Leveraging Your Enterprise Content with IBM Content Discovery

Aaron Brown, Program Director, ECM Content Discovery

Act.Right.Now.

IBM INFORMATION ON DEMAND 2007
October 14 – 19, 2007
Mandalay Bay
Las Vegas, Nevada

IBM INFORMATION ON DEMAND 2007

Act. Right. Now.

The Enterprise Information Challenge...

- Do decision makers have the right information at hand to take timely, effective action?

- Is enterprise content being leveraged effectively to improve business performance?

- How much time do employees spend looking for the right answers?

- Is content being utilized to gather insight on operational effectiveness of the organization?

Act.Right.Now.

There is a Solution: Business Insight from Content

- Turn content into <u>actionable information</u>

- Deliver breakthrough <u>business insight</u>:

  - **Discover** new insights into customers, suppliers, operations, and more

  - **Differentiate** from the competition

  - **Empower** knowledge workers to make the right decision

  - **Improve** business results

... all enabled by IBM Content Discovery solutions.

Leader in 2007 Gartner Information Access Magic Quadrant!*

*Gartner, "Magic Quadrant for Information Access Technology, 2007," Sept. 5, 2007

IBM INFORMATION ON DEMAND 2007

Act.Right.Now.

How Content Discovery Solutions Leverage Content

Improve Content Access

Organize Unstructured Content
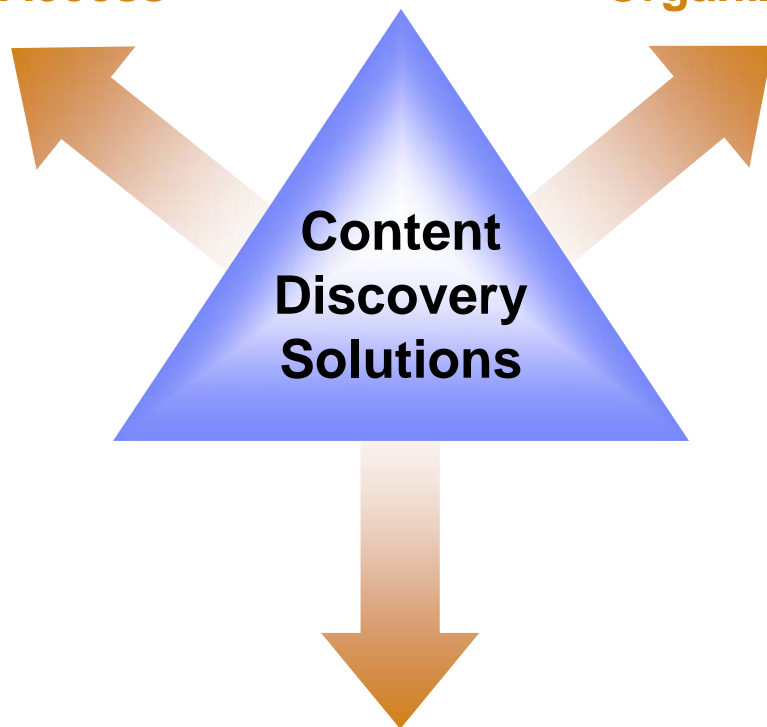
Content Discovery Solutions

Derive Business Insight

Act. Right. Now.

How Content Discovery Solutions Leverage Content

Improve Content Access

IBM OmniFind Search

Content Discovery Solutions

Find the right content at the right time to influence key business decisions

"What do our widget contracts with ACME Corp. say about liability?"

- *Rapid access to managed and unmanaged content*

- *Quickly find relevant content*

- *Respect and enforce native security*

How Content Discovery Solutions Leverage Content

Organize Unstructured Content

IBM Classification Module

Content Discovery Solutions

- *Maximize content utility by accurately cataloging it*
- *Increase productivity by automating decisions*

Integrate taxonomies and "folksonomies" to provide a coherent view of content

"How can I standardize my taxonomy to make it easier to find all relevant content?"

Organize content and records to manage legal exposure

"How can I ensure users are classifying records properly?"

"What records should I examine first for this case?"

7

Act.Right.Now.

How Content Discovery Solutions Leverage Content

IBM Content Analytics

Content Discovery Solutions

Understand the meaning of content and gain breakthrough insights

"What makes our customers happy or unhappy about our services? What are the trends?"

Derive Business Insight

- *Discover trends and insights to optimize your business*

- *Turn unstructured content into an asset by extracting actionable knowledge*

Act. Right. Now.

Content Discovery Solutions Drive Broad Value Across an Organization

Information Management Initiatives

- Advanced Search
- Semantic Search
- Self-service Search
- Integrating Managed and Unmanaged Content
- Dynamic Warehouse insight & analysis
- Vertical solutions for Customer Insight, Quality Early Warning, Law Enforcement analytics, Public Opinion Monitoring, Healthcare, and more...

ECM Initiatives

- Content Classification
- Taxonomy Automation
- e-mail Supervision
- eDiscovery
- BPM Decision Automation
- Content Insight / Analytics
- Process analytics
- Additional vertical solutions...

Collaboration Initiatives

- Intranet Search
- Portal Search
- eCommerce Search
- Self-service Search
- Web Site Search
- Departmental File Search
- Notes/Domino Search
- Integrating Community Content
- ...

Act.Right.Now.

Topics

- Driving value from content with Content Discovery

- The three pillars of Content Discovery

  ➡ Search

  – Classification

  – Analytics-driven Business Insight

- Take Action: Roadmap for leveraging Content Discovery

Act.Right.Now.

Improved Content Access – via Search

- Search is the most familiar, effective metaphor for users trying to find information
  - Accurate, relevant search provides rapid access to content
  - Enables access to information regardless of where stored
  - Improves productivity, reduces cost, and optimizes customer experience

Act. Right. Now.
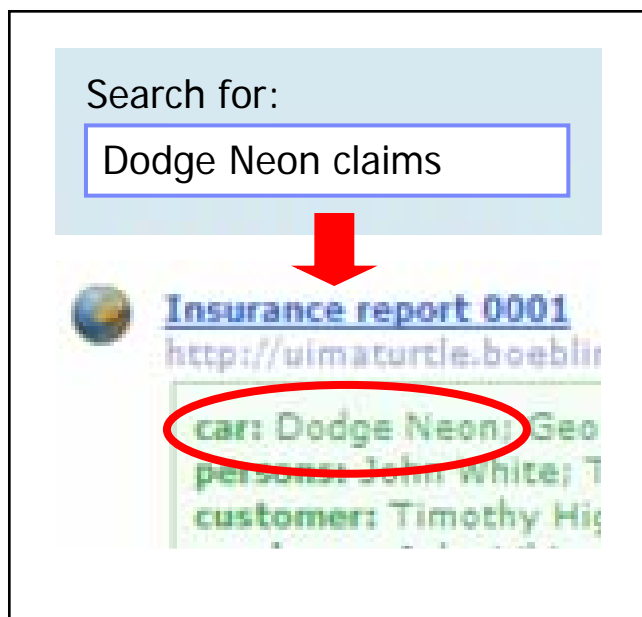
Business Value of Search

Key Business Drivers			
Improve Productivity ①	**Make Better Information-based Decisions** ②	**Reduce Support Costs** ③	**Maximize Online Revenue** ④
Enable employees to more quickly find information needed to complete their business activities ▪Rapid, accurate access to the right document ▪Draw content together from managed and unmanaged sources ▪Secure access to content	**Find critical information by concept, not just keyword, providing key insight to make better decisions faster** ▪ Semantic search to enable linguistically-rich concept extraction ▪ Drill-down and navigation to explore results by concept ▪ Trend detection enables proactive action	**Empower customer and partners to support themselves and perform their own research** ▪Understand customer intent, regardless of form of interaction (keyword, phrase, sentence) ▪Rich user interface for intuitive discovery of solution	**Ensure customer can easily find products and services, driving higher sales and increasing customer retention** ▪Contextually sensitive access to the right products ▪Targeted cross-sells and up-sells ▪Business user empowerment to manage the customer experience

← *Internally-facing* → ←→ *Externally-facing* →

OmniFind Enterprise Edition

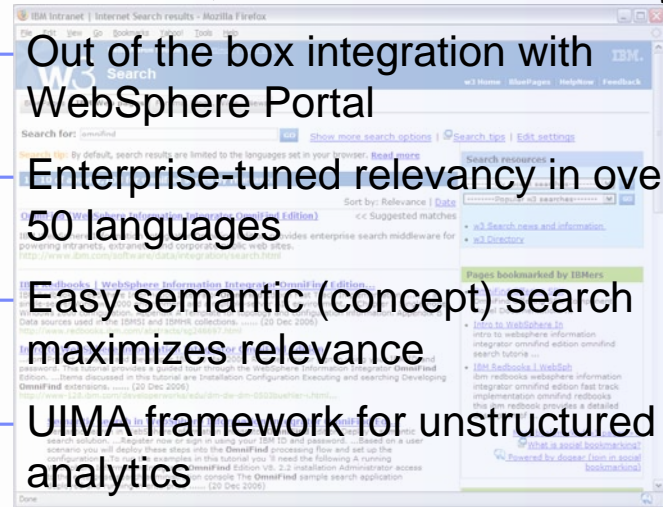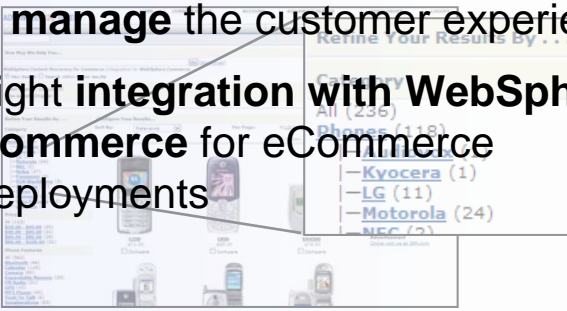| IBM Products and Solutions		
Improve Productivity **1**	Make Better Information-based Decisions **2**	**OmniFind Enterprise Edition** – Secure, scalable enterprise search – Support for 30+ content sources, including P8, CM8, Domino, web, file system, and more – Real-time, document-level security
Enable employees to more quickly find information needed to complete their business activities ▪Rapid, accurate access to the right document ▪Draw content together from managed and unmanaged sources ▪Secure access to content	**Find critical information by concept, not just keyword, providing key insight to make better decisions faster** ▪ Semantic search to enable linguistically-rich concept extraction ▪ Drill-down and navigation to explore results by concept ▪ Trend detection enables proactive action	– Out of the box integration with WebSphere Portal – Enterprise-tuned relevancy in over 50 languages – Easy semantic (concept) search maximizes relevance – UIMA framework for unstructured analytics

← *Internally-facing* →

Act. Right. Now.

