iPhrase Technologies

iPhrase One Step offers natural language processing as "the first truly dynamic Web site navigation and precise search platform. The company was one of the first to claim its system used "adaptive learning" in order to exploit "content intelligence" and move from keyword search to user experience via natural language processing.

iPhrase blossomed when organizations waned a "silver bullet" for search and retrieval for eCommerce. The company added enterprise search to its One Step platform. In a short time, iPhrase offered a "framework" upon which a licensee could, in theory, hang any search-enabled application. iPhrase and Endeca were quite similar. Both companies went into business about the same time. Both companies drew on the Massachusetts Institute of Technology for technologies. Both companies offered facets. Both companies were based in the Boston area. Both companies sprawled across multiple information markets. The iPhrase system could support almost any information service provided that the licensee had the money and appetite for large-scale enterprise software projects.

The iPhrase approach required significant hardware and storage resources. IBM was a partner of iPhrase' and acquired the company in 2005. When the deal was announced, iPhrase had filed for a number of patents and seemed to provide IBM with eCommerce, natural language processing, and other advanced functions that could "snap into" the OmniFind products. As it turned out, IBM had to buy a taxonomy management company to deal with the poorly managed metadata operations in iPhrase. Then IBM had to purchase Vivisimo to acquire federating technology that was more modern than the first-generation iPhrase approach. Shortly after the 2005 acquisition was finalized, the iPhrase and One Step brands began to dissipate.

Author's note: This is a 2006 draft. It will not be updated.
Stephen E Arnold, January 7, 2014

Introduction

Privately-held iPhrase opened for business in 1999 by a group from the Massachusetts Institute of Technology's Laboratory for Computer Science. Instead of keyword search, iPhrase offered a "vision" of a "one step search portal framework." The company embraced enterprise search, but the main customers applied iPhrase's system to eCommerce. The company integrated the financial jargon of ROI or return on investment into the company's marketing collateral. The slogan for iPhrase search was "One Step to ROI." iPhrase, like other search vendors, had an opportunity to provide hard data that One Step delivers more revenue and lower costs than directly competitive systems like Endeca, a company whose technology is somewhat similar to iPhrase's.

Table 1: Hummingbird Search Server: A Bird's Eye View

Product Thumbnail	
1 Search Brand	One Step (initially branded as "Smart Language") and Web Resolve, a self-service customer support system
2 OS Supported	IBM operating systems, AIX, Linus, and Microsoft operating systems.
3 Est License Fee	$500,000 and up. A custom price quote is provided to prospective licensees License fees do not include servers, storage, and engineering services.
4 Functions	Automatically recognizes phrases as phrases, adapts to misspellings, seeks related terms for user queries, understands searches expressed as a question, and infers meaning based on content structure. The system includes cross-sell and up-sell functions when One Step is used for eCommerce. Stores content processed by the system in a content repository anticipating the functions of an Extensible Markup Language data management system.
5 Claimed Features	The system displays "answers" not just a results list. Adaptive learning, natural language processing, content intelligence, built-in report builder, a management console or "dashboard" in addition to standard search and retrieval operations
6 Downsides	The system required significant computing and storage resources to perform with low latency. Licensees should be prepared to invest in significant hardware and network infrastructure to minimize latency within the system.
7 Similar To	A kissing cousin to Endeca. The similarities, particularly in faceted searching and the focus on eCommerce, are quite interesting and mostly ignored by licensees and analysts.
Product Close Up 	iPhrase asserts that it has developed the first truly dynamic Web site navigation and precise search platform, providing users with one-step, direct access to all of a site's content. Coincident with Autonomy's emphasis on "understanding" content, iPhrase asserts that One Step interprets the concept behind a user's query. Then the system matches the query to the terms and the metatags in the index. The output is a "tailored response." These operations are performed in "one step." The product is a suite of software, linguistic and analytic processes, and administrative tools to provide enterprise search, integrate content into an Intranet or an Internet portal, and power a self-service customer support The One Step platform combines natural language processing, multi-source retrieval, dynamic presentation and analytics to simplify access to information.

In 2005, IBM purchased the company. The iPhrase technology provided structured and unstructured content processing and search functions for IBM's Apache-based line up of content management systems. iPhrase emphasized that its system performed many functions automatically, thus

"**For organizations taking control of their taxonomy and metadata initiatives, iPhrase One Step leverages this additional information to fine tune search precision and facilitate dynamic refinement of search results. One Step's unique capability to understand user queries and to present results in ways that make it easy for users to take action improves the overall user experience and effectiveness of information requests."—** Paul Esdale, SVP of business development at iPhrase Technologies. Source: BusinessWire, May 25, 2005, at http://bit.ly/1e8pPaU

reducing certain costs accrued by less sophisticated search-and-retrieval systems.

iPhrase, in my opinion, was a reworking of concepts pioneered by Fulcrum Technologies and Verity, among others. iPhrase offered One Step, a search system that was to eliminate the many different and complex procedures required by competitor's information retrieval systems. iPhrase offered:

- A system that can answer a user's questions expressed in natural language; for example, "What are IRAs that return more than 5 percent per year?"
- Output generated for each user with "dynamic summaries" of the most relevant documents
- Automatic indexing of structured (database) content and unstructured information (email and Word files)
- Support for standards like Java, XML (Extensible Markup Language) CORBA, ODBC, etc.
- Classified results and related content and suggestion displays.

iPhrase added some verbal twists to positioning its system; for example, One Step included "adaptive learning" and "content intelligence." I am not sure what "content intelligence" means, but it differentiates iPhrase from its competitors which index "content." iPhrase has its own jargon, including "knowledge connectors" for filters, "toolkits" instead of application programming interfaces, "dynamic presentation" instead of "results", and "direct navigation" instead of hot links to related content.

iPhrase was one of the first search systems to package natural language processing, federated search, and knowledge as a "framework" for an organization. To top it off, iPhrase asserted that its system was easy to install, although the time required was four weeks or more.

