Entopia (Offline 2006)

Entopia's approach depended on a backbone metaphor. Search was a core feature and specialized functions were plugged in. Entopia was one of the first vendors to offer search-based applications.

Author's note: This is an unpublished, preliminary draft of a description originally destined for a client report. The information is provided as part of ArnoldIT's archiving project. The information in this draft may not be used without prior written permission. The information in this document was written before Convera went out of business with the sale of its remaining assets to Vertical Search Works.

Entopia was one of the first vendors to combine search, text processing, user monitoring, and work flow in one massive system. The company's approach was to bring an IBM or SAP R/3 style re-engineering to information. The diagrams of the Entopia architecture suggest the complexity of the company's system. The merging of many hard-to-define features created problems for Entopia's marketing and sales team. I listened to a number of presentations given by Entopia and came away with the sense that closing a deal and getting annual license renewals would be difficult. The company shut down in 2006. The expansive vision of Entopia is a good example of seeing the future and not being able to deliver value for a client.

This information is a rough draft and is frozen.

Introduction

Founded in 1999, Entopia (formerly KangarooNet) is one of the new breeds of enterprise search system vendors that describe themselves as "third-generation technology." In this context, third generation means that search includes indexing information about document authors and users as well as document categorization, not just simple key word indexing. Entopia's marketers try not to pigeonhole the company as a vendor of basic search technology. Entopia's vice president of product management, David Hickman, says, "The next inflection point in information discovery comes from using activities to provide personalization schemes." No search needed. because discovery, social nuances, business functions, work activities, and metrics do the heavy lifting. Entopia's technology, not the user, does the hard work to make search easy.

Table 1: Entopia K-Bus: A Bird's Eye View

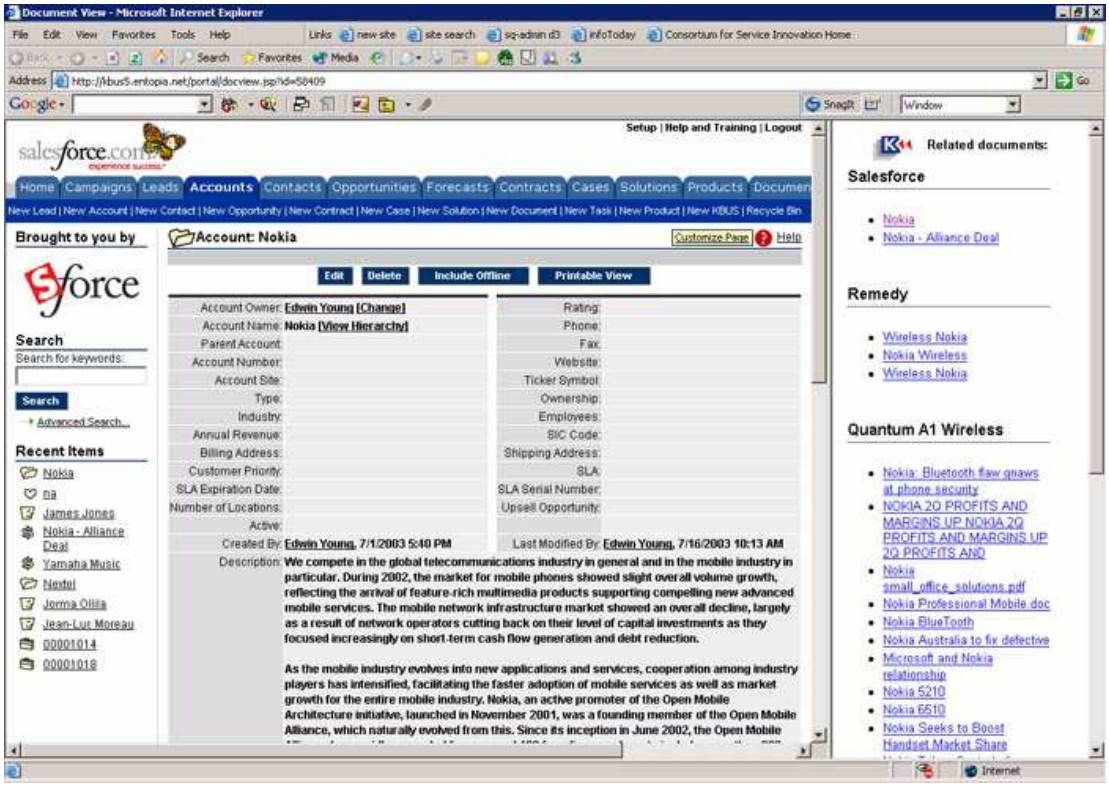
Product Thumbnail	
1 Search Brand	K-Bus and Quantum
2 OS Supported	Unix, Solaris
3 Est License Fee	Pricing begins at $250,000 per year
4 Functions	Search, work flow, expert identification, collaboration, relationship mapping, visualization, support for Salesforce.com's customer relationship management system
5 Claimed Features	Contextual indexing, semantic indexing, automatic document classification, rich metadata generation, text mining, analytics, generation of "knowledge maps"
6 Downsides	The network and system resources required are significant. The cost and time for installation and tuning is comparable to an SAP R/3 system.
7 Similar To	Autonomy IDOL, Convera RetrievalWare, OpenText LiveLink, Fast Search & Transfer ESP
Product Close Up 	Entopia's management harvested technical and management trends from business schools and competitors. The company bundled a wide range of features, functions, and services as an enterprise solution. Not surprisingly, the time required to close a deal and the cost of configuring a extremely complex system frustrated investors and licensees. The company requires additional investment capital to remain a viable business as lower-cost and less complex options flow into the market.

Entopia argues it delivers "truly-relevant search, expertise identification, content visualization, social networks analysis and content connectivity functions can be added to any business system, portal framework, or software infrastructure to create robust knowledge-enabled applications."