OmniFind Discovery Edition

| IBM Products and Solutions		
OmniFind Discovery Edition	Reduce Support Costs **3**	Maximize Online Revenue **4**
– Deliver content to end customers via **rich, managed search experience** – Faceted navigation, contextual & natural language queries **guide customers to their goals** – Graphical **layout editor** makes it easy to design rich search UIs – Business owners can easily **monitor & manage** the customer experience – Tight **integration with WebSphere Commerce** for eCommerce deployments	**Empower customer and partners to support themselves and perform their own research** ▪Understand customer intent, regardless of form of interaction (keyword, phrase, sentence) ▪Rich user interface for intuitive discovery of solution	**Ensure customer can easily find products and services, driving higher sales and increasing customer retention** ▪Contextually sensitive access to the right products ▪Targeted cross-sells and up-sells ▪Business user empowerment to manage the customer experience

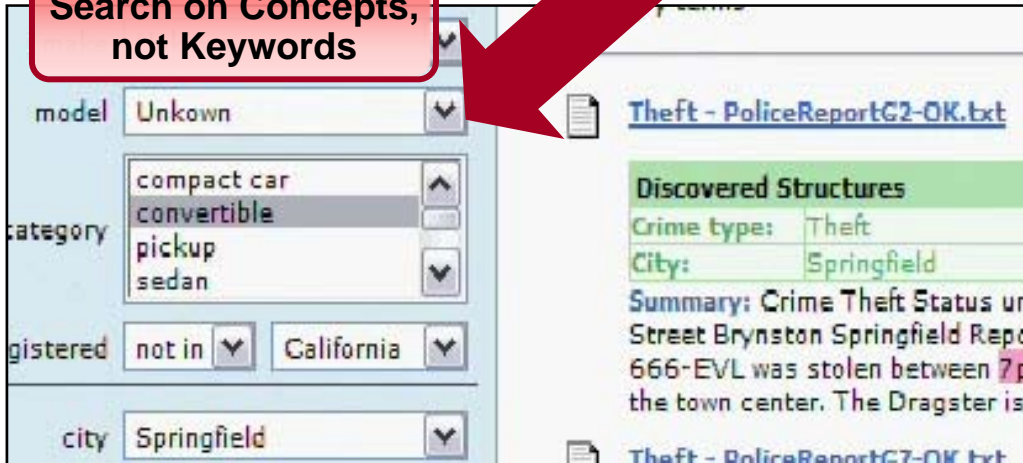*Externally-facing*

OmniFind EE/DE vs. Embedded Search
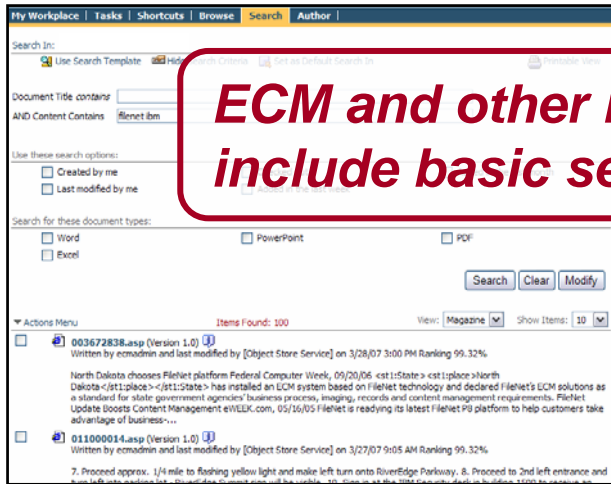
ECM and other IM solutions include basic search
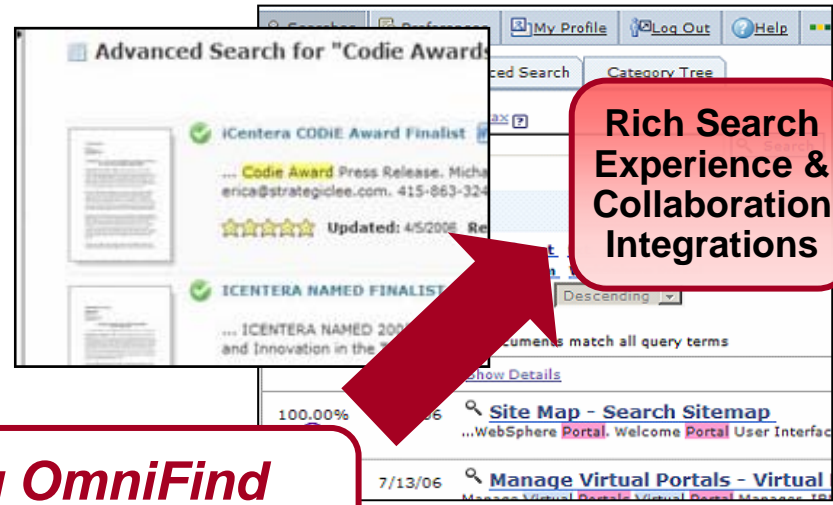
Rich Search Experience & Collaboration Integrations

But adding OmniFind enables much more value

Search on Concepts, not Keywords

Integrate Managed & Unmanaged Content

Act.Right.Now.
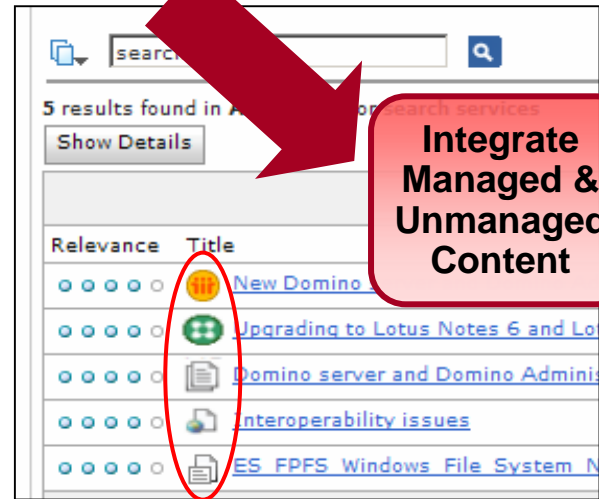
A Starting Point for Taming Unmanaged Content
IBM OmniFind Yahoo! Edition – <u>No-Charge</u> Search for Web and Files

- **Breakthrough simplicity in no-charge basic
  site and intranet search for SMB and departments**

  - **Easy to deploy and configure** – in as
    little as 5 minutes

  - Customizable **Yahoo! Search**
    interface that **unifies enterprise and
    web info**

  - Built on **Lucene** open-source
    technology providing best of
    community and IBM innovation

  - **Try it!** No-charge download
    http://omnifind.ibm.yahoo.com
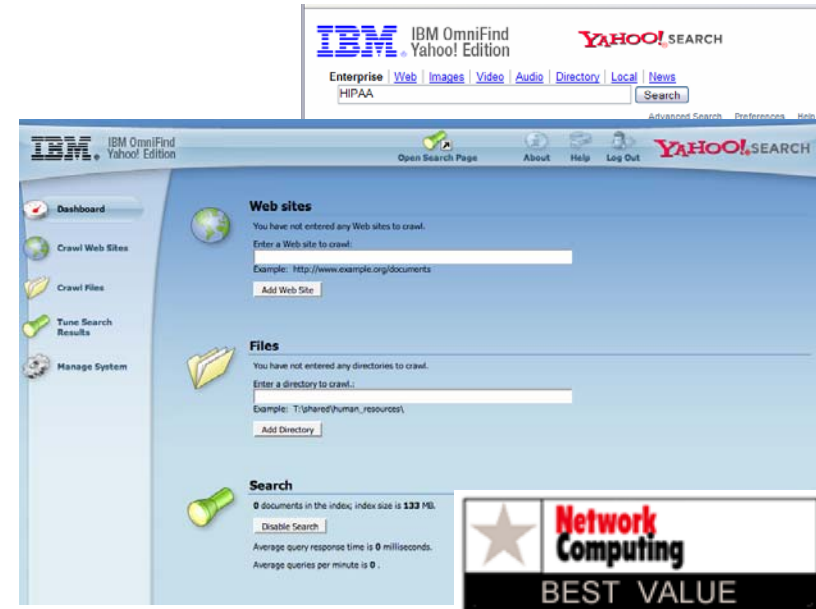
Decision Critical uses OmniFind Yahoo! Edition to provide healthcare institutions
with rapid access to key medical training resources

A **growing partner ecosystem** delivers value-add solutions built upon
OmniFind Yahoo! Edition

Act.Right.Now.

Topics

- Driving value from content with Content Discovery

- The three pillars of Content Discovery
  - Search
  - Classification
  - Analytics-driven Business Insight

- Take Action: Roadmap for leveraging Content Discovery

Act. Right. Now.

Where Do I Start?

Organizing the explosion of unstructured content becomes critical:

- We've got 600 GB of content from basic content services all over the enterprise.
  How can we get this content efficiently mapped into our ECM taxonomy?

- We've been managing our content without classifying it for a few years now.
  How can our users navigate amongst this existing content in a way that's intuitive for *our* business?

- The lawyers have to review 400,000 electronic documents for their case.
  How can we make sure they don't waste their time?

Act. Right. Now.

Classification: Invest to turn Chaos into Order

Accelerate the time to value in your investment in ECM

Ensure more accurate content catalogs

Free up your subject matter experts

Make your content easier to find and leverage

CLOUDMARK

Cloudmark's use of IBM Classification Module made a return on their investment in three months

Act. Right. Now.
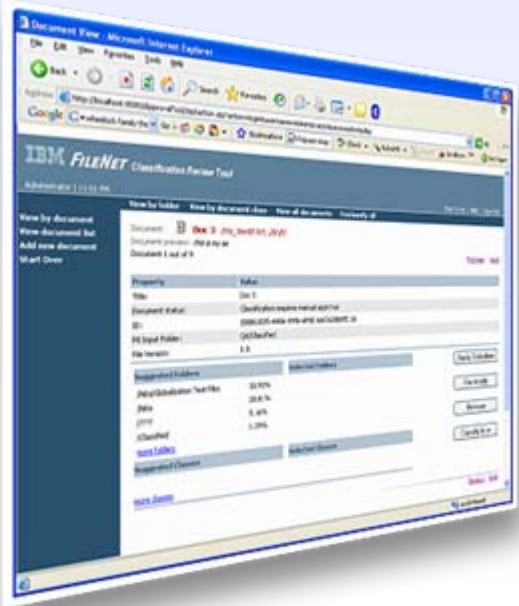
Business Value of Classification

Key Business Drivers			
ECM Taxonomy and Classification **1**	**Compliance, Records, Legal Discovery** **2**	**In Process Classification** **3**	**Message Tagging, Classification and Monitoring** **4**
Increase accessibility of content under management	**Increase legal discovery review effectiveness while reducing risk**	**Increase worker productivity and automate content related decisions**	**Reduce inquiry costs, automate message routing and increase customer satisfaction**
▪ Automated, High Scale Classification ▪ Classify at ingestion and/or re-classify over time ▪ Taxonomy Evolution Tools ▪ Enhanced Accessibility ▪ Taxonomy Proposer	▪ Legal Discovery Prioritization and Workflow Assignment ▪ Records Classification and Exception Handling ▪ Storage and Retention Policy Assignment	▪ Ad Hoc Category Suggestion ▪ Content-Based Workflow Selection ▪ Content Based Decision Making	▪ Email, Chat Routing ▪ Agent Response Suggestion ▪ Email Supervision and Monitoring ▪ Automatic Customer Response

Act. Right. Now.

