



Kaona: Deep Searching and Curating Safety Reporting Systems

Hailey Pham
University of California,
Berkeley

Christopher Hong
Oregon State University

Carlos Paradis, Bryan Matthews
KBR @ NASA Ames Research Center

Misty Davies, Becky Hooey
NASA Ames Research Center

Metadata

Safety Reporting Systems commonly use coding taxonomies. Each narrative stored is manually annotated, offering a different search method to users. **Metadata** is an optional search mechanism in Kaona, defined by the standardized metadata format.

The Narrative, Metadata and Embedding Pre-Production Models are updated more rarely, and define a standard format so other datasets can be used interchangeably in Kaona by modifying a [configuration file](#).

Search

Users can perform keyword search, metadata search or semantic searches in Kaona. Semantic searches can leverage any (trained) machine learning model enabled in the tool.

Curate

Searched narratives can be saved in a Search Task, and **curated** by being starred and hidden. Subsequent searches for the same search task will leverage prior search information to improve the search results.

Narrative

Kaona's basic unit of analysis are uniquely identified documents, such as **narratives**. A standardized document format is used so any conforming datasets can be used in Kaona for search, curation and synthesis.

Embeddings

To enable semantic search, i.e. the retrieval of narratives which keywords are equivalent in meaning but not an exact match to the query's keywords, both narratives and query are converted to **embeddings**.

Embedding Models

Kaona supports a plug-in architecture, where **models** trained to a specific dataset (e.g. word2vec in ASRS) can be specified to not be displayed when Kaona is used in other datasets.

Curated Discussion

If users curate searches under a search task, they can choose the search task to be used as context to initiate a **Curated RAG Discussion**.

Synthesis and Discussion

Either **searched narratives** or **curated search tasks** can be used to synthesize Chatbot responses for discussion. For example, a set of curated narratives for a newsletter can be selected for discussion.