Unlike companies nurtured with a U.S. government contract, Entopia was conceived and delivered by entrepreneurs from Omind, an Israel-based tech-

nology incubator and venture firm. The language used to describe Entopia's search and retrieval system makes figuring out what the technology does and how certain functions are performed an interesting and somewhat challenging exercise.

Entopia's approach to search addresses the criticism that many clients of well-known search vendors level; namely, search is isolated from other business processes in the organization. Entopia's solution is to create an enterprise-wide information system. The K-bus product provides a pipeline for information into which virtually every authorized user taps. The company's Quantum product provides an information manufacturing system. When both the content management system (Quantum) and the knowledge finder (K-Bus search system) are in place, information management is integrated.



Entopia's search system seamlessly integrates with the Salesforce.com Web service. A user can manage his sales contacts and search for information and experts germane to a proposal or a prospect's requirements from one screen.

Entopia provides a total search solution much in the way that SAP provides a solution to enterprise resource planning or SAS Institute addresses data mining. These companies require that its customers buy into the overall framework of their systems. Because these systems subsume information, a homogeneous information environment is created. With standardization, the

enterprise application can integrate, slice and dice, and perform the type of bird's-eye view analyses senior managers want from their information systems.

Entopia is a company name derived from the Greek word *entopizo*. For those of you whose Greek is a rusty, the verb means to locate or bring to light. Entopia's senior technologists stress that their K-Bus and Quantum systems allow a licensee to locate and make use of information that would otherwise be lost to decision makers. The name is significantly more tech sounding than KangarooNet.

The founders of Entopia were individuals with a nose for technology trends. Entopia is privately-held, with the principal stake in the hands of Omind. In 2001, Entopia garnered an $19 million in funding from Vivendi Universal and a handful of other sources. Global Catalyst has provided funding since the company's inception. The roster of investors includes The Invus Group, QuestMark Partners, Vertex Management, Crystal Internet Ventures, and Walden International. It is difficult to get a fix on ownership. What's clear is that Entopia is likely to be nudged toward an initial public offering or possibly a sale. Omind's principles, Kamran Elahian of Global Catalyst Partners and Lionel Baraban (associated with Omind), are among the key stakeholders.

Entopia's solutions start by addressing the need of employees to gather information more effectively, particularly now that the amount and sources of raw data are overwhelming. Entopia talks about its solution as a tool, de-emphasizing the modules that have been woven together to make the system work.

Search Becomes Knowledge Management

By de-emphasizing the bits and pieces, Entopia simplifies its selling proposition. In a nutshell, Entopia sells search and retrieval packaged as a tool to collect, organize, and share information in a personal and easy way. The payoff from search is increasing the personal productivity of Entopia users. Beyond productivity, Entopia stresses that information processed and accessible in its system puts the information in context. Entopia delivers intellectual capital to its licensees.

When Entopia is compared to another search solutions such as Autonomy-Verity's, Endeca's or FAST Search's, Entopia drives home the point that its system does not disrupt normal work processes. Entopia, the company argues, is a lean, mean information machine. Its competitors' solutions are unwieldy. Online Magazine described Entopia in this way:

> Entopia stands out from the knowledge management pack because they recognize that much of an organization's intelligence exists both at the desktop level and in a variety of repositories across the

> enterprise. They [sic] tap this collective intelligence to help leverage it across workgroups or across the enterprise.

The marketing message is working. Entopia has been capturing customers and growing in the last year. Licensees may wish to keep in mind that Entopia's owners will, at some point, leverage the keen interest in search via a sale or IPO.

The company has its principal offices in Redwood Shores, Calif. and maintains a research and development facility in Israel.

A developed application is to integrate a set of Social Networks Analysis applications, using the traditional input of any search engine, and extract a picture of interconnectedness between entities or "knowledge workers". Following this approach we are able to provide a clear insight on information flow and knowledge transfer within an organism, that is not hindered by any recollection or framing biases. Furthermore we also use these relations to render a dynamic expertise profiling solution.--Dr. Igor Perisic, Entopia (Source: http://goo.gl/DDGtGV)

Explaining K-Bus: A Challenge

As one might imagine, convincing an enterprise to make a comprehensive commitment to an information management system is a tough sales job. Entopia, according to **Igor Perisic**, Entopia's chairman of the company's advisory panel, "is doing very well. This year [2005] has been very strong for us."

Entopia stressed that it had a strong customer base. One headline firm deploying K-Bus is the Saab Group, a Swedish aerospace and defense company. Among the organizations the company said were using, K-Bus was a fruit salad of organizations; for example:

- Cukierman & Co. Investment House, an investment firm providing some financial services to Entopia
- DGA, U.K. software consultancy
- Evesham Technology, a British supplier of personal computers
- Gate5 AG, a German geographic information system company
- George Washington University
- INPI (Institut national de la propriété industrielle), provider of patent information.
- Jackson State University, U.S., four year university
- Ministère de Défense Nationale, French military entity
- QCSI, a health care company in Phoenix, Arizona
- Space and Naval Warfare Systems Command, U.S. military entity.

Entopia was an aggressive marketer of search wrapped in buzzwords.

The MBA and venture capital heritage is evident in Entopia's positioning, its marketing, and its technical descriptions of its complex system. The company laces its white papers, Web site, and technical information with liberal helpings of trendy phrases, spoonfuls of knowledge management jargon, and neologisms like: "Decision Orchestration. Search has morphed into information discovery or knowledge extraction. Indexing becomes semantic analysis. Only aficionados of the enterprise search game will recall that the

company entered the U.S. market as KangarooNet. The K-Bus was the SmartPouch."[1] The idea is that an organization and a user would have information in one, easily-accessed service.

By 2001, the Kangaroo metaphor had hopped into history and Entopia focused on enterprise search. Like Endeca, Entopia positioned itself by not emphasizing search and retrieval. Entopia's approach was to offer customers two bundles of technology.

