GIANT OAK

We see the people behind the data.

Beyond Google

> The purpose of this white paper is to explain the strengths of GOST® versus traditional public search engines.

The State of Search

Over the years, traditional search engines like Google and Bing have transformed the way risk and compliance professionals perform screening, monitoring, and investigations of customer, correspondent, and vendor populations by opening the internet to the public. However, there are four primary reasons that traditional search engines remain imperfect tools for large-scale screening and vetting missions.

First, traditional search is not optimized for the returns needed by professionals to identify illicit actors. This stems from the fact that other search engines are designed to optimize revenue, primarily by selling ads to consumers. These search engines return information according to what would be of greatest interest to the commercial browser. This method of search return adds unnecessary work for professionals tasked with screening and vetting, as they provide popular pages but not pages relevant to finding the derogatory information required for proper due diligence.

Second, manual searching takes a massive amount of time, with little return; and third, general-purpose search engines are not efficient. These two challenges go hand-in-hand: though on average each case takes more than twenty minutes to complete and document, the efficiency rate of identifying illicit behavior is as good as random.

Fourth, KYC and due diligence practices are not auditable or replicable using general purpose search tools. Analysts must know the right combination of keywords to identify risk, fraud,

and illicit behavior. This knowledge varies from person to person, each with their own biases. Results are inconsistent when using general-purpose search engines, compounded with user bias.

The internet is a dynamic and living place, where information appears, disappears, and then reappears at a moment's notice, further complicating the challenge of replicability. The dynamism of the internet makes screening and vetting a labor-intensive process, which is a gamble for financial institutions. Relevant information may vanish before it can be reviewed or addressed, and illicit actors may intentionally produce new content to bury negative information deep in the search results. The inability to compare and replicate results makes onboarding and customer refresh costly in both time and money, and unnecessarily risky.

Financial institutions need a solution that allows risk-mitigation and regulatory compliance professionals to efficiently search an entire population, at scale, in a repeatable and auditable way. This solution must also have continuous vetting capabilities. GOST, by Giant Oak, was built specifically for this purpose.

Innovation for Search and Retrieval

GOST meets the four challenges identified above. GOST is a bespoke screening tool that re-indexes the internet for fraud, risk, and compliance professionals. This reorganization brings forward the derogatory information professionals need. GOST enhances the best features within Google and Bing- their reach and depth- by using machine-learning algorithms to

prioritize cases. GOST's ability to re-index the internet to show results that are most relevant for analysts and investigators addresses the main shortcomings of relying on ad-hoc search in a world where results are distorted by ad-driven algorithms. GOST's searches are also uniform for all entities, reducing the liabilities associated with human bias while minimizing false positives. GOST saves time, making the tool a force multiplier for any financial institution.

1. Domain-Specific Search

GOST empowers compliance and risk professionals. Among its most powerful features is its ability to perform domain-specific search and retrieval at scale.

To help understand this concept, it is worth remembering that the internet is a large and messy place. Without the right tools to locate and sort the billions of websites online, the internet would be unusable. Commercial search engines like Google and Bing simplify the internet by indexing large slices of the web and making this information available to the public, free of charge. The availability of web pages on the internet comes to us via online advertising models.

In most people's minds, the internet is comprised entirely of these commercial search engines. But the truth is that commercial search engines are dependent on ad revenue to remain commercially viable. This means that their search results are optimized to promote the websites which generate the highest number of clicks for ad-buyers. While this prioritization is fine for the average person interested in buying goods online or accessing social media platforms, it creates problems for investigators searching for derogatory information. Search engines do not focus on derogatory information because derogatory information does not create revenue.

