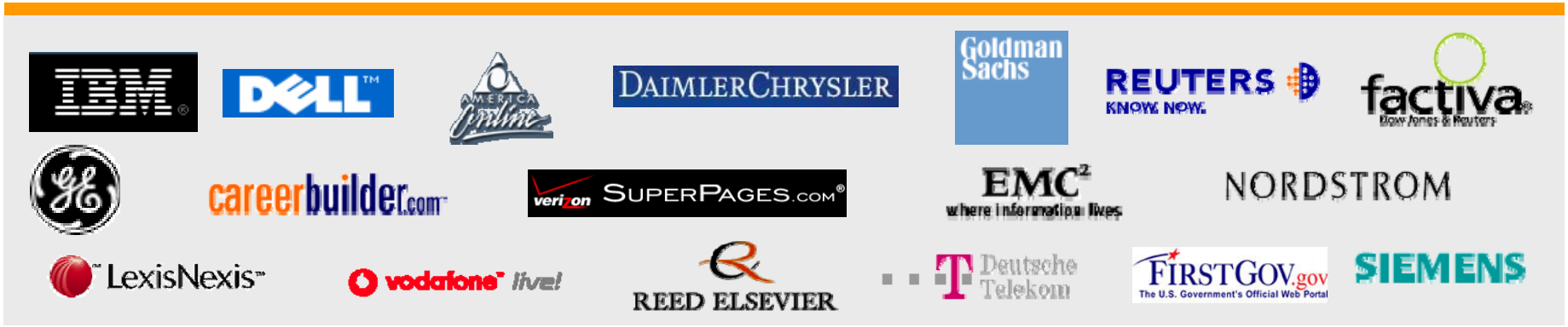
- #1 in revenue growth
- Over 3,500 installations

Exceeding Expectations

- 99% Customer Satisfaction

fast

Mission Critical Applications



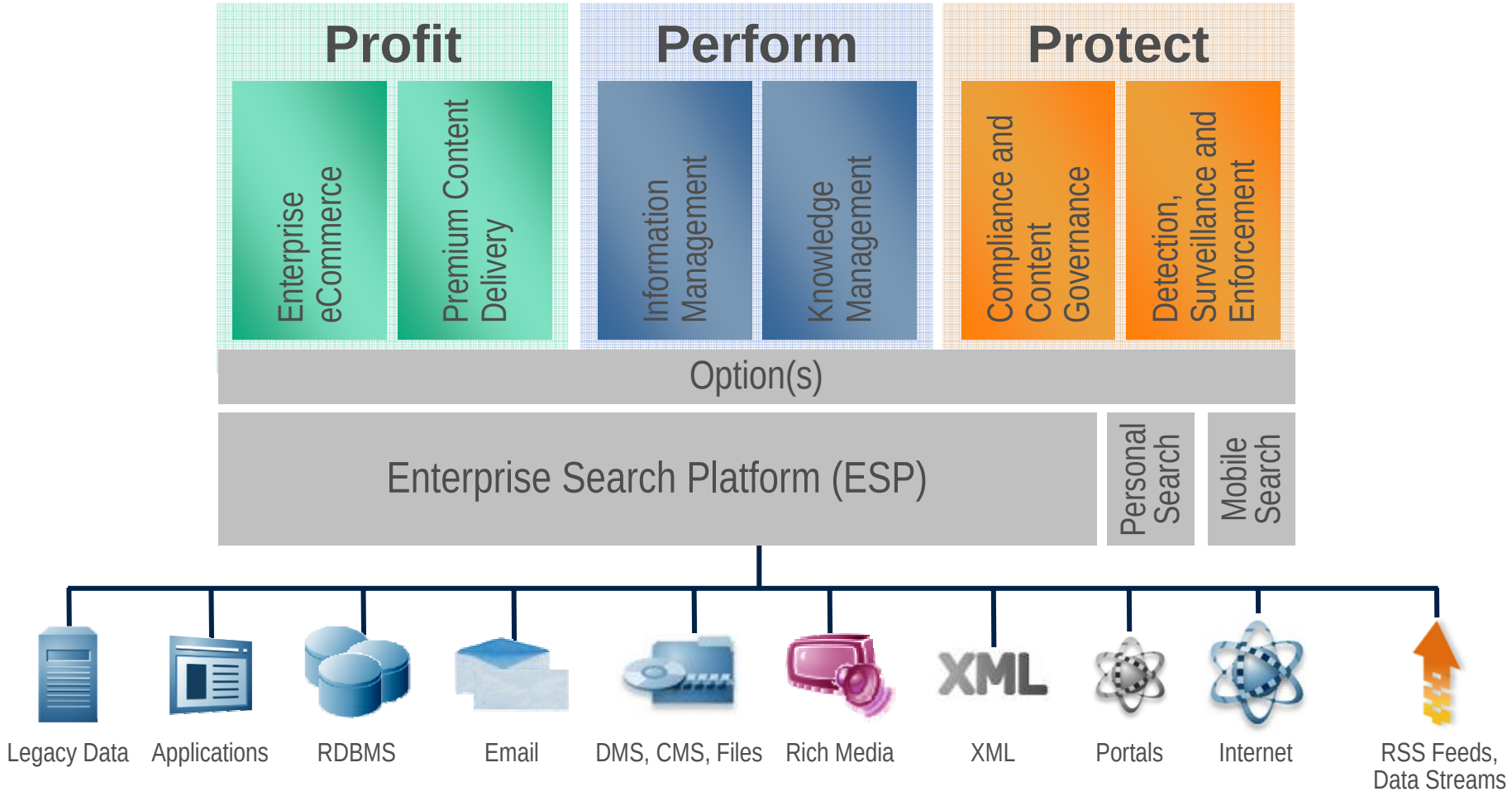
OEMs

Content Applications	Messaging Applications	Storage Applications	Business Applications
EMC² documentum Partner	Orchestria	EMC² where information lives®	SIEBEL
BROADVISION	ZANTAZ®	ARCHIVAS digital archiving	LexisNexis™
STELLENT	AXS one®		Chordiant.
Computer Associates®			

