
AnswerWorks®

Better Answers Make Better Business

**Natural Language Development System for
Local and Web-Based Help Systems and
Knowledge Bases**

**by Vantage Software Technologies
www.vantagesoftwaretechnologies.com
September 2004**

© Copyright 2004, Vantage Software Technologies. All rights reserved.

Vantage Software Technologies
110 Terry Drive Suite 100
Newtown, PA 18940
215-579-8390
FAX 215-579-8391

www.vantage-software-tech.com

Natural Language Development System for Local and Web-Based Help Systems and Knowledge Bases

Table of Contents

AnswerWorks Value Proposition	3
AnswerWorks Return on Investment.....	8
How AnswerWorks Works.....	10
Recursive Refinement Using AnswerWorks Analytics.....	13
Comparison with Other Technologies	16
Word on the Street.....	17
Who Makes AnswerWorks.....	18

> AnswerWorks Value Proposition

Information is your most valuable asset (...but only if you can find it)

"Knowledge is power."

—Francis Bacon

It was true when Bacon wrote it four centuries ago, and it's even truer today in our information-driven business world: Knowledge is the fuel that powers business — your knowledge of your customers' needs, their knowledge of your products, your employees' problem-solving knowledge and your partners' specialized expertise.

Ready access to this most important business asset benefits everybody. With needed information close at hand and easily found, your customers are happy, your employees are productive, and your business is efficient.

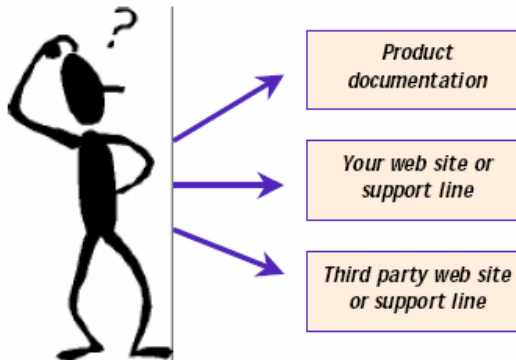
The converse is also true: People who can't find the information they need are unhappy customers and unproductive employees, and your business sputters and stumbles under the weight of increased costs and decreased sales.

Accessible Information = GOOD BUSINESS
Inaccessible Information = BAD BUSINESS

Supercharge Your Business with Better Information Access

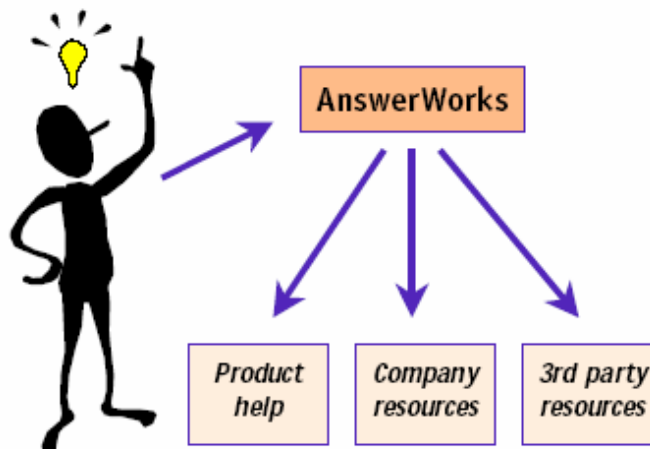
A question-answering engine that works for local, web-based, and hybrid self-help can tip the equation in your favor. Using this engine, people can get the information they need, when they need it — and they can get it by themselves.

How Your Customers Probably Get Help Now



In most companies, the information customers need is divided among several sources, typically including the product's printed documentation or online help, the company's web site or support line, and third party web sites or support lines. Not only is the information split into these separate fiefdoms, but the means to find it may differ at each location: perhaps an index at one, full-text search at another, and a drill-down hierarchy at the third. At best, a customer embarking on this answer odyssey will find it time-consuming and frustrating. At worst, long before finding an answer the customer will abandon the search — and quite possibly your company's products and services.

How a Hybrid Search can Provide Customers with a Better Experience



With a hybrid natural language search, a customer can query any and all of your company's knowledge repositories from a single, intuitive, natural language interface.

