

# Northern Light

## *A New Model for the Enterprise Information Portal*

*An IDC White Paper*

*Analyst: Susan Feldman*

### **Profile of the Enterprise Information System**

Enterprise information portals (EIPs) are high-end systems that provide a foundation for decision making and action. They must act as a central gateway to any information the enterprise generates or needs. These portals consist of an information system infrastructure, the enterprise information system, that may stand alone or support additional knowledge management applications such as data mining. The enterprise information system comprises all the functions that create and manage collections: acquiring information from diverse sources such as Web crawlers or authoring tools, indexing materials, organizing those materials into a structure that is pertinent and understandable to the

### **IDC Opinion**

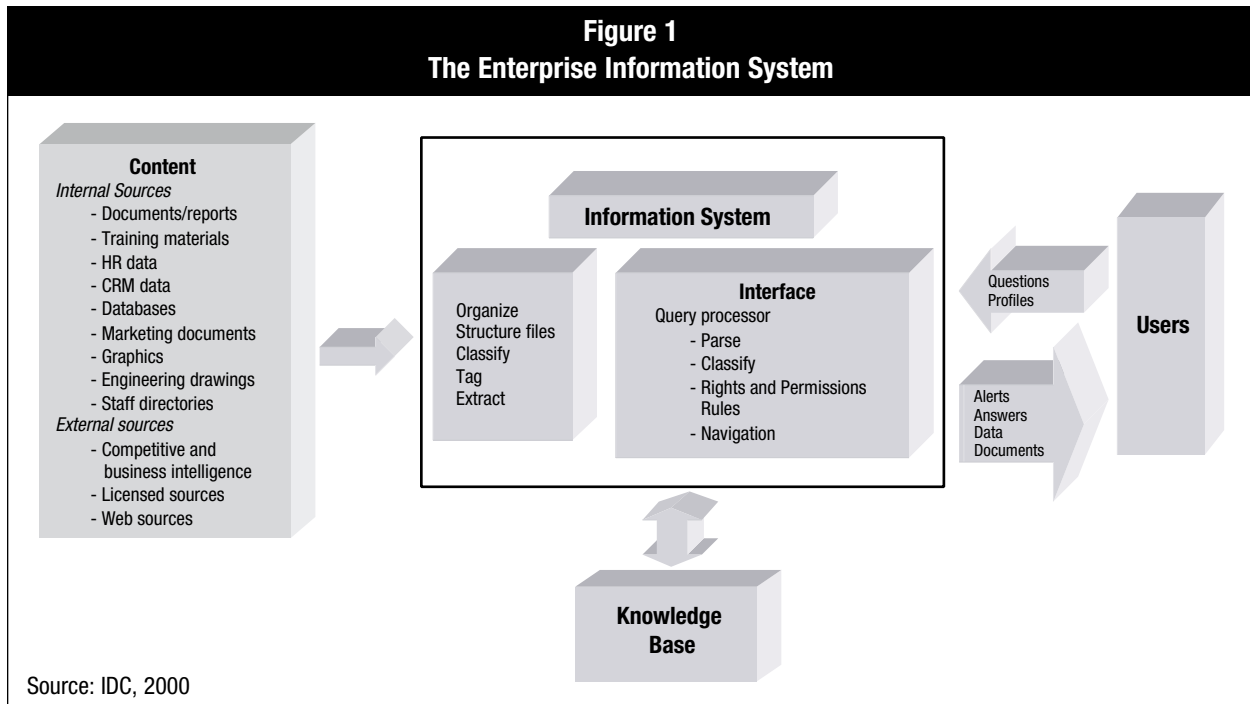
*Will Northern Light's business model for enterprise portal software disrupt the enterprise information portal (EIP) market?*

The EIP stands at the center of any knowledge-based organization. It should be the single point of entry to all the information an enterprise needs to function. It must provide easy access to information, such as unstructured text, databases, image files, presentations, training materials, video or audio files, in internal or external repositories. The EIP may be located in multiple geographic locations and may require different kinds of access for different employees, but it should still be a priority for organizations because it is the single enduring repository of corporate knowledge.

For that reason, we will see increasing emphasis placed on developing and expanding its functions. Companies will seek good retrieval technologies and demand improved navigation and usability.

The Northern Light EIP business model is an intriguing one, and it may well constitute a disruptive model for this market. It is one part application service provider (ASP) and one part outsourced service. It is scalable and quick to implement. In a fast-changing technology environment, this model could be a relief and a boon to enterprises that need to provide excellent and swift access to information.

organization, maintaining versions and archives, and, most importantly, making the information readily accessible through advanced, easy-to-use search and browsing tools (see Figure 1).



Users of enterprise information systems run the gamut from IT and information professionals, to knowledge workers in R&D or market and competitive analysis, to support staff who need basic facts and figures. Alternate views for different kinds of information needs and roles will make the system more user-friendly. Unlike the public Web environment users, enterprise users can be identified and trained. They are a captive audience that can be expected to be more patient than the ordinary consumer on the Web. They are driven by enterprisewide goals and business rules, and they are chronically short of time. In addition, they are embedded within the enterprise culture and are driven by its politics. Any enterprise system will affect and be affected by the environment in which it is used. It should be flexible enough to be redesigned as unforeseen uses are discovered, or as technologies or needs change.

As users become more sophisticated in their use of new retrieval tools, we can expect their queries to become more elaborate and specialized. While documents will still be retrieved by author, number, or date, we can also expect that range searching by date or numbers may be needed, or that searchers will be looking for information in the

Copyright © 2000 IDC. **Reproduction without written permission is completely forbidden.**

External Publication of IDC Information and Data—Any IDC information that is to be used in advertising, press releases, or promotional materials requires prior written approval from the appropriate IDC Vice President or Country Manager. A draft of the proposed document should accompany any such request. IDC reserves the right to deny approval of external usage for any reason.

Printed on  
recycled  
materials



context of a specific industry or market (e.g., a search on “ABC,” where it is known to mean “activity-based costing” and not a major television network). Subject queries should become more complex (e.g., “Methods for placing a value on intellectual assets”), or they may demand an answer instead of a set of documents (e.g., “Who was director of marketing in 1994?”).

