

Separating the Wheat from the Chaff: The Exclusion Method for Classifying Law Firm Documents

Richard Robbins, Epiq Managing Director, Applied Artificial Intelligence

# Introduction

Law firms face the complex task of managing their expansive document repositories. Consider the example of a law firm exploring advanced classification APIs, like Zuva’s, to improve the metadata contained in its document management system. With the overwhelming volume of documents, the goal is to focus on the most relevant documents. However, this is often more challenging than it appears due to the nuanced nature of what constitutes 'relevance' and challenges in identifying the same.

This blog post discusses the challenges of isolating high-relevance documents and proposes a more practical, exclusion-based approach as an effective strategy.

# The Challenge of Identifying High-Relevance Documents

Identifying which documents are most relevant for classification and term extraction is a daunting task in large, varied document sets. The criteria for 'relevance' can vary significantly depending on the context and specific firm or practice-group needs, making it a highly nuanced and often inefficient process to pinpoint the most relevant documents directly.

# Adopting the Exclusion Approach

Given the complexity of identifying high-relevance documents, a more pragmatic approach involves exclusion. This method focuses on eliminating documents that are much less likely to be relevant, to concentrate efforts on a group that includes the most relevant documents. To be conservative, when in doubt, we assume the documents associated with a matter are relevant.

This approach assumes that every document is associated with a matter number and that it is easy to identify a matter type from the matter number. For example, firm administrative matters may appear with a unique prefix identifying the matter as a firm administrative matter, so to with personal matters. Finally, we assume that a matter type is associated with every client matter number so, for example, a matter can be identified a corporate acquisition as compared to litigation.

* **Exclude Administrative Material**. Firm administrative documents are not likely the focus of the firm’s classification and term extraction efforts. So, begin by excluding all documents associated with firm administrative matters.
* **Exclude Personal Workspaces.** Documents contained in personal workspaces are, by their nature, not intended to be shared. These documents are typically not relevant to a firm’s classification and term extraction project. Exclude them.
* **Filtering by Matter Types**. In organizations with varied practice areas, focus on documents relevant to your specific classification and extraction objectives by excluding unrelated matter types. For example, if we assume that a firm is most interested in applying document classification and term extraction to transaction related documents, you might exclude documents associated with matters that are clearly litigation related. The firm’s matter taxonomy will help here. As noted above, if a matter type is ambiguous, do not exclude the documents associated with that matter type.
* **Prioritize Most Recent Versions.** Consider analyzing only the most recent version of a document where there are multiple versions.
* **Applying A Time-Based Approach.** Very old documents or documents associated with dormant matters are often not particularly relevant. Consider only analyzing documents associated with matters that have had time billed to them in the most recent three or five year span.

# The Importance of Refinement

As your efforts progress you may identify collections of documents that you excluded that should be included. For example, the firm may have collected a series of useful reference documents or templates that would have been highly relevant if viewed as client work product, but which were set aside in an administrative matter for knowledge management purposes. These collections, when identified, should be included.

# Conclusion

In conclusion, the journey towards more efficient document management in law firms, especially concerning classification and term extraction, is less about directly pinpointing high-relevance documents and more about intelligently excluding the less relevant ones. This exclusion-based approach, which emphasizes filtering out administrative materials, personal workspaces, and unrelated matter types, allows firms to focus on a more manageable set of potentially relevant documents. The process, however, is not static. It requires ongoing refinement, adapting to new insights and realizations about the relevance of certain documents initially overlooked. Ultimately, this method paves the way for more focused and effective use of classification and extraction technologies, turning a daunting task into a manageable and fruitful endeavor.