

**Text Analytics and Smart Business Process Applications:
Empowering knowledge professionals in the 21st century**

WHITE PAPER

ABBYY®

“Knowledge worker productivity is the biggest challenge of the 21st Century”

-Peter Drucker

Executive Summary

The biggest challenge facing organizations today is how to increase the productivity of their “knowledge workers.” The primary job of knowledge workers is to combine information and their own knowledge to accomplish critical business goals. Knowledge work requires making decisions under tremendous time pressures. It is rarely a dearth of information that holds back knowledge worker productivity, but rather information overload – they just don’t have the time to find the facts, events and stories they need to achieve real insight and action information.

Because of the criticality of knowledge work, a technology stack to handle its content, processes and workflows has grown organically over the years. At the base of the stack, Enterprise Content Management (ECM) systems provide the storage platform for the collection of documents and data used by knowledge workers to manage the lifecycle of documents. The next layer, Business Process Management (BPM) platforms meet the need for predictable, repeatable processes to achieve business objectives. Next, Case Management Frameworks (CMFs) provide business domain solution components that help organize knowledge work into “cases.” The topmost and newest layer is comprised of Smart Process Applications (SPAs) to actualize the CMF process, connecting content to cases to decisions within industry-specific business activities.

The knowledge work technology stack provides a complete IT infrastructure. But nothing more. It provides the context for productivity, but not the missing key to productivity. For knowledge workers to action information and increase productivity they must have insight and intelligence into “case” information to drive the process and make decisions. Language is the key to insight. All of our documents, our communications, the information upon which knowledge workers make their decisions is based in language. To fully optimize the knowledge worker stack we need to plug in technology that can understand language the same way that we do.

ABBYY’s Compreno technology can empower knowledge workers within even the most advanced ECM/BPM/CMF/SPA technology stack in any content-intensive, knowledge-intensive environment. Compreno “understands” and “dissects” language within content, and uses this understanding to provide comprehensive text analytics that automatically and accurately identify entities, facts and events in text to reveal the stories within. Compreno drives efficiencies throughout the stack, and it is at the top of the stack, with SPAs, that Compreno is most critically necessary. Compreno provides the intelligence and insight needed to make Smart Process Applications truly “smart” by providing the right information – exactly, to accelerate and improve results.

Text Analytics and Smart Business Process Applications

Empowering knowledge professionals in the 21st century

Contents

Executive Summary

4 The Way We Work Today

5 The Knowledge Work Technology Stack

7 Getting Smart with Text Analytics

8 How Knowledge Workers can Action their Information

11 Conclusion

The Way We Work Today

First, knowledge work is goal-oriented, not process-determinative

Second, knowledge work requires making decisions under demanding, if not crushing, time pressure

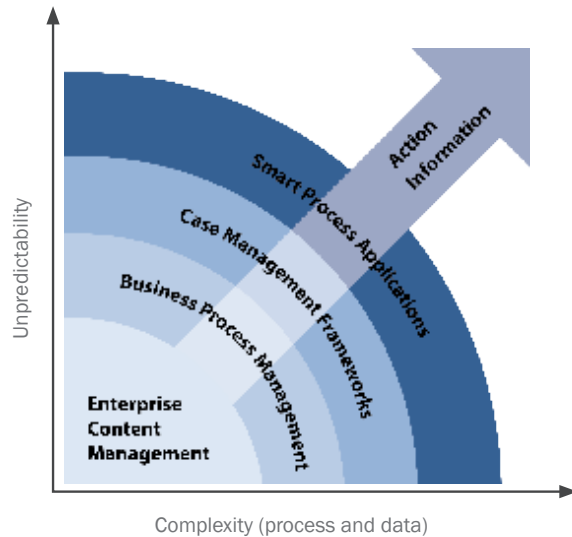
Since Peter Drucker coined the phrase “knowledge worker” in 1964 the number of those who are employed to produce or analyze ideas or information has grown to an estimated 200 million worldwide¹. Knowledge workers are found in many of the most critical roles in organizations today. Pioneering business process author Thomas Davenport gives us some examples:

- Management
- Business and financial operations
- Computer and mathematical
- Architecture and engineering
- Life, physical, and social scientists
- Legal
- Healthcare practitioners
- Community and social services²