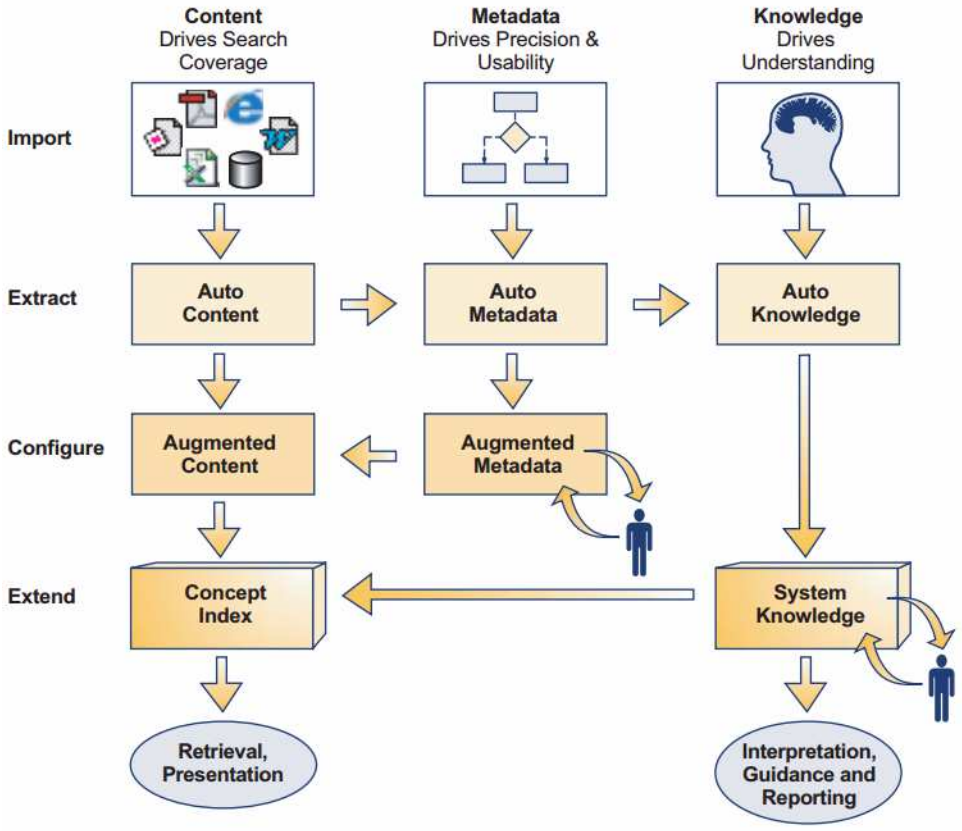
iPhrase, based on its July 2002 white paper, "A Technical Introduction to iPhrase One Step: The Market Leading Search and Navigation Platform that Delivers Self-Server," trumpeted:[1]

- One step to ROI (return on investment, a financial calculation that continues to be quite difficult to apply to search-and-retrieval expenditures)
- Conceptual understanding quite similar to the assertions made for the Autonomy IDOL system
- Multi-source retrieval; that is, converting documents to a format that resided in an iPhrase repository. The stored documents were then available to users without having to retrieve them from the server on which

1. The paper, written by Patrick Fitzgerald, is at http://www.pfitz1000.com/Samples/iPhrase_Technical_Whitepaper.pdf

the original content resided. iPhrase, like Fulcrum, transformed source documents into a format suitable for the iPhrase system.

- Dynamic presentation. The iPhrase system shipped with a library of "report" templates. Search results would be displayed to a user in one of the templates. Licensees can edit the templates and create additional templates.
- A library of "connectors" so the iPhrase system can acquire, transform, and process source content from relational databases, semi-structured Web pages, and unstructured information like a Microsoft Word document or an email. Connectors are available for some content management systems like Interwoven.



At first glance, iPhrase's system delivers a fully automated solution for information retrieval. The system "augments" and delivers "auto knowledge." The well-crafted concepts do not make clear the lack of metadata management tools, the resource requirements for optimal performance, or the lack of a graphical administrative tool.

"As Web sites have gotten richer, effective search and navigation has become a major differentiating factor. We're thrilled to have Yahoo! Finance, an industry leader and usability pioneer, recognize the power of iPhrase One Step to improve the user experience on their site for both beginning and advanced investors." —Noam Ben-Ozer, president, CEO and co-founder of iPhrase. Source: Yahoo news release, December 2001

In a span of 72 months, the marketing angle worked. The company sold out to IBM in 2005 at a time when rumors of a financial slowdown began surfacing at search engine conferences. iPhrase is a supporter of IBM UIMA (unstructured information management architecture, and its complex configuration and appetite for massive computing resources seem to be an ideal match for IBM's service and software business model

The deal is closing as I write this profile of iPhrase. IBM's Jon Prial, vice president of content management and discovery stated:

> iPhrase's technology combines a real-time understanding of user queries with application context, guiding people to the most relevant results tailored to their individual needs.[2]

Selected Executives

iPhrase was founded by a Bain consultant and three Ph.D. colleagues from the Massachusetts Institute of Technology. These individuals are:

- Noam Ben-Ozer, co founder and the firm's first chief executive officer. Mr. Ben-Ozer worked for Bain, a consulting firm.
- Jane Chang, chief scientist and co-founder. Dr. Chang has a Ph.D. from the Massachusetts Institute of Technology. Dr. Chang founded Banter, a natural language processing company which iPhrase acquired in 2004, an interesting interaction among different companies at the same time.
- Raymond Lau, chief technology officer and co-founder. Dr. Lau earned a Ph.D. from the Massachusetts Institute of Technology.
- Michael Kyle McCandless, co-founder and chief architect. Dr. McCandless earned a Ph.D. from the Massachusetts Institute of Technology. After 2005, Dr. McCandless shifted from proprietary search technology to open source search technology

iPhrase, like other search vendors, has tapped a number of other professionals to help the firm acquire clients and, its turns out, help sell the company to IBM:

- Joseph Krivickas, president and CEO after Noam Ben-Ozer left the position in 2002. Mr. Krivickas had been the chief operating officer at Bluestone, an Internet infrastructure company.
- Dan Keshian, chief executive officer in 2003 after Mr. Krivickas left his position
- Stephen Wietrecki, chief financial officer
- Robert Rinaldi, executive vice president of sales. Hired in July 2000

---

2. Barney Beal, "IBM Finds Match in Search Company," Search Business Analytics, November 3, 2005 at http://bit.ly/Jjpftl

- Paul Esdale, senior vice president, business development
- Randy Laughlin, senior vice president
- Larry Angelo, senior vice president, worldwide operations
- Jake Mizrahi, vice president
- Simanta Chakraborty, director of engineering and professional services
- Tony Frazier, vice president, product management in 2002
- Steve Berczuk, principal software engineer.

Financial Performance

From 1999 to 2005, iPhrase had received $36 million via two rounds of financing. The firm's investors include Charles River Ventures, Greylock Management, Reed Elsevier Ventures, RSA Ventures, Sequoia Capital, TD Capital Technology Ventures. and Bain Capital (Noam Ben-Ozer's former employer).

At the time of the deal, iPhrase had about 70 employees. Arnold Information Technology estimates that the firm's revenue is about $19 to $24 million. At a multiple of four to six times revenues, Arnold Information Technology estimates that IBM paid more than $150 million for iPhrase.

Selected Clients

iPhrase has a number of high-profile licensees. By December 2004, the company had more than 100 licensees, although it is not clear if these were one-year or multi-year customers. iPhrase has been able to attract high-profile, resource-healthy licensees who can afford the iPhrase system and attendant infrastructure. Selected clients are:

- Ameritrade
- Charles Schwab & Co.
- Computer Associates
- Country Financial
- Gateway Computers (eCommerce)
- GEICO
- Intuit
- LexisNexis (Reed Elsevier. Reed invested in iPhrase. According to Enterprise Apps Today, "The agreement, comes three months after LexisNexis' parent, Reed Elsevier, led iPhrase's $25 million second round of financing. In 2001, Reed Elsevier expects broader collaboration marrying iPhrase's search technology with LexisNexis and other Reed Elsevier's information services.)[3]

- Mellon Financial
- Neiman Marcus (eCommerce)
- Radio Shack
- Restoration Hardware (eCommerce)
- Staples (eCommerce)
- TD Waterhouse (customer self help)
- US Robotics
- Verizon
- Wells Fargo.

