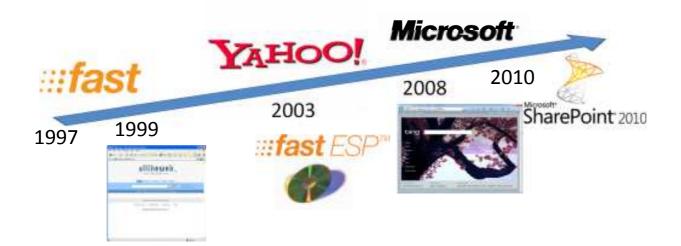
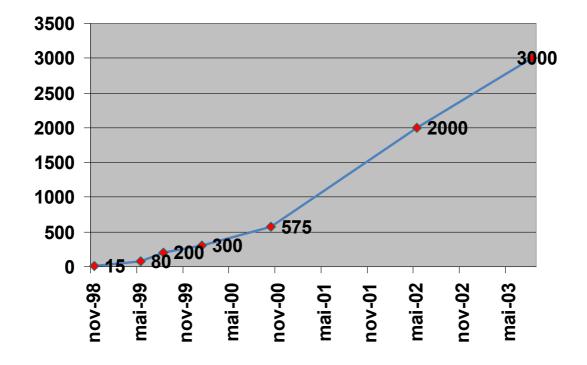