The flagship software was the bundle of systems marketed as Entopia K-Bus or what Entopia calls third generation information discovery. Think of K-Bus as a suite of technology to acquire, normalize, index, and make searchable the information in an organization.

The second service was Entopia Quantum, described by the company as a collaborative work space. Quantum allows system users to assemble, discuss, and manipulate the information processed by the K-Bus. This is the original SmartPouch technology that allows a user to gather information and keep it in a virtual work space.

Entopia seeks to compete in enterprise search with Lotus Notes-style functionality. Entopia Quantum is a workgroup solution that enables individuals to collect information from any digital source, allows teams to collaborate in project or shared spaces and empowers users to capitalize on this accumulated knowledge—from one convenient location—by locating documents, experts, and information sources. Unlike other collaborative content management solutions, Entopia Quantum's bottom-up approach creates intellectual capital—both explicit and tacit—as a natural by-product of personal and workgroup activity. This analysis will not discuss the Quantum product.

Scrape the surface of the buzzwords away, and Entopia has assembled the basics of search and retrieval technology and categorization. These are bundled with a wide range of built-in utilities to allow users to locate individuals with an expertise or an interest in a particular topics, point-and-click reports to display relationships among information, and other functions that convert a plain vanilla results list into views and reports.

Search and Retrieval

The search function supports a keyword or phrase search. A semantic search allows a user to look at documents related to a heading in the taxonomy displayed to that user. Entopia keyword searches allow greater precision. Semantic searches allow greater recall. Entopia calls a semantic search a fuzzy search.

1. It is amusing to recall that Stratify (once part of Iron Mountain and then folded into Autonomy) was called "Purple Yogi."

Entopia search packs three functions into its system:

- Full text indexing, including words and phrases
- Contextual indexing; that is, the business unit for the document
- Semantic indexing; that is, categorization of the document.

This is a multiple pass document process. The goal of this “rich metadata” indexing process is to allow the licensee to expose indexed content by business activity, key authors or experts, relationships among objects as well as traditional keyword search techniques.

Metadata

Beneath the verbal fireworks of Entopia’s sales presentations, Entopia indexes the words and phrases in a document and generates metadata about the information. Not content to classify an information object into a categorization system, Entopia’s system notes who created a document, people mentioned in the document, and those in the organization looking for experts on a topic.

The metadata generated for each document or file processed by the Entopia system consists of one or more of these items:

- Keywords and bound phrases
- Concepts
- Security flag associated with the document; that is, role level access rights
- Document properties; that is, size, date stamp, format
- Embedded metatags; that is, any tagged element appearing in an SGML, XML, or XHTML document
- Context; that is, department of the document creator, functional unit of the business, etc.
- Activity; that is, the business activity to which the document relates; for instance, accounting, legal, marketing
- Life cycle; that is, version information about the document, destroy by date, etc.
- Custom metadata created by Entopia engineers for a client; for example, function to associate a project number with a department.

One distinguishing feature of this metadata is that Entopia is one of a small number of enterprise software applications that include metadata about the document owner and the people accessing the document.

Social software systems have typically operated in isolation from search and retrieval systems. Entopia’s tools allow the licensee to link a database con-

taining employee contact details, department, and pictures to a document. Entopia, like Tacit Software, Inc., recognizes the importance of individuals and their expertise and interest.Entopia, unlike Tacit, has built its social software functions on top of search and retrieval. Tacit has integrated search into its social and user tracking system.

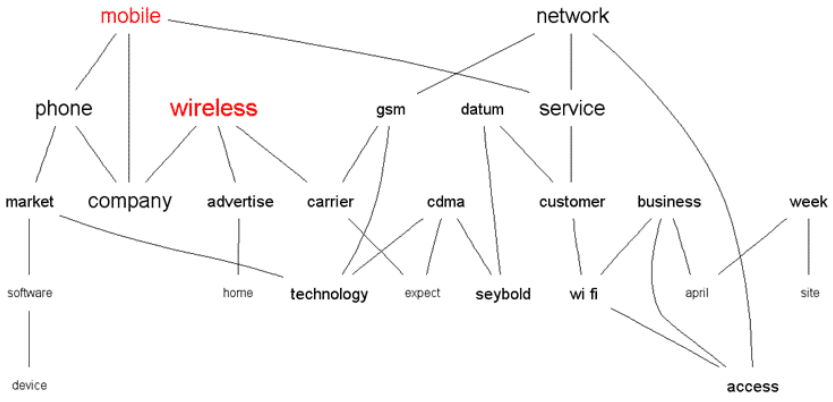
In terms of enterprise search, Entopia has been successful in putting document indexing into a broader operational context. It is too soon to determine if Entopia can raise the bar in enterprise search so that its competitors must add social software functionality to the now-standard line up of classification functions.

Advanced Functions

Entopia says that its approach allows search results to have less noise than competing systems. In effect, Entopia's approach to indexing delivers higher precision and recall than its competitors' systems.

Entopia's indexing technology can automatically create metadata from the content in connected enterprise repositories, applications, intranets and Web sites. The content repositories can be located behind a firewall, on a wide area network, or on Web sites.

The Entopia indexing identifies the semantic value—that is, concepts—in a document. Entopia can discover these concepts and make use of licensee-supplied taxonomies if available. In addition, Entopia identifies and coverts to metadata the social activity of system users. Entopia captures and converts to metadata reads, writes, edits, comments, and accesses to documents. In addition, Entopia also extracts from available repositories personal, organizational, and application-related contextual information around documents or content.



A knowledge map provides users with a way to view relationships among topics in a keyword query or in a list of related content.