Davenport’s list is broad and diverse, and at first there seems to be little in common between these examples. Yet, there are two critical commonalities. First, knowledge work is goal-oriented, not process-determinative. According to Keith Swenson, VP of Research and Development at Fujitsu America, knowledge workers are “People who have a high degree of expertise, education, or experience and the primary purpose of their job involves achieving goals in a way optimized by that knowledge.”³ Second, knowledge work requires making decisions under demanding, if not crushing, time pressure: the business leaders who needs to make a decision quickly before a window of opportunity closes, the healthcare first-responder who needs to decide upon a treatment before the patient worsens, and the lawyer who has to advise a client before a legal situation deteriorates. We live in a world of information overload, and so while knowledge workers need information, they rarely lack for it. What they need instead is the right information – exactly: the facts, events and stories to make the best decision in the shortest possible time.

Considering the central role of knowledge workers it becomes clear that organizations must provide them with the core requirements they need to perform their work. At base, this means a flexible, information-centric work environment. Yet, to meet the biggest challenge of the 21st century, organizations must go beyond document-centric processes and provide systems that adapt – not just to varying types of work, but to work that varies over time and context – so that knowledge workers can focus on their end goals.

The Knowledge Work Technology Stack



Software evolution to meet the demands of increasingly complex processes and information

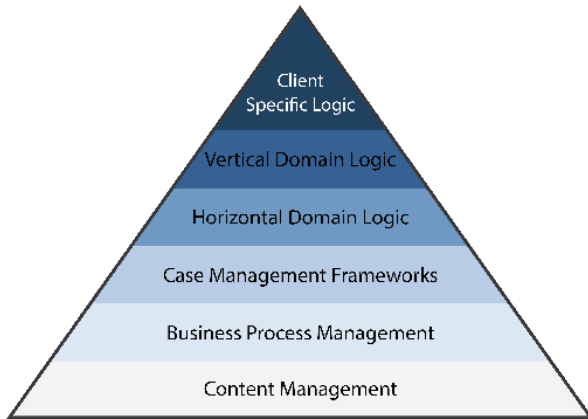
The key to understanding the needs of knowledge workers is grasping how they use, generate and manage information. To accomplish business objectives, knowledge workers must identify, weigh and choose critical information before they can reach necessary decision points.

Enterprise Content Management (ECM) systems adopted by many organizations in the last few decades provided a storage platform for the collection of documents and data used by knowledge workers. However, ECM systems cannot structure content, whether in terms of classification or the intelligent extraction of entities, attributes and relationships, except through labor-intensive manual means. And because they are document-centric ECM systems largely lack any process or workflow capabilities beyond providing document lifecycle management.

The need for predictable, repeatable processes to achieve business objectives led to the creation and adoption of Business Process Management (BPM) platforms as a layer above ECM. BPM platforms provide a way for organizations to measure and control operational processes as well as apply IT solutions directly to those processes. Organizations gain visibility into the ongoing work processes through scheduling, tracking and monitoring. The key to these benefits is rigorous documentation and planning. Because this requires a strong degree of predictability the vast majority of business processes that are modeled in BPM systems are linear decision trees used to guide the action and decision of workers.

Unfortunately, while the objectives of knowledge workers are predictable, the processes that they use in accomplishing those goals are not. The linear, pre-defined processes created by BPM systems lack the flexibility needed by goal-oriented knowledge workers whose process is typically undefined at the start or even emerges entirely through the work project itself.

The key to understanding the needs of knowledge workers is grasping how they use, generate and manage information. To accomplish business objectives knowledge workers must identify, weigh and choose critical information before they can reach necessary decision points, communicate that information with others, make decisions, and then generate content deliverables. Pre-IT era solutions involved storing all of this in a centralized location, a “case,” “folder” or “file.” In the digital world this has morphed into the virtualized form of case management. As case management pioneer Michael White describes it:



Smart Process Application technology pyramid

Case Management is the management of long-lived collaborative processes that coordinate knowledge, content, correspondence and resources to progress a case to achieve a particular goal; where the path of execution cannot be predetermined in advance of execution; where human judgment is required to determine how the end goal can be achieved; and where the state of a case can be altered by external out-of-band events.⁴

Case Management Frameworks (CMFs) provide an out-of-the box architectural pattern and business domain solution components to sit above content-centric applications such as ECM and BPM in the knowledge work technology stack. CMFs are by design partial solutions used to jump-start creation of a non-linear, unique, and customized solution using plug-in tools to provide a best-of-breed solution. Organizations can use CMF components to create flexible processes to match their business objectives.

