Distributed Web Crawler COMP 599 - Graduate Seminar, Spring 2019

Rihan Pereira, MSCS

Department of Computer Science California State University, Channel Islands

April 9, 2019

References

Example
Single node crawler (block diagram)
Indexer
Challenges faced when crawling

Web Crawling

type something to see how the font looks

References

Example

Single node crawler (block diagram) Indexer Challenges faced when crawling

Example

3/14

References

Example
Single node crawler (block diagram)
Indexer
Challenges faced when crawling

Single node crawler (block diagram)

References

Example
Single node crawler (block diagram)
Indexer
Challenges faced when crawling

Indexer

References

Example
Single node crawler (block diagram)
Indexer
Challenges faced when crawling

Challenges faced when crawling

The vision

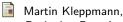
Distributed web crawler

Distributed Indexing

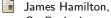
Presentation tier

Thoughts on infrastructure setup

11/14



Designing Data Intensive applications, 2017



On Designing and Deploying Internet-Scale Services, 2007.



Automatic Management of Partitioned, Replicated Search Services, 2011.



http://aosabook.org,

The architecture of open source applications.

http://book.mixu.net/distsys,

Distributed Systems: fun and profit.

13/14

Thank you! Questions?