Lextek International

Lextek International is a search vendor with a stealth profile. Onix is a collection of functions a licensee assembles to provide information retrieval for users.

2014 comment

Author's note: This is a 2006 draft. It will not be updated.

Stephen E Arnold, January 7, 2014

Lextek is one of the lower profile search system developers. The company's technology is used by tens of millions of people, but the brand "Lextek Onix" is essentially invisible. Lextek is interesting because like other early 1990 information retrieval systems, the company provides nuts-and-bolts functionality. What is different in the Lextek Onix approach is that the company does almost zero marketing. The few licensees of the Onix tool kit generate sufficient revenue to keep the company in business. The Lextek Onix system does not capture the attention of consulting firms engaged in pumping companies which pay to get coverage by "analysts." The Onix system is not a product that one downloads and begins to use like dtSearch or the pre-Lexmark version of ISYS Search Software. Onix is a collection of software components, numerous configuration files, and the digital equivalent of a box of Legos. A licensee like Adobe Corporation assembles the components to provide a search-and-retrieval system that meets the licensee's goals. Anyone who has used the search function in Acrobat or Acrobat Reader can appreciate that a basic search system delivers a bare bones search experience. Lextek licensees assume responsibility for the search experience. Couple the tool kit approach with no marketing and one can almost understand why the Gartner- and Forrester-type of consulting firm ignores Lextek. The founders and owners of Lextek, which operates from a suburban home in Provo, Utah, have turned their attention to artisanal chocolate. I compared the Lextek search system Web site with the chocolate company's Web site. The chocolate company Web site is more modern and more passionate. For historical purposes, I have profiled the Lextek system. It provides some insight into the engineering approach of engineers in the early 1990s. For the right licensee, Lextek may be the obvious choice. For most organizations, Onix may be a poor fit.

Stephen E Arnold, March 19, 2014

Introduction

Lextek's founders—physicists Art Pollard and Clark Noble—developed the Onix information retrieval system in the early 1990s. In 1993, the duo founded Lextek International and then in 2006 launched an old-fashioned chocolate company in 2006.

Table 1: Lexmark Onix: A Bird's Eye View

Product Thumbnail: Onix	
1 Search Brand	Onix Full Text Indexing and Retrieval Toolkit
2 OS Supported	Microsoft, Linux
3 Est License Fee	$3,000 (est.) Custom price quote required.
4 Functions	C++ code library for indexing and searching data, including very small and fast indexing, quick retrieval, flexible queries including proximity operators and wildcards, stemming, multiple languages and Unicode support.
5 Claimed Features	The company provides "free support." High-performance content and query processing, proprietary data compression, support for Boolean, free text, and natural language queries. Optional classification and summarization modules are available.
6 Downsides	Onix is not an out-of-the-box solution. Locating information about Lextek and Onix can be frustrating. Other companies have used the same names. The company does not provide applications comparable to Autonomy's Virage system for rich media, for example.
7 Similar To	Convera, Fast Search & Transfer
Product Close Up 	C++ code library for indexing and searching data, including very small and fast indexing, quick retrieval, flexible queries including proximity operators and wildcards, stemming, multiple languages and Unicode support. Handles data from a few hundred K to hundreds of Gigabyte, and includes free technical support to licensees. The State of Utah reports that the company has one to four employees; therefore, Lextek's size may limit the search system's appeal for some organizations.[a]

a.The State of Utah document is located at http://bit.ly/1aPLAv3

The company maintains a low profile and does comparatively little marketing when compared to loud voices like Attivio, Coveo, and LucidWorks (originally Lucid Imagination). There is steady background buzz about Onix32.dll. Onix provides Adobe Acrobat with search and retrieval. Some computer savvy professionals spot the mysterious "dll" and think the code is nefarious. Onix32.dll is more well known than Lextek or the 20 year old information retrieval system.Since Acrobat 5, Onix has been the search system for Adobe. Onix, it seems, replaced Verity by 2003 as the information retrieval system for Acrobat-produced files.