Smart Process Applications (SPAs) provide the topmost level in the stack by actualizing the CMF process within industry-specific business activities and connecting content to cases. SPAs support processes that are people-intensive, highly variable, loosely structured, and subject to frequent change – all aspects associated with knowledge workers. SPAs bring together:

- Content capture, output and management;
- BPM tools for effectively executing upon tasks, visibility and accountability;
- Content-rich structuring and search tools to allow knowledge workers to educate themselves on the case information, whether from the beginning, end or middle of the matter;
- Embedded analytical tools to support efficient and effective decision-making; and
- Collaboration capabilities for creating content deliverables.⁵

Yet even with these all of these advances, the 21st Century challenge of knowledge worker productivity is still not yet fully met. ECM systems can provide access to content. BPM platforms can provide repeatable workflow, and CMFs can extend this to decision-intensive, non-linear knowledge work. SPAs bring us further ahead with their ability to organize tasks and the complex, even unique, interactions between them into an adaptable workflow. From an IT operational perspective the stack seems complete, but for knowledge workers it is still far from fully optimized: information is still overwhelming, time demands are still crushing and it is too hard and manually-intensive for them to analyze “case” documents and uncover the hidden fact patterns. **Something is missing from the stack: intelligence and insight.**

Process	Workflow	Content
Goal-Oriented	Solves a repeatable situation	Cases do not die
Decision-intensive	Non-Linear, non-pre-determined workflow	Rich content and knowledge intensive
Uses process fragments or tasks	Unique complex interactions	Analytics for insight and patterns

Key characteristics of Smart Process Applications

Getting Smart with Text Analytics

Language is the key to insight. Our documents, our communications, the information upon which knowledge workers make their decisions is based on language.

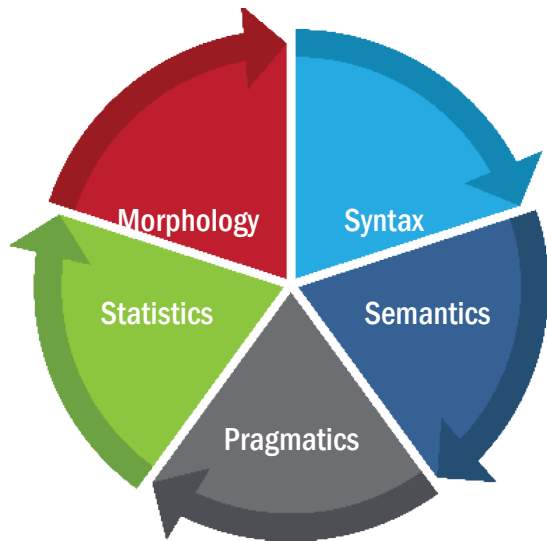
Language is the key to insight. Our documents, our communications, the information upon which knowledge workers make their decisions is based on language. As explained by Allim:

Language gives form to and enables us to express and communicate knowledge. All semantic analysis technologies begin by working with language because thoughts and ideas are formally expressed using language.⁷

Semantic analysis technology to leverage language is not new. Latent Semantic Analysis (LSA) and Support Vector Machines (SVM), both available for many years, use the number-crunching power of computers to manufacture a mathematical index of linguistic content using statistics. However, LSA and SVM provide a mathematical interpretation of language – they can tell you what words are on a particular page, where they are located and what other words often appear near them in context – but they cannot understand what those words actually mean, and how they ultimately describe facts, events and stories. LSA and SVM can support the predictable, repeatable processes contained within ECM and BPM at the bottom of the knowledge work stack. But when we want to action information within unpredictable processes at the level of CMFs and SPAs the mathematical modeling of language shows its limitations.

In summary, knowledge workers need context. We understand information through facts and events, not through statistical correlations. We understand tasks as they fit within a story, not through boxes and lines on a workflow diagram. We analyze and uncover patterns in those facts and events and build them into stories to find intelligence and insight and to make decisions. We do this all through language. To fully optimize the knowledge worker technology stack we need to plug in technology that can understand language in the way that we do.