If the full blue space represents all of the publicly available electronic information (PAEI) about an entity,

this is what a General Purpose Search Engine displays,

... this is what GOST® displays,

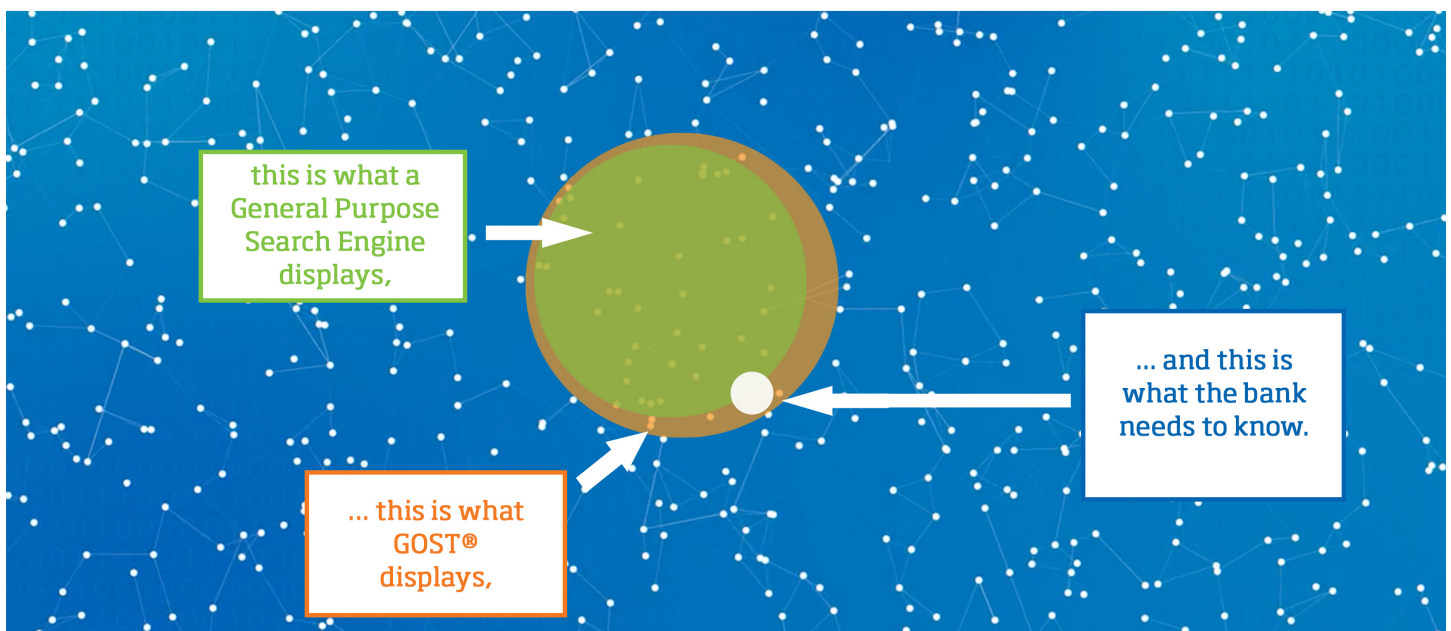... and this is what the bank needs to know.

Figure 1) Traditional search engine internet index versus GOST internet index.

> Leveraging behavioral science and machine learning, GOST creates custom information domains of the internet for an organization, allowing users to prioritize results that are most meaningful to their mission. Each GOST deployment provides a bespoke internet for each institution.

To complicate the task, not everyone in risk and compliance screens for the same kinds of derogatory information. Different organizations with screening needs are interested in distinct behaviors. For example, one institution might search for money launderers, while another might search for human traffickers.

Leveraging behavioral science and machine learning, GOST creates custom information domains of the internet for an organization, allowing users to prioritize results that are most meaningful to their mission. Each GOST deployment provides a bespoke internet for each institution. If your institution needs to identify money launderers, GOST will index the internet specifically for those behaviors. If you are more interested in human traffickers, GOST will do that as well. Giant Oak builds domains by first considering the behavioral patterns that produce different kinds of internet data. GOST then uses machine learning to construct algorithms that organize the desired information across countless webpages. GOST continually refines the quality of the results by using artificial intelligence (AI), allowing users to indicate whether the information returned provided value. This interaction by the users of GOST further trains the underlying algorithms to improve GOST's ability to bring back more germane results after each use.

2. Time and Cost Savings

GOST saves massive amounts of time on screening and vetting processes.