We capture the context and user activity around documents. We know who wrote it, how many times it's been read, and whether anyone took the time to save, print or e-mail it.--David Hickman, vice president, Entopia (Source: http://goo.gl/6s5xRl)

The next-generation features of the Entopia search system include:

- Semantic and full-text search for documents, people and sources
- Access by concepts and categories or key word queries
- Results display in ranked lists in the form of documents, people with related expertise and sources of information on a single screen
- Quasi-parametric search by metatags; for example, limit results by time, specific user, a department, document author's job title, file type, language or other factors implemented with customization
- Content visualization or a "knowledge map" of the relationships between concepts within a specified set of results. According to Entopia, the knowledge map is "useful for quick navigation through a large amount of information and for uncovering hidden knowledge within a company"
- Automatic document summarization using concept terms as well as words and phrases found in the source document.
- A more like this function to allow one click access to related information
- Support for a search launched from a Microsoft Word or other document, a Web page, or another enterprise application. Running a query from an enterprise application allows Entopia to be used to replicate Endeca-like functionality

Entopia's Architecture

Entopia processes content, creates its indexes, and then points to the original documents on the server where they reside. The approach is similar to that taken by Verity in the 1980s.

K-Bus requires business process engineering in addition to the standard identification of content to be processed. Entopia's system design is notable because it combines the complexity of an SAP solution with a wide range of search applications like identification of experts in a licensee's organization.

K-Bus borrows a concept from electrical engineering. A bus is an electrical conductor that serves as a common connection for two or more electrical circuits. Entopia has applied this notion of a common connector to plug information into a search connector. Instead of fiddling with incompatible bits and pieces, the Entopia bus creates a common connector for the myriad processes that are needed for an enterprise search system.

The customer buys into the notion that Entopia delivers an integrated system. One system provides search, content management, classification, and analytic functions plus "knowledge management," "collaborative func-

tions," and other information-centric services. The Entopia system is an environment that requires a commitment to the firm's engineering philosophy. Just as customers of SAP's R/3 enterprise software buys into the SAP approach to process control, a licensee of Entopia makes a commitment to the Entopia approach.

Not surprisingly, the complexity of the Entopia system is evident in the firm's "basic block diagram."

Extending K-Bus

The architectural core of Entopia is a software infrastructure that is built upon J2EE technologies, including JSP (presentation layer), EJB (business logic layer), JAAS (pluggable authentication modules), JDBC/JCA (data access layer), and utilizes SOAP as the inherent communication protocol. This open standards-based architecture provides OEM partners and some licensees with the flexibility needed for easy integration and customization of the Entopia functions.

In addition, the Entopia architecture is multi-tiered, including a database layer, an application (server) layer and a semantic processing pipeline layer. Each of these tiers can be deployed on separate/multiple physical machines and can therefore be scaled independently to meet the specific needs of OEM Partners and their customers.

Entopia has invested considerable effort to make its search system mesh with enterprise software requirements. For example, the system ships with adaptors to permit the K-Bus and Quantum products acquire information from Open Text LiveLink repositories, Documentum, IMAP4 email servers, mainstream relational database systems such as Oracle and DB2, IBM Domino servers and Lotus Notes' repositories, managed systems services such as Salesforce.com, BMC Remedy service management systems, and more than 200 file types.

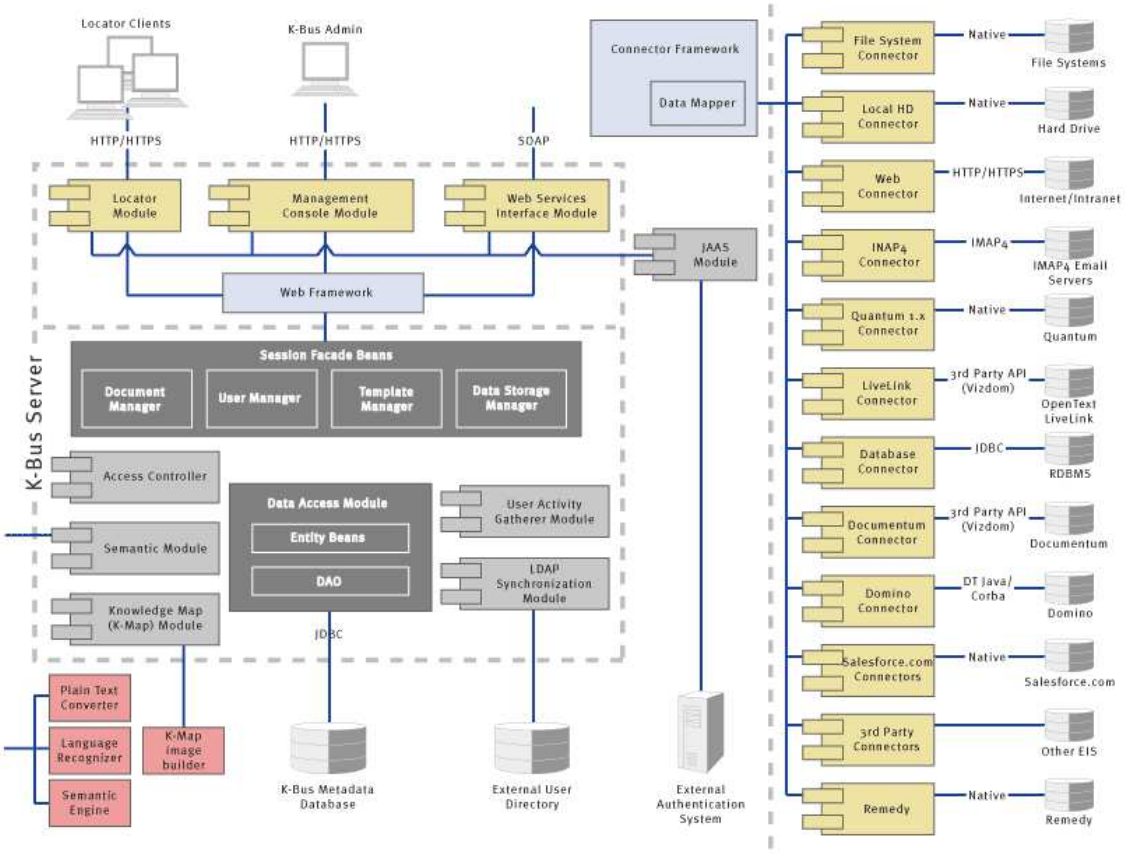
The Entopia products ship with pre-built JSR-168 portlets. Customization of Entopia templates and reports is handled via CSS style sheets.

