



Subject: FINRA Machine-Readable Rulebook

Date: Feb. 21, 2023

To: pubcom@finra.org

From
Evgeny Likhoded
CEO and Founder, Clausematch
30 Churchill Place
London, E14 5EU, UK

To whom it may concern:

The purpose of this letter is to provide feedback and support towards the FINRA Machine Readable Rulebook initiative which highlights the ability for firms to more effectively and efficiently consume regulatory rules that are relevant and applicable.

The need for uniform digitization in the regulatory environment has never been higher, as companies see more and more new regulations come to light every day. Clausematch is also focusing on this push for digitization across the regulatory value chain: this month, it [published in open source its knowledge graph](#), which will fundamentally transform the approach to the very understanding of regulation and how it can be delivered.

The Clausematch knowledge graph has been years in the making with other regulators, and its framework could prove very beneficial to FINRA, as the agency is already on the right track with its release of the Machine Readable Rulebook initiative. Both initiatives apply machine learning to read and consume regulations.

We have identified five areas of improvement of the FINRA Machine Readable Rulebook initiative that we believe should be taken into consideration if further expansion is deemed appropriate:

1. Tagging is on the rule level: this poses an issue from the Machine-Learning point of view, as it is challenging to automate tagging with such data. Ideally, tagging should be in-text for many tags, so as to include personally identifiable information.
2. Releasing tagged rules: this is a step in the right direction. Ideally, everyone will benefit far more from the release of the tagging model made by a regulating agency such as FINRA. However, with just 40 tagged rules, this does not provide enough data to train the models.
3. Knowledge graph is a way to organize tagged content into a structure convenient for machine learning models to consume. We suggest in-text tagging is the best way to

tag regulations. You can review our scientific paper, explaining the steps and effects, [here](#).

4. Make FINRA initiative available as SQL/data frame dump: it is understandable and standard that FINRA made its rulebook initiative available through API, but it would be better to have it through data frame, as it makes it easier to be analyzed from a data science perspective.
5. Comprehensive taxonomy by FINRA is key: this allows for the standardization of rules, which allows companies like the ones Clausematch assists to map rules to policies effectively. Without it, there is a taxonomy discrepancy, on top of the meaning discrepancy, during the mapping process.

In summary, the financial services industry will benefit immensely from the release of a machine learning model built by a regulator. But there needs to be greater capabilities of tagging and structuring any firm's internal text. For that, more clarification will be needed.

Again, it is imperative to note that Clausematch has been working for years on a similar initiative, the knowledge graph, with ADGM, and has released the graph in open source. This is a similar experiment, on a high level, between a regulator and our company with the hopes of pushing regulation into the digitized world. The technological tools, along with visualizations, are available for anyone to test, and we encourage FINRA representatives to do just that.

This is an exciting time to be in the regulatory technology space, and we are thrilled FINRA is diving deep into digitization of regulation. We welcome any collaboration efforts or request for expertise.

Regards,

Evgeny Likhoded