

Mastering search: the semantic edge

The US government has a search problem. It generates billions of pages every year, from tax audits and budget reports to the president's daily intelligence brief. Classified data alone is growing at an estimated rate of 50 million documents per year.¹

When searching for information, no one has a tougher job than intelligence analysts. Crucial information is scattered among data silos, each with their own specific search tools. Locating the right information is the most time-intensive part of their day, leaving a thin sliver of time to read, synthesize, and write.

Artificial intelligence can help analysts find the information they need, particularly with a new technique called semantic search. But to understand the analysts' dilemma, one must understand the tool they use today: boolean search.

A search nightmare

Crafting a good boolean query is a painfully learned art. Consider this boolean query used to find relevant documents about Ukraine's relationship with NATO member states:

```
(("Ukraine foreign relations"-5 OR "Ukraine diplomatic relations"-5 OR "Ukraine diplomacy"-5 OR "Ukraine international affairs"-5 OR "Ukraine foreign policy"-5 OR "Ukraine international diplomacy"-5 OR "Ukraine-Russia conflict"-5 OR "Ukraine military cooperation"-5 OR "Ukraine security cooperation"-5 OR "Ukraine defense cooperation"-5 OR "Ukraine military equipment"-5 OR "Ukraine arms trade"-5 OR "Ukraine sanctions"-5 OR "Ukraine peace negotiations"-5 OR "Ukraine conflict resolution"-5 OR "Ukraine ceasefire agreement"-5 OR "Ukraine regional security"-5 OR "Ukraine alliance building"-5 OR "Ukraine strategic partnership"-5 OR "Ukrainian foreign relations"-5 OR "Ukrainian diplomatic relations"-5 OR "Ukrainian diplomacy"-5 OR "Ukrainian international affairs"-5 OR "Ukrainian foreign policy"-5 OR "Ukrainian international diplomacy"-5 OR "Ukrainian military cooperation"-5 OR "Ukrainian security cooperation"-5 OR "Ukrainian defense cooperation"-5 OR "Ukrainian military equipment"-5 OR "Ukrainian arms trade"-5 OR "Ukrainian sanctions"-5 OR "Ukrainian peace negotiations"-5 OR "Ukrainian conflict resolution"-5 OR "Ukrainian ceasefire agreement"-5 OR "Ukrainian regional security"-5 OR "Ukrainian alliance building"-5 OR "Ukrainian strategic partnership"-5) AND (North Atlantic Treaty Organization) AND (United States OR United Kingdom OR Ukraine OR England OR "Britain" OR France OR Germany OR Canada OR Italy OR Spain OR Poland OR Netherlands OR Belgium OR Greece OR Turkey OR Denmark OR Norway OR Portugal OR Hungary OR Czech Republic OR Slovakia OR Romania OR Bulgaria OR Estonia OR Latvia OR Lithuania OR Slovenia OR Croatia OR Albania OR Montenegro OR North Macedonia OR Bosnia And Herzegovina OR Serbia OR Kosovo OR Cyprus OR Washington, D.C. OR London, England OR Paris, Ile-De-France OR Berlin OR Ottawa, Ontario OR Rome, Latium OR Madrid OR Warsaw, Mazovia OR Amsterdam, North Holland OR Brussels, Brussels Capital OR Athens, Attica OR Ankara OR Copenhagen, Capital Region OR Oslo OR Lisbon OR Budapest OR Prague, Praha OR Bratislava, Bratislavsky Region OR Bucharest, Bucuresti OR Sofia, Sofia-Capital OR Tallinn, Harjumaa OR Riga OR Vilnius OR Ljubljana OR Zagreb, City Of Zagreb OR Tirana, Tirane OR Podgorica OR Skopje, Karpos OR Sarajevo, Federation Of B&H OR Belgrade, Central Serbia OR Pristina OR Nicosia) AND (Vladimir Zelensky OR Volodymyr Zelenskyy OR Volodymyr Zelenskyy OR Volodymyr Zelenskyy OR Russian Stefanchuk OR Ivan Bakanov OR Dmytro Razumkov OR "Oleksandr Danyliuk" OR Ihor Zhovkva OR Oleksiy Arestovych OR "advisor" OR "adviser") AND NOT (title:(live updates) OR title:(ukraine latest))
```

In boolean search, an analyst can't simply search for "Ukraine foreign relations" because that will retrieve only documents with those exact words. Intelligence analysts spend hours honing their boolean searches over and over until, finally, they get the results they need. Good booleans get passed down from analyst to analyst like treasured family heirlooms.

Dream search

What if our analyst could search with a simple plain language query? She simply types "What are Ukraine's relationships with NATO member states?" and the documents that come back are about exactly that, regardless of whether they include those exact words. Even better, the most relevant parts of the retrieved documents are highlighted. This is semantic search. Rather than having the user define what counts as relevant, the search system "understands" the meaning of the user's query and retrieves documents with truly relevant content.

