Security in Knowledge Management

[Insiders as Security Risks]

Dr. Rajat Mukherjee

Principal Architect, Verity, Inc.
rajatm@verity.com

HP WORLD 2002
Conference & Expo

Verity, The Market Leader

"Verity #1 Vendor in Portal Infrastructure by Market Share"
– Delphi Group 5/01

"Verity is the Established Market Leader in Enterprise Search"
– OVUM 7/01

Verity is #17 on Fortune 100 Fastest Growing Companies
– Fortune Magazine 8/01

"Verity is the Market Share Leader in Text Search and Retrieval"
– IDC 9/01 Report

Verity Leader in the Gartner 2002 Enterprise Search Magic Quadrant
– Gartner Group 2/02

Verity Corporate Overview

- 1500+ Corporations
- 175 E-Commerce Sites
- 220 OEM Applications
- 430+ Staff Worldwide
- NASDAQ (VRTY)
- $245 Million in Cash and investments
- 4th Consecutive Profitable Year

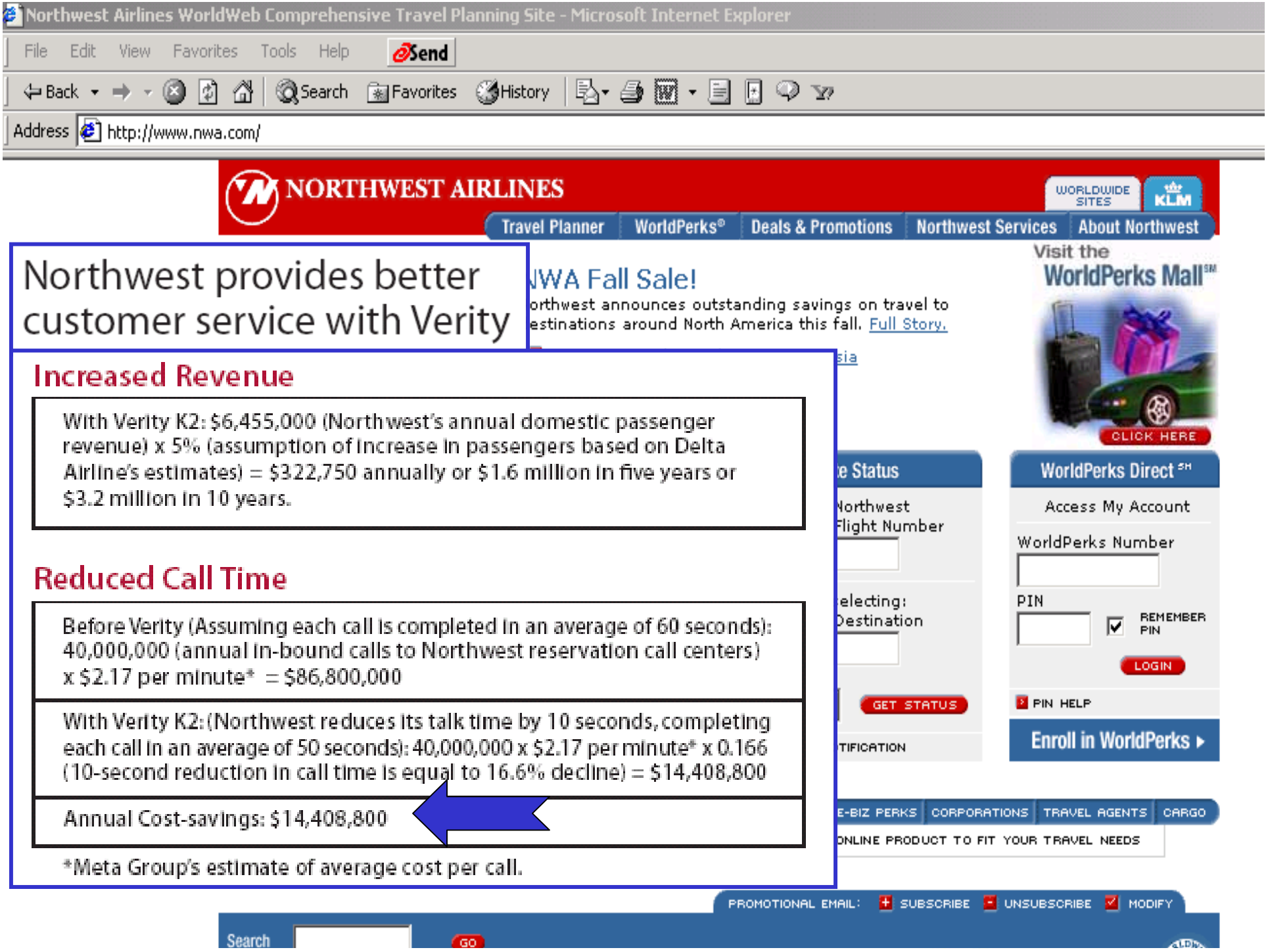
Return On Investment

Knowledge Workers (KW) spend 15% of their time searching for information. – AIIM (Assn for information and Image Mgmt.)

Knowledge Workers spend 25% of their time searching for information. – Ford Motor Company

Companies without a portal "waste $4.8 million per year. As the volume and velocity of document-based information increases, the problem becomes even more severe."

– Gartner Group



Northwest provides better customer service with Verity

Increased Revenue

With Verity K2: $6,455,000 (Northwest's annual domestic passenger revenue) x 5% (assumption of increase in passengers based on Delta Airline's estimates) = $322,750 annually or $1.6 million in five years or $3.2 million in 10 years.

Reduced Call Time

Before Verity (Assuming each call is completed in an average of 60 seconds): 40,000,000 (annual in-bound calls to Northwest reservation call centers) x $2.17 per minute* = $86,800,000

With Verity K2: (Northwest reduces its talk time by 10 seconds, completing each call in an average of 50 seconds): 40,000,000 x $2.17 per minute* x 0.166 (10-second reduction in call time is equal to 16.6% decline) = $14,408,800

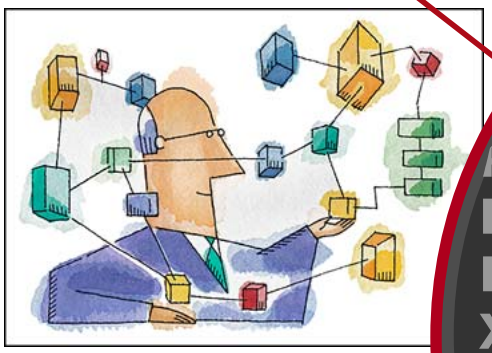
Annual Cost-savings: $14,408,800

*Meta Group's estimate of average cost per call.

The Opportunity

Online Data in the World

4,000,000+ TB



World Wide Web

~50+ TB

Databases

~8,000+ TB

Verity Organizes All This And More …With Security

Document Level Security

Strategic Planning (Upcoming Gartner Report):

- **Through 2005, enterprises that incorporate security as an element of search installations will realize 50 percent greater return on investment than those who do not.**

- **Through 2006, 40 percent of enterprises that choose initially not to include security as an element of a search installation will revise the initiative to do so within 24 months of installation.**