### **Challenges**

*The ideal enterprise information system should be a single point of access to one virtual data store.*

The ideal enterprise information system should be a single point of access to one virtual data store. Otherwise, employees run the risk of forgetting or ignoring important information sources, thus missing key data that might swing a decision one way or another. Unfortunately, keeping track of information within a large enterprise is a staggering task. Each department or business process generates collections that are unique in their structure and content.

Information systems are complex. Because of the nature of retrieval technologies, searching well across disparate types of data has become one of the greatest challenges to information scientists. What is the value of a single fact from a customer relationship management (CRM) database when it is weighed against a paragraph from a 10-page, full-text document? Which should be served first to the user? Can the system provide alternate views to each individual, depending on his or her role in the company? Can weights be changed to reflect how the company uses data in its decision-making process? Is the system capable of providing several views into the same data so that it can be examined from different angles? Can access to sensitive financial or personal information be limited?

*Enterprises need advanced systems that can be installed and used with minimal knowledge or effort.*

This complexity demands more than a basic search system. Enterprises need high-end tools for involved information needs. Without them, the data analysis upon which good decision making rests is dangerously incomplete. Yet, keeping up with new retrieval technologies requires a depth of understanding that is outside the expertise or interests of most organizations. Investment in the wrong technology dooms employees to unmitigated frustration. Worse, lost productivity and poor decisions are the result. Enterprises need advanced systems that can be installed and used with minimal knowledge or effort.

### **The Ideal Enterprise information System**

Key elements are required in order for an enterprise information system to be effective.

#### **Search**

Good retrieval is an elusive goal. It depends not only on the sophistication of the search engine but also on the preparation of the information and the underlying structure of contents and the index. In addition, the aptitude, skills, and patience of the users can determine success or failure.

*Good search is crucial to any information system, but in an enterprise, good search may mean the difference between making well-grounded decisions and poor ones.*

Good search is crucial to any information system, but in an enterprise, good search may mean the difference between making well-grounded decisions and poor ones. It has a quantifiable impact on the success of the organization. A good search system returns the information you need, when you need it. While “good” is impossible to define functionally, there are technical capabilities that contribute to making a search system good. An enterprise information system must:

- Return the most relevant documents first
- Search across multiple formats and repositories
- Find documents about a topic even if the author’s and the searcher’s terminology don’t match exactly
- Alert users to new information on topics they are monitoring

### *Relevance Ranking*

Proper relevance ranking determines whether a user feels that a system is effective. If the system finds relevant documents but does not present them early in the search results, the user feels that his or her question has not been interpreted properly or that the system is returning materials haphazardly. Yet, relevance ranking is somewhat in the eye of the beholder. It is difficult, based on a one-, two-, or three-word query, to infer what will be most useful to a particular searcher.

Most systems determine relevance by counting the number of occurrences of a query term in a document. Those documents that have the most occurrences of the most query terms are deemed to be most relevant. But other factors must also be considered. For instance, long documents have more words in them, so the number of occurrences must be averaged over the length of each document. The location of the words in the document may determine how important they are to that text. If they appear in the title, the lead paragraph, or the conclusion, they are more likely to be important than if they appear in the middle of the text in an illustrative anecdote. Many systems take these factors into account.

Relevance ranking is sorely tested when it must return materials of unequal length or different structures. For instance, words in a bulleted list should probably carry more weight than words in a long paragraph. Text in captions for illustrations or within a table is usually quite brief, but it is probably worth more in determining the importance of a subject than an entire paragraph of unstructured text. Systems must compare the content of a PowerPoint presentation with that of a technical report or a news release. A good system would examine the type of document and adjust the relevance weighting for document type and content. The importance of each of these varies within the context in which the search is performed. Ranking algorithms should be adaptable to the priorities of the organization.

### *Integrated Search*

Most knowledge-centered enterprises have amassed a diverse collection of sources. These comprise text documents in various formats, presentations, training materials, graphics and illustrations, Web pages, structured databases, and the contents of CRM and enterprise resource planning (ERP) systems. They are typically distributed all over the company.

Two streams of external content need to be added to the enterprise information system. The first is high-quality, carefully selected content such as market reports, technical reports, business intelligence, and industry studies. This content must be selected and licensed for internal distribution, and it typically carries a high price tag. The second stream is Web content that is pertinent to the enterprise. This content may include competitors' Web sites, product announcements, research in specific areas, or new advances in CRM or human resources. While access to the entire riches of the Web should be available, delivery of focused content ensures better retrieval of the most important materials. Therefore, any system should be able to configure crawlers for tailored Web coverage.

*The ideal information system would be able to search across all content sources and all formats in order to deliver a single answer to a query.*

The ideal information system would be able to search across all content sources and all formats in order to deliver a single answer to a query.

### *Language*

Search systems match questions to textual content. Typically, more than 75% of the materials that an organization churns out are text documents. Since language is notoriously imprecise, perfect matching is a tall order. The problem is compounded by users who rarely ask the right question to retrieve the information that they need. While it is difficult to ask a question on a subject that one knows very little about, matching off-center query terms to the precise language of a subject expert makes it even more difficult to achieve a perfect match. In addition, there are myriad ways to express the same idea. Any search system that can offer synonyms, as well as clues about related topics and terminology, immediately improves the likelihood of searching success.

Some search systems ignore the language problem and return exact matches or nothing. Others measure similarity between documents and queries statistically. These approaches are inadequate for any knowledge-centered organization. Instead, the search system should incorporate advanced retrieval capabilities that:

- Automatically identify phrases in the query and match them against phrases in documents (Phrases are better carriers of meaning than are separate words. Automatic use of phrase searching eliminates many poor or puzzling search results.)
- Identify name and term variants and map them to a single concept
- Eliminate documents that contain the right word with the wrong meaning (Since most words have more than one meaning, exact match systems often return documents that are not relevant but

that contain the terms in the query. Systems that classify documents by topic, or that examine the context of a query and a document to determine the concept or idea, improve the quality of the results.)

