

# The Stratify Discovery System™

*The Stratify Discovery System automatically organizes, classifies and presents document-based information. By adding structure to unstructured data, the system allows companies to find and leverage textual information in the same way a relational database allows them to leverage numerical data.*

*The ever-growing volume of unstructured information available to an enterprise makes it difficult, if not impossible, for workers to find what they need. Computing advances and the World Wide Web allow us to store information easily, but organizing and accessing it is very time-consuming and complicated. With a Stratify Discovery System, information finds you — not the other way around — so you can know more about what is happening in your business, how your customers and partners are being served and how you can increase revenue and profits.*

*The Discovery System combines superior technology with human control to create an information hierarchy of topics, classify internal and external information into this hierarchy, and proactively present this information to users. With it, an unorganized wasteland of documents and Web pages becomes a wealth of useful information.*

## Automatic Hierarchy Building

For business people to use unstructured information effectively, documents must be organized, classified and presented in a uniform and consistent manner. Organizing the documents is the most labor intensive and expensive step in the process. The Discovery System begins its work here by analyzing words, their frequency and patterns in documents. Starting with a sample set of documents, the Hierarchy Builder, a key component of the Discovery System, automatically organizes them into a set of topics specific to your business — with no human intervention required. Its algorithms, based on statistical methods and information theory, group documents into categories and then organize these categories into a hierarchy. By uncovering the natural structure in unstructured documents, the Discovery System eliminates the weeks or months you could spend hand-tagging or organizing information manually.

What a relational database does for structured information, Stratify's Discovery System does for unstructured information. The resulting hierarchy provides a consistent, global view of information relevant to your business — a common framework where unstructured documents are easily classified and accessed. The Discovery System extracts information from file systems, the Internet, intranet Web servers and subscription feeds and organizes it into a single, easy-to-navigate view that is relevant to your business.

*Unlike other products, the Discovery System uses multiple independent classifiers that reinforce each other and provide fine-grained, accurate classification.*

*The Discovery System provides a powerful, flexible interface between people and technology. It combines machine efficiency with the intelligence that only human judgment can provide.*

## Accurate Classification

Once the information hierarchy is created, the Discovery System collects text and data about documents, identifies the main ideas in the text, and classifies the documents into the hierarchy. Unlike other products, the Discovery System Classifier uses multiple independent classifiers that reinforce each other and provide fine-grained, accurate classification. Using parallel, complementary classification methods, it evaluates the results of each independent method and, through our patented Combiner technology, evaluates the best possible match between the document and subject matter. As a result, the Discovery System classifies documents more accurately than single-method classifiers and delivers more relevant results to users.

The Discovery System automates the labor-intensive, repetitive activities associated with classifying documents. People need not complete forms, write summaries, or move documents into directories. As a result, your company can access the critical information it needs to stay competitive and develop new products – without expensive and time-consuming manual classification.

## Information Presented in a Hierarchy

Recent studies show that users find information 50% faster when it is organized into topics.<sup>1</sup> The Discovery System information hierarchy allows users to see the relationship between documents and topics more easily. Instead of presenting a flat list of article titles, the Discovery System organizes search results and related documents by topic area, quickly directing users to the subject areas and information they want to see based on their interests and behavior.

Information organized into a hierarchy is more accessible and more easily navigated. The Discovery System prevents information loss because users can look for information by subject area rather than by constructing complicated queries with Boolean operators and elaborate syntax. Discovery System users are more productive because they can more easily access well-organized information.