- A single search finds all the relevant information, regardless of its source.
- The customer only needs to use one retrieval mechanism.

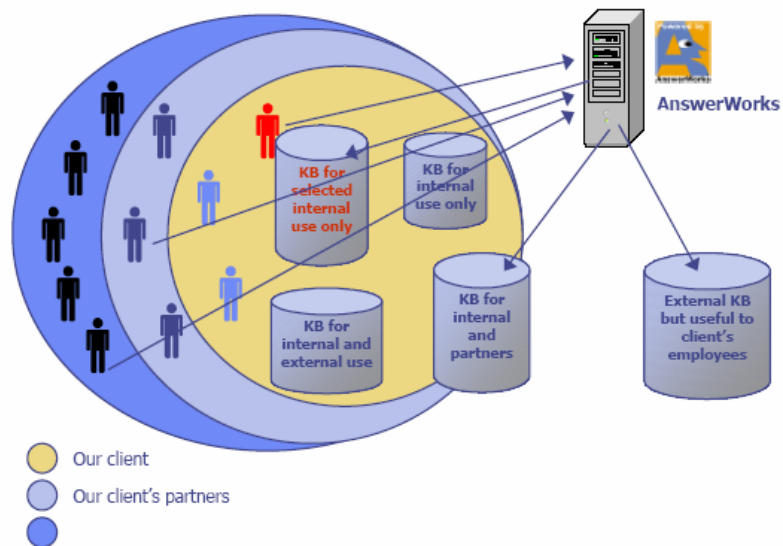


No Boolean logic required!

- The interface is simple and intuitive: The customer does not need to know what terms the documentation uses, understand Boolean logic, discern the organization of the documentation or bring any other particular expertise or experience to bear. Customers simply ask questions using natural language, as they would if speaking to a colleague or a company representative — and with a cross-lingual system they don't even have to ask in the same language as the documentation.

The system can even “know” which product the customer is using and only display relevant answers without the customer having to choose from a menu or otherwise explicitly specify the product.

Other Ways Natural Language Search Can Help



Hybrid natural language search helps more than just your customers. Your employees and business partners also benefit from the search's powerful, easy-to-use access to vital information, even if that information is privileged. A scalable flexible solution should work with knowledge stores that are internal or external, public or private or anything in-between — and in any combination.

How Your Company can Learn More about its Customers

Do you know what issues generate the most frustration for your customers and the most calls for your customer service department? Do you know what parts of your documentation or web site fall short and need to be expanded or revised? Do you know which customers' queries suggest that they need another product or service your company offers?

If your company used a robust question analytics tool, you'd know. A good system provides to-the-point reports detailing exactly what is causing trouble for your customers, pointing the way for your company to improve your

products, documentation, and ultimately customer experience.

What to Look for in a High-Quality Question-Answering and Analytics System

Sure, information is important, but there are plenty of search engines and other information retrieval tools you can use — what makes AnswerWorks so special? Listed here are some of the most significant features that set AnswerWorks apart from other tools. For more details, see “Comparison with other Technologies” on page 17.

<p><i>Stochastic analysis uses statistics and probability in text analysis. AnswerWorks' combination of stochastic and linguistic analysis yields far better results than either alone.</i></p>	Feature	Benefits
	Automatic linguistic and stochastic analysis	Improved customer satisfaction and reduced support costs Rapid deployment and easier maintenance
	Integrated cross-lingual support (eight languages)	Reduced international support costs
	Simultaneous access to multiple local and web-based repositories	Users only need to go to one place to find the answers they need, reducing time and support costs
	Role-based support for employees, partners, customers	Streamlined access to the best answers, saving time and money in resolving problems
	Linguistic normalization analytics	Identifies customer pain points and content gaps
	Enterprise, self-hosted, ASP, Web Service solutions available	One-stop shopping for the solution that's best for your company

Benefits of Adopting AnswerWorks

*"Knowledge is the most democratic source of power."
—Alvin Toffler*

AnswerWorks is the most democratic source of knowledge.