*The unpredictable nature of unstructured text makes it a poor candidate for relational database structuring.*

### *Appropriate File Structure*

Systems that retrieve unstructured text such as documents and reports require a different functionality and structure from those that deal primarily with structured data. To search large collections of text well, a system must be able to search across entire documents at once, including the metadata and illustrations. When these elements are stored separately, relevance ranking is impaired. Calculations must be performed on each of the separate elements, and then the results must be joined. In a small system, this is not a problem, but text-based systems grow quickly, and extensive processing slows them down. In addition, the unpredictable nature of unstructured text makes it a poor candidate for relational database structuring. Inevitably, the text is placed in a table as a single object, obviating the usefulness of relational database operations. Enterprise information systems must support such searches as:

- What are current best practices for risk assessment in the insurance industry?
- Notify me when the following companies release a new product.
- Find all instances of airline mergers that have not been approved.

Large databases present some unique problems to an information system. So much data may slow down a system unless it is optimized to handle large amounts of text as well as many simultaneous queries. Large text databases require specialized file structures to enable searching by proximity, phrase identification, entity identification, and other linguistic features. Distributed databases must be searched simultaneously and then have their separate results merged and compared for relevance to the query.

Large databases by their very nature return large numbers of documents in response to a query. These documents must be presented to the user so that the entire retrieved set can be comprehended at a glance and navigated easily. Their size mandates that they will be slower to search unless searching techniques are optimized.

### *Alerting Features*

Most employees have continuing needs for the same kind of information: “What are new techniques for risk assessment?”; “Send me all changes in management at my competitors’ companies”; “Notify me about new trends in benefits packages”; and “Alert me to any new research on theophylline.” A good information system should be able to collect preset profiles of employees’ interests and run them against

*A good information system should be able to collect preset profiles of employees’ interests and run them against all incoming new materials.*

all incoming new materials. It should also be able to establish a repeating query from any query that a user stipulates, without lengthy procedures or forms.

### ***Categorization and Clusters of Related Terms***

Good categorization creates the foundation for a good search system. Categorization adds indexing information to documents as they are entered into the system. This indexing metadata should include title, author, date, format or document type, and other identifying data. It should also include subject headings so that documents about a single concept are grouped together. The process must be automatic in order to avoid bottlenecks. Additionally, it must be intelligent enough to avoid “dumb computer” errors.

Using categorization, a search system can find related materials even if all the query terms are not in the document. Good categorization also builds clusters of related terms and synonyms. These are used to find documents that may be about a topic even if they don't use the terms in the query. They expand a query to enrich it with appropriate additional terms. The clusters are also valuable for eliminating terms that appear in the query but are used with a different meaning in the document.

Categorization can be used to create “instant directories” that show the user the contents of a set of search results by categories so that the most useful results can be navigated to quickly. Most importantly, categorization enables a search system to hone its results, eliminating many of the irrelevant hits that plague systems that do not categorize.

Automatic directory building also requires categorization. Since some users prefer to browse rather than search, directories are a valuable information finding tool that improve any information system.

### ***Usability and Good Interface Design***

Usability is a critical component of any system. Design, reliability, predictability, and understandability are all facets of usability. A system must make its features known and understandable so that users can comprehend how to use it as well as how it works. A system must be available when needed, and it must return the same answers to the same question, barring updates to content.

Usability requires an easy-to-use interface that will accept all kinds of queries — questions, sentences, keywords, or Boolean commands — and interpret them as they are intended to be interpreted. The interface must be easy to navigate and offer visual navigation features and multiple views for filling different kinds of information needs. Ease of use is important for end users, system administrators, and content contributors.

Good information finding requires many kinds of functionality within a system if it is to serve the varied needs of an organization. Both

*Usability requires an easy-to-use interface that will accept all kinds of queries — questions, sentences, keywords, or Boolean commands — and interpret them as they are intended to be interpreted.*

browsing and searching should be supported. While searching is best for finding answers to specific questions, it does not answer questions such as “What is in this information system?” or “What marketing materials or company profiles have already been written that I can use for my next press release?” Thus, a good information system should support browsing and searching by creating and keeping a directory of information that can be viewed from several perspectives (e.g., by subject, type of materials, author, or date). Based on the extensive categorization described above, the directory should present materials in a hierarchy that makes sense to the user community.

### **Flexibility**

The needs of any enterprise must be reflected in the structure of its information system. Terminology should match internal use. Relevance ranking must reflect the priorities of the organization. Therefore, any good search system should be adaptable and customizable as follows:

- The taxonomy must be easy to change and modify.
- Access should be customizable for different parts of the organization, based on business rules and need for sensitive information.
- Search boxes should accommodate different types of queries -to retrieve employee records by date or name or number, to ask for facts, to launch Boolean queries, to ask natural language queries of any length. A good system should let the user cut and paste an entire passage into the search box in order to locate the source document, as well as any other documents in which it was used or paraphrased.
- Relevance ranking should be adjustable to meet the priorities of the enterprise, giving higher weight to types of documents or repositories that are central to the organization.
- Personalization of the interface or the type of information delivered should be possible based on an employee’s role in the company or need for information.
- Customization should be possible of the interface appearance, of the external content searched, and of the internal crawler to crawl sites according to enterprise-set schedule.

*A good system should let the user cut and paste an entire passage into the search box in order to locate the source document.*

### **Administration Tools**

Any system requires tools to enable system administrators, content administrators, and appropriate staff to organize materials and analyze usage. These include:

- Reporting tools for tracking usage, analyzing queries, and determining which information sources are of most value to employees

- Workflow features to control and track input of and changes to content
- Taxonomy construction and maintenance tools
- Versioning and rollback capabilities

### ***Security***

Secure access to sensitive or proprietary information is a major concern for all enterprises. In addition to protecting its intellectual capital, an organization must ensure the privacy of its employees. The problem is compounded by a mobile workforce that requires access from within or from outside the firewall. Therefore, an enterprise information system must contain:

- Built-in rights and permissions processes that use established business rules
- Hacker-resistant architecture
- Secure, private access to data from any location worldwide

### ***Scalability***

Information systems grow quickly. The more successful they are, the more quickly they grow. What begins as a useful tool for a single department may eventually spread to an entire enterprise as well as to its customers and suppliers. Some information architectures are more scalable than others.

