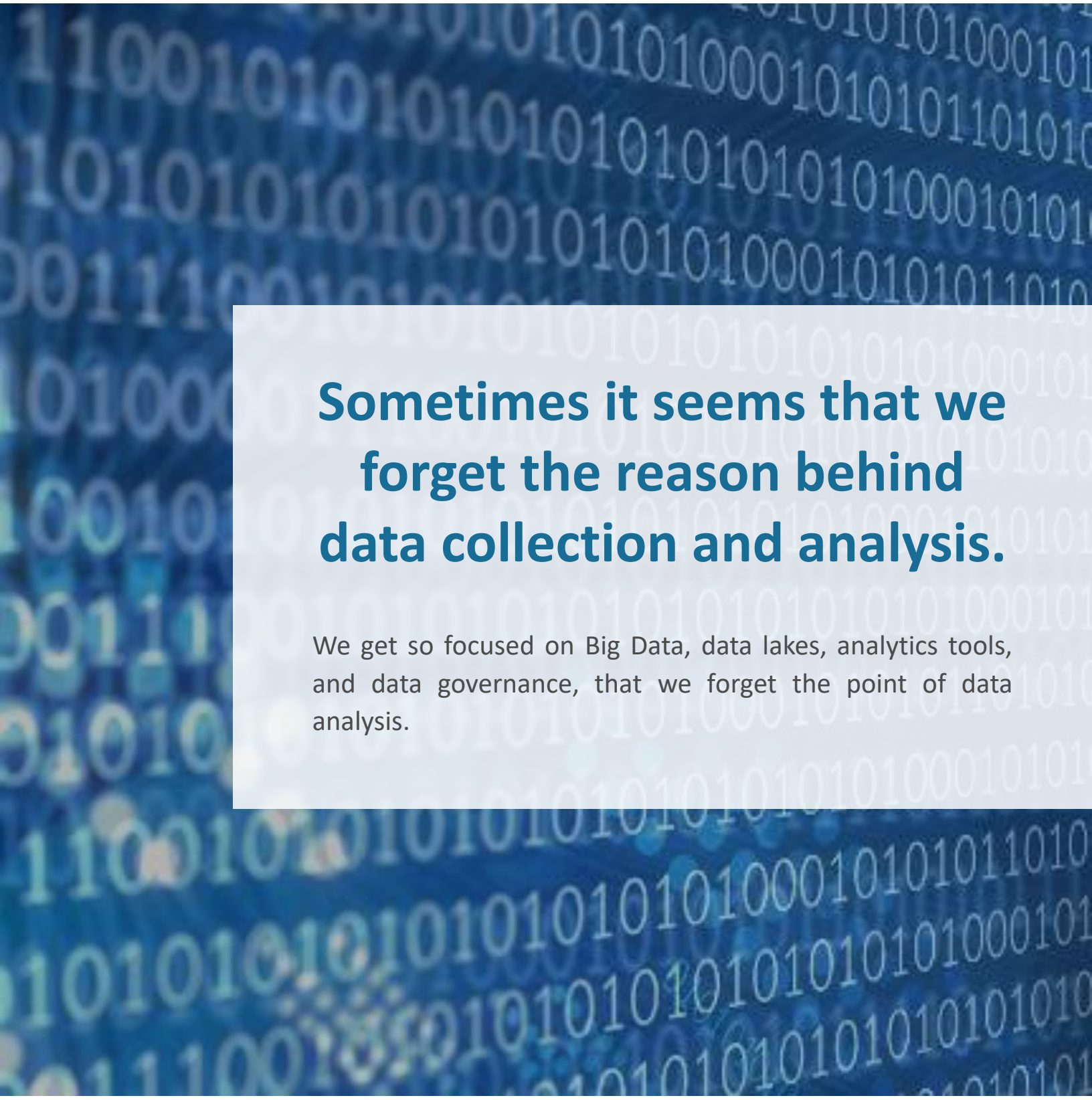
YSEOP Compose

Natural Language Generation

The Last Mile in Analytics and Business Intelligence

Contents

NATURAL LANGUAGE GENERATION



Sometimes it seems that we forget the reason behind data collection and analysis.