IBM Classification Module for ECM

IBM Products and Solutions			
ECM Taxonomy and Classification **1**	Compliance, Records, Legal Discovery **2**	In Process Classification **3**	Message Tagging, Classification and Monitoring **4**
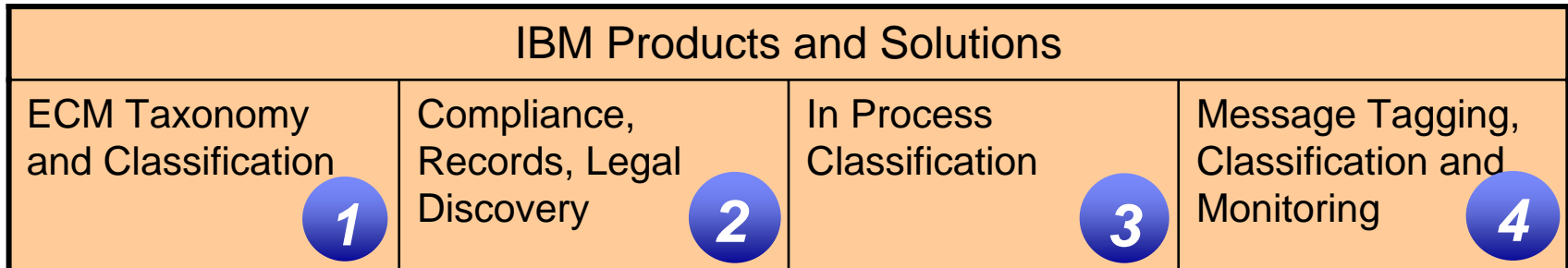
IBM Classification Module

- Automated classification and filtering
- Combines text analytics understanding with rules
- Acquires domain specificity from your own content
- Unique learning technology for adaptive classification
- Suggests new categories or even seeds an entirely new taxonomy
- Rectifies conflicting taxonomies
- Market proven, scalable platform

Act.Right.Now.

Understanding Your Content with Text Analytics
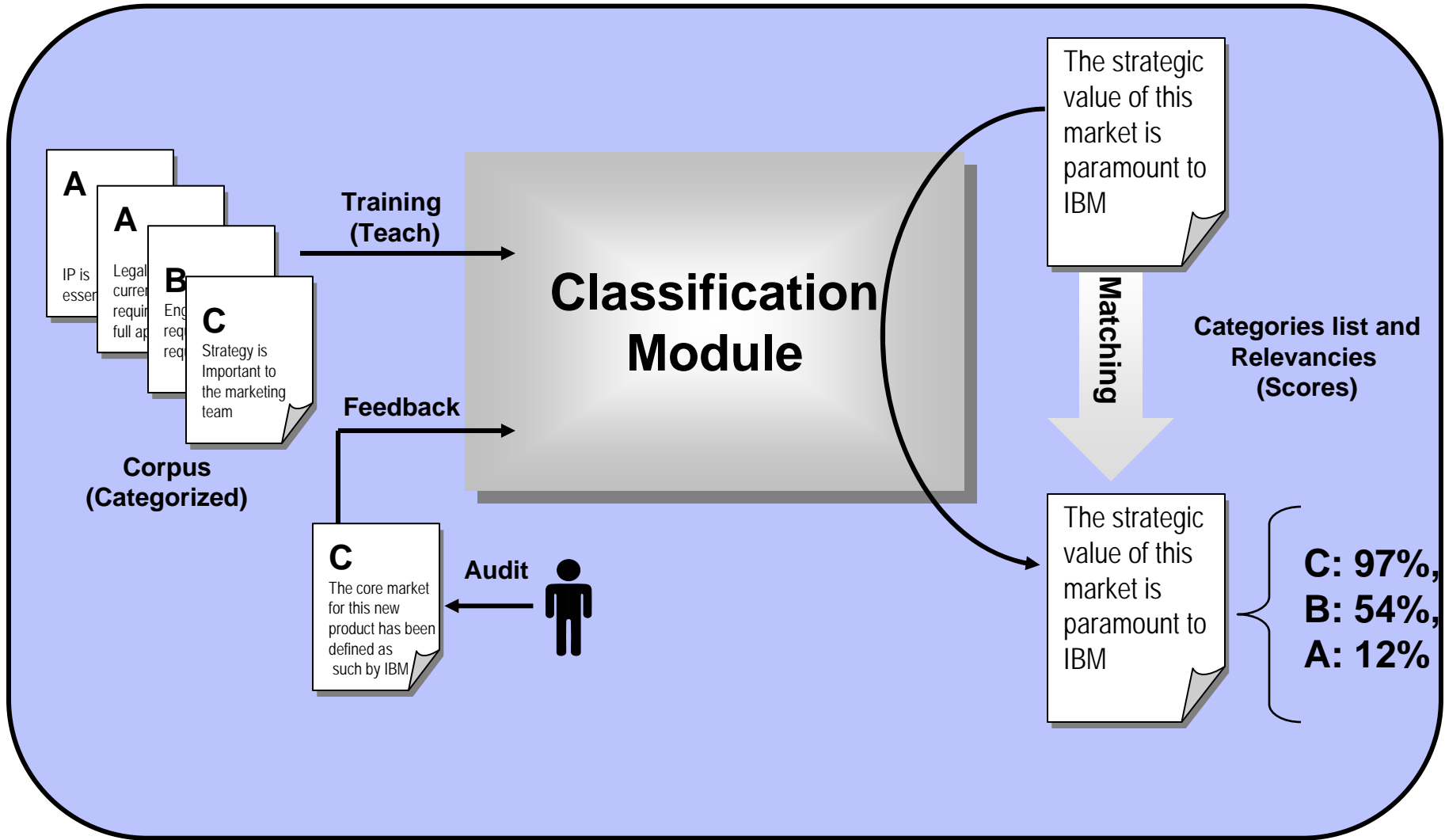
Example Classification Flow: Accelerating Content Ingestion

File Shares

Basic Content Services

e-mail
(future integration)

IBM Classification Module

Review & Audit, Taxonomy Proposer

Document Class: Marketing

Document Class: Legal

Document Class: Finance

Document Class: Research & Development

Records Management

Business Process

e-mail/ Content Archive

Act. Right. Now.
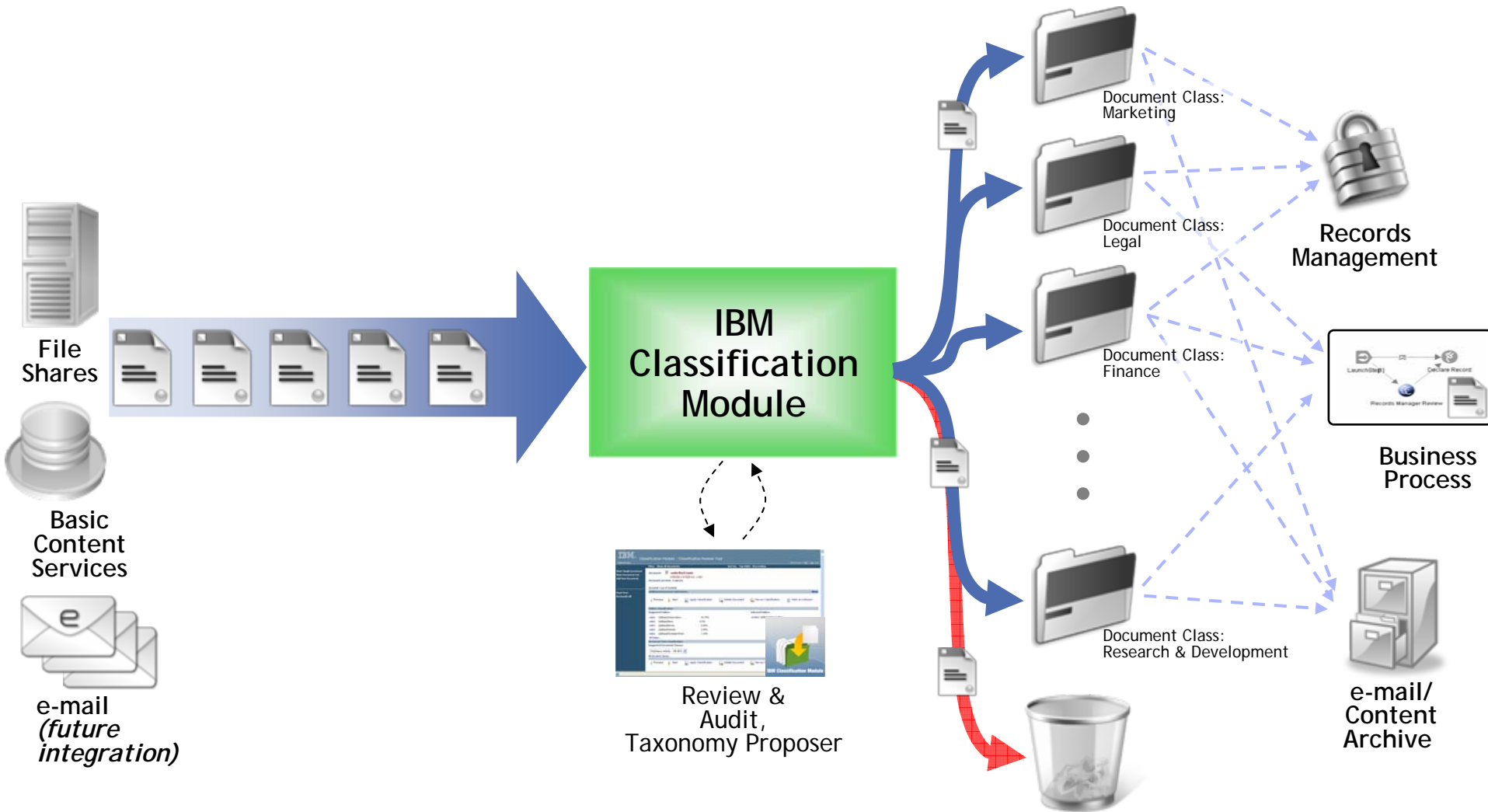
Example Classification Flow: Reclassification