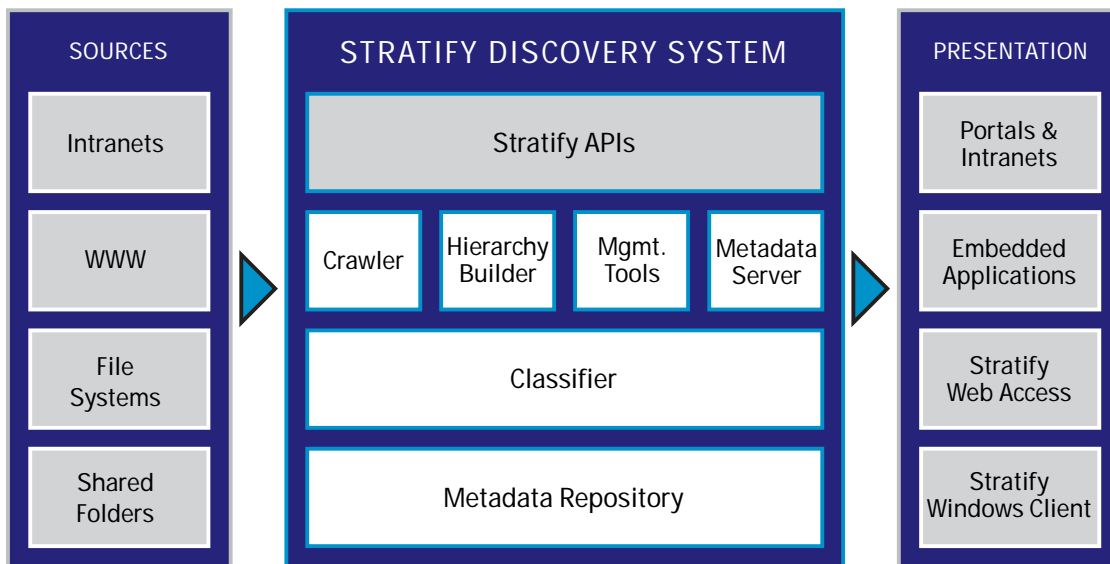
## Machine Efficiency with Human Oversight

Many automatic classification products allow very little control over the process of organizing and classifying information. In contrast, the Discovery System provides a powerful, flexible interface between people and technology. It combines machine efficiency with the intelligence that only human judgment can provide. By opening the classification process to human oversight, the Discovery System automates classification while allowing people to control the processes as needed.

<sup>1</sup> "Bringing Order to the Web: Automatically Categorizing Search Results", H. Chen and S. T. Dumais, (2000) *Proceedings of CHI'00, Human Factors in Computing Systems*, pp. 145-152.

## Discovery System Components

The Stratify Discovery System has an open, modular architecture, including components designed to meet the flexibility and scalability requirements of large organizations and global implementations:



The Discovery System has a modular architecture that crawls documents from many locations, organizes them into a hierarchy and presents them through easy to use applications and interfaces.

**Crawler** — Crawlers collect text-based information from internal and external sources and extract the text within them, regardless of formatting and graphics.

**Hierarchy Builder** — Using our patent-pending technology based on clustering and pattern matching algorithms, the Hierarchy Builder automatically creates a structured information hierarchy from unstructured documents.

**Management Tools** — The Management Tools facilitate review and input to the hierarchy building and classification processes.

**Classifier** — The Classifier identifies the main ideas in text and classifies documents into a topic-based hierarchy defined uniquely for each organization or business function.

**Metadata Repository** — An open, SQL-standard relational database stores document metadata and classifications for easy access by any Discovery System application.

**Metadata Server** — The Discovery System's middle-tier enables applications to present related documents and combine structured data with unstructured documents in context.