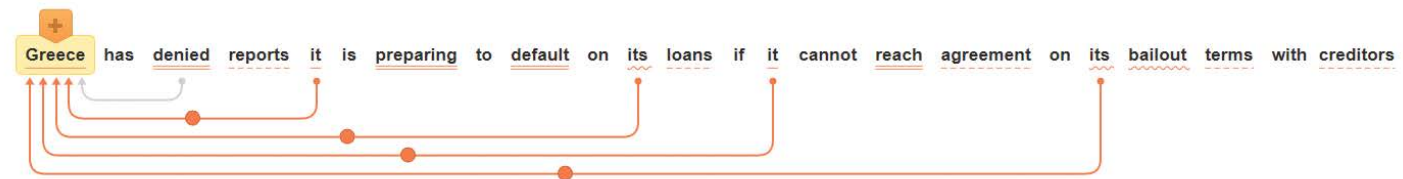
How Knowledge Workers can Action their Information



ABBYY's Compreno language-based technology provides comprehensive text analytics

Information is rooted in language. ABBYY provides customers with a unified language-based computing approach to text information that addresses and captures both the physical representation of language and the meaning in language. Powered by deep syntactic, semantic and ontological analysis, ABBYY's **Compreno technology "understands" the meaning of words and reveals the relationships between them within content.**

Compreno uses this understanding to provide comprehensive text analytics that automatically and accurately identify facts and events to understand the stories within textual documents. Compreno can do this regardless of document complexity and word location within the documents, or the shortcuts that we use in natural language. Compreno detects relationships among words in a sentence and between sentences, including relations between words that stand far apart, that refer back to an already mentioned object, (e.g. by using pronouns – known as "anaphora") that use other words (known as "co-reference") or that have been omitted from a sentence (known as "ellipses"). Compreno can extract the meaning needed to empower knowledge workers within domain specific vocabularies and linguistic usage, without the need to construct complex rule sets or large statistical training sets.



ABBYY InfoExtractor uses the power of Compreno to identify and extract events and facts between related entities, with a speed and comprehensiveness that no human can match, to reconstruct narratives and uncover fact-patterns within and across documents in a case. And it does all of this

“Compreno’s ability to compare documents on different levels can be very powerful.”

Keith Swenson
-Fujitsu

Core capabilities of ABBYY InfoExtractor



Accurately extract entities, events, facts



Identify relationships between entities & events



Add customized entities for specific cases



Use custom ontologies for industry solutions



Work with text regardless of source

automatically, embedded in enterprise systems to analyze incoming content in the background after it is captured, to accelerate knowledge workers’ time-to-discovery and time-to-decision. In short, ABBYY InfoExtractor provides knowledge workers with the intelligence and insight that they need to do their work more efficiently.

ABBYY InfoExtractor infuses insight into the knowledge work technology stack from the bottom up. Using a simple REST API, ABBYY InfoExtractor integrates easily with ECM systems at the base of the technology stack. Compreno’s deep syntactic, semantic and ontological analysis of unstructured text found across enterprise data stores, email communication and form-based processes enables knowledge workers to discover, filter and action information. Most importantly at the highest level of the stack, SPAs have tremendous power to organize tasks within the CMF workflows, but without the proper awareness of case-based context that power can be wasted through false positives, near misses and unnecessary tasks.

Smart Process Applications need to provide knowledge workers with the right information, in the right tasks at the right time – exactly, or they will become just more noise that highly skilled professionals will work around. That level of precision to deliver the right information can only come from a thorough understanding of the content so that it can be applied to the right task. ABBYY InfoExtractor leverages Compreno’s understanding of the content to provide the insight and intelligence needed for that level of precision.

This is where the power of Compreno’s natural language processing technology is critical. ABBYY InfoExtractor provides the needed understanding for SPAs from the content itself. **ABBYY InfoExtractor structures the unstructured data, so that your knowledge workers don’t have to. In this way, ABBYY InfoExtractor drives efficiencies and enterprise value.** Let’s look at an example, as we discussed with Keith Swenson from Fujitsu:

As anyone who has had to deal with a serious medical issue knows, the healthcare industry runs on information, lots and lots of information. All of that information needs to be captured and then contained within the patient chart, the healthcare version of a “case folder.” Properly capturing that content is, of

“Sorting incoming content would be very useful for applications. Right now categorizing and organizing incoming content is tedious and takes a long time. It would be very useful as well if we can use AI to extract concepts, not just text. Lots of products extract text. Not a lot do nuance.”

