Semantic Overlap Vs. Density: Finding The Balance That Wins Retrieval via @sejournal, @DuaneForrester

Marketers today spend their time on keyword research to uncover opportunities, closing content gaps, making sure pages are crawlable, and aligning content with E-E-A-T principles. Those things still matter. But in a world where generative AI increasingly mediates information, they are not enough.

The difference now is retrieval. It doesn’t matter how polished or authoritative your content looks to a human if the machine never pulls it into the answer set. Retrieval isn’t just about whether your page exists or whether it’s technically optimized. It’s about how machines interpret the meaning inside your words.

That brings us to two factors most people don’t think about much, but which are quickly becoming essential: *semantic density* and *semantic overlap*. They’re closely related, often confused, but in practice, they drive very different outcomes in GenAI retrieval. Understanding them, and learning how to balance them, may help shape the future of content optimization. Think of them as part of the new on-page optimization layer.

Image Credit:: Duane Forrester

Semantic density is about meaning per token. A dense block of text communicates maximum information in the fewest possible words. Think of a crisp definition in a glossary or a tightly written executive summary. Humans tend to like dense content because it signals authority, saves time, and feels efficient.

Semantic overlap is different. Overlap measures how well your content aligns with a model's latent representation of a query. Retrieval engines don't read like humans. They encode meaning into vectors and compare similarities. If your chunk of content shares many of the same signals as the query embedding, it gets retrieved. If it doesn't, it stays invisible, no matter how elegant the prose.

This concept is already formalized in natural language

processing (NLP) evaluation. One of the most widely used measures is **BERTScore** (https://arxiv.org/abs/1904.09675), introduced by researchers in 2020. It compares the embeddings of two texts, such as a query and a response, and produces a similarity score that reflects semantic overlap. BERTScore is not a Google SEO tool. It's an open-source metric rooted in the BERT model family, originally developed by Google Research, and has become a standard way to evaluate alignment in natural language processing.

Now, here's where things split. Humans reward density. Machines reward overlap. A dense sentence may be admired by readers but skipped by the machine if it doesn't overlap with the query vector. A longer passage that repeats synonyms, rephrases questions, and surfaces related entities may look redundant to people, but it aligns more strongly with the query and wins retrieval.

In the keyword era of SEO, density and overlap were blurred together under optimization practices. Writing naturally while including enough variations of a keyword often achieved both. In GenAI retrieval, the two diverge. Optimizing for one doesn't guarantee the other.

This distinction is recognized in evaluation frameworks already used in machine learning. BERTScore, for example, shows that a higher score means greater alignment with the intended meaning. That overlap matters far more for retrieval than density alone. And if you really want to deep-dive into LLM evaluation metrics, this article is a great resource.

Generative systems don't ingest and retrieve entire webpages. They work with chunks. Large language models are paired with vector databases in retrieval-augmented generation (RAG) systems. When a query comes in, it is converted into an embedding. That embedding is compared against a library of content embeddings. The system doesn't ask "what's the best-written page?" It asks "which chunks live closest to this

query in vector space?”

This is why semantic overlap matters more than density. The retrieval layer is blind to elegance. It prioritizes alignment and coherence through similarity scores.

Chunk size and structure add complexity. Too small, and a dense chunk may miss overlap signals and get passed over. Too large, and a verbose chunk may rank well but frustrate users with bloat once it’s surfaced. ***The art is in balancing compact meaning with overlap cues, structuring chunks so they are both semantically aligned and easy to read once retrieved.*** Practitioners often test chunk sizes between 200 and 500 tokens and 800 and 1,000 tokens to find the balance that fits their domain and query patterns.

Microsoft Research offers a striking example. In a 2025 study analyzing 200,000 anonymized Bing Copilot conversations, researchers found that information gathering and writing tasks scored highest in both retrieval success and user satisfaction. Retrieval success didn’t track with compactness of response; it tracked with overlap between the model’s understanding of the query and the phrasing used in the response. In fact, in 40% of conversations, the overlap between the user’s goal and the AI’s action was asymmetric. Retrieval happened where overlap was high, even when density was not. Full study here.