*Some information architectures are more scalable than others.*

### ***Integration***

This ideal EIP does not exist. Enterprise needs vary. So do vendor offerings. The information system is the foundation, infrastructure, and heart of any enterprise portal. However, enterprises may need to add CRM or ERP systems, data mining systems, and collaborative applications for conferencing and collaborative work environments. The degree to which these elements can be integrated easily determines how extensible the system will be. It is vital that the eventual enterprise portal be presented to users as a single, integrated whole. It must be the only logical starting place for working within the enterprise so that pertinent information is found and used as a matter of course.

### ***Minimum Upkeep***

Changes to any system are time consuming and therefore costly. In addition to installing and distributing software, and making sure that older content is compatible with changes in document structures, taxonomies, or terminology, changes affect the users of a system. Multiply the learning time required for each user by the number of users in the enterprise, and the hours lost to productive work are astronomical. Changes which can be invisible to the user and which do not require extensive in-house IT staff time are therefore preferable.

### **Enterprise Information System Checklist**

- High-quality search with:
  - Sophisticated relevance ranking
  - Integrated search of all formats, repositories, and data types
  - Concept searching
  - Alerting features
- Categorization for improved searching, navigation, and automatic directory building
- Ease of use (e.g., interface, browse and search capabilities, predictable, reliable, and understandable features)
- Flexibility (e.g., of taxonomy, interface design, ranking algorithms, addition of new content sources, and crawler)
- Personalization features
- Use of business rules to customize delivery of appropriate information
- Tools
  - For taxonomy building and maintenance
  - For workflow and versioning (e.g., check-in/checkout and permissions)
  - For administration and usage data analysis
- Security
- Scalability
- Integration
- Minimum upkeep

*IDC predicts a compound annual growth rate (CAGR) of 46% for knowledge management access software by 2004.*

### **The Enterprise Portal Market**

The market for enterprise portal applications is growing rapidly. IDC predicts a compound annual growth rate (CAGR) of 46% for knowledge management access software and a CAGR of 31% for knowledge management infrastructure software by 2004. Preliminary figures for the search and retrieval segment of the document and content technologies market show a CAGR of 66% by 2004.

Clearly, this market is growing rapidly. We believe that the market for enterprise information systems is only beginning to expand. New retrieval and analysis products now in development, as well as increasing demand for good information systems, will fuel growth through the next five years. Products within this market include knowledge management infrastructure applications that build and manage collections of structured or unstructured data, and knowledge management access applications that provide access to a knowledge base, including search and data analysis. Companies within this market offer software and custom services, which are used to develop internal information

systems for enterprises. They have entered the arena from diverse areas of expertise. Document management companies bring strong workflow and permissioning capabilities. Data warehousing vendors understand how to store and retrieve structured data. Web commerce companies bring strong design and transaction capabilities as well as skill with HTML, XML, and internetworking. Full text retrieval companies address the core of the problem: They create and structure repositories of full text that are appropriate for advanced text searching and mining. Content providers provide the external materials, sometimes with specialized intranet information system capabilities as well. We can expect both competition and partnerships to develop among these groups in the next five years. Vendors that offer products which compete in this market include, in addition to Northern Light, Autonomy, Verity, Plumtree, AltaVista Business Services, Inktomi's Ultraseek, BroadVision, Interwoven, LeadingSide (formerly Dataware), Excalibur, Hummingbird, IntraNet Solutions, FileNET, Documentum, Factiva, and Dialog.

*Northern Light is an advanced information system that has confronted the technical challenges of retrieval, categorization, and usability successfully.*

### **Northern Light Technology**

Northern Light is an advanced information system that has confronted the technical challenges of retrieval, categorization, and usability successfully. It provides excellent search capabilities, and its relevance ranking has been proven in tests using the U.S. Government's NIST TREC (Text Retrieval Conference) methodology. As an established Web search engine, it has demonstrated that it can search across very large and diverse collections of materials in multiple formats. It can handle large numbers of simultaneous queries in many formats, including keywords, Boolean expressions, and complete paragraphs of text. This success is due to Northern Light's willingness to harness and combine techniques from retrieval, artificial intelligence, indexing and cataloging, usability and cognitive studies, and plain common sense.

### **Search**

Northern Light searches across any and all sources of information to create a single, integrated search result. Formats searched include the text contained within bitmaps as well as all text and structured database contents. Searching on Northern Light appears easy because the complexity of the process is handled within the system. The user need not think of every possible expression of an idea because the system considers alternatives automatically. Queries can be enriched from the clusters of synonyms and related terms in the classification system. The system contains built-in lexicons and phrase identification capabilities. In addition to the lexicons and the categorizer, Northern Light is improving its phrase recognition. New tools analyze queries for significant and non-significant words. False drops are kept to a minimum by recognizing when a term is used with a different meaning, a process called "disambiguation."

Ranking algorithms are complex. In addition to word frequency, frequency of terms in the whole database, and document length, the Northern Light system considers the type of document and weights its contents accordingly when establishing relevance rankings. Metadata and implicit phrase recognition are also used to determine a document's relevance. Web pages are given higher weights if they are linked to by authoritative sources, a process called "link analysis." The system can treat subsections of documents separately so that tables and chapters can be retrieved at a smaller level of granularity.

Data from structured databases can be mapped into Northern Light's system so that it can retrieve across data types and formats. Additionally, Northern Light is capable of extracting the text content from bitmapped images so that the text is searchable. Enterprise customers can expect Northern Light will map their internal and external sources into the system in order to develop appropriate algorithms for searching across all data types. The enterprise can also change weighting internally to deliver specified types of content first.

