

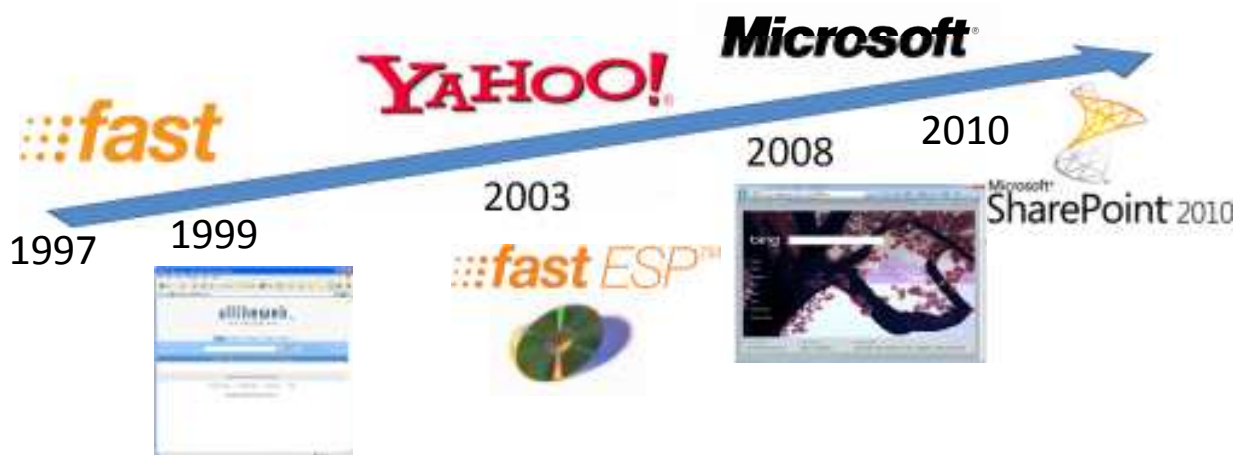
Web Search

Rolf Michelsen

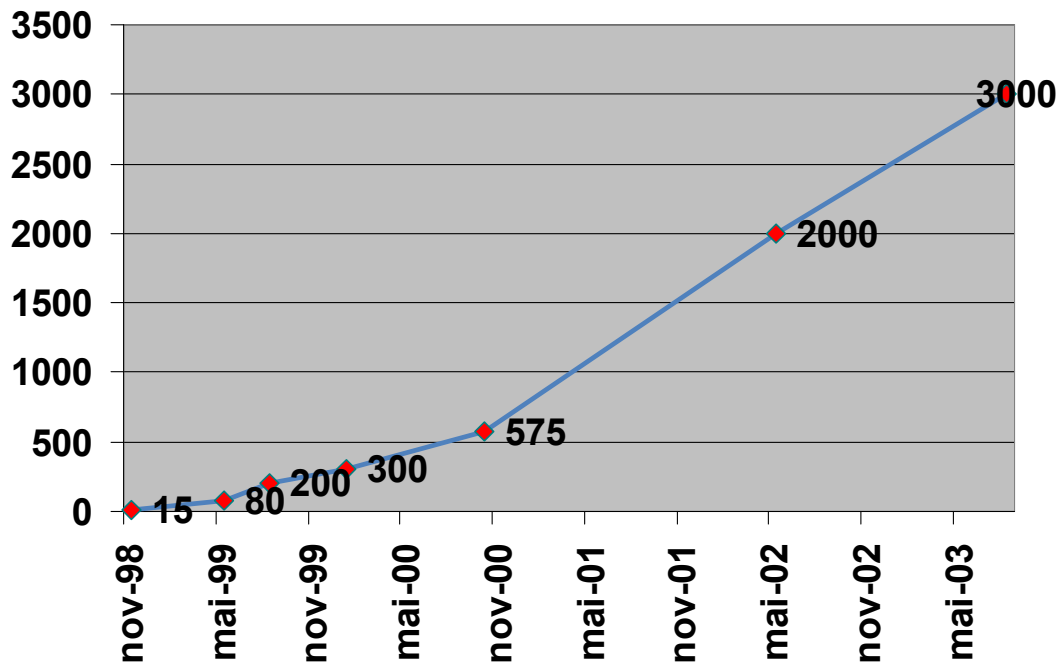
Microsoft Development Center Norway

2011-04-04

Background



Web Search Development

