How To Design Defensible Keyword Searches

Learn to design effective and defensible keyword searches for e-discovery. This guide covers term selection, syntax, negotiation, documentation, and common pitfalls, ensuring your search methodology is robust and withstands scrutiny.

e-discovery.uk, Computer Forensics Lab. Published 2026-08-30.
https://e-discovery.uk/library/how-to-design-defensible-keyword-searches

KEYWORDSEARCHDESIGN · A GUIDE FOR UK LAWYERS How to Design Defensible Keyword Searches Terms, Syntax and the Method That Survives the Other Side's Scrutiny COMPUTER FORENSICS LAB

§ ABOUT THE AUTHOR PREPARED BY COMPUTER FORENSICS LAB E-DISCOVERY TEAM ESTABLISHED 2007 · LONDON ISO 17025-ALIGNED PROCEDURES SEARCH PROTOCOL DESIGN

§ CONTENTS In this guide 01 Executive summary 02 The problem in plain English: the list drafted from memory 03 Where terms come from: case theory and custodian vocabulary 04 Syntax that works: operators, wildcards and proximity 05 What keywords miss: the known failure modes 06 The negotiation: DRD tables, hit counts and refinement 07 Documentation: making the list defensible 08 Source architecture: where else the evidence lives 09 Worked examples 10 Common mistakes and technical limitations 11 Questions to ask · Suggested wording 12 Checklist and red flags · When to involve a digital forensic expert 13 Frequently asked questions 14 Glossary · References · Disclaimer · How a specialist laboratory can assist

§ 01 · ORIENTATION Executive summary THE HEADLINE POINT: TERMSCOMEFROMTHEDATA ' SOWNVOCABULARY, SYNTAXIS ENGINEEREDPERPLATFORM, ANDEVERYCHOICEISTESTEDANDDOCUMENTEDBEFORE IT IS AGREED

§ 02 · FIRST PRINCIPLES The problem in plain English: the list drafted from memory

§ 03 · THERAWMATERIAL Where terms come from: case theory and custodian vocabulary

§ 04 · THEENGINEERING Syntax that works: operators, wildcards and proximity

§ 05 · THEBLINDSPOTS What keywords miss: the known failure modes

§ 06 · THETABLESTAKES The negotiation: DRD tables, hit counts and refinement

§ 07 · THEPAPERTRAIL Documentation: making the list defensible

§ 08 · THEWIDERMAP Source architecture: where else the evidence lives WHEN WIDER OCR & KEYWORDS TERMS / TEXT RETURN VARIANTS REPAIR PARTICIPANT STRUCTURED-ANALY TICS ADJACENT & DATE PULLS DATA QUERIES & TAR CHANNELS

§ 09 · IN THE WILD Worked examples EXAMPLE1 ·

THECODENAMETHEPLEADINGSNEVERKNEW EXAMPLE2 · THEWILDCARDTHATATETHEBUDGET EXAMPLE3 · THENULLRETURNTHATWASANOCRPOSTURE

§ 10 · WHEREITGOESWRONG Common mistakes and technical limitations Common mistakes Technical limitations

§ 11 · INTERROGATORIES & DRAFTING AIDS Questions to ask · Suggested wording Ask your client Ask your opponent Ask your e Discovery / forensic provider SUGGESTED WORDING · SEARCH METHODOLOGY PA R AG RAPHFORTHEDRD / PROTOCOL

§ 12 · QUICK CONTROL Checklist and red flags · When to involve a digital forensic expert The search-design checklist Red flags When to involve a digital forensic expert

§ 13 · COMMON QUESTIONS Frequently asked questions How many search terms should a list have? The other side proposes terms hitting hundreds of thousands of documents. Must we accept? Our searches returned nothing on a central issue. Can we certify that? Do keyword searches work on Whats App and Teams data? Are keyword searches still acceptable now TAR exists? Who should draft the list: lawyers or the provider?

§ 14 · REFERENCE Glossary Sources and authoritative references DISCLAIMER

§ HOW A SPECIALIST LABORATORY CAN ASSIST Working with Computer Forensics Lab Speak to a forensic examiner, not a salesperson. INSTRUCTTHELAB NEWENQUIRIESEMAILE - DISCOVERY

Questions and answers

How many search terms should a list have?

As many as the issues need and no more: quality of construction beats length: a tested thirty-term list with proximity engineering out per form s an untested hundred every time, and costs less to run, review and defend. The DRD conversation goes better with a short list you can justify term-by-term than a long one you cannot.

The other side proposes terms hitting hundreds of thousands of documents. Must we accept?

No: you counter with arithmetic: hit counts, sampled precision, review cost per term: and propose refinements (proximity, intersections, date limits) that keep the concept while shedding the noise: Example 2's renegotiation is the template, and courts consistently reward the party holding numbers rather than adjectives. Acceptance under deadline pressure of an untested term is the one concession that cannot be walked back cheaply.

Our searches returned nothing on a central issue. Can we certify that?

Only after the §8 routing: is the relevant estate actually text-searchable (Example 3's question first), did the terms speak the custodians' register, has a participant-and-date pull been run as the safety net? A null certified over those checks is solid; a null certified instead of them is a future application with your name on it.

Do keyword searches work on Whats App and Teams data?

They run, but the vocabulary and the unit change: in form a l spelling, fragments-across-messages, emoji and voice notes defeat email-register terms: so messaging sources get adapted terms from the guide 63 harvest, conversation-level retrieval, and honest declaration in the protocol that the methodology differs per source. Searching chat with the email list and reporting the nulls is a recognised way to miss the case.

Are keyword searches still acceptable now TAR exists?

Fully: PD 57AD contemplates both, and mature practice combines them: keywords for the nameable (parties, projects, references), TAR and analytics for the conceptual and the unnamed, with the DRD recording which tool covers what: guide 89 treats the TAR half. What is fading is not keywords but untested keywords presented as a complete methodology.

Who should draft the list: lawyers or the provider?

Both, in sequence: lawyers own the issues layer and the legal judgements (scope, NOTs, proportionality); the provider owns the dialect, the testing and the engineering; the custodians own the vocabulary nobody else has. The protocol file is where the three meet: and a list drafted by any one of them alone carries that author's characteristic blind spot. cflab. u k · e-disc ove r y. u k