In addition to extracting and adding metadata for title, author, date, location, language, source, major names, subject, and text structure, Northern Light "chunks" the documents into logical pieces so that they can be retrieved based on the relevance of the "chunk" or as part of the larger work. This process improves retrieval of longer works, since a single chapter may be relevant to a search while the entire work may not be. In addition, tables are treated separately. New data sources are analyzed before they are added to the system in order to optimize their retrieval.

*Alerting and monitoring profiles can be established for either individuals or departments on any schedule.*

Alerting and monitoring profiles can be established for either individuals or departments on any schedule. In addition to profiles, any query may be used as a standing alert to monitor a topic continuously. If users miss an alert, they can view all those they did not check the next time they log on. Alerts can be set for any interval.

### ***Classification***

The Northern Light classification system is an amalgamation of traditional, well-established taxonomies of world knowledge, specialized taxonomies, current terminology, and indexer expertise. Northern Light uses its extensive classification system to improve searches and to create browsable directories automatically. Good classification eliminates ambiguous results. It organizes the "banks" in "river banks" into one category and the financial institutions in another. It also puts "hypertension" and "high blood pressure" into the same subject, making it possible to retrieve by concept or idea rather than by terms alone. Northern Light uses its taxonomy of knowledge as a basis for classifying every document it searches. New terminology is added judiciously by Northern Light indexing experts as well as from suggestions by customers who seek additional granularity for their systems. New terms

are also identified as new content is added. While the categories have been developed by experts, the system uses artificial intelligence judiciously to develop and apply classification rules to each document. The system can thus index the unimaginably large contents of the Web.

Search results are classified to display results visually in subject folders, as well as by source or document type. This is a notable and patented feature. Users are offered a complete overview of their search results at a glance. They can browse the top documents or examine a subject folder separately to narrow the scope of the retrieved set without having to craft another query.

*Customers can have their own taxonomies mapped into the Northern Light taxonomy.*

The taxonomy is extensible. Customers can have their own taxonomies mapped into the Northern Light taxonomy, with rules created for classifying their documents. Northern Light maintains a staff of indexers for this purpose. They catch term variants, create lexicons, review sample results to hone and improve classification rules, and cluster related terms. Taxonomy construction tools are also available for enterprise customers.

### ***Usability and Interface Design***

Northern Light's hallmark custom folders make it very usable. The folders are created on the fly, based on the classification system. They present search results in a manageable, visual display so that the user can see the contents of a search at a glance. This visual approach is invaluable. It saves time and offers ideas for further searching. Since a document can appear in several folders, it is possible to hone a search without being so exclusive that materials are missed.

Searchers who prefer other search methods are offered alternative search screens — by industry, Boolean command, geographic location, or publication name. However, any search screen will accept any type of query. The choices are clear and easy to follow.

The flexibility of this approach makes this interface one of the most usable we have examined.

### ***Customization and Flexibility***

Northern Light will customize its ranking algorithm, interface, content, and taxonomy for its enterprise customers. While a “back office” exists to govern rights and permissions, the enterprise may elect to build in its LDAP and business rules to augment access rules.

*Northern Light also creates custom Web crawlers for enterprise customers in order to deliver focused, selected Web content for vertical markets.*

Northern Light also creates custom Web crawlers for enterprise customers in order to deliver focused, selected Web content for vertical markets. It will process data in any format — images, TIFF files, PDF, or paper — in order to make it searchable.

Geosearch and other special views can be used by the enterprise to limit searches by locality or region, publication type, industry, or publication name.

### **Security**

The system is secure. Enterprise customers can opt for a virtual private network (VPN), a leased line, or HTTPs to secure all content. IP validation or user/ID password log-ons integrated with an existing customer authentication system can be used to govern rights and permissions. Additionally, Northern Light offers secure, low-cost transaction processing and performs authentication and IP validation.

### **Tools**

Analytic and reporting tools enable enterprise customers to analyze usage of both the system and the content. The user interface is customizable, and Northern Light either provides staff to aid in design or works with the customer's design staff. The Northern Light staff will also obtain content licenses for external content or maintain them.

Taxonomy building and maintenance tools are available. In addition, Northern Light works with customers to map their internal taxonomies to its own. However, the company can also host a proprietary taxonomy without mapping it to its own.

Transaction processing tools enable enterprises to maintain subscriptions, track spending limits, identify passwords, track usage or spending by department or user ID, and create transaction histories.

### **Content**

Northern Light will license content for the customer, building on the process it has developed in order to grow Special Collection, an online searchable library of high-value, published content. Web content, of course, can be accessed through its Web search system. Northern Light has also developed 300 vertical collections in order to focus the Web content within an industry or topic, and the company builds custom verticals to suit the needs of a specific customer. Enterprise portals can choose to search all the Web or to confine the search within the vertical collection.

### **Reliability and Scalability**

Northern Light has been in existence since 1996. It was launched in 1997, which makes it an old timer among Web search services. The system, which is robust, has been tested and continues to be tested by the demands of the Web.

*Depending on the content types and sources, implementation ranges from a few weeks to less than three months.*

In addition, the system is quick to implement. Depending on the content types and sources, implementation ranges from a few weeks to less than three months.

## The Northern Light Business Model

Building and maintaining an appropriate information system within an enterprise is a complete business in itself. Enterprise portals require an understanding of information science that is beyond the scope of the normal IT department. Therefore, the outsourcing model offered by Northern Light is an intriguing alternative. It offers the enterprise advanced retrieval technologies across all of its collections of information, but it makes no demands on the internal IT department to understand, build, and maintain a complex information system. Instead, it benefits from the experience of an established information enterprise that has developed technologies to handle exactly the problems the customer faces, but on a much larger scale (e.g., diverse data streams; multiple formats; taxonomy development and maintenance; indexing, metadata extraction, and categorization; ease of use for many levels of end users; maintaining archives; licensing access to external information providers; security of internal content; and merged search across collections).

The Northern Light business model is a fusion of application service provider (ASP), software vendor, and business outsourcer. The typical ASP model provides one-to-many services. It serves as a central access point to numerous products, providing a simpler solution for end users through alliances and partnerships. Typically, it owns the customer relationship, but not the intellectual property of the software, which it licenses from software vendors.