The company is decades old and appears to operate from a suburban home in Provo, Utah. Lextek International is a vendor of tools that can be used to *build* information retrieval solutions. The company is best known for providing a downloadable stopword list. Like Autonomy and Verity, Lextek

licenses its text indexing technologies for use in third-party products. One Lextek customer support professional told me:

> The company is innovative.

That may be accurate, but the company has allowed its name and the name of its brand (Onix) to erode. A query for Lextek returns hits to motorcycle parts and an Illinois-based legal news publication. The search brand Onix has suffered similar encroachment. Lextek is an example of a company surviving by licensing a text-centric information retrieval tool kit to organizations that make a concerted effort to find and experiment with a very low profile company.

Like other information retrieval systems dating from the 1990s, Messrs. Pollard and Noble, co-founders of the company, built statistical and linguistic components. A developer uses these components to provide search and retrieval in a manner that meets the needs of the developer's client. Almost 15 years after the company opened for business, a skilled developer can replicate a system that can meet the needs of customers like Adobe Software. Adobe bakes Lextek Onix into each shrink-wrapped box of Acrobat software. A Lextek licensee, therefore, has to be able to write code and integrate an Onix system. Onix is not a product that one installs and uses like dtSearch or a Google search appliance.

The best demonstration of what a developer can do with Onix is the Lextek International Web site at http://www.lextek.com. The site is located at www.lextek.com. Lextek's The Amano Chocolate Web site, owned by the founders of Lextek, does not offer a search function. See http:/www.amanochocolate.com.

Anyone with a copy of Acrobat can experiment with the Lextek search system. Searches can be on a local computer or any computer on a LAN. The Adobe Acrobat implementation is a good demonstration of the strengths and weaknesses of the Onix technology.

Selected Executives

The company appears to be of modest size.

Identified executives are:

- Art Pollard, co-founder
- Clark Goble, co-founder.

These individuals met during their university experience and continue to work together from a suburban home in Provo, Utah.

Financial Performance

Lextek is a privately-held company. With operations in a suburban home and a comparatively low profile, Lextek provides few clues about its financial position. Based on the Adobe deal and a modest list of clients, ArnoldIT estimates that Lextek's revenues are in the $3 million range although more robust revenues are possible.

Selected Clients and Partners

Lextek does not provide a complete list of clients. Examination of the companies identified as Lextek Onix users has been unproductive. Licensees of Lextek include:

- Above the Treeline (Edelweiss Online Catalog)
- Adobe
- Apple Computer
- Glaxo Wellcome
- Integration New Media (INM VizionDB)
- Motorola
- National Instruments (DIAdem)
- Prodigy.

Lextek has a select group of partners. These include:

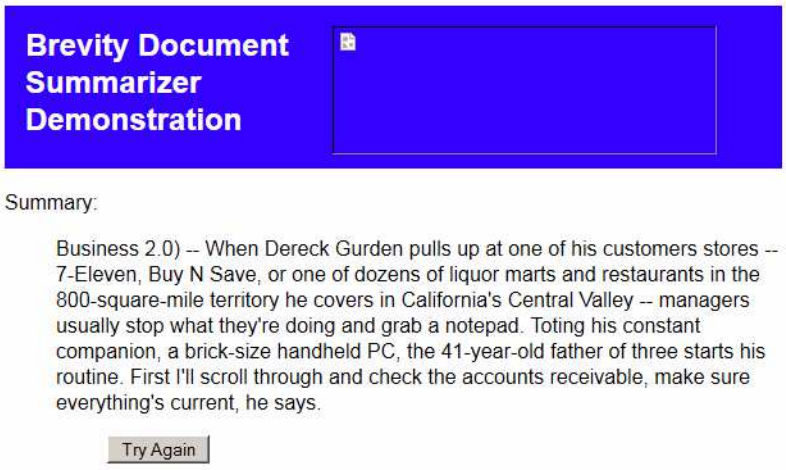
- C2C, email archiving and management solutions
- Embarcadero
- Iknow, an integrator in New Jersey.