Alliances

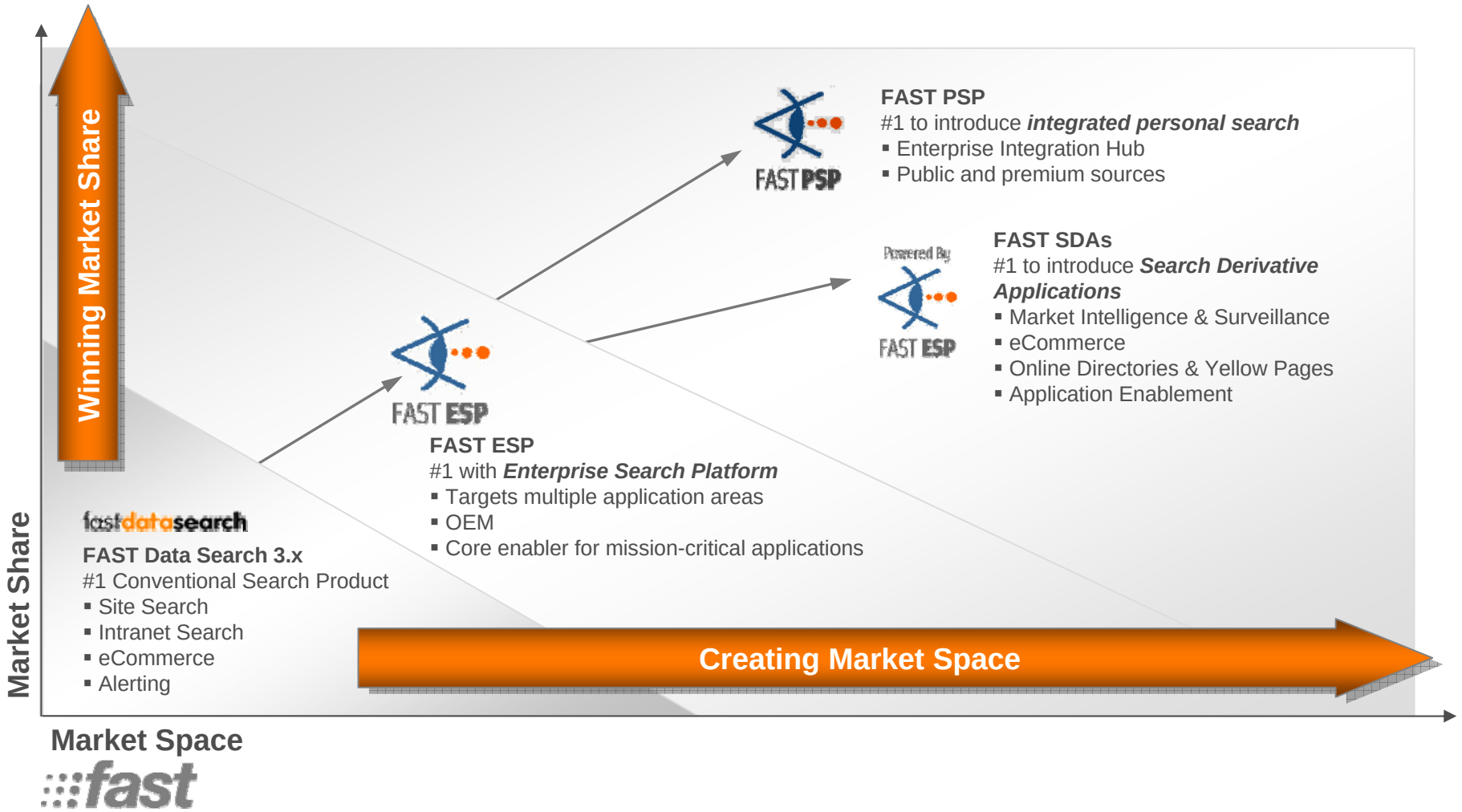
bea® · CAP GEMINI ERNST & YOUNG · accenture High performance. Delivered. · Microsoft · INTERWOVEN · NORTHROP GRUMMAN · LOCKHEED MARTIN · Satyam · grokker · danet · Sun microsystems · TietoEnator · Deloitte. · hp invent · comperio · IBM BCS & LS · ORACLE · EDS · ERNST & YOUNG Quality In Everything We Do · WIPRO Applying Thought

One Platform, Many Applications



Exponential Innovation

Technology Innovation & Market Innovation



Trustworthy Search

Silvija Seres, FAST

Abelia Innovasjons
Fagnettverk for Informasjonssikkerhet

Trustworthy Computing

The term **Trustworthy Computing** (TwC) has been applied to computing systems that are inherently secure, available and reliable. The Committee on Information Systems Trustworthiness' publication, Trust in Cyberspace, defines such a system as one which:

"does what people expect it to do – and not something else – despite environmental disruption, human user and operator errors, and attacks by hostile parties. Design and implementation errors must be avoided, eliminated or somehow tolerated. It is not sufficient to address only some of these dimensions, nor is it sufficient simply to assemble components are themselves trustworthy. Trustworthiness is holistic and multidimensional."

Source: wikipedia

fast

Trusted vs. Trustworthy

The National Security Agency (NSA) defines a *trusted* system or component as one "whose failure can break the security policy", and a *trustworthy* system or component as one "that will not fail".

Trusted Computing has been defined and outlined with a set of specifications and guidelines by the Trusted Computing Platform Alliance (TCPA). These include: secure input and output, memory curtaining, sealed storage, and remote attestation.

Trustworthy Computing aims to build consumer confidence in computers, by making them more reliable, and thus more widely used and accepted.

Source: wikipedia

fast

Microsoft's Definition of TwC

Four areas are initiative's key "pillars".

1. Security
2. Privacy
3. Reliability
4. Business Integrity

But…

– History speaks for itself
– So why is trustworthiness so difficult to solve? Some ideas for discussion are:
  - Breadth of definition
  - Legacy code or dependency on 3rd parties
  - General complexity of software systems relative to human engineering capacities
  - Inherent difficulty of security – correct policies, and their perfect implementation
– What is good enough?

FAST's Vision

A service is **trustworthy** to the extent that it continues to function as expected, despite:

- attacks
- Failures

Trustworthiness is relative to assumptions about:

- numbers / kinds of failures
- the attacker

System construction is thus relative to some assumptions.