A single SDK contains APIs for

- Enterprise Search
- Expertise Identification
- Social Networks Analysis
- Enterprise Content Connectivity.

The Entopia Software Development Kit (SDK) allows application developers to access Entopia functionality from other enterprise applications or extend the functionality of a specific Entopia function. Entopia Web APIs adhere to the Web service recommendations of the W3C and do not rely on any proprietary extensions. The openness of the APIs allows for tight inte-

gration at a granular level, while the flexibility ensures support for applications written in other development languages, including Java.



The complexity of the Entopia system is fascinating. An organization adopting K-Bus must make a significant commitment to the platform and its associated costs.

The Entopia SDK consists of Web service application programming interfaces (APIs) that have been modularized into functional capabilities. This means that developers can add only the K-Bus functions needed for a specific application or business need. Existing functional API modules include:

Two of these—search and connectivity—are discussed below.

Enterprise Search SDK

The Entopia SDK for Enterprise Search includes a robust set of information discovery functions to meet a wide variety of business-specific requirements. APIs are available so that the licensee can:

- Hook search into other enterprise applications such as enterprise resource planning or content management systems

- Implement “more like this” search from within other applications or documents
- Provide point-and-click access to reports that show relationships among people, documents, and expertise as well as licensee-defined entities such as patent numbers or purchase order codes
- Set filters to limit displays by one or more metatags such as department, business function, or business process.
- Integrate an automatic document summary when information is requested. Entopia generates summaries on the fly from the key words and metadata associated with the document during indexing.

Content Connectivity SDK

With Entopia SDK for Content Connectors, developers can rapidly create new native connectors to structured and unstructured repositories to enable the crawling and indexing of content by the Entopia K-Server metadata repository. The functions of this API provide the conduits to transfer the information object, history of the information object, access control lists and any other attached metadata to the K-Server so that the information can be utilized by business knowledge applications.

Entopia comes with connectors for dozens of enterprise applications, notably Microsoft SharePoint, Microsoft Office, Salesforce.com, and Lotus Notes. The content connectivity SDK provides hooks and sample code to integrate non-supported file types into Entopia.

Search Becomes Knowledge Management

“Knowledge management” is a term that is difficult to explain. Entopia has embraced it. Entopia provides a comprehensive suite of enterprise software, but in terms of search or what Entopia sometimes calls knowledge location, Entopia is trying to move well beyond keyword searching. A query will return documents as well as experts and other sources of knowledge within the enterprise. Even though a document may not contain a specific word, a document about a subject can be related to a user's query via the categorization metatags.

Entopia’s approach to search returns hits in a relevance ranked list. The user will be also able to view a list of individuals who are expert in the subjects related to the user’s query. If the user wants to see documents that are conceptually related to the query, the system can display these items as well, also in a relevance ranked list.

In short, a search on Entopia can return a list of documents, a list of people, and a list of related documents. In order to assist the user in navigating the

information, an individual user can set a wide range of personalization controls so that the display of information is tailored to that user's needs.Entopia furthermore provides point-and-click access to visualization tools that display key items with lines connecting those that are related to key people or important concepts.

The key to this type of information retrieval is traditional indexing. In order to provide the conceptual and expert information, however, Entopia performs additional indexing processes to obtain metadata to support these functions. Compared to some other search systems, Entopia generates more metadata about document than some other systems profiled in this report.

A Taxonomy of Enterprise Knowledge

Entopia uses a relaxed taxonomy. An entry can contain a category or subcategory name as well as a unique term assigned to an existing category. Entopia's approach to taxonomy allows a standard taxonomy to be modified over time by the licensee's users. The taxonomy allows individualized views of content; thereby accommodating specific jargon within a department or an organization. Entopia permits the information to be sliced and diced by key word, category, or one or more of the metatags such as by date, file type, etc.

Entopia allows user to take an active role in categorizing information. However, if users do not input categories, the Entopia system will assign the document to a category and a zone based on information generated by rules and other actions taken by users of the system in a particular organization.

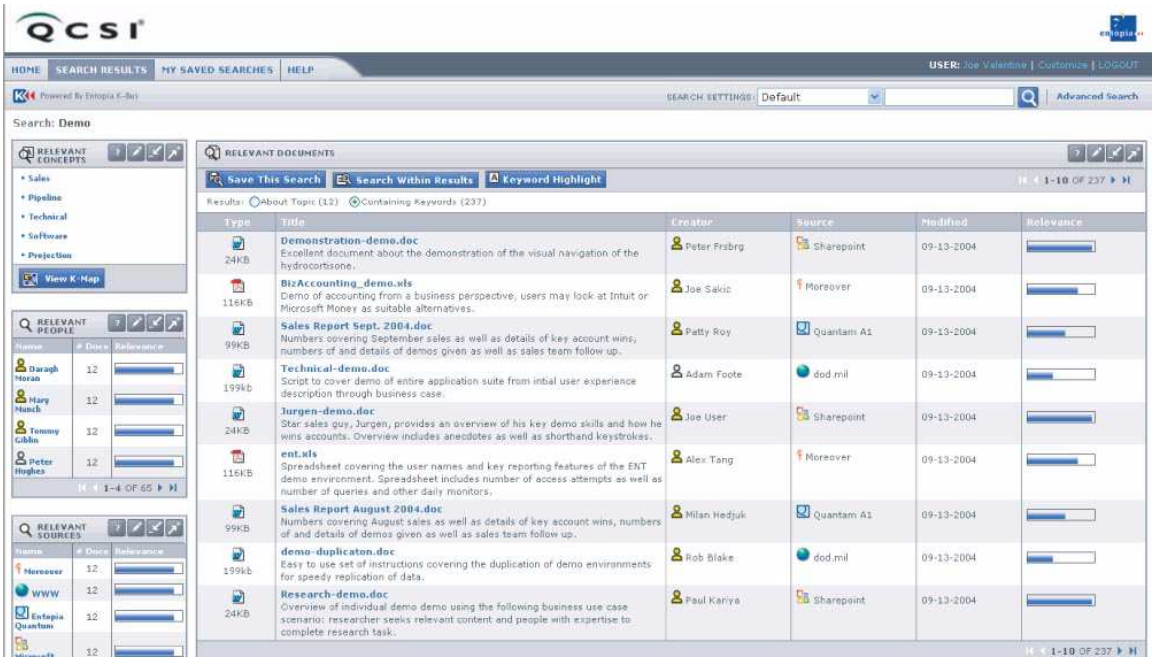
Entopia's taxonomy contains terms for business, financial services, consulting, pharmaceuticals, biotechnology, government, and education. Even though the taxonomy is large, the documents processed by the system populate only the needed portions of the taxonomy.

