

Meeting Notes

Project: LTU Search Engineee

Meeting: 2025-12-18

Attendees:

Emma, Cilla, Linnea, Jean-Paul, Nattarintra

Agenda

1. Divide work between members

Discussion and Notes

We focused on dividing the tasks for the search engine project. Everyone has been assigned a specific area to research or implement.

Action Items

- **JP:** Query syntax
- **Emma:** What to include in an index, upload meetingnotes to Git Repository
- **Linnea:** How to limit crawler rate limit
- **Cilla:** Investigate different algorithms for searching
- **Nattarintra:** Test

Meeting: 2025-12-17

Context: Updated directives from Malte (replaces parts of original spec).

Attendees:

Emma, Cilla, Linnea, Jean-Paul, Nattarintra

1. Technical Requirements (Scope)

- **Database:**

- Free choice (Postgres, MariaDB, SQLite are all okay).
- *Important:* Avoid MS SQL Server (due to Linux compatibility issues). Choose something easy to install/run.
- Storage: 512 GB is available on the VM, but we likely only need a few hundred MB.

- **Web Crawler Behavior:**

- **Whitelist:** Strict. Only crawl `ltu.se` and specific links provided. Do **not** follow external links (e.g., Facebook, Instagram).
- **Robots.txt:** Generally respect it, *unless* it blocks "contacts" or "users". We must index contact pages so staff can be found.
- **Content:** Skip images to reduce complexity. Focus on text.
- **Rate Limiting:** No hard limit, but "don't DDoS". A few hundred requests at a time is acceptable.

- **Search Algorithm & Ranking:**

- Use a combination of **PageRank** (link popularity) and **Word Frequency**.
- We are encouraged to experiment with algorithms to find the best results.

2. Documentation & Requirements (Milestone 0)

- **Requirements Document:**

- Add a **third column** for "Test Cases". (Be specific: How do we prove the requirement is met?)
- Remove "Test Coverage" from the requirements doc (keep it in code, but not as a functional requirement).

- **Diagrams:** Create initial drafts for Class Diagram, ER-Diagram (Database), and UML Diagram. Share these early ("Share everything").

- **Tools:** Documentation in LaTeX, Code in Git.

3. Project Structure & Team Roles

Total team size: 5 members.

- **1 Lead Developer** (Responsible for overall architecture).
- **2 Frontend Developers**.
- **2 Backend Developers**.

Workflow: Use Code Reviews and Pull Requests (PR). Clear separation of tasks between Frontend and Backend.

4. Reporting & Next Steps

- **Weekly Report:** Send an email **every Friday**.
 - Content: "Activity report" (what we did) and "Next to-dos" (what we will do).
 - *Note:* Send one immediately before the holidays.
- **Next Meeting:** January 8th at 15:00.

Meeting: 2025-12-16

Attendees:

Emma, Cilla, Linnea, Jean-Paul, Nattarintra

Agenda

1. Meeting for preparation

Discussion and Notes

We had a shorter meeting to prepare questions regarding tomorrow's meeting with Malte.

Meeting: 2025-12-15

Attendees:

Emma, Cilla, Linnea, Jean-Paul, Nattarintra

Agenda

1. Meeting for preparation

Discussion and Notes

We had a shorter meeting to prepare questions regarding tomorrow's meeting with Malte.

Meeting: 2025-12-11

Attendees:

Emma, Cilla, Linnea, Jean-Paul, Nattarintra

Agenda

1. Daily meeting we discussed what to focus on, Jean-Paul set up a Req document and we started to fill it up with Test and priorities them.

Discussion and Notes

We had a shorter meeting to prepare questions regarding tomorrow's meeting with Malte.

Meeting: 2025-12-10 (Project Kick-off)

Attendees: Team and Malte (Client)

Agenda

1. Review of project scope and limitations
2. Technical requirements clarification
3. Expectations for Milestone 0

Decisions & Directives (Summary)

1. Scope & Crawler Behavior

- **Whitelist:** Strictly *.ltu.se. Do not follow external links (Facebook, etc.). Future domains like islab.se might be added later.
- **Robots.txt:** strictly adhere to it for Milestone 0. (Exceptions for contact pages might be added later).
- **Dynamic Content:** LTU's dynamic content is server-side rendered, so no "headless browser" is required.
- **PDFs:** Crawler must detect PDF links. Indexing the *content* of PDFs is optional (nice-to-have).

2. Search Functionality

- **Queries:** Optimized for simple queries (2–3 words). Ranking priority: All terms > Some terms > One term.
- **Pagination:** Required (e.g., 10, 20 results per page).
- **Performance:** Response time should be under 10 seconds.
- **Optional features:** Wildcards (*), category filters, and boolean operators are *not* mandatory.

3. Technology & Deployment

- **UI:** Plain HTML/CSS/JS is perfectly acceptable. No complex frameworks (React/Vue) needed unless desired.
- **Deployment:** Local execution is enough. No server deployment required yet.

Deliverables for Milestone 0

- **Requirements Document:** Must include Stakeholder requirements, Functional requirements, and Non-functional requirements.
- **Diagrams:** UML Class diagram + ER Diagram (Database).

/newpage

Meeting: 2025-12-09

Attendees:

Emma, Cilla, Linnea, Jean-Paul, Nattarintra

Agenda

1. Get to know eachother
2. Send a mail to Malte
3. Set up a Github project

Discussion and Notes

We had a presentation round and talked about the project and sent a email to Malte for scheduling a meeting. Finally we set up a Github Project