Ideal Design: Allows adjustment of system parameters to accommodate changes in assumptions about environment.

fast

Trustworthiness Properties

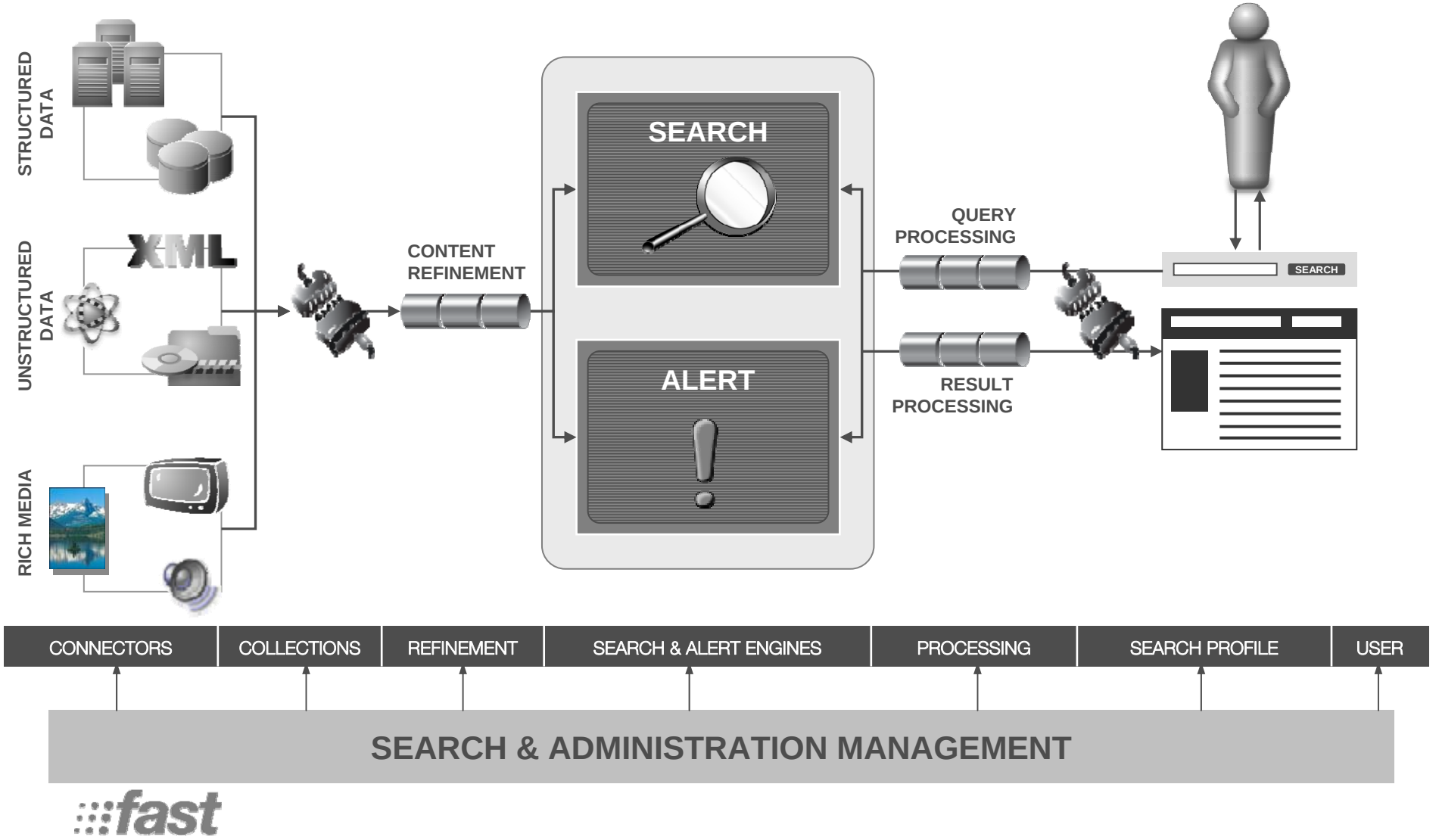
Trustworthy systems do no more and no less than expected:

- **Availability**: The system is there when you need it.
- **Integrity**: Accurate reporting of values.
- **Confidentiality**: Secrets don't get leaked.

Easy: Support for availability and integrity

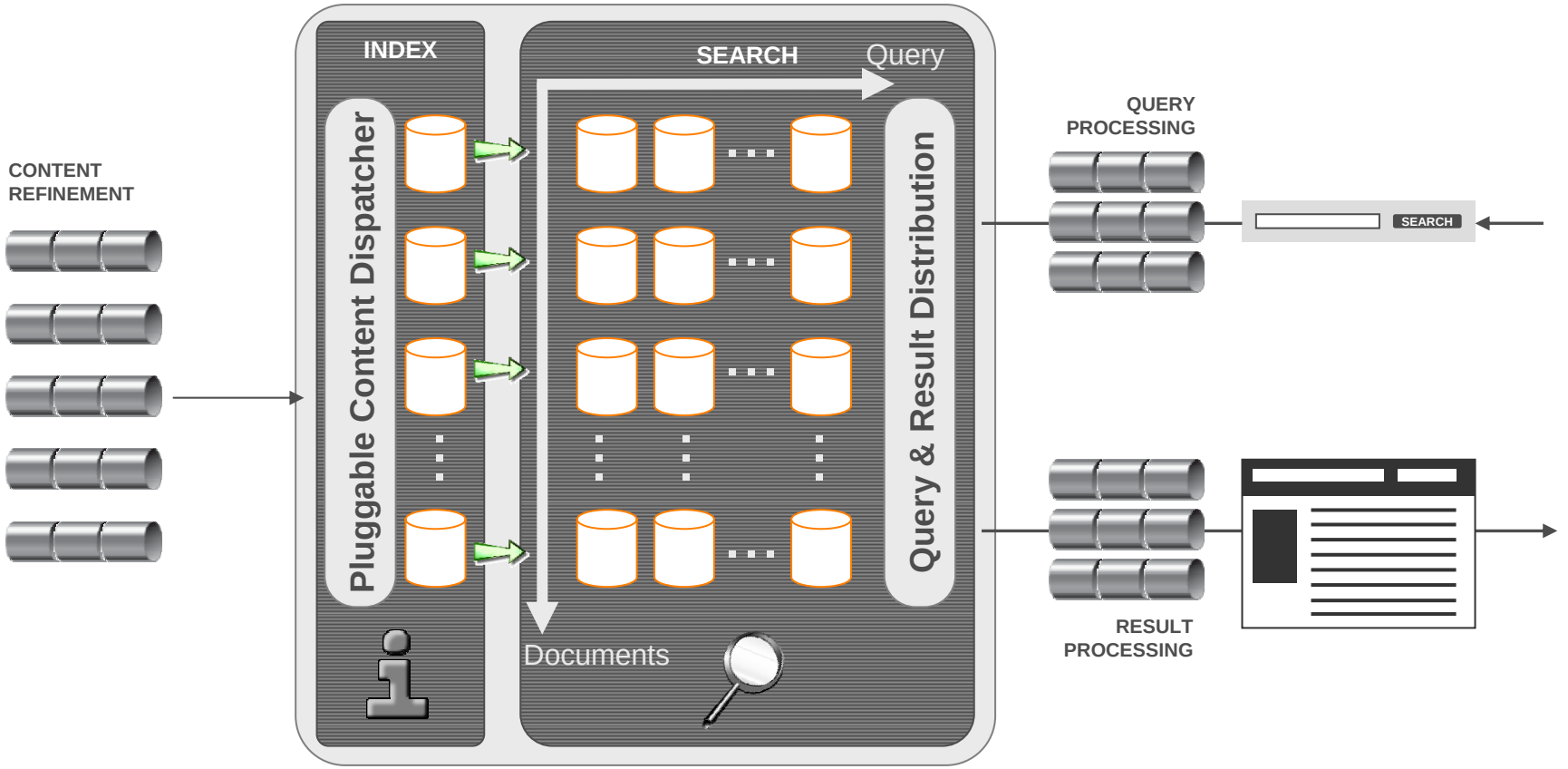
Difficult: Support for confidentiality (more than the Search Engine is involved)