"Searching with indexes is still present in Acrobat 6. I now this for assurity as I (along with others) wrote the search technology used in Acrobat 6.0... Oh, one thing is that you will need to recreate the index as the indexes themselves are not compatible."—Art Pollard, Founder, Lextek International. Source: Planet PDF, June 4, 2003

Marketing

The company maintains a Web site at www.lextek.com. Marketing activities are comparatively modest. In fact, among search vendors tracked by ArnoldIT, only dtSearch has a comparable "out of sight, out of mind" approach.

Lextek's major marketing program is its stop words list. The company provides two lists of stopwords. According to Lextek, "Stopwords are words which have very little informational content. These are words such as *and*, *the*, *of*, *it*, *as*, *may*, *that*, *a*, *an*, *of*, *off*, etc." These are located at: http://www.lextek.com/manuals/onix/stopwords1.html and http://www.lextek.com/manuals/onix/stopwords2.html.



The source input was the February 25, 2004 article in Business 2.0, "66,207,896 Bottles of Beer on the Wall" by Kevin Kelleher. The 1,500 word article was reduced to a 400 word summary. Notice that the Brevity tool did not present the bibliographic information for the story. To generate a useful summary, I manually deleted advertisements, links, and other content presented on the Web page containing the story. The original story is located on the CNN.com Web site at http://www.cnn.com/2004/TECH/ptech/02/25/

Case Examples

The Onix search system surfaced during research into the Acrobat search technology. With Acrobat 5, Adobe adopted Lextek's Onix technology and used it in some other Adobe products; for example, the first version of Adobe Creative Suite.

The Acrobat search function was acceptable. But the system delivered exact string matches. These was no synonym expansion or other procedures to

"Our technology is found throughout a wide variety of industries all of which want indexes that are generated quickly and even more importantly, searches that are conducted accurately and in reasonable time. People do not want to have to sit around waiting for the results of their search"——Art Pollard, Founder, Lextek International. Source: Planet PDF, June 4, 2003

assist a user unfamiliar with Adobe's terminology and jargon. Because of Adobe's idiosyncratic language, locating information via a search box was often a frustrating, needlessly time-consuming exercise.

Adobe continues to rely on Lextek. With Acrobat Reader a free and ubiquitous software, readers can try out the Onix search system and draw his or her own conclusion about the system.

The company asserts:

> Onix is used in a wide variety of computer applications, from electronic publishing, document management, and imaging to web crawlers, search engines, and mail routers.

The challenge for a procurement team is locating specific examples of the successful deployment of Onix in these applications. My efforts to contact the company were met with suspicion. A potential licensee seeking information about Onix may encounter a different reception.

Lextek Products

Lextek offers for license a number of tools. These primary product is the Onix Full Text Indexing and Retrieval Toolkit.

The firm's other products include:

- A Profiling Engine SDK. The licensee can compare a subset of documents against documents in a Lextek index or recently indexed documents. The "matches" permit routing, classifying, and analyzing of information. The company asserts that the Profiling Engine can handle inflows of content, not just matching against static indexes.
- RouteX. Document Classifier and Router. This component classifies and touts documents. The rules-based approach requires a person to set up and manage the rules. The manual approach incurs on-going maintenance costs.
- Brevity. This component generates summaries of documents. Lextek offers an online demonstration of the system. The system is located at http://www.lextek.com/brevity/brevtest.html. I inserted the text of a Business 2.0 article into Brevity. The summary is acceptable.
- Lextek Language Identifier. This component automatically identifies the language of a document submitted to the system. The tag generated by the system can be used to route documents based on language. This product is available as freeware at http://www.freedownloadmanager.org/download/character-encoding-in-pagemaker-5000250.html
- PhonMatch. This toolkit is a phonetic matching toolkit. Unlike most matching algorithms PhonMatch uses advanced phonological (sound) methods.

