

Business Search vs. Consumer Search

Five Differences Your Company Can't Afford to Ignore

Ultraseek White Paper

Ultraseek

Business Search vs. Consumer Search

Five Differences Your Company Can't Afford to Ignore

January 2006

Table of Contents:

Business Search vs. Consumer Search	3
Five Differences You Can't Afford to Ignore	3
1. <i>Return Role-based Results</i>	3
2. <i>Provide Multiple Methods of Searching</i>	4
3. <i>Search All Corporate Information Repositories</i>	4
4. <i>Support Multiple Languages and File Formats</i>	5
5. <i>Enforce Corporate Security Models and Compliance Policies</i>	5
'Why I Chose Business Search over Consumer Search'	6
Customization is Critical	6
'Great for Quick Demos, but Not for Real Applications'	7
Ultraseek: Search for Business Users	7
Key Functionality	7
Highly Targeted Search	7
Business-strength Security	7
Multiple Methods of Search	7
Tunable to Your Business	7
Multiple Language Support	8
Multiple IT Platform Support	8
Search Analytics	8
Distributed Infrastructure Support	8
Unstructured Data	8
Downloadable	8
Affordable	8
Built for Business	8
Other Autonomy Products	9
Autonomy's Intelligent Data Operating Layer™ (IDOL)	9
IDOL K2	9
IDOL Enterprise Desktop Search	9
Contact Autonomy	9
About Autonomy	9

Publisher's Note: Information contained in this document is intended for guideline purposes only. Autonomy product documentation supersedes information contained in this document. The situations described in this document are offered as examples; actual configurations and results will vary from system to system.

Business Search vs. Consumer Search

IDC Warns: 'Requirements are Inherently Different'

Selecting the right search engine for business can challenge even the most seasoned technology buyer. The choices are many. The decision is pivotal. And the purchasing process is often complicated by requests from business end-users who want the same tools they use to search for information on the Internet deployed at the office.

But be careful what your end-users wish for. Not all search is created equal. What constitutes effective search in one situation can prove highly ineffective in another.

“The business requirements for finding information are complex and may require a spectrum of tools that extend beyond a simple search engine,” says Susan Feldman, IDC’s Research Vice President for Content Technologies. “These tools may include easy interfaces to help tune relevance to increase online sales, reporting tools to monitor usage and uncover questions that aren’t answered, as well as extensive security.”

If your business is serious about its search strategy, you need to know the fundamental differences between searching for information on the Internet and finding the right document quickly inside your corporate intranets, public websites and partner extranets. The implications of understanding these differences, and being able to evangelize them in your enterprise, are significant. Because the cost of deploying the wrong search tools are staggering.

IDC says the average employee spends 3.5 hours every week on searches that fail to locate the desired document and another three hours every week recreating content that couldn’t be found. Over the course of a year, the salary costs alone of ineffective search are more than \$9.7 million for a 1,000-employee business.¹ And these numbers don’t take into account the cost of employees finding and acting on out-dated or incorrect information.

This white paper provides an in-depth look at the unique business search needs of today’s commercial enterprises and government agencies.

Specifically, the paper:

- Identifies five critical requirements that no company can afford to ignore when comparing business search and consumer search.
- Provides first-person accounts from IT professionals who compared business search and consumer search.
- Describes the capabilities of Ultraseek, the world’s leading business search engine.



Five Differences You Can't Afford to Ignore

Although employees frequently seek Web-like search experiences in their workplace, there is a myriad of reasons why the consumer tools used to search for information on the Internet have limited success for businesses.

Business is different, and its requirements are specialized, from the way information is stored to the way information is used and secured. The unique search requirements of commercial enterprises and government organizations highlight five important differences between business search and consumer search.

- 1 Return Role-based Results**
The tasks for which employees use information vary widely, depending on their department and their role within their company.