In contrast, business outsourcers manage services on a one-to-one basis, but the customer owns the license to the software.

Northern Light's enterprise portal product fits within both categories. Like an ASP, it creates a single access point to multiple applications and content streams through a combination of software and content alliances. Like the business outsourcer, it customizes its product for each customer. However, Northern Light, unlike the typical ASP, owns both the customer relationship and the software technology. And, while it creates customized products, it also offers one-to-many access to its retrieval system. Therefore, its business model differs from others we have examined in the market, and its potential impact on the market is intriguing.

Enterprise information systems may be well suited to this kind of business model. Information retrieval can be considered a meta-application. It takes the results of multiple applications — authoring tools, CRM and ERP applications, data warehousing applications — and performs a set of operations across all of them in order to merge the results of their output within the context of a specific question or information need. This requires the kind of extensive integration work, along with numerous partnerships and alliances, that is characteristic of ASP models. IDC predicts the ASP market will increase at a CAGR of 91% over four years, reaching \$2 billion by 2003. Therefore, it is a promising choice.

*IDC views the Northern Light business model as a potentially disruptive innovation within the knowledge management enterprise portal market because it provides a logical and simpler solution to the customer.*

IDC views the Northern Light business model as a potentially disruptive innovation within the knowledge management enterprise portal market because it provides a logical and simpler solution to the customer. Enterprises need integrated access to their information. It should be deceptively simple to use that access system so that end users can get the information they need quickly, without devoting too much time to information seeking. Few IT departments have the depth of knowledge in taxonomy building, indexing, user information seeking behavior, and search technologies to build such a system. Because of the complex interaction between the tools which build and manage the collections and the tools which search them, integration of unrelated applications may give less than optimum search results unless extensive integration work is devoted to making both the input and the delivery sides of the information system work together. An end-to-end integrated system is therefore preferable. Few such systems exist, which, in addition, can point to a proven track record of handling the day-to-day demands of indexing and categorizing incoming information and searching it well.

IDC has been tracking new business models in the software industry for some time. It is apparent that the increased complexity of networked applications, the speed at which software applications evolve, and the demands that this process places on overtaxed internal IT resources create a strong demand for a new approach to enterprise information systems.

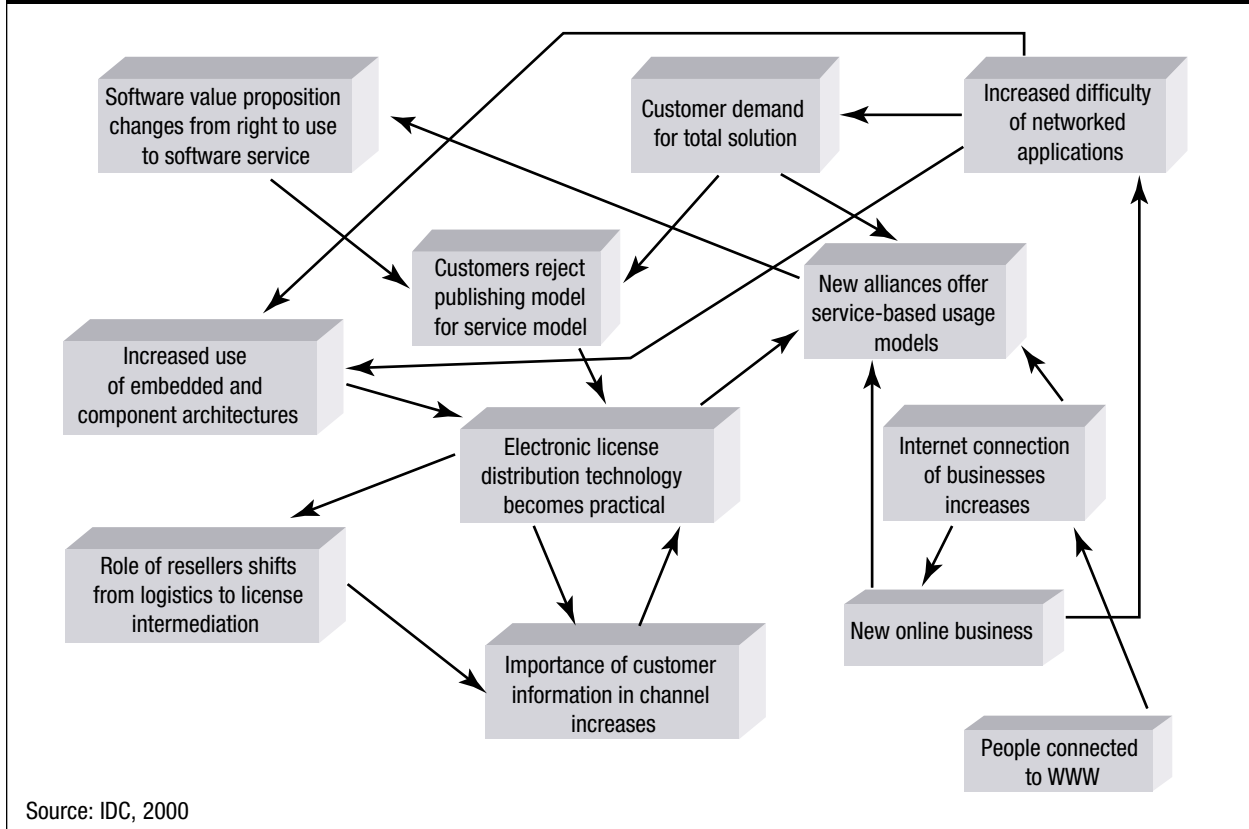
Figure 2 shows a complex interaction of factors that results in the demand for an outsourced service.

We see that the complexity of the software and the use of embedded and component architectures, coupled with increased reliance on the Internet, result in customer demand for a service model rather than for purchasing software outright.

This model is particularly true for the online content sector. Licensing content for an enterprise is a minefield, fraught with subtle legal implications and very little definitive case law. Restrictions on copying and distribution conflict with the knowledge worker's demand and need for unlimited information delivered to the desktop any time and any place. This is the bottleneck which has constricted the development of the online information industry since its inception.