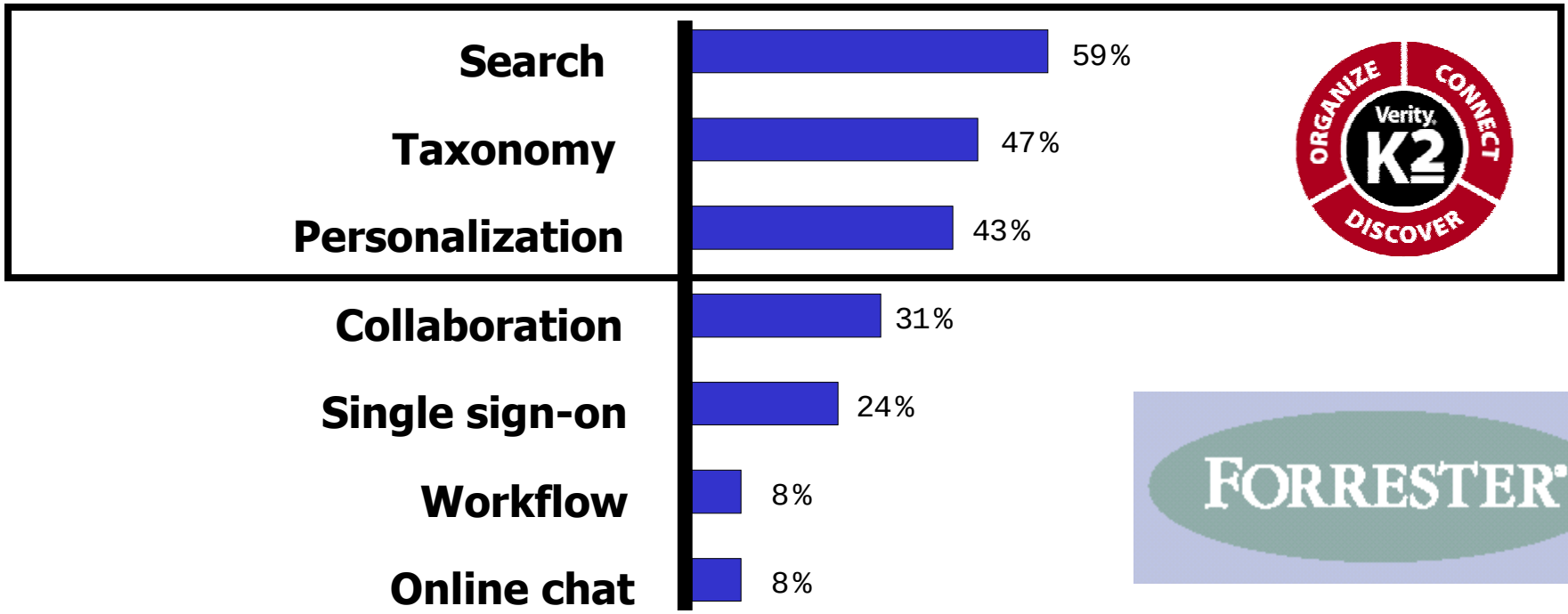
Document Level Security

The Forrester Report

AUGUST 2001

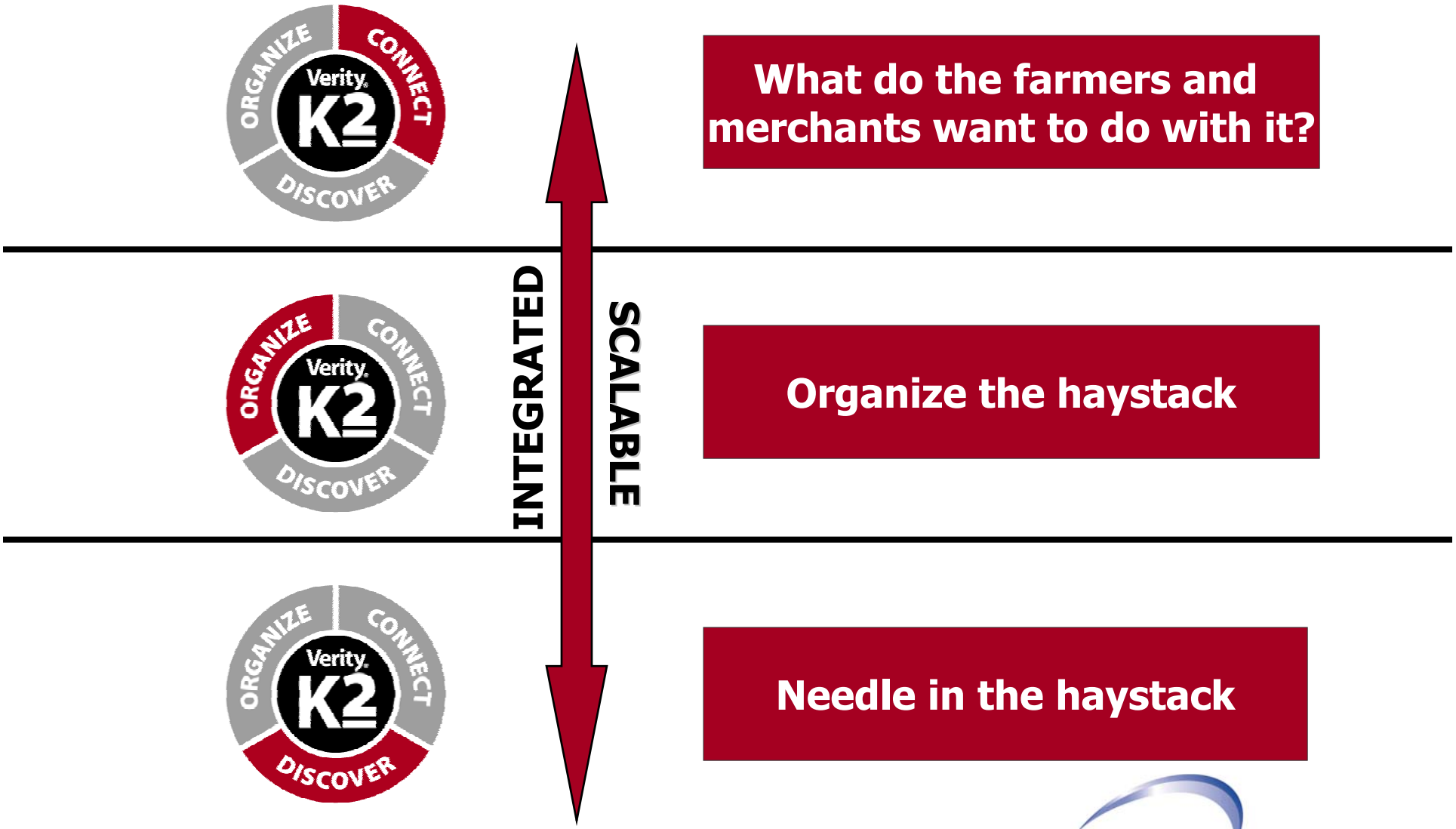
Making Enterprise Portals Pay

"What are the Key Enabling Features of The Enterprise Portal?"



Survey of Global 3500 (49 Interviewed)

3 Tiers of Knowledge Management



3 Tiers of Knowledge Management

Popular Ranking	Expert Location
Recommendation	Community
Profiling	Peer to Peer

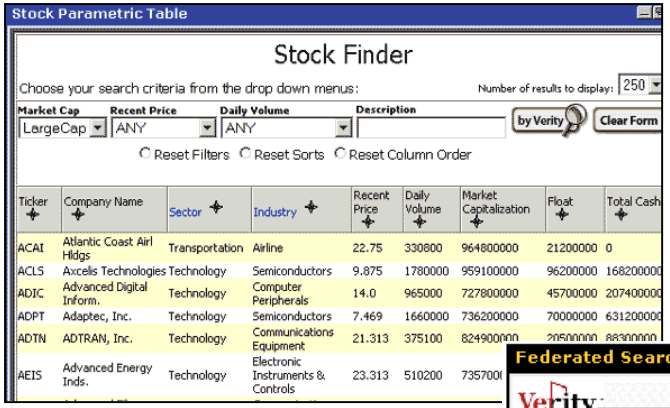
Automatic Rules	Business Rules
Concept Maps	Relational Taxonomy

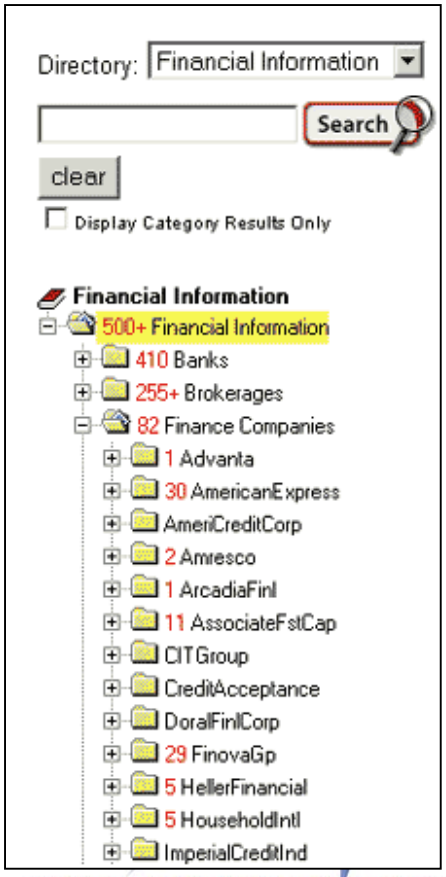
Advanced Search	Universal Index
Parametric Search	Secure Gateways
Structured Search	Federated Search
Keyword Search	Multiple Languages

