
The funnel of relevance

Abhishek Tiwari 

Citation: *A. Tiwari*, "The funnel of relevance", Abhishek Tiwari, 2024.
[doi:10.59350/2dpr6-vrf72](https://doi.org/10.59350/2dpr6-vrf72)

Published on: September 15, 2024

Anyone who has worked with advanced search systems knows the intricacy of boolean operators. AND, OR, NOT - these seemingly simple words become powerful tools in the hands of search specialists. Add in proximity operators like WITH and NEAR, throw in some parentheses for nesting, and you've got yourself a query that can pinpoint exactly what you're looking for... or so we thought.

In practice, these queries can become unwieldy beasts. Imagine a search query for a government minister that spans multiple pages. Now imagine having to update that query every time there's a cabinet reshuffle. It's a time-consuming process that requires constant attention and expertise.

This approach worked well in the era of traditional media - newspapers, magazines, and broadcast news. These sources had standardized structures and predictable content. But in the world of online and social media, where content is diverse, unstructured, and ever-changing, even the most meticulously crafted boolean query can fall short.

The funnel of relevance

Enter the funnel of relevance. This approach, supported by Natural Language Processing (NLP) technology, offers a more nuanced and adaptable way to determine relevance. Instead of relying solely on rigid boolean logic, the funnel of relevance processes information in stages, each refining the results further.

The process begins with a broad capture using simpler boolean queries. This casts a wide net, ensuring high recall. The captured content is then analysed using NLP techniques, involving the identification of key entities, understanding of sentiment, enriching with audience data, and grasping of context. Advanced algorithms then classify and rank the content based on its relevance to the user's needs based on keywords and attribute-based filtering. Finally, the results are further refined, balancing precision and recall to deliver the most pertinent information.

This funnel approach allows for a more flexible and comprehensive understanding of relevance. It can adapt to the nuances of language and context that are prevalent in online and social media platforms, capturing information that might be missed by traditional boolean queries.

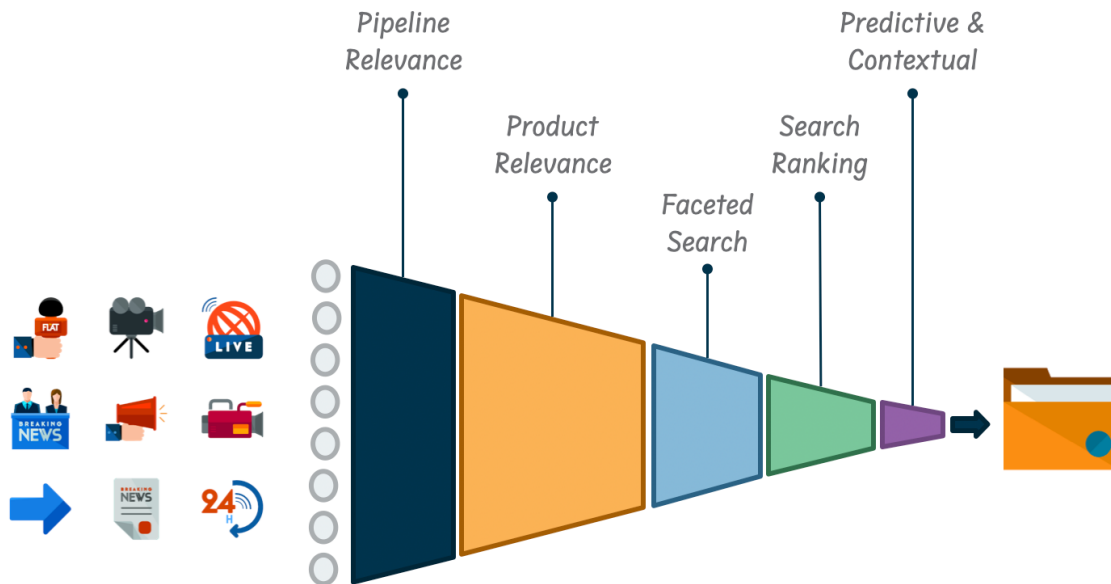


Figure 1: Funnel of relevance to balance the precision and recall

Enriching content with audience and web ranking data

An essential part of the funnel of relevance is the enrichment of content using audience and web ranking information. This step adds another layer of sophistication to the relevance determination process.

For audience data, the system incorporates metrics such as readership, viewership, or social media engagement. This information helps gauge the reach and impact of a piece of content. For instance, an article from a niche industry publication might be highly relevant in terms of content, but its limited readership could affect its overall importance in the media landscape.

Web ranking information, on the other hand, provides insights into the authority and credibility of the source. This can include factors like domain authority, page rank, and inbound link quality. A story from a highly respected news outlet with a strong web presence might be given more weight than a similar story from a less established source.