We get so focused on Big Data, data lakes, analytics tools, and data governance, that we forget the point of data analysis.

Fundamentally, we collect and analyze data to empower data-driven decision making to help our businesses succeed. To paraphrase Jean-Paul Sartre's famous remark, "it is only in our decisions that data is important."

How could we lose sight of such a fundamental point? The answer lies in the complexity of data collection, analysis, and governance. Entire divisions of companies have spent much of the last decade trying to build their data infrastructure, so you can’t blame them for losing sight of the ultimate goal.

The Last Mile: "No one ever made a decision because of a number. They need a story."

The quote above comes from Daniel Kahneman, a psychologist, Nobel Laureate, and one of the leading thinkers in the area of decision analysis, specifically how human beings make decisions.[1] In his book, "Thinking, Fast and Slow," Kahneman states that humans make decisions based on two "Systems":

1) an automatic, instinctual thought process that happens quickly; and

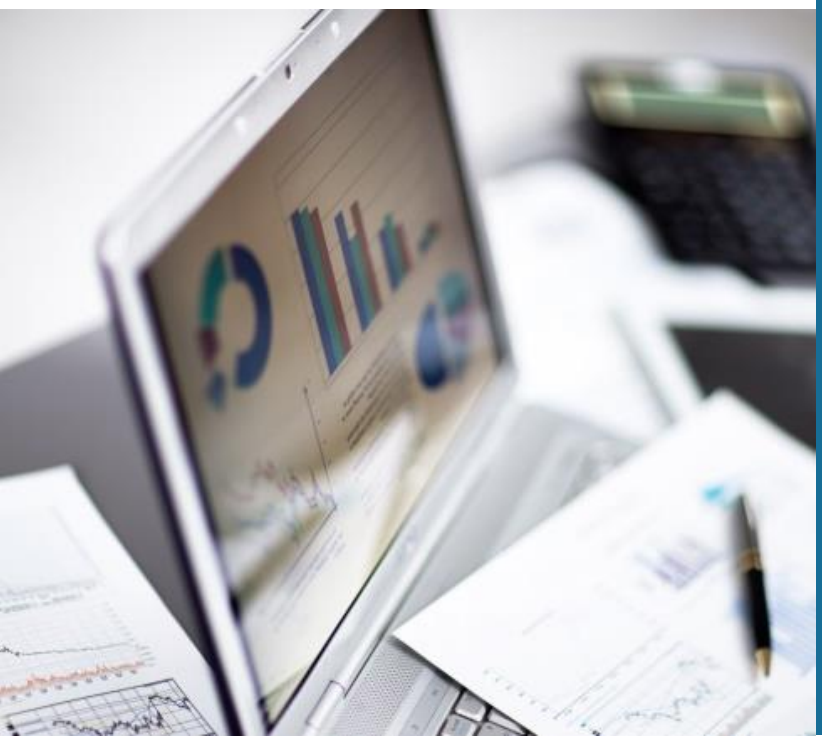
2) a slower, more deliberate thought process.

His basic premise is that our instinctual side often overpowers our slower, more deliberate thought process (for many reasons). This simple, yet novel idea seems quite familiar to many, because who hasn't heard a CEO say "I'm going with my gut" in response to numbers? Then how can a company see a strong ROI from Business Intelligence?

1. Kahneman, Daniel "Thinking, Fast and Slow" Penguin Psychology, New York, 2012.

The Last Mile: “No one ever made a decision because of a number. They need a story.”

Fundamentally, data is about numbers, but human beings don’t speak data, we speak English (or whatever language).



The answer is so simple that the founding thinkers of Business Intelligence could see it, but for us today who are so focused on technology, we tend to miss it. If I ask you to tell me how your company is preforming, how do you respond? Do you email me a graph? Do you send me an excel file? No, you use a combination of data, graphics, and a written explanation to paint the full picture.