The quality of the information that is used for decision making is an additional problem. With the Internet as the primary content delivery vehicle, the information professional has lost the ability to control the flow of online information to the enterprise. Today, individual knowledge workers forage for their information directly. This creates a dilemma for the enterprise: how to get high-quality information into the hands of the people who need it to make crucial business decisions. Knowledge workers are usually unaware of information sources they cannot find through the public Web. Unhappily, these sources are

**Figure 2**  
**Software Industry Ecosystem Showing Factors That Affect Software Licensing**



usually of higher quality than the average Web page and are invisible to the Web searcher. Even if the enterprise maintains multiple subscriptions to high-quality online services, these sources get lost in the knowledge worker's expectation that everything is available through the Web. Centralized access and merged search is the only answer to this dilemma.

For this reason, it makes sense to create a complete information system that can deliver software and content together. Development, installation, service, licensing, and information are delivered without taxing the enterprise's resources. Updates and functionality arrive automatically.

*Knowledge workers spend at least 50% of their time looking for information, leaving them only 50% to actually use what they have found. Any improvement to this situation must have a remarkable impact on performance and profits.*

While quantifying the savings from such a system is difficult, some studies indicate knowledge workers spend at least 50% of their time looking for information, leaving them only 50% to actually use what they have found. This is not only a frustration, it also represents a tremendous expense for the organization. No one has quantified the loss to an organization through poor decisions based on incomplete or wrong information. Any improvement to this situation must have a remarkable impact on performance and profits.

By installing a high-quality information system, and by outsourcing its development and maintenance, the enterprise will realize:

- Lower collaboration and transaction costs because of standard Internet-based protocols
- Access to content and software for the whole enterprise, both for frequent and infrequent users
- An instant virtual presence on the Web for the enterprise
- Quick implementation
- Reliable support for complex systems; no pain in updating software
- Reliable, tested products
- High levels of security
- Savings in staff time and in costs
- Better information for decision making

### **Challenges for Northern Light**

The elegance of the Northern Light system is concealed from the ordinary user. The major challenge for Northern Light, then, is to explain its capabilities to potential customers. How do you explain the effectiveness of the system's outstanding relevance-ranking capabilities without demonstrating its performance in a head-to-head competition with its competitors? Without comparing Northern Light's relevance ranking on the same set of documents, evaluators would have difficulty determining the value of its ranking algorithms. While the custom folders make Northern Light's classification capabilities somewhat visible, they only reveal the tip of the iceberg: The classification system is one of Northern Light's distinguishing features; it improves retrieval immensely.

IT staff, who are often involved in software and services decisions, are typically well grounded in structured database design concepts. It is difficult to explain the difference between information retrieval technology for unstructured text and relational database searching. In fact, vendors such as Oracle claim to do both. Yet, a good text retrieval system that can also search across the text in internal databases is crucial to an enterprise information system.

Software costs typically quoted by competing vendors appear comparable. A more difficult argument will be to convince security-conscious enterprises to trust their internal documents to a hosted service. However, the security of internal systems has been shown to be vulnerable to outside penetration. In fact, a VPN may be more secure.

The challenge is to make Northern Light a household name. While it is known as a Web search engine, its role as an enterprise information system is not widely recognized. It will be competing against more recognized names such as Autonomy, Verity, Plumtree, and Viador.

*Outstanding search capability and relevance ranking, together with careful classification and a host of customizable features, make Northern Light a worthy competitor in the EIP market.*

## **Conclusion**

Northern Light is an entire integrated information system: content and software are rolled into one package. Outstanding search capability and relevance ranking, together with careful classification and a host of customizable features, make Northern Light a worthy competitor in the EIP market. The company can provide the robust infrastructure for an extensive enterprise portal by adding CRM, data mining, or ERP functionality to its system. As a Web search engine, Northern Light has demonstrated its ability to handle extremely large databases with high numbers of simultaneous queries reliably. The outsourced model makes this a persuasive offering that may disrupt the current market in Northern Light's favor.

## NORTH AMERICA

### Corporate Headquarters

5 Speen Street  
Framingham, MA 01701  
508-872-8200

### IDC Canada

36 Toronto Street, Suite 950  
Toronto, Ontario, Canada M5C2C5  
416-369-0033

### IDC Irvine

18831 Von Karmen Ave, Ste 200  
Irvine, CA 92612  
949-250-1960

### IDC Mtn. View

2131 Landings Drive  
Mountain View, CA 94043  
650-691-0500

### IDC New Jersey

120 Wood Ave South, Suite 509  
Iselin, NJ 08830  
732-632-9222

### IDC New York

2 Park Avenue  
Suite 1505  
New York, NY 10016  
212-726-0900

### IDC Texas

100 Congress Ave, Suite 2000  
Austin, TX 78701  
512-469-6333

### IDC Washington

8304 Professional Hill Drive  
Fairfax, VA 22031  
703-280-5161

## ASIA/PACIFIC

### IDC Asia/Pacific (Hong Kong)

12/Floor, St. John's Building  
33 Garden Road  
Central, Hong Kong  
852-2530-3831

### IDC Asia/Pacific (Singapore)

71 Bencoolen Street, #02-01  
Singapore 189643  
65-226-0330

### IDC Australia

Level 4, 76 Berry Street  
North Sydney, NSW 2060, Australia  
61-2-9922-5300

### IDC China

Room 611, Beijing Times Square,  
88 West Chang'an Avenue, Beijing,  
P.R. China, 100031  
86-10-8391-3456

### IDC (India) Limited

Cyber House  
B-35, Sector 32 - Institutional  
Gurgaon - 122002, Haryana, India  
91-124-6381673 to 80

### IDC Japan

10F The Itoyama Tower  
3-7-18, Mita Minato-ku  
Tokyo 108-0073, Japan  
81-3-5440-3400

### IDC Korea Ltd

Suite 704, Korea Trade Center  
159-1, Samsung-Dong, Kangnam-Ku  
Seoul, Korea 135-729  
82-2-55-14380