Imagine that t equals the time that it takes for a financial institution's compliance team to complete an advanced search, plus the time it takes to subsequently capture the relevant information returned from that search into a case management or customer account lifecycle system.

For example, consider a large bank that annually searches over one million customers and associates. These can include daily onboarded clients, high- or low-risk populations under periodic review, or counterparties to transactions. For mathematical simplicity, assume an extremely conservative

estimate of t = 6 minutes to run a comprehensive advanced search and to capture all relevant information on a single entity. We can see that searching:

10 entities takes 1 hour,
100 entities takes 10 hours, and
1,000 entities takes 100 hours.

Using GOST reduces this time dramatically. Completing 1,000 searches and retrieving relevant information takes just 10 minutes.
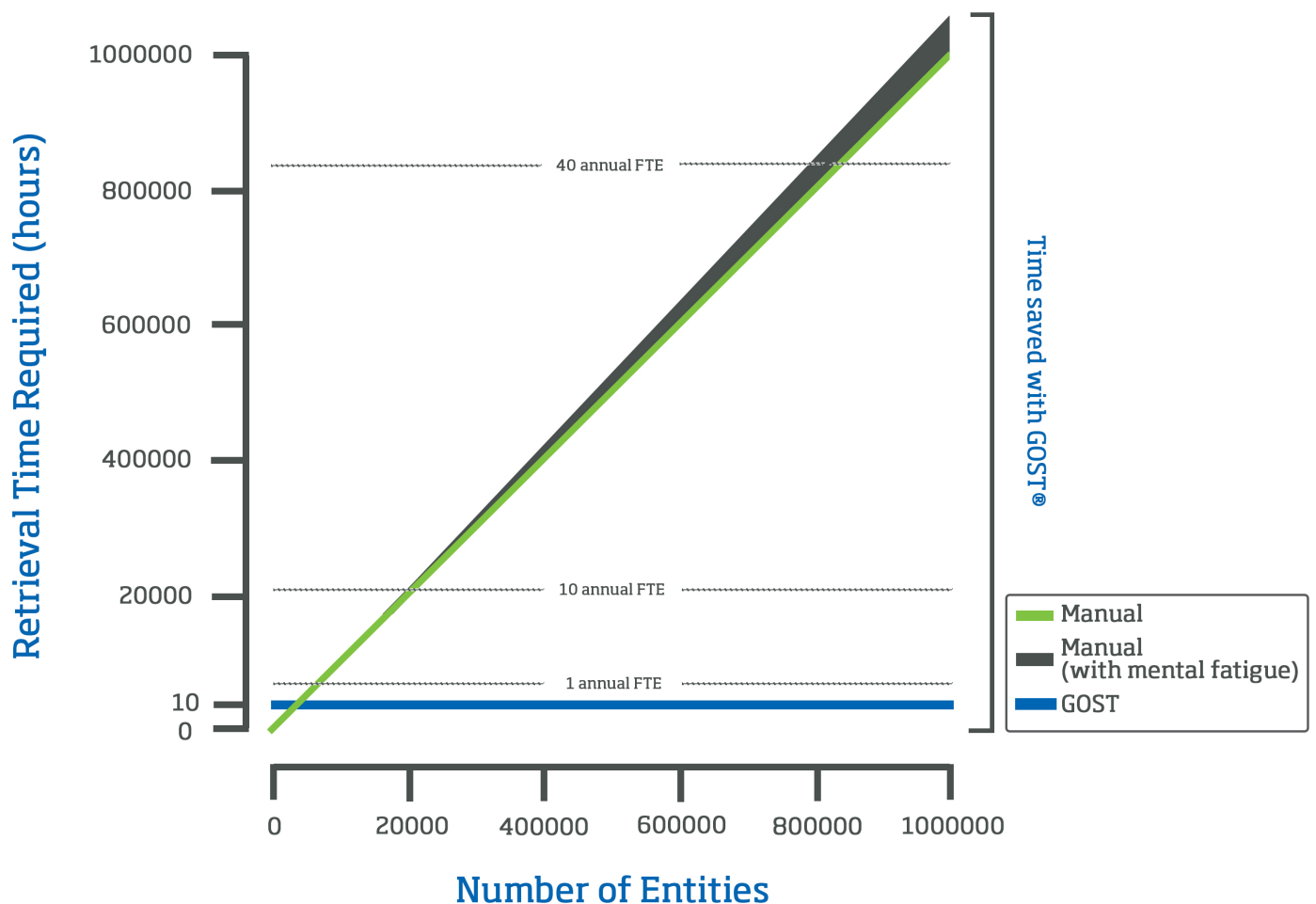
Retrieval Time Comparison