Platform Architecture

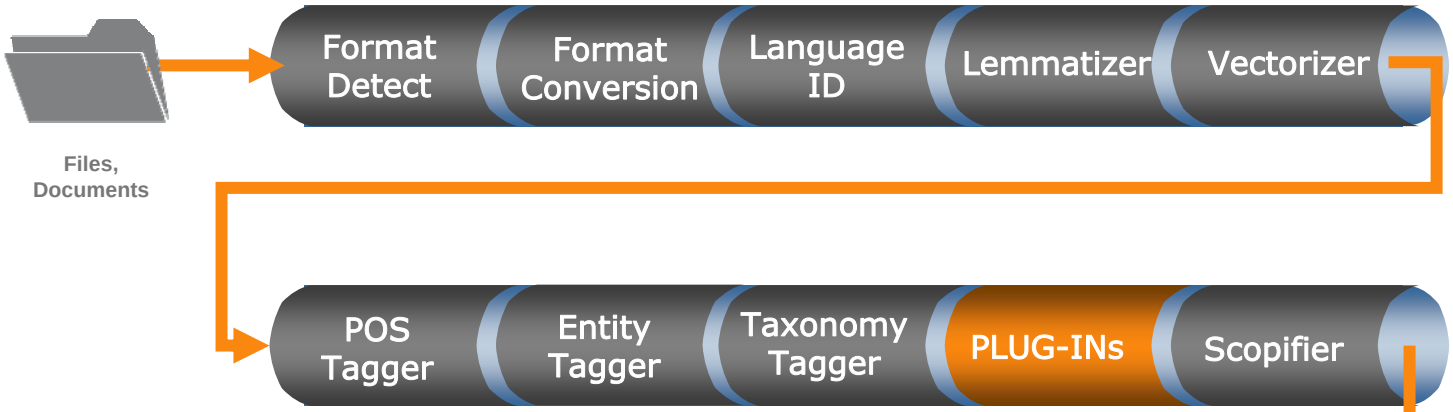


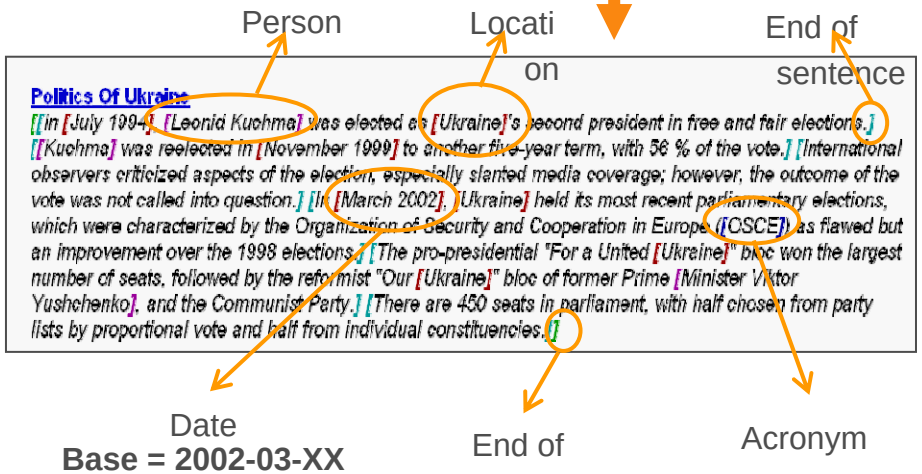
Integrity and Availability

Handle Volume and Linear Scaling – Through Careful Distribution



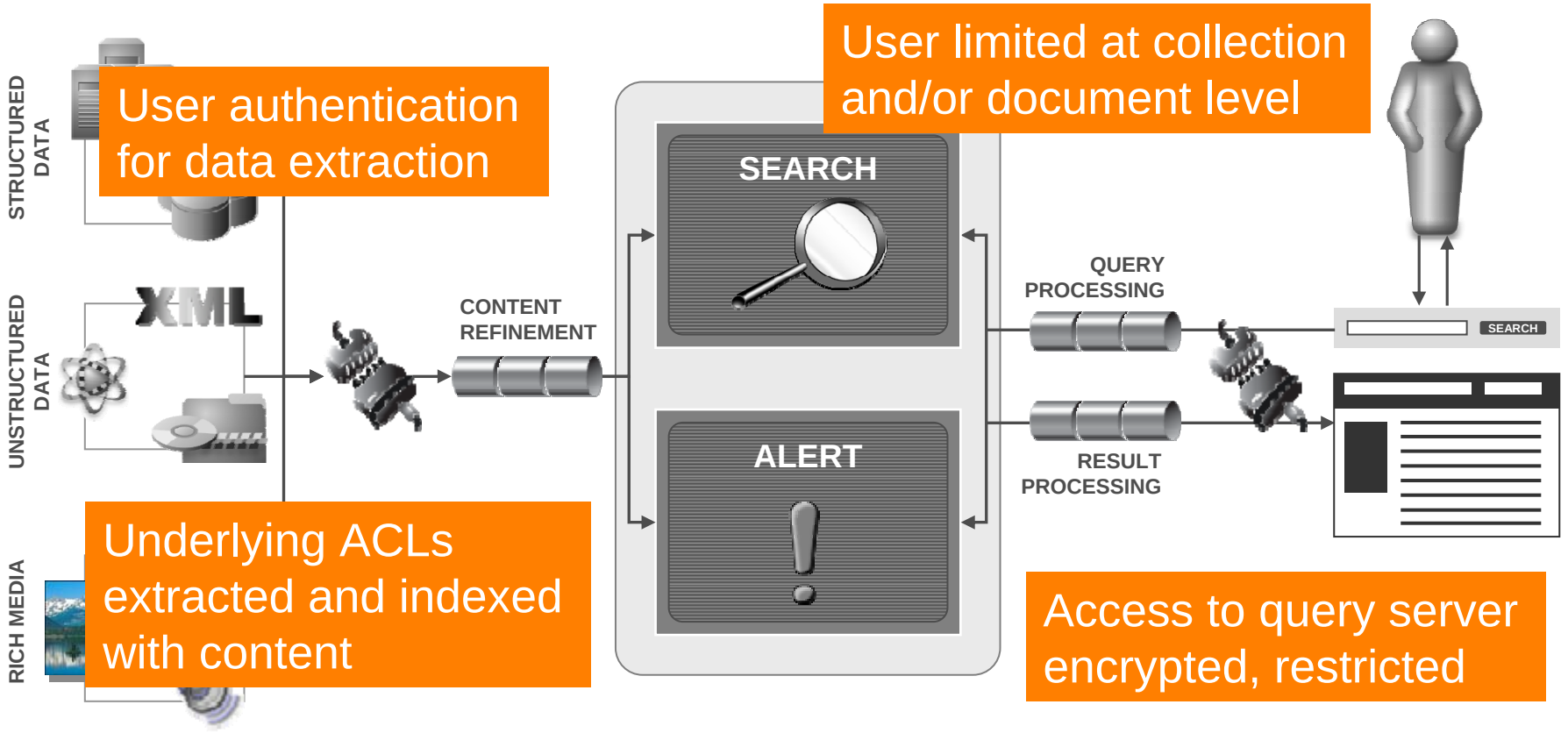
Integrity
Give Users Rich Search – Through Content Refinement





Security

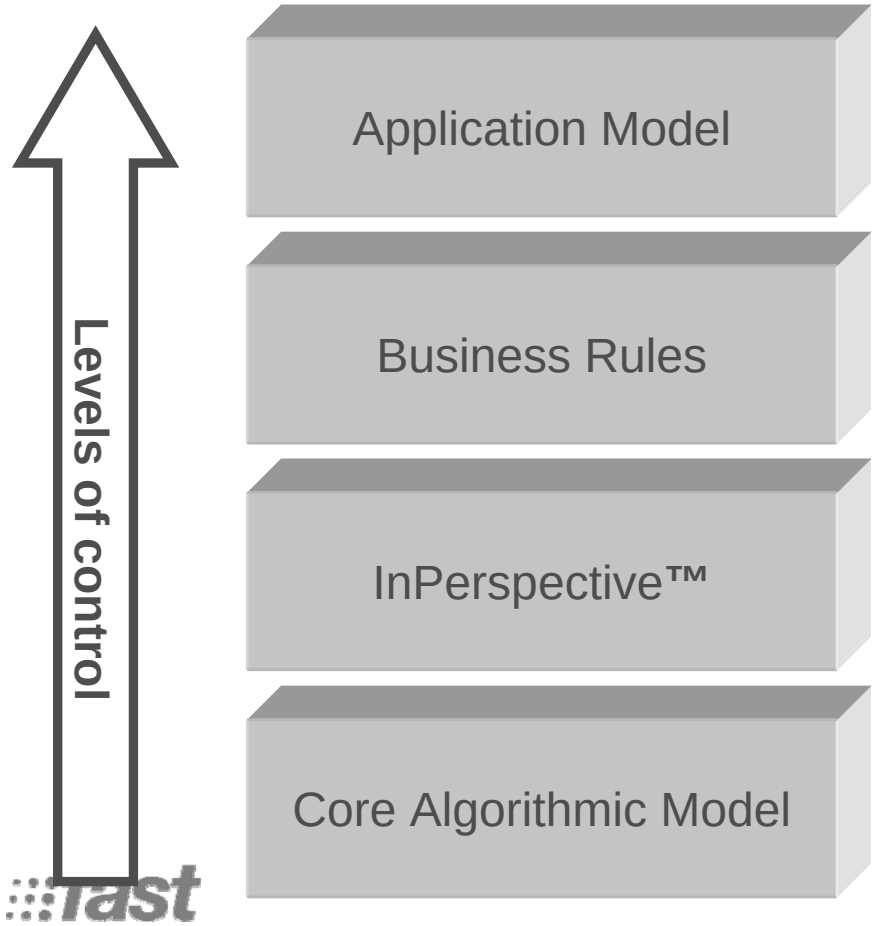
Protect Privacy – Through Authentication at All Levels



More Integrity

Provide Expected Results – Through Relevancy

Multiple levels of control



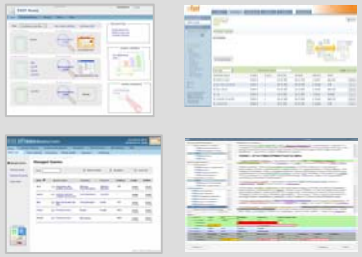