Business Search

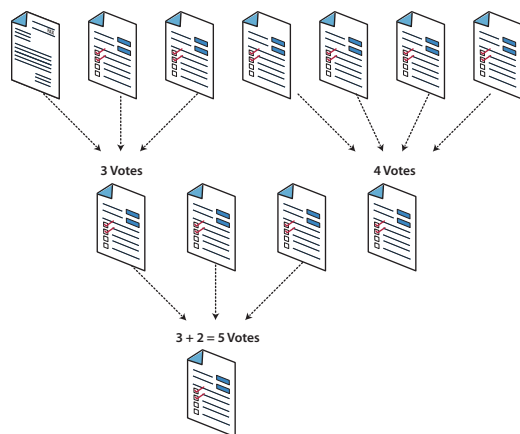
Business search recognizes that everyone in a company is different. This means that search relevance can be tuned to better meet the unique needs of employees in any department or level. Content can be organized to reflect the way employees view information. And answers can be returned so employees see information in the context of the question or problem they are trying to solve. This is important because when someone

¹IDC Content Technologies Study, 2004-2005.

in Finance searches for “hedging short positions,” they’re looking for specific information in a time-sensitive manner. What they don’t need are proposals Facilities received from landscapers to plant shrubs in front of the office.

Consumer Search

Consumer search treats all users the same. The rich linking structure—that some consumer search engines rely on to determine relevancy—pays little regard to an employee’s role or unique search needs. In essence, the more links to a page from other pages and websites, the more relevant the creators of those websites think that page is. In a business environment, this rich linking doesn’t exist. How many links are there between word processing documents, PDF files and spreadsheets on your intranet? In most corporate environments there aren’t any, or are very few.



How Consumer Search Works: Cumulative links into a page determine relevance. For example, the file above with five votes, would score highest with a consumer search engine that relies exclusively on link analysis.

2 Provide Multiple Methods of Searching

Standards for search relevance is higher in business. Employees want a single, correct answer to their information request.

Business Search

As many as 80% of all search queries in a business target documents an employee has already seen or even authored². This means search results in the enterprise must provide more than a good answer. They must provide the right answer.

With multiple ways of searching for information, business search meets the high demands that employees have for pinpointing relevant information. An employee can find information up to 50 percent faster³ by navigating through aptly named categories. Users can also expedite the discovery of relevant information through capabilities such as advanced full-text search, federated search, spelling suggest and relevance tuning.

For example, when equipped with information classified into relevant categories, a service agent at an airline’s call center can quickly retrieve schedules of an affiliated airline’s

flights, along with the affiliated airline’s special diet meal policies. When a customer asks for a ticket from “Geneva” to “the Bay Area,” concept-based search capabilities return flights from Geneva, Switzerland, to San Francisco, California, on the airline’s affiliate, Zurich Air. Using the same interface, the service agent drills down through standard categories and subcategories labeled “Affiliated Airlines” and “Luggage Policies” to immediately locate the affiliate’s carry-on restrictions.

Consumer Search

Consumer search engines provide general keyword search—and little else. While this one-size-fits-all approach to finding information may work for a user looking for a potato salad recipe on the Internet—who has time to sort through results and refine searches—it is insufficient for the complex requirements of business.

Take the example of an employee at a door and window manufacturing company who needs information on troubleshooting the Microsoft Windows 2000 operating system on his computer. If the intranet uses a consumer search engine, a search for “Windows 2000 troubleshooting” would deluge the employee with instruction manuals for the hundreds of types of glass windows the company manufactured in the year 2000. But with a business search taxonomy, the employee could either browse directly through the IT category to the Microsoft Windows 2000 subcategory, or he could dramatically reduce the number of results returned by limiting his search for “Windows 2000 troubleshooting” to the IT category, rather than the entire intranet.

3 Search All Corporate Information Repositories

Corporate information is spread across a host of specialized secure business applications, databases, content management repositories, email systems and Web servers—all of which require special interfaces.

