



---

## Term Testing – a Case Study

IE Discovery, Inc.

### 1. Purpose and Background

The litigation world has many examples of cases where the volume of Electronically Stored Information (ESI) demands that litigators use automatic means to assist with document identification, classification, and filtering. This case study describes one such process for one case. This case study is not a comprehensive analysis of the entire case, only the Term Testing portion.

Term Testing is an analytical practice of refining match terms by running in-depth analysis on a sampling of documents. The goal of term testing is to reduce the number of false negatives (relevant / privilege document with no match, also known as “misdetections”) and false positives (documents matched but not actually relevant / privilege) as much as possible.

The case was an employment discrimination suit, against a government agency. The collection effort turned up common sources of ESI: hard drives, network shares, CDs and DVDs, and routine e-mail storage and backups. Initial collection, interviews, and reviews had revealed that a few key documents, such as old versions of policies, had not been retained or collected.

Then an unexpected source of information was unearthed: one network administrator had been running an unauthorized “just-in-case” tracer on the email system, outside the agency’s document retention policies, which created dozens of tapes full of millions of encrypted compressed emails, covering more years than the agency’s routine email backups. The agency decided to process and review these tracer e-mails for the missing key documents, even though the overall volume of relevant documents would rise exponentially.

The agency had clear motivation to reduce the volume of documents flowing into relevancy and privilege reviews, but had concerns about the defensibility of using an automated process to determine which documents would never be reviewed. The case litigators and Subject Matter Experts (SMEs) decided to use a process of Term Testing to ensure that automated filtering was both defensible and as accurate as possible.

### 2. Term Testing Process

The Term Testing process is an iterative approach to refining match terms. A subset of documents is reviewed definitively by SMEs for relevance and privilege, then run through the first pass of match terms to discover false negatives and false positives. Terms are refined and re-run until the results are within

limits of acceptability as defined by the client and the circumstances of the case. The rest of the case study explains the steps in depth, and gives detailed numbers to show affected volumes.

## **2.1. Identify Document Sample**

The complete collection of tracer e-mails is estimated to be approximately 10 million documents (exact numbers are unknown because the e-mails are stored in a compressed format, each compressed file unpacks to a different number of e-mails, and not all tapes were uncompressed before the project was put on hold).

Given that the tracer e-mails were collected by date (versus most collections, which are primarily by custodian or location), date was the best criteria to ensure a representative sample. Uncompressed files were selected from each year in the collection, covering different months in case there was an unexpected seasonal factor. The files from the chosen dates ended up comprising 15,220 documents, approximately 0.15% of the overall collection.

The decision to keep the sample small was practical as well as statistical in nature. For practical purposes, because all the documents in the sample required in-depth human analysis from a small team of experts, a collection of more than about 15,000 could jeopardize the target duration of three weeks for the Term Testing process. Statistically, the attorneys hoped to identify the key documents within the first 1,000,000 documents processed, which gave a sampling of approximately 1.5% of the overall collection, a much more common sample size.

The process to identify the documents only took one day.

## **2.2. SME Review**

Two SMEs reviewed each document for relevance and privilege, in a double-blind review so neither SME knew how the other had marked the document. Documents with conflicting markings were then reviewed by a panel of SMEs charged with resolving conflicts. Fortunately, only 326 documents were in dispute (2.1% of the sample), so the resolution process took only one day. These SME review decisions became the standard of correctness for the rest of the process, so the results of the automated term matching were compared against the SME review decisions to determine false negatives and false positives.

## **2.3. Search Terms List 1**

The SME team created a first draft of search terms, known as "List 1". Terms were identified for both relevance and privilege simultaneously, since the match technology only needed to be run once to determine hits on both sets of terms. One SME made a first draft of the term list, and other team members added to the list. For the first pass, all ideas were included.

Then technical staff translated the English-version of the term to a technical, regular-expression format of the term. The regular expression technology allows for one term to cover multiple spellings where

required. For example, the name “Stephen” is a regular expression of “Ste(v|ph)e(n)?” which catches any of the following versions of the name: Steve, Steven, Stephe, Steven.

The creation of List 1 happened before the other steps in the Term Testing process began, so did not add any days to the process.

## **2.4. List 1 Analysis and List 2 creation**

List 1 was run against all the sample documents for search term hits. The SME team then analyzed both false negatives (relevant /privilege document with no match) and false positives (documents matched but not actually relevant / privilege). The agency had already decided that the risk of omitting potentially relevant documents outweighed the costs of reviewing a higher volume of documents, so the primary focus was on eliminating false negatives down to zero.

The group was pleasantly surprised to find that the first pass only resulted in 921 documents which were false negatives (6.1%), and 1,892 documents which were false positives (12.4%). The SME group distributed the false matches and combed through the documents to find new terms, and to analyze terms as candidates to be removed.

The technical staff analyzed the term hits to see if any terms were unnecessary, which would have been true one of three ways: either (a) the term did not match any documents in the sample, or (b) the term only matched documents where other terms also matches, so could be unnecessary, or (c) the term produced so many false positives that it was not helpful to include. No terms fell into category (a). The few terms which fell into category (b) were deemed too necessary to discard. Only 4 terms were considered in category (c), but only 2 terms were actually deemed unnecessary.

The total change between List 1 and List 2 as the addition of 12 terms, and the elimination of 2 terms, for a total net increase of 10 terms.

This analysis phase of the process took longer than any other portion; it only took a couple of hours to run the terms against the documents, but the analysis of results and creation of List 2 took about five business days to complete.

## **2.5. List 2 Analysis and List 3 creation**

List 2 was run against all the sample documents. The group was very pleased that no documents were found to be false negatives (0.0%), but somewhat discouraged that the false positive rate rose to 4,120 documents (27.1%). The team repeated the same analysis performed after List 1 ran, and agreed on removing 5 of the new terms from the list.

This second analysis only took 3 business days to complete.

## **2.6. List 3 Analysis**

List 3 was run against all sample documents. The result of 0 documents with false negatives (0.0%) was achieved, so even though the false positives were 2,133 documents (14.0%), List 3 was finalized as the list of record to start processing the entire collection.

This third analysis only took 2 business days to complete.

## **2.7. Ongoing Evaluation**

On an ongoing basis, two SMEs dedicated one day per month to review non-relevant documents. Documents were identified for the non-relevant review by randomly selecting approximately 1.0% of the documents processed during the previous month with no relevancy match terms. The goal was to ensure that additional terms were not needed. In the three months where these reviews occurred, no false negative documents were identified, so no term changes were made.

Had new terms been identified, the new terms would have been run against all files, including files which had been analyzed with List 3 terms.

## **3. Outcome and Summary**

As often happens in complex litigation, the case changed mid-stream, and in this case, because case strategies paid off. The tracer e-mail effort started by focusing on a specific 6-month time window most likely to uncover missing key documents. As hoped, some of the missing key documents were found quickly, and produced to opposing council immediately. The revelations from those key documents changed the nature of the matter so fundamentally that the litigators decided to suspend further tracer e-mail efforts indefinitely to focus resources elsewhere. Although the case has not yet been fully resolved, the tracer e-mail effort continues to be on hold.

The team considers the Term Testing to have been successful because, in conjunction with the time window strategy, the right key documents were uncovered quickly, the risk of missing key documents was significantly reduced, and the defensibility of the match terms was greatly improved.