opentext™

Artificial Intelligence for Virtual Assistant

IDOL for Virtual Assistant, is an AI (Artificial Intelligence) powered solution, that allows organizations to offer their customers and employees access to information and processes using an automated human-like operator which can engage in natural language dialogues.

To address an increasingly diverse range of information requests, IDOL uses our unique Natural Language Question Answering (NLQA) technology to allow natural conversations between end users and a virtual assistant, to perform queries against: curated (FAQ) responses (Answer Bank), data tables and data services (Fact Bank), and unstructured documents (Passage Extraction); and to follow configured processes to facilitate knowledge discovery.

Question Answering

The information retrieval process is at first glance well established—a user enters some search terms, a Boolean expression, a document, or a natural language query, and a set of relevant documents are returned—but in many situations this workflow can be optimized further. If a user enters a direct question, for example, then it is generally more appropriate to respond with a direct answer rather than a document or documents that may contain the answer. The technology must be able to understand, process, and answer direct questions to

streamline the retrieval process, which allows information to be obtained in a more convenient and user-friendly fashion. Moreover, it should offer an interface to allow configuration of that process to ensure existing human knowledge is allowed to create and train the system to answer questions optimally. In general, there are three independent steps to accomplish this.

Answer Bank

Many administrators of support or user-help systems have an existing set of frequently asked questions that human support agents are trained on, or the help pages are populated with. For example, if a user encounters a particular problem on his mobile phone, then the manufacturer has established steps that the user should follow to try to correct the problem. The system should be loaded with these answers to provide the best one when a given question is asked or a particular problem is encountered. In addition, the system's natural language processing (NLP) should ensure that differently worded variants of the same questions are all directed to the relevant answer. This knowledge base is generally referred to as the **answer bank**.

Fact Bank

Users frequently have tables of information that are commonly matched via structured (SQL) queries. The system should be able to exploit

such databases of information to allow natural language questions to receive direct factual answers. This is achieved by processing and understanding the question and mapping it to an appropriate structured query that will in turn respond with the desired answer. An example of this would be databases of financial prices over time, allowing a query such as, "What was the EPS of HPQ in Q3 2016?" to receive the precise answer. This database of structure information is generally known as the **fact bank**.

In addition, the system should have the ability to understand and extract information (for example names, numbers) from unstructured data such as free-form documents. It means that databases can be automatically populated with a rich set of structured data from a corpus of unstructured documents. For example, ingesting a set of corporate annual reports could create tables of financial data so that a question such as, "What was BP's revenue in North America in the second half of 2015?" will return the correct value pulled from such a table.

Passage Extraction

In many cases, the information requested is simply not present in either an FAQ data set or a structured database, and an extended approach is required. It is at this point that the system must be able to process human information effectively. Once again, the question should be processed to achieve a basic understanding

of what is being requested and then, the system must form a query against its corpus to find the most relevant documents. The system must process the selected portions of those documents to determine the short passage that it believes most accurately answers the original question. Statistical and probabilistic thresholding is used to determine whether an answer is relevant enough to answer the question, so that if a valid answer simply isn't present in the corpus then none is returned. Once again, the system must have the language-handling capability to allow this process to work in any language.
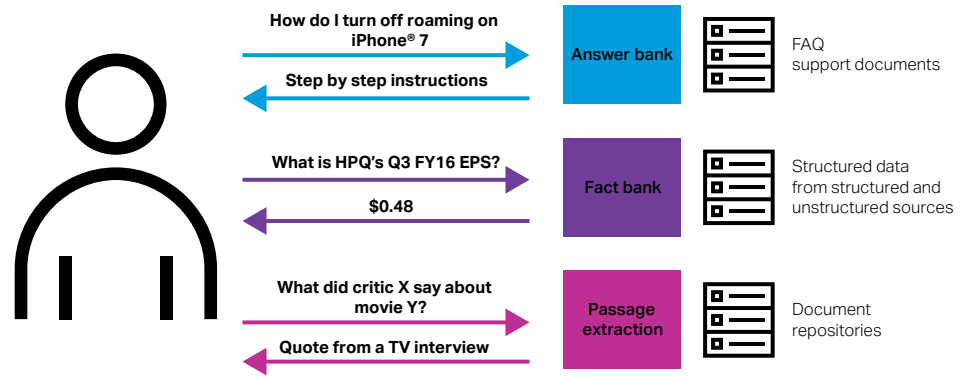
The above steps combine to form a powerful data analytics system for processing questions of any type. The typical workflow is to first check whether the query is appropriate for question processing as many queries are not of this type. If it is, then it matches the question against the answer bank to see if a predefined answer has been set. If none is found, it will then try the fact bank to see whether a precise answer is present. And again, if no answer is found, then it will pass the query to the system's unstructured index to see if an appropriate passage can be found that answers the question.