Tier 1 Discovery

Discovery Paradigm

- I don't know what I want…
- I sort of know what I want…

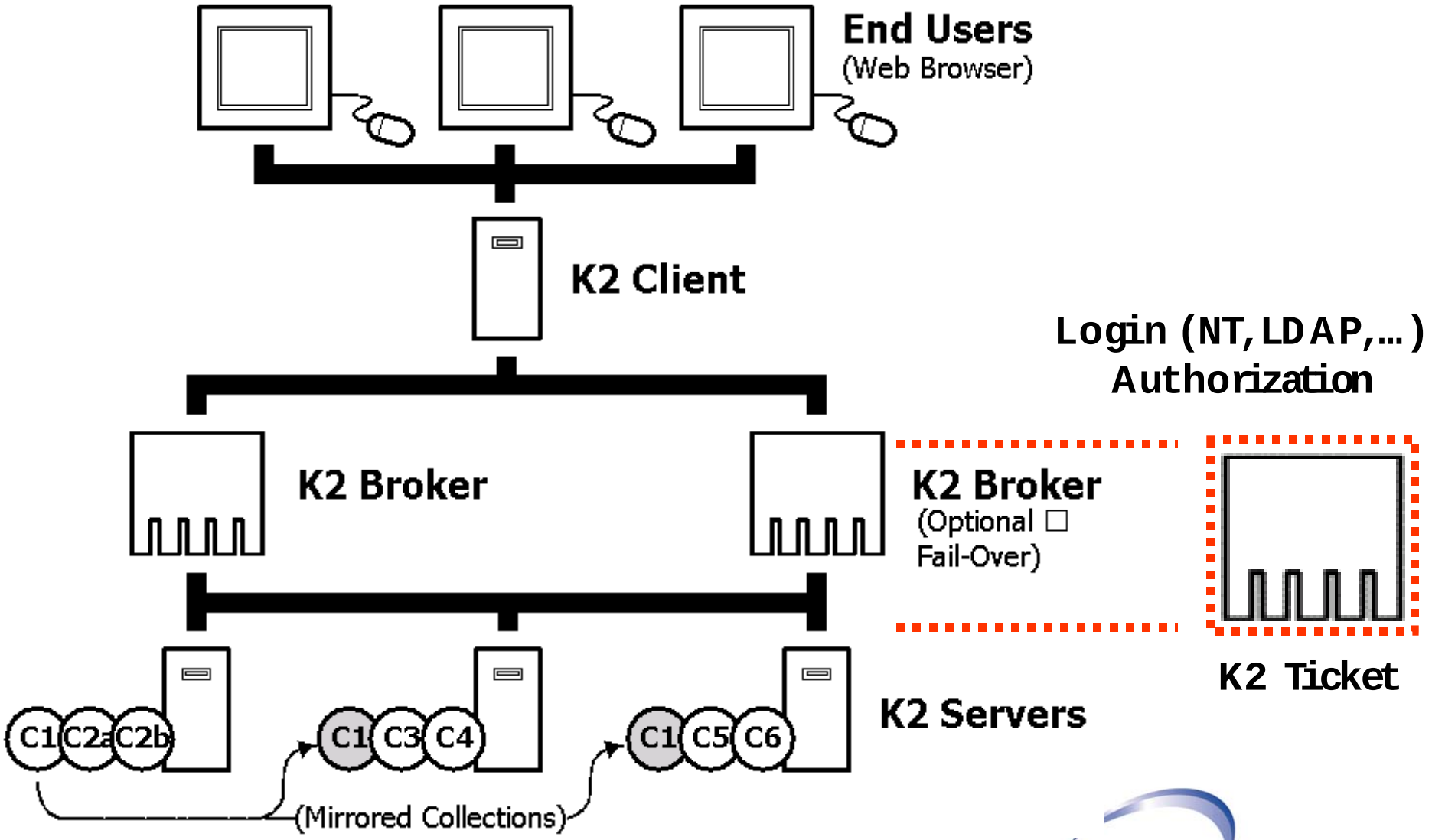




Gateways & Filters

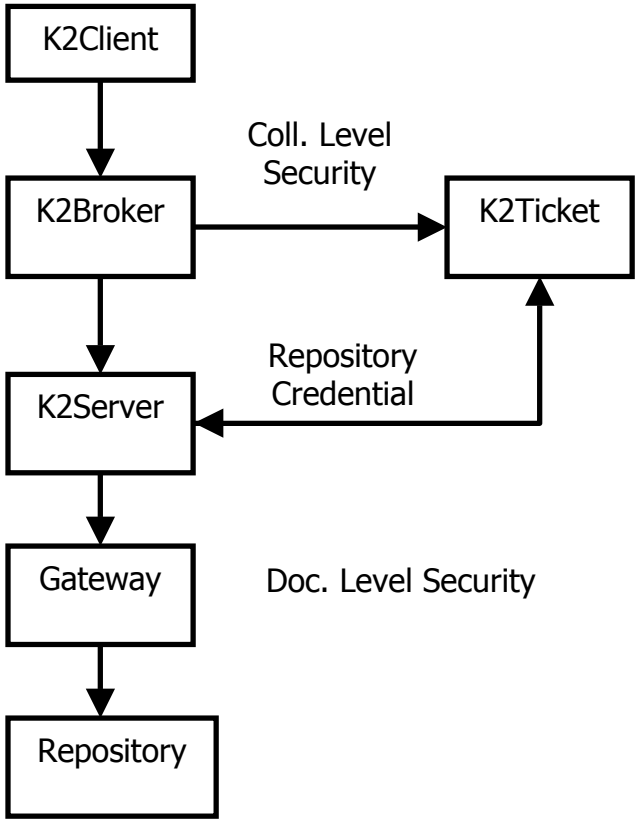
- **SECURE Gateways for Indexing**
  - File system
  - HTTP
  - MS Exchange
  - ODBC
  - Documentum
  - Lotus Notes
  - Others…
- **Filters for dealing with different formats**

Scalability, Availability, Security

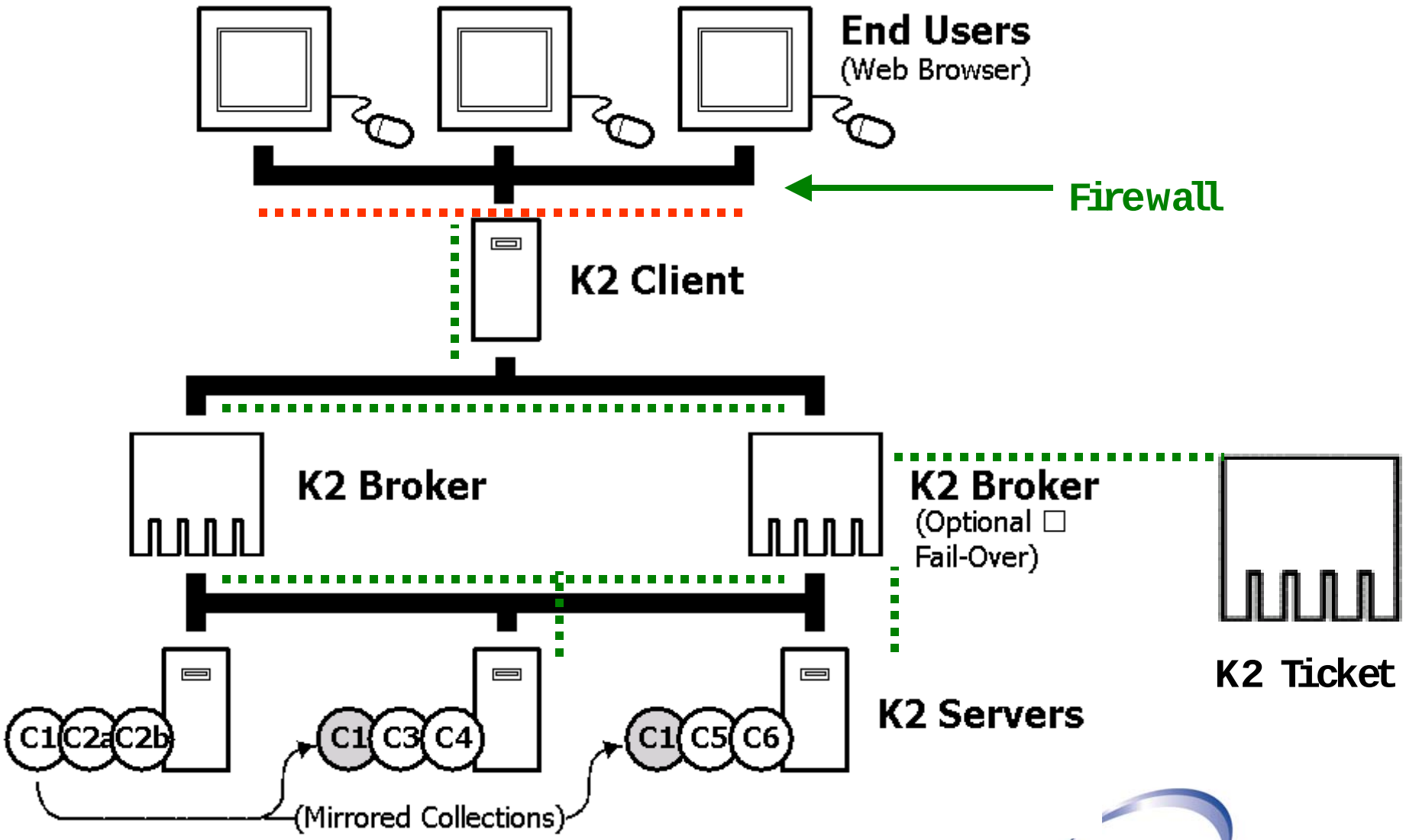


Node, Document Level Security

- **Ticket**
  - **A Ticket represents a user in the k2 search system.**
  - **Used For caching Gateway certificates in K2Server**
  - **Obtained via K2UserLogin**
- **Node level security**
  - **The node level security is enforced by K2Ticket based on a Ticket**
- **Document level security**
  - **Doc. Security is enforced through the gateway**



Full Wire Encryption



Federated Infrastructure

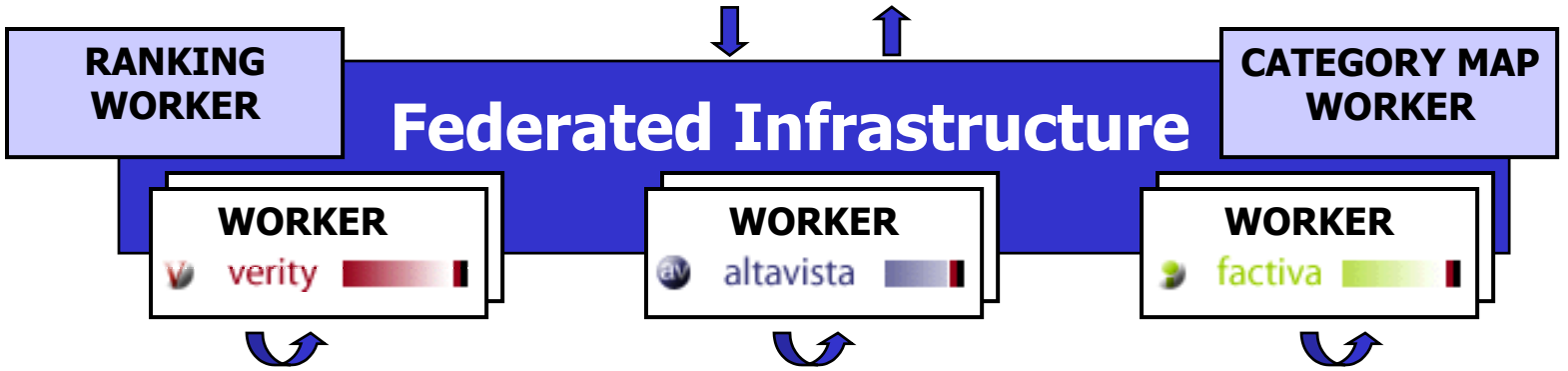
- Single point of "discovery"
- Integrate multiple sources of information
- Allows user login credentials via parameters for each source
- Filtering, Merging, Re-ranking of results
- Infrastructure for enterprise application integration (EAI)
- Infrastructure basis for P2P Technology (Personal to Portal)

Merged Uncategorized Results

World Wide Arts Resources - Your First Stop for Arts Online! - Artists, Museu
Summary: Access to the visual arts, artists, art galleries, art museums, Arti images, painting, sculpture,...