Business Search

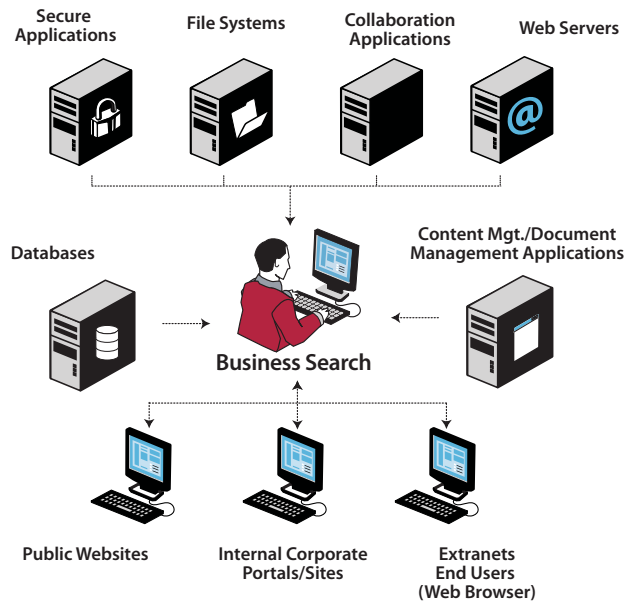
Increasingly, business decisions depend upon the ability to find and correlate unrelated content from totally different areas of the enterprise. With pre-built connectors and/or federated plug-ins to a wide variety of repositories and applications, business search can index database records, documents in file systems, email servers, content management systems, pages on highly targeted Web sources, and so on. This gives employees a 360-degree view of their business and the content spread across it, so they can search it with a single query.

For example, with business search a product safety officer at a global industrial company can unify search across multiple sources—including the corporate intranet, R&D documentation in a content management system, customer complaints and warranty claims in databases and the reports from the public website of a government testing laboratory. The

² Rajat Mukherjee and Jianchang Mao, “Enterprise Search: Tough Stuff,” April 2004.

³ H. Chen and S.T. Dumais (2000), “Bringing Order to the Web: Automatically Categorizing Search Results.” An evaluation of generic user interfaces used to present search results. Proceedings of CHI’00, Human Factors in Computing Systems, pp. 145-152.

results are merged and ranked for relevance, which enables the product safety officer to quickly ferret out information on defects in the design and manufacturing of the complex products his company makes.



Consumer Search

Consumer search is limited in its ability to aggregate information from disparate business repositories. Non-secure Web servers can be indexed out of the box. But, integrating the information from databases, file systems and content management applications into a consumer search engine is considerably more complex—and in some cases impossible. This means employees frequently face the prospect of using consumer search for their intranet content, and the native search engines for each of the other repositories in their business. The result is that employees waste time submitting multiple queries to multiple systems, and manually assimilating the results to put everything in perspective. And if they don't know that a system or application exists, the valuable information in it is missed entirely.

4 Support Multiple Languages and File Formats

Employees need to access business documents in any language and from a dizzying array of word processing, spreadsheet, presentation, graphic, multimedia, compression and encoding formats.

Business Search

More than 80% of the content on a business' network is unstructured information. And in the typical business, less than 10% of that content is found on Web pages⁴. The rest could be in any one of hundreds of unique file formats, diverse applications and repositories, written in multiple languages and protected by any number of security models. Business

search can filter and index text as well as several metadata fields including author, date and title out of all of the file formats and applications that office workers use everyday for their email, presentations, word processing documents or electronic forms. Business search can also index content from virtually any language an organization conducts business in.

This means a single high-risk email containing phrases in multiple languages and attachments with multiple MIME-types can be easily identified and searched to determine if it violates privacy or SEC regulations.

Consumer Search

Consumer search was built to search uniform HTML content on the Internet. It provides only limited abilities to filter, index and view many of the unique file formats in use in a business environment. As a result, businesses must install and license a myriad of native applications on every employee's desktop in order for them to view the returned non-HTML documents. In truly heterogeneous business environments where multiple languages are used inside the multiple file formats, consumer search becomes an even greater obstacle. Consumer search engines do not provide advanced language support such as stemming that most businesses with a global profile require.