Folder: Finance

IBM Classification Module

Review & Audit

Folder: Finance/Accounting

Folder: Finance/SEC Filings

Folder: Finance/Staff Meetings

Folder: Finance/Quarterly Earning Reports

Act. Right. Now.
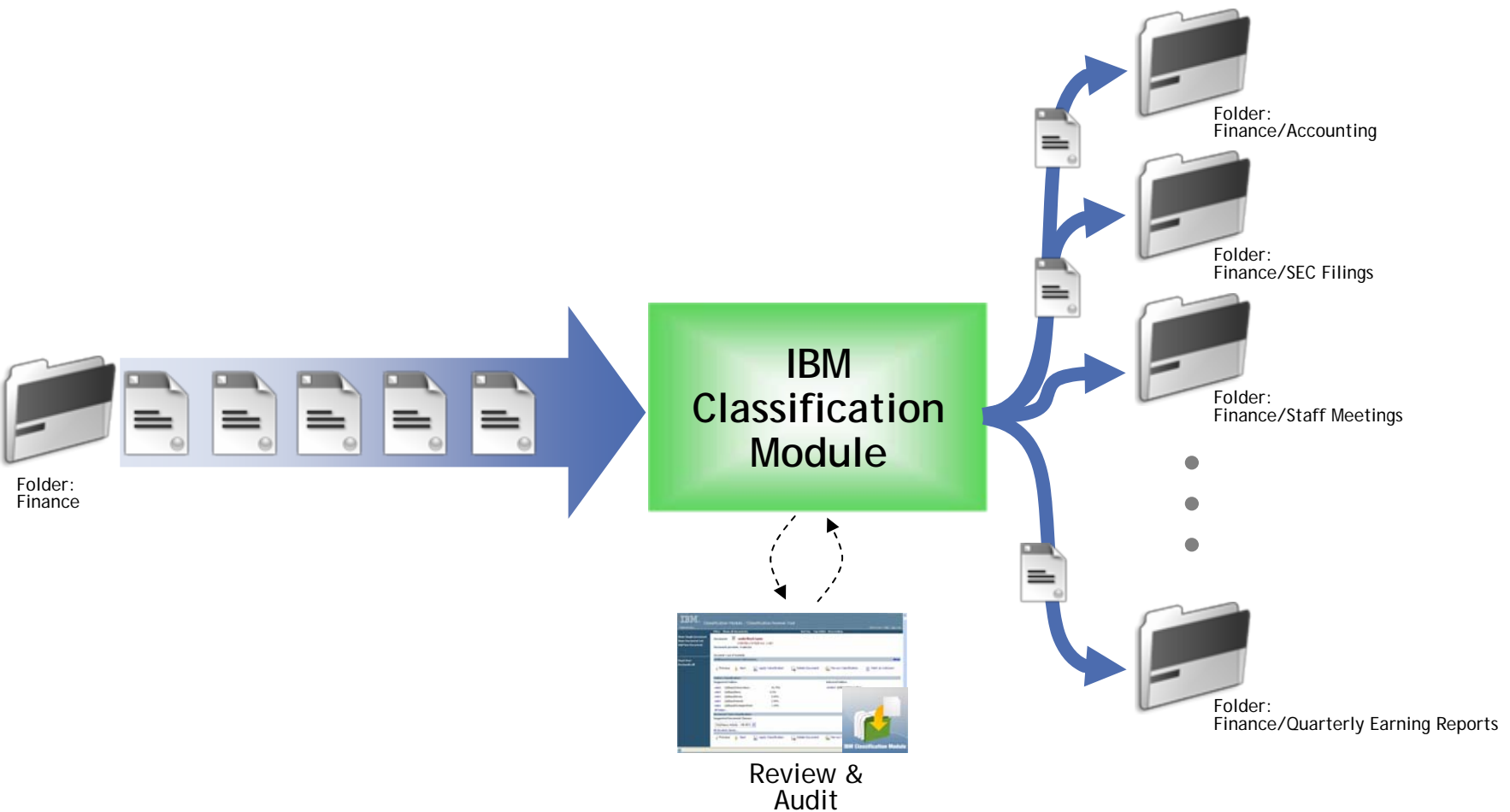
Cost Savings from Automation: Records Declaration Example

According to a recent study by Cohasset Associates, the typical enterprise spends 15 seconds declaring each record, by staff costing $75K/year, leading to – the law of small numbers:

A cost of 15 cents per declaration	at an example enterprise of 1000 employees	each declaring 18 records per day	over the course of 250 working days a year	Results in almost $170,000 in potential cost savings per month

ROI can be made in just a few months!

Act.Right.Now.

Summary: IBM Classification Module is a Platform for Bringing Order to Content Chaos
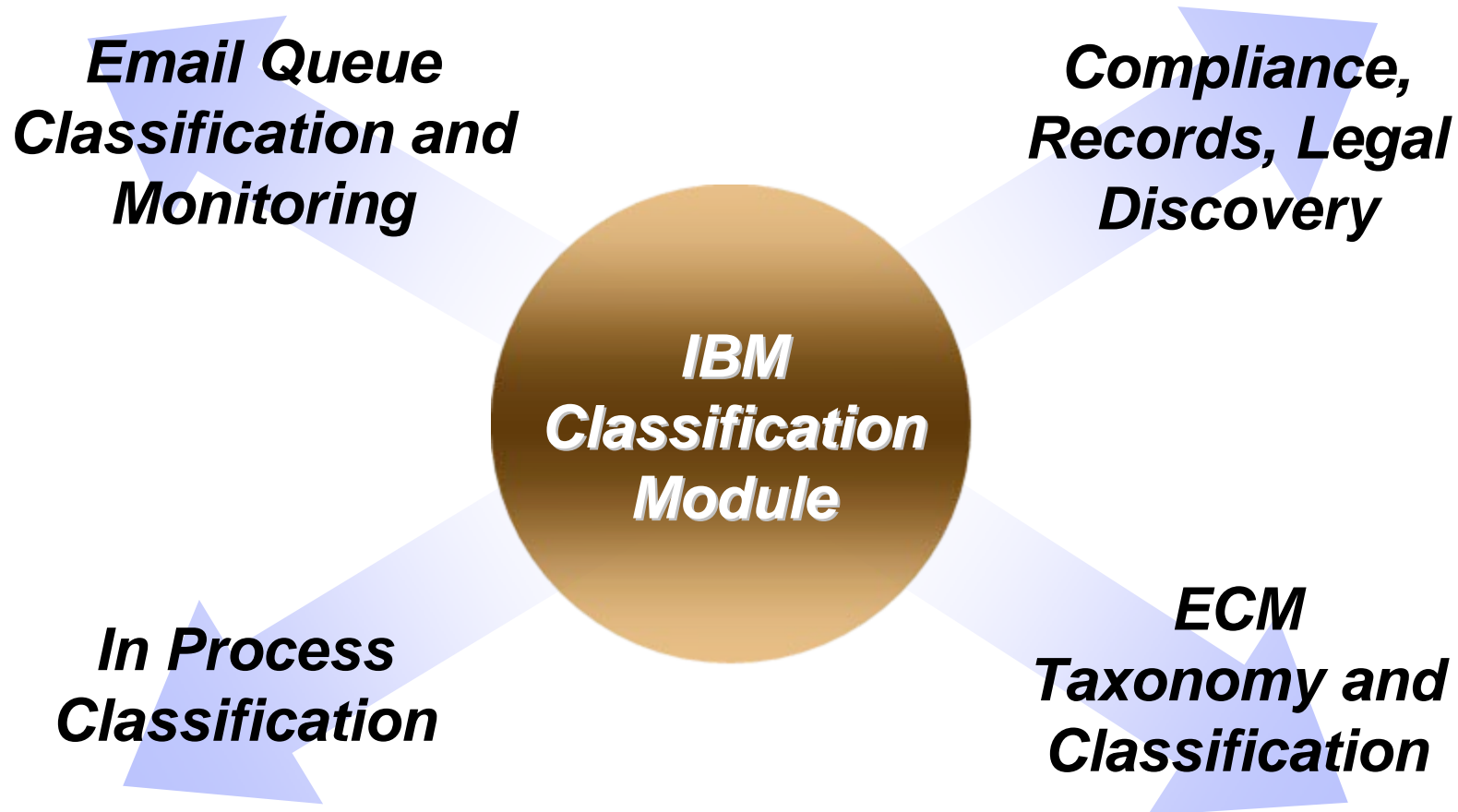
Email Queue Classification and Monitoring

Compliance, Records, Legal Discovery

IBM Classification Module

In Process Classification

ECM Taxonomy and Classification

Act. Right. Now.

Topics

- Driving value from content with Content Discovery

- The three pillars of Content Discovery
  - Search
  - Classification
  - Analytics-driven Business Insight

- Take Action: Roadmap for leveraging Content Discovery

Act. Right. Now.

Delivering Active Business Insight from Content

- What business insights are locked away in your content repositories?

- What if your ECM solution could understand the <u>meaning</u> of unstructured content?

  - Help identify trends, patterns, and correlations?

  - Proactively detect potential fraud, liability, or product or service issues?

  - Help discover new markets and opportunities?

- **You could discover new business insights, differentiate from your competitors, and improve business results**

Act.Right.Now.
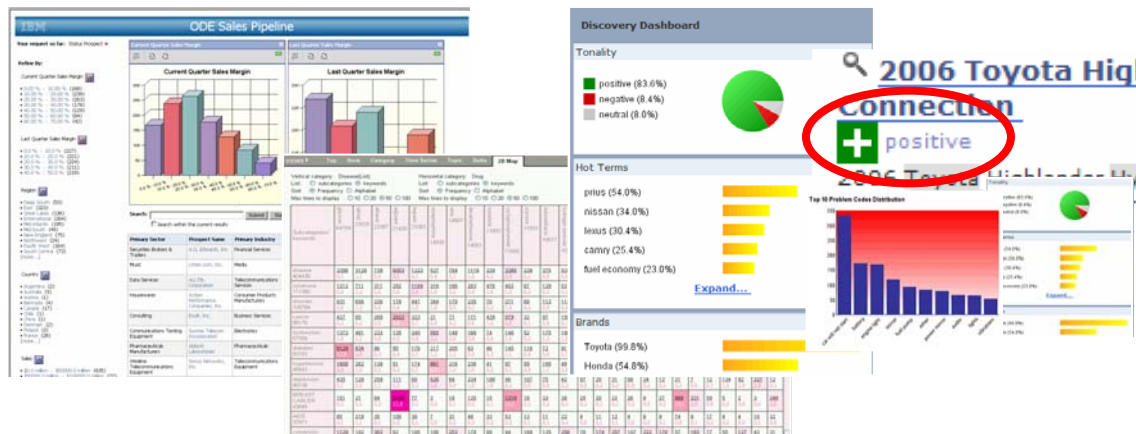
Enabling Capability: Content Analytics

Unstructured Content → **Text Analysis** → **Index/DB/ Catalog** → Search

ECM Apps

Interactive Mining & Viz.

Example

</Organization>   <Located-in>   </Name>

The **associated press** reported today that, **Ms Brown,** CEO of Textract Corp, had been recently spotted at the Summit meeting in **Zurich**,…………At, **42**, Ms. Brown, is the youngest **CEO** at the Summit,…