National Endowment for the Arts, a federal agency
Summary: Homepage for the National Endowment for the Arts, a federal age

Americans for the Arts: Home Page
Summary: Americans for the Arts is the national organization that supports public resource development, leadership ...



verity

CALENDAR(The Fort Worth Star-Telegram)
Summary: JANUARY Southwestern Exposition and

Symphony orchestra wa[illegible]s Canberra alive with the
Summary: Margaret Reid: shares concern for the c

The arts have a key role in making our society a be
Summary: The arts have a positive role to play in Minister Tessa Blackstone said today. They can als

altavista

About - Arts & Humanities
Summary: About arts and literature. Get hot ne human Guides.

arts.indiaabroad.com: arts & entertainmer
Summary: arts.indiaabroad.com: arts & er

Meta-Arts.com
Summary: Welcome to meta-arts.com. Enter He YourNameFree.com - Register your domain...

factiva

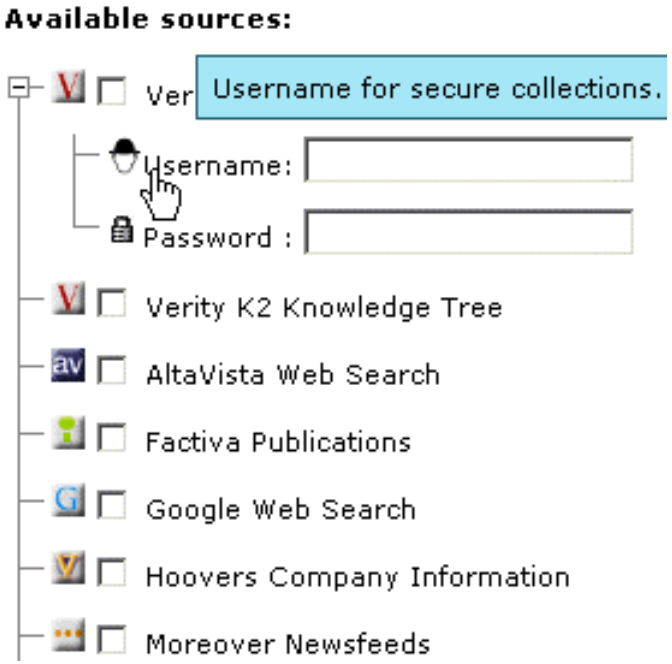
Uncategorized Results

National Endowment for the Arts, a federal agency
Summary: ... 2001; Ten National Design Competitior Bush Sends Greetings to Arts Supporters; New Summa

The New York Times: Arts
Summary: ... took to Carnegie Hall on Tuesday, is lik Then Crashing Its Best-Seller List By ALAN RIDING ...

World Wide Arts Resources / absolutearts.com - Your
Summary: ... Your First Stop for Arts Online! This is