Indexing Dictionaries and Relevance

Entopia is a comparatively complex system, because the company's technology returns hits in response to a query that typically includes:

- A relevance ranked list of documents matching a query.
- The people in the company that the Entopia system has identified as an expert in the topic of query.
- Documents identified as germane to the user's query, which may not have the words or phrases in the user's query in them.

- Visual representations of the relationships among documents, experts, and related topic areas.



The Entopia search result screen shows the document type, document summary, creator, source, date last modified, and relevance in the main display. The left hand column displays: [a] relevant concepts, relevant experts, and relevant sources.

To deliver this type of rich results, Entopia uses many technologies and tools. For the purposes of this report, however, three are pivotal to the knowledge location function in Entopia search:

A taxonomy that consists of more than 1.5 million categories and terms covering business, finance, pharmaceuticals, biotechnology, and government. This taxonomy can be modified and extended automatically as Entopia processes user requests or directly by individual system users. When humans must interact with a large dictionary, the time and cost of index maintenance can be significant. Ignoring manual tuning of the dictionary means that relevance of queries is eroded due to language drift.

Black box algorithms sometimes described as neural networks by Entopia that look at the key words and outputs from Entopia's statistical routines. These black box algorithms identify the categories into which a document will be placed if the user accepts.

A user-interaction mechanism that allows an individual user to reassign a document to a category or create a new category. For example, in addition to assigning a document to a category, the user can assign the document to a

use zone: personal access, group access, and enterprise access. Unmoderated index updates can have an impact on search results. Users often assign broad terms instead of more precise terms. Users who tag documents with jargon can cause problems for users who are not familiar with the overly precise language.

Entopia's indexing ingredients deliver classification of documents via the built-in taxonomy, linguistic, and statistical routines that assign a document to a category in the built in taxonomy, and a mechanism to allow each user to accept or modify the system's categorization suggestion. As users interact with an Entopia system, the system "learns" and modifies the taxonomy accordingly.

The approach allows Entopia's engineers to say, "Taxonomy is a means to an end." The "end" is to take a standard query and return documents, suggestions, and people in the organization who are experts in the topic of the query.

Multi-Pass Indexing

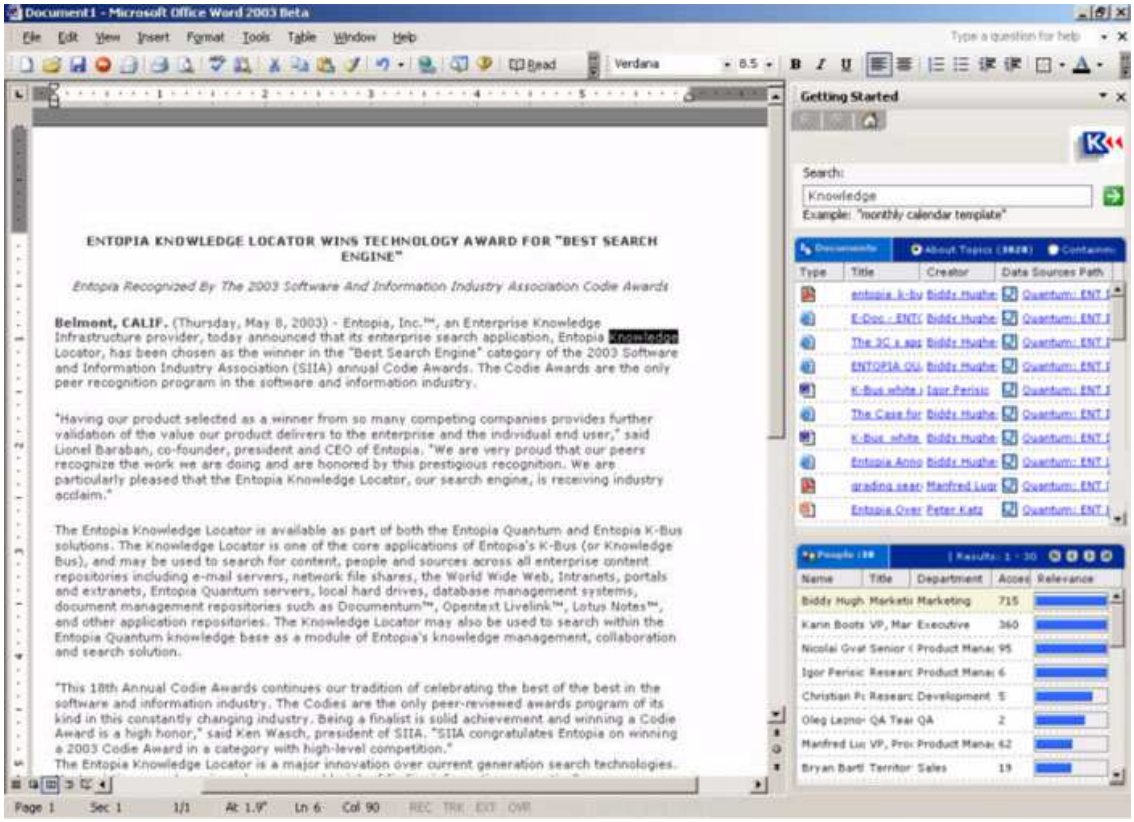
In most systems, the indexing process occurs as part of the document processing function. In an Entopia system, when a document enters the system, the document processing system performs key word extraction and assigns the document to one or more categories in the taxonomy. In parallel with these functions, Entopia extracts certain metadata about the document and writes that metadata to a file. Some of the tags are related to the file itself such as file type, date, and time. Launching a separate process that identifies the author of the document, the department in which the author of the document works, and the various category tags to which the document has been automatically assigned creates other tags.