</Occupation>   </Is-a>   </Age>

Act.Right.Now.
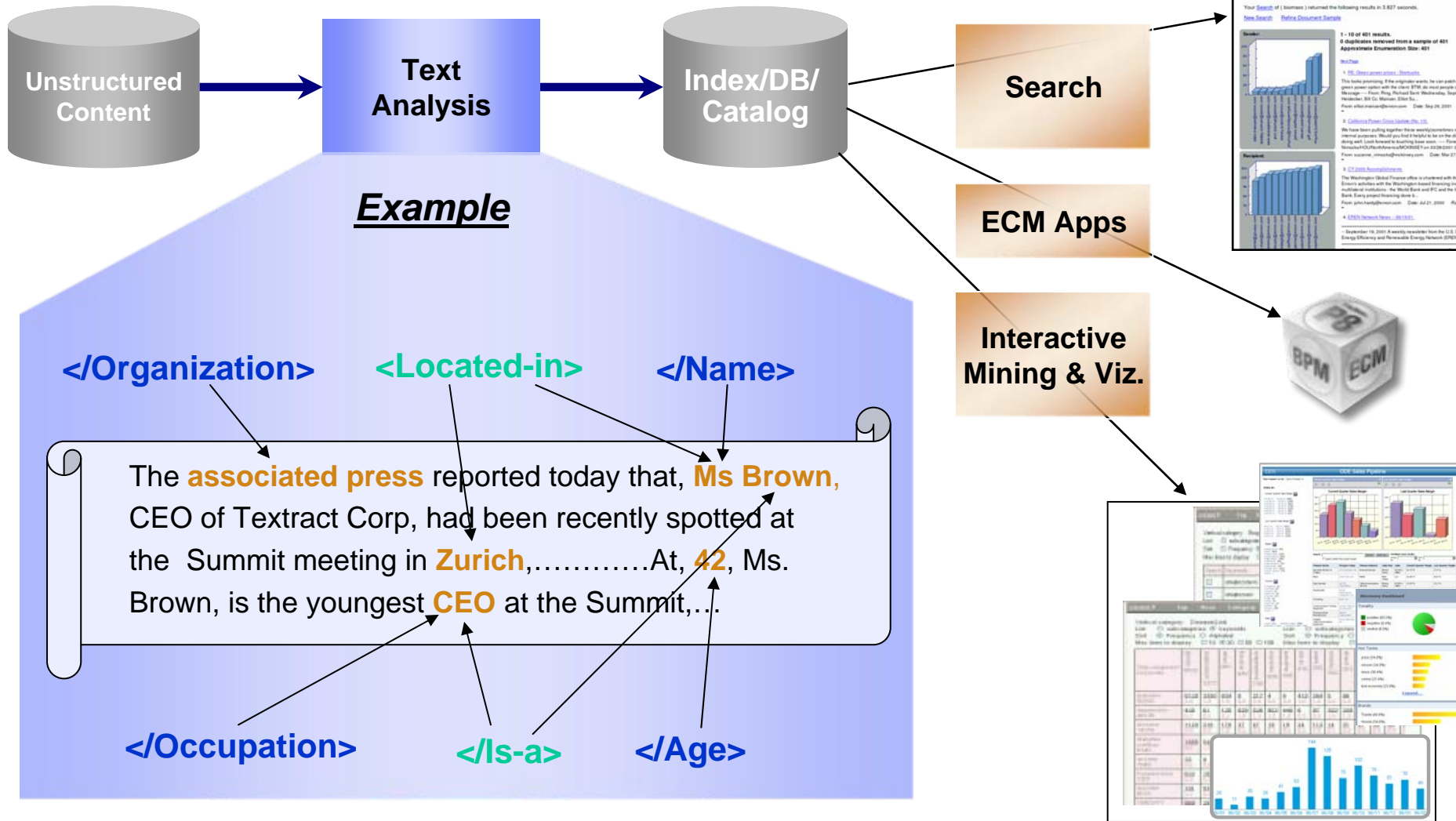
The Importance of Insight from Content

- Insight from content delivers key business value:

Industry	Content analyzed	Enabling...	Business value
Manufacturing	Technician notes, warranty claims	Proactive quality management	Reduced warranty costs & improved customer satisfaction
Insurance	Claims materials	Proactive fraud & claim trend detection	Proactive business response to fraud & shifts in claim workload
Finance	E-mails	eDiscovery for potential fraud	Reduced legal and financial exposure
Law Enforcement	Police records, 911 calls, statements, ...	Rapid crime solving and crime trend analysis	Safer communities and optimized force deployment
Media, FSS, Government	Public opinion sources (news, blogs, forums)	Customer sentiment analysis	Improved products & increased marketing effectiveness
Pharma, Life Sciences	Patent and scientific literature repositories	Improved drug development strategy	Higher return on drug R&D investment
Healthcare	Provider, patient, treatment information	Dashboard view of providers	Stronger negotiating position with providers
Telecom, FSS	Customer calls and e-mails	Churn prediction and FAQ auto-generation	Improved customer retention

30

Act. Right. Now.

Business Insight Case Study: Media & x-Industry

- A media company analyzes public sentiment to determine customer response and marketing effectiveness

Business Insight Case Study: Financial

- A financial services company analyzes and visualizes emails for compliance monitoring and forensics
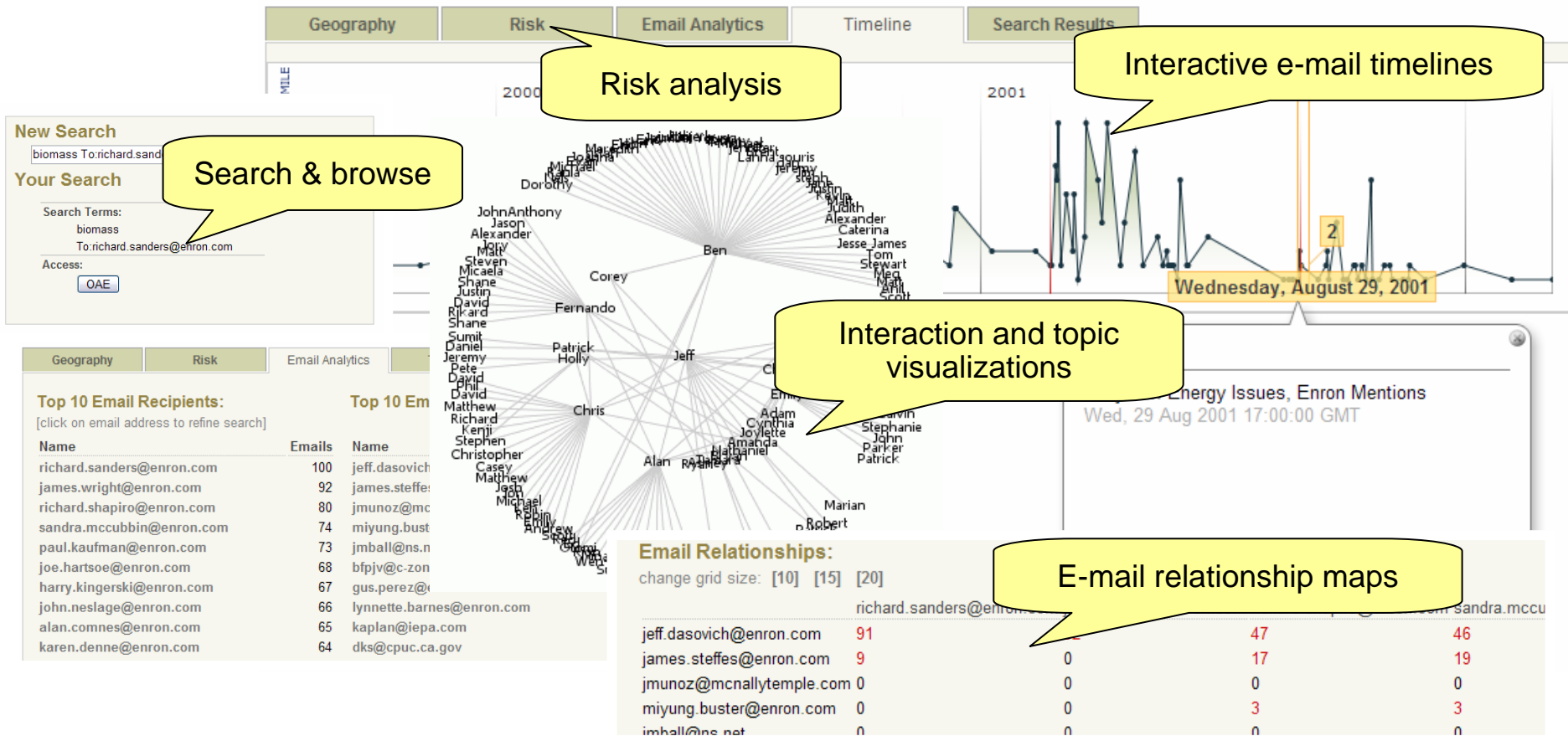
Business Insight Case Study: Telecom

- A major mobile phone carrier optimizes customer retention, improves offerings, and reduces support costs by mining customer contact transcripts

Customer

Voice
Email
Web

Employees

FAQs

VOC

Other sources

FAQ

Business System

Analysis System

Act. Right. Now.

Business Insight Case Study: Healthcare

- Healthcare insurance companies are starting to optimize coverage models by mining and correlating treatments, outcomes, and costs from unstructured records

VIEWS ▶ | Top | Docs | Category | Time Series | Topic | Delta | **2D Map**

Vertical category: Disease(List) Horizontal category: Drug
List: ○ subcategories ⦿ keywords   List: ○ subcategories ⦿ keywords
Sort: ⦿ Frequency ○ Alphabet   Sort: ⦿ Frequency ○ Alphabet
Max lines to display: ○10 ○20 ⦿50 ○100   Max lines to display: ○10 ○20 ⦿50 ○100