Scott Fischer

-Director, Information Governance and
Legal Technologies at Aurotech Corp

course, a well-known battle, but perhaps lesser known is the equally critical battle over how to best present that information within the patient chart, because a medical practitioner needs to be able to pick up a chart and make an immediate decision based upon what they have time to review.

A CMF/SPA system can organize that information –and even action it to some degree, but not to the degree needed when lives can be on the line. The information has to be structured on a highly granular level so that medical personnel can make critical decisions, sometimes on critical patients. The structure also needs to be adaptable to a highly-granular and fluid context: it is not enough to designate the needs of each user simply by macro titles (“physician,” “nurse,” “technician”) because the roles each one has to play within a process will change rapidly and dramatically as the context changes. The primary person, often a primary physician, driving a process has their own ontologies and the way they want to think of a case, but all of the other knowledge workers (the other doctors, the nurses, the techs) have their own views of the case. All of these workers need similar information, but they will need to see it differently depending upon how their roles within the context.

How does a system determine that context with the precision necessary to empower efficiencies? Not from the users themselves, they don't have the time to organize the massive volumes of content. The CMF/SPA system needs to determine the context and deliver the right information – **by understanding the facts and events within the content, by uncovering fact patterns and by building the storyline. That is the power of ABBYY InfoExtractor – exactly.**

ABBYY InfoExtractor gives knowledge workers the intelligence and insight that they need; nothing less will drive productivity in our knowledge-centric economy. ABBYY InfoExtractor provides the otherwise missing link within the knowledge work technology stack, and it is the necessary component to leverage the true power of CMF/SPA systems.

Conclusion

Accelerating decision-making and driving organizational efficiency starts – and some would even go so far as to say ends – with creating the proper information environment for knowledge workers. Big (and ever increasing) Content is with us to stay, and will have a deleterious impact on productivity, because it strikes at the heart of what knowledge workers do: find, discover and analyze information to make critical business decisions.

Not surprisingly, many organizations have built out a specific IT infrastructure stack to support their knowledge workers. Unfortunately, shortcomings at every level of that knowledge work technology stack prevent it from enabling the desired level of efficiency. Each component of the stack, whether ECM, BPM, CMF, or first generation Smart Process Applications come close, but do not hit the mark because they lack the necessary precision.

For knowledge workers to action information and increase productivity they must have insight and intelligence into information to drive the process and make decisions. That insight can only come from systems that can understand language the same way that we do.

ABBYY's Compreno technology can empower knowledge workers within even the most advanced knowledge work technology stack. Compreno “understands” language to provide comprehensive text analytics that automatically and accurately identify entities and their properties, facts and events in text to reveal the stories within. Compreno provides the intelligence and insight needed to make Smart Process Applications truly “smart” by providing the right information – exactly, to accelerate and improve results.

References and Links

- ¹ Jacques Bughin, Michael Chui, and James Manyika, "Ten IT-enabled business trends for the decade ahead" McKinsey Quarterly, May 2013
- ² Thomas Davenport, "Thinking for a Living: How to Get Better Performances And Results from Knowledge Workers" 2005, at 6
- ³ AIIM, "State of the Art in Case Management 2013"
- ⁴ Michael White, "Case Management: Combining Knowledge With Process" BPTrends, July 2009
- ⁵ Derived in part from Forrester, supra.
- ⁶ AIIM, "2015 Training Course, Content Analytics" 2015

ABBYY®

ABBYY International Headquarters

Otradnaya str. 2b/6 127273
Moscow, Russia

Tel: +7 495 783 3700
Fax: +7 495 783 2663

Email: office@abbyy.com

ABBYY European Headquarters

Elsenheimerstrasse 49, 80687
Munich, Germany

Tel: +49 89 69 33 33 0
Fax: +49 89 69 33 33 300

Email: sales_eu@abbyy.com

ABBYY Australia

Citigroup Building, level 13,
2 Park Street
Sydney, NSW, 2000, Australia

Tel: +61 (02) 9004 7401

Email: sales@abbyy.com.au

ABBYY North American Headquarters

880 North McCarthy Blvd., Suite #220
Milpitas, California 95035, USA

Tel: +1 408 457 9777
Fax: +1 510 226 6069

Email: sales@abbyyusa.com

ABBYY 3A

(Asia, Africa, South America)

Otradnaya str. 2b/6 127273
Moscow, Russia

Tel: +7 495 783 3700
Fax: +7 495 783 2663

Email: sales_3A@abbyy.com