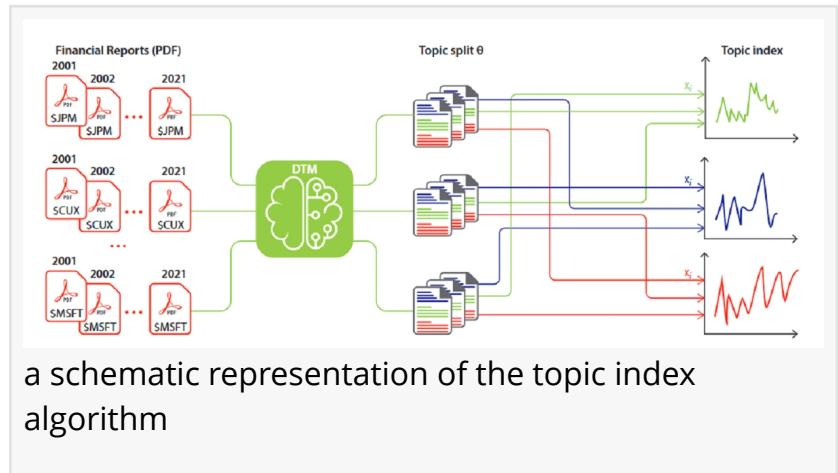


Unsupervised generation of tradable topic indices through textual analysis

FAYETTEVILLE, GA, UNITED STATES,
April 27, 2025 /EINPresswire.com/ --

Scientists developed a new way of investing in stocks using natural language processing. Using dynamic topic modeling, a variant of Latent Dirichlet Allocation, the new model uncovers hidden risk factors directly from company reports and translates them into tradable indices with minimal human oversight. Investors can now trade these risk factors directly and track industry trends using only the information contained in words.



A recent article in *The Journal of Finance and Data Science* introduces an innovative method for constructing investment instruments directly from financial reports — without the need for human intervention.

This novel approach employs dynamic topic modeling ([DTM](#)), a variant of Latent Dirichlet Allocation (LDA), to analyze annual and quarterly reports from companies, uncovering hidden risk factors and translating them into tradable indices.

"The beauty of this method lies in its simplicity and transparency; it combines several established algorithms to achieve what previously was not possible," says co-author Marcel Lee. "By automating the process, we eliminate biases and provide a cost-effective alternative to traditional index construction."

This unsupervised technique automatically selects optimal parameters, discovering implicit risk factors through the semantic analysis of corporate publications, thereby creating a new class of investment instruments — thematic indices.

The study describes the model's capacity to dynamically track economic and industrial trends, illustrating that sectors considered static are in reality constantly evolving. This method captures the fluid nature of industries more accurately than traditional static classifications like GICS or

ICB.

"We're observing the industrial landscape through a much sharper and multicoloured lens, enabling investors to tap into nuanced market themes and risk factors previously inaccessible," adds co-author Alan Spark.

In several cases, the research demonstrated that these newly created thematic indices closely mimic established indices, yet are derived without the predefined biases of manual classification systems. "This not only paves the way for a more unbiased benchmarking tool but also reveals industry trends and vocabulary shifts over time, offering a fresh perspective on sectoral dynamics," says Lee.

One notable challenge acknowledged by the researchers is the approach's reliance on a 'bag-of-words' model, which, while instrumental in parsing large datasets, overlooks the nuanced relationships between words. "Future iterations of this work aim to incorporate more complex models that capture these subtleties, potentially enhancing the predictive power of thematic indices on corporate actions and industry shifts," shares Spark.

References

DOI

[10.1016/j.jfds.2025.100149](https://doi.org/10.1016/j.jfds.2025.100149)

Original Source URL

<https://doi.org/10.1016/j.jfds.2025.100149>

Lucy Wang

BioDesign Research

[email us here](#)

This press release can be viewed online at: <https://www.einpresswire.com/article/803674587>

EIN Presswire's priority is source transparency. We do not allow opaque clients, and our editors try to be careful about weeding out false and misleading content. As a user, if you see something we have missed, please do bring it to our attention. Your help is welcome. EIN Presswire, Everyone's Internet News Presswire™, tries to define some of the boundaries that are reasonable in today's world. Please see our Editorial Guidelines for more information.

© 1995-2025 Newsmatics Inc. All Right Reserved.