However, when a user creates a document or interacts with information retrieved in response to a query, the Entopia system captures additional information about the document. Some of the information is usage data that updates the file capturing the number of times a document has been accessed. Other processes accepts user input regarding categories to which the document is assigned. Entopia captures these user inputs and behaviors, using the data to adjust.

As a result, Entopia implements a multi-phase indexing process. This approach stands in sharp contrast to systems that index a document one time. User behavior and metatag information about various aspects of the document are ignored. Entopia, on the other hand, pays attention to these items of information, using them as training data for the system. It is for this reason that Entopia describes its indexing as capturing the "business context of a document." When a user restricts a document in some way to a department, the Entopia system uses this information as part of its access control service to limit viewing of the document to an employee of that particular depart-

ment. Entopia thus provides access control tagging that exists within the organization's broader access control system.

Some indexing takes place during document processing and other indexing occurs when users interact with a document. The Entopia system does not discard any of the metatag data.



The default display of search results within Word 2003 shows the highlighted search term knowledge the documents matching the query and the experts Entopia identified.

These key processes are hidden in the unusual word choice and phraseology of Entopia's marketing professionals. Indexing is described as a semantic process. The automatically generated metatags for classifying a document are described as invisible. The net result of automatic processes and user input is part of the knowledge process. When the Entopia system adjusts to user input, the function is called Smart Classification.

Most users do not provide metadata when prompted. Entopia's approach allows user interaction but can categorize a document if the user provides no metadata. Entopia, therefore, automatically places a document into its 1.5 million term taxonomy. When a user provides inputs, Entopia uses these cues and adds the information to its internal taxonomy tables. As users work

with an Entopia system, the Entopia system adjusts to the terminology of its users. Entopia does not discover a classification scheme from ground zero. It arrives with a ready-made taxonomy and a mechanism for modifying it on the fly. The approach is in contrast to companies that extract a taxonomy from a training set and then use specialists to finalize the taxonomy.

Semantic Processing

One of the pivotal functions of Entopia search is semantic indexing. Entopia's terminology, as noted elsewhere in this analysis, is a magical blend of hyperbole and metaphor. Semantic indexing is not exception.

For Entopia to create knowledge maps or deliver a slice of information based on a job function, Entopia has to process the document to "discover" the metadata. Discover is a problematic word. For our purpose, discover means extract key words and phrases and then identify relationships that have a high probability of applying to the document. In addition, known information such as the author's name is used as a query in order to locate more information about that author. These data are stored in the Entopia index. Instead of indexing the words in a document, Entopia indexes the context in which the document exists such as the author's business unit. The notion of discover is less like Vasco de Gama's setting sail into the unknown and more like collecting available information that is ignored by other indexing systems.

There is little argument that Entopia's scouring of an organization for information that makes a single document more findable is of considerable value. Generating "rich metadata"—to use one Entopia engineer's favorite phrase gobbles considerable computing horsepower, however.

A closer look at what Entopia does to generate its "rich index" provides two benefits. First, the basics of the Entopia categorization system comes to light and, second, the complexity of the so-called third-generation search systems provides useful information about what Entopia's competitors are likely to include in their search systems.

User Tracking

Like Tacit Software, which was acquired by Oracle, Entopia monitors users' actions. These "actions" are captured, analyzed, and used to generate reports on individual user activities.

Since Tacit's ground breaking effort, a number of enterprise search systems—most notably Google—have strengthened their analytic tools. Instead of semi-structured log files chock full of numbers, enterprise search vendors offer point-and-click reports.

Entopia provides a wide range of usage tracking options for its licensees. Because Entopia keeps track of documents at the author level, the combination of who looks at a document and data about the authors gives Entopia's reporting an added boost; for example:

- Document creators can be ranked from most viewed author to least viewed author.
- Authors of similar documents can be displayed because of Entopia's linking of authors to topic categories, business units, and business function.

Users of documents can be analyzed in terms of users with a similar interest, creating a ready-made affinity group, for the person with access to this function. Other views of the data allow mixing and matching within departments or across business units. Entopia says that its system can answer this type of question: "Who in the Seattle office is the go-to person for solid state engineering?"

One way to conceptualize what Entopia's social metatags can provide is to think of retrieving a document or documents by a particular author accessed in the last 48 hours related to the product price changes in the server for headquarters's marketing department.

Instead of crafting a Boolean query, Entopia allows its licensees to make these types of searches a matter of pointing and clicking on drop down lists, check boxes, and radio buttons. A potential licensee of the Entopia technology will want to keep in mind that most enterprise users of search do not want to type into a naked search box. The preference is to display certain information with a search box option. Entopia's wealth of metadata require that the licensee to continuously invest in dictionary maintenance.

ArnoldIT Opinion

The Entopia business context comes at a cost. Entopia requires that its licensees make a commitment to the Entopia way and provide sufficient resources to allow the computationally intensive processes to crunch the data quickly. Without adequate CPUs, storage, and RAM, Entopia's system can slow to a crawl.