5 Enforce Corporate Security Models and Compliance Policies

Access to corporate content must be securely managed in the face of a new matrix of government regulatory mandates and privacy concerns.

Business Search

Just as important as making all of your information accessible is ensuring that sensitive information remains secure. Business search engines enforce existing security systems and honor access profiles for information sources across an enterprise. This protection is extended to result lists and category views to ensure that no information is leaked by displaying document titles or summaries to individuals that don't have authorization to access documents. Without this protection, businesses may be forced to avoid indexing sensitive content—making quick and reliable access for those who need this information virtually impossible.

As an example, business search security can be used to protect sensitive human resources content. If a business indexes all of its human resources information into a single collection, its administrator can control the access rights for employee groups across the enterprise. In the case, the administrator probably wants to restrict access to select staff in the human resources department. This way, if an employee in manufacturing queries for enterprise-wide salary information, the business search engine searches only the collections this employee can access. The query will not return any information from the HR index, and therefore, no salary data appears in the search results list.

⁴ Enterprise Search: Tough Stuff, Rajat Mukherjee and Jianchang Mao, April 2004

Consumer Search

Consumer search provides open access to most documents—a potential hazard for businesses needing to keep proprietary information under wraps, respect privacy concerns or enforce regulatory mandates. Without document-level security, businesses have no choice but to avoid indexing sensitive content—making quick and reliable access for those who need the information virtually impossible. Equally problematic are the inabilities of consumer search to provide a 360-degree view of the information scattered across the business and the advanced search tools to pinpoint potentially compromising documents.

Summary of Key Differences	
Business Search	Consumer Search
<ul style="list-style-type: none">• Tunable to unique needs of users• Multiple methods of search• Supports multiple repositories, file formats and languages• Compliant and secure	<ul style="list-style-type: none">• One size fits all• Keyword search• Limited support beyond beyond uniform HTML content on non-secure Web servers• Open access

'Why I Chose Business Search over Consumer Search'

As IT managers in organizations across the world evaluate search products for use on their intranet, extranet and public websites, they learn firsthand about the significant differences between consumer search and business search. Here, are two of their stories.

Customization is Critical

By John Luu, IT professional from Canada

As a software engineering professional at one of Canada's largest corporations, I was asked by my manager in 2003 to find an effective search engine for our intranet. The outdated engine in use had serious issues. More often than not, when employees searched for information, they would not find what they were looking for until three or four pages into the results list.

Selecting a new search engine for our intranet became a top priority and came down to a choice between two solutions—a business search engine and a consumer search appliance. The right choice would install easily, provide high search relevance and be customizable to the complex business and technology requirements of our company.

We gave each of the products a list of the top URLs and evaluated them on their ability to spider five levels deep. We also tested the products on their ability to integrate with our portal technology. This is where business search came out the clear winner. Business search enabled us to rapidly deploy multiple search interfaces and result pages tailored to the look, feel and usability requirements of our company.

The consumer search appliance, on the other hand, was essentially a black box. There was no easy way to modify any of its templates. The only way to get anything changed is to actually hook up an Ethernet jack to point to the outside world so the vendor of the consumer search appliance could reconfigure the box to our satisfaction. This simply wouldn't have worked for us. We wanted the tools to control all aspects of our intranet.

Business search became the standard on the corporation's 450,000-document intranet. It helps more than 25,000 employees from across Canada find the information they need to do their jobs more effectively.

We made the right decision. I've since changed jobs and was pleased to see that my new employer, the largest national newspaper in Canada, also successfully deploys a search engine—internally and on its public websites—that is built for business, not the Internet. If I had to make the choice all over again, I would definitely recommend business search over consumer search.

'Great for Quick Demos, but Not for Real Applications...'

