HYDRA: An Open-Source Platform for Automated Mining and Visualization of Regulatory Data using Natural Language Processing and Machine Learning

Nathan S. Ivey, PhD (CDRH/OPEQ) and Pavel Takmakov, PhD (CDRH/OSEL)

Introduction

About 80% of the approximately 190,000 medical devices marketed in the United States are cleared through the 510(k) Program. Our team was interested to investigate whether we could link all medical devices that have been cleared in the 510(k) Program and if we could then query and stratify across all of the public data provided in medical device summaries.

The idea is to exploit the product classification scheme as well as the previously uncharacterized predicate device linkages to develop a foundational architecture from which more advanced analytics can eventually be run.

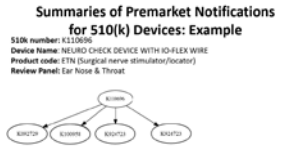

Figure 1. An image showing an example of medical device predicate relationships.

HYDRA is a broad platform that has the potential to enhance The regulation of medical devices. As its underpinning, HYDRA utilizes a foundation that exploits the linkability of the medical device universe through both product classification (product codes) and file numbers (Kxxxxxx) to foster unprecedented analytics

Materials & Methods

HYDRA is built on an open-source Python-based platform, and has the capability to read and extrapolate full text of most of the public FDA-generated data surrounding medical devices. Once fully realized, it will build on the foundation demonstrated here to analyze / query across multiple possible data streams. For example, we may include PMA devices

HYDRA utilizes trainable algorithms to allow us to identify emerging technologies in product types or to help us to understand how and why product evolve over time. Once fully validated, HYDRA will have the capability to then apply machine learning algorithms to 'hunt' within the public FDA data streams for important targets.

Some current applications:

-Instantly establishing medical device lineages
-Recategorizing and understanding all devices in a generic device type or regulation based on queried attributes.
-Possible automated classification of new devices
-Potentially understand risk/benefit in certain emerging or established technologies (e.g., a specific material).

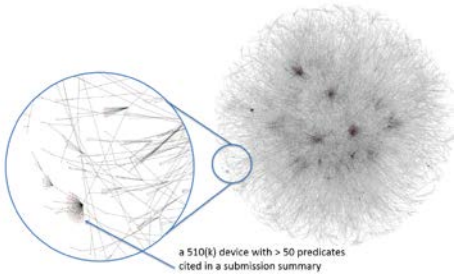

Figure 2. An image showing the interconnected relationships between medical devices cleared under the 510(k) pathway

"In the next five years, it will be entirely feasible to run sophisticated queries and validated analytics across all public unstructured medical device data in FDA's universe."

Scan the QR code to access additional information online

Results

We were able to link about 55,000 medical devices, and allow us to query across the unstructured data of the published summaries in an intuitive and visually appealing manner.

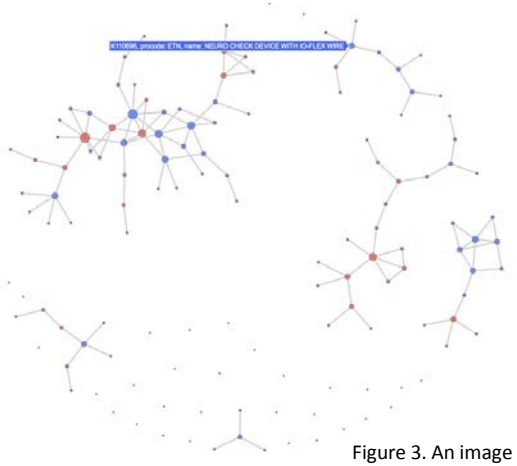

Figure 3. An image showing an example of medical device predicate relationships within a product code.

Conclusions

We were successful at linking about a third of all marketed medical devices and their associated data. HYDRA needs additional development, validation, and error correction in order to realize its fullest potential to enhance out overall mission of improving public health.

While its development is meant to enhance the ability of review staff to understand the medical device universe, we hope to eventually publish the code and offer this agile platform to the public. In this way, we hope to deliver a system to crawl and analyze data rapidly, delivering it in a visually appealing way and enable broad strata of users to navigate complex data flows.

References

Kumsa, D. , Steinke, G. K., Molnar, G. F., Hudak, E. M., Montague, F. W., Kelley, S. C., Untereker, D. F., Shi, A. , Hahn, B. P., Condit, C. , Lee, H. , Bardot, D. , Centeno, J. A., Krauthamer, V. and Takmakov, P. A. (2018), Public Regulatory Databases as a Source of Insight for Neuromodulation Devices Stimulation Parameters. Neuromodulation: Technology at the Neural Interface, 21: 117-125. doi:10.1111/ner.12641