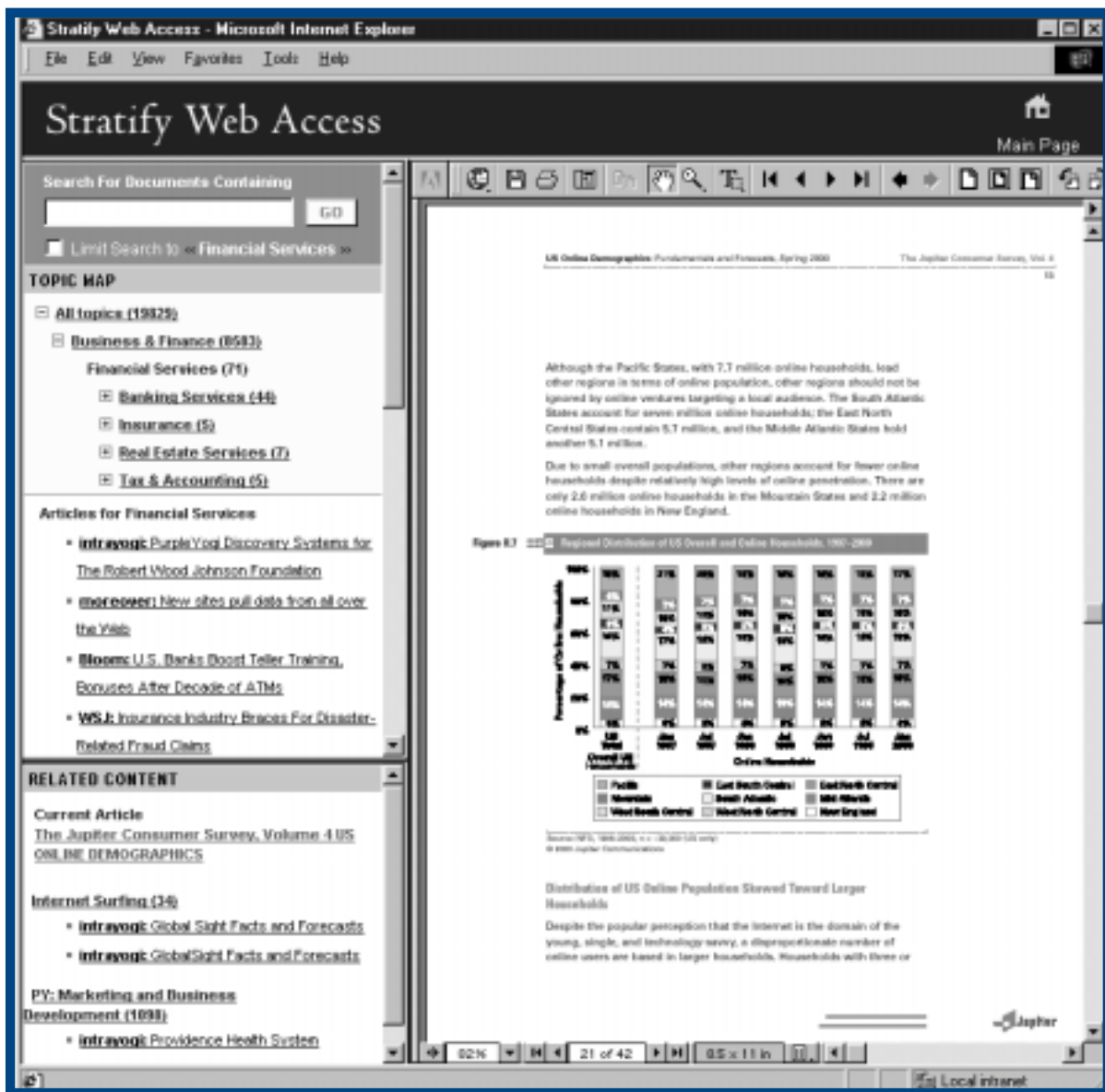
**Stratify Web Access™** — This browser-based interface allows users to navigate the information hierarchy, find related documents and view search results organized into the same hierarchy.

**Stratify Windows Client™** — A Windows-based application that supports automatic user-interest profiling, related document delivery, real-time classification of Microsoft Word, PowerPoint, Adobe Acrobat and HTML documents and offline access to selected documents.

# The Discovery System in Action:

Organizing unstructured corporate information is usually a complicated process. When using the Discovery System, organizing and classifying information involves three basic steps:

1. Building an information hierarchy to organize your documents
2. Classifying documents based on the hierarchy
3. Presenting the classified information to users



The Stratify Discovery System offers a rich variety of presentation features and interfaces to deliver an organized hierarchy of unstructured information. Stratify Web Access, shown above, is the browser-based user-interface.