Selected iPhrase Partners

IPhrase's partners are market leaders in customer support and the Web Infrastructure space. These include Siebel, PeopleSoft, Avaya, Kana, Microsoft, and Stellent, among others. The hook for partners is that iPhrase provides the "last mile" of information access because iPhrase's search system can automatically leverage the organization and aggregation of content. iPhrase deserves high marks for finding a clever way to position search as complementary to other enterprise systems.[4]

In 2005 iPhrase formed a strategic partnership with SchemaLogic, an action required because the iPhrase automatic indexing system generated index terms and provided no built-in controlled term management tool. Because iPhrase used dictionaries, "discovered" index terms had to be integrated into the existing dictionaries to permit iPhrase's synonym expansion and concept functions to deliver accurate results.

IBM is an iPhrase partner and was well aware that an iPhrase licensee needed hardware and professional services to make the iPhrase system deliver on the promise of ROI. IBM purchased iPhrase and learned that the lack of a metadata management subsystem was a liability. In 2006, IBM sought to acquire Unicorn, a metadata management software developer.[5]

Shortly after the deal closed, IBM rolled iPhrase into its OmniFind product and began marketing a "discovery" edition and an "eCommerce" edition. This blending of a product that used open source and proprietary technologies provided IBM with a complex, computationally intensive suite of software. IBM includes iPhrase technology in the Websphere Metadata Workbench and the Websphere Business Glossary. Like other aging search

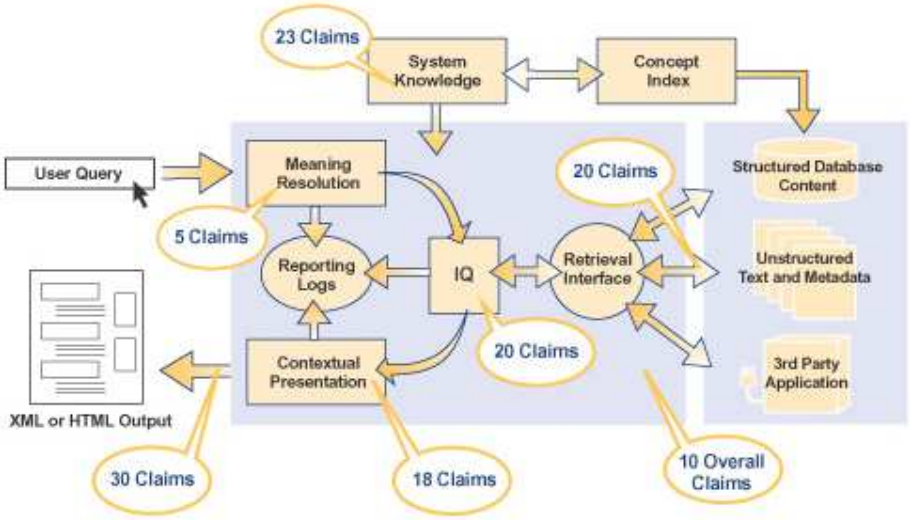
---

3. See http://bit.ly/1durjaP
4. This remarkable bit of marketing poetry appeared in the Gilbane Report, September 2002 on page 16.
5. For background, see the Gartner news item at http://gtnr.it/1bFaneN

systems, iPhrase technology is likely to be sold as up to date as the systems age and become less and less in step with more modern technologies.

Selected iPhrase Patents

iPhrase touted the system's underlying technology. Within a year of opening for business, the MIT-trained founders generated a number of patent applications. One question I have is, "When did iPhrase inventors first develop their breakthroughs?" Was knowledge work done when the principals were employed at other companies or at MIT's research facilities? If not, iPhrase's founders were indeed exceptional technologists and fast-cycle inventors.



iPhrase's technology touches a number of different aspects of search-and-retrieval. However, iPhrase is beginning to package different modules of its technology in order to create solutions. One initiative is the firm's new products for business and competitive intelligence functions.

The company's approach pivoted on methods for understanding the meaning of source content. iPhrase was in step with the state-of-the art in indexing and natural language processing in the 1980s and 1990s. iPhrase's approach echoes somewhat similar claims made by Autonomy, Endeca, and Verity in the period from 1999 to 2004. Will IBM continue to invest in iPhrase's technology post-acquisition?

Third, despite significant intellectual effort in creating iPhrase's "information access methods," one of the founders embraced open source search technology after the sale of iPhrase to IBM. The question key inventors of iPhrase's breakthroughs shifted from proprietary search technology to open source search technology. The question that arises is, "Did the iPhrase expe-

""This round of funding represents another important validation for the power of our solutions and our business model, which are enabling iPhrase to create a truly great enterprise software company."—Noam Ben-Ozer, CEO and co-founder of iPhrase. Source: http://www.enterpriseappstoday.com/call-centers/article.php/795311/iPhrase-Talks-Up-VCs.htm

rience inform Dr. Michael McCandless that proprietary search-and-retrieval systems were not commercially viable?"

Several iPhrase patents provide insight into the mechanics of the "framework;" for example:

- US 6,704,728, granted March 9, 2004. The invention is a method of accessing information from a collection of data. The patent includes receiving a query, amplifying the query via a form of synonym expansion, generating an inverse index of the processed content with a canonical non-terminal representations of the data, and generating results to the query. One of the interesting features of this invention is that source content is stored as "prose" and the user receives "prose" from the iPhrase system; that is, iPhrase makes use of a repository and representations of the content in the form of key words, entities, and metadata.
- US 6,711,561 granted March 23, 2004. The invention discloses a method for query processing.
- US 6,714,905, granted March 30, 2004. This invention discloses the parsing method used to disambiguate a query.
- US 6,745,181, granted June 1, 2004. This invention converts "hits" in the index to a result set.
- US 7,127,450, granted October 24, 2006. This invention sets forth the method of decomposing documents that match a query. The result set deduplicates the result set and summarizes the non-duplicative results. The notion of "space" allows the system to fit the result space into an iPhrase results template.

The selected inventions highlight the key differentiating points for the iPhrase system. These inventions help explain why latency was a persistent challenge for licensees. The iPhrase work flows have more steps than a basic keyword matching system.

The One Step Product Line Up

iPhrase licenses a framework or complete information storage and retrieval system. The firm's engineers work with customers to integrate the iPhrase solution into the customer's enterprise computing environment. A typical deployment takes a month or longer depending on the client's requirements.

iPhrase's solutions are based on eCommerce and now extend to traditional enterprise search and customer support. Products and modules include:

- Web self-service, called Web Resolve. One Step Web Resolve is an online self-service system for customers. No customer support agent is involved. The system permits search and personalization. The person using Web Resolve to address an issue can send an email or chat with a

person. The One Step Web Resolve server runs on Windows and several versions of Unix, including Solaris, AIX, and Linux.

