



Exploit the Treasures of Unstructured Big Data

## The Challenge

Patricia has a challenge. As a technical service engineer, she needs to respond quickly to customers' operating problems. Her customer has experienced a sudden drop in process performance. To effectively respond to the customer, she must tap her company's vast expertise quickly. She can either wait to consult with the few remaining experts in her firm who tend to be unavailable, or she can search through large in-house repositories of information herself. These repositories of reports, manuals, presentations, publications, and other unstructured files contain the majority of knowledge and expertise of her company, but they are arduous to explore because they are a large and growing body of *unstructured data*.

If only Patricia could interact with this unstructured data as easily as she would with a human expert to unearth the buried answer that she needs or to find and organize all the scattered data she needs.

The best way to discover, collect, and deploy this intelligence is with cognitive applications. Cognitive applications empower businesses to exploit in-house accumulated experiences and intelligence specific to their industry domain, firm, team, and use-case.

Exploiting your company's treasure of reports, papers, and other forms of text has traditionally been limited to the elementary and unappealing search capabilities of document management applications. The exponentially increasing volumes of data, especially unstructured data, has made it impractical to mine valuable information quickly through conventional document management tools today. While there are many powerful data analysis technologies to tackle structured data designing in databases, there have been limited options to tap the intelligence hidden in volumes of unstructured data, until now.

The advanced capability of cognitive applications can create significant competitive advantages for companies as their employees can make better decisions faster. The capability is an opportunity as well as an emerging competitive dynamic that organizations cannot afford to ignore.

## Meeting the Challenge

Companies can meet the challenge of tapping unstructured data through an advanced text analytics based cognitive application. Cognitive applications allow employees to instantly access buried

intelligence and scattered data from unstructured files with unprecedented ease. This drastically expedites regular workflows as well as unplanned troubleshooting efforts, boosting efficiency while reducing the possibility for error in decision-making.

Cognitive applications can provide this capability because (a) they are able to comprehend natural language and grasp the intent of the body of unstructured files they peruse as well as the intent of a user's question, (b) they can process large volumes of information extremely quickly (many hundreds of thousands of times faster than a human expert), and (c) they can return information to users in valuable formats (responses in natural language or data organized into templates matching user workflows).

A well-known example of a cognitive application is Amazon's Alexa. Users can ask Alexa a wide variety of questions ranging from inquiring about the weather to searching topics on the internet to finding specific products on the Amazon store. This class of cognitive application is called a *virtual advisor*.

Virtual advisors tend to excel when designed to focus on a specific domain. They take the user experience beyond the confines of the conventional menus and structured layers to an intuitive human-like experience useful in quickly answering questions with information contained in large bodies of unstructured data. Such an application can help people like Patricia get their answers readily and to *make better decisions faster*. These virtual advisors are ideal in situations when prior firm intelligence must be tapped to resolve a pressing issue. They become a source of wisdom for the firm – a trusted second opinion, potentially preventing mistakes and rounding out fuller and more comprehensive answers. They effectively raise the expertise of employees by allowing them to exploit accumulated in-house experience and intelligence to make better decisions faster.

A less well-known but equally important class of cognitive application is the *unstructured-to-structured data populator*. Rather than take user questions as inputs, this newer class of cognitive application instead takes in unfilled templates of structured data. For example, a user might provide such a cognitive application with an unfilled data table with marked row and column headers. The cognitive application, using similar processing as the virtual advisor, would then scour a provided body of unstructured textual files to return all the relevant data appropriate to fit each cell of the data table.

Like virtual advisors, data populators tend to excel when designed to focus on a specific domain. They remove then need for the typical tortuous and time-consuming manual scouring of digital files for relevant data. They replace that process with the instant and intuitive process of automatic data population. The user has the option to filter the body of unstructured files that the populator scours, and the user can instantly see the specific passages from which data was drawn to fill in the provided structured data template. This allows the user to quickly control the quality of the drawn data, even across very large bodies of unstructured files. Data populators enable practical data analysis across organizations with large bodies of unstructured files, and they improve existing workflows by drastically reducing the required time to collect and organize data from across many diverse tables and documents.