By Joe Kuefler, Principal Engineer for Computer Associates

Concord Communications, now wholly owned by Computer Associates, was building an internal search application to index engineering content (especially source code). We were attracted to the simplicity of popular consumer search sites and wanted to model that behavior inhouse.

Our first approach was to attempt to implement our application on a consumer search product directly. Although we were able to demonstrate some early success, this approach ultimately failed, for a number of reasons, once we moved beyond a simple proof-of-concept. At that point we had to deal with the more complex realities of our business requirements including:

- Custom results pages
- Both HTTP and file system crawling
- Heterogeneous authentication schemes
- Database queries
- The cost of an unlimited collection size
- C/C++ APIs for feeding and querying the collections
- "Crontab" driven batch jobs

We essentially needed to build a "real application" on top of a search engine that the very high-level and simplistic approach taken by the consumer search engine couldn't support.

That's why we selected a platform for specialized business search applications. Using the BIF feature, we could tailor the platform to index only the content we wanted. The platform also gave us exactly what we needed in terms of a low-level C/C++ API into the resulting collections. In addition, the business search platform has plenty of tuning options and runs on both UNIX and Windows. The consumer search engines only offered Web interfaces and no underlying APIs.

The quick demo that we created with the consumer search engine allowed us to prove the concept of the application and to essentially create a budget for the real application. After purchasing a business search platform in November 2003, we were under tremendous pressure to finish our implementation quickly. Because of the reliability and consistency of the product, we were able to go live only two months later and quickly became the most successful search application at Concord Communications. Previously, it could take as long as 45 minutes to recursively "grep" across our massive code base. With our business search application, we can perform these same searches in a few seconds and interleave search results from documentation, tech support cases, customer data and personal knowledge bases. The application remains live as of the publishing of this white paper and sets the standard for search within Computer Associates.

Ultraseek: Search for Business Users

More than 3,500 customers across the globe—from small to mid-size organizations to global enterprises—rely on Ultraseek, an award-winning business search engine from Autonomy Corporation. Ultraseek helps you find the information you are looking for, whether it's on an intranet page, a presentation on a file server or a report in a content management system. To see Ultraseek, and the features described below, download the full-version of Ultraseek for free from www.ultraseek.com, and be up and running in about 15 minutes.

Key Functionality

Highly Targeted Search

Simultaneously search for relevant information inside and outside your business from a single window—including fast, straightforward integration with existing portal, content management, targeted Internet sites and other systems.

Business-strength Security

Enforce your organization's specific security and compliance models. Security permissions provide precise access control by job function or role.

Multiple Methods of Search

The most complete range of business search engine tools available—including numerous full-text search options, federated search, and navigation—to dynamically lead employees to the information they need.

Tunable to Your Business

Customize indexing and search relevance to different functions and roles across the organization. Data can also be organized to reflect the way individual workers view information.

Superior Relevance

Ultraseek provides a variety of search options including keyword, phrase, wildcard, field, range, implied Boolean, and advanced operators. Relevance is determined through an optimized combination of enhanced linguistics techniques, stemming, link analysis, and weighting based on term location, frequency, and rarity. Thesauri and custom dictionaries can also be integrated to support specific industries and business objectives.

Relevance Tuning

Ultraseek's search results can be tuned to match information structure and end-user needs with relevance tuning options that deliver comprehensive control over the relative weighting of metadata fields. In addition, Ultraseek's exclusive Quick Links feature provides editorial control over search results, allowing keywords to be manually associated with specific URLs that are returned above normal search results.

Multiple Language Support

Index documents from virtually any language your organization conducts business in. Advanced linguistic capabilities including language recognition, part-of-speech tagging, stemming, and tokenization of sentences and phrases.

Multiple IT Platform Support

Ultraseek fits with your organization's existing IT infrastructure and leverages technology investments by running on many platforms, including Windows, Linux, AIX, HP-UX, and Solaris.