- eCommerce and online shopping
- Portal search and navigation
- Traditional customer support packaged, including the iPhrase Call Center Agent Assist module, Call Center Search and Navigation module, and Contact Classification Plug In, “Email Deflection, and Contact Classification Server
- Enterprise search including structured and unstructured content, user metrics, and alleged near real-time indexing

iPhrase, however, is a newer code base, and the company has a patent pending for more than 30 claims. Since much of the core of search has been available for more than 30 years, iPhrase’s claims to innovation are aggressive.

iPhrase can integrate with provided adaptors or via customized routines Web pages, databases, content management system (CMS) repositories, and other content stores. iPhrase provides a spider, a library of connectors, and a document content integration application programming interface. The platform creates a unified representation of these diverse content stores, treating each Web page, database row, and CMS entry as a unique content item inside the index or “build pipeline” as iPhrase describes these functions.

Case Examples

Neiman Marcus

Neiman-Marcus, according to iPhrase, abandoned traditional search box technology for iPhrase's natural language and rich development environment for a single reason: money. With a search box from a mainstream supplier of search systems, more than half of the visitors abandoned shopping because the results of a search did not match what the user wanted.

Reed Elsevier

In 2001, iPhrase reported that Reed Elsevier invested $11 million in iPhrase. The professional publishing company is looking for a search system that delivers a solid pay off.

The first deployment of iPhrase in the Reed Elsevier empire is at LexisNexis. iPhrase One Step solution was selected for its new global information service in a directory application. LexisNexis officials said that the new platform would be available in 2003, but the roll out was moved to 2004, more than 36 months after the investment. Keep in mind that iPhrase deploys quickly.[6]

LexisNexis, like its competitor Westlaw, has been struggling to regain the growth rates of the 1980s. It will be a good indication of the iPhrase technology to see how One Step can respond to the very different search requirements of LexisNexis legal, government, and business customers.

This announcement is interesting because both Autonomy and Verity have licensing deals with units of Reed Elsevier.

The iPhrase Framework

Like Autonomy and Verity, iPhrase presented its information retrieval system as a framework upon which applications could be built. Like Endeca, iPhrase stressed that the indexing process made it possible for a user to navigate with a mouse click to related information.

iPhrase referred to hyperlinks as "anywhere to anywhere" navigation. The idea is that the iPhrase system would return highly-relevant information with related information a click away. Instead of browsing a list of links and going through the click-open-review process, iPhrase presented information the user can use without additional querying or drilling down into results.

Via connectors, iPhrase places content processed by the system into a warehouse or repository. The system displays documents, snippets, or facts from the repository. A user's query, therefore, taps relevant information regardless of its original file type or location on the network. iPhrase eliminated the requirement for the user to know if a query was appropriate for structured data in a relational database or in an unstructured file type like an email. iPhrase's system could acquire and process content from enterprise content management systems, including Interwoven.

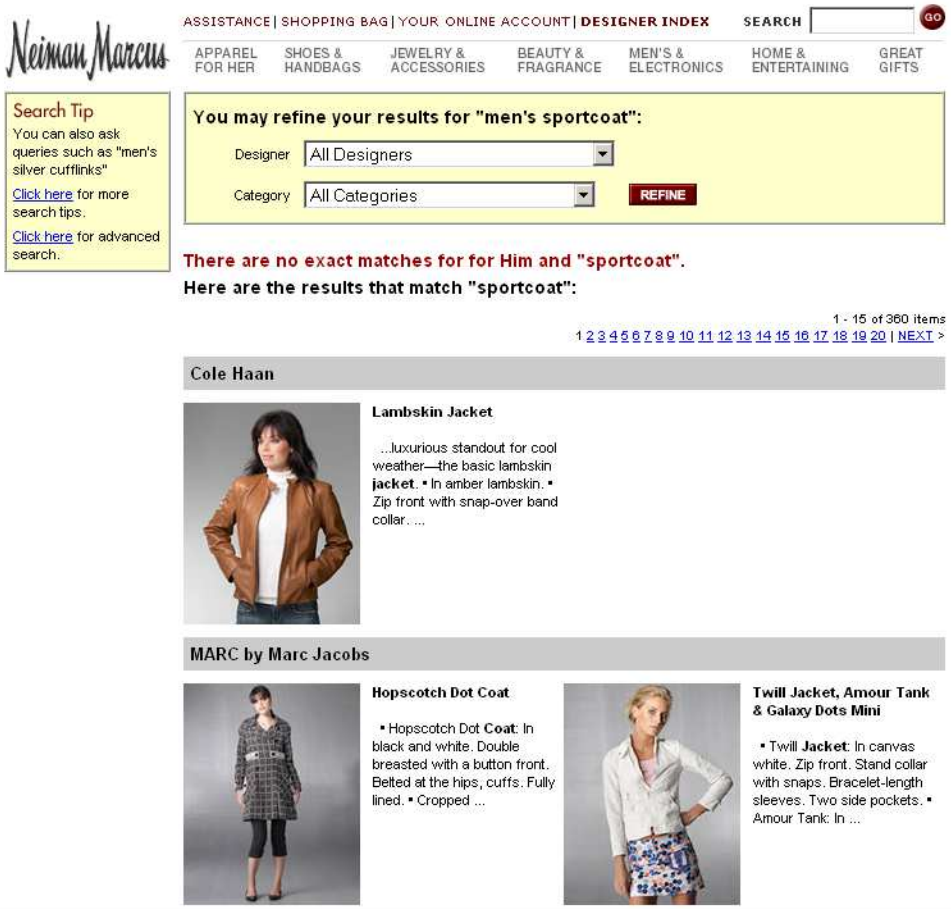
iPhrase places the result sets in an iPhrase template. The templates are suitable for eCommerce results with images as well as tables for database information, traditional results lists for documents, or a report consisting of snippets from relevant, longer documents. iPhrase describes this insertion of results as a dynamic, personalized approach to results display. Licensees can modify the templates to match the graphics of the licensee's Web site or Intranet. In 1999, the rendering of pages on the fly was a novelty.

iPhrase One Step allows a user to enter a query without Boolean operators; for example, instead of entering men and black and sweater, the user could enter "men's black sweater." iPhrase would stem "men's" and automatically insert the Boolean AND as part of the query processing

---

6. Barbara Quint, "LexisNexis Applies iPhrase Natural Language Search Software to Key Directory Files, Information Today, October 29, 2001 see http://bit.ly/19w9JE6.

iPhrase allows the user to phrase questions in everyday language, and is able to intelligently interpret a user's query by understanding both the concepts and the context behind the request. When a user's result set is displayed, the iPhrase system provides a summary of how the iPhrase system interpreted the query. iPhrase calls this "play back." The assumption is that the user will either craft a revised query if the results are off the mark or learn how to interact more effectively with the iPhrase query processing system. The idea is that with feedback to the user, One Step goes beyond other natural language processing systems. Instead of "associations" or brute force "matching," iPhrase is enriching and interpreting the user's query.