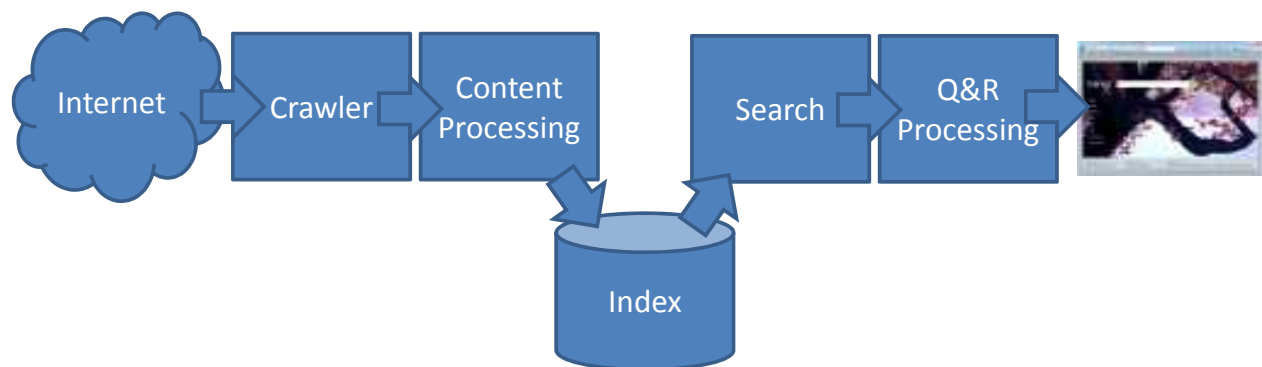


Topics

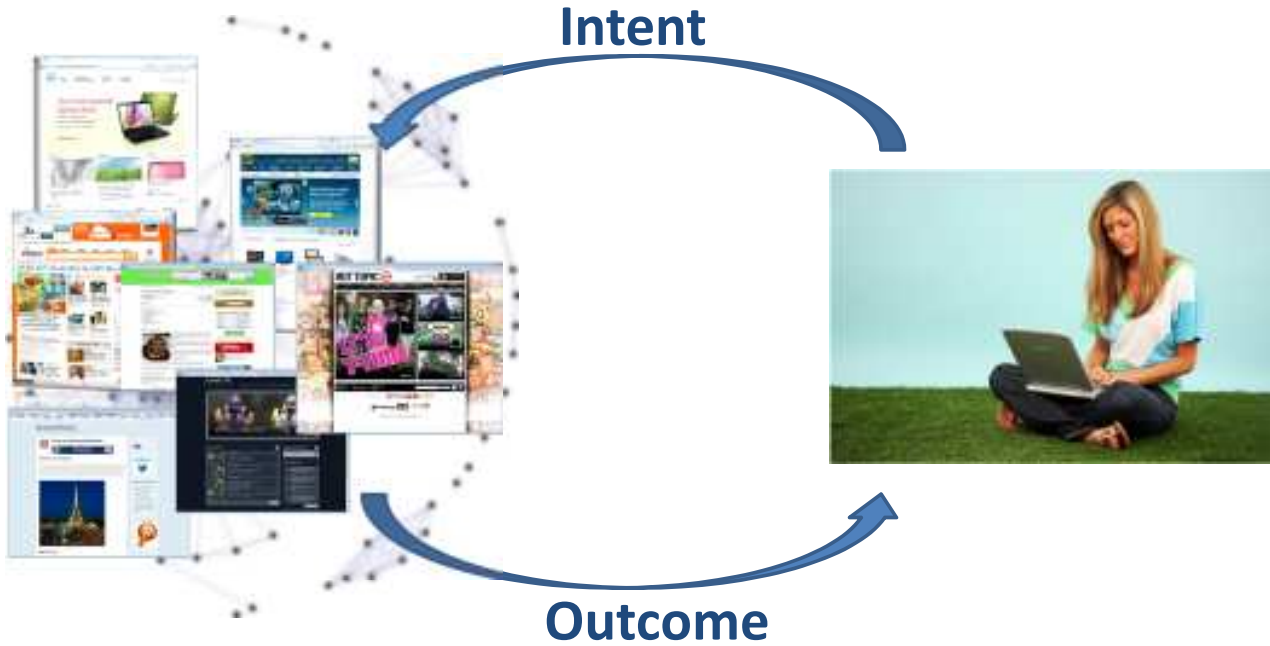
- Web search.
- Major real-world challenges.
- Engineering complex systems.
- ...less about algorithms.