The Last Mile: “No one ever made a decision because of a number. They need a story.”

The last mile in business intelligence is language.

Whether you have prescriptive or predictive analytics abilities, embraced data lake and Big Data, and have state-of-the-art analytics tools that process data in real time, you still need to explain your results, including the reasoning process, in a language the reader understands.

The Last Mile: “No one ever made a decision because of a number. They need a story.”

This written explanation can take the form of a report. The problem is that this can cost a company millions to do manually and are slow to generate, negating the value of real-time data. Conversely, this can also take the form of a conversation. How many times has the C-suite called analysts to say “what does this mean?” This approach has its problems too. **Your company can’t hire enough analysts to explain all the data either verbally or in written form. So what do you do?**

The answer is Natural Language Generation (NLG).

The Last Mile: “No one ever made a decision because of a number. They need a story.”

In January 2017, Forbes called NLG one of the hottest trends of the year. Gartner has described NLG as the last-step in the Data Discovery and BI & Analytics processes.[2] Essentially, Natural Language Generation is software that knows how to write automatically. It connects to data and Big Data tools and explains the results of the analysis in plain English.

Moreover, tools like Yseop Compose are fully self-service, so the analysts who manually write the reports can configure Yseop Compose so it scales their work, generating their reports but at the speed of thousands of pages per second.

So with Natural Language Generation, the last mile in the data-to-data-driven decision process is fully automatable.

But why haven’t we heard more about NLG? Why isn’t it integrated in all BI tools from the beginning? Interestingly, if you look back at the history of Business Intelligence, Natural Language Generation appears very early in the thinking process.

2. Sallam, Rita, Gartner, “Smart Data Discovery Will Enable a New Generation of Citizen Data Scientists” June 2015.

A Brief History of Data Analytics and Business Intelligence

The term “Business Intelligence” dates back to the 1860s when it appeared in R.M. Devens’ Cyclopædia of Commercial and Business Anecdotes. Devens describes how a somewhat unscrupulous business man, named Sir Henry Furnace, kept ahead of his competition with a vast network of people who acted as data collectors. These collectors reported directly to him as quickly as possible, thus enabling him to make decisions based on the best possible data. No doubt these sort of Business Intelligence networks have existed as long as competition between enterprises.

A Brief History of Data Analytics and Business Intelligence

In was in the 1950s, however, that Business Intelligence began to be industrialized. In 1958, an IBM researched posited:

"An automatic system (…) to disseminate information to the various sections of any industrial, scientific or government organization. This intelligence system will utilize data-processing machines for auto-abstracting and auto-encoding of documents and for creating interest profiles for each of the "action points" in an organization. Both incoming and internally generated documents are automatically abstracted, characterized by a word pattern, and sent automatically to appropriate action points."[3]

3. Luhn, H.P. A Business Intelligence System, IBM Journal, October 1958.

Luhn, the author of the paper, called this system a Business Intelligence System. He lays out the exact technical requirements for an end-to-end Business Intelligence System that would provide executives with all the data they need to make decisions.

But critically, Luhn didn't stop there, he argued the Business Intelligence System would also generate written documents and action points, essentially explaining the results of the data collection and analysis within the system.

In short, at the advent of contemporary BI, Natural Language Generation was already seen as the last step in the process, the last mile in the BI & Analytics process.

Criteria for Luhn's Original End-to-End Business Intelligence System

Let's dig a bit deeper. Using modern terminology, we can break down Luhn's system into a number of existing components.