One Step provided the search function for Neiman Marcus online store. For this query, the system returned women's wear, not men's sport coats. Other test queries were correctly processed by the system. The mismatch calls attention to the importance of managing the metadata and the index terms used by the One Step system.

iPhrase supports open standards. The idea is that integration of One Step with other software is less burdensome. iPhrase makes use of Extensible

“Even the most well-designed sites require users either to dig blindly through layers of information or to struggle with imprecise search tools. iPhrase's dynamic platform addresses this `information access crisis' head-on by offering the most efficient and intuitive way to connect users with the content they want. We believe the launch of iPhrase ushers in a dramatically improved experience for users and a mission-critical, infrastructure technology for Web businesses.”
—Noam Ben-Ozer via BusinessWire, June 19, 2000 at http://bit.ly/1h3Z8lp

Markup Language. Source content is stored in iPhrase XML in the One Step repository. XML versions of source content are more verbose than some source content. Storage for the iPhrase processed content and indexes can be an issue in some deployments. Decompressing XML can add latency to the system.

iPhrase differentiates itself from its competition by asserting that iPhrase is designed for rapid deployment. The idea is that iPhrase embodies a flexible architecture and permits easy customization. The assertion is a good one because systems from Autonomy and Endeca often require considerable time to set up and deploy. Endeca’s approach is consultative. Autonomy’s approach requires training the IDOL system. However, the prospective One Step licensee will want to keep in mind that iPhrase’s system is as complicated as those from Autonomy, Convera, Endeca, and Verity.

With each release of its framework, iPhrase takes steps to improve its administrative tools. iPhrase lags behind Endeca in this important search feature. As of mid-2005, One Step offers a Web-based management console that allows authorized users to manage certain aspects of their search. The management console permits system testing, accessing report analytics, to configuring certain portions of the iPhrase system to their particular needs or their department's requirements. However, the system administrator will require coding skills and a thorough knowledge of the iPhrase system to configure and maintain the One Step system.

iPhrase’s more advanced features can be more easily deployed in iPhrase’s portal framework. iPhrase, like Endeca, is blurring the line between search and creating an access environment in which search and advanced, “intelligent” features can be made available to users.

Unlike Endeca, iPhrase wants their personalization, analysis, and NLP functions to allow the system to present to the user the information he / she seeks. Endeca, on the other hand, relies on presenting each user with screens that guide the user to the information he / she needs to answer a question.

Which approach is better or cheaper? There is no definitive way to “prove” that one approach is better than the other. The search-related features of the iPhrase system are based on monitoring user behavior and the user’s “location” or “context” in the online system. As the user interacts with the system and the “hits,” the iPhrase system displays information that meets the user’s needs based on interpretive analysis of user actions.

The “Search Portal Framework” anticipates the user’s possible next steps to guide without pre-programmed “screens” to related resources, products, or services.

Search Portal Framework allows the user to access information from across the enterprise. The framework dynamically selects the most appropriate products, services, and information from the data, document, and application resources indexed in the system.

"Our focus is on the corporate Intranet. In the more structured world of intranets, web crawlers are not the way to go. fare are trying to differentiate ourselves by ease of use," Mr. Eddison states. "Our Intuitive Searching allows the user to grab an interesting paragraph, and say: "Find me more like this. We firmly believe that the user is the smartest part of the process, and we provide tools to help the user find information, wherever it is."—Tony McKinley "Intranet Content at AIIM, no date.

To answer the question "What is a large cap fund?" Search Portal Framework will extract information from a glossary, an FAQ, and the Fund database to present a targeted, highly-relevant response and offer the user a clear, easy-to-grasp view of the available information resources and relevant documents or data.

The informing purpose of the One Step the framework is to allow a user an experience similar to the type of dialog a user might expect to have with a knowledgeable colleague or an information specialist." Getting a computer to "understand" a human utterance is a very difficult problem, and I am not sure if the cumbersome One Step approach provides the solution the company's marketing collateral presents in a compelling way.

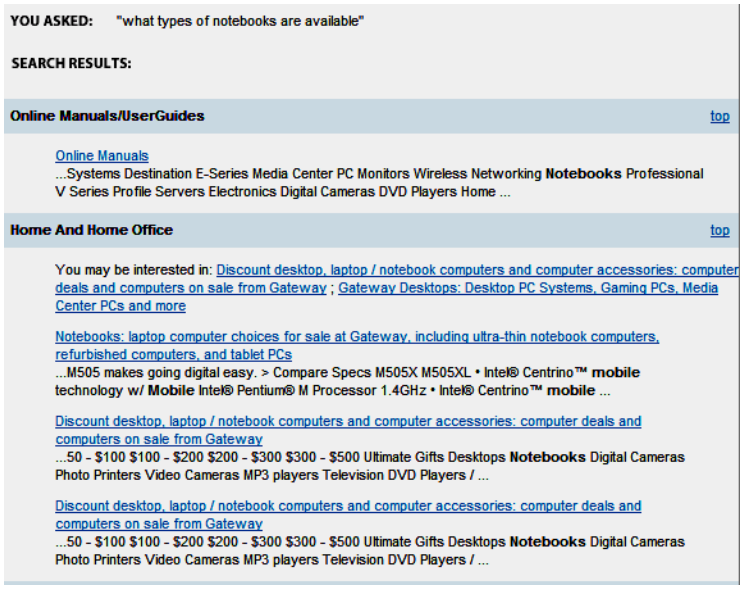
The Natural Language Component

Like other search systems, iPhrase processes the query and passes the reformed query against the One Step indexes. The iPhrase natural language processing subsystem relies on these different types of metadata to "understand" the user's query. Keep in mind that iPhrase's personalization component taps the user's profile to narrow search results based on the profile data.

Once the queries have been converted, the iPhrase system performs proprietary functions to match the values of the query string with the data in the One Step indexes. NLP systems make an attempt to "understand" the intent of the user by factoring in to the retrieval and processing of hits the user's context or location in the interface, the various sources of data available to that particular user, and the metadata generated when the document or content is indexed. These processes are proprietary to iPhrase, and they allow the search-and-retrieval system to:

- Display related links that the user may find helpful in looking for information across the data in the system or specific links that are displayed for a particular employee or business partner such as previous orders or a product display of special offer products.
- Cluster hits from each topic domain available to that particular user.
- Relevance ranking of the hits within cluster and displaying a text summary generated by the system at the time of the query. Each dynamically generated page of "hits" can also display "See also" references to other

information in the iPhrase system that match a particular query each time the user refines a query.



The iPhrase response provides links that answer the user's question from different vantage points. The Gateway example above provides hyperlinks to specific products for a home user and for a business or professional user.

iPhrase's search solution accepts a question such as "What is an IRA? iPhrase understands that this question requires an explanation of individual retirement accounts. The search engine will take the user to a Web page that displays this specific information plus related information such as links to IRA investment options. In iPhrase's approach, the first hit will be labeled as a definition of an IRA. If a user types question on the Gateway.com site similar to "What types of notebook computers are available?", iPhrase displays a page that presents headings with specific links as shown below:

One of the key components is the index and associated knowledge bases. The One Step system does not include an administrative tool for managing the index terms and the "associations" used for query expansion. Indeed, maintenance of term lists and mapping relationships among the knowledge bases provided to the system and automatically generated by One Step when content is processed is a significant and on-going task. The provided analytics and tuning toolkit provides a mechanism for enhancing iPhrase; however, the licensee will have to integrate a third-party metadata management system like Unicorn or SchemaLogic. Metadata management is a non-trivial issue.

Stemming (lemmatization) identifies the root term for each keyword. Respelling accounts for query misspellings relative to terms in the index.