In a democracy, power rests with the people; AnswerWorks provides comparably popular access to the power of your company's critical information. The “populace” — your customers, employees, and partners — enjoy all the advantages that free-flowing information brings. Your company also enjoys benefits, including:

- Better experience for your customers, reducing support costs and increasing loyalty
- Better insight into customer needs
- Global support, not just in English
- Fewer calls to call center and the call center will improve first call resolution

› AnswerWorks Return on Investment (ROI)

Business decisions today aren't just about features and benefits. Any service needs to prove its value with a solid return on investment (ROI). To gauge the ROI of AnswerWorks, consider the following quantifiable and qualifiable metrics.

Quantifiable Metrics

Reduction in call volume. Our clients have told us that they are seeing very significant reductions in call volume after using AnswerWorks. One major client (with over two million users) reports that since implementing AnswerWorks in their Help and Support system, call volume decreased by 35%. How much of this can be attributed to AnswerWorks? Did adopting AnswerWorks result in 25% fewer calls? 20%? 15%?

We suggest taking a very conservative number — 5% — and computing the cost savings associated with that 5% reduction in call volume. We're confident that AnswerWorks will reduce call volume by much more than that, but even with a very conservative number, the return on investment can be significant.

Qualifiable Metrics

The list below includes items with value that are difficult to measure. It is up to you to decide whether these features are worthwhile to your organization, and just how much value they bring.

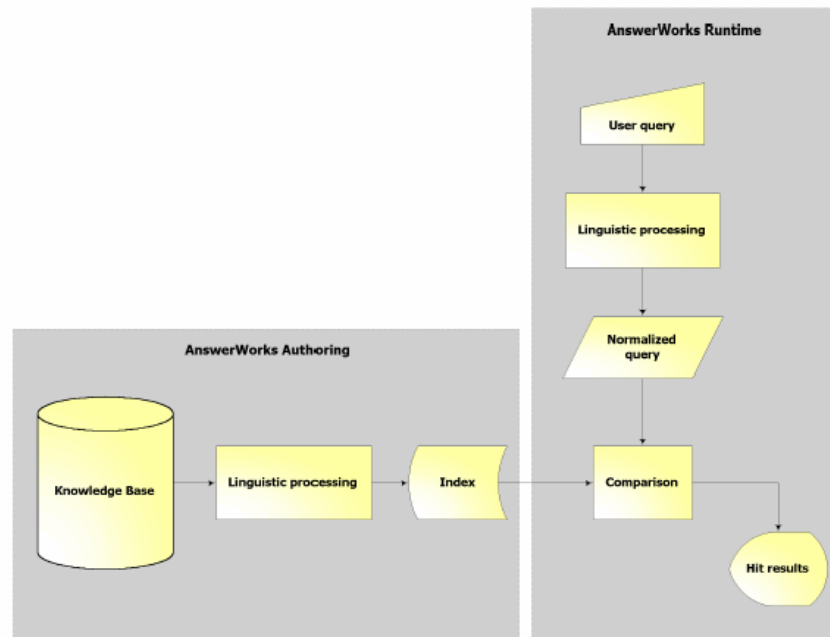
- **Customer satisfaction.** How much value do you attribute to the satisfaction your customers get in being able to answer questions quickly, without having to make a phone call?
- **Employee satisfaction.** Many of our clients use AnswerWorks internally to find information within their help systems and web support pages. We encourage you to consider how valuable this would be to your organization.
- **Value-added.** Consider how valuable it is to be able to tout ubiquitous natural language access to all relevant data within your company's products and services.
- **Multilingual and cross-lingual support.** Consider the benefits of being able to support customers who speak Dutch, French, German, Italian, Japanese, Portuguese and Spanish, as well as English.
- **Understanding the customer.** Our analytics module offers a unique and potent way to really understand customer pain points as well as reveal content gaps. How much is this worth to your organization?

- **Track record.** Vantage Software Technologies has a proven track record of delivering superior technology and service. What's the value of working with a partner who is not just reliable, but offers a level of service and support (and customer empathy) not found with other companies?
- **Ease of implementation.** Consider how valuable it is to deploy our solution "right out of the box" without having to dedicate developer resources.

> How AnswerWorks Works

The AnswerWorks engine actually runs at two different points in the information cycle, although it performs similar duties both times.