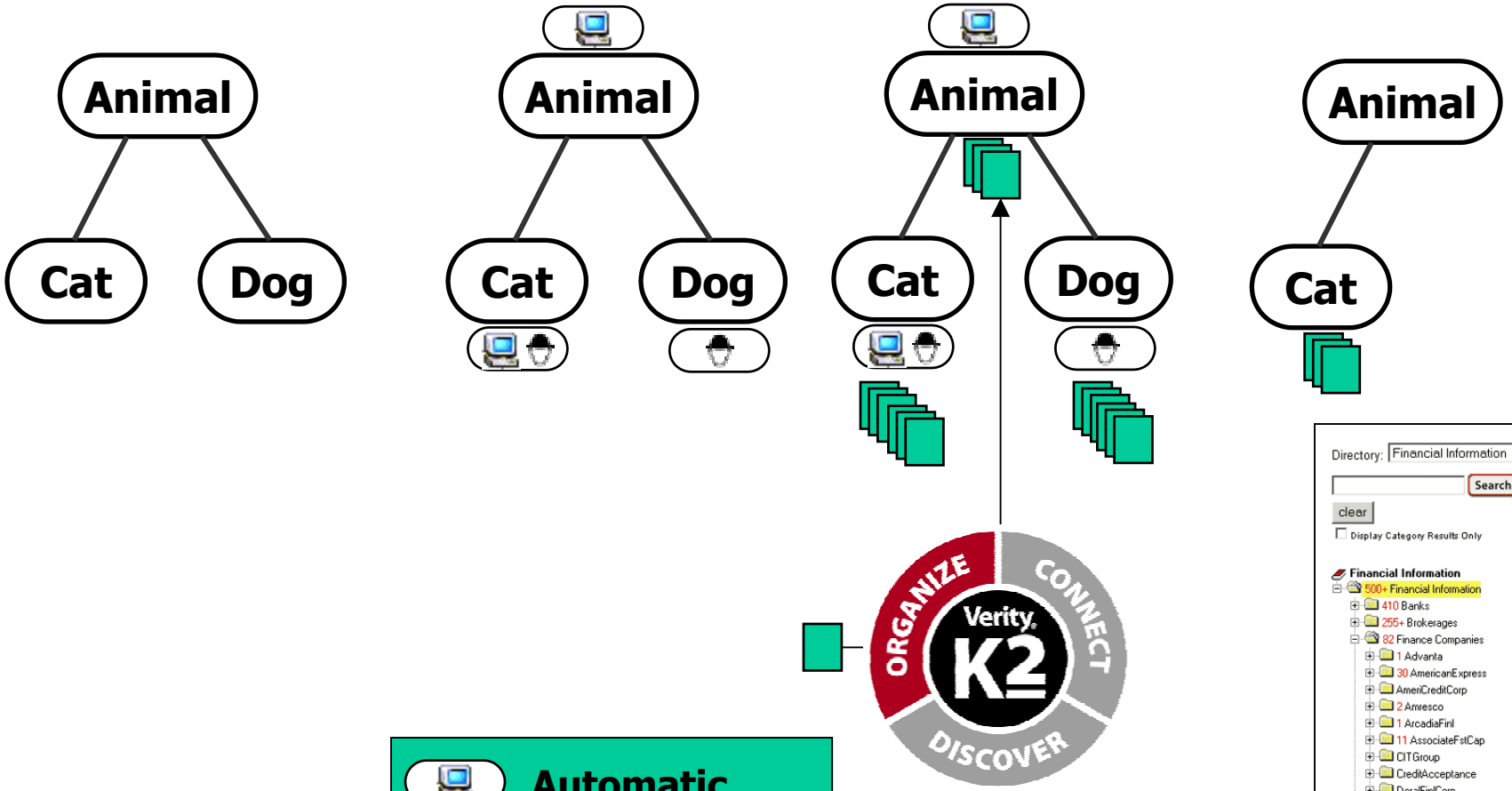
Federated Infrastructure



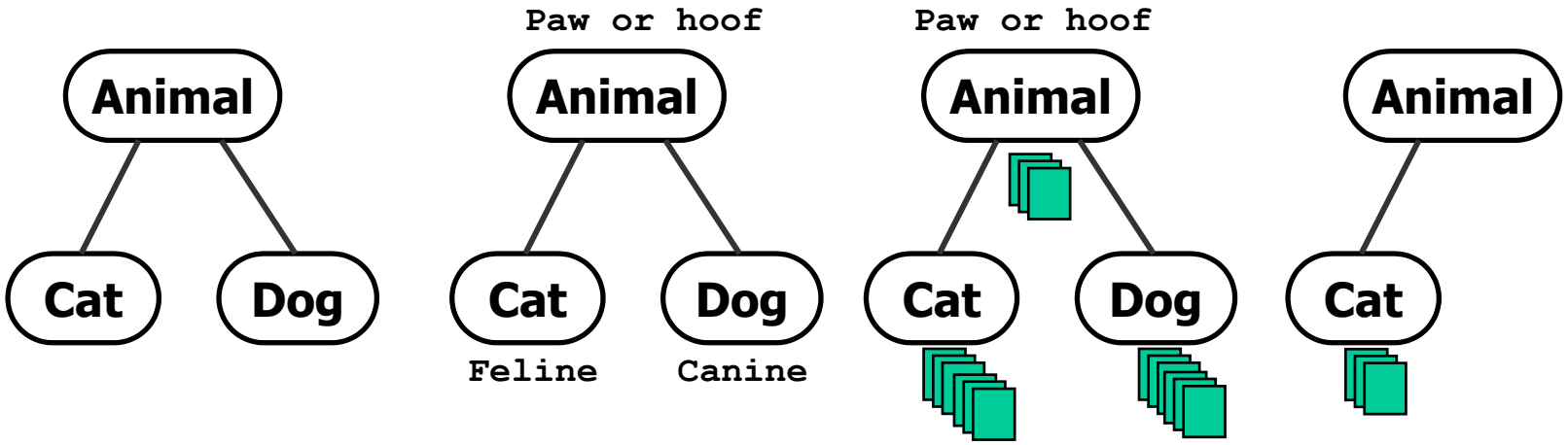
Tier 2 Content Organization

Taxonomies

Build Taxonomy	Build Model	Populate Taxonomy	Use It

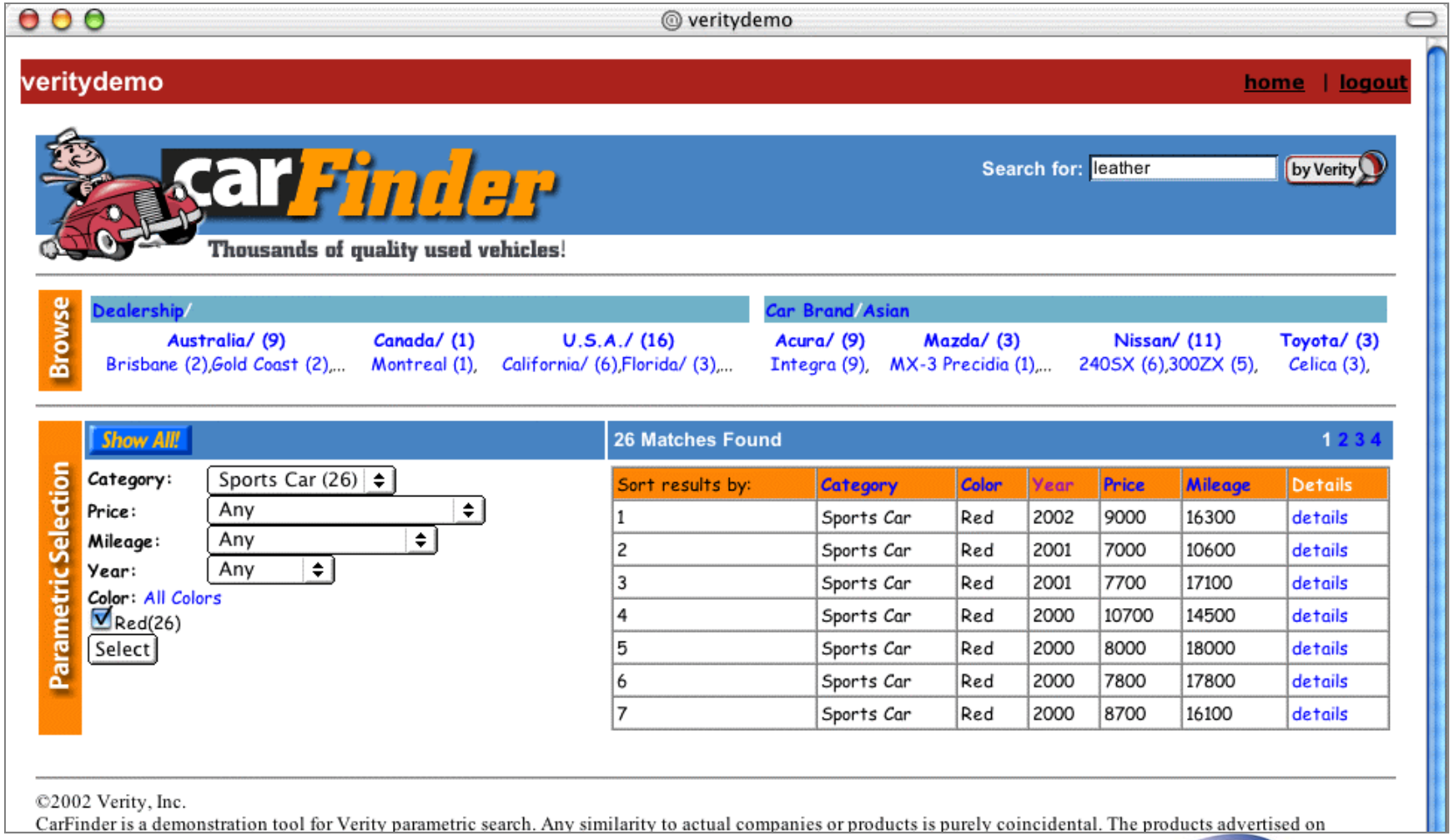


Taxonomies

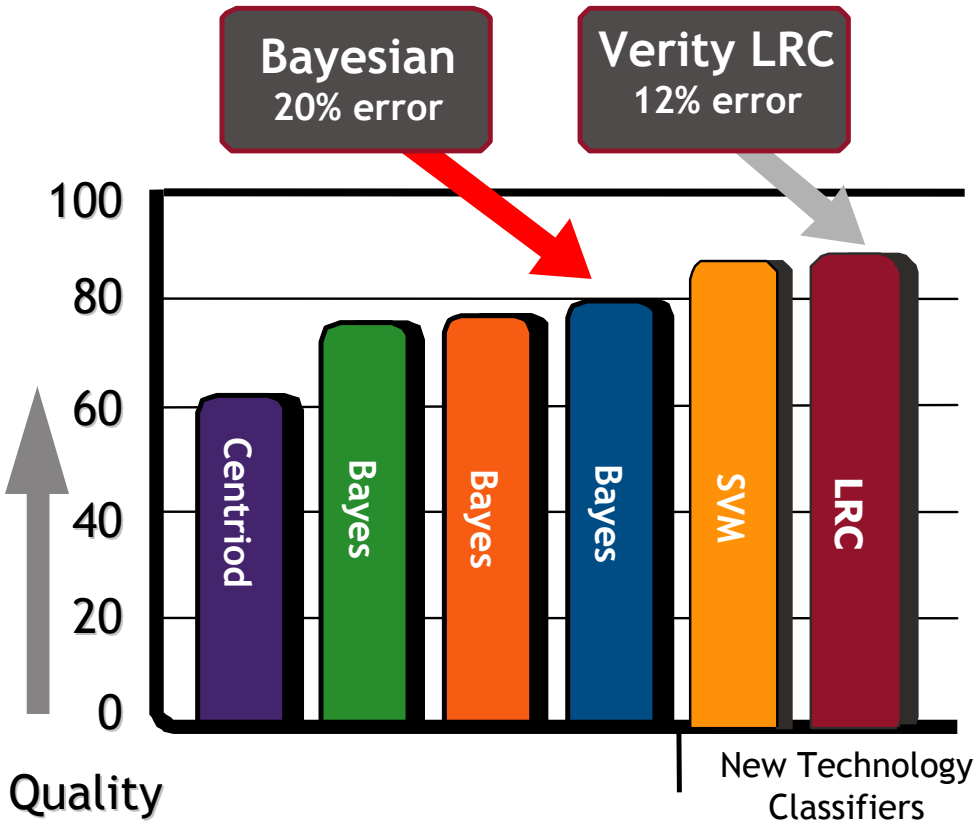


Build the Taxonomy	Build a Model for Each Node	Populate the Taxonomy	Use the Taxonomy
Automatic Naming ✓	Automatic Classification (training) ✓	Automatic ✓	Browse ✓
Import ✓	Import ✓	Manual ✓	Search & Browse ✓
Industry Taxonomy ✓	Industry Taxonomy ✓		Relational Taxonomies ✓
Concept Extraction ✓	Business Rules ✓		
Manual ✓			

Relational Taxonomies



Automatic Classification



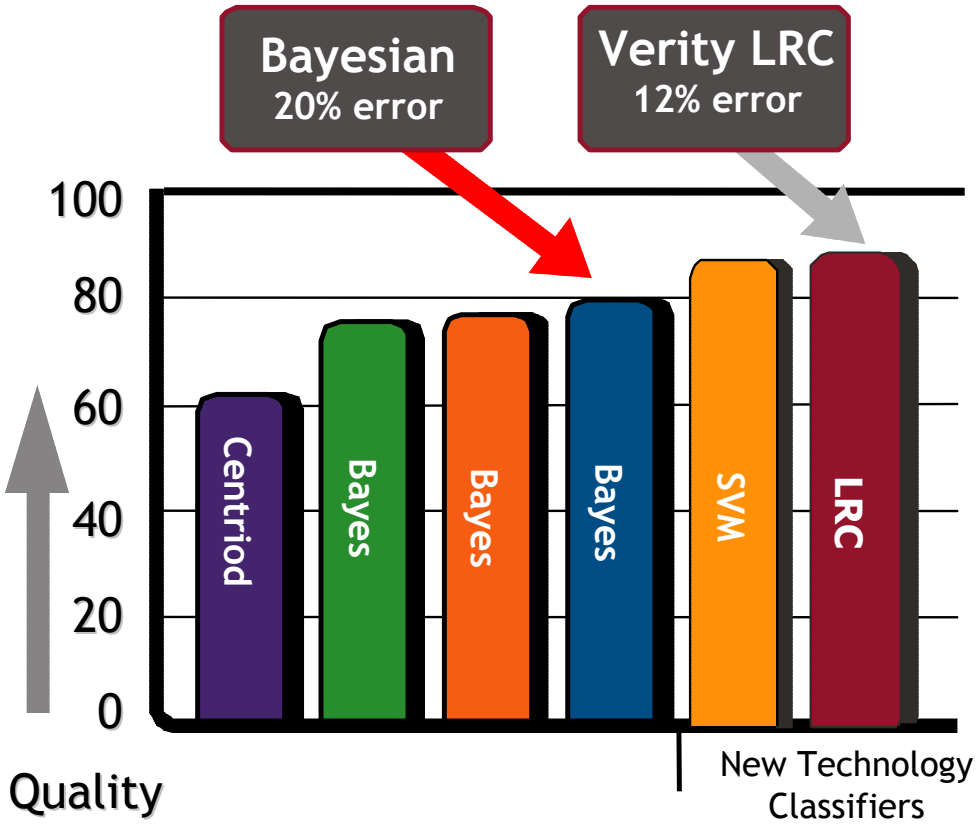
Findsim	61.7
Naïve Bayes	75.2
Taper2 (IBM)	77
BayesNet (MS)	80
SVM (MS)	87
LRC (Verity)	88.2