## Organize

Starting with any number of documents from inside your enterprise or external sources, the Hierarchy Builder matches documents by subject, groups them into clusters and organizes the clusters into a Topic Map. It then converts groups of documents into mathematical models that represent the subject area associated with each node in the Topic Map. Using the System Tools, administrators can label, combine, split and adjust the hierarchy and subject area models, allowing the Discovery System information hierarchy to reflect any business process or information organization need.

*The Hierarchy Builder matches documents by subject, groups them into clusters and organizes the clusters into a Topic Map.*

## Classify

Crawlers extract text from documents, parse and filter it into the most essential and relevant information, and send it to the Classifier to be categorized by subject area. They extract and store the document's title, original location identifier, source and date/time stamp. For each document classified, the Classifier creates database entries consisting of metadata and the matching classifications and stores them in the Metadata Database. This database of unstructured information forms the foundation for enhancing existing enterprise applications and for building new generations of decision-making and insight-producing applications, not previously possible.

*This database of unstructured information forms the foundation for enhancing existing enterprise applications and for building new generations of decision-making and insight-producing applications, not previously possible.*

## Present

The Metadata Server uses the document metadata and classifications to power a wide variety of applications and to present related articles to the end user. The Discovery System includes two unique end-user interfaces. Stratify Web Access provides a server-only based solution for navigating the hierarchy inside a web-based application or portal page. It integrates easily with search engines, presenting search results by subject area or from a single area of the hierarchy. The Stratify Windows Client is a robust, client-based user interface that profiles user interests based on what they read or select from the information hierarchy and automatically delivers more, relevant information matching this profile.

## Conclusion

Relational database technology has allowed corporations to harness and manage structured-data and to build applications that manage their enterprise business processes. With the Internet and the development of the online, "e-business" channel, unstructured, text-based information has grown to enormous volumes and is the source of valuable insights and information that help people make important business decisions. The Discovery System tames this wilderness of unmanageable information and provides a platform upon which new generations of company-making business applications will be built. With a Stratify Discovery System, users will search less and discover more, saving time and optimizing the use of their company's intellectual capital.

# Discovery System Feature Summary

- ❑ Automatically builds a custom hierarchy from corporate documents, producing a global, consistent view of information tailored for your business.
- ❑ Automatically collects and classifies corporate documents and external information, eliminating manual tagging.
- ❑ Provides intuitive navigation, information discovery and proactive retrieval of classified information by matching related documents to current work in process.
- ❑ Allows users to share and manage information throughout the enterprise.
- ❑ Incorporates advanced personalization features that learn users' interests automatically and alerts users to new documents matching their profile.
- ❑ Includes a choice of presentation and integration options so users can always get the information they need.
- ❑ Supports offline folders that help remote and mobile users access and share information while traveling.
- ❑ Open, flexible architecture and APIs support industry standards including SQL, HTTP, and XML and makes for simple, streamlined application integration.
- ❑ Optionally supports a subscription-based topic hierarchy of business, industry and current events covering over 12,000 subjects.

## System Requirements

### Hardware:

Database Server – One Pentium III, 700 MHz, 512 MB RAM, 20GB, 7200 RPM RAIDED disk storage or better

Database – IBM DB2, Microsoft SQL Server

Operating System – Linux or Solaris for DB2,

Windows 2000 for Microsoft SQL Server

Application Servers – Two Pentium III, 700 MHz, 512 MB RAM, 20 GB disk storage or better running Microsoft Windows 2000

Desktop System – Microsoft Windows NT 4.0 or Microsoft Windows 2000 Professional Workstations

### Software:

Microsoft Windows NT 4.0 or Microsoft Windows 2000

To learn more about the Stratify Discovery System, visit [www.stratify.com](http://www.stratify.com).



Stratify Headquarters  
201 Ravendale Drive  
Mountain View, CA 94043

Ph: 650.988.2000  
Fx: 650.988.2159  
[www.stratify.com](http://www.stratify.com)