

Exploit the treasures of unstructured big data – The Delfin Way

A White Paper by Astron International, Inc.

The Challenge

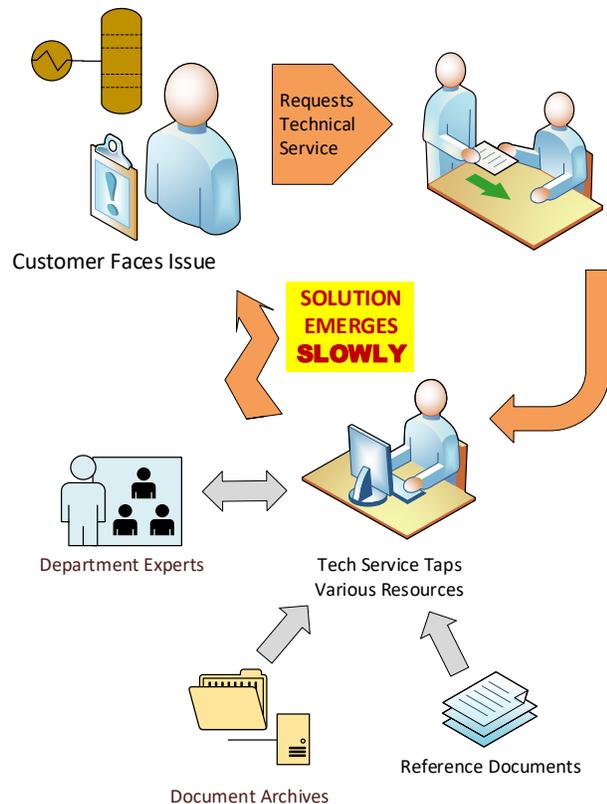
Steve has a challenge. As a technical service engineer he needs to respond swiftly to a customer's operating problem. His customer has experienced a sudden drop in process performance. To effectively respond to his customer, Steve must tap his company's vast expertise quickly. He can either consult with experts in the company who are few and not always available or search large in-house repositories of information. Most of the knowledge and expertise of his company resides in hundreds of internal reports, papers and various other documents collectively called *unstructured data*.

If only Steve can interact with this unstructured data as easily as he would with a human expert and unearth the answer he wants.

The best way to unearth, use and deploy this buried expertise within the company is with cognitive applications. This empowers businesses to exploit in-house accumulated experiences and intelligence specific to their industry domain and organization.

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. While there are many powerful data analysis technologies to tackle, structured data residing in databases, there were limited options until recently to tap the expertise hidden in volumes of unstructured data such as documents.

This advanced capability of cognitive applications can create a significant competitive advantage for companies as their employees can make better decisions faster. It is an opportunity as well as an emerging competitive dynamic one cannot 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 interact with computers as easily as they interact with each other to gain insight.

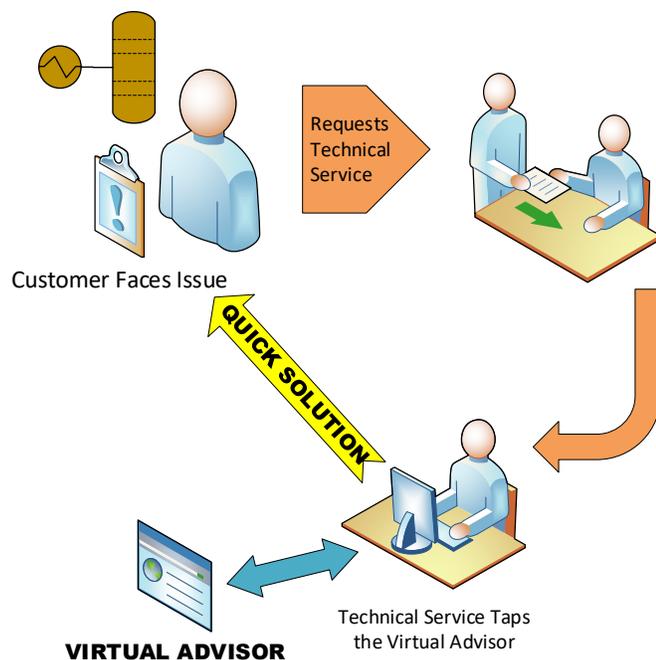
Cognitive applications can provide this capability because (a) they are able to comprehend natural language and grasp the intent of the user's question, (b) quickly go through large volumes of unstructured data and (c) provide a response back to the user in natural language.

A well-known example of a cognitive application is iPhone's Siri. Users can ask Siri a wide variety of questions ranging from inquiring about the weather to searching topics on the public Web to finding features in their iPhone device. Siri appears to know it all!

Cognitive applications can bring similar advisory capability and ease to help your employees exploit buried expertise of your company and make better decisions faster. Such advisory applications form a new class of products called *Virtual Advisors*.

The benefits of a Virtual Advisor are many with many more yet to be identified.

