

# ChatNoir Reloaded

Web Search Engine

Winter term 2012/13

# ChatNoir Reloaded – Web Search Engine

# ChatNoir Reloaded – Web Search Engine

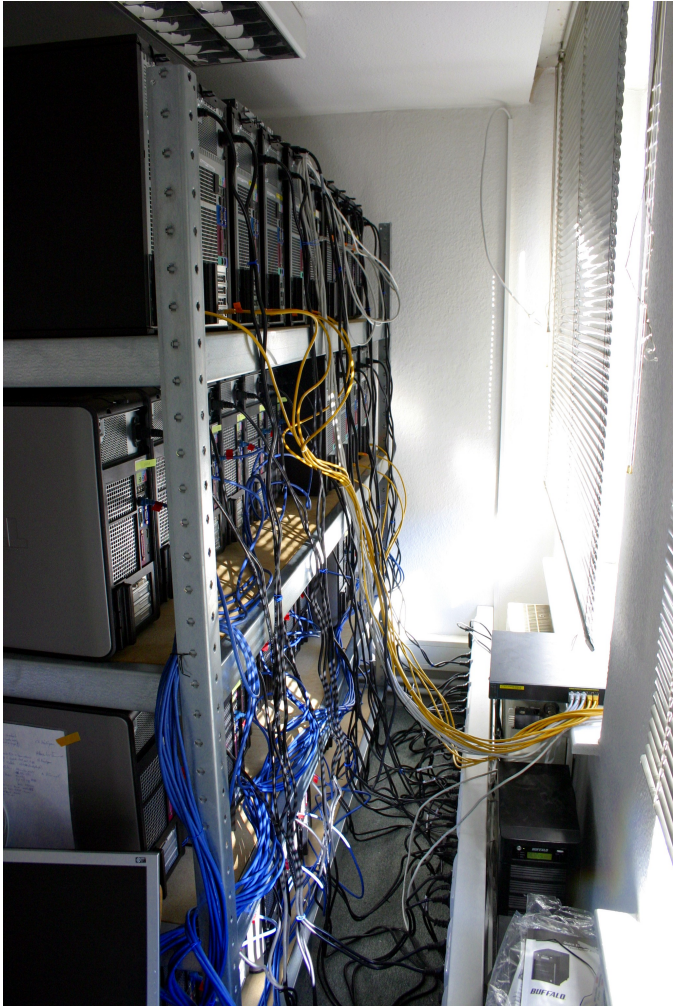
Data



25 TB Web Crawl

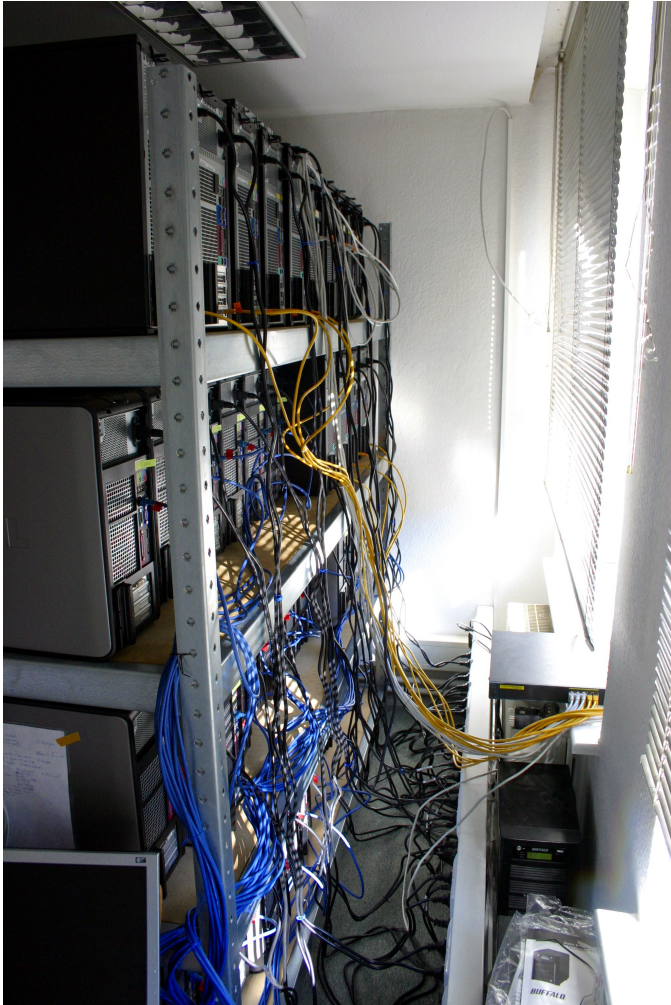
# ChatNoir Reloaded – Web Search Engine

## Cluster



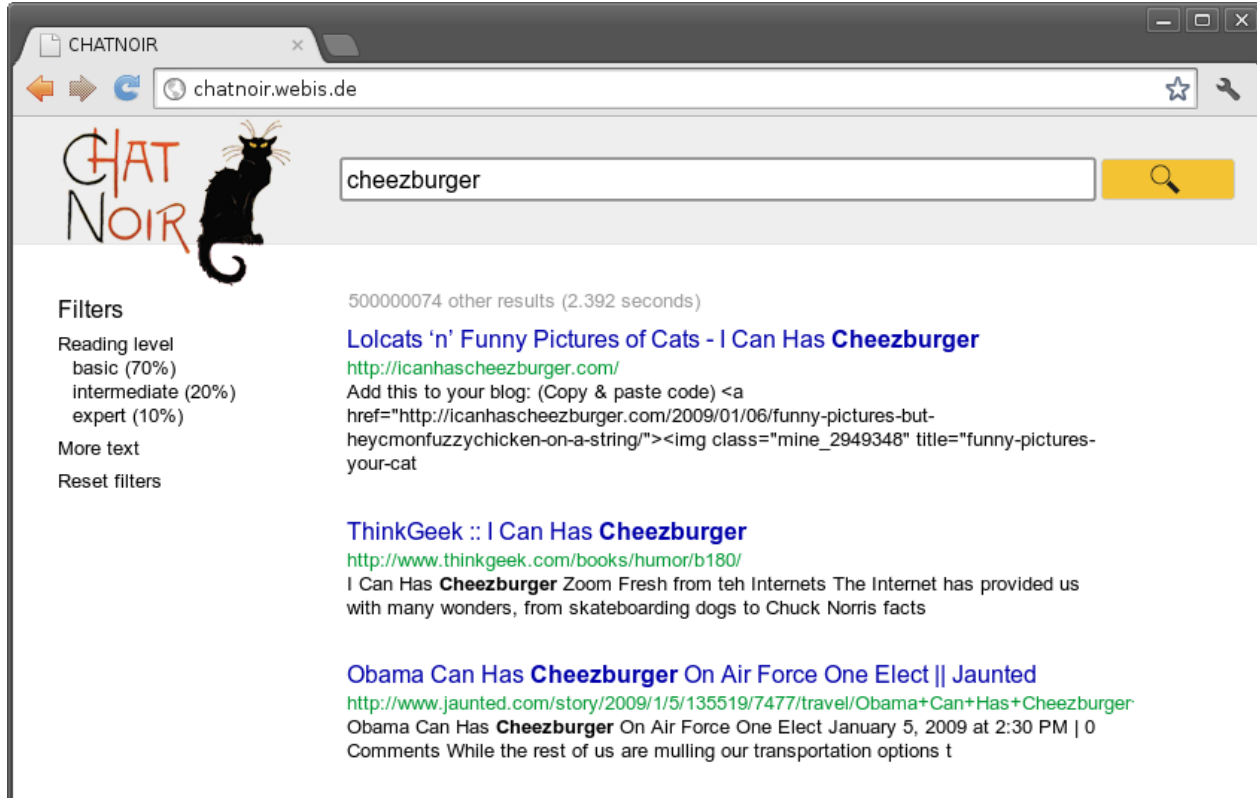
# ChatNoir Reloaded – Web Search Engine

## Cluster



# ChatNoir Reloaded – Web Search Engine

Mission: Improve ChatNoir



- + Phrasal Search
- + Spelling Correction
- + Spam Classification

- + CommonCrawl (60TB)
- + Compressed Index
- + Learning to Rank

# ChatNoir Reloaded – Web Search Engine

## Project goals

Learn and implement the technology underlying web search engines.

## Information Retrieval

- ❑ Distributed inverted indexing (sharding, compression)
- ❑ Query processing techniques (phrases, spelling)
- ❑ Learning to rank (click-through information)
- ❑ Evaluation (TREC experiments)

## Software Engineering

- ❑ Cluster Computing (Hadoop, Map-Reduce)
- ❑ Programming languages (Java, C++, GWT)