AnswerWorks Authoring Process



AnswerWorks authoring and runtime processes

- First the Help author invokes AnswerWorks to analyze the Help file, construct a linguistic image and index of the file, and recompile the file to include the natural language interface.
- Later on, when a user poses a question to the Help file using the natural language interface, AnswerWorks again springs to action, analyzing the question and seeking a match with the information already on hand about the contents of the Help file.

These processes are described in more detail below.

How AnswerWorks Analyzes Help

The AnswerWorks process (Figure 1) begins with the Help author and a completed Help file. AnswerWorks analyzes the Help file to form a *linguistic image* (.LIM) file. This linguistic image is a special kind of index of the Help file, but it has been reduced to the core concepts and facts. AnswerWorks derives these core concepts by parsing the text of the Help file using intricate linguistic rules and algorithms and extensive tables of synonyms. AnswerWorks also takes into consideration additional information found in the topic keywords (K footnotes). While the parsing mechanism used is general, the rules and tables that are used are language-

specific. (Most of these rules and tables are built into AnswerWorks, but there are certain customizations that can be made by the Help author.)

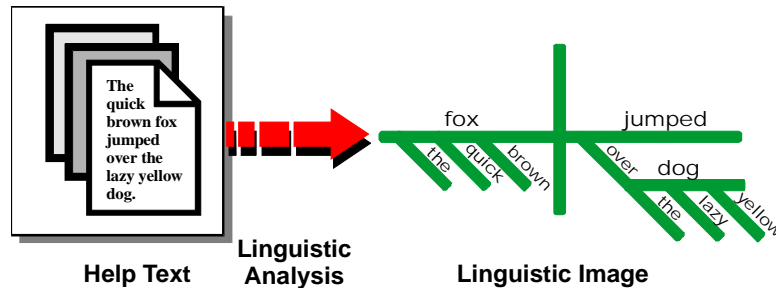


Figure 1. Author's side of the AnswerWorks process

Some of the processes performed include the following:

- The morphology of each word – the changes in word form depending on how it is used in the sentence – is analyzed so that the word can be reduced to a root form. For example, *am*, *are*, *is*, *was*, and *were* are all forms of the verb *to be*, and *geese*, *gander*, and *gosling* are all forms of the noun *goose* (which also has a slang verbal form that we won't get into here). Understanding the morphology of a word is also crucial in determining its role in the sentence. *Play* can be either noun or verb, and so can *plays*, but *played* must be a verb.
- Sentence placement and context is analyzed. For example, the word *the* preceding a word typically indicates use of the word as a noun (or a modifier to a noun). The sentences *I like to play* and *I like the play* have very different meanings just from the change of *to* to *the* (and, of course, the corresponding change in the meaning of *play*).
- If a word has multiple definitions, analysis of a broader context – not just the current sentence, but the paragraph and beyond – may reveal which use is intended.
- Words which are not necessary for the core concepts are filtered out. In some cases, every instance of a common word such as *the* is removed. In other cases, a word is filtered out if it occurs so frequently in a file as to be useless in selecting a particular topic.
- In some cases, words are replaced by synonyms. Consistent use of synonyms improves the speed of searches, and enhances the likelihood of getting desired search results.

After all this, AnswerWorks has a fully parsed, pared-down, pocket index version of the document which it compiles into a .LIM linguistic image file. With this in hand, the Help file is recompiled incorporating the references to the .LIM, as well as other files necessary for the end user to run AnswerWorks.

How AnswerWorks Analyzes Questions

When an end user poses a question to an AnswerWorks-enabled Help file, many of the same processes that were performed on the Help file itself are performed on the user's query:

- The morphology of each word is analyzed, reducing it to a root form.
- A word's sentence placement and context is analyzed.
- Words which are not necessary for the core concepts are filtered out.
- In some cases, words are replaced by synonyms.

The pared-down, parsed query is then compared with the linguistic image (Figure 2). Linguistic analysis is again run to determine the closest matches, which are then listed in qualitative order; that is, with the matches that AnswerWorks considers best listed first. With a well-written Help file and a reasonably well-posed question, accuracy levels above 90% can be obtained.