Then concept Identification finds meaningful synonyms and grammar relationships. A “Relationship Identification” process finds other semantic relationships and passes on processed concepts, phrases, and keywords to the results acquisition module.

iPhrase claims that One Step treats phrases as phrases, seeks related terms, and infers meaning based on content structures. However, these and advanced operations such as automatically narrowing or relaxing a user’s query require that the indexes and knowledge bases be consistent and the mappings meaningful.[7]

iPhrase has developed a “Precision Query Optimizer.” When a user enters a query, the Query Optimizer automatically interprets and enhances every query to ensure that users get the best possible responses. The query optimizer expands single word queries and narrows the results sets. This feature can be tested at Gateway.com.

When the query is passed against text--called Unstructured Repositories--iPhrase searches content and affiliated meta-data using its text retrieval algorithm using prioritized concepts, keywords, and phrases identified when the query is processed. For example, the query “IRA early withdrawal penalties” will return documents that include “Individual Retirement Accounts” or “IRA” and the phrases “withdraw” or “withdrawal penalties”

The “secret sauce” of iPhrase occurs when content is indexed and the associated metadata are generated. The metadata and indexing are used to produce classification, term frequency, inverse database frequency, length, timeliness, keyword prominence and positioning, The indexing and metadata provide data for filtering, analyzing, and comparing hits on multiple attributes. Search results can be displayed in almost any way via style sheets. The system administrator can activate routines that can extract sections of source data and show them in table form.

The system includes search tuning capabilities. The system administrator can use these tools to add spelling correction for certain terms, alter stemming features, and add synonyms to the dictionary. The tuning tools also include an interface to manipulate weightings for certain content or hits, adjust relevancy criteria, and other business rules against content and metadata for improved precision control. iPhrase provides options to add metadata automatically or programmatically in order to influence search results and relevance.

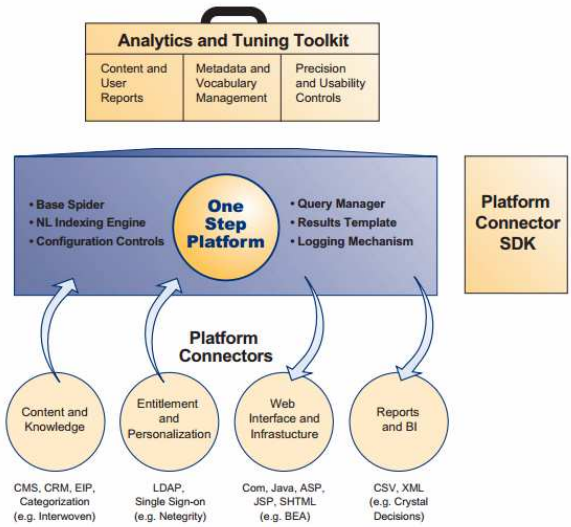
---

7. See the iPhrase technical paper “A Technical Introduction to iPhrase one Step, July 2002, pages 6 and following.

Selected One Step Features

Data loaded into the system is transformed into structured records in row and column format. These records are linked to with associated textual content, metadata and optional access control restrictions. Complex architecture is becoming the “new normal” for search and content processing. With proprietary technology, iPhrase includes a number of components in One Step.

One Step manifests a “unified content architecture.” iPhrase, like TeraText, indexes content in a consistent, XML-based format. The approach allows the licensee to manage the system within a single framework. The often-sluggish response with querying content across multiple servers distributed geographically is side-stepped. Like TeraText and other repository-centric enterprise search engines, One Step erases the distinction between structured data and unstructured content. Content in a One Step index and repository can be sliced and diced, counted, analyzed, and displayed in many ways. One Step's unified architecture places each piece of content, regardless of type and source, into an extensible unified schema.

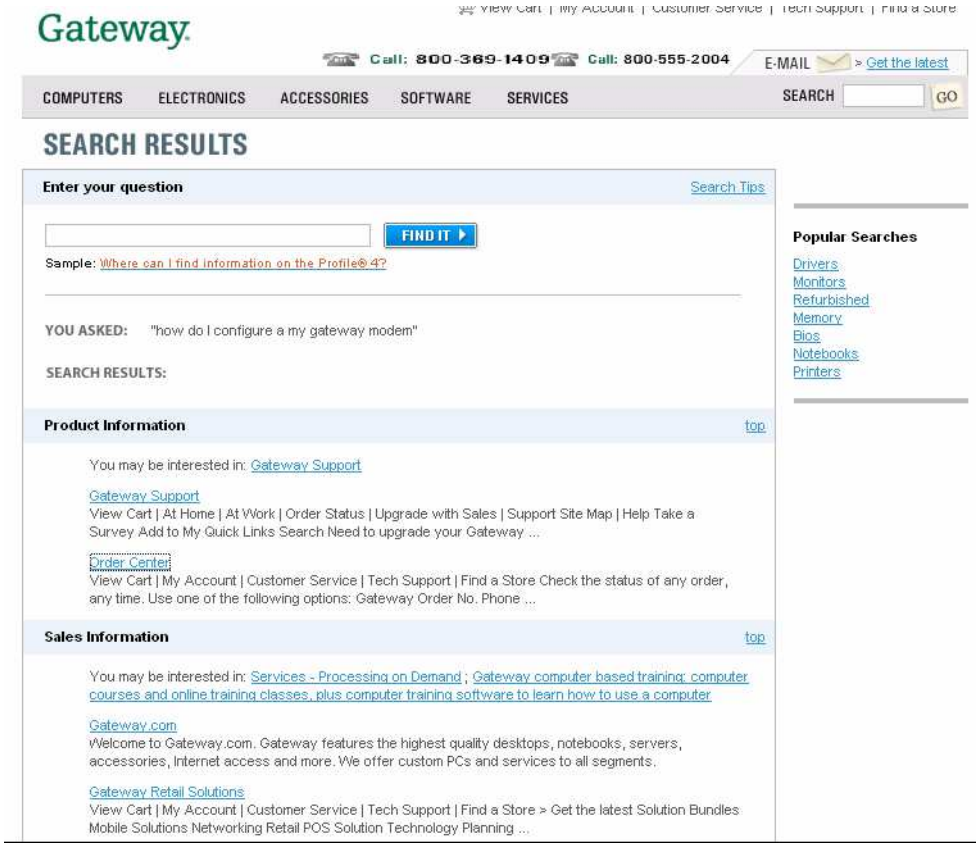


The simplification of the iPhrase system obscures the multiple subsystem ems required by One Step.

A Metadata Pipeline

iPhrase uses what it calls a “metadata pipeline.” When a document is processed by One Step, the information content and metadata for each content item are extracted and assigned. “Extraction” means that iPhrase indexes key words in the content object. “Assigned” means that One Step uses con-

trolled term lists, dictionaries bound phrases like “White House”, and other lists to index the content. Pointers to categories and other metadata are written to the index during content processing. Keep in mind that the source document or database is transformed to an XML encoded object and stored in the iPhrase repository.



The Gateway Computer One Step eCommerce system provides a list of hits, a statement of how the iPhrase system interpreted the user’s query, and a list of popular searches.