This reflects a structural truth of retrieval-augmented systems. Overlap, not brevity, is what gets you in the answer set. Dense text without alignment is invisible. Verbose text with alignment can surface. The retrieval engine cares more about embedding similarity.

This isn’t just theory. Semantic search practitioners already measure quality through intent-alignment metrics rather than keyword frequency. For example, Milvus, a leading open-source vector database, highlights overlap-based metrics as the right

way to evaluate semantic search performance. Their reference guide emphasizes matching semantic meaning over surface forms.

The lesson is clear. Machines don’t reward you for elegance. They reward you for alignment.

There’s also a shift in how we think about structure needed here. Most people see bullet points as shorthand; quick, scannable fragments. That works for humans, but machines read them differently. To a retrieval system, a bullet is a structural signal that defines a chunk. What matters is the overlap inside that chunk. A short, stripped-down bullet may look clean but carry little alignment. A longer, richer bullet, one that repeats key entities, includes synonyms, and phrases ideas in multiple ways, has a higher chance of retrieval. In practice, that means bullets may need to be fuller and more detailed than we’re used to writing. Brevity doesn’t get you into the answer set. Overlap does.

If overlap drives retrieval, does that mean density doesn’t matter? Not at all.

Overlap gets you retrieved. Density keeps you credible. Once your chunk is surfaced, a human still has to read it. If that reader finds it bloated, repetitive, or sloppy, your authority erodes. The machine decides visibility. The human decides trust.

What’s missing today is a composite metric that balances both. We can imagine two scores:

Semantic Density Score: This measures meaning per token, evaluating how efficiently information is conveyed. This could be approximated by compression ratios, readability formulas, or even human scoring.

Semantic Overlap Score: This measures how strongly a chunk aligns with a query embedding. This is already approximated by tools like BERTScore or cosine similarity in vector space.

Together, these two measures give us a fuller picture. A piece of content with a high density score but low overlap reads beautifully, but may never be retrieved. A piece with a high overlap score but low density may be retrieved constantly, but frustrate readers. The winning strategy is aiming for both.

Imagine two short passages answering the same query:

Dense version: "RAG systems retrieve chunks of data relevant to a query and feed them to an LLM."

Overlap version: "Retrieval-augmented generation, often called RAG, retrieves relevant content chunks, compares their embeddings to the user's query, and passes the aligned chunks to a large language model for generating an answer."

Both are factually correct. The first is compact and clear. The second is wordier, repeats key entities, and uses synonyms. The dense version scores higher with humans. The overlap version scores higher with machines. Which one gets retrieved more often? The overlap version. Which one earns trust once retrieved? The dense one.

Let's consider a non-technical example.

Dense version: "Vitamin D regulates calcium and bone health."

Overlap-rich version: "Vitamin D, also called calciferol, supports calcium absorption, bone growth, and bone density, helping prevent conditions such as osteoporosis."

Both are correct. The second includes synonyms and related concepts, which increases overlap and the likelihood of retrieval.

This Is Why The Future Of

Optimization Is Not Choosing Density Or Overlap, It’s Balancing Both

Just as the early days of SEO saw metrics like keyword density and backlinks evolve into more sophisticated measures of authority, the next wave will hopefully formalize density and overlap scores into standard optimization dashboards. For now, it remains a balancing act. If you choose overlap, it’s likely a safe-ish bet, as at least it gets you retrieved. Then, you have to hope the people reading your content as an answer find it engaging enough to stick around.

The machine decides if you are visible. The human decides if you are trusted. Semantic density sharpens meaning. Semantic overlap wins retrieval. The work is balancing both, then watching how readers engage, so you can keep improving.

More Resources:

This post was originally published on Duane Forrester Decodes.

Featured Image: CaptainMCity/Shutterstock

https://www.searchenginejournal.com/semantic-overlap-vs-density-finding-the-balance-that-wins-retrieval/554251/