WHAT IS WEB SEARCH?

Web Search



Web Search



Web Search



Key Challenges

- Knowing the application.
- Knowing the environment.
- Engineering complex systems.

THE APPLICATION

Knowing the application

- What makes a good consumer experience?
- What makes the ecosystem work?

Relevance



vs.



Metrics

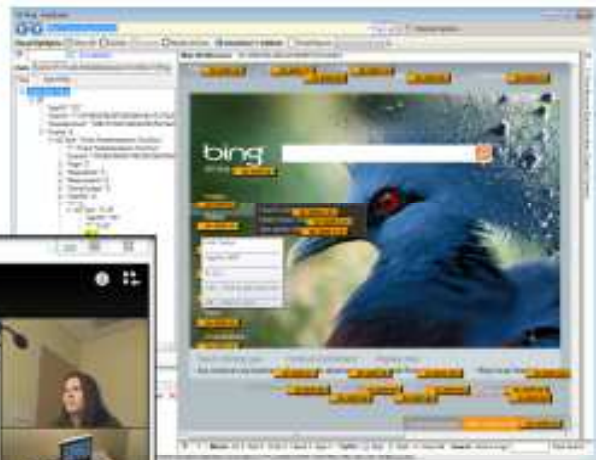
Hard metrics



Soft metrics

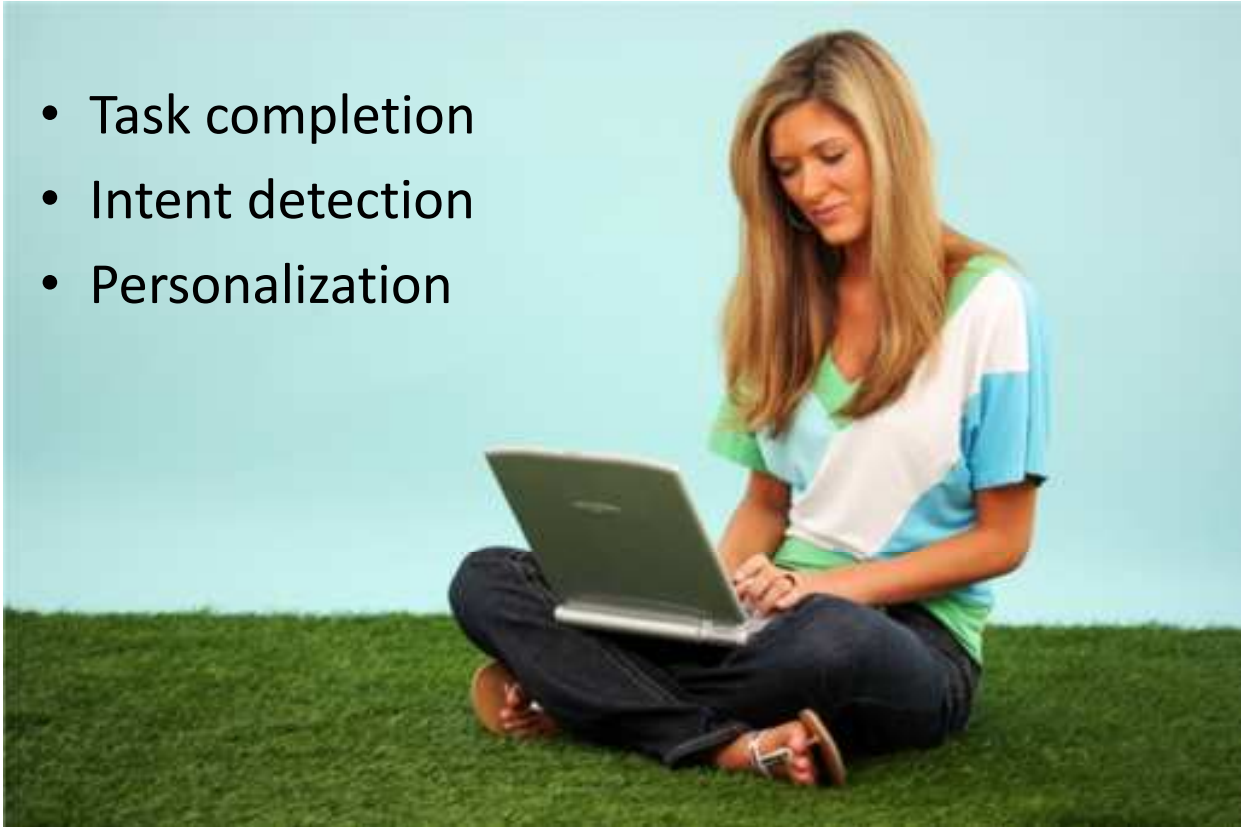


User Studies



The Future

- Task completion
- Intent detection
- Personalization



THE ENVIRONMENT

What does the web look like?

- What is its size?