1. Data Collection

2. Data Analysis

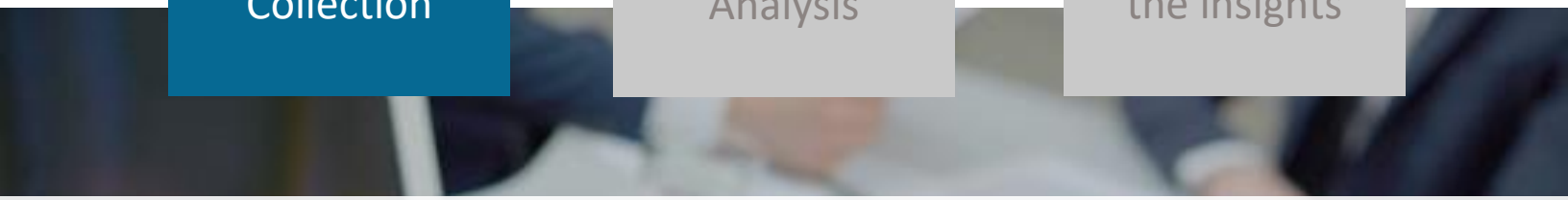
3. Explanation of the insights

Criteria for Luhn's Original End-to-End Business Intelligence System

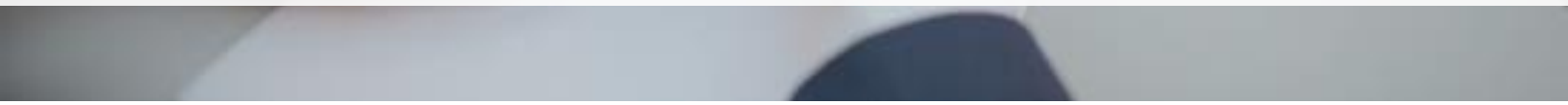
1. Data Collection

2. Data Analysis

3. Explanation of the insights



Luhn is a little unclear on what he generally calls the "data processing" portion of this system, and it makes sense. His thinking was decades ahead of the technology. Here we would break down the data into two categories: structured and unstructured data. For the structured data, we would firstly talk about an MDM plan and data governance for data integrity. Then we would look at state-of-the-art systems like Teradata or Oracle for the actual databases. With regard to unstructured data, we would look at Natural Language Understanding systems like Nuance, which transform the unstructured data into structured data.



Criteria for Luhn's Original End-to-End Business Intelligence System

1. Data Collection

2. Data Analysis

3. Explanation of the insights

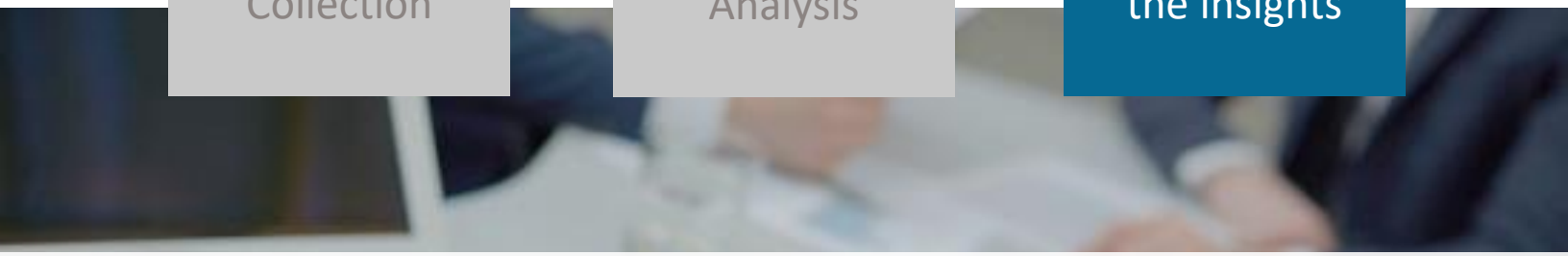
Luhn again seems to assume that the analysis portion is obvious, though he does focus a little on the data analysis workflows. However, he doesn't focus on the volume of data, nor does he seem to foresee what we would call Big Data. Today, the analytics market is full of powerful data analysis solutions and many of the leading BI vendors contain analytics engines more powerful than anything Luhn could have dreamed up.