The system also considers social media metrics, such as shares, likes, and comments. This data offers a real-time measure of a piece of content's resonance with audiences. A viral social media post, for example, might be deemed highly relevant due to its widespread engagement, even if it comes from a non-traditional source.

By integrating these additional data points, the funnel of relevance can provide a more holistic view of a piece of content's significance. It's not just about what the content says, but also about who's reading it, how it's being shared, and where it sits in the broader web ecosystem.

"iPhone 14" AND ("iPhone 14 Apple"~10 OR "iPhone 14 iOS"~10 OR
 "iPhone 14 smartphone*"~10 OR "iPhone 14 Verizon"~10)

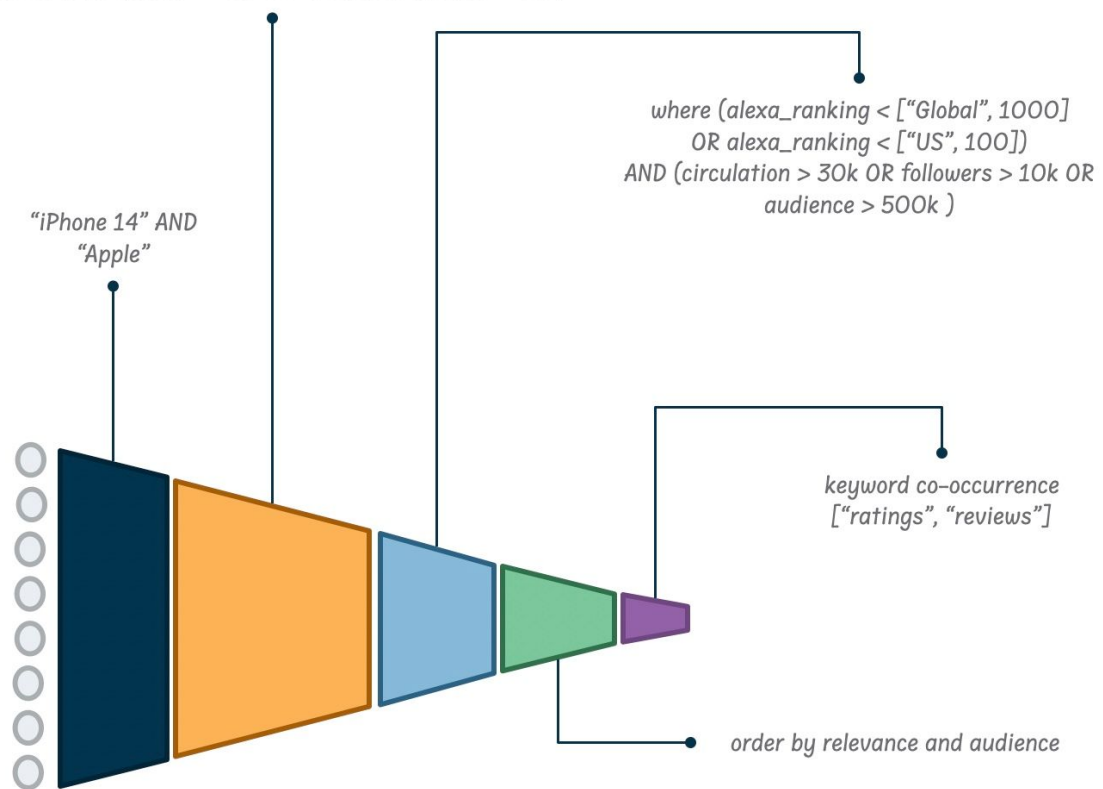


Figure 2: Replacing complex booleans with simpler boolean queries and attribute based filtering

Balancing precision and recall

One of the key advantages of the funnel of relevance is its ability to balance precision and recall. High precision ensures that the results are highly relevant, but can sometimes miss important information. High recall captures a wide range of potentially relevant information, but may include some irrelevant results. The funnel approach allows for a careful calibration of this balance. By processing information through multiple stages and enriching it with audience and web ranking data, it can maintain high recall in the initial stages while progressively increasing precision.

The impact on media intelligence

For media intelligence professionals, this approach offers several significant benefits. The system can quickly adapt to changes in the media landscape without requiring extensive query rewrites. It can effectively capture relevant information across a wide range of sources, from traditional press to social media, while also considering the impact and reach of that content. The automated nature of the process allows for real-time analysis and insights, incorporating not just the content itself but also its performance in the digital sphere. Perhaps most importantly, search specialists can focus on strategy and analysis rather than constantly updating complex queries.

As the media landscape continues to evolve, so too must our approaches to search and relevance. The funnel of relevance, powered by NLP, machine learning, and enriched with audience and web ranking data, offers a promising path forward. It combines the precision of traditional boolean search with the adaptability needed for today's dynamic media environment, providing media intelligence professionals with a powerful tool for navigating the information age. By embracing this approach, we can ensure that we're not just finding information, but finding the right information at the right time, and understanding its true impact in the broader media ecosystem.