- What content does it contain?
- How is content linked?
- How is content updated?
- Does it grow? How fast?
- Duplicates and links.



Odd stuff...

Crawler playing chess



Robots.txt

```
Please do not crawl this
site.
```

Evil stuff...

- What is spam?
- Motivation for spamming
 - Economic gain.
 - Pranks, vandalism and experiments.
- Types of spam
 - Content
 - Structure
 - User behaviour
- How bad is it?
- Other bad stuff



How to cope?



ENGINEERING

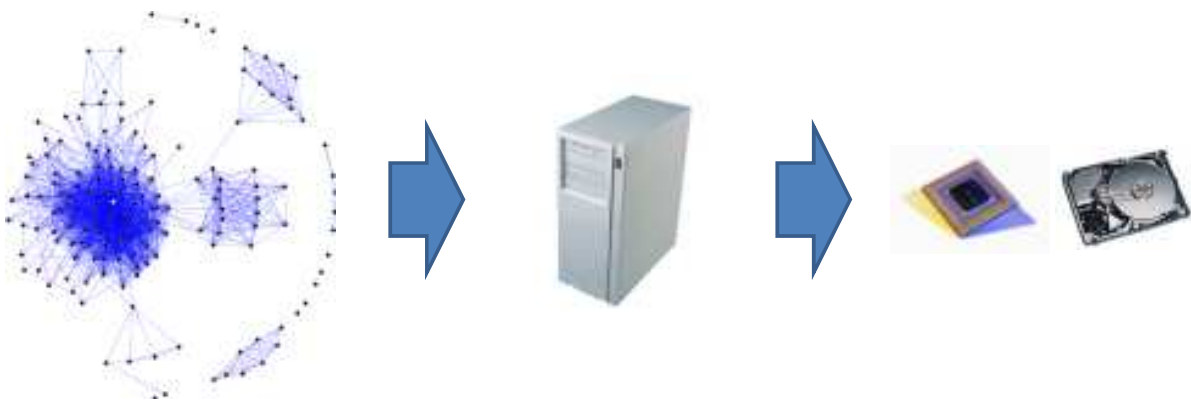
Engineering Complex Systems

«Anyone can build a fast CPU. The trick is to build a fast system.»