In many industries, including the energy industry, highly experienced staff are leaving the industry faster than they are succeeded by subordinates, causing an increasing dearth of expertise. A key benefit of both the virtual advisor and the data populator is to capture and retain the knowledge of highly experienced staff and make that knowledge available on demand to employees.

More universally, using the expertise of top performers to enhance the proficiency of others is a benefit of cognitive applications that all industries can exploit.

An additional valuable benefit of cognitive applications is that they can become tools to train less-experienced staff. Since these applications have internalized in-house knowledge, they can very effectively be used to train others with minimum supervision.

To provide the above benefits, cognitive application may be used through direct vocal interaction like Alexa, through a typed chat interface, or through embedding into business applications.

While cognitive applications have the potential to ingest large volumes of varied information sources and provide insight into their content, they need to be specialized to the client domain within the client industry vertical in order to work effectively and provide the greatest value. Therefore, users like Patricia need cognitive applications that understand their specific business domain, make sense of their firms' unstructured files (which are not available to the public), and provide swift and accurate answers for quick and reliable decision-making.

Successful cognitive applications for the enterprise require not only the capability to comprehend and address the complexity of unstructured data but also the capability to consider domain-specific terminologies and nomenclature which vary greatly between verticals such as oil refining, reservoir engineering, healthcare, finance, and other business segments. To deliver significant value, a cognitive application needs to have the following three abilities:

1. The ability to understand all the terms and their nuances of the target domain. For example, for oil refining, the application needs to understand not only the key refining products such as gasoline and its properties, but also the technologies involved in producing it such as fluid catalytic crackers and all the dependencies of their optimal performance. This corpus of knowledge forms the domain specific knowledge foundation, which is the core of the application for a specific vertical – oil refining.
2. The ability to ingest and make sense of large volumes of unstructured information. Sources of information such as reports, technical papers, and communications unique to the client, and which contain proprietary terminologies related to the company's products and their attributes, must be ingestible and processable.
3. The ability to integrate with the client firms' business applications and workflows. Examples of such applications are customer service applications, technical data analysis tools, proposal applications, and others. Input files formats and output data templates must be matched. This integration provides a fully-embedded seamless experience for each client.

Developing a virtual advisor and data populator platform that understands and is primed for your business domain is a challenging and resource-intensive undertaking. Simply treating it as an IT initiative is fraught with risks and pitfalls. Your IT department is likely to contract a big box vendor which appears to have command over the building blocks of cognitive applications but probably does not know or truly comprehend your business domain, which may very well lead to an inadequate application. Simply put, if the cognitive application delivered does not know your business, it is unlikely to correctly interpret the nuances and complexities of your unstructured data.

Given the emphasis on domain knowledge, it would be prudent to select a cognitive application provider with a strong background and experience in your business domain. Such a provider would have a successful track record of delivering applications to leading clients in your domain that handle data, structured or unstructured, like yours. For example, if you are looking for a Virtual Advisor to master the knowledge in your technical reports and papers then the solution provider should have extensive experience with technical data like yours.

Firms interested in cognitive applications need a provider with a well-developed process deployable over a reasonable period without consuming too much of your capital and human resources. Leveraging existing expertise in the solution and in the deployment prevent the inordinately time and cost intensive process of reinventing the wheel to build a new application from scratch.

Another important consideration is the ease of updating and maintaining the cognitive application with new data. New unstructured files are continually created through regular functioning of your business, and these need to be incorporated into the solution regularly. Regular deployment of this capability through internal IT resources is challenging, and doing it on a consulting basis with third party provider is expensive. Cognitive application platforms with built-in tools overcome both challenges, allowing business users to easily maintain the solution and to keep it updated with new data.

### **Introducing the Delfin Platform**

Delfin is a software platform that provides bespoke virtual advisors and unstructured-to-structured data populators to the energy industry. The platform can be rapidly adapted to the client's specific use case, rapidly deployed to create instant value without disrupting existing workflows, and can be maintained to allow continual additions to the referenced body of unstructured files. Once deployed, the cognitive application knows the client team's specific domain, continues to learn over time, and can process large bodies of unstructured files including textual passages of natural language.

