What Is Latent Semantic Indexing & Why It Won't Help Your SEO by @clarkboyd

You have most likely heard of Latent Semantic Indexing (LSI).

The term has been appropriated by some in the SEO industry, including some well-known influencers. They claim that its application is key for organic search success.

In fact, a search for [latent semantic indexing] on Google reveals well-known sites such as HubSpot claiming that a little LSI will give your SEO a boost and take it to the next level.

So, what exactly is LSI?

And is there any actual *evidence* that LSI can help your SEO performance?

To answer these questions, let's explore the origins of LSI and what it does (or rather, doesn't) mean for SEO in 2018.

There is a surprising amount of misinformation about this topic out there; this article debunks the theory that using "LSI keywords" will positively impact your SEO and suggest some more effective strategies you can apply instead.

What is Latent Semantic Indexing?

Latent semantic indexing, sometimes referred to as latent semantic analysis, is a mathematical method developed in the late 1980s to improve the accuracy of information retrieval. It uses a technique called singular value decomposition to scan unstructured data within documents and identify relationships between the concepts contained therein.

In essence, it finds the hidden (latent) relationships between words (semantics) in order to improve information understanding (indexing).

It provided a significant step forward for the field of text comprehension as it accounted for the contextual nature of language.

Earlier technologies struggled with the use of synonyms that characterizes natural language use, and also the changes in meanings that come with new surroundings.

For example, the words ‘hot’ and ‘dog’ may seem easy to understand, but both have multiple definitions based on how they are used. Put both of them together and you have a whole new concept altogether.

So how can we train a machine to adapt to these nuances?

This is a problem that has troubled great minds for centuries and LSI has helped computers to start understanding language in use.

It works best on static content and on small sets of documents, which was great for its initial purposes. LSI also allows documents to be clustered together based on their thematic commonalities, which was a very useful capability for early search engines.

Latent semantic indexing can be summarized as follows:

- A technology developed in the late 1980s for information retrieval, in response to earlier technologies that could not understand synonymy or polysemy.
- A specific approach that tries to grasp the underlying structure of meaning in language.
- Capable of inducing from these findings the hierarchical categories into which terms and concepts fall.
- Originally useful for working on small sets of static documents.

Latent Semantic Indexing & SEO

If we take away from the previous section of this article that LSI would allow a search engine to understand synonyms, it follows logically that using synonyms throughout a document could, in fact, help a search engine to understand your content.

And if a search engine can understand the content, it can index and rank it for your target queries, too.

Moreover, using synonyms may strengthen the thematic relevance of the overall piece of content, which must be good for SEO, right?

Distilled down to its purest essence, the proposition is that including synonyms for your target keywords within a piece of content will help SEO performance. These are sometimes even called “LSI keywords.”

The problem?

There is no evidence to support this.

LSI has potentially played a part in the development of early search engines.

As Roger Montti put it: “LSI is training wheels for search engines.”

But there is no reason to believe that this has been the case at any recent time.

Nonetheless, there has been an assumption in some quarters that the respective paths of Google and LSI have converged evermore over time, when in fact one could argue with greater conviction that the contrary has occurred.

Undoubtedly, Google wants to understand the context of any piece of content. The field of semantics (the study of meaning in language) is a fundamental part of this approach.

However, it is quite an assumptive leap to conclude that the presence of “semantic” in both “latent semantic indexing” and “semantic search” reveals some direct and underlying link between the two.

There is good reason to believe that Google has evolved far beyond this and uses much more sophisticated, machine learning-led technology for document indexation and information retrieval.

JR Oakes delivered an enlightened and enlightening presentation at TechSEO Boost in late 2017, which dispels some myths about modern information retrieval and replaces them with some evidence-based approaches to understanding how Google works:

And yet, as Google has developed its ability to deliver semantic search through new technologies, some in the industry have found cause to promote LSI even more.

Ironically, this is exactly the kind of linguistic confusion that Google is trying to clean up with its semantic search technology.

What Should You Focus on Instead of LSI?

Optimizing content for organic search visibility has evolved in line with Google’s advancements.

Equally, search engines still have challenges when trying to understand the meaning of words in context.

There are better ways to achieve this rather than by adding “LSI keywords,” though.

First of all, structured data is an essential component of a modern SEO strategy. By labeling data clearly, we can help search engines to index and serve our content in rich results across multiple devices.

The concept of co-occurrence is also increasingly important, as search engines identify words that are typically used together to understand how they relate and interact with each other to alter meaning.

We can identify some of these terms by researching the products or services that we wish to promote and including accurate terminology.

Rather than muddy the waters with synonyms, we should all aim for greater clarity.

Bill Slawski is an excellent resource on this topic. Make sure to read his excellent post on SEO by the Sea: Does Google Use Latent Semantic Indexing? (Spoiler alert: no.)

Also, his presentation below provides a great balance of theory and practice:

Stop the LSI Madness!

LSI is a specific technology that provided a leap forward in the field of information indexing and retrieval.

In the 1980s.

How much tech from the 1980s do *you* still use?

Exactly.

Search engines, by their very nature, are in the business of indexing and retrieval.

However, there is no proof that Google uses LSI.

To believe so is an archetypal example of flawed syllogistic reasoning.

There is an argument that, even if there is no evidence that Google uses LSI, adding synonyms throughout your content won't do any damage. As such, it's worth giving it a try, since there also isn't hard evidence that Google definitely doesn't use LSI.

The counterargument runs that many activities fall into the same category, but that doesn't give them any merit.

You could change all your text to Comic Sans font, just in case Google gives a 1 percent ranking boost for sites with the bravery to use it.

Actually, please don't do that!

Perhaps its seductive appeal lies in its combination of a scientific name and an unscientific application. It sounds like it's very advanced, but it really comes down to adding synonyms and related words, which anyone can do.

Furthermore, while the theory is nigh-on impossible to verify,

it’s also tricky to falsify.

All the evidence points to the fact that any effort expended chasing “LSI keywords” would be better spent understanding the true functioning of semantic search instead.

There are practical ways to apply this knowledge, too.

Using structured data and understanding how co-occurrence can benefit content indexing will be of much more value than the addition of LSI terms.

Conclusion

Many smart people have been misled about LSI.

If we want to foster an industry of well-informed SEO professionals and digital marketers (and I very much assume that we do), we need to focus on building trust through the sharing of evidence-based findings.

Promoting the concept of “LSI keywords” may not be doing anyone any damage in a quantitative sense, but it can start to erode trust and lead the way to more fallacies in future.

More SEO Resources:

Image Credits
In-Post image: Pixabay

http://tracking.feedpress.it/link/13962/8472873