Figure 2) Retrieval time comparison between a manual review process using Google, and a semi-automated process using GOST.

[1] Note that these estimates do not account for analysis time or investigator fatigue, but represent time purely spent on consistent and repeatable search and retrieval. Accounting for these conditional lags will result in increased search times.

Even with these most conservative time estimates, going through search and retrieval for 1,000 cases in real time while using manual processes and tools would take 12-18 regular working days. This time is accompanied by a requisite mental and physical fatigue. Unless more fulfilling and creative tasks are interspersed between the search and retrieval workflow, this sort of monotonous work is not sustainable for a staff of professionals.

Figure 2 illustrates the time and cost savings that GOST provides. Using GOST, the time needed for an analyst to screen and rank a large number of entities remains constant. A professional completing the task manually requires additional hours to screen growing numbers of entities. Notice that the GOST line remains at 10 minutes, while the manual line grows linearly.

Even with these most conservative time estimates, going through search and retrieval for 1,000 cases in real time while using manual processes and tools would take 12-18 regular working days. This time is accompanied by a requisite mental and physical fatigue. Unless more fulfilling and creative tasks are interspersed between the search and retrieval workflow, this sort of monotonous work is not sustainable for a staff of professionals.

In addition to the time this manual process takes, each professional's cognitive biases also impact the screening. One example of a common cognitive bias is the availability heuristic. An availability heuristic is a mental shortcut that relies on immediate examples that come to an analyst's mind when evaluating a specific topic, concept, method, or decision. For example, if a professional had just completed a large fraud case, they would be more likely to report the fraud information than the money laundering information, even if their search turned up equal amounts of data on both activities.
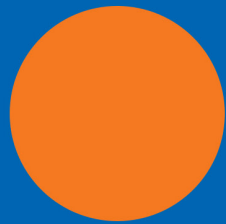
Figure 3) A visual representation of analyst bias.

A professional's efficacy and efficiency are impacted by both mental fatigue, which slows them down, and cognitive biases, which impact their decision-making processes.

GOST addresses these challenges and serves as a cost-and time-saving resource for compliance professionals in the financial industry. GOST:

1. Lowers full-time employee (FTE) costs dedicated to manual search and retrieval tasks,
2. Saves money by preventing losses due to illicit actors missed in the screening process,
3. Allows search and retrieval on more entities than is currently feasible at current budget levels, and
4. Frees up investigator time to focus on high-value human tasks, such as judgement and analysis.

3. Increased Efficiency

Once GOST provides the custom index and algorithms, high-speed prioritization of cases makes large-scale screening and vetting incredibly efficient. GOST provides two scores that help professionals prioritize information relative to illicit actors:

Ranking measures the relevance of the information GOST identifies about a specific entity for users. It is determined by the specific purpose of a domain. For instance, a domain built to identify money laundering will provide higher scores to search results containing information about money laundering, and lower scores for information dealing with other behaviors.

Reliability measures GOST's confidence in entity resolution for that case. This reduces false positives by providing professionals an additional way to prioritize cases.

For screening, these scores allow professionals to prioritize cases immediately. For continuous vetting, GOST provides threshold alerts, enabling analysts to monitor changes in data associated with individuals over time, without having to manually review entire populations.

Used together, Ranking and Reliability scores enable users to quickly process entities by prioritizing only those results that have information relevant to the domain of interest, like money laundering, drug trafficking, etc. Surfacing those entities that return the most derogatory information reduces labor costs and allows analysts to focus on higher-order tasks. By allowing professionals to review a higher volume of cases, GOST enables them to look more thoroughly at entities that would normally be ignored, or never considered for screening, because of the massive cost of manual review.