Figure 1. Question answering in natural language

Dialogue Systems

For many users contacting a company for help with a problem or an inquiry, a search box is insufficient. Even if the system is capable of understanding and processing direct questions, the user is often unable to accurately describe the problem or need or fails to provide enough information. The natural response in such situations is to want to talk to a human operator who wil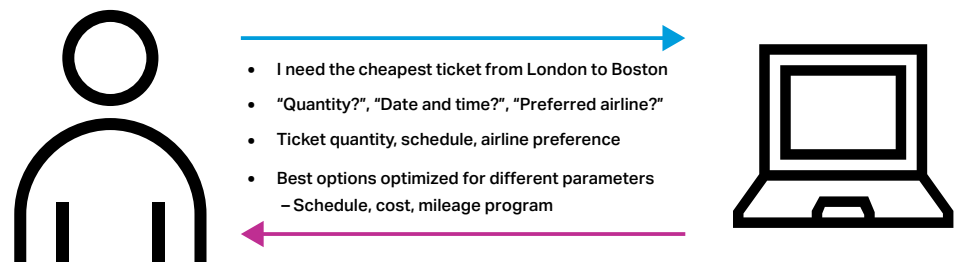l guide them to the correct solution or answer. However, large teams of human operators are costly and clients are often frustrated by having to wait to talk to someone. So, how can we accelerate the right information to the right user in the right context?

A Natural Language Question Answering system can be extended to create a more natural system of customer services. Rather than a single search box, an IM-style interface allows a user to enter an initial question or problem description. In many cases, the problem can be answered directly and a single answer or solution provided, but in many other cases, a conversation between the system and the user is required to gather more information on how to diagnose and solve the problem.

This can be achieved by training the system with classes of problems and the type of information that is needed in each case. The system will then respond with follow-up questions until it believes it has enough knowledge to suggest a solution. Here is a very simple example: user enters, "Cheapest ticket from London to Boston." System asks, "Quantity, date and time, preferred airline?" User responds with answers. System presents best options optimized for

Figure 2. Dialogue system

different parameters: schedule, cost, mileage program.

Simple Curation of Question/ Answer Pairs in Answer Bank

While it is natural to expect a manually intensive process for curating question/answer pairs (FAQs), OpenText™ IDOL comes with an intuitive GUI tool to simplify such a process. This tool allows administrator to easily refine the answers' availability, accuracy and relevancy with a quick clicks.

Connecting Relevant Information to Humans

Technology is only as good as the value realized by its users. With the alluring promise of Big Data, comes some of the biggest barriers to success—user adoption and productivity. True impact will come when humans can truly connect to technology.

To learn more, please visit **www.microfocus. com/idol**
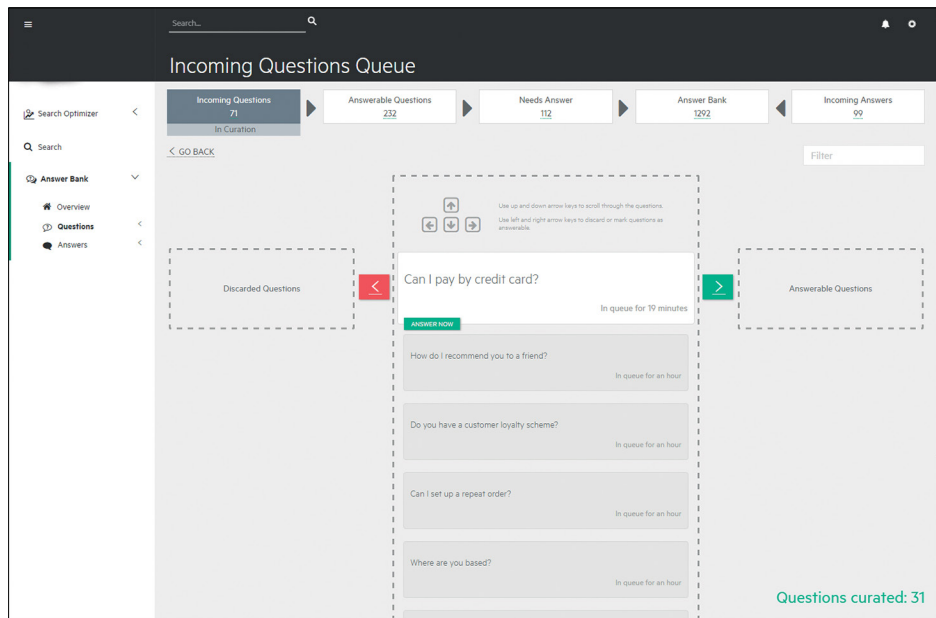
www.microfocus.com/opentext

Figure 3. Question/Answer pairs in answer bank

opentext™