

Keyword-based data reduction is often maligned as being imprecise and ineffective. But with experts at the helm who know how to construct keyword search that can achieve simultaneously high recall and precision (finding what's needed without over-inclusion), data volume can be significantly reduced and the refined dataset can be searched to more readily find the documents that matter most. H5 offers its clients such expertise. The case studies below are just a few examples of how H5 has helped clients reduce cost and risk using keyword-based data reduction.

## CASE STUDIES

### Case Study 1

- ▶ **Matter Type:** SEC Investigation of Possible Federal Securities Violations
- ▶ **Collected Data Volume:** 900,000 documents
- ▶ **Volume after H5's Keyword Reduction:** 22,847 (97% volume reduction)

**Upshot:** From the review population, H5 also identified a subset of documents as potentially privileged to help prioritize counsel's second-pass and privilege review. Counsel reviewed the 22,847 documents over the course of three weeks, resulting in the production of 16,125 documents, approximately 70% of the review set.

### Case Study 2

- ▶ **Matter Type:** Pharmaceutical Industry Responsive Review for Patent Litigation
- ▶ **Collected Data Volume:** 332,296 documents
- ▶ **Volume after H5's Keyword Reduction:** 43,981 (87% volume reduction)

**Upshot:** Using H5's keyword list, the client avoided an estimated \$230,652 in contractor review expense.

### Case Study 3

- ▶ **Matter Type:** Telecom Industry Responsive Review for Patent Litigation
- ▶ **Collected Data Volume:** 972,681 documents
- ▶ **Volume after H5's Keyword Reduction:** 141,264 (85% volume reduction)

**Upshot:** Using H5's keyword list, the client avoided an estimated \$665,132 in contractor review expense.

### Case Study 4

- ▶ **Matter Type:** Internal Investigation
- ▶ **Collected Data Volume:** 600,000 documents
- ▶ **Volume after H5's Keyword Reduction:** 27,900 (a 95% reduction)

**Upshot:** Using H5's keyword list, the client avoided an estimated \$457,680 in contractor review expense.

## H5's Keyword-Based Data Reduction Methodology

Keyword-based searches are often employed to reduce collected document populations prior to review. Typically, the goal is to reduce the document population in a principled manner, while fulfilling the obligation to review and produce what is required. Successfully achieving this goal requires a sound methodological approach to the development of keyword searches. Outlined below are the primary steps in H5's *court-approved* keyword-based data reduction methodology.

### Step 1: Create topical subject matter outline

- Express subject matter as entities (nouns) and processes (verbs)
- Group related subject matter into topic areas

### Step 2: Create searches for each topic and associated sub-topics

- Map each search to the topic/sub-topic it addresses
- Determine if any searches are mere-mention responsive
- Pair non mere-mention searches with conceptual anchors focused on responsive contexts
- Leverage H5 organizational experience to optimize utilization of operators during pairing of keywords and conceptual anchors

### Step 3: Account for linguistic and computational linguistic variability

- Develop searches that address variability in:
  - Word-order: sentences with similar meaning may benefit from bi-directional operator to address syntactic variability
  - Word morphology (e.g., number (plural vs. singular), tense, derivation)
  - Word-choice (e.g., synonyms, formal vs. colloquial terms, industry vs. lay usage, brand vs. generic)

### Step 4: Test and Iterate (as required)

- Run searches over document collection
- Review sample from “hits” for non-responsive documents
- Revise searches responsible for hitting non-responsive documents
- Review sample from “no hits” for responsive documents
- Create supplemental searches for responsive documents found
- Iterate as required

For information regarding how H5 can help you use keyword search for data reduction, contact us at 1.866.999.4215.