We can see these capabilities demonstrated in Figure 4, which illustrates the use of Ranking and Reliability for prioritizing cases to dramatically increase performance.
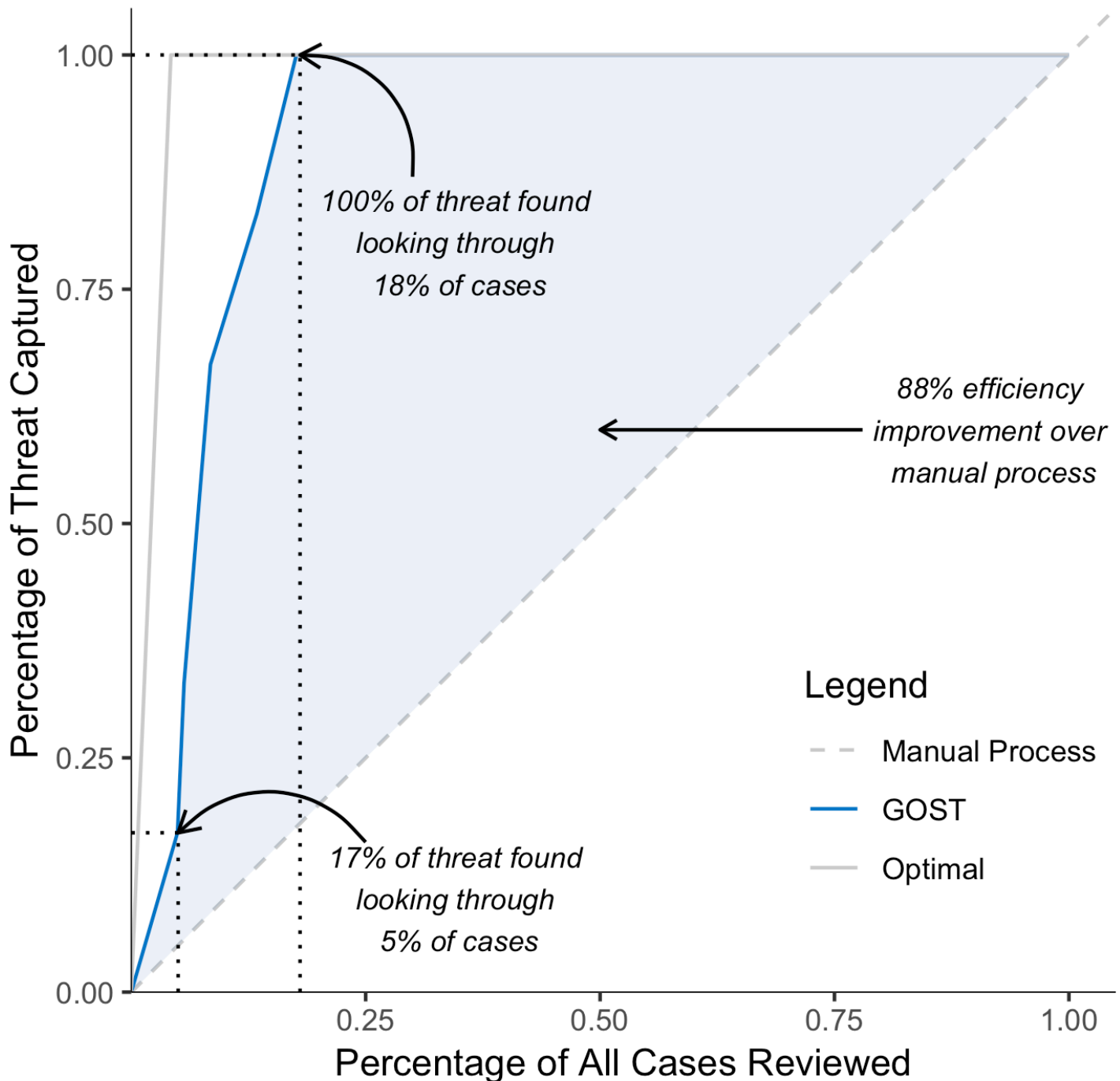
Figure 4) Efficiency of a manual review process using Google, and a semi-automated process using GOST®.

