

This thesis documents the process of analyzing, designing, and implementing a software tool able to accept files in various formats, inspect them in depth, and produce a graph in the Resource Description Framework that represents their metadata. Such a description may be useful to any person or system capable of understanding RDF, to provide insight into large sets of files or archives, to allow searching using SPARQL based on concrete domain criteria, or to identify common or distinct entities across different datasets. The results of this thesis may be used by any individual or organization wishing to process files in a semantic and extensible way, to offer users of file hosting sites a wide range of search options, to provide analysts a way to work with metadata in a compact and detailed form, detached from the original source, or to improve systems for processing files with greater control over what kind of data is accepted and processed.