Virtual Advisors take the user experience beyond the confines of conventional menus and structured layers to an intuitive human like experience to answer questions quickly from volumes of unstructured data. Such an application can help people like Steve get their answers readily and *make better decisions faster*. It becomes a source of wisdom for the business - a trusted source for a second opinion, potentially preventing mistakes. Effectively it raises the expertise of employees by allowing them to exploit in-house accumulated experiences and intelligence to make better decisions faster.



In many industries, such as oil & gas, highly experienced staff are retiring and leaving the industry faster than they are succeeded by subordinates causing an increasing dearth of experiential expertise. A key benefit of a Virtual Advisor is to capture and retain the knowledge of highly experienced staff and make it available on demand to employees.

More universally using the expertise of top performers to enhance the proficiency of others is a benefit of Virtual Advisors all industries can exploit.

An additional valuable benefit of Virtual Advisors 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.

Virtual Advisors may be used through direct interaction like Siri or a chat interface or embedded in business applications to provide the above benefits.

While Virtual Advisors have the potential to ingest large volumes of varied information sources and provide insight into their content, they need to be specialized to your domain or industry vertical to work effectively and provide the greatest value. Therefore, users like Steve need Virtual Advisors that understand his specific business domain, make sense of his company's unstructured data, much of which may be proprietary, and give swift valuable answers for quick and reliable decision-making.

Hence, employees need Virtual Advisors that understand them and their business. This includes not only comprehending and addressing the complexity of unstructured data but also the domain specific terminologies and nomenclature which vary greatly between verticals such as oil refining, healthcare, finance, and other business segments. To deliver significant value a Virtual Advisor needs to have the following three characteristics:

1. *Understand all the terms and their nuances of your domain.* For example, for oil refining, the application needs to understand all the key refining terms such as gasoline, its properties, the technologies involved in producing it such as FCC, and all other dependencies. This corpus of knowledge forms the domain specific knowledge foundation, which is the core of the application, for a specific vertical – oil refining.
2. *Ability to ingest and make sense of large volumes of varied information.* Sources of information such as reports, technical papers, and communications unique to your company, and which contain proprietary terminologies related to the company's products and their attributes.
3. *Integrate with the company's business applications.* Examples of such applications are customer service applications, technical data analysis tools, proposal applications, and others. This integration provides a fully-embedded seamless experience.

Developing a Virtual Advisor 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 Virtual Advisor 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 Virtual Advisor provider with a strong background and experience in your business domain. Such a provider

would have a long and successful track record of delivering applications to 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.

Another important consideration is the ease of updating and maintaining the Virtual Advisor with new data. Since new data are continually created these need to be deployed to the Virtual Advisor regularly. The question to ask is - does this effort require IT resources or can it be done by average business users? Virtual Advisors that have built-in tools to allow its business users to easily maintain and keep it updated with new data are time and cost efficient.

Last but not the least you need a provider with a well-developed solution and process to deliver a Virtual Advisor for your business domain in a reasonable period and without consuming too much of your capital and human resources. This is so you don't reinvent the wheel to build a Virtual Advisor from scratch taking inordinate amount of time and cost.

Introducing *Delfin*[™] – The *Sia*[™] Virtual Advisor Creator

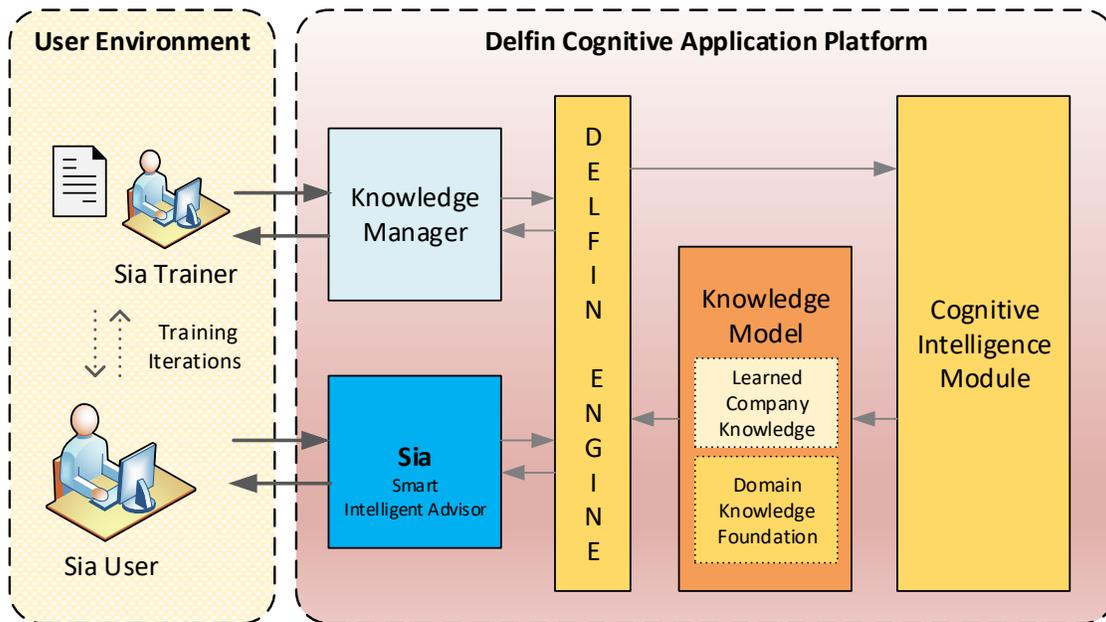
Delfin is a software platform from Astron International, Inc. that delivers the promise of Virtual Advisors by providing a framework of tools to create your Virtual Advisor called *Sia*, short for *Smart Intelligent Advisor*[™]. Once created *Sia* knows your company's domain, its proprietary information, learns over time, and understands questions and provides answers in natural language.