Search Analytics

The Ultraseek Reporting Manager generates detailed reports that analyze search usage and provide insights into how you can improve relevance in real time.

Standard reports include:

- Top queries
- Queries with no results
- Queries with no click-through
- Most requested documents
- Query trend analysis
- Complete server usage summary

Distributed Infrastructure Support

Ultraseek is easily configurable so you can delegate the control of search interface features to business-line managers. Distributing administration duties enables local customization to create targeted user experiences, and eliminates the scalability obstacles resulting from centralized control.

Unstructured Data

Ultraseek's intelligent, adaptive spider can index all of your content—including as many as 300 unique file formats—into a common index so all of the information important to your business can be searched with a single query. Ultraseek can even spider JavaScript, allowing it to find and index dynamically generated Web pages. Structured and unstructured

information sources can be included in a single index through Ultraseek's optional lightweight database connectivity

Downloadable

Ultraseek is downloaded directly from www.ultraseek.com. Install, crawl and index, and start searching with Ultraseek in a matter of minutes. There are no installation CDs or hardware to order, and you can even download and install Ultraseek as a free, 30-day trial version.

Free 30-day Evaluation

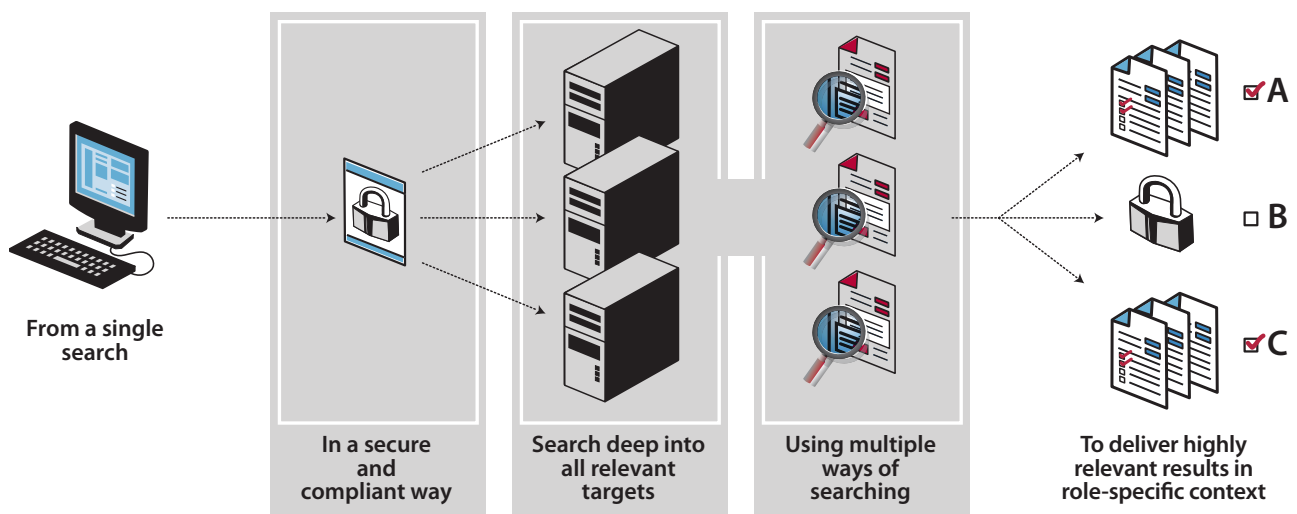
Ultraseek can be downloaded, installed and used to index and search up to 25,000 documents for 30 days, making true enterprise-class search available to businesses of all sizes. And combined with Ultraseek's extremely low ongoing maintenance requirements, this makes evaluating market-leading enterprise search on your intranet or website virtually risk free.

Built for Business

When you select Ultraseek—or any other Autonomy product—your investment is backed by the stability of a profitable, market-leading provider of unstructured information management for businesses. Autonomy has the real-world experience you need to ensure that your project is successful. This experience can be seen in the depth and breadth of our products, our world-class professional services organization, comprehensive technical support and our base of more than 17,000 customers around the world.