- SpellWright. This is a spell check toolkit.

Onix Search

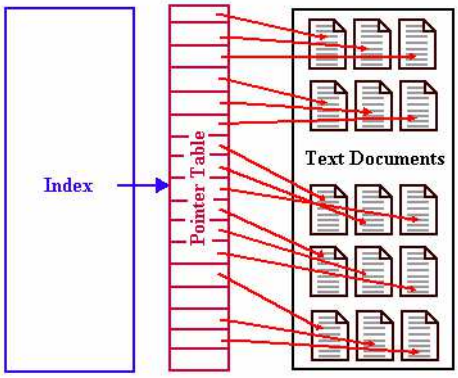
Onix is not a product. Lextek licenses code modules. These have to be configured and integrated into the licensee's system.

Developers can use the Lextek tool kit to create customized information retrieval applications. Lextek's approach is to provide basic text retrieval with the expectation that licensees will assemble their own provides the underlying text-searching technologies, allowing other developers to build their specific forensic application around them. Lextek offers a wide range of document management systems using text-indexing technologies.

The Onix system is a collection of components. The components can be configured to perform the information retrieval functions the licensee needs. The functions that Onix makes available include, among others routines for:

- Creating and dispose of the index manager
- Creating indexes
- Getting and setting the location of temporary files and get their sizes
- Opening and closing indexes
- Indexing
- Implementing distributed indexing
- Sorting routines
- Storing and retrieving an arbitrary piece of data that is associated with every record
- Managing record IDs
- Providing statistics about the index
- Managing the deletions of records
- Processing of standard queries
- Accessing and navigating the word list
- Manually processing queries
- Supporting Unicode
- Highlighting query terms
- Assisting Web crawlers parse robots.txt
- Managing the stop word list
- Managing and getting status information.

Using these and other functions, a licensee can add search and retrieval to almost any enterprise application or software product.



If content classification is required, Lextek offers a profiling engine. Content processed by Onix will be automatically classified, routed, and filtered. The Profiling engine provides licensees with a way to create standing queries that match a user's information requirements. When new content is processed by the system, the profiling engine can generate an output such as an email and send the new information to the individual. If licensees require multi-language support, Lextek offers a language identification component. Onix supports 260 language and encoding modules.

What Lextek licenses is a C++ code library. Lextek asserts that the system incorporates features that speed up content and query processing. In addition, Onix consumes fewer resources than some other enterprise information retrieval systems. Licensees, according to Lextek, receive unlimited technical support. Many vendors charge for support after a period of time.

Core Features

The principal features of Onix include, according to the company's documentation:

- Indexing speed, even on systems where memory use is critical.
- A proprietary index compression and storage technology
- Support for a variety of index formats, allowing developers to tailor the toolkit to fit their application.
- Support for standard Boolean operators
- Range, phrase searches and proximity queries.

Building Search Systems with Onix

"I am the founder and president of Lextek International. We specialize in developing high performance indexed search systems for developers. Typically, people take our technology and integrate it into their applications be that a Web-based application or an end user application such as Acrobat. While I end up performing many duties as president of Lextek, I still remain the chief architect of our search products due to my long experience in this field."——Art Pollard, Founder, Lextek International. Source: Planet PDF, June 4, 2003

The instructions for building search systems with Onix appear in "Onix Text Retrieval Toolkit." This is an application programming interface reference document. The API reference states, "You are looking at the manual of one of the fastest full text indexing engines available." The reader is assured:

> Users of Onix have been amazed at its fast indexing speeds as well as its flexibility and ease with which it can be integrated into projects.[1]

The system, however, has limits. According to the API documentation, the Onix system can manage "small text databases" or "large text databases approaching one terabyte."