Reuters Benchmark:
9603 training docs,
3299 test docs

Micro-Averaged
Precision, Recall

Ref: Susan Dumais (MS research)
published in IEEE Intelligent Systems

Automatic Classification



- 1 level taxonomy: 20% error
- 2 levels: 36% error
- 3 levels: 49% error

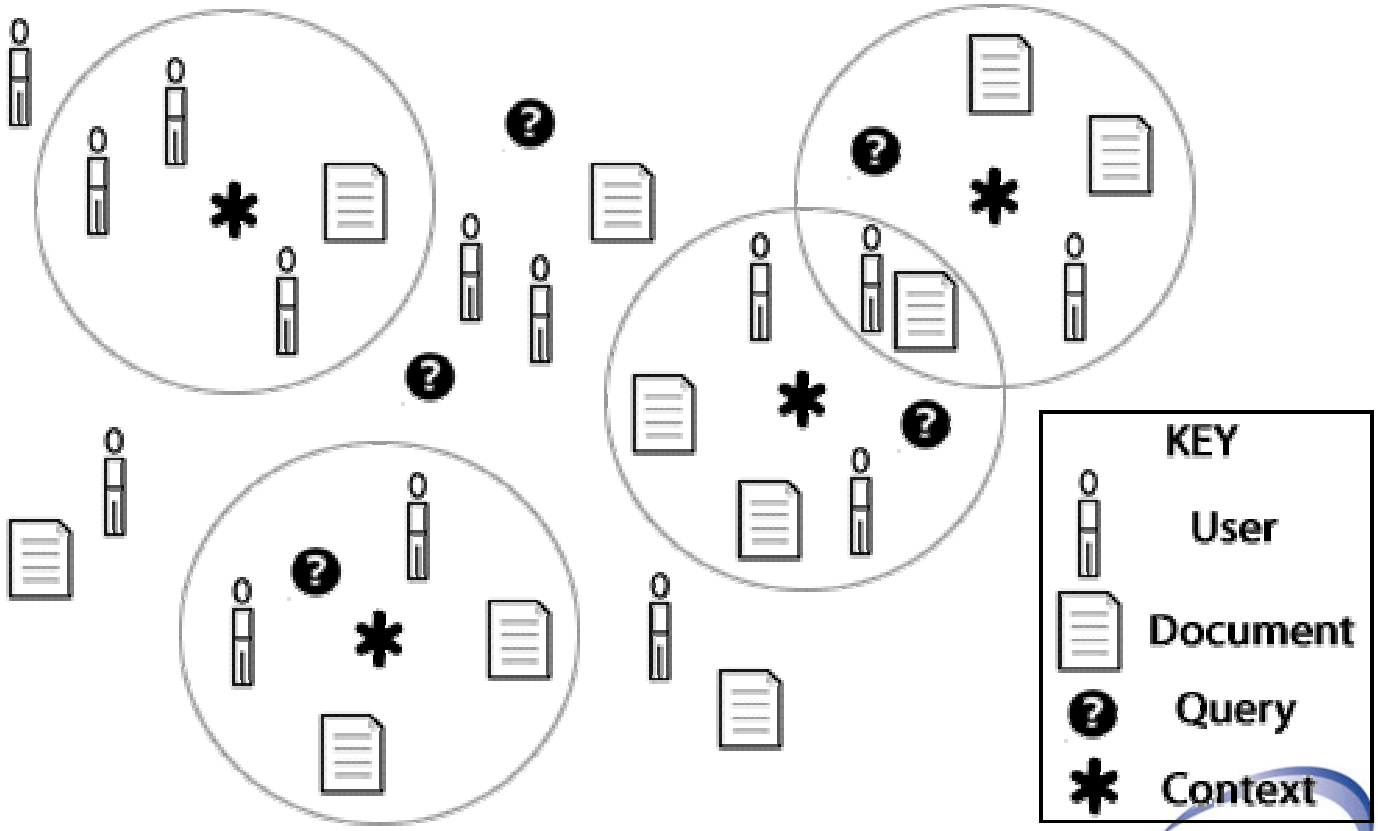
- With LRC (1 level) : 12% error
- With LRC (3 levels): 32% error
- With business rules: 1-5% error

Findsim	61.7
Naive Bayes	75.2
Taper2 (IBM)	77
BayesNet (MS)	80
SVM (MS)	87
LRC (Verity)	88.2

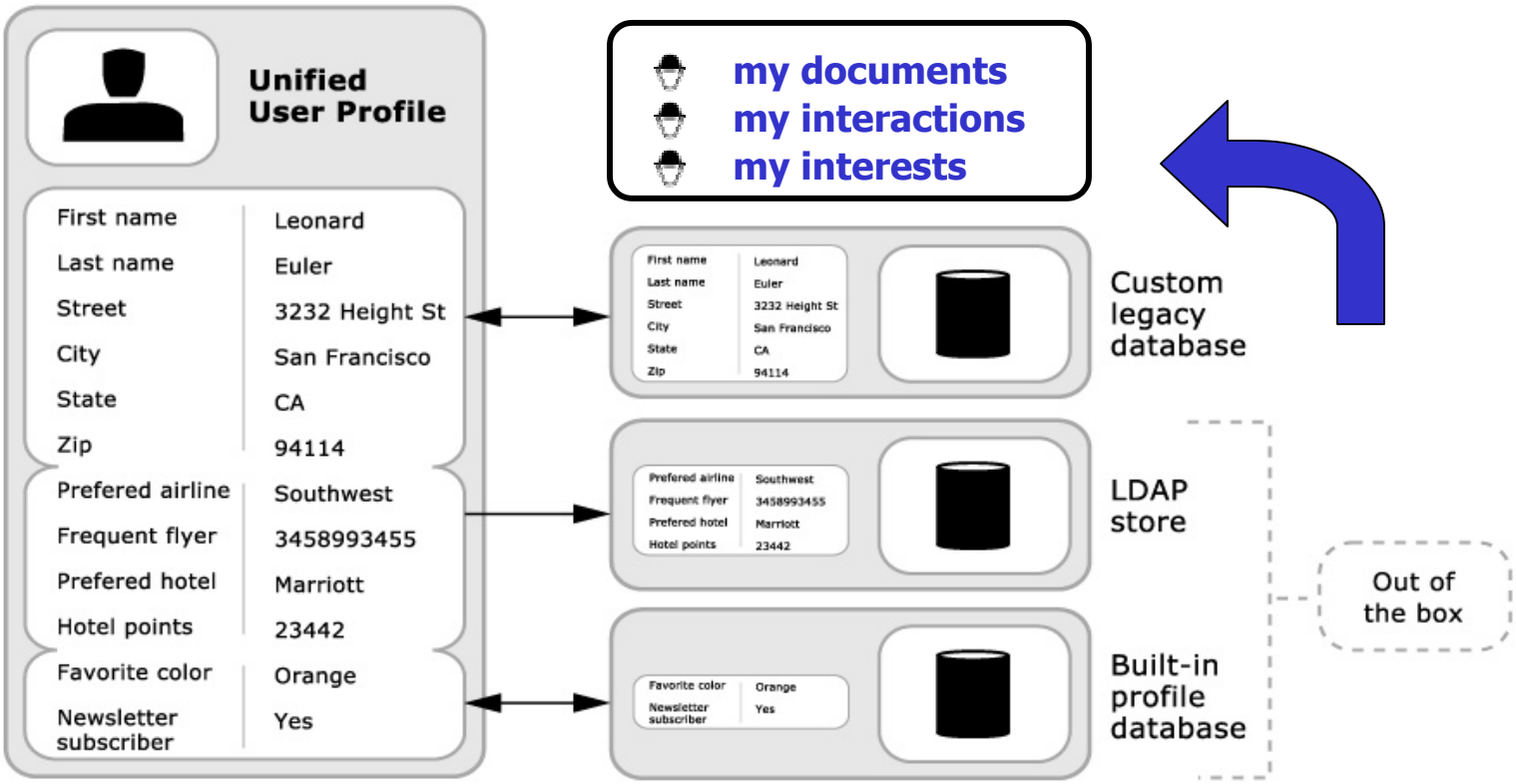
Tier 3
Social Networks

Social Networks – Implicit Links

- **Represents users, documents, categories and queries in a dynamic multidimensional Tensor Space**
- **Model continually learns explicit and implicit associations between these entities**



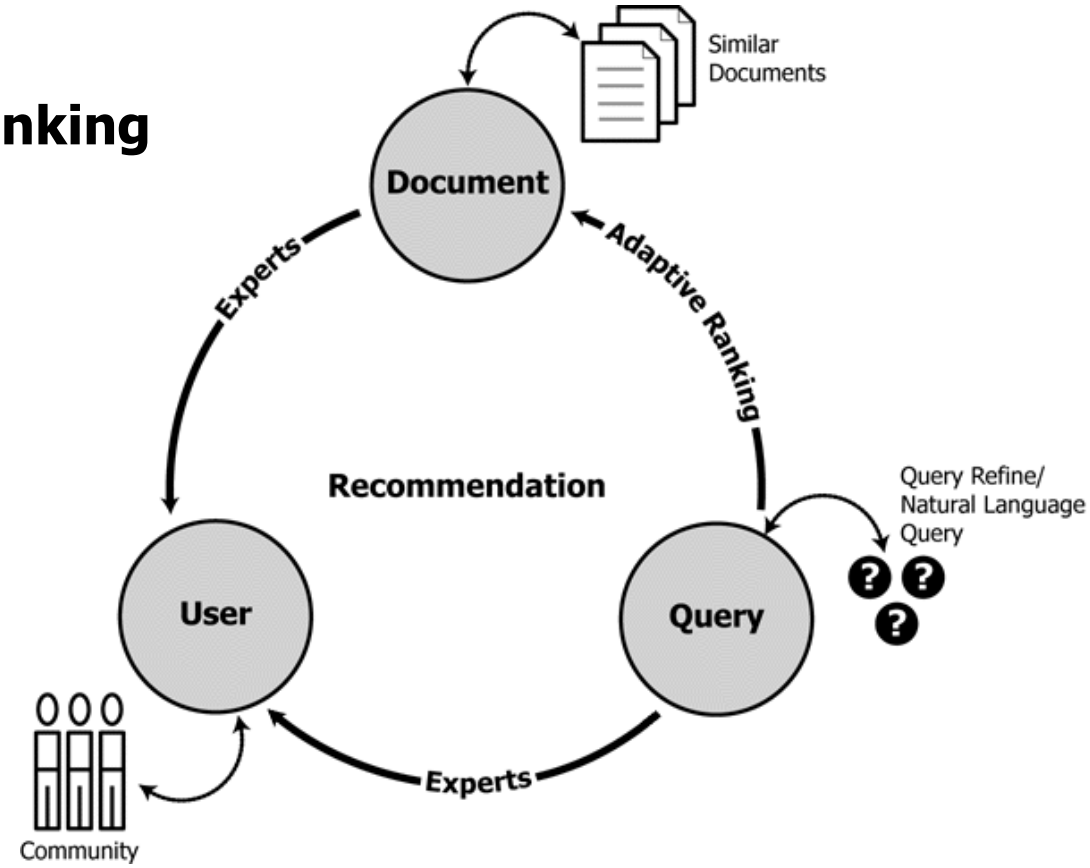
Content-based Personalization



Verity's Recommendation Engine

Functions:

- **Popular (Adaptive) Ranking**
- **Document Recommendation**
- **Document Similarity**
- **Community**
- **Expert Location**
- **User Transactions**



Engine ensures that secure documents are not recommended.

