Trustworthy Email Search: Verifiable Completeness and Evidence in KMail

Stavros Kousidis

Bachelor thesis · Published: 2026-10-25

Research question

How can a KMail search system make index coverage, result provenance, and synchronization correctness observable and verifiable at practical archive sizes?

Description

An email search may silently miss messages that were not downloaded, could not be indexed, or became stale. This thesis develops a **trustworthy, read-only search prototype** integrated with KMail's Akonadi infrastructure. The goal is verifiable retrieval, not an AI chatbot.

Scope and evaluation

Define a precise searchable corpus (e.g., locally accessible messages in selected Akonadi collections). Distinguish **index completeness** from **retrieval effectiveness**: even a complete index can miss relevant answers.

Implement full-text search and coverage reporting for indexed, pending, unavailable, encrypted, failed, and stale items. Link each result to its source message and matching passage where possible. A companion interface is acceptable if native plugin integration proves impractical.

Keep the index synchronized as messages arrive, change, move, or disappear; recover from interrupted processing. Test with reproducible mailboxes and known expected results. Measure coverage, reference correctness, failure detection, query latency, index size, and resource use at increasing archive sizes.

Do not persistently index decrypted private correspondence or cryptographic keys by default; make security-related exclusions visible.

Expected outcome

A documented KMail/Akonadi integration, working search prototype, automated consistency tests, and a quantitative evaluation of reliability and scalability. Semantic search and LLMs are not required.

Mandatory open-source publication

Open-source publication is a condition for taking this topic. Before registration, students must agree to release thesis-developed code, tests, and reproducibility tools under an agreed, compatible OSI-approved license by project completion. Development may remain private; confidential mail, keys, and credentials must not be published. **If you cannot accept this condition, please choose another topic.** See the publication and licensing policy for explanations of OSI, commercial use, and third-party licenses.

Starting points

- Akonadi API; SQLite FTS5.
- KDE licensing policy.

Related research

The complementary Master's topic investigates agentic retrieval and confidentiality. Both topics are independently feasible.

Contact: stavros.kousidis@th-koeln.de
Code publication: open source by default; exceptions require prior agreement (policy).