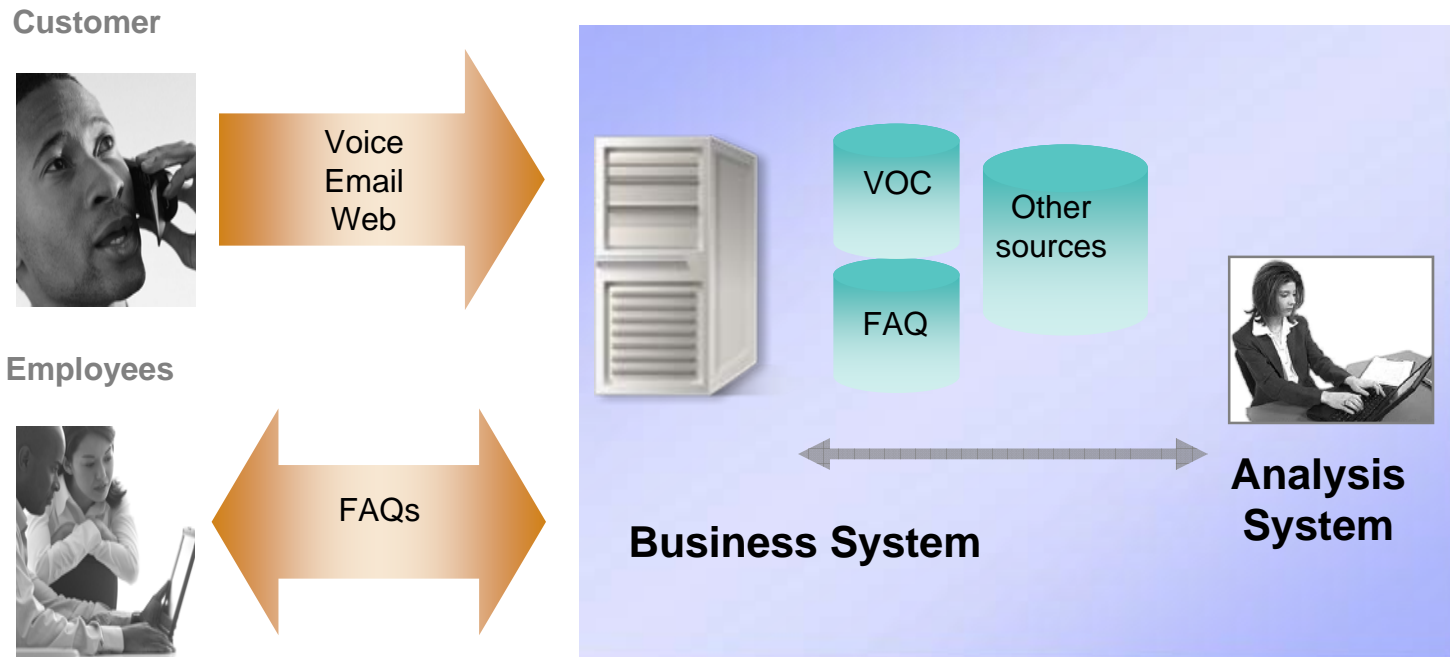| Subcategories/keywords	glucose 64154	retinal 23525	ethanol 23387	adjuvant 21420	heparin 21083	norepinephrine 14935	urea 14807	dexamethasone 14001	morphine 11693	cyclophosphamide 11656	sucrose 11503	indomethacin 10517	hydrogen peroxide 10256	aspirin 9895	methanol 9604	cyclosporine 9147	forskolin 8942	glycerol 8629	EDTA 8614	doxorubicin 8481	retinoic acid 8401	agarose 8339	epinephrine 8280	carbon dioxide 7710	nicotine 7531	etoposide 7306
disease 404438	3398 / 0.7	3128 / 1.7	739 / 0.4	6053 / 3.7	1223 / 0.7	527 / 0.4	784 / 0.7	1116 / 1.0	220 / 0.2	3380 / 3.8	239 / 0.2	375 / 0.4	531 / 0.6	1339 / 1.7	139 / 0.2	1249 / 1.7	151 / 0.2	166 / 0.2	327 / 0.4	1553 / 2.3	409 / 0.6	391 / 0.6	354 / 0.5	348 / 0.5	422 / 0.7	1788 / 3.1
syndrome 111380	1212 / 0.9	711 / 1.3	311 / 0.6	292 / 0.6	1104 / 2.4	316 / 0.9	166 / 0.5	283 / 0.9	476 / 1.8	402 / 1.5	67 / 0.2	129 / 0.5	53 / 0.2	658 / 3.0	24 / 0.1	226 / 1.0	34 / 0.1	48 / 0.2	42 / 0.2	102 / 0.5	212 / 1.0	43 / 0.2	201 / 1.0	86 / 0.4	145 / 0.8	119 / 0.6
disorder 105784	831 / 0.6	656 / 1.3	230 / 0.4	178 / 0.4	447 / 1.0	344 / 1.0	170 / 0.5	235 / 0.7	70 / 0.2	271 / 1.0	68 / 0.2	112 / 0.4	113 / 0.4	215 / 0.9	44 / 0.2	207 / 1.0	43 / 0.2	79 / 0.4	56 / 0.2	39 / 0.2	102 / 0.4	78 / 0.4	144 / 0.7	100 / 0.5	255 / 1.5	43 / 0.2
cancer 98178	427 / 0.3	85 / 0.2	269 / 0.5	2022 / 3.0	323 / 0.7	21 / 0.0	77 / 0.2	171 / 0.6	428 / 1.8	579 / 2.5	32 / 0.1	97 / 0.4	156 / 0.7	232 / 1.1	78 / 0.3	86 / 0.4	17 / 0.1	44 / 0.2	70 / 0.3	699 / 4.2	218 / 1.2	94 / 0.5	28 / 0.1	57 / 0.3	119 / 0.7	557 / 3.8
dysfunction 57558	1372 / 1.9	481 / 1.7	233 / 0.5	135 / 0.9	240 / 0.9	565 / 3.2	148 / 0.8	166 / 0.9	74 / 0.4	146 / 1.0	52 / 0.3	175 / 1.1	198 / 1.5	218 / 1.8	20 / 0.1	258 / 2.3	93 / 0.8	45 / 0.3	71 / 0.6	115 / 1.0	27 / 0.2	25 / 0.2	183 / 1.7	109 / 1.1	87 / 0.8	58 / 0.5
diabetes 50183	8129 / 13.4	834 / 3.5	96 / 0.3	95 / 0.4	176 / 0.7	217 / 1.3	205 / 1.3	83 / 0.5	46 / 0.3	145 / 1.1	116 / 0.9	72 / 0.6	91 / 0.7	304 / 2.9	19 / 0.1	162 / 1.6	64 / 0.6	123 / 1.2	82 / 0.8		20 / 0.0	49 / 0.4	135 / 1.4	18 / 0.1	37 / 0.4	1 / 0.0
hypertension 49943	1806 / 2.9	262 / 1.0	139 / 0.5	51 / 0.2	174 / 0.7	961 / 5.5	216 / 1.3	236 / 1.6	41 / 0.3	97 / 0.5	59 / 0.4	195 / 1.7	49 / 0.4	474 / 4.7	23 / 0.2	443 / 4.7	71 / 0.6	26 / 0.2	59 / 0.5	12 / 0.1	5 / 0.0	20 / 0.1	315 / 3.6	62 / 0.6	78 / 0.8	14 / 0.1
depression 48128	435 / 0.7	128 / 0.5	259 / 1.1	111 / 0.5	80 / 0.3	526 / 3.6	69 / 0.4	234 / 1.6	198 / 1.6	46 / 0.3	107 / 0.8	75 / 0.4	42 / 0.3	87 / 0.8	25 / 0.2	31 / 0.2	88 / 0.9	24 / 0.2	13 / 0.1	21 / 0.2	7 / 0.0	12 / 0.1	134 / 1.5	92 / 1.0	227 / 2.9	12 / 0.1
BREAST CANCER 43849	151 / 0.2	21 / 0.1	84 / 0.3	2129 / 17.4	77 / 0.3	3 / 0.0	18 / 0.1	125 / 0.9	15 / 0.1	1219 / 12.1	18 / 0.1	33 / 0.3	38 / 0.3	25 / 0.2	20 / 0.1	23 / 0.2	26 / 0.2	8 / 0.0	27 / 0.2	968 / 13.1	221 / 2.8	59 / 0.6	5 / 0.0	2 / 0.0	2 / 0.0	246 / 3.6
AIDS 37971	85 / 0.1	218 / 1.1	35 / 0.1	100 / 0.5	30 / 0.1	7 / 0.0	31 / 0.2	48 / 0.3	33 / 0.3	53 / 0.5	13 / 0.1	11 / 0.1	22 / 0.2	4 / 0.0	11 / 0.1	12 / 0.1	4 / 0.0	8 / 0.0	9 / 0.1	74 / 0.9	9 / 0.1	17 / 0.2	4 / 0.0	4 / 0.0	10 / 0.1	32 / 0.4
conversion	1129	192	363	82	185	106	253	178	89	44	169	135	256	78	174	257	107	222	170	57	163	77	55	127	43	31

IBM OmniFind Content Analytics Solutions
Powered by OmniFind Enterprise & Analytics Editions

- **OmniFind-based solutions enable insight from unstructured content**

  - Sophisticated **natural language extraction** from unstructured content

  - Rich **mining** UI to analyze, correlate, and search unstructured and structured data

  - **Targeted solutions** for Customer Care, Quality Insight, Forensic Compliance, and Public Image/ Media Analysis use cases

  - **Additional solutions** in development

  - Based on **10 years of experience** in transformative analytics engagements from IBM GBS & Research

  - A key part of IBM's **Information on Demand** initiative

  - Solutions leverage various combinations of **OmniFind Analytics Edition*** and **OmniFind Enterprise Edition**
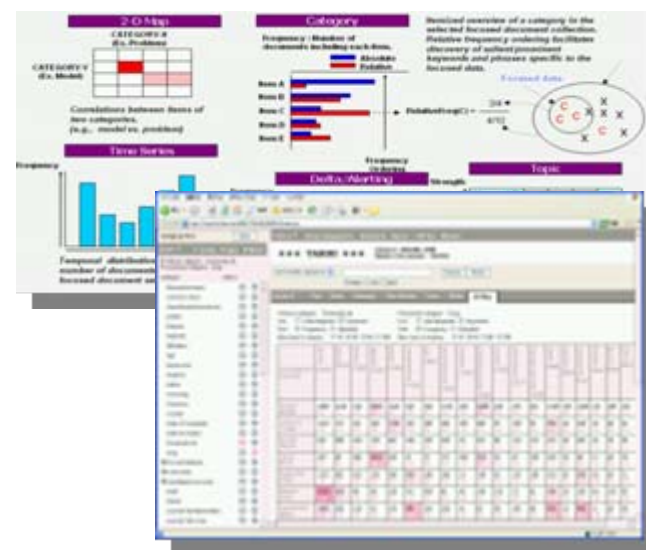
* OmniFind Analytics Edition currently available in North America, UKISA, and Japan only

Act. Right. Now.

Take Action: Get More Value from your Content with IBM Content Discovery Offerings

Improve Content Access

Organize Unstructured Content

IBM OmniFind Enterprise Edition

IBM OmniFind Discovery Edition

IBM OmniFind Yahoo! Edition

Content Discovery Solutions

IBM Classification Module

Derive Business Insight

IBM OmniFind Analytics Edition

IBM OmniFind Enterprise Edition

Partner + Analytics Plugins

SAS

openNLP

CLEARFOREST
TEXT-DRIVEN BUSINESS INTELLIGENCE

inxight

attensity

SPSS

SchemaLogic

TEMIS

nstein

GATE

36

Act. Right. Now.

Recap: The Keys to Leveraging Content

Improved content access...

- **IBM OmniFind Enterprise Edition**
  - Platform for secure, scalable search-based content access
  - Advanced concept/semantic search
- **IBM OmniFind Discovery Edition**
  - Deliver searchable self-service content directly to customers
- **IBM OmniFind Yahoo! Edition**
  - Free and easy way to start taming unmanaged content, complementing ECM

Better organizing content...

- **IBM Classification Module**
  - Proven auto-classification solution
  - Enables new ECM value from taxonomy automation, compliance, BPM and more

Delivering active business insight...

- **Insight solutions based on IBM OmniFind**
  - Leading analytics and content mining capabilities
  - OmniFind Enterprise Edition
  - OmniFind Analytics Edition

Act. Right. Now.

Questions?

Aaron Brown

Program Director, ECM Content Discovery

IBM Information Management Software

abbrown@us.ibm.com

Act.Right.Now.

More Discovery, Classification, & Analytics Sessions – Thursday & Friday

Thursday

Session #	Title	When	Where
HOL-1993A	Automating Document Classification in FileNet P8	09:30 AM - 12:30 PM	Breakers H
TCM-2134A	Secure Enterprise Search for WebSphere Portal and Lotus Domino	10:00 AM - 11:00 AM	Lagoon C
TCM-2136B	Get More Value Out of Your Content with Content Discovery and Search	10:00 AM - 11:00 AM	Lagoon L
TCM-2133A	Search & Text Analytics as Part of an IOD Solution from Warehousing to BI	11:15 AM - 12:15 PM	Lagoon A
TCM-2135A	Making BI as Simple as Shopping	02:30 PM - 03:30 PM	Lagoon C
BOF-2601A	Classification and Taxonomy Automation: Business and Technical Challenges	05:30 PM - 06:30 PM	Lagoon L

Friday

Session #	Title	When	Where
TTF;TCM-2138C	Content Discovery Technology Trends and Directions	08:00 AM - 09:00 AM	Lagoon D

Act. Right. Now.

Appendix

Act.Right.Now.