How *Delfin* works.

1. *Delfin ingests and learns* volumes of the company's unstructured data contained in a variety of documents. *Delfin* has a *pre-built domain knowledge foundation* relevant to your company's industry domain which gives it a head start. It utilizes this domain knowledge to evaluate an initial set of your company's documents and identifies terms or entities contained that may be of significance to the business such as product names, properties, related processes, operating conditions and more. *Delfin* focuses its custom knowledge learning of the company's information on these entities.
2. *User feeds and Delfin ingests* a larger set of the company's documents to create a company specific knowledge model based on the ingested documents. This knowledge model works in conjunction with *Delfin's* domain knowledge foundation. The text in the company's documents are processed and analyzed, 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 of the documents.* This knowledge model then serves as the foundation for *Delfin's* Virtual Advisor, *Sia*, which allows business users to quickly unearth the knowledge they need to be successful.
3. *Sia is driven by the Delfin created knowledge model.* *Sia* is the smart natural language interface for users to ask questions to and receive answers from the knowledge model.

4. *Sia's answering capability is sharpened to provide reliable answers.* The trainer, a typical business user, prepares a comprehensive list of questions and presents them to *Sia*. *Sia* processes each question and produces candidate answers 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.
5. *Sia is now ready for use.* New documents continually generated in your company can be used to keep the knowledge model and *Sia* up to date using *Delfin*.

A schematic of the *Delfin* platform with *Sia* is presented below.



Applications of *Sia*

Sia is the interface through which the user consumes the knowledge extracted from the documents loaded into the *Delfin* knowledge model. It offers multiple options for the user to exploit the knowledge model such as:

- a. Chat Interface – The user submits a question in natural language in a text box and receives a response.
- b. Voice Capability – The user speaks - asks questions and receives an audible response.
- c. Application Integration –
 - i. A more sophisticated way to consume the Knowledge Model is to integrate or embed *Sia* into the company's business applications, such as a technical or customer service application, where *Sia* provides valuable insight into issues and enables the user to make better decisions faster.

- ii. Add data analytics capabilities to the Knowledge Model from *Delfin* to increase *Sia*'s scope to cover numeric data in addition to textual information.

Revisiting the Challenge Armed with *Sia*

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

Steve's data analysis application can identify other process changes that possibly caused the variable deviation but he is unsure which changes are at the root of the problem. He 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 his company had concluded about this issue and (c) how it was resolved. Steve proceeds to ask *Sia* these questions.

The knowledge model powering *Sia* 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. *Sia* draws upon information across this variety of content sources to answer Steve's questions.

Sia identifies an earlier technical service report prepared a few years ago which addressed the same issue facing Steve's 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 Steve asks *Sia* to search for relevant literature references. *Sia* identifies a publication that discusses the negative effects of similar feedstock components on process performance. Armed with this information Steve quickly responds to his customer with the recommendation to minimize the suspect component in the feedstock to resolve their issue. Steve accomplishes all of this in minutes instead of hours or days with a technical depth deeper than was otherwise easily possible.

Why Astron International, Inc.

Founded by experienced chemical engineers, Astron International offers a unique blend of refining industry domain knowledge, business process expertise, and proven application development skills. The company has over 20 years of experience delivering strategic Web based software applications and services involving technical data, technical applications, and complex industry specific business rules to the petroleum refining and chemical industries. Astron International has also delivered mobile applications and solutions for smartphones and tablets to these industries.

With its skills and experience Astron International, Inc. has leveraged the latest developments in machine learning and text analytic to develop *Delfin* and deliver *Sia* for industrial applications.

Conclusion

The emerging technology of cognitive applications based on natural language has resulted in an exciting new class of products called Virtual Advisors with huge potential. Virtual Advisors give businesses the potential to access and exploit information and expertise locked in volumes of unstructured data. To realize this potential Astron International, Inc. offers its *Delfin* platform to create a virtual advisor called *Sia* that knows the customer's unstructured data and interfaces with its business users through natural language.

For more information visit delfinSia.com.