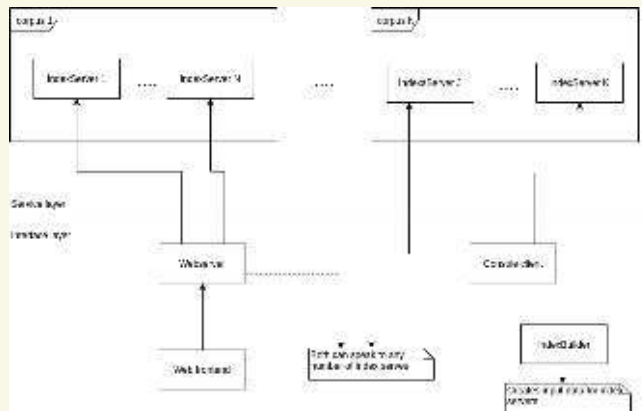


Indexing of Big Text Data and Searching in the Indexed Data

Overview

- ▶ Semantic search engine for KNOT.
- ▶ Searching based on the meaning of words.
- ▶ Powerful query language EQL.
- ▶ Scalable distributed system.
- ▶ Built on top of MG4J.

Components of the platform



- ▶ 5 components built from 13 modules.
- ▶ Each of them can be replicated.
- ▶ Currently running on over 50 servers.

EQL – Enticing Query Language

- ▶ Query entities and their attributes.
- ▶ Context constraints for paragraphs or sentences.
- ▶ Global constraints for relationships between entities.
- ▶ Compiler with extensible semantic analysis toolkit.

Other features

- ▶ Smart search bar.
 - Syntax and semantic analysis.
 - Feedback in the GUI.
- ▶ Easily extensible.
 - Different indexing libraries can be added.
- ▶ Custom configuration DSL.
- ▶ Built using modern technologies.
 - TypeScript with React & Redux on the frontend.
 - Kotlin with Spring on the backend.

location.surname: 'Smith'

Performance measurements

EQL Query	Average[ms]	Deviation[ms]	min[ms]	max[ms]
water	711.4	65843.12	337	1823
nertag:person	155.4	2263.76	104	421
nertag:location	177.73	24321.48	93	835
water nertag:person	503.0	35445.24	355	1259
water nertag:person ctx:sent	423.15	8891.11	321	771
nertag:person nertag:location	130.24	1919.32	96	433
nertag:person nertag:location ctx:sent	287.09	2490.72	220	463
water nertag:person nertag:location	453.82	14603.47	331	1056
water nertag:person nertag:location ctx:sent	4317.23	16767.16	4071	4828
nertag:person nertag:person	119.28	1071.40	96	403
nertag:person nertag:person ctx:sent	124.48	828.09	100	330
a=nertag:person b=nertag:person	117.31	699.074	97	287
a=nertag:person b=nertag:person ctx:sent	130.61	667.64	104	247
a=nertag:person b=nertag:person && a.url != b.url	126.33	801.48	98	263
a=nertag:person b=nertag:person ctx:sent && a.url != b.url	353.91	10034.46	250	801

Table: Early results of performance tests

- ▶ Platform is still being optimized.

Current state

- ▶ Core implementation finished.
- ▶ Ongoing testing and optimizations.
- ▶ Many possible extensions.
 - Different indexing libraries.
 - Autocomplete for the search bar.
 - Native mobile client.