Giant Oak's standard diagnostic for assessing GOST's performance is the effort curve. The effort curve measures the share of the population examined (the x-axis) in order to find a given share of bad actors (the y-axis). The effort curve compares how quickly analysts using GOST find illicit actors relative to a standard review process.

Using GOST to prioritize cases with Ranking and Reliability scores, the team would be able to find 100 negative entities, or 100 percent of the illicit actors within the sample population, while reviewing 18 percent of the sample population of 10,000 - or 1,800 entities.

The solid grey line shows theoretical maximum efficiency, in which all illicit actors are identified before any others in the dataset are reviewed. The solid blue line shows the share of illicit actors identified by using GOST's Ranking score to triage and prioritize entities for review. The dashed grey line shows the efficiency with a standard randomized review process, or in simpler terms, the manual review process. The shaded blue area represents how much more efficient GOST is relative to the manual review process.

We can use a short thought experiment to illustrate this advantage. Imagine that an institution with an entity population of 1 million has a money-laundering incidence rate of 1 percent. This means that of those 1 million entities, 10,000 launder money.

Of that 1 million entity population, investigators or analysts might review a sample of 10,000. With an incidence rate of 1 percent, they can expect to find 100 money launderers in that sample. However, an analyst won't complete 10,000 searches.

In contrast, using GOST to prioritize cases with Ranking and Reliability scores, the team would be able to find 100 negative entities, or 100 percent of the illicit actors within the sample population, while reviewing 18 percent of the sample population of 10,000 - or 1,800 entities.

Overall, using GOST in this production model results in an 88% efficiency improvement compared to the status quo manual process. Ranking and Reliability scores make this improvement possible.

Because GOST is an AI system that elicits feedback from human reviewers, over time the blue GOST line in Figure 4 will move closer and closer to the Optimal Efficiency line. The Optimal Efficiency line is tantamount to having a tool that lets you perfectly classify good behaviors and bad behaviors, with no wasted effort. GOST takes the manual search and retrieval process and brings you 88% closer to that

4. Auditable and Replicable

GOST's search and retrieval workflow is a repeatable and automated process.

To be successful, manually-completed traditional search processes are reliant on very specific advanced search logic. Figure 5 illustrates an example of one of these searches. Each component is intentionally chosen by the individual assigned to the entity.

if you search this ...

"site:rt.com, ~weapons, "Viktor Bout", -Tajikistan, 2008 ... 2012"

... this is what it means.

| SITE: only searches the pages of that site | " " Searches for the exact phrase, not each of the words separately | ... shows all results from within that designated time range |