Basic content processing involves:

- Tokenization. In this step, One Step identifies each atomic unit in a document. The first step in document processing. This is the equivalent of tagging each structural unit in a source document. The approach is more fine-grained than Fulcrum Ful/Text’s identification of a document “zone.” Not surprisingly, content processing is more computationally intense than key word indexing.
- Lexical analysis. One Step processes the stream of characters that make up the content object. The system reads each character and groups “lexemes”. A “lexeme” is a word-like element of a content object; for exam-

ple, keywords, identifiers (tags), and punctuation. The lexemes are then passed to the One Step parser.

- Named entities. The system "reads" the lexemes and identifies elements that match entries in the knowledge bases for the system.
- Categories. Content objects are then classified and assigned to one or more topics
- Event templates. A process similar to that used by AskJeeves.com to answer user questions then matches the processed document's index with schemas that say, for example, "When this happens, then obtain this information."
- Ontologies of concepts. One Step system then identifies and generates a hierarchical structure of vocabularies and phrases necessary for clear communication within knowledge domains known to the One Step system. One broad way to think of this process is to assign a document to the Accounting Department as an Order.
- Mapping. One Step then produces tables that map a content object's words to categories and ontologies.

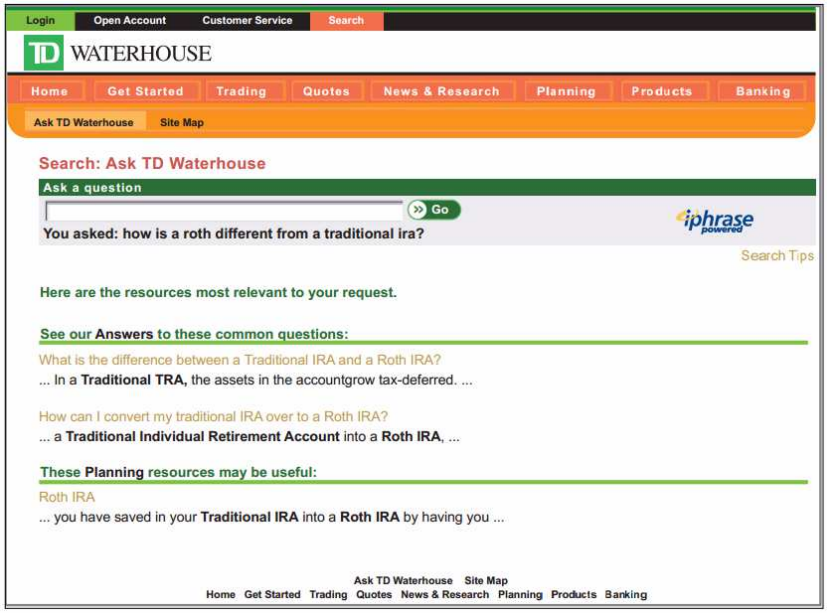
A One Step "linguistic analyzer" operates when content is indexed or "built" in iPhrase's jargon. In the indexing process, the system automatically creates deployment-specific dictionaries. Automatic dictionaries can be extended through custom grammar and ontologies for use during indexing and query processing. Note: these extensions can use third-party tools or be constructed by human indexers. The point to keep in mind is that the provided knowledge bases and the generated metadata are not automatically controlled. As a result, iPhrase produces different word lists and assigns terms from knowledge bases. Without rigorous management, iPhrase indexes can assign metadata that may mislead, not assist, a user. An example of this is evident in the Neiman Marcus eCommerce screen shot. A query for a product for a man returns hits about products for a woman.

iPhrase also produces statistical data about the content processed by the system. So, an iPhrase index contains statistical outputs, a linguistic representation of text, and a parametric representation of the content processed, automatically generated metadata, and index terms drawn from knowledge bases.

Interfaces

One Step interfaces are customizable. Licensees are able to provide suggestions or "See also" references, lists of users' most frequent queries, bread crumbs hyperlinks so users can back out of a series of clicks, and similar

"guided navigation" features. One Step provides basic interface assembly tools.



The user results are categorized. Notice that short category headings are not provided by the system. The snippets are called "dynamically generated" excerpts. These are brief and do not convey much more than the user's query terms in a portion of a sentence.

Integration

iPhrase integrates with adaptors included with the iPhrase system. Connectors allow the system to process data from relational databases, Web pages, standard office file types, and content management systems. iPhrase includes a content integration application programming interface.

Content processed by the iPhrase system is transformed and stored in a repository as "a unified representation." iPhrase uses the phrase "build pipeline" to describe the transformation and indexing process. When a source document is required, the system regenerates the document from the XML representation in the One Step data warehouse. Each document, Web page, database row, and CMS entry becomes a unique content item inside the One Step repository. From a practical standpoint, the iPhrase system eliminates the need for a source document to be retrieved over the organization's network. However, source data are duplicated. The information remains in the source system and is replicated in the One Step repository.

Application Programming Interfaces

iPhrase supports the major programming languages. In addition, iPhrase provides a packaged ability to integrate with Netegrity Siteminder and Java environments, iPhrase includes SOAP, Java, and.NET support for integration with Web Services, Java Server Pages / Active Server Pages, and a range of application servers. iPhrase includes an entitlements and access Control API, a Netegrity SiteMinder Connector, and Java and Dot Net plug ins.

The system administrator can use the development environment to interact with the concept-based index.Reed Elsevier Deal:

Personalization

One Step offers a personalization module. It allows organizations to display an interface or content to a particular user. When used in self-service customer support applications, One Step generates a personalized self-service view for every individual user based on the intent of each query. With suggested content, a person looking for information can click to see additional information or product alternatives. In an enterprise environment, iPhrase can allow a customer with access to a product catalog to see Web pages that include links to cross-sell and up-sell products or services.

Analytics

iPhrase has put considerable emphasis on capturing and making use of log file data. One Step templates allow the system to present information based on the user's context as coded into the system. Access controls, template preferences, and the type of information accessed are captured and can be used to tune the One Step system.

iPhrase's analytics are one strong point. They underpin the intelligence of the system and provide useful information about the behavior and preferences of users.

The One Step analytics toolkit includes:

- Summary Reports These reports show the number of queries during a specified period broken down according to the type of response users received. These reports summarize how One Step exposed content over a specific time period.
- Usage Intelligence Reports. These reports summarize queries by topic, providing insight into user needs, expressions used in submitting requests, and content provided to resolve them.
- Content Utilization Reports These reports indicate what content is used by those with access to the system.
- Gap Reports These reports provide data to measure unfulfilled requests.

Licensees can generate a profile of aggregated users' needs, which is useful when looking at usage within a department or a customer segment. If feedback has been provided in a self-help One Step deployment, the reports can extract snippets of text from a user's comments. The customer support manager can get a sense of the customer's viewpoint. The One Step system generates log file data that include metadata about every facet of users' interactions, including the key topics and concepts that users requested.

ArnoldIT Opinion

Enterprise search and eCommerce were less mature in the late 1990s than in 2005. iPhrase packaged a remarkable number of systems and methods as a framework. The company leaves no stone unturned when attributing capabilities to its One Step system. However, stripping away the jargon iPhrase is a repository and an index. The natural language processing passes the user's query against metadata that expands the query. The query then is matched to content stored in the repository. The approach is similar to that used by another MIT-infused, Boston-area vendor—Endeca. In fact, the similarities between the two systems are striking; for example, both permit faceted browsing; that is, the system displays hot links of related information. Both systems made sales to companies struggling with online store fronts. Both systems use metadata to deliver more on point results than a search system based on keyword matching.