Accessible to…	Control Mechanisms
End Users	Sorting order, navigation, relevance feedback
Business Managers	Query and document “boosting” (BMCP)
Administrator	“Rank Profile”
Developer	Algorithm “weights”

Availability

Provide Flexible Solutions – Through Pipelines & SOA

Application Layer

(Eclipse, Web, Command Line)



<Develop new GUIs/SDAs>

Service Layer

(Exposed as Java & Web Services)

<Write new custom services>

Domain Logic Layer

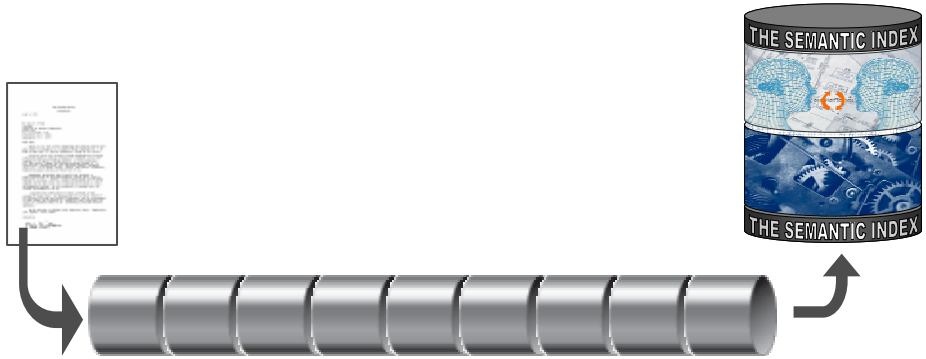
ESP Core

Content SDK
C++, Java, .NET

Search SDK
C++, Java, .NET

Example:

Add *3rd party or custom analysis*



Identify Object Type

Calculate Final Pricing

Part-of-speech semantics

My Analyzer

fast

Security Notes

1. A system is only as secure as its weakest link; don't spend time and money on traffic encryption and yet allow staff to leave print-outs of confidential data lying around
2. Document-level security means that individual documents cannot be accessed by other authorized users of the system
3. System-wide security means locking down against unauthorized access using encryption, IP filtering, and OS-level security
4. Index-based ACL mapping is faster and more scalable than post-query filtering
5. Search engines typically rely on being installed within a secure environment to maintain complete data integrity

A search solution must be integrated into the security fabric of the organization or of the host IT shop

fast

Availability Notes

1. Computers fail for many reasons: hardware, software, power, connectivity, etc
2. Service degradation can be abrupt or graceful
3. High availability is ensured with redundant systems
4. Different parts of a system can have varying degrees of fault tolerance
5. Downtime costs money, but so does redundancy

The extent of built-in redundancy required needs to be measured against the opportunity cost of a failure.

Use Cases

Silvija Seres, FAST

Abelia Innovasjons
Fagnettverk for Informasjonssikkerhet

Use Cases

There are Many Search Flavors

Corporate Search

Provide all employees with a single and simple information access point

Consumer Search

Provide all viewers, listeners and other consumers with a great and complete user experience

Structured Search

Provide faster, cheaper, more flexible access to all structured relational data

Paid Search

Provide potential customers with quick and precise access to digital content they will want to buy

Intranet:
Statoil

About Statoil

- Statoil is an integrated oil and gas company with 23,899 employees and activities in 29 countries
- Total revenues 2004: USD 50 billion. Norway's largest company.
- One of the world's biggest sellers of crude oil, Statoil is also a major supplier of natural gas and has substantial industrial operations. It has services in the Scandinavian countries, Ireland, Poland, the Baltic states and Russia.

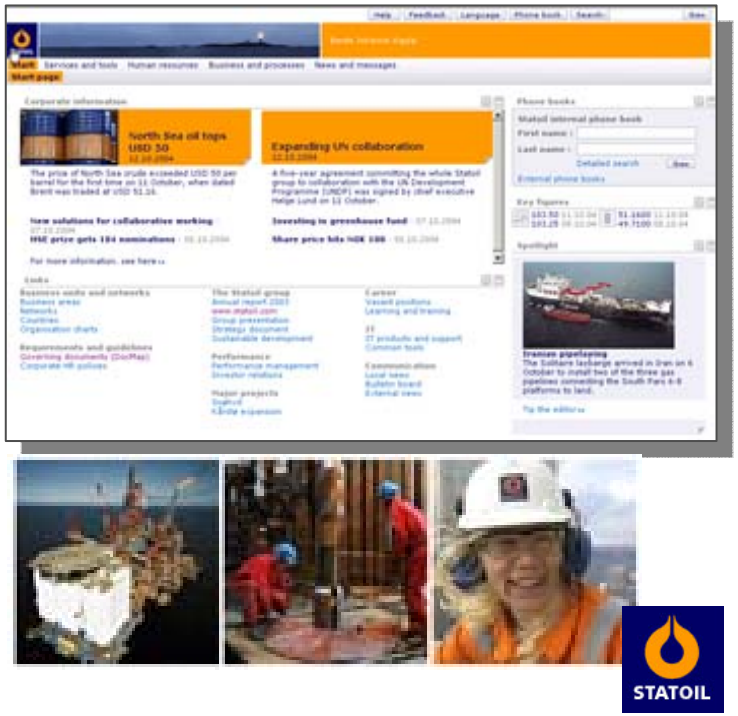
Challenge

– New intranet search, with multifaceted end user experience, task oriented usage, across large data volumes and queries
– Information security, sharing and management
– Data migration and maintenance, integration with legacy technology

Solution

– Collaboration and information sharing across organizational and geographical boundaries
- Traceability and easy, correct and secure access of information through the information life cycle and with respect to legal requirements
- Facilitating effective change management for widespread acceptance of new information infrastructure.

Multinational Energy Company



Results with FAST ESP

20TB of data, growing at 10TB per year
Reliable access to huge data volumes, 24/7, varied users
Complexity control**, in terms of volume, bandwidth, maintenance**

fast

End user requirements: I love IT!

Monitoring and Patrolling:
Toll and Excise Authority

About TAD

- The leader in Europe in detecting financial transaction patterns which indicate money-laundering and other illegal activity
- With the new anti-money-laundering and foreign exchange transaction registration laws enacted and effective Jan 1 2005, all financial entities in Norway are required to report transactions over 3000 Euros conducted by foreigners in Norway, international transfers, and credit-card transactions by Norwegians abroad