Criteria for Luhn's Original End-to-End Business Intelligence System

1. Data Collection

2. Data Analysis

3. Explanation of the insights

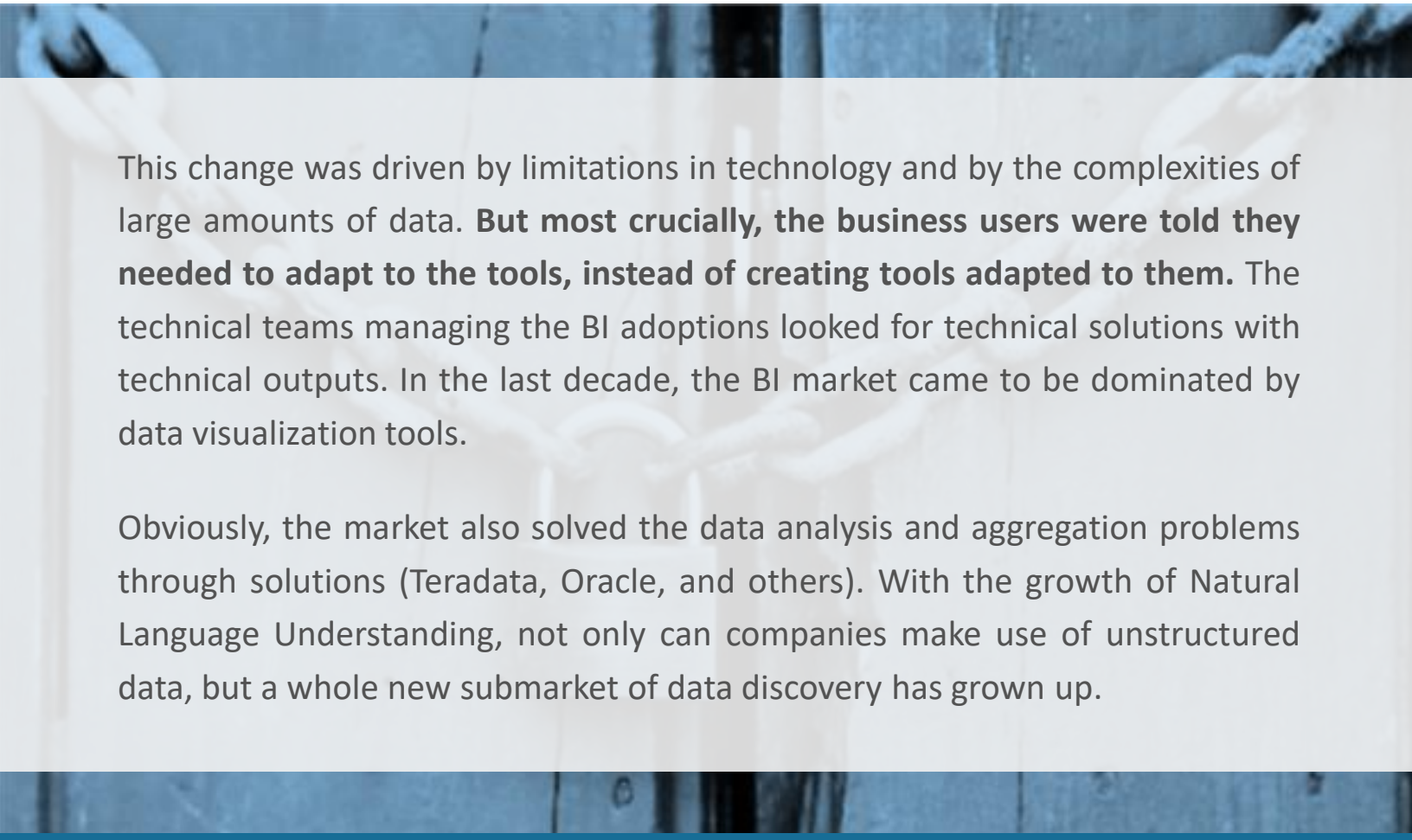


Luhn discussed automated summaries or abstracts of data, along with action items and the generation of written documents. The focus, for Luhn, is narrative describing the results of the analysis. Today, we know the technology that automates the writing of written documents is called Natural Language Generation (NLG). NLG is the last step in Luhn's Business Intelligence workflow, the last mile required to explain, in a universally understandable way, the results of the data analysis.