| site:rt.com | ~weapons | "Viktor Bout" | -Tajikistan | 2008 ... 2012 |

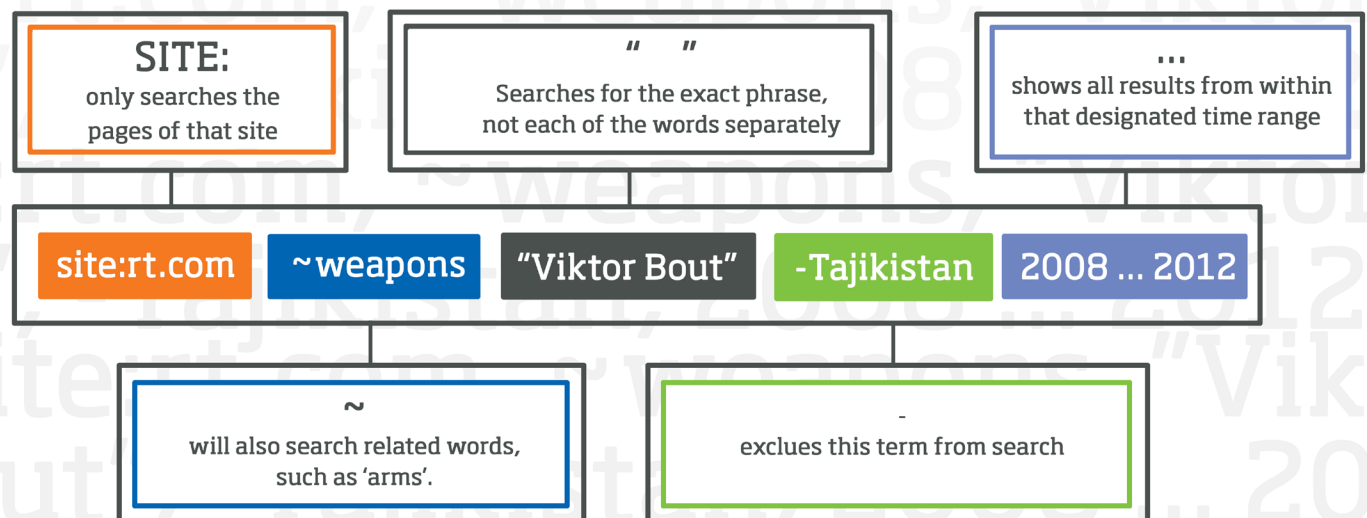| ~ will also search related words, such as 'arms'. | - exclues this term from search |

Figure 5) An example of a very specific advanced Google search, typical of a manually-completed traditional search process. These are almost impossible to replicate across departments.

> Additionally, the platform can place large populations into a state of continuous vetting. Continuous vetting is a state in which GOST periodically screens at an interval set by managers with the intent of looking for changes in patterns of behavior over time.

While these highly-specific search equations can provide results, they are not replicable across risk mitigation and regulatory compliance departments. Each search is completed by an individual, with each of the multiple components determined individually. In this manner, each professional will most likely form unique search equations. This makes it difficult for a search to be auditable across a department, and nearly impossible for it to be replicable.

Using GOST, the initial screening process is not only almost instantaneous, but also replicable. Every entity is screened in the same method, eliminating search method inconsistencies like the one illustrated above.

Additionally, the platform can place large populations into a state of continuous vetting. Continuous vetting is a state in which GOST periodically screens at an interval set by managers with the intent of looking for changes in patterns of behavior over time.

Since continuous vetting is automated, GOST will immediately alert users to important changes in an entity's online presence that merit scrutiny. This enables customer refresh at a faster pace, or across a large population. These continuous searches represent snapshots of the internet that GOST captures, eliminating human bias in the selection decision about what information to retain. Professionals can rest assured knowing that the same way a bank continuously evaluates suspicious transactions and interactions, GOST does the same with information that exists outside of the bank.

Value Proposition of GOST® at Your Institution

General-purpose search engines fail to meet the needs of the screening and vetting mission in four primary ways.

1. Traditional search is not optimized for the returns needed by professionals to identify illicit actors,
2. Manual searching takes a massive amount of time, with little return,
3. General-purpose search engines are not efficient, and
4. KYC and due diligence practices are not auditable or replicable using general purpose search tools.

GOST offers a more effective solution for less cost. Using machine learning and artificial intelligence capabilities, GOST optimizes returns for the user by re-indexing the internet based on specific behaviors. GOST is fast and return time remains consistent, while Ranking and Reliability scores result in an 88 percent efficiency increase over manual search tools. Lastly, GOST is replicable and auditable, unlike general purpose search tools.

What Can GOST® Do For You

In addition to the added benefits of strong security, data encryption, non-attribution of searches, and continuous vetting, using GOST allows humans to complete the critical analysis and judgement tasks of evaluating the information, instead of the rote, mechanical task of search and retrieval.
By allowing professionals to review a higher volume of cases, GOST enables them to look more thoroughly at entities that would normally be ignored, or never considered for screening, because of the massive cost of manual review.

---

[2] Back-of-the envelope calculations suggest it would take anywhere between 480-29,000 FTE to do repeatable and auditable search and retrieval on 100 million entities using Google Advanced Search and permutations of search inputs.