In semantic search, every chunk of text in your documents and your query is assigned an address in a high-dimensional "semantic space". The system then gathers up documents that are in the same semantic neighborhood as her query. The closer they are in the semantic space, the higher the relevancy. The breakthrough that makes this possible is an AI tool called a language model.

Voilà! Now you understand the core concept behind semantic search.

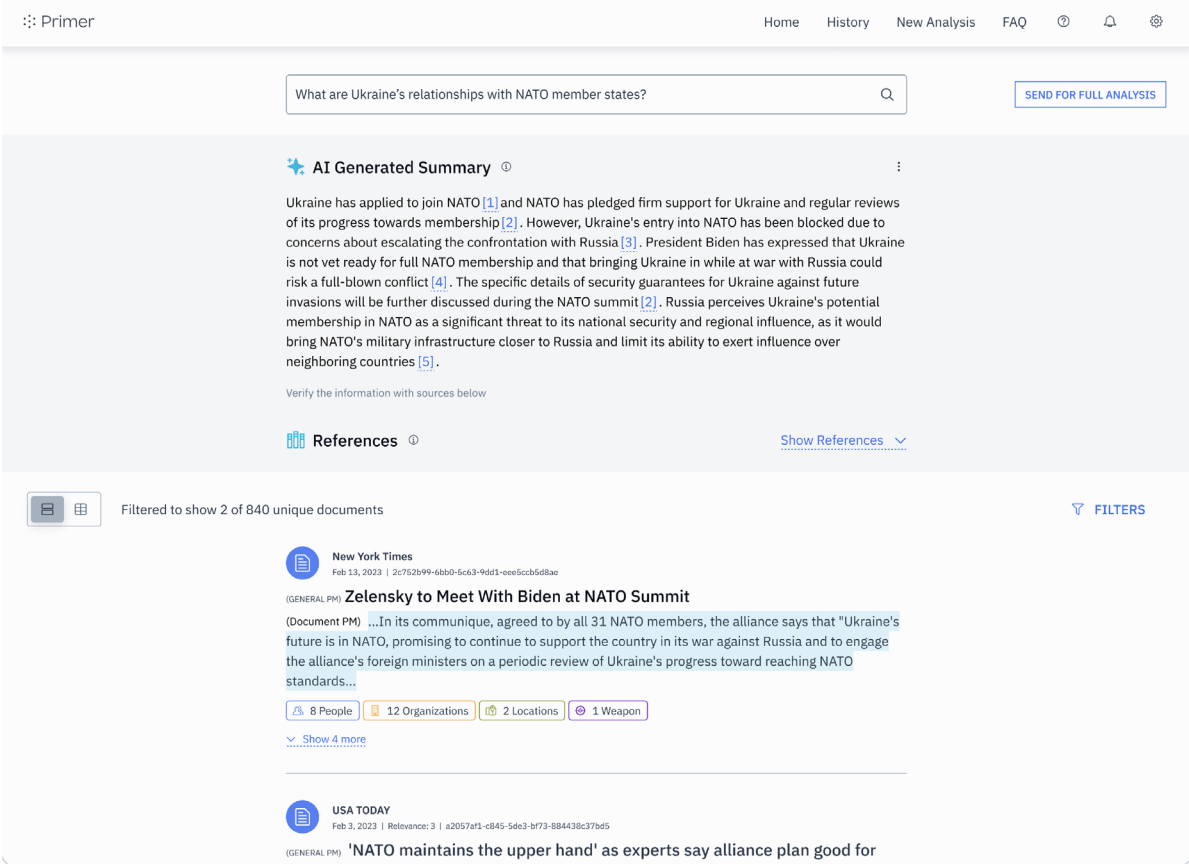
Even better search

What if instead of search, our analyst could just ask the question that motivated her search in the first place?

This is semantic question-answering. The trick is to add a large language model to the end of the search flow. Taking the user's question and search results, the model generates a direct answer to the user's question, including in-line citations to sources.

Primer has semantic question-answering deployed in products serving US Defense and IC customers. To the right is a screenshot showing it in action.

With the help of AI, analysts can now engage with the vast corpus of government data in a more natural way, unburdened by booleans. This translates to faster analysis and faster decision-making.



1. <https://www.nytimes.com/2023/01/27/briefing/classified-documents-government.html>

About Primer

Primer exists to make the world a safer place. We believe that AI is critical to accelerating intelligence and decision cycles that keep us safe. Our AI technology is deployed by the U.S. Government, strategic allies, and Fortune 100 companies to extract timely insight and decision advantage from massive datasets. Primer has offices in Arlington, Virginia and San Francisco, California. For more information, please visit primer.ai.