Luhn's Business Intelligence System actually mimicked the way a human being works: we collect data, analyze it, and then explain the results in written or spoken form. Simple, right? Then how did we lose our way? Why isn't Luhn's BI tool in every business today?

Modern Business Intelligence Approach Misses the Last Step!

As technology advanced, the market moved away from Luhn's vision.



This change was driven by limitations in technology and by the complexities of large amounts of data. **But most crucially, the business users were told they needed to adapt to the tools, instead of creating tools adapted to them.** The technical teams managing the BI adoptions looked for technical solutions with technical outputs. In the last decade, the BI market came to be dominated by data visualization tools.

Obviously, the market also solved the data analysis and aggregation problems through solutions (Teradata, Oracle, and others). With the growth of Natural Language Understanding, not only can companies make use of unstructured data, but a whole new submarket of data discovery has grown up.

Despite all these advances, BI vendors and users alike have forgotten about the last step and the real point of analytics and BI: explaining insights from data as clearly and quickly as possible to support data-driven decisions.

The advent of data visualization tools was both a quantum leap in the Business Intelligence market and an acknowledgement that Natural Language Generation, which could write reports explaining the data and action items, wasn't yet available. Admittedly, in the early 2000's, NLG technology wasn't scalable, so the growth in data visualization made sense even as late as 2010.

But why are companies still relying only on graphics?

Is the CDO the Savior of Business Intelligence?

We've already discussed the disconnect between BI and the actual business. This disconnect has been highlighted by Gartner and others who've pushed companies hard to adopt a Chief Data Officer (CDO). In fact, Gartner believes that most large organizations will have a CDO by 2019. **The CDO's job is data monetization, data governance, and making sure that data serves the business and business users.**

So far the CDO seems to mostly exist in the largest international enterprises. Many of the CDO's are relatively new, so the long-term impact is hard to measure. However, anecdotally, CDO's do seem to be refocusing BI and analytics efforts on the business, specifically the last mile in Business Intelligence, where they can finally see the ROI for their other investments. However, a CDO isn't necessary to adopt an A-Z approach.

When You Need the Last Mile in BI and Analytics

By definition, the last mile isn't necessary for every business at every time. There are some first steps you need to take, such as an overarching data governance strategy and some basic data aggregation if you are dealing with Big Data.

However, there is also a simple set of three questions you should ask yourself:

1. **Do you pay analysts or others to write data-driven reports explaining, summarizing, and advising based on data?**
2. **Do you sell data?** If you are selling data to another company, the chances are they are writing content about it and then reselling it at a 1,000% mark-up. That's revenue you are missing out on.
3. **Are you failing to see the ROI you expected from your BI investments and adoptions?**

If the answer to any of these questions is yes, then you should be seriously considering Natural Language Generation to help you reach that last mile in your analytics and BI solution.

About US

Yseop offers a powerful artificial intelligence software that automates knowledge work, dialog, and generation of written content in multiple languages. We believe the next industrial revolution will be the automation of the service sector. Just like robots in manufacturing boost productivity and enhance capacity, Yseop's software suite automates expert customer interaction, either as a guide for employees or directly to the customer. **Our software also generates written reports, analysis, and advice automatically using a client's best practices.** Now, companies can use their data in real time and guarantee expert customer service for all, no matter the account size or channel.

Secure
Install securely on your server or personal cloud

Tried & True
50,000 Users in production in Fortune 1000

Global
Multilingual software with offices in EMEA, US, and APAC.

Autonomous
Only enterprise, self-service, language generation software

Learn More: **https://compose.yseop.com/**

404 5th Ave, New York, NY 10018 | (214) 780-9730 | compose@yseop.com