Other Autonomy Products

Whether you select Autonomy for business desktop search, the Ultraseek business search engine, or our most advanced platforms to automatically analyze and process any piece of information, your enterprise can expect a wide range of meaningful and measurable benefits.



information. Using this off-the-shelf solution, organizations can process digital information automatically and communicate with multiple applications without the need for manual processing or metadata. IDOL is entirely data agnostic and scalable, thereby allowing large organizations to manage vast quantities of information regardless of format or storage location.

K2

IK2 grants full access to IDOL functionality including video and audio search capabilities while faithfully maintaining the established and award-winning K2 functionality such as parametric selection, relational taxonomies, faceted navigation, Topics, VQL, K2 Taxonomies and K2 vector-based categories to name but a few.

IDOL Enterprise Desktop Search

IDOL Enterprise Desktop Search runs on any PC in order to make the corporate desktop entirely searchable. A secure, enterprise version of desktop search, IDOL Enterprise Desktop Search integrates with all other Autonomy modules making content from secure corporate networks, intranets, local data sources, the Web as well as information on the desktop, such as email, office documents readily and easily accessible.

Contact Autonomy

If you are evaluating the effectiveness of your search or unstructured information management strategy and would like additional details about which Autonomy solutions are right for your specific needs, we can help.

Call 800.935.6246 to speak to a sales representative and schedule a consultation.

About Autonomy

Autonomy Corporation plc (LSE: AU. or AU.L) is a global leader in infrastructure software for the enterprise. Autonomy's technology powers applications dependent upon unstructured information including Call Center Solutions, Customer Relationship Management, Business Process Management, Knowledge Management, Enterprise Portals, Enterprise Resource Planning, Online Publishing and Security applications.

Autonomy's customer base comprises more than 16,000 global companies and organizations including, among others, ABB, AT&T, Bristol-Myers Squibb, BAE Systems, Ford, Ericsson, Shell, Nestlé, AOL, BBC, Reuters, Hutchison 3G, Royal & Sun Alliance, Sun Microsystems, Philips, Boeing, Schneider Electric, Coca Cola, GlaxoSmithKline, Hewlett-Packard, Citigroup, ABN AMRO, Deutsche Bank, Nomura, the New York Stock Exchange, Daimler Chrysler, Kraft Foods, Lloyds TSB, PricewaterhouseCoopers, the U.S. Department of Homeland Security, the U.S. Securities and Exchange Commission, NASA and the U.S. Department of Energy. Strategic reseller and OEM partners include leading companies such as BEA, Business Objects, Citrix, EDS, IBM

Global Services, Novell, Veritas, Vignette, Supportsoft and Sybase. The company has offices worldwide.

The Autonomy Group includes: Aungate, a leader in technology for Real-Time Enterprise Governance; Cardiff, a leader in business process management; Virage, a visionary in Rich Media Management technology; eTalk, a leading provider of enterprise-class contact center products and Ultraseek, a leading provider of business search engines.

Autonomy Inc.

One Market, Spear Tower, 19th Floor,
San Francisco, CA 94105, USA

Tel: +1 415 243 9955

Fax: +1 415 243 9984

Email: info@us.autonomy.com

Autonomy Systems Ltd

Cambridge Business Park,
Cowley Rd, Cambridge CB4 0WZ, UK

Tel: +44 (0) 1223 448 000

Fax: +44 (0) 1223 448 001

Email: autonomy@autonomy.com

Other Offices

Autonomy has additional offices in Boston, New York, Sunnyvale, Vista and Washington DC, as well as in Amsterdam, Beijing, Brussels, Hamburg, London, Madrid, Milan, Munich, Oslo, Paris, Rome, Shanghai, Singapore, Stockholm and Sydney and Taipei.



www.autonomy.com