IBM INFORMATION ON DEMAND 2007
October 14 – 19, 2007
Mandalay Bay
Las Vegas, Nevada

Business Value of Classification – One Example

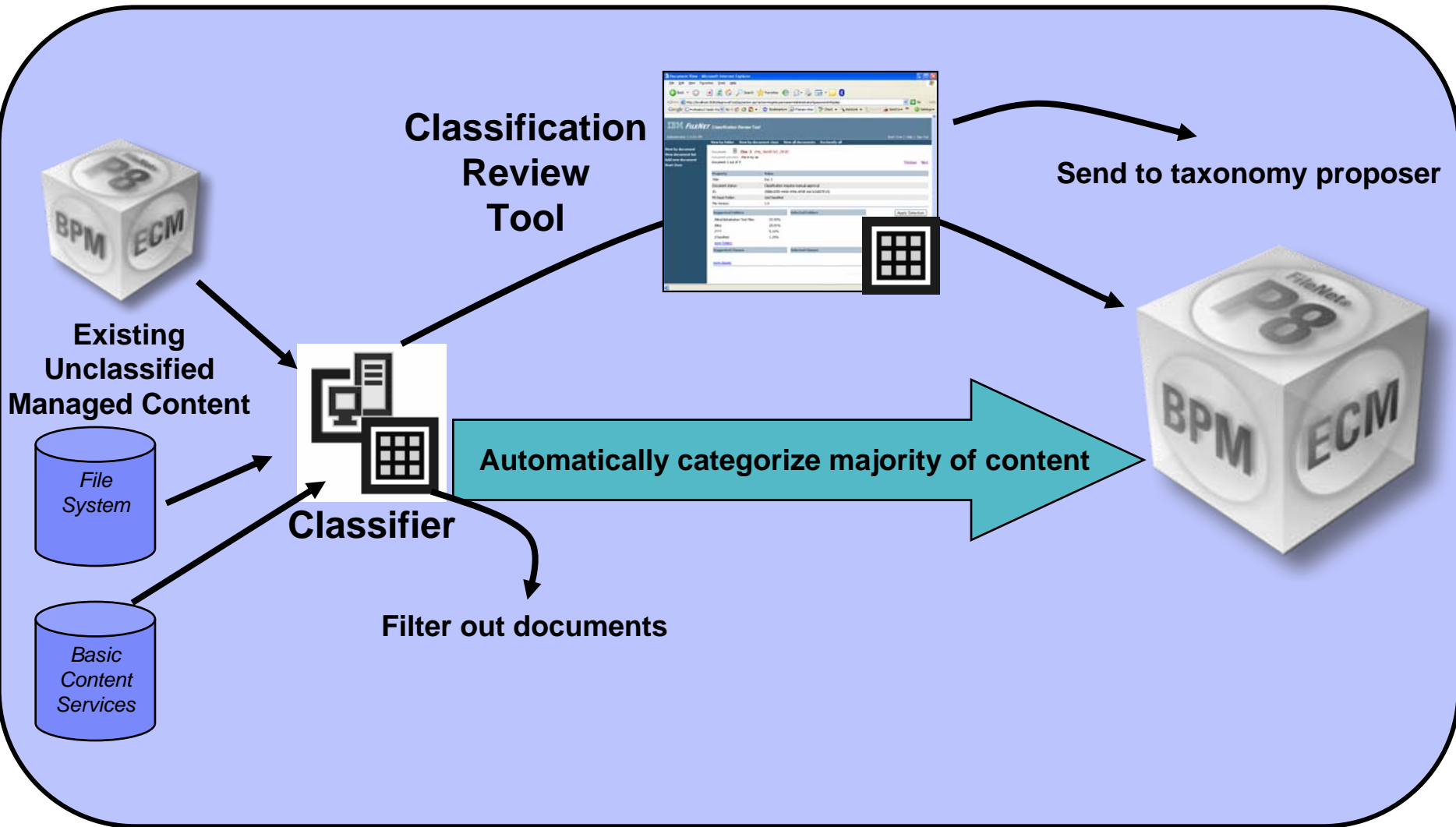| Key Business Drivers	
ECM Taxonomy and Classification ⓵	**Most organizations face content taxonomy pains – especially as they standardize around ECM** – Mapping content to taxonomy or records plan during ingestion – Reclassifying content under management – Evolving taxonomies as new types of content emerge – Integrating folksonomies (SharePoint) into a master taxonomy
Increase accessibility of content under management ▪ Automated, High Scale Classification ▪ Classify at ingestion and/or re-classify over time ▪ Taxonomy Evolution Tools ▪ Enhanced Accessibility ▪ Taxonomy Proposer	

Act. Right. Now.

Example Classification Workflow: Accelerating Content Organization

Classification Review Tool

Send to taxonomy proposer

Existing Unclassified Managed Content

File System

Basic Content Services

Classifier

Automatically categorize majority of content

Filter out documents

Act. Right. Now.

Our View: The Information Access Market

Access ←————————————————————————————————→ **Actionable Insight**

Component	Simple Utility	Infrastructure	Experience-Driven Solutions	Insightful Solutions

Basic, Unsecured Web Site Or Departmental File System Search

Component Search Engine Or Libraries

- OEM

- Simple site search
- SMB/basic intranet

Scalable, Secure Enterprise Search With Broad Reach To Many Content Sources

- Secure intranet search
- Integration of enterprise content
- Intranet portals
- Platform for search applications

Tunable, Business Controlled Search With Application-specific Interfaces, And Reporting

- Self Service
- e-Commerce
- Contact Centers

Solutions that deliver breakthrough Business Insight by Discovering, Analyzing, Mining, & Integrating Unstructured and Structured Enterprise Information

- Customer Care & Customer Insight
- Quality Early Warning
- Public Image Monitoring
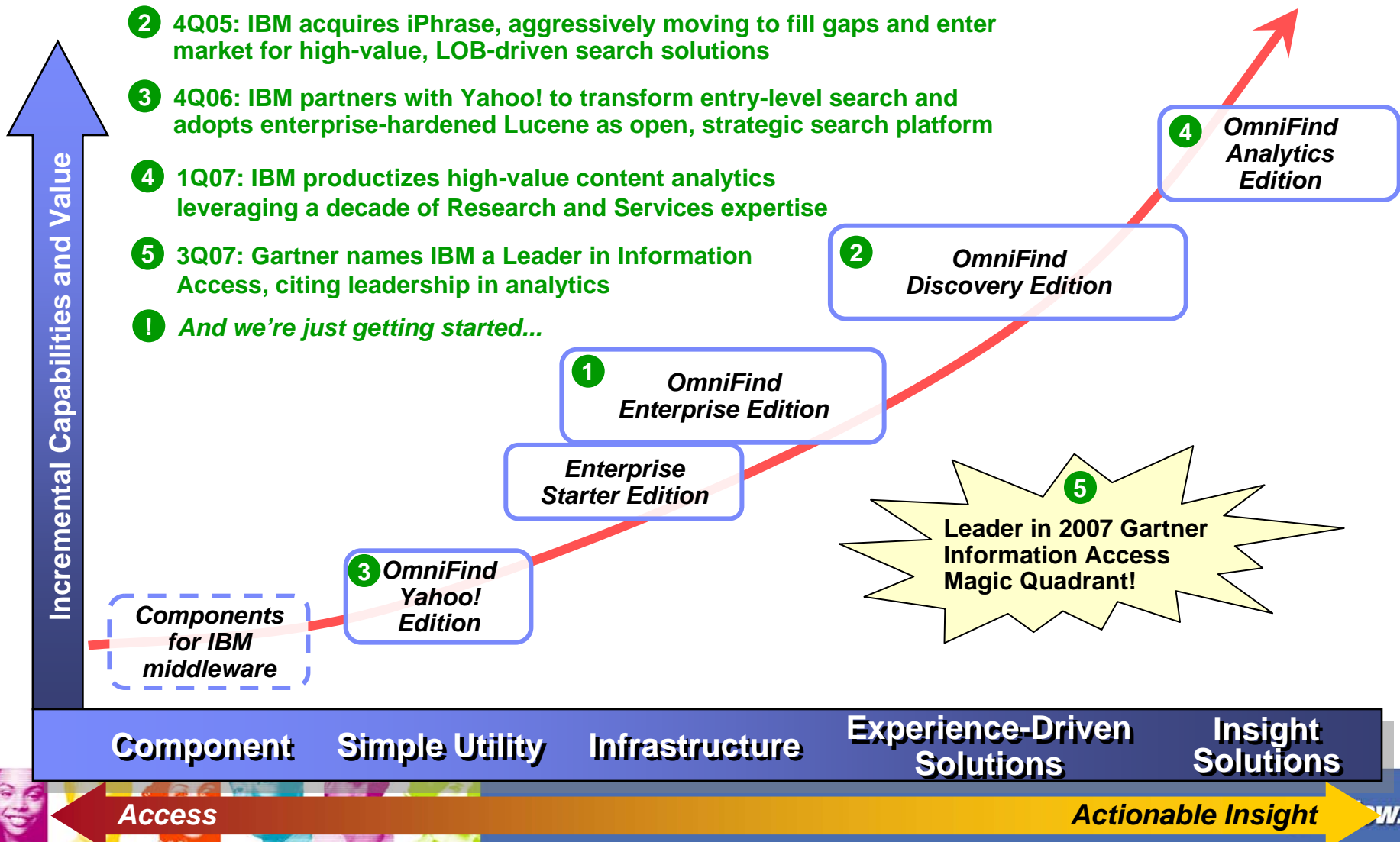- Compliance & Legal Discovery
- BI for the Masses