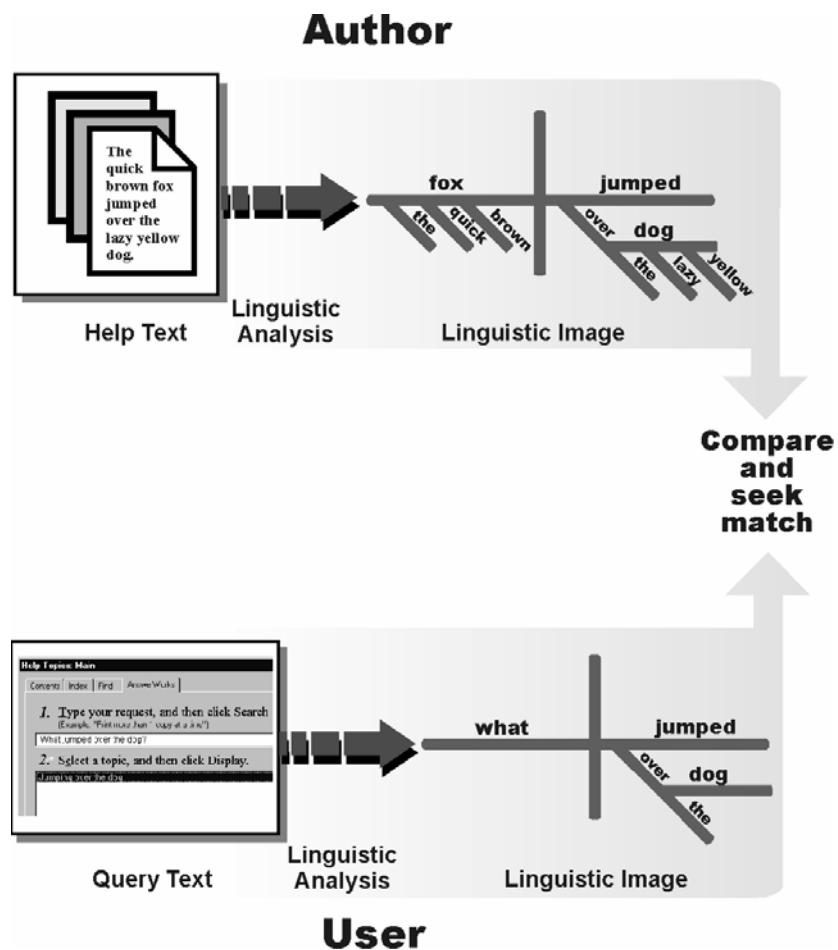
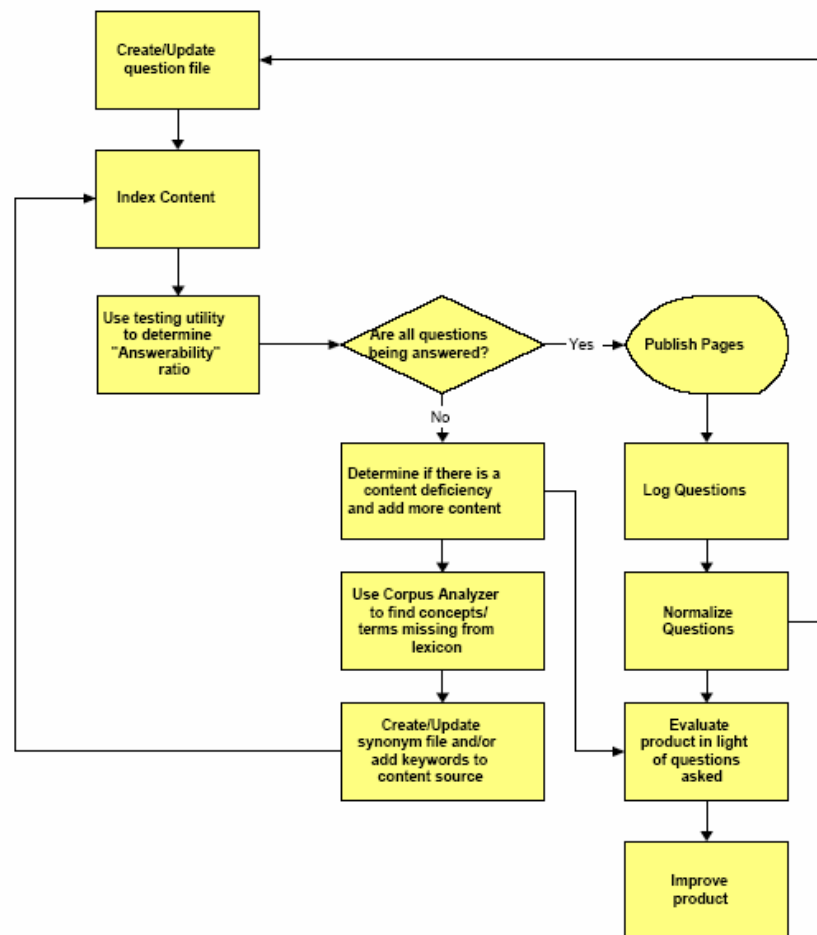