There are two Onix APIs: one for CD-ROM projects and one "aimed at text databases which are either of very large size or require periodic updating of the index."

The Onix system pivots on several important concepts. The first is "words." Licensees define what "a word is inside the index." Onix allows the licensee to normalize words via a stemmer. Lextek provides a starter list of stop words. As mentioned the Lextek system uses the Porter Stemmer, a lemmatization routine developed by Dr. Martin Porter in the 1980s.

Onix makes use of a concept called "records." Onix indexes chunks of source content. The analogy used in the API documentation is a page in a book. Onix indexes the page in which the word occurs. The company says:

> Choosing how large a record should be is an important choice when building your application... Certain operators such as the Boolean operations AND, OR, and NOT operate on the record level returning which records match your Boolean expression.

Onix opens and closes an index, indexes files, conducts queries, and allows deleting a record. These operations work at the record level. Sizing a chunk of content as a record is an important step in setting up Onix. Get the record size wrong, and the results list will point the user to a chunk of text that can be either too large for the user to scan quickly or too short so the retrieved item does not answer the user's question.

Content Processing

During the indexing phase, it is common for a licensee to write a pointer file which stores information on how to find the record indexed. This pointer file can be as simple as a list of four byte integers specifying the offset into a file a record begins or it can be as complex as two different files – one specifying

1. "Onix Text Retrieval Toolkit: API Reference," Lextek, page 1.

a variable length field (such as a file name) and the other specifying how far into the other pointer file the variable length field begins.

Onix is sufficiently flexible to allow the licensee to determine what to store in the pointer file. Unlike a system such as Google's or dtSearch's, a search engine engineer can specify specific attributes and values to include in a pointer file. A good example of this is Adobe's inclusion of hyperlink data in its indexes.

Onix also allows the licensee to define what a word is in the text that is indexed. A word can be composed of, for example, simply a sequence of characters a-z or, it can contain upper ASCII/ANSI extended characters or it can be Unicode or any other sequence of binary data (during the indexing process a developer must specify what the binary data is and how long it is).

Some people want to "normalize" words before they are indexed. A common way to do this is to use a stemming algorithm that normalizes all forms of a word into a standardized form – which may or may not be a real word. Onix currently has the Porter Stemmer as part of its toolkit for the English language. The Porter stemming algorithm is considered by many to be one of the better stemming algorithms developed. Stemming has its share of advantages and disadvantages and only a specific licensee's application will dictate whether it is desirable to stem words before indexing. It is important to keep in mind, however, that if a licensee stems a word before the document containing the word is indexed, the search terms must be stemmed before launching the query. Lextek provides a detailed discussion of the stemming issue in its API documentation.

Onix expects the licensee to divide the indexed text into "records" during indexing. A record of text can be just about any size. Choosing how large a record should be is an important choice when building the application – though the circumstances make it fairly easy to decide how to divide up the text.

In terms of storage, the smaller the records, the larger the Onix index will be – when compared to the original text size. The larger the records, the smaller the index will be as a percentage of the original text size. The reason for this is that there is a minimum amount of space that is needed in the index for each record. Larger records allow this space to be amortized over larger amounts of data than smaller records would.

These technical issues underscore the need for the licensee to have skilled technical professionals available to configure the Onix system. The specific approach to configuring Onix depends on the licensee's application.

Query Processing

Onix provides as part of its query processor a query optimizer that evaluates the query and deduces the optimal way that a query should be performed by

the query processor. This helps give Onix the ability to solve queries in a shorter time than could be done otherwise and is also one of the reasons why Onix may prove to be faster than many other systems you will test.

The result set features in the Adobe implementation a document rendering and term highlighting. The “search box” is visible, so a user does not have to hit the “Back” arrow to modify a query or to make use of a special feature implemented by a developer such as a “Bookmark” or a hyperlink to another document or a Web site.