How Delfin works:

1. Delfin ingests and learns volumes of the client's unstructured data contained in a variety of documents. Delfin has a pre-built domain knowledge foundation relevant to the industry domain, which gives it a head start. It utilizes this domain knowledge to evaluate an initial set of client documents, and it identifies terms or entities contained that may be of significance to the business. These include product names, properties, related processes, operating conditions, and more. Delfin focuses its custom knowledge learning of the company's information on these entities.
2. The client provides Delfin with a larger set of unstructured files to create a company-specific knowledge model. This knowledge model works in conjunction with Delfin's domain knowledge foundation. The text in the company's documents is processed, analyzed, and then fit to its own custom knowledge model. Numerous machine learning models then build statistical and domain-specific layers in the knowledge model to represent the semantics, intents, and topics contained in the source files. This knowledge model serves as the foundation for Delfin's cognitive applications, which allow business users to quickly unearth the buried knowledge and organize the scattered data they need to be successful.

3. The bespoke cognitive application's capabilities are further sharpened. If necessary, the trainer, a typical business user, prepares a comprehensive list of input examples and presents it to the cognitive application. The application then processes each input example and produces candidate answers or data tables which the user ranks and scores. Delfin uses this feedback to refine the knowledge model to improve the quality of answers produced. This question and answer training is repeated until Sia produces valuable answers reliably.
4. The cognitive application is now ready for use. It can continually ingest new source files generated in the client firm to ensure quality and comprehensive output.

### **Revisiting the Challenge Armed with Delfin**

Returning to Patricia's challenge, she studies her customer's operating data charts in her technical analysis application to investigate their problem. She notices a deviation in a key variable which raises the possibility that it could be the root of her customer's problem.

Her data analysis application can identify other process changes that possibly caused the variable deviation, but she is unsure which changes are at the root of the problem. She would like to know (a) if such a deviation has been reported before for this or any other unit, (b) if so what the experts in her company had concluded about this issue and (c) how it was resolved. Patricia proceeds to ask Delfin's virtual advisor these questions.

The knowledge model powering the virtual advisor covers a wide array of content sources including technical service reports prepared by the company for its customers, domain related literature – internal and public, and communications with experts. It draws upon information across this variety of content sources to answer Patricia's questions.

Patricia identifies an earlier technical service report prepared a few years ago, which addressed the same issue facing her customer. In that report, the technical service engineer had concluded, after much analysis and deliberations with colleagues, that the performance degradation was likely caused by the inclusion of a particularly low-quality component in the feedstock by the customer.

Eager to research the deleterious effect of the suspect feed component on process performance, Patricia asks the virtual advisor to search for relevant literature references. The advisor identifies a publication that discusses the negative effects of similar feedstock components on process performance. Armed with this information, Patricia quickly responds to her customer with the recommendation to minimize the suspect component in the feedstock to resolve their issue. She accomplishes all of this in minutes instead of hours or days, and she does so with a deeper technical depth than was otherwise practically possible.

### **Why DelfinSia, Inc.**

Founded by experienced software developers and oil and gas professionals, DelfinSia, Inc. offers a unique blend of industry domain knowledge, data science expertise, and software application development skill. Our founders have over 80 years of combined experience delivering strategic web and mobile applications and services involving technical data, technical applications, and complex

industry-specific technical considerations. Our founders are also thought leaders in the data science and natural language processing domains. With its skills and experience, DelfinSia, Inc. has leveraged the latest developments in machine learning and text analytics to develop the Delfin platform.

## **Conclusion**

The emerging technology of cognitive applications based on natural language processing has resulted in exciting new classes of products called virtual advisors and data populators, both demonstrating enormous value-added capacity. These cognitive applications give businesses the potential to access and exploit information and expertise locked away in volumes of unstructured files. To realize this potential, DelfinSia, Inc. offers its Delfin platform to seamlessly handle client unstructured files and interface with client business users. The Delfin platform optimizes workflows, reduces redundant work efforts, frees up valuable subject matter expert time, allows smaller teams to accomplish more, reduces the likelihood of error, and enables consideration of additional valuable data in the decision-making process. DelfinSia, Inc. drives delivers this superior platform-as-a-service in the fastest time and at the cheapest price, yielding the highest-value results.

For more information visit [delfinSia.ai](https://www.delfinsia.ai).