

# Requirements Specification - LTU Search Engine

Fullstack Developer .NET

December 17, 2025

## 1 Functional Requirements

Table 1: Functional Requirements List

RQ ID	Description	Test Method
<b><i>Crawler (1000 Series)</i></b>		
FRQ-1001	The system shall crawl web pages starting from a given seed URL.	Test Crawl starts from provided seed URL
FRQ-1002	The system shall follow internal links recursively.	
FRQ-1003	The system shall limit crawling speed max 300 requests/minute (limit the site requests).	
FRQ-1004	The crawler must parse and adhere to the “robots.txt” file of the target domain.	
FRQ-1005	The system shall restrict crawling to a configurable whitelist of domains.	
FRQ-1006	The crawler shall avoid crawling the same URL more than once per execution.	
FRQ-1007	The crawler shall ignore non-relevant resources (images, CSS, JS).	
FRQ-1008	The crawler must only search within the domain specified in the whitelist.	
FRQ-1009	The system must support adding new domains to a whitelist automatically.	
FRQ-1010	The crawler must detect linked PDF files and include them in the index.	
FRQ-1011	The crawler must not follow/crawl links found inside PDF documents.	
<b><i>Indexing (2000 Series)</i></b>		
FRQ-2001	The system shall extract textual content from HTML pages.	
FRQ-2002	The system shall store indexed terms together with page references (inverted index).	
FRQ-2003	The system shall support incremental updates of the index.	
FRQ-2004	The system shall ignore non-textual content (images, videos, binaries).	

Continued on next page

**Table 1 – continued from previous page**

RQ ID	Description	Test Method
<b><i>Searching (3000 Series)</i></b>		
FRQ-3001	A query containing 2-3 words should return the correct result.	
FRQ-3002	A query must have an option to match characters strictly (AND).	
FRQ-3003	A query must have an option to include non-mandatory characters (OR).	
FRQ-3004	Use same syntax as other providers (standard search syntax).	
FRQ-3005	Results of queries must be paginated when more than 10 results are found.	
FRQ-3006	The results should contain the headline of the context found.	
FRQ-3007	The results should contain a snippet with keywords highlighted.	
FRQ-3008	Search results must be ranked by relevance.	
FRQ-3009	The system shall provide a search API for the UI.	
<b><i>User Interface (4000 Series)</i></b>		
FRQ-4001	The system shall provide a web-based search interface.	
FRQ-4002	The system shall allow users to enter search queries and view results.	
FRQ-4003	The UI shall indicate the current page (pagination state).	
FRQ-4004	The UI shall display a clear message when no results are found.	

## 2 Low-Priority Functional Requirements

RQ ID	Description	Test Method
L-FRQ-5001	The query should be able to handle wildcards.	
L-FRQ-5002	An estimation of the completed query should be suggested (Autocomplete).	

## 3 Non-Functional Requirements

RQ ID	Description	Test Method
NFRQ-6001	A query should take no longer than 10 seconds.	
NFRQ-6002	The search engine should search the whole LTU-Domain (with exceptions).	

NFRQ-6003      The system shall only index publicly available HTML pages.

---

## 4 Non-Testable RQ

FRQ-2005 The system shall use a clearly defined data structure (e.g., ER diagram).