Figure 2. The complete AnswerWorks process.

➤ Recursive Refinement Using AnswerWorks Analytics

The “third half” of the AnswerWorks process is the ongoing improvement of products, information, and the AnswerWorks implementation itself, based on AnswerWorks’ analysis of the questions customers ask, and the answers they find. This is where AnswerWorks goes beyond being just a search engine or information retrieval tool, and becomes a partner in the improvement of your business. AnswerWorks Analytics can identify weaknesses in your products or information so that you can address them, and can also indicate any improvements that may be required by your installation of AnswerWorks.

The following flowchart describes the AnswerWorks development cycle and how questions posed to an “AnswerWorked” site can be used to improve the site and the product.



AnswerWorks development and improvement cycle.

Create or Update a Question File

We recommend that an organization start by assembling the questions that it feels are most likely to be asked by its clients and that produce the greatest number of phone calls (and the largest amount of user frustration). The goal here is to know that AnswerWorks will answer these questions correctly.

Index Content

After the first list of questions is assembled, the content that is supposed to answer these questions should be indexed. In addition, indexing should occur whenever new content is added to the knowledge base. AnswerWorks can be configured so that it automatically indexes every night, or even whenever new information is published.

Testing Utility Determines “Answerability” Ratio

Before going “live” with an AnswerWorked web site or help system, it’s important to make sure that the questions you want to have answered are indeed being answered correctly. Which questions are and are not being answered correctly is readily determined by batch feeding the questions into the AnswerWorks Content Testing and Batch Question Analysis utility.

If Questions Aren’t Being Answered Correctly

If a review of the Content Testing utility results indicates that not enough of the questions are being answered correctly, the first step is to determine if there are gaps in the content. For example, if the question “how do I print sideways” does not yield a good hit, perhaps it is because a discussion of how to print in landscape orientation is nowhere to be found within the Help system or knowledge base.

If the content is indeed present but AnswerWorks is not producing good hits, this suggests there are concepts and terms with which AnswerWorks is unfamiliar. While AnswerWorks has an extensive built-in corporate and technology-based lexicon, it will not recognize certain industry and company specific terms.

Vantage Software Technologies can perform a corpus analysis on content to extract terms with which AnswerWorks is unfamiliar. These terms can then be included in synonym lists or, in some cases, added to the core lexicon to include concepts and terms beyond what can be achieved by adding synonyms.

After adding new content and/or adding synonyms and expanding the lexicon, the content is again indexed and then re-tested to see if questions are now being answered correctly.

If Questions Are Being Answered Correctly

If questions are being answered correctly, then you can take the “AnswerWorked” system online, confident that the questions you have identified as being the most important will be handled properly.

However, you now have the opportunity to learn about your product - and your users – by analyzing the questions they are asking and adding these questions to the Content Testing test suite.