44

IBM OmniFind Offers Solutions for Every Market Need
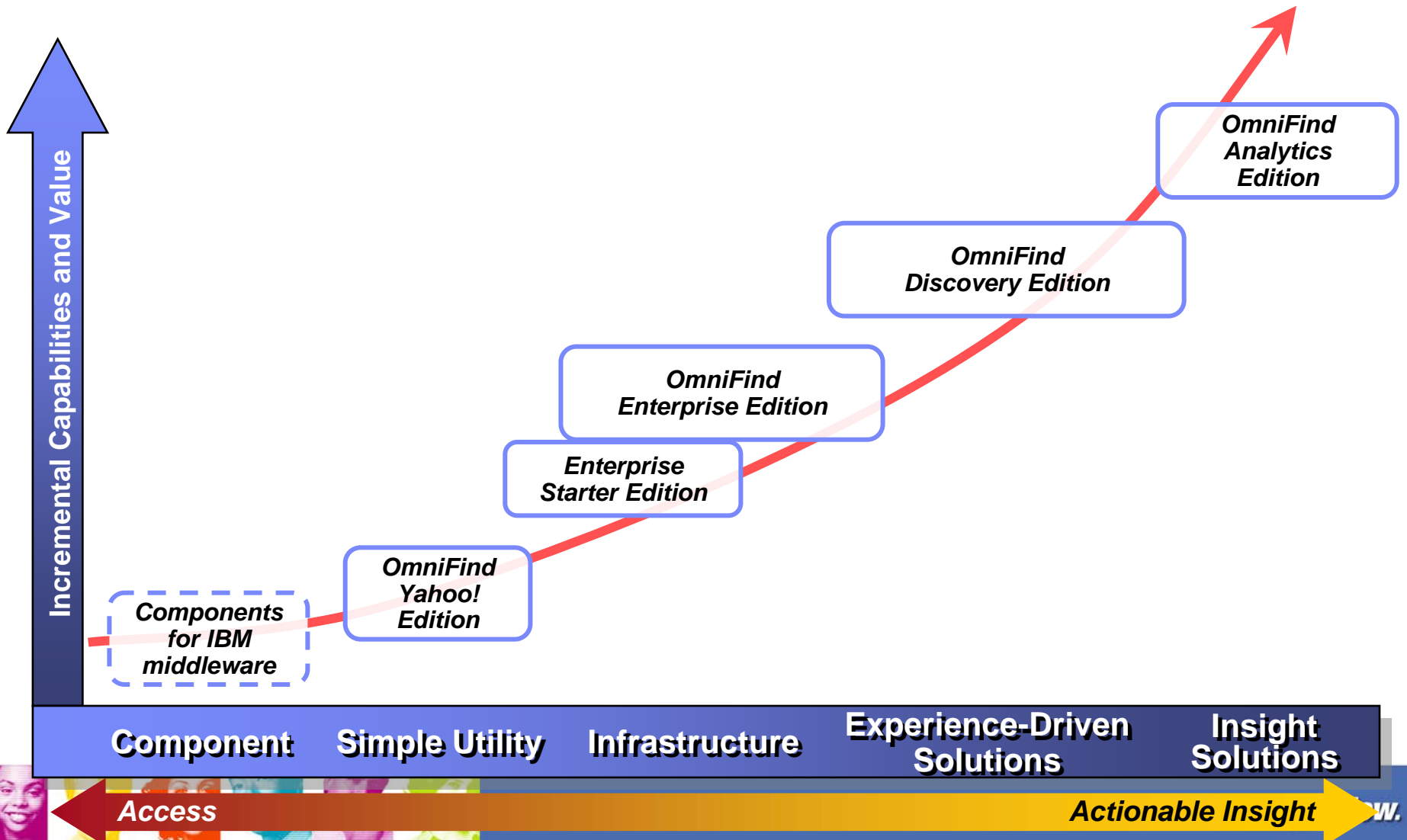
The Evolution of OmniFind: IBM's Thrust for Leadership

① **4Q04:** IBM enters enterprise search market with home-grown OmniFind EE

② **4Q05:** IBM acquires iPhrase, aggressively moving to fill gaps and enter market for high-value, LOB-driven search solutions

③ **4Q06:** IBM partners with Yahoo! to transform entry-level search and adopts enterprise-hardened Lucene as open, strategic search platform

④ **1Q07:** IBM productizes high-value content analytics leveraging a decade of Research and Services expertise

⑤ **3Q07:** Gartner names IBM a Leader in Information Access, citing leadership in analytics

! *And we're just getting started...*

Incremental Capabilities and Value

④ *OmniFind Analytics Edition*

② *OmniFind Discovery Edition*

① *OmniFind Enterprise Edition*

Enterprise Starter Edition

⑤ **Leader in 2007 Gartner Information Access Magic Quadrant!**

③ *OmniFind Yahoo! Edition*

Components for IBM middleware

| Component | Simple Utility | Infrastructure | Experience-Driven Solutions | Insight Solutions |

Access — *Actionable Insight*

Search History

- **IBM has been involved in search since the 1970's**
  - 1970s: System 370 text search offering (STAIRS) defined keyword search
  - Continued investment in research and product through present day
  - 2004: IBM re-enters the modern search market
  - 2004 – 2007: Rapid acceleration of IBM's search offerings
    - 2004: OmniFind Enterprise Edition introduced, powers IBM intranet
    - 2005: IBM acquires iPhrase, broadens portfolio capabilities and enters market for high-value, LOB-driven search solutions with OmniFind Discovery Edition
    - 2006: IBM partners with Yahoo! and releases OmniFind Yahoo! Edition, enabling every organization to get started with enterprise search at no charge
    - 2007: OmniFind Analytics Edition expands high-end analytics capabilities

- **Today: IBM offers a complete portfolio of search solutions**
  - Leading security, content reach, and integrations with IBM middleware
  - Leading content analytics capabilities
  - IBM continues to invest heavily in leading-edge search innovation

- **Recognized by Gartner as a Leader in the 2007 Information Access Magic Quadrant**

Act. Right. Now.

Classification History

- Banter develops leading-edge text classification software

- iPhrase acquires Banter in 2004
  - Banter technology marketed primarily for customer interaction use cases such as contact center routing & e-mail response

- IBM acquires iPhrase in 2005
  - Continues to ship classification technology for customer service and contact center use cases

- IBM acquires FileNet in 2006
  - Portfolio assessment reveals new synergies in content classification

- *2007: IBM ECM announces new content classification solutions for ECM customers*
  - Unlocking new value via IBM ECM's new leadership position
  - Leveraging leading technology with a long proven history in-market

New!

Act.Right.Now.

IBM has delivered analytics value for 10+ years!

- Mid-1990s: IBM begin delivering first-of-a-kind content analytics solutions
  - Delivered by IBM Research and Services, offered early adopters a taste of transformative business insight from unstructured data
- 1998: TAKMI project first delivered by IBM Research – precursor to OmniFind Analytics Edition
  - Advanced text analytics and mining tools for unstructured data
- 2004: Unstructured Information Management Architecture (UIMA) framework released to market
  - Open platform and SDK for content analytics
- 2004: OmniFind Enterprise Edition productizes UIMA
  - Becomes basis for commercial deployments of analytics-enhanced search solutions
- 2004 – 2007: UIMA adoption grows
  - Widely used by open community, adopted by 15+ software companies
  - Standard architecture for DARPA-funded research
  - Delivered to open community first via SourceForge, then as an Apache incubator project
  - Currently undergoing standardization by OASIS
- 2007: OmniFind Analytics Edition productizes TAKMI
  - Brings advanced UIMA-compliant annotators and mining UI to widespread commercial use

Act.Right.Now.

Business Insight Case Study: Manufacturing

- An automotive company mines technician notes and warranty claims to preemptively address quality problems

Semantic Search and Drilldown Interface
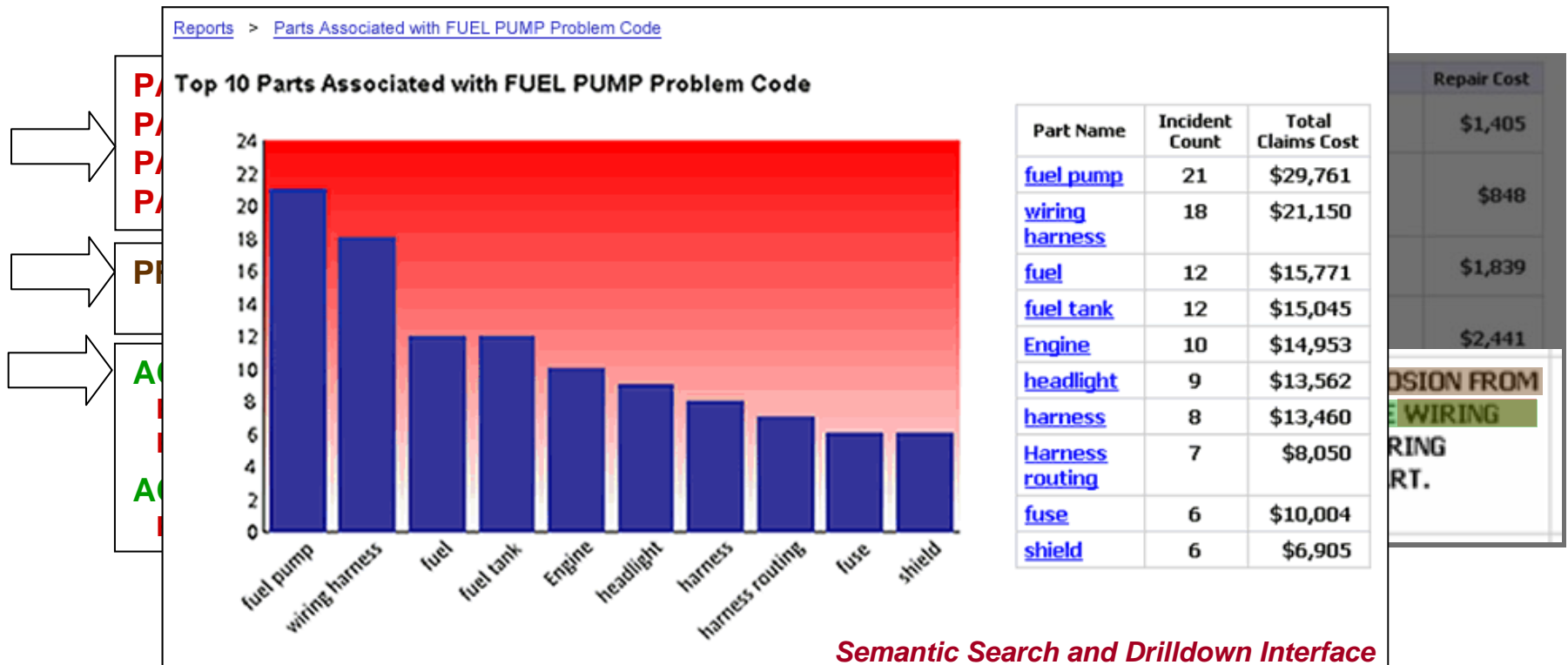
Meaning from Content

- IBM OmniFind solutions use advanced content analytics to extract deep meaning from unstructured content
  - Leveraging decades of advanced R&D and field experience from IBM Research and Global Services
  - Based on open, standard Unstructured Information Management Architecture (UIMA) framework

Behind the Scenes: Sentiment Analysis

Sentiment analysis is enabled by "annotators" that perform linguistic analysis of text in content items:

1. **Recognize words that are "positive" and/or "negative"**

   - bargain, successful, ..., bad, pollution,...

   - One word can have different meaning:
     fine (adjective) vs fine (noun)

   - Based on standard + app-specific word lists

2. **Recognize syntax in order determine the sentiment of a paragraph**

   - A bargain → positive

   - (Is) A bad bargain → negative

   - (Is) **not** a bad bargain → positive

   - Underpinned by sophisticated understanding of linguistic rules

3. **Assign overall sentiment score to paragraphs and content items**
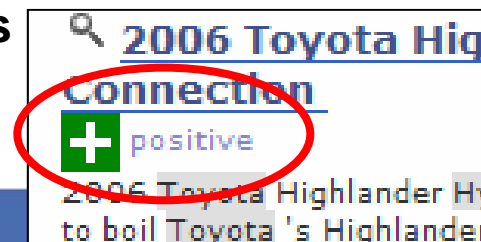
The IBM Difference

- A complete portfolio covering every content discovery need
  - Solve today's business problem...with room to grow tomorrow
  - Open platforms, standards, and APIs enable unprecedented customization and preserve your investments
- Leading content analytics capabilities to maximize business value
  - Go beyond keyword search to improve findability and extract new insight
  - Leverage decades of investment by IBM Research and Global Services
- The best integrations with IBM middleware and content sources
  - Faster time to market with reduced risk
  - Broad support for third-party sources as well, with easy extensibility
- Rapid access to the latest leading-edge search innovation
  - IBM invests ~$50M annually in research and development for search and text analytics (200+ people)
  - IBM holds over 200 patents in search with more each year
  - Our worldwide development team is focused on delivering that innovation to drive value for OmniFind customers

Act. Right. Now.