

# COLLECTIVE WISDOM

## CHAPTER 7 EXCERPT

## How Advanced Search Features Work

Understanding the challenges of guiding users to the content they need for service and support, technologists have been hard at work building different models for making the user's interaction with search more intuitive—and, we hope, more likely to result in a good answer. We review the most popular of these advanced features here.

### “Advanced Search”

This feature is by far the oldest of the lot. “Advanced” refers to the user, not to the search capability. When a user performs an advanced search, they're using special terms to control precisely how the retrieval portion of the search algorithm performs. The most important advanced search functions include:

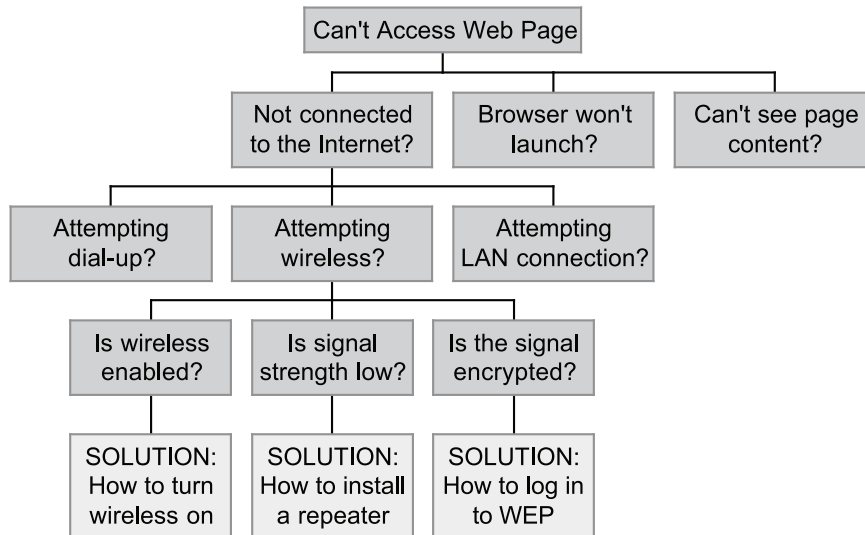
- Boolean operators, named after George Boole, the mathematician who first laid out the principles of the true/false logic that drives modern computers. These include:
  - AND, meaning the query term before and the query term after the AND have to both be in the document or it won't be returned. The query “wireless AND network” will return a document that says “Wireless network channels...” but not the one that says “802.11(b) network channels...”
  - OR, meaning either or both of the query term before and the query term after the OR must be in the document. The query “wireless OR network” returns both documents in the above example.
  - NOT, meaning the query term after the NOT can't be in the document, or it won't be returned. Adding “NOT channels” to the queries above will exclude both documents, since they both contain “channels.”
- Phrases, generally identified with double quotes. A search for “reboot error” in quotes will match “Reboot error occurs after installation...” but not “Reboot device to work around the error.”
- Proximity operators, which are like phrases but less precise. “Application NEAR Installation” would match “Application and driver installation...” Some systems give the searcher finer-grain control over how many words or sentences apart the terms can be.
- Parametric search options, such as pulldowns for version numbers or operating systems, or a date range for when the document was created. These take advantage of structured metadata attached to the content to filter results. (As we said in Chapter 4, use these with caution.)

- Wildcards, such as \*, which can match any number of characters. This lets searchers say 802.11\* which will match 802.11, 802.11(b), 802.11(g), and so on.

*Net-net: Not many users will want to use these advanced search options, but some will, and search technology should respect these experts' desires.*

**Decision Trees**

Some knowledge applications allow users to set up diagnostic scripts or procedures. They are similar to the troubleshooting guides frequently found in the back of manuals for cars or appliances: “Is the power light on? If not, verify that the device is plugged in. If so, go to step 3.” These scripts are called decision trees, because at each point, the answer a user gives to a question takes him down a particular branch. At the end, the tree designer attaches a document that presumably answers the user’s question.



*Figure 7-1: Example decision tree subset. In a real decision tree, each path would be filled out to a set of solution documents at the bottom.*

Decision trees represent a good first step in the process of guiding users to a resolution, rather than just hoping they'll type in smart query strings. And they can work well for issues that are very well defined and come up again and again.

However, decision trees have significant drawbacks in their real-world application.

For one thing, they're a great deal of work to maintain. Rather than "simply" maintaining content, the support organization now has to maintain procedures—in effect, software. Because decision trees can shoot out branches so easily, they can become large and unwieldy. The result of this challenge is that trees often are out of date and don't reflect the latest content about an issue.