The Secret Site showcases some of the Onix functionality. The site is located at http://64.38.176.115/search.html

Onix takes all its queries in a normalized form, which allows it to search on any data that was indexed – even binary data. It does this by representing the query terms in hexadecimal. For example, a search for the term “whale” would be represented as: “0x7768616c65” and a search for “white & whale” would be “0x7768697465 & 0x7768616c65”. Onix provides support routines to help a licensee convert queries into this normalized form.

Onix Query Language

Onix has its own query language; the licensee’s engineers need to learn it. The Onix Query language consists of a series of statements. Each statement is a query expression, a variable being set to a query expression, or the definition of a function containing other statements.

Query expressions resemble the search queries in simple search tools. They consist of a series of operands (terms) combined with operators (symbols representing logical operations). A statement can span multiple lines and is terminated by a semi-colon character (;), just as in many programming languages. To send several statements to the query processor the developer simply creates a string with a semi-colon separating each statement.

Sample query consisting of multiple statements:

```
a = ‘apple’ & (‘oranges’ | ‘orange’);
c = a | ‘pear’;
```

To have Onix return to the licensee the results of a statement the programmer must make the last statement an expression. The results of this expression will then be stored in what is called a “vector.” A vector for Lextek is simply

"As a search engine vendor we try to develop a file format that has a mix of performance and other features that we feel will meet the market's demands. Then we develop an indexing-and-search system to along with it that is capable of fulfilling people's needs. Unfortunately, what this all means is that the different vendors are not generally capable of reading each others indexes. We are no exception."——Art Pollard, Founder, Lextek International. Source: Planet PDF, June 4, 2003

a list of "hits." Each hit is a document and word number with the weight of that hit.

In most Onix search implementations, each query will represent a category or concept the licensee has created. So each "hit" tells the users those documents that fit the category and how well they "fit" the category. Term weighting controls allow the programmer to tune the algorithms governing "fit" or the statistical distance permitted in the retrieved document set.

Onix offers a broad range of operators. The Boolean operators may be combined with the relevancy ranking capabilities in a very flexible way. Queries may be Boolean, ranked or a combination of both where the results must match the Boolean query but are ranked according to the ranked query.

Onix Modules

Licensees can select from specific information processing modules. The principal search-and-retrieval modules available from Lextek include:

Lextek Profiling Engine SDK

Lextek's Profiler is an indexing engine customized for the needs of comparing a few documents at a time against hundreds or thousands of complex queries. Conversely, many competitors' profiles focus on running a limited set of queries against large document sets. Onix's technology focuses on pre-existing profiling against new content indexed by the system, not against relatively static indexes. This is designed to meet the unique needs of the routing, categorizing, and analysis markets rather than, say, the particular preferences of an individual searcher in the enterprise.

To deal with these very different needs, Lextek designed its Profiling Engine to create memory-based indexes; a technique used in extremely high-performance search applications for wireless applications. Much of the time taken in querying a traditional index is taken up in searching through the index on the disk. That takes time, not only to wait for the disk to transfer the information but also to compute where the desired data is. In general, accessing a disk is several orders of magnitude more lengthy than going directly to memory. By keeping the index in memory Lextek's profiler works more rapidly. However, we believe that Lextek's approach is no more memory intensive than other advanced search systems, including those from Google or Verity.

The Lextek Profiling Engine SDK also has its own query language similar in format and syntax to the Onix toolkit. This is a full programming language allowing code reuse. It also has ranking methods, deferred evaluation of queries, and named subqueries.

RouteX Document Classifier

RouteX Document Classifier and Router comprise a toolkit designed to allow a licensee to classify documents or take actions on a document. The system is rule-based. The licensee will have to create these rules. The module supports forwarding documents to specific people based on criteria the individuals specify or simply organizing documents based on their content.

Brevity Document Summarizer Toolkit

Brevity is a toolkit that allows a licensee to automatically generate document abstracts. It is roughly comparable to similar autosummarization tools in Verity and other competitors. The Brevity module also enables a licensee to highlight key sentences within a document for display or as links to other documents.