Auto-Initialization

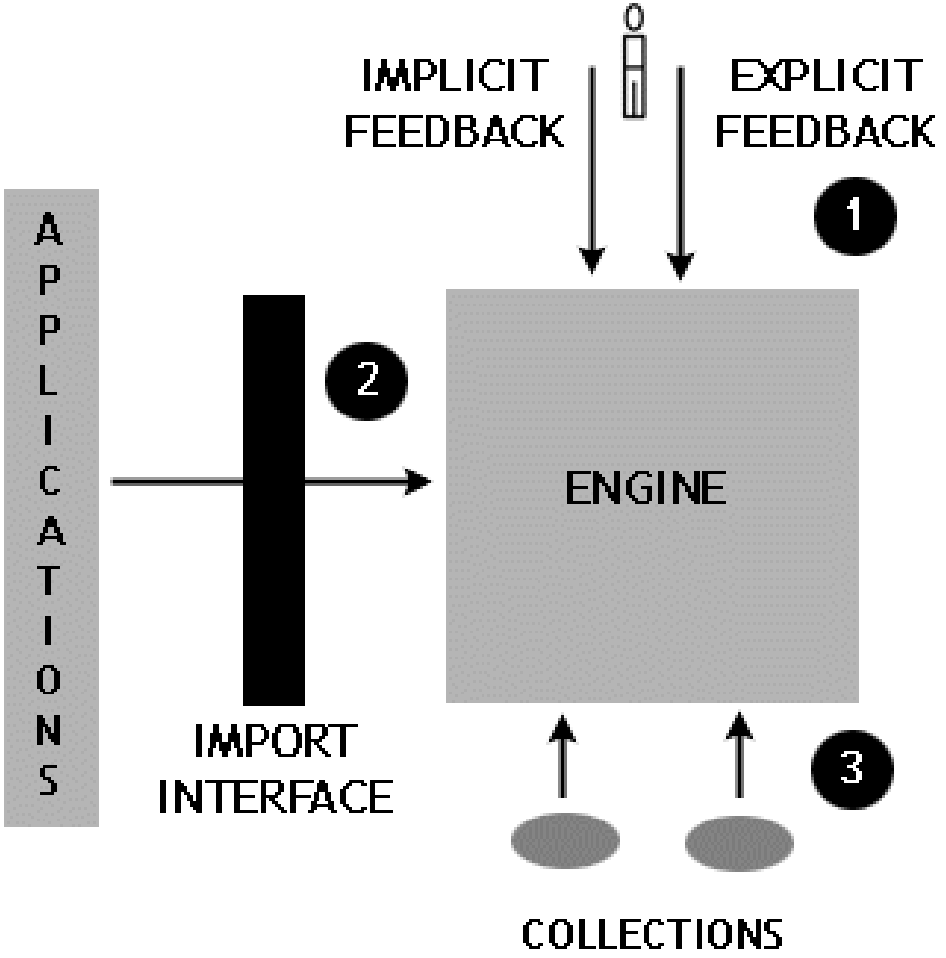
User Feedback

Weblogs

Customer Databases

LDAP

Databases



Email Forums Document Collections

Sample Applications

- **Corporate Intranets**
  - **Recommendation, finding domain experts,…**
- **Intellectual Property**
  - **Finding related research, consultations,…**
- **Medical Applications**
  - **Matching patients to doctors, clinical trials,…**
- **Customer Relationship Management**
  - **Matching customer problems to internal experts in Technical Support, Development,…**
- **Recruitment**
  - **Matching job postings to resumes, creating profiles from resumes,…**
- **Criminal Profiling**
  - **Matching records, crime descriptions, etc.**

Privacy and Trust

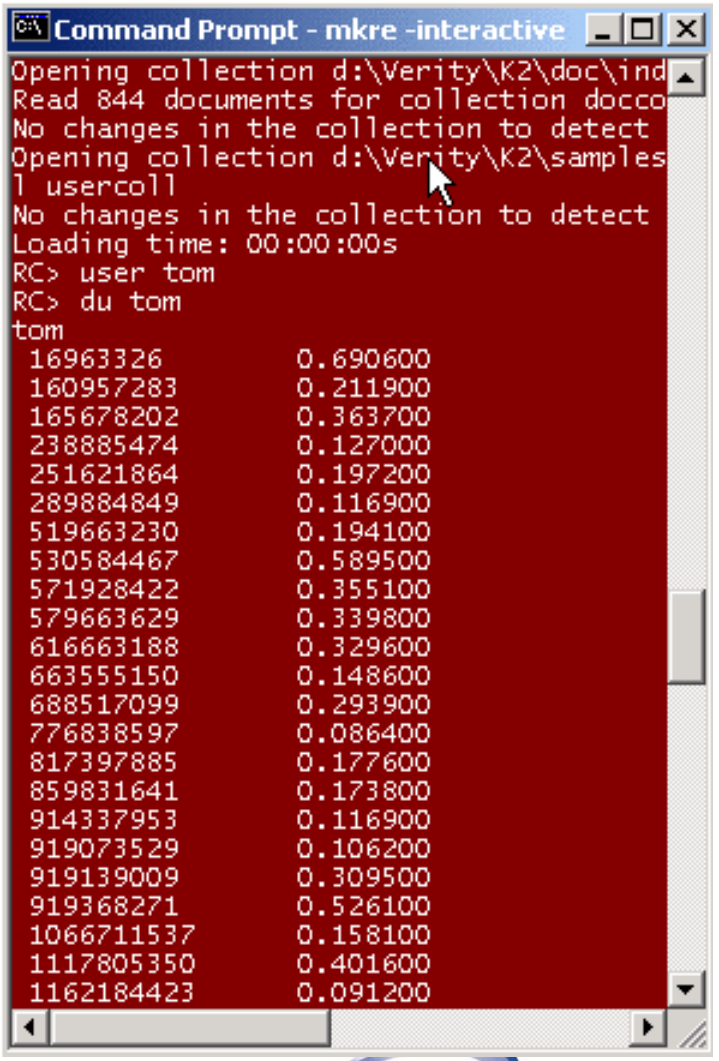
- **Provides an environment that enables privacy and builds trust.**
- **Provides capabilities that adhere to the 4 pillars of trust:**