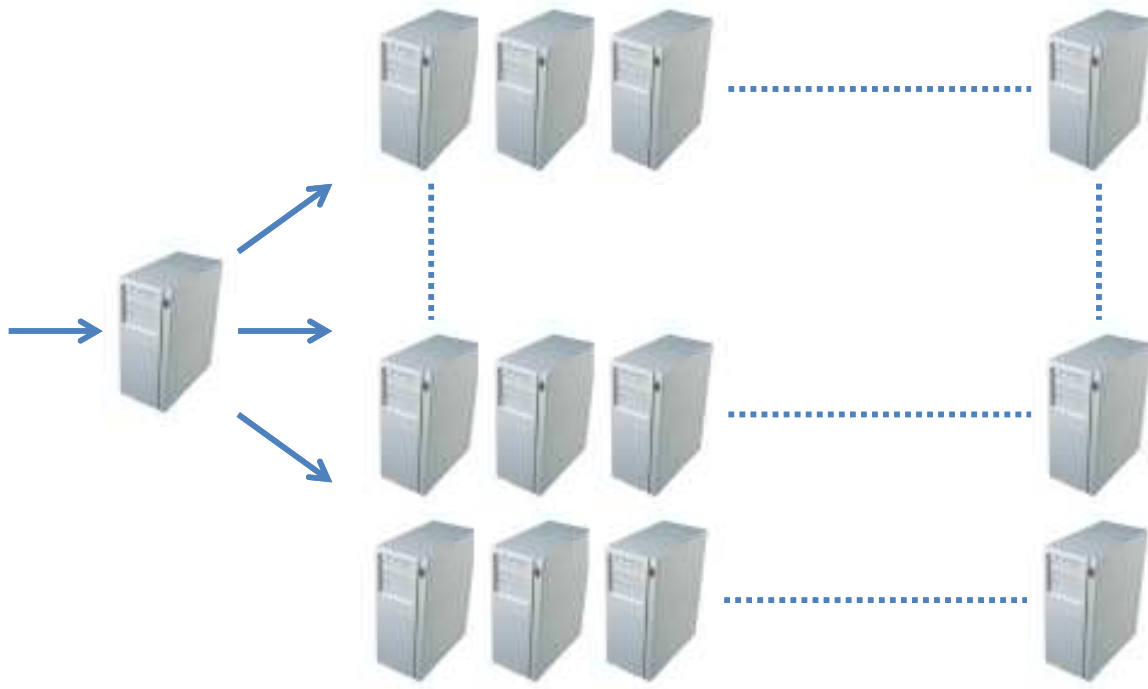
— Seymour Cray



Engineering for High Performance



Scale Drives Complexity



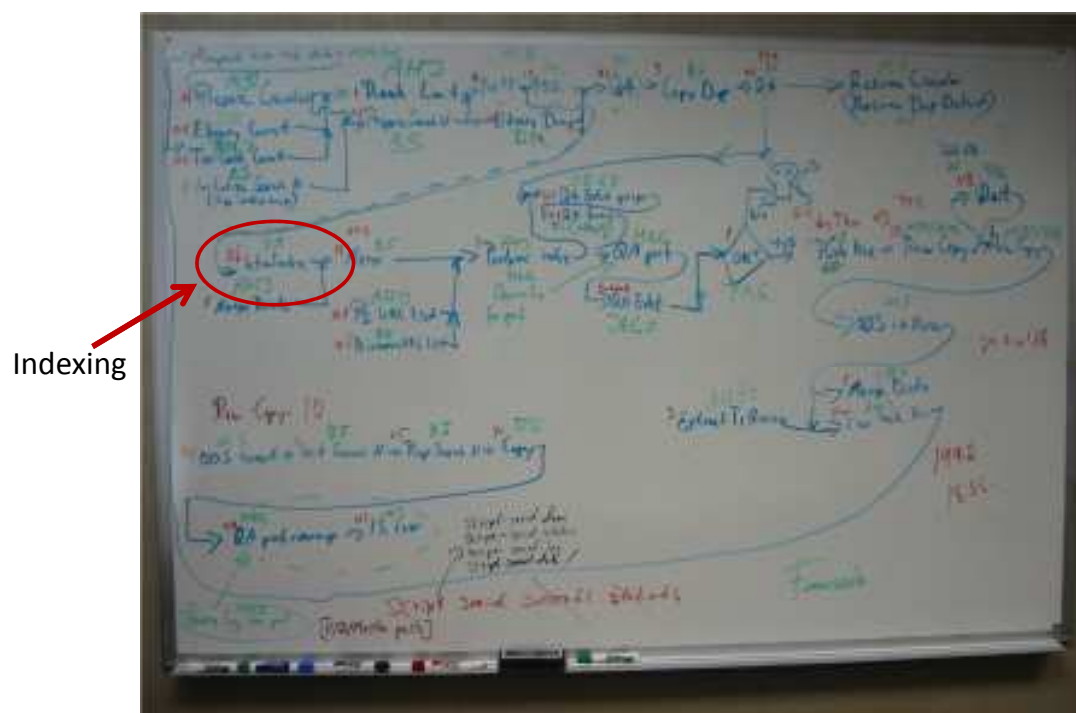
Datacenter



New Challenges

- Synchronization and consistency.
- High availability and fault tolerance.
- New bottlenecks.

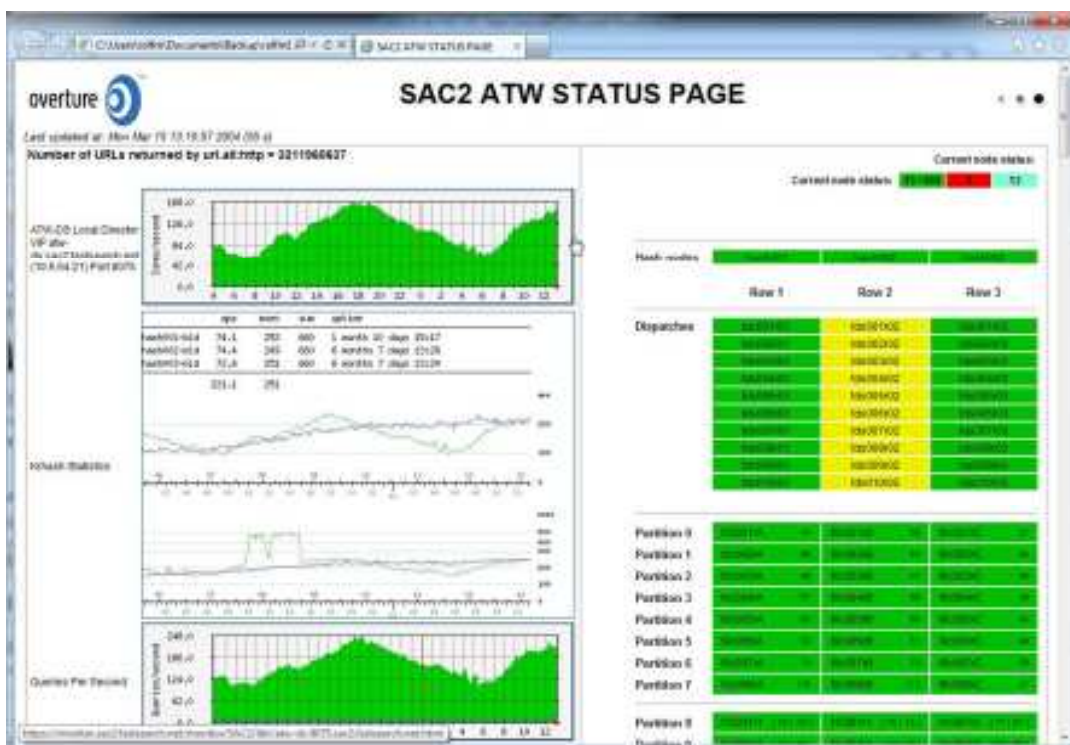
Indexing — How hard can it be?



Operations



Monitoring Operations



Modern datacenter



CONCLUSION

Advice from Old People

«The only constant is change.»

— Heraclitus (ca 535-475 BC)

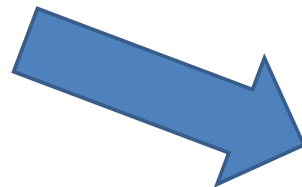
«Simplicity is the ultimate sophistication.»

— Leonardo da Vinci (1452-1519)

Real-World Application



Web Search



- Application
- Environment
- System
- Disciplines
- Change