Background



Web Search Development

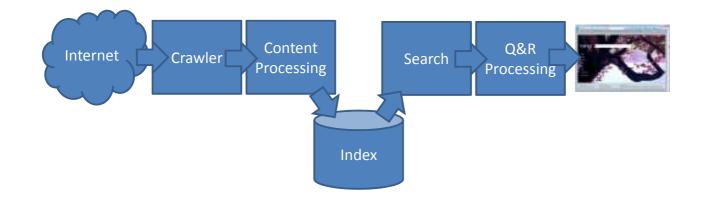


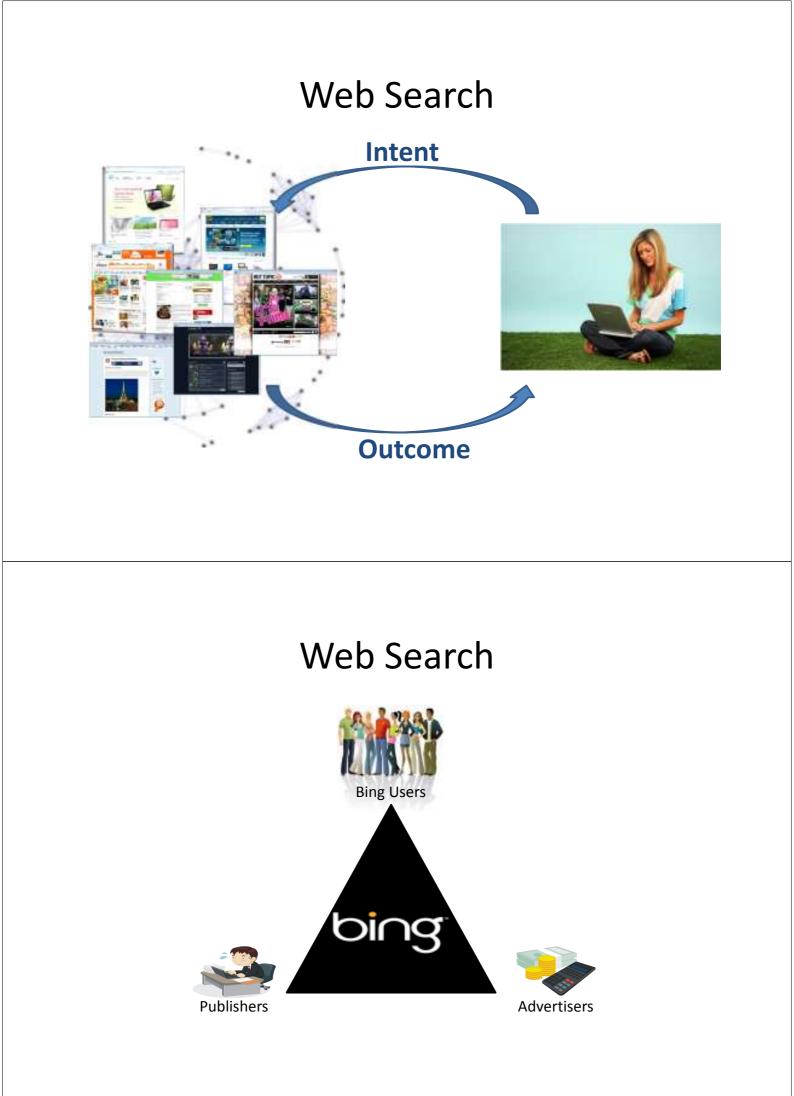
Topics

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

WHAT IS 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



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





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?



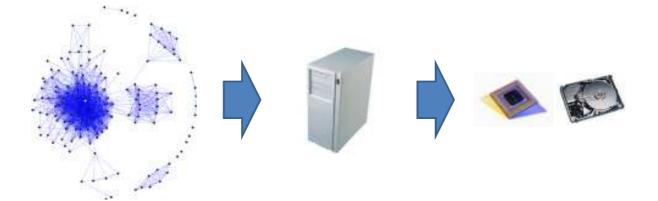
ENGINEERING

Engineering Complex Systems

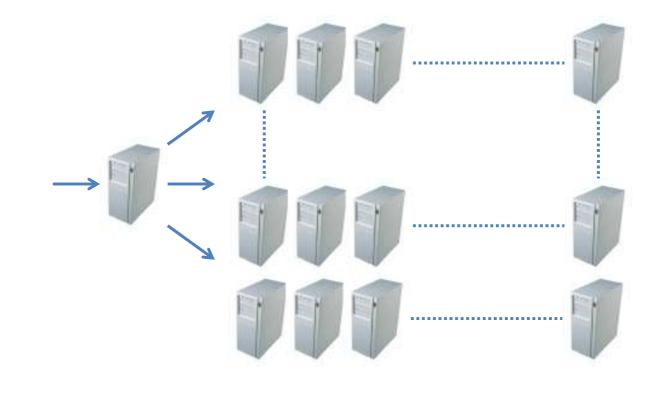
«Anyone can build a fast CPU. The trick is to build a fast system.» — Seymour Cray



Engineering for High Performance



Scale Drives Complexity



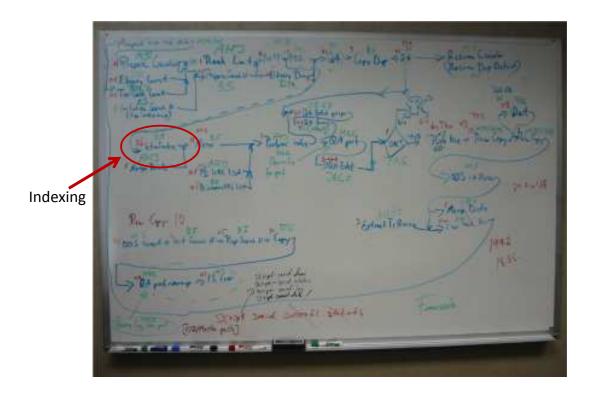
Datacenter



New Challenges

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

Indexing — How hard can it be?



Operations



Monitoring Operations

No-OP Long Description Image: Description of the Descriptio	werture 🧿	SAC2 ATW ST	ATUS PAG	θE		
	Uniber of URLs re	samed by unit all map = 3211946437	Carrent sole eletes Carrent sole eletes			
Nation Provided 24.1 201 800 5 and	/P 204- 10 2017 10:00 2004 000 10 7.04 21) P 91 8075	A see	Handi window	Rev 1	Row 2	Rev 3
Person 1 Per	out takine	and the set of the set	Disparshwa	anti) (10 Marine Antine Antine Salari Salari Salari Salari Salari Salari Salari Salari	Hantichtod Hantichtarta Hantichtarta Hantikhenig Hantikhenig Hantikhenig Hantikhenig	
And the second s			Partition 1 Partition 2 Partition 3			and the second se
	artes Per Tyrini (THE REAL PROPERTY AND ADDRESS OF THE PARTY O	Partition 5 Partition 6	Second 1	NUMBER OF	Anna Anna Anna Anna Anna Anna Anna Anna

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