The benefit of the Entopia system, according to Entopia, is that disconnected information residing on users' and departmental servers is indexed. The drawback is that if the source document is deleted from the user's personal machine, the document is not stored in a central repository and, therefore, lost. Entopia does not include a repository function.

Table 2: Entopia K-Bus Checklist

Attribute	Entopia Asserts	ArnoldIT Comment
1 Platform	Unix, Solaris, Linus	Significant hardware and bandwidth resources are required
2 Key word search	Entopia asserts that it implements traditional key word indexing as well as semantic methods	Boolean and free text search supported
3 Text mining	The system can use metadata to generate relationship maps and provide data about user behaviors	Primitive metadata tallies
4 Automated indexing	Yes. The K-Bus system makes use of dictionaries.	Controlled term lists and dictionaries require continuous updating and maintenance by trained indexing staff.
5 Personalization	The system is supposed to understand the context of each user's query based on work tasks	The buzzword is "context" as the user's actions relate to a work process. Organizations have to nail down work processes for the system to work.
6 Workflow	K-Bus includes a proprietary work flow system	
7 Interface	Browser-based. Results can be integrated into proprietary enterprise systems	Entopia integrates with Salesforce.com and provides a cloud and o-premises functionality
8 Hosted service	No. On premises installation required	
9 Administrative interface and tools	Graphic and command line interfaces provided	Programming expertise required
10 Application programming interface	Multiple APIs are provided	Documentation is sketchy. A knowledge of coding and scripting is required for system administrators
11 Professional services	Yes	
12 Security	Basic security functions are provided.	Customized security functions may be integrated via the API
13 Connectors	More than 200 file types are supported.	The system ships with third-party conversion tools
14 Support for structured data	Entopia references database content	Entopia is based on processing of unstructured content
15 Relevance ranking	The system performs key word and concept relevance ranking	Entopia asserts that it ranks using traditional and conceptual methods
16 Video	K-Bus indexes video metadata	
17 Federated search	Entopia returns results from content stored on multiple distributed servers	Deduplication operations are primitive. Licensees may speed content processing by breaking up source documents into collections

Attribute	Entopia Asserts	ArnoldIT Comment
18 Fielded search	No	
19 Content crawler	Yes	The system does not include a crawler for Internet content
20 Price	Pricing begins as $250,000	

Anticipated Benefits

Entopia provides a search and retrieval system that bundles basic search with categorization, social software, and reporting tools that allow a user to see relationships related to the indexed information. Other benefits include:

- A homogenous framework that allows a licensee to focus on manipulating the information that is acquired by the Entopia system, not assembling separate pieces into an integrated system.
- When properly resourced, Entopia can provide the licensee with a wide range of access and information manipulation options. Although the company packages its technology into two convenient products, the licensee can integrate search into business processes in numerous ways. Entopia pushes the boundary of what search and retrieval typically includes.
- For organizations with a highly-educated workforce, Entopia allows a relatively straightforward way to determine who knows what in an organization.

Possible Drawbacks

The sweep of features, functions, and capabilities of Entopia's system is remarkable. In addition to the time and cost of deploying a complex system with many moving parts, there are other considerations to weigh:

A commitment to the Entopia vision. Although Entopia says it supports standards, the core processes of Entopia are proprietary and certain types of Entopia information about indexed data will be challenging to reuse in another enterprise system. Like SAP, Entopia requires that the licensee give up some flexibility in data reuse for the homogeneity of functional delivered by the framework.

Multiple passes of content are required in order to extract the metadata required by the Entopia system. Significant document processing resources are required in order to keep an indexing backlog from growing to an unmanageable size.

Entopia's system attempts to combine advanced content processing methods to deliver knowledge management features to an an organization's employees. The complexity of the system poses a significant sales challenge to Entopia's marketers.--Stephen E Arnold, ArnoldIT, 2005

The number of metatags available for search or slicing and dicing the indexed information requires that considerable attention be applied to the user interface.

Integrating Entopia with other enterprise or third-party applications requires custom scripts or custom programming.

Other issues include:

- The licensee should not assume that Entopia's document processing and index updating occur in near real time. The computational demand may create a backlog even on robust systems when operating under peak load conditions.
- Search is a relatively minor component of the Entopia solution. Entopia is representative of a type of search and retrieval system that includes social software, text mining, and fine-grained usage analysis.

Conclusions

Entopia provides convenient packaging of sophisticated technology that extends search and retrieval in some interesting and (to some organizations) potentially useful ways. Entopia is not a search system for the small company.

The key point is that Entopia and many other third-generation enterprise search companies have shifted from simplistic key word matching as the bread and butter of search to concepts and categorization. The limitations of keyword searching and the intimidating nature of a search box are well known. Entopia is one of the companies that focuses on knowledge as opposed to finding documents.

Entopia's greatest value will be to organizations that have a high concentrations of highly educated professionals. A pharmaceutical company, for example, could use Entopia to good advantage in its research, marketing, and product development units. A similar argument can be made for Entopia's value in a consulting, financial services, or research-centric operation. In these organizations, finding a document is important, but getting information about who knows what or who is an influence center for a particular topic may be of even greater value.

Entopia, like an Autonomy-Verity or Fast Search solution, can easily top $1 million in start up costs. Entopia requires appropriate support personnel, servers, storage, and bandwidth. The Entopia system does not ship on a single CD-ROM and goes from load to deployment in less than one day.

The successful Entopia installation requires careful analysis of content types, custom set up of the system, and tuning. As a result, the licensee will have a staff of at least one or two full time equivalents dedicated to the Entopia system. In addition, Entopia professionals will also be called upon to handle certain integration and set up tasks.

Entopia's framework delivers some interesting search functionality; however, those enhancements require a strong commitment to what might be called the Entopia way. We anticipate that Entopia's competitors will monitor the client wins and financial performance of Entopia. If Entopia succeeds in blazing a trail with its packaging of search with complementary functions, other companies will emulate what Entopia provides.

Stephen E Arnold
Minor edits to a rough draft on October 3, 2013