The features that differentiate iPhrase are its use of a repository to hold transformed content. When a user's query is passed against the iPhrase indexes, snippets of content can be pulled from the repository. When a document is requested by a user, the iPhrase system generates a replica of the source document or populates a report containing the data.

Like other search systems, iPhrase relevance ranks hits that match the user's query. iPhrase is currently focusing on solutions to enhance the customer service and technical support experience as well as improved integration with call center operations. The system has not been widely adopted as an enterprise-wide information retrieval solution. Like Endeca, the iPhrase system performs in an optimal manner when tightly constrained content is managed by the system. The more dissimilar the corpus content, the greater the need for manual intervention in the metadata, controlled term lists, and relationship tables for the processed content.

Table 2: iPhrase One Step Checklist

Attribute	Verity Asserts	ArnoldIT Comment
1 Platform		
2 Keyword search		
3 Text mining	Basic analytics are available	
4 Automated indexing	Yes	Makes use of controlled term lists and discovered metadata. Management of index subsystem requires a subject matter expert
5 Personalization	Yes	The focus is on tailoring content for eCommerce
6 Workflow	Yes	
7 Interface	Graphical	
8 Hosted service	No	A hosted service for Web sites is supported
9 Administrative interface and tools	Graphical and configuration files	
10 Application programming interface	Yes	
11 Professional services	Yes	The system is not designed for a person unfamiliar with iPhrase technology
12 Security	Multiple security options available	iPhrase integrates into the client's security system
13 Connectors	Support for most common file types and data in relational database management systems	SGML and XML are supported by One Step
14 Support for structured data	Yes	
15 Relevance ranking	Yes	
16 Video	Metadata for rich media supported	
17 Federated search	Yes	iPhrase renders content from its repository so mixed content types can appear in a results list or a report
18 Fielded search	Yes	
19 Content crawler	Yes	Content can be pushed and uploaded to an iPhrase server
20 Price	Begins at $500,000	A custom price quote is required. Costs typically exceed seven figures. License fees do not include professional services, customizing the system, or storage and computational resources.

"Until someone's gone through the effort to deploy solutions, it's difficult to gut all that and bring them together, mainly because you have so many different components that have been stitched together."— iPhrase Technologies Senior Vice President of Marketing Tony Frazier. Source: 1to1 Magazine at http://bit.ly/1c7i5nV

Anticipated Benefits

For an organization that is information savvy, iPhrase's framework makes sense for an eCommerce site or for a customer support application.

The iPhrase system administrator can set up iPhrase to adapt results based on the behaviors of the user; namely, what content the user viewed. The iPhrase templates make it easy to present information in a report or display appropriate to a job type like customer support agent or online shopper.

iPhrase addresses some of the challenges presented by indexing structured and unstructured information. The One Step system includes software that permits the iPhrase engine to acquire, transform, and index content. The One Step repository is a partial solution to latency some systems exhibit when displaying content from distributed servers.

The benefits of the iPhrase approach include:

- Permitting users to enter queries without Boolean operators
- Merging structured and unstructured information in a template
- Including connectors for standard file types, database content, and information in enterprise content management systems
- A functional personalization subsystem

iPhrase offers what it calls "pre-packaged best practices." These are procedures and software that allow a specific sequence of tasks to be implemented quickly for a customer support or e-commerce application. iPhrase adds to its "best practices" bundles periodically. However, if the One Step application a licensee requires is not a "best practice", the licensee will have to hook up the iPhrase modules to obtain the functions required.

Possible Drawbacks

iPhrase One Step is a computationally intensive system. What this means is that the licensee will have to provide adequate servers, storage, and network capacity to realize the benefits of the complex iPhrase system.

There are two significant issues associated with the 2005 version of One Step.

First, iPhrase does not provide a suitable metadata management tool. Because iPhrase generates its own indexes and metadata, the tables containing these entries and associated mapping grow quickly in size and number of entries. iPhrase also makes use of licensee-provided knowledge bases. Keeping the metadata organized is a major job that requires subject matter specialists and specialized controlled term and taxonomy management systems. iPhrase signed a partnership with SchemaLogic, a vendor specializing in metadata management. However, the SchemaLogic system is as complex as

the iPhrase system and requires dedicated professionals to maintain the metadata integrity.

Second, iPhrase lacks comprehensive graphical administrative interfaces. Licensees have to edit scripts and configuration files. In addition, the system technical staff will have to modify scripts and write original code. The likelihood of an untrained programmer making an error is extremely high. iPhrase offers professionals services, and these are essential to the successful deployment and optimization of an iPhrase installation.

My view is that the time and expense required to set up an iPhrase system may have been one reason why IBM finds the iPhrase technology attractive. As IBM shifts to services, a complex system provides an on-going revenue stream. iPhrase is resource intensive. IBM has numerous opportunities to sell additional hardware and support services to feed the appetite an iPhrase system has.

Other drawbacks of the iPhrase One Step system include:

- The system requires dedicated technical staff. One Step is not a set it and forget it system
- The new S-PLUS business intelligence system and it uses iPhrase's version of a programming language widely used in the statistics and business intelligence market known as "R". This is a new effort and is, as I write this profile, unproven in the market.
- The cost of the iPhrase system ranges from $500,000 to millions of dollars. With competition for information retrieval solutions increasing, the expense of iPhrase may limit its market appeal.

In short, the key question is, "Can the company support multiple initiatives and maintain the quality of its product and support services?"

Net Net

The iPhrase search system is the first framework to shift the focus from speed and keyword retrieval to user experience. iPhrase is notable because it knits together a stunning line up of jargon as the key to search ROI or return on investment. The payoff from a six or seven figure search system is interesting to contemplate. Because IBM has acquired the company, we will never know if iPhrase delivered on its ROI promise. I am skeptical.

Next-generation systems require significant management commitment, robust technical and financial resources, and a generous, flexible deployment schedule. Will these large-scale solutions fit into an increasingly competitive business environment?

Although great strides are being made in monitoring user behavior and adapting dynamic displays of content to that behavior, it is early days for "intelligent systems." Consequently, a search of the Neiman Marcus Web

site for “men’s sport coat” returns zero hits and illustrations of women’s garments. The Neiman Marcus implementation of iPhrase after several years of operation cannot determine that a query that returns a zero set for “men’s” should display other man-oriented products. Similarly, the Neiman Marcus site cannot automatically correct a misspelling of “neckwear” from the incorrect form “neckwear” to the correct spelling.

iPhrase may be a nearly perfect fit for IBM. As an independent vendor, iPhrase was successful at raising money and luring a number of large clients. Its marketing was as impressive as its fund-raising ability. The firm’s technology, like that of Delphes, embodies many research insights and provides additional evidence that search alone is not enough to generate sustainable organic growth for a company.

Stephen E Arnold
Minor edits to a rough draft on January 7, 2014