

## Accelerate discovery with Diffeo's AI-powered research assistant

*Diffeo performs data reconnaissance, scouting out several steps ahead to bring back data you didn't know existed*

### EXECUTIVE SUMMARY

While data creation and accessibility has grown dramatically, our ability to consume and utilize that data has not kept up. Humans are limited by time and language skills; however, full automation is not an option as machines lack the intuition and broader understanding of research goals that a human researcher brings to the table. Diffeo's automated assistant helps humans discover connections by recommending information the analyst hasn't yet seen. It continuously improves its recommendations by tracking what users add to working notes or highlights in a browser. With computer speed, Diffeo uncovers surprising and vital information that would otherwise go undiscovered. With all-language coverage a priority, Rosette's multilingual offerings are key to covering all sources.

### KEY HIGHLIGHTS

- **An AI research assistant:** The Diffeo agent combines the teammate role of digital assistants with the data and insights of a business intelligence tool for an entirely new research experience.
- **Better results, faster:** Diffeo automatically recommends game-changing data in a fraction of the time a person would spend creating complex search queries.
- **Global coverage:** Key data can be hidden in text in any language and the foundation for Diffeo's magic is powerful, native-trained, multilingual entity extraction. Diffeo turns to Rosette for powerful linguistic analysis, especially in complex, difficult languages like Arabic, Chinese, Korean, and more.

## The tech talk that produced a company

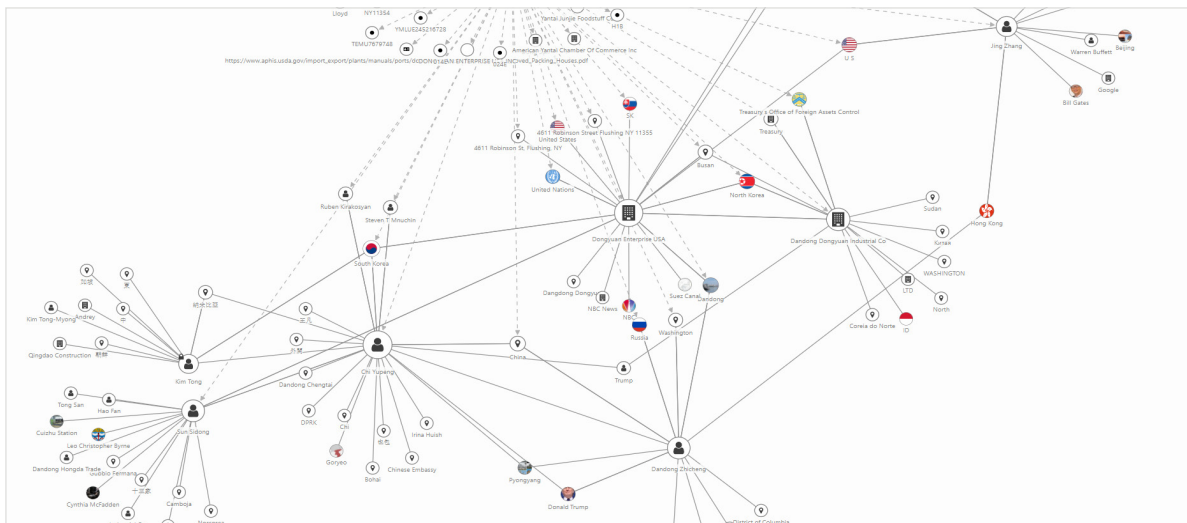
Diffeo began as a brainstorm amongst cofounders John Frank, Max Kleiman-Weiner, and Dan Roberts at a Hertz Foundation summer workshop. The team wondered how to use technology to help Wikipedia authors write faster. How can we accelerate the assimilation of knowledge? The reliance on public participation fueled Wikipedia's rapid growth and continues to enable its adaptability, however it simultaneously means that the creation of new pages is haphazard rather than strategic—subject to the whims and interests of its most active editors.

That conversation grew into the Text Retrieval Conference Knowledge Base Acceleration project (TREC KBA), which Diffeo organized for the National Institute of Standards and Technology (NIST) and DARPA Memex. The project evaluated algorithms for helping people discover new knowledge to add to a knowledge base like the crowdsourced Wikipedia. The trio started to think of a computer-human collaboration that would help Wikipedia contributors produce better content more quickly.