Lextek Language Identifier

The Lextek Language Identifier is capable of automatically recognizing in which language and character encoding a document was written. The module supports most widely-used languages and character encodings.

PhonMatch

PhonMatch is an advanced phonetic matching toolkit.

The module allows a licensee to know if two or more words *sound* the same, Unlike other algorithms that have been developed by non-linguists as “quick and dirty” solutions, PhonMatch is designed to provide a high-quality matching for applications where the ability to tell if two words sound alike is the most demanding, both in terms of performance and in terms of linguistic ability. The rules that govern PhonMatch’s matching algorithm were developed by leading phonologists for Lextek.

Applications of PhonMatch include:

- Product Directories
- Address Directories
- Phone Books

SpellWright

SpellWright is a spell checking toolkit. It provides advanced suggestion algorithms to help ensure that an ideal correction is returned. The dictionary technology provides automatic corrections, memory efficiency and multiple dictionary types.

Unlike many other spell checkers, SpellWright returns suggestions in the order of how probable it is that they are the word the user is looking for.

SpellWright allows an application to have up to 32 dictionaries open at one time, each with several different dictionary styles including:

- Main Dictionary
- Skip Dictionary
- Document History Dictionary
- Suggestion Dictionary
- Auto Replace Dictionary

Licensees can assemble their own, custom dictionary.

Technology

Onix is written in ANSI C++ and runs on a wide variety of compilers and platforms. Onix is available for Windows, Macintosh, Linux, Solaris, and BSD Unix, among others.

As you might expect, Onix includes an extremely wide range of functions within its APIs. This range may obviate much of the manual coding required to integrate other vendors' search tools into applications.

The documentation for the API runs to hundreds of pages, and it is, therefore, not possible in this report to do much more than provide an indication of some of the functions available to a licensee. The core functions in the API provide basic features for opening and closing an index as well as more esoteric actions. Examples include:

- Supplemental routines for performing distributed indexing
- Four separate routines for storing and retrieving an arbitrary piece of data that is associated with every record
- Seven functions for deleting records
- Eleven functions for handling standard queries
- Five Unicode support routines
- Eight functions for managing the crawler's interaction with content to be indexed.

The broad range of functions means that a licensee can integrate Onix into a variety of portal, content management, application server, or desktop applications that need search technology.

A licensee is able to classify documents based on security. During indexing, a "hidden" keyword – that is, one that does not occur in the text – describes a security group. When the query is passed to the query processor, the hidden keyword is added to the query to ensure that only the records that come back have that keyword for a specific or security group. Although on the surface

Lextek's approach seems primitive, the licensee can extend the security model to meet the specific requirements for the search system deployment.

ArnoldIT Opinion

Onix is a set of software modules. Licensees, like Adobe, are responsible for the search system assembled from these widgets. Nevertheless, I have been dissatisfied with the Adobe search service for many years. Although I try to remain objective, Lextek flashes a yellow caution light for the ArnoldIT team. The sparse marketing collateral, the failure of the company to make the Onix brand findable via a simple Google search, and the very low profile the founders have in the information retrieval community make me somewhat cautious with regard to Onix.

Table 2: On the surface, Onix provides everything an organization might want. The support for Unicode is robust. The company makes provisions for automatic classification of documents, includes workflow capabilities, and supports multiple languages. Lextek appears to be an outlier compared to other vendors. In certain procurements, the company's invisibility could be either a help or a hindrance.

Anticipated Benefits

For an organization that wants to build a custom enterprise or software search solution, Lextek may be the ideal solution. Adobe found the Onix technology an acceptable solution for its products. Other organizations are likely to agree. Onix can be configured to deliver the quite specific features and functions. Unlike the Google Search Appliance that arrives ready to plug in and use, Onix has to be constructed from its comprehensive enterprise search toolkit.