  1. Identification
  2. Notice
  3. Transparency
  4. Control

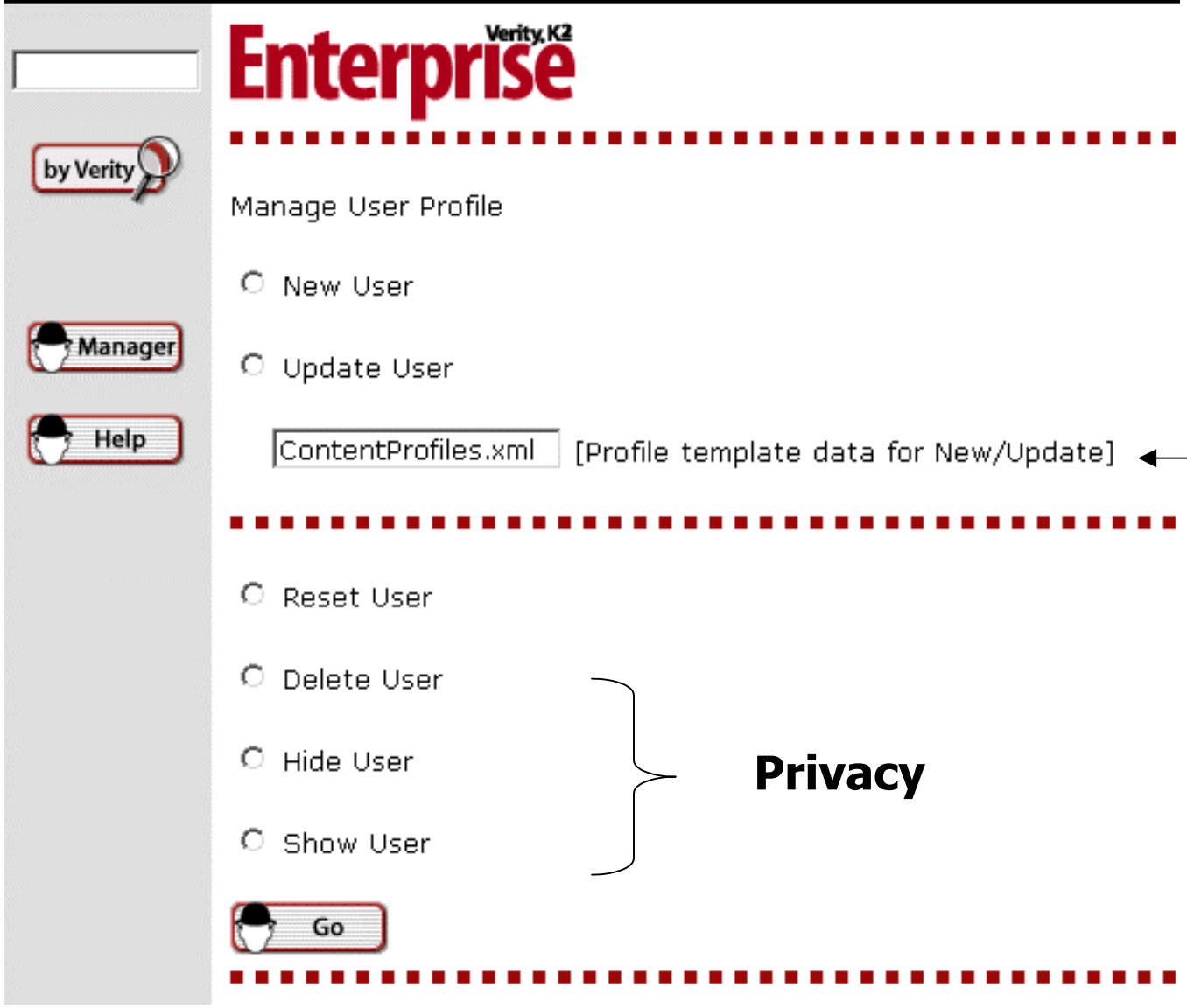
Transparency

- **Secure Mode: User profile unreadable – default product**
- **Open Mode: Authorized personnel able to export User Profiles in the clear, if required by organization.**

User profile →



User Control



Opt-in/opt-out possible

Easy to configure canned user profiles (XML)

Verity K2 Enterprise

Query: "tree"
Found in: 55 documents out of 296
Collections: pe_coll

1 to 20

Rank	Document
1	Knowledge Tree Management Utility (ktmgr)
2	Knowledge Trees and Workspaces
3	K2 Server Configuration Options
4	Testing Search on a Server-Examples
5	Classification Wizard
6	Understanding Workspaces
7	Table of Contents
8	Concepts of Utilities
9	Understanding Knowledge Trees
10	Optimized Knowledge Tree Search Utility (qsr
11	Testing Search on a Server
12	Topic Editing Tutorial
13	Index
14	Using Verity Intelligent Classifier
15	Knowledge Tree Search Utility (ktsrch)
16	Detailed Operators and Modifiers Reference
17	Panes Within the Intelligent Classifier Windo
18	Setting Intelligent Classifier Options
19	Index
20	About Topics and Topic Sets

[PREV] [NEXT]

Verity K2 Enterprise

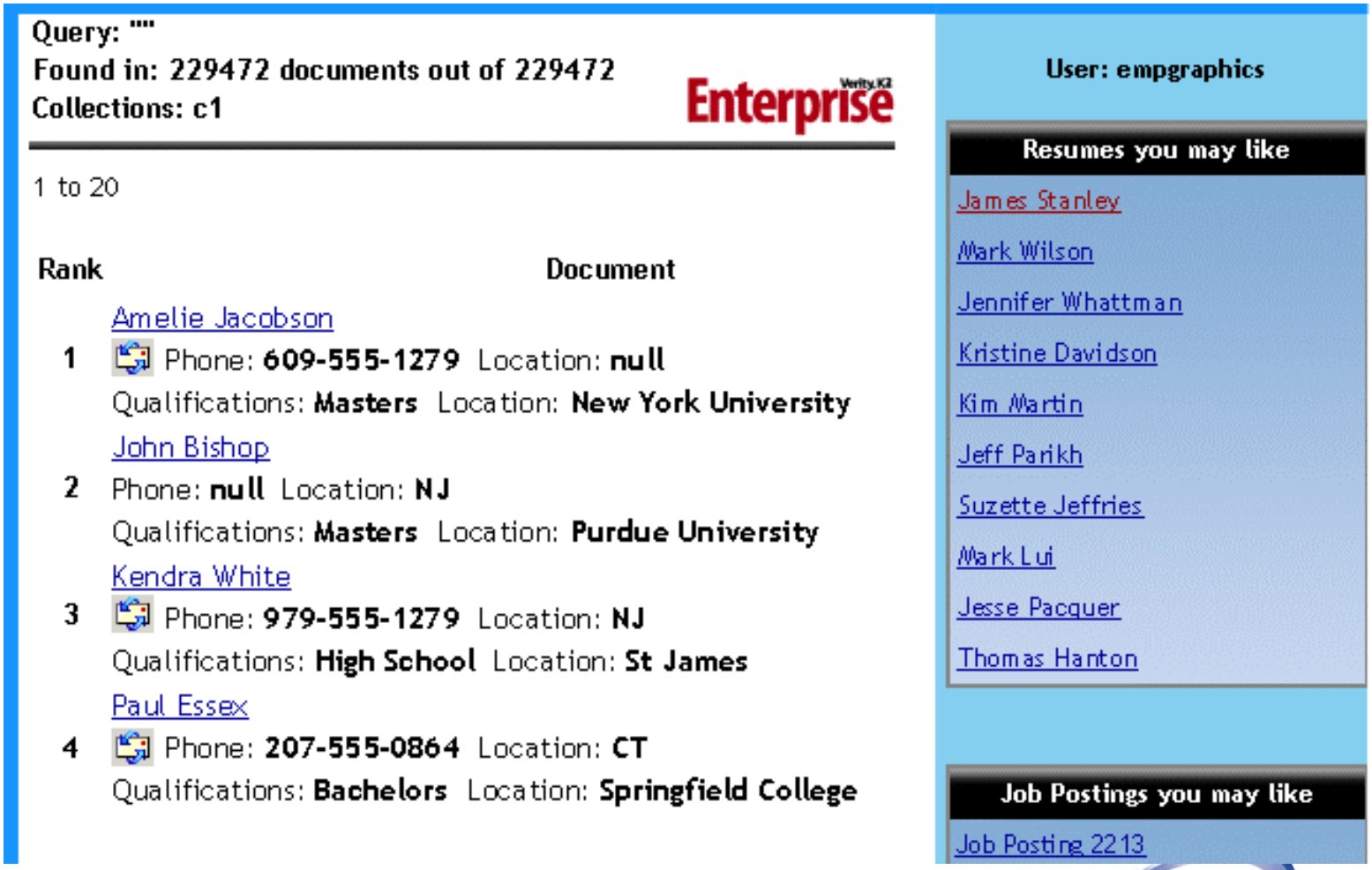
Query: "tree"
Found in: 55 documents out of 296
Collections: pe_coll

1 to 20

Rank	Document
1	Knowledge Tree Management Utility (ktmgr)
2	Knowledge Trees and Workspaces
3	K2 Server Configuration Options
4	Testing Search on a Server-Examples
5	Classification Wizard
6	Setting Intelligent Classifier Options
7	About Topics and Topic Sets
8	Understanding Workspaces
9	Table of Contents
10	Concepts of Utilities
11	Understanding Knowledge Trees
12	Optimized Knowledge Tree Search Utility (qsrch)
13	Testing Search on a Server
14	Topic Editing Tutorial
15	Index
16	Using Verity Intelligent Classifier
17	Knowledge Tree Search Utility (ktsrch)
18	Detailed Operators and Modifiers Reference
19	Panes Within the Intelligent Classifier Window
20	Index

[PREV] [NEXT]

Previous
Next
Index
Contents
Search
Manager
Help
Rate this page:
Personalized Discovery
The personalization feature incorporated in Verity K2 Enterprise automatically connects your users to subject experts within your organization, and recommends documents based on the individual users' past and present queries and/or documents that are similar to selected documents. This highly-advanced technology is based on Verity's proprietary analysis engine that combines business rules with latent patterns in user behavior. It puts simple queries into the context of the social networks created by the interaction of users, information and queries in your organization.
K2 Enterprise's Personalization and Adaptive Ranking features all contribute to personalizing your users' search experience.
K2 Enterprise Recommendation Engine Feedback ...
File Edit View Favorites Tools
ORGANIZE CONNECT Verity K2 DISCOVER
Thank you for your feedback tom
re allows administrators to
users. One of the features of
analyzes group behavior to
ancy ranking by incorporating
ocument, the higher it appears
sers consistently find the most
wledge discovery process.
Powered by Verity
User: tom
Similar Documents
Administering the Personalization Engine
Personalization Engine Internals
Introduction
Personalization Engine
System Architecture
Interest Group
tom
ben
leo
george
ashley


Query: ""
Found in: 229472 documents out of 229472
Collections: c1
Enterprise
1 to 20
Rank
Document
Amelie Jacobson
1 Phone: 609-555-1279 Location: null
Qualifications: Masters Location: New York University
John Bishop
2 Phone: null Location: NJ
Qualifications: Masters Location: Purdue University
Kendra White
3 Phone: 979-555-1279 Location: NJ
Qualifications: High School Location: St James
Paul Essex
4 Phone: 207-555-0864 Location: CT
Qualifications: Bachelors Location: Springfield College
User: empgraphics
Resumes you may like
James Stanley
Mark Wilson
Jennifer Whattman
Kristine Davidson
Kim Martin
Jeff Parikh
Suzette Jeffries
Mark Lui
Jesse Pacquer
Thomas Hanton
Job Postings you may like
Job Posting 2213

Summary

- Security issues do not stop outside the firewall
- Node-level, document-level security are critical within enterprise
- Huge security lapses can occur without document level security
- Encryption, privacy, trust are important in some settings
- Infrastructure must have security built into it.

Thank You

rajatm@verity.com