*The Diffeo agent combines the teammate role of digital assistants with the data and insights of a business intelligence tool like the Bloomberg terminal.*

## A collaborative, all-source research agent powered by AI

The Diffeo team wanted to create a tool that would understand and mimic how people behave when performing research. The human brain makes connections constantly, sifting through vast amounts of data, drawing conclusions, and connecting that information to related and tangential concepts. Diffeo combines that abstract



thinking with the power and speed of artificial intelligence to create an automated research assistant that allows researchers to access more sources and produce better analyses, faster.

## An automated research partner

Modern operating systems are not designed to integrate data across applications. Even if Chrome and Outlook are running at the same time, the two programs are not communicating with one another. When users download the Diffee agent, they invite it into all of their computer's programs, acting as a bridge between them. This universal access allows the agent to make cognitive connections between data across browser tabs and different desktop applications, much as a human would. The Diffee sidebar recommends related data regardless of whether it comes from online articles, local database documents, or your email.

Diffee also offers knowledge "boards"—like whiteboards—where researchers can edit and interact with the Diffee agent like a colleague. Each Diffee Knowledge Board is rendered from an underlying knowledge operations journal. Operations journals were made famous by the multi-person editing capabilities of Google Docs. Diffee has extended the idea of operations journals to the knowledge exploration process. Each knowledge board shows suggested changes from the collaborative agent as well as from other users working on that board.

The Diffee agent combines the teammate role of digital assistants with the data and insights of a business intelligence tool like the Bloomberg terminal. This makes Diffee an ideal research partner.

## Research without complex queries

Under the hood, the Diffee agent uses machine learning to track researcher choices, so that it can model what the person knows and what topics they're interested in pursuing further. The sidebar updates with new suggestions by watching what the researcher decides to open, ignore, or dismiss. Further, the agent remembers what information the user has already seen and highlights new information in red and repetitive information in blue, allowing the eye to quickly skip to the most useful bits.

*Diffee does the work for the researcher, automatically creating and altering queries to get to the most pivotal bits of information.*

When researching complex subjects, one must formulate complex queries for many search tools. While you can have a conversation with a colleague to flesh out a research project, traditional search engines cannot participate in that free wheeling exchange of ideas. The flexible nature of the Diffee agent lets the researcher have a dialogue of sorts with it. Instead of writing complex Boolean queries into search form fields, Diffee does the work for the researcher, automatically creating and altering queries to get to the most pivotal bits of information.

## Data you didn't know you were missing

The Diffeo agent also performs gap analysis, finding information that fills in knowledge gaps for the user. Instead of supplying more-of-the-same articles, the Diffeo agent seeks surprising information that expands the researcher's horizon.

*Diffeo's recommendations are "categorically different" because they show "new documents that mention key concepts that the user did not yet realize could be found."*

How do you know what you don't know? It's easy to miss critical information. Often, you'd need to know the query terms ahead of time. With so much information publicly available on the internet, it is impossible for a researcher to compose exactly the right query for all of the information they don't know yet. It's hard to uncover unknown unknowns.

For example, one of Diffeo's subject matter experts (SMEs) in intelligence analysis was studying a known group of North Korean money launderers. While parts of the network had already been sanctioned, the Diffeo researcher wondered what other parts of the network might still be unknown. Bad actors often create new front companies. Sure enough, as the analyst took notes on the known entities in the network, Diffeo recommended related entities. One of these turned out to be one of the few apple (fruit) shipping companies with a license from the U.S. Department of Commerce for importing apples into the United States. Surfacing this kind of surprising connection is Diffeo's forte.

Sample shipment record for **Dongyuan Enterprise Usa Inc**

Dongyuan Enterprise Usa Inc imports from **Yantai Junjie Foodstuff Co., Ltd** in **South Korea** through the port of **New York/newark Area, Newark, New Jersey**

BILL OF LADING		OPERATOR	KODLE240001556A
Manufacturer / Shipper	Consignee	ARRIVAL DATE	2017-01-18
YANTAI JUNJIE FOODSTUFF CO., LTD XUNJKIANG COUNTRY SIEWOPO TOWN CA XIA CITY SHANDONG CHINA ANHUI, CHINA	DSONGYUAN ENTERPRISE USA INC 15052 28TH VENUE SUITES 301 FLUSHING NY 11354 CALIFORNIA UNITED STATES	WEIGHT (LB)	41,888
Notify Party	Marks and Numbers	WEIGHT (KG)	19,040
-NOT AVAILABLE -NOT AVAILABLE	N/M	QUANTITY	1904 BT
Master Bill of Lading Number	Container Number	CONTAINER(S) COUNTRY	1
YMLUE240061656	SZLU6890377	VOYAGE NO.	024E
Product Description		PORT OF LOADING	PUSAN
APPLES		PORT OF DISCHARGE	NEW YORK/NEWARK AREA NEWARK, NEW JERSEY
		VESSEL NAME	HOUSTON BRIDGE