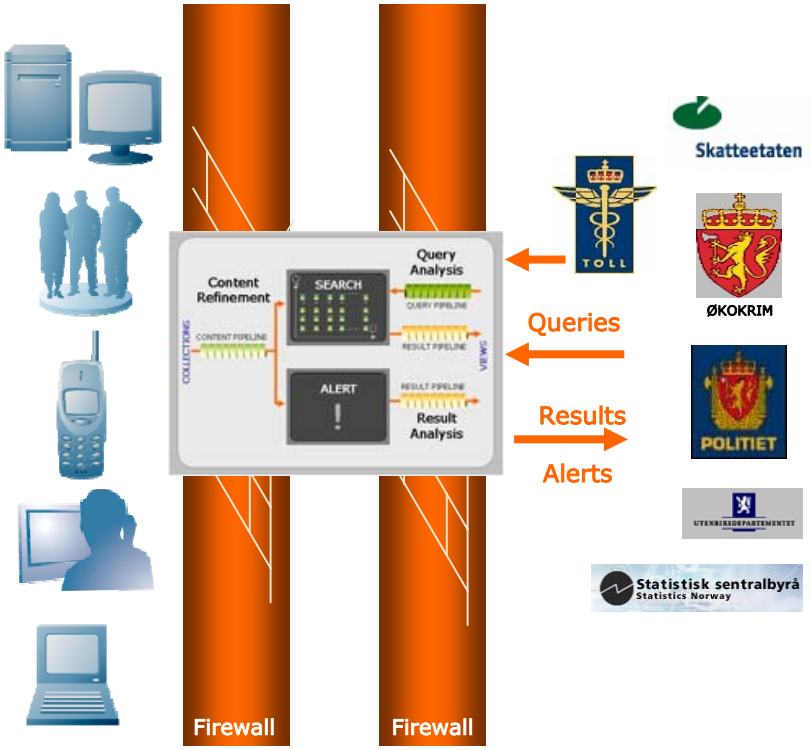
Challenge

- Existing solution was batch-oriented, expensive to administer, required components with separate licenses, and did not allow data mining linked to BI & Portals, had severe scaling and performance limitations, and little value was gained from unstructured data

Solution

- All data is stored in FAST ESP, and searchable securely by a variety of authorities with varying degrees of access. The National Statistical Bureau can extract aggregate reports on foreign exchange movements, while police authorities can use the system to investigate specific cases down to the level of individual transactions.
- The system was operational ahead of schedule, and is expected to contain up to 10TB of information over the lifecycle of the program.

Norwegian Customs Office



Results with FAST ESP

Full corporate/organizational data consolidation, unstructured data annotated and tagged

The speed, cost-effectiveness and navigation provided by Search supplanted an oracle-RDBMS-based design that was projected to cost 3-5x more.

eCommerce:
Dell

Premier Provider of Information Technology

About dell.com

- **One of the worlds leading eCommerce sites with $**47.3 billion in revenues **and approximately 53,000 team members**
- Over 50% of Dell's revenue now comes through Dell.com

Challenge

- **Drive revenues by allowing brand managers to control their own products**
- **Enhance consumer experience**

Solution

- **FAST ESP 4.x with Live Analytics**
- **Advanced linguistics and real-time files**

FAST Differentiators

– Business Studio for better merchandising **(promote specific products)**
– **More complete results by integration of multiple data sources**
– **Global presence, multiple languages**

We feel that the integration of FAST's software has enhanced the overall search experience on Dell.com.

Susan Sheskey
VP, I/T Sales and Marketing Services, Dell

Results with FAST ESP
Each brand manager optimizes their own product collections
Dell.com supports 5.5 million visits per week
Query search time reduced by order of magnitude (10 sec to sub-second)
FAST Data Search is used for retail, B2B, self-service support and internal customer service rep support

eMerchandizing:
Best Buy

About Best Buy

- Best Buy Co., Inc is North America's number-one specialty retailer of consumer electronics, personal computers, entertainment software and appliances.
- Fiscal 2005 earnings from Continuing Operations increase 17% to $934 million. Fortune 100 company with more than 830 retail stores across the United States and in Canada
- In 2004, Forbes named Best Buy as its company of the year, and . Best Buy held the No. 1 position on the Stores ranking of the Top 100 Specialty Retailers.

Challenge

- Grow online sales, by creating a competitive differentiato and defensive tool against major online vendors (e.g. Amazon)
- Provide a complete information platform, extendible to many verticals, with a unique set of features and scalability characteristics – cannot be a ‘me too’ solution that is easy to copy by other vendors.

Solution

- FAST Enterprise Search Platform technology (FAST ESP) across multiple applications throughout the organization, including the company's online retail sites

fast

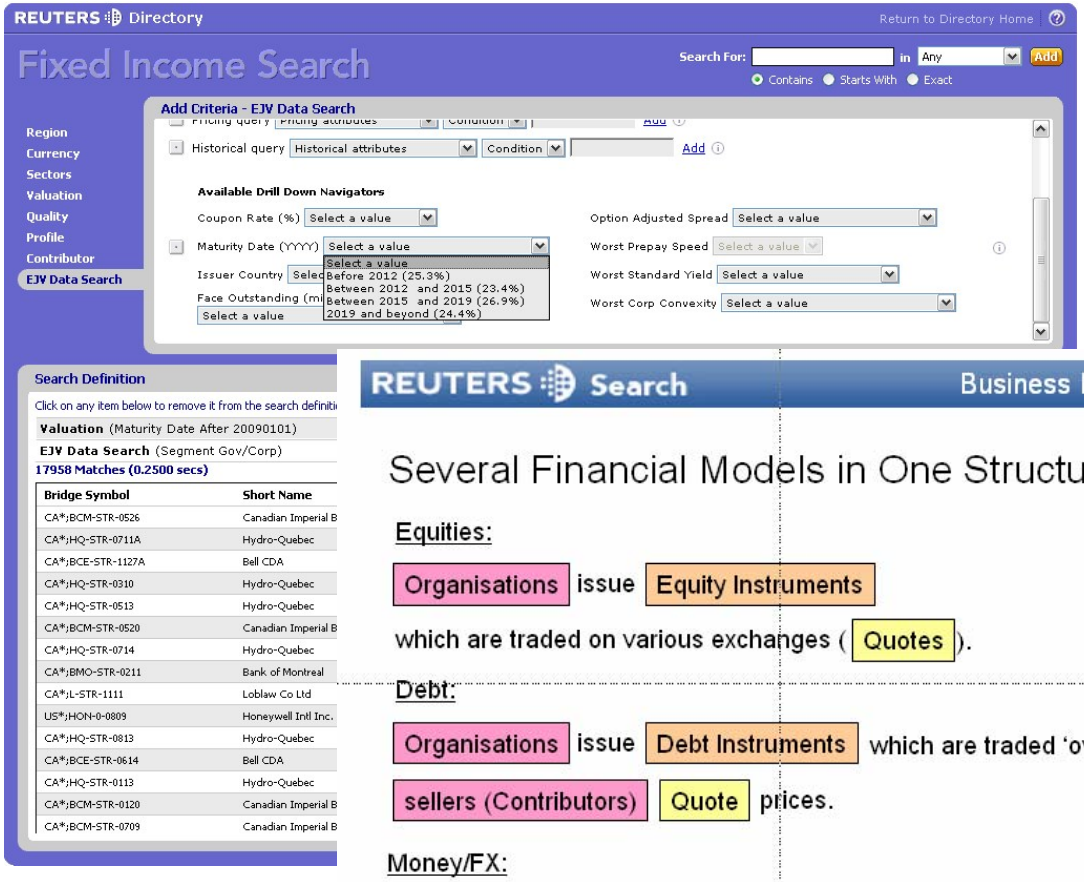
Largest US Retailer of Consumer Electronics

FAST ESP and FAST ImPulse enable online retailers to easily fine-tune the search experience to benefit both buyer and seller

Results with FAST ESP

The percent of searchers seeing a "no results" page when searching has dropped dramatically from 20% to 5%.