Secondly, decision trees can only handle situations that are well defined in advance. For all but the simplest products and services, however, support gets involved when the problems aren't well defined. In these situations, telling the user to check the power cable can be insulting.

**Net-net: For highly repetitive issues in very low-complexity support environments, consider small, easily maintained decision trees. Otherwise, invest your energy elsewhere.**

### Natural Language Processing for Question Intent

Short, common words like "how," "do," "not," "the," and "many" are often stop words: they're completely ignored by keyword search. So, "how do I install the driver," "how many drivers do I install," and "The driver does not install" all are treated as more-or-less the same query, although the user's implied question is very different in each case.

In comes natural language processing (NLP) to the rescue.

NLP algorithms for question intent are designed to interpret the user's intention based on patterns of word usage. For example, "how do I" signals that the user needs "how-to" information, whereas "how many" requires a quantitative response, and "does not" or "can't" is suggestive of a break-fix issue. Based on these patterns, and using other statistical techniques, the ranking algorithm will favor content that appears to provide the answer desired—either from its source, metadata tags, or an analysis of the content itself. While perfect natural language understanding by computers is far off, the results of NLP search in many cases are very impressive.

#### Search Results

**YOU ASKED:**  
How many MPG does a Civic get?

**ANSWERS**

Model	Trim	5-Speed Manual (City/Highway)**	5-Speed Automatic (City/Highway)**
CIVIC COUPE	DX	30/38	30/40
CIVIC COUPE	LX	30/38	30/40
CIVIC COUPE	EX	30/38	30/40

*Figure 7-2: NLP search knows that "how many" questions should be answered with numbers.*

NLP is also commonly used to describe the smarter handling of alternative word usages afforded by advanced knowledge representations (discussed below) and can be used to describe almost any way of processing language (for example, through stemming or spelling correction) that's smarter than literal keyword search.

While NLP search is helpful in some cases, it is by no means a panacea. For one thing, true natural language requests are relatively rare: the majority of queries simply use keywords. For another thing, it's unusual that support questions have a short, simple answer. The fact that an NLP search can provide a direct answer to the question "How many miles per gallon does a Civic get?" on the Honda site is impressive, but there are few analogous questions in support. Using our example above, getting to "driver installation" documents from any of the above questions is a good start.

One additional caveat: almost every search vendor describes their search as NLP, no matter how much of this kind of natural language pattern matching it actually does. If these question intent capabilities are important to you, ask the vendor to demonstrate your idea of NLP functionality on deployed sites.

*Net-net: NLP processing for question intent provides modest benefits to the support search user. Don't let NLP drive your strategy, but if a vendor provides NLP capabilities that don't require much work on your part, it's definitely a nice-to-have.*

### Chatterbots and Virtual Agents

There is a connection that happens when two humans talk, even if one of them is an irate customer and the other is in a far-away support center. Some vendors have advanced the theory that self-service adoption would be increased if people talked with technology that acted more like a human. Thus were virtual agents (derisively known as "chatterbots") born.

Virtual agents are generally accompanied by human-looking faces, or avatars, and use simple natural language processing to attempt to converse with self-service users.

### Widgeteer

<b>CBOT</b>	Tell me about your problem or ask me a question. (First time users)
<b>User</b>	Parallel Strobe crash after installation
<b>CBOT</b>	Are you telling me that you need help installing your 25bit Parallel Strobe?
<b>User</b>	Yes
<b>CBOT</b>	There is a complete installation guide available at the following link <a href="#">Click Here</a> for step by step installation instructions.
<b>CBOT</b>	I'm learning, was this useful?
<b>User</b>	No
<b>CBOT</b>	<a href="#">Click here to escalate</a>
<b>CBOT</b>	Select one of the options below or describe the problem a bit further. <input type="radio"/> 1) You need help installing your Strobe Pro. <input type="radio"/> 2) The scanner "locks up" during initial configuration. <input type="radio"/> 3) You need to uninstall your scanner from Windows 2000. <input type="radio"/> 4) You need the strobetext.dl file. <input type="radio"/> 5) You need to uninstall your scanner from Windows XP. <input type="radio"/> 6) Display all relevant information. Enter your selection (1 through 6 out of 18) or describe your problem a bit further. <input type="radio"/> None of the items in the list are my problem

*Figure 7-3: A virtual agent powered by decision trees.*

Vendors go to great lengths to modulate the affect of the avatar, with happy looks when answers are being delivered, quizzical looks when input is not understood, and so on. Often their interaction is driven by a complex maze of decision trees, as in this example. These two factors make maintenance of chatterbots very labor intensive.