Recommendation: **Yantai Junjie Foodstuff Co., Ltd**  
[www.importgenius.com/cn/stu...](http://www.importgenius.com/cn/stu...)  
 Yantai Junjie Foodstuff Co., Ltd 通过 New York/newark Area, Newark, New Jersey 港出货给 Orchard Direct, Inc. Call +1 (480) 744 2559 在线注册 Yantai Junjie Foodstuff Co., Ltd 的提单样本 Yantai Junjie Foodstuff Co., Ltd 通过 New York/newark Area, Newark, New Jersey 港

## Surface key data in any language

Diffee's recommendation feature is particularly powerful when in a language the researcher doesn't speak.

Suppose an intelligence officer is researching political discontent in a city in Syria. The officer is fluent in Arabic and Russian, and can easily exploit open source data like social media in those languages, however he'll miss data in Urdu without help. The Diffee agent takes care of this work, and matches entities across languages, automatically formulating multilingual queries for the researcher and returning all relevant results regardless of the researcher's language ability.

*Instead of supplying more-of-the-same articles, the Diffee agent seeks surprising information that expands the researcher's horizon.*

Human translation is extremely costly in terms of resources and time. Diffee's multilingual fluency saves time and money by flagging multilingual content for closer analysis. The researcher can then machine translate to get a gist and then selectively forward to the human translator only data they already know will be valuable.

## Rosette powers dialogue and discovery

As the Diffee founders built their AI-powered research agent, they knew that entity extraction would be a foundational component. It needed to be highly accurate, fast, comprehensive and cover many languages. Luckily, Diffee was in a position to know exactly which commercial entity extractor was the best as it had performed "bake-offs" between commercial and open source entity extraction tools for a sophisticated customer building identity resolution and entity co-referencing. Diffee concluded that Rosette was the most usable tool with the broadest language coverage (20 languages) and most consistent quality across languages—a priority for any global analysis projects.

*"We trust Basis Technology to be responsible for certain difficult language categories...With ready access to a slew of languages, we can accelerate adding new languages to Diffee for greater competitive advantage."*

— John Frank, Diffee CEO & Co-founder

Rosette delivers a text analysis pipeline of capabilities, first identifying the language of each document and then applying the correct language-specific morphological analysis to produce tokens (“words”), part-of-speech tags, and lemmas (dictionary form of a word). From there entities (people, places, organizations) are extracted, and if some names are in challenging languages such as Arabic, Chinese, Korean, or Russian, Rosette transliterates them for English-speaking analysts to see what entities are mentioned in foreign language documents.

“We trust Basis Technology to be responsible for certain difficult language categories,” says Diffee CEO and co-founder, John Frank. “With ready access to a slew of languages, we can accelerate adding new languages to Diffee for greater competitive advantage.”

## Diffee powers financial and intelligence analysis

Diffee found a ready market for their AI-research agent in the financial and government intelligence industries where sophisticated analysts are constantly updating research. Similar to Wikipedians, they work from hundreds of pieces of source material. For example, an intelligence analyst may be looking to compile all OSINT (open source intelligence) data on a recent protest, and will want to pull from historical documents, as well as real-time social media posts. An analyst at a Wall Street bank may need to study a private company or physical infrastructure assets in an emerging market. Such counterparty analysis draws on internal documents as well as public data from the deep web. The Diffee agent helps these researchers quickly and thoroughly understand all of the connections.

When working with a new client, Diffee’s team of subject matter experts (SMEs) helps the customer measure the value of the Diffee agent by setting up an internal evaluation of the tool. Such evaluations are most easily accomplished as task completion measurements: an analyst will work on a test topic without Diffee until they cannot find any more. Then, the user invites Diffee to help. Frank says, the number of citations (sources discovered) typically doubles and prospective buyers remark that Diffee’s recommendations are “categorically different” because they show “new documents that mention key concepts that the user did not yet realize could be found.”