Financial:
Reuters



185.000 Companies
2.100.000 Different Stocks
2.800.000 Stocks/Stockmarkets

REUTERS Search Business Model

Several Financial Models in One Structure:

Equities:

Organisations issue Equity Instruments

which are traded on various exchanges (Quotes).

Debt:

Organisations issue Debt Instruments which are traded 'over the counter'

sellers (Contributors) Quote prices.

Money/FX:

Sellers (Contributors) Quote prices for Money Market Instruments.

432 Facts
83 Dimensions

Fact = stock broker
>550 updates / sec
>2000 query / sec

fast

Marketing Intelligence:
DnB NOR

Major Nordic Bank

About DnB NOR

- Norway's leading financial services group with total assets of over NOK 1200 billion. Principal bankers to 60 percent of Norway's 300 largest companies
- Norway's largest life insurance company, capital markets operation, asset management operation, and Norway's leading real estate broker
- Commitment to Knowledge-based Innovation: annual award for new solutions in knowledge-based business development, consistently rated "best-in-class" by European financial press

Challenge

- Discover market-moving information 'in front of the curve.'
  - Get more information, Get better information
  - Get information sooner, Get precise information

Solution

- Support personalization of information flow, with analyst focus on premium sources, e.g. Reuters, "Live" sources – e.g. forums, chats, discussions, blogs and information from correspondent network institutions, DnB Research studies.
- Correlation of price movements with related news, analysis of news flow for market-moving potential, 'on-the-fly' analytics, juxtaposing structured and unstructured information
- Communication of actionable information with minimal 'latency', personalized taxonomies, views for analyst research with analyst profiles for 'low-latency' alerting.

The Bottom Line:

Timely Knowledge is Money

Reuters used to stream by, now we mine it for information.

DnB NOR Markets

Results with FAST ESP

Brokers and analysts have more time to do what they do best – sell/trade and identify opportunity

DnB NOR Markets delivers better results, leveraging timely information

'Early warning' reduces risk, opens additional opportunity

Monitoring and Patrolling:
Asian Police Agency

About Customer

- One of the technically most advanced police agencies in the world, in an area with one of the highest internet usage in the world.
- The police authority was looking for a way to effectively monitor the internet to uncover child pornography images.

Challenge

- Using a database of images classified as illegal, the police is scanning public web spaces (chat rooms, web sites, etc) to discover and alert officers to images that resemble these.
- Images are often significantly manipulated to avoid such programmatic recognition.

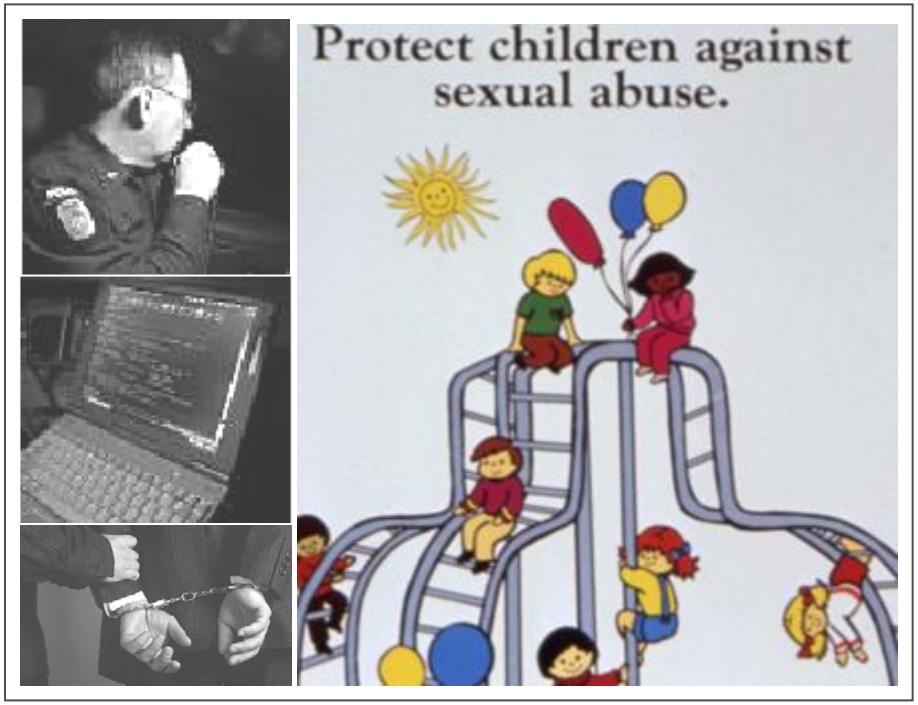
Solution

- FAST's advanced imaging tools recognize images that are similar to the offending ones, through statistical and heuristic methods taking in to account many image aspects, including hue and color.

> *Children's protection is our responsibility, but too often we have offered children nothing more than vague non-profits and feel-good rhetoric.*
>
> PROTECT

fast

Major Asian Country's Police Agency



Results with FAST ESP
Effective monitoring of the Internet space and discovery of illegal content
Several arrests made as result of FAST technology's monitoring capabilities

Conclusions

Silvija Seres, FAST

Abelia Innovasjons
Fagnettverk for Informasjonssikkerhet

Do you Really Want Less Information –
Or Just Better Information?

~~Information Overload?~~

Information Underload

The Future of Search

universal

Search will be everywhere all the time. Technical boundaries will disappear (Internet vs. Intranet, structured data vs. unstructured data). Information will be source and device independent (server, desktop, mobile).

pervasive

Search will permeate more and more areas of people's lives. It will follow you around and be contextually aware of your surroundings. It will be proactive rather than reactive. It will suggest and alert and inform.

necessary

You simply won't be able to live without search. Consumers demand more, governments require more, and competitors know more about you. With search, a higher standard emerges with greater expectations.

fast

… so, it must also be Trustworthy!

Thank you

Would you like to know more about FAST?

Please mailto: silvija.seres@fast.no