Getting Better, Not Older

Once AnswerWorks is up and running, it's time to go beyond the initial set of expected questions, and see what questions are *actually* being asked and how well they're being answered. AnswerWorks Analytics presents this material in several different reports that highlight what issues cause the biggest problems for your customers, employees, and partners. These reports include:

- **Most frequent questions.** This report normalizes questions – grouping together different phrasings under one umbrella question – showing at a glance what is most on people's minds. In turn, their concerns can point up product flaws, usability and documentation issues, compatibility problems and similar areas requiring your company's attention.
- **Questions that get the worst hits.** Questions that don't turn up a good answer indicate either a gap in content, or a flaw in AnswerWorks (such as unknown jargon). These problems are typically the easiest to resolve; AnswerWorks synonymy is readily customizable; creating and publishing new content is straightforward.
- **Breakdown by product.** Knowing which products generate the most questions enables you to make informed decisions about product R&D, customer service training, service releases, and a host of other business issues.
- **Breakdown by document/Final document viewed.** AnswerWorks lets your customers find the answers they need; this will tell you *where* they found them.
- **Question activity.** Displays the volume of questions asked over time.
- **Word count.** Indicates whether questions are being asked, or if only keywords are being entered. Results are much better with complete sentences instead of keywords, as AnswerWorks is able to perform proper disambiguation and determine the actual intent of the query. It has proven quite simple to train users to input more than just a keyword through a message and sample questions accompanying the results.
- **Returning users.** Shows whether or not users come back to use AnswerWorks again.
- **Action items.** Perhaps the most important of all the AnswerWorks Analytics reports, the Action Items report includes listings of the most frequent issues, content gaps, areas in which AnswerWorks implementation for the site may be improved, and To-Do lists for all parties responsible for improving content and access to content. These reports contain the kind of information that often leads to business building insights.

> Comparison with other Technologies

Natural language versus other search methods

If you've got good information on your web site or in your help system, and you supply an index or full-text search, what do you need AnswerWorks for? Isn't having the information available enough?

In a word, no.

There's no doubt that having the necessary information and making it available is important, and a big step. But people have to be able to find it, and other information access tools just may not be up to the challenge:

- A **table of contents** is terrific for an overview of documentation, but it's unwieldy for finding a particular piece of information – and the less rigidly structured the documentation is, the harder it is to find information solely by using a table of contents. A natural language search, on the other hand, deftly and precisely finds needed information regardless of the documentation's structure (or lack thereof).
- An **index** or **keyword search** is probably the most popular search tool. However, it's only as good as the index, and a good index requires a knowledgeable indexer and plenty of time – and even then, you are still likely to find alternate terms and forms that the index does not include. A natural language search tool's extensive built-in synonymy and morphology tables easily give it a breadth and depth difficult for any conventional index to match.
- **Full text search** works well if you know the precise term you're looking for (“Was it ‘car’ or ‘automobile’?” are not allowed) and if there aren't too many “noise” instances of the term, where it is used in passing. Natural language search easily surpasses full text search with its contextual analysis and automatic handling of synonymy.

> Word on the Street

“Can AnswerWorks natural language question-answering system bring reduced costs and better service to Autodesk’s and Intuit’s help systems and Internet knowledge bases? The answer is yes. AnswerWorks seems to have succeeded in satisfying the users of Autodesk’s AutoCAD and Intuit’s QuickBooks help systems. AnswerWorks has marshaled the combined capabilities of robust natural language-processing technology and state-of-the-art retrieval techniques to enable users to simultaneously search both local and remote resources.”

**— IDC
Speech and Natural Language Bulletin
December 2001**

Autodesk surveyed customers after adding AnswerWorks and found the response very favorable. In the Autodesk survey:

78% felt AnswerWorks was a useful addition to online information retrieval

36% felt AnswerWorks helped them significantly in finding answers

51% would use AnswerWorks as primary search method

Intuit also enjoyed very positive results after adding AnswerWorks to QuickBooks. Intuit noted that AnswerWorks:

- Reduced tech support calls significantly
- Improved user experience
- Proved to be an invaluable research tool for development and support staff
- Exposed redundant content

“This feature makes QuickBooks very friendly. I didn’t search for most items any more than 10-15 seconds.”

— QuickBooks user

> Who makes AnswerWorks?

Vantage Software Technologies provides linguistic-based technologies for spell checking, grammar checking, language recognition, natural language search, desktop tools and self-service solutions which enable our customers to improve productivity, efficiency and organizational performance.

Companies such as Microsoft, AOL, Autodesk, Intuit, Lucent Technology, Sun and other world class organizations have chosen Vantage Software Technologies to help improve the quality of the products and services they provide their customers.