### IDC Malaysia

Suite 13-03, Level 13, Wisma KiaPeng  
No. 3, Jalan Kia Peng  
50450 Kuala Lumpur, Malaysia  
6-03-2163 3715

### IDC New Zealand

Level 7, 246 Queen Street  
Auckland, New Zealand  
64-9-309-8252

### IDC Philippines

7F, SEDCCO 1Bldg  
Rada Street Corner  
Legaspi Street  
Legaspi Village  
Makati City, Philippines  
632-894-4808

### IDC Taiwan Ltd.

10F, 31  
Jen-Ai Rd, Sec 4,  
Taipei 106, Taiwan, R.O.C.  
886-2-2731-7288

### IDC Thailand

27 Soi Charoen Nakorn 14  
Charoen Nakorn Road, Klongtongnai  
Klongsan Bangkok 10600, Thailand  
66-2-439-4591-2

### IDC Vietnam

37 Ton Duc Thang Street  
Unit 1606  
District-1 Hochiminh City Vietnam  
84-8-910-1235

## EUROPE, MIDDLE EAST, AND AFRICA

### IDC Austria

c/o Loisel, Spiel, Zach Consulting  
Mayerhofgasse 6  
A-1040 Vienna, Austria  
43-1-50-50-900

### IDC Benelux (Belgium)

29 Avenue Louis Gribaumont  
B-1150 Brussels, Belgium  
32-2-779-46-04

### IDC Benelux (The Netherlands)

A. Fokkerweg 1  
1059 CM Amsterdam  
The Netherlands  
31-20-669-2721

### IDC Central Europe (ECE)

Male Namesti 13  
Praha 1 110 00, Czech Republic  
420-2-2142-3140

### IDC Central Europe (Germany)

Nibelungenplatz 3, 11th Floor  
60318 Frankfurt, Germany  
49-69-90502-0

### IDC Central Europe (Switzerland)

Niederlassung Züerich  
WTC, Leutschenbachstrasse 95  
CH - 8050 Züerich  
Switzerland  
41-1-307-1000

### IDC Egypt

39 Iraq Street  
Mohandesseen, Cairo, Egypt  
20-2-336-7355

### IDC France

Immeuble La Fayette  
2, Place des Vosges, Cedex 65  
92051 Paris la Defense 5, France  
33-14-904-8000

### IDC Hungary

Bajcsy-Zsilinszky út. 57  
Building 3, Rooms 103-104  
H-1065 Budapest, Hungary  
36-1-153-0555/ext. 165, 166

### IDC Israel

4 Gershon Street  
Tel Aviv 67017, Israel

### IDC Italy

Viale Monza, 14  
20127 Milano, Italy  
390-2-284-571

### IDC Nigeria

House 2, 'C' Close  
403 Road, 4th Avenue  
New Extension, Festac Town  
Lagos, Nigeria  
234-1-883585

### IDC Nordic (Denmark)

Jagtvej 169B  
DK-2100 Copenhagen, Denmark  
45-39-162222

### IDC Nordic (Finland)

Jarrumiehenkatu 2  
FIN-00520  
Helsinki, Finland  
358-9-8770-466

### IDC Nordic (Sweden)

Box 1096 Kistagången 21  
S-164 25 Kista, Sweden  
46-8-751-0415

### IDC Poland/ProMarket

Wrobla 43  
02-736 Warsaw, Poland  
48-22-644-4105

### IDC Portugal

Av. Antonio Serpa, 36 Piso 9  
1050-027 Lisbon  
Portugal  
351-21-796-5487

### IDC Russia

c/o PX Post, RDS 186  
Ulitsa Zorge 10  
Moscow 125525  
Russian Federation  
7-501-929-9959

### IDC South Africa

c/o BMI-TechKnowledge  
3rd Floor, 356 Rivonia Blvd.  
PO Box 4603, Rivonia, 2128  
South Africa  
27-11-803-6412

### IDC Spain

Ochandiano, 6  
Centro Empresarial El Plantio  
28023 Madrid  
34-91-7080007

### IDC Turkey

Tevfik Erdonmez Sok. 2/1 Gul Apt.  
Kat 9D; 46 Esentepe  
Istanbul, Turkey  
90-212-275-0995

### IDC U.K.

British Standards House  
389 Chiswick High Road  
London W4 4AE  
United Kingdom  
44-20-8987-7100

## LATIN AMERICA

### IDC Miami

Latin America Headquarters  
5301 Blue Lagoon Drive  
Suite 490  
Miami, FL 33126  
305-267-2616

### IDC Argentina

Trends Consulting  
Rivadavia 413, 4th Floor, Suite 6  
C1002AAC, Buenos Aires, Argentina  
54-11-4343-8899

### IDC Brasil

Alameda Ribeirão Preto, 130 cj 41  
01331-000 São Paulo  
SP Brazil  
55-11-253-7869

### International Data Corp. Chile

Luis Thayer Ojeda 166 Piso 12  
Providencia, Santiago 9, Chile  
56-2-231-0111

### IDC Colombia

Carrera 40 # 103-78  
Bogota, Colombia  
571-533-2326

### IDC Mexico

Select - IDC  
Av. Nuevo Leon No. 54 Desp. 501  
Col. Hipodromo, Condesa  
C.P. 06100 Mexico, D.F.  
52-5-256-1426

### IDC Venezuela

Calle Guaicapuro  
Edif. Torre Seguros Alianza  
Piso 6, Ofc. 6-D, El Rosal  
Caracas 1060, Venezuela  
58-2-951-3270

IDC delivers accurate, relevant, and high-impact data and insight on information technology to help organizations make sound business and technology decisions. IDC forecasts worldwide IT markets and adoption and technology trends, and analyzes IT products and vendors, using a combination of rigorous primary research and in-depth competitive analysis. IDC is committed to providing global research with local content through more than 500 analysts in more than 40 countries worldwide. IDC's customers comprise the world's leading IT suppliers, IT organizations, and the financial community. Additional information on IDC can be found on its Web site at <http://www.idc.com>.

IDC is a division of IDG, the world's leading IT media, research, and exposition company.