Onix is, therefore, not a product like Coveo or ISYS Search Software. Onix provides tools and code components. For companies with strong developer talents and requirements for specialized, high-performance search applications, Lextek can be an excellent choice.

Table 3: Since 1993, Lextek has added a range of statistical, natural language, and proprietary algorithms and information processes routines. An Onix licensee has direct control over every aspect of a search solution.

Table 4: Onix Checklist

Attribute	Verity Asserts	ArnoldIT Comment
1 Platform	Linux, Microsoft, Unix	
2 Keyword search	Yes	Boolean, free text, and natural language supported
3 Text mining	Basic analytics are available	Third-party routines can be integrated with the system
4 Automated indexing	The system indexes words	Lextek Onix does not perform synonym expansion unless the licensee codes these operations
5 Personalization	No	Lextek Onix does not perform personalization unless the licensee codes these operations
6 Workflow	No	Lextek Onix does not perform operations unless the licensee codes these operations
7 Interface	Customizable	The licensee codes the interface
8 Hosted service	Yes	The system can be configured by the licensee to operate as a hosted service
9 Administrative interface and tools	Command line interface	
10 Application programming interface	Yes	A 700 page reference manual documents the hundreds of options, functions, and operations Onix makes available
11 Professional services	Yes	Due to the small size of the company, engineering and support resources may be stretched thin at times
12 Security	No	
13 Connectors	Text and Web content connectors provided	The licensee can code or integrate content connectors as required
14 Support for structured data	No	The system can be adapted by a licensee to process records in a structured form
15 Relevance ranking	Yes	
16 Video	No	
17 Federated search	No	
18 Fielded search	No	
19 Content crawler	Yes	The system includes a Web crawler component
20 Price	Negotiated	

Possible Drawbacks

An organization looking for a search system that is a ready-to-run product may want to look at an appliance or a hosted search system. Onix is a collection of components. These have to be configured, assembled, and integrated into a solution. Lextek promises unlimited customer support. With the low profile the company has, a licensee may want to make certain that its developers have the expertise to implement an Onix solution. “Learning by doing” may work in some situations, but for an enterprise search system, the approach is likely to be time consuming and more expensive than an appliance or a hosted service.

If that’s not the profile of your firm, consider looking elsewhere. Other search vendors can provide their customers with plug-and-play solutions. For licensees without solid search requirements and a strong engineering team, Lextek is likely to be a solution that has a steeper learning curve than many other competitors’ products, and some of its approaches (such as its query language) may take some time to learn and use without experimentation via trial and error.

Other potential considerations are:

- Lextek Onix has a low profile. Locating experienced engineers who can work with the Onix tools may be a challenge.
- The founders have not been particularly active on the enterprise search conference circuit. Combined with the lack of technical publications and minimal marketing program, Lextek is one of the least visible vendors in the information retrieval market.
- The company’s management has shifted focus to a chocolate business. For mission-critical deployments, potential licensees may want to consider a search vendor with partners, resellers, and an in-house technical support staff that can handles a large-scale search implementation.
- The company operates from a suburban home. After more than a decade in business, the lack of a formal office in which to receive prospects and hold meetings may be a negative for some organizations.

Net Net

The Onix toolkit is a fine choice for constructing a customized enterprise search solution. Lextek imposes comparatively few constraints on platforms, making it possible to integrate Lextek search into portals or develop a completely customized solution for very specific search applications.

With the flexibility comes the need for commitments. The licensing organization must be prepared to support the product with hardware, network infrastructure, and technical staff. Lextek is not right for every major enterprise. But when the requirements for search cannot be easily or economi-

cally met using an “out of the box” search solution, Lextek may well be the optimal choice to explore with an in-house pilot and head-to-head testing against other customizable systems from TeraText, Autonomy, and others.

Stephen E Arnold
Minor edits to a 2008 draft finalized on March 18, 2014