Accelerating Text Analytics Queries on Reconfigurable Platforms

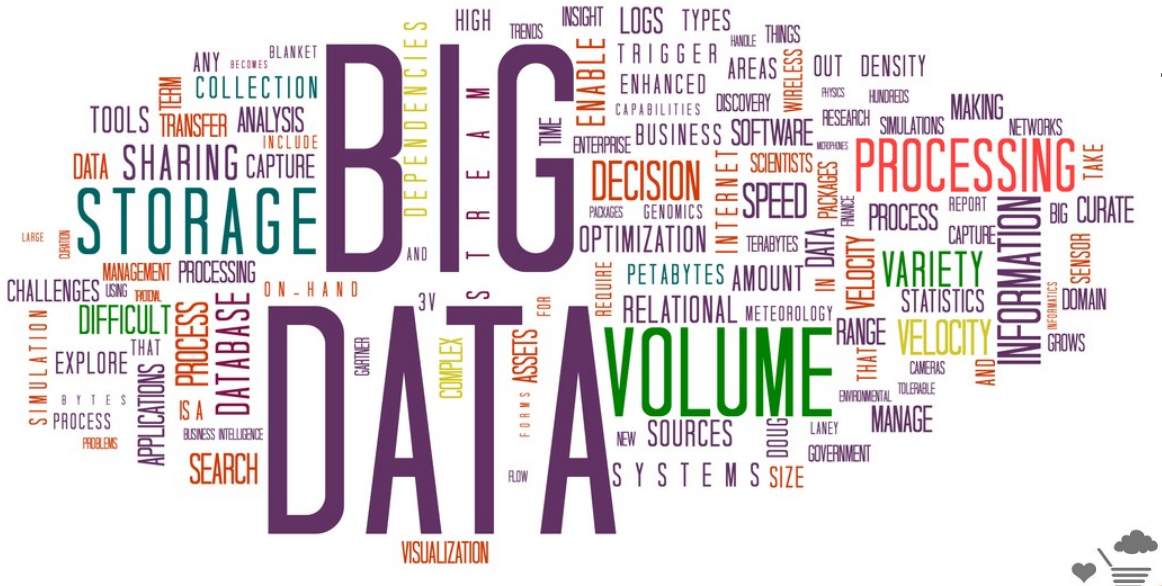
Kubilay Atasu, Raphael Polig, Christoph Hagleitner, H. Peter Hofstee
Laura Chiticariu, Frederick Reiss, Huaiyu Zhu, Cesar Berrospi

June 14, 2015 @ CARL Workshop

Outline

- **Introduction**
- Text analytics use cases
- SystemT text analytics software
- HW-accelerated SystemT
- HW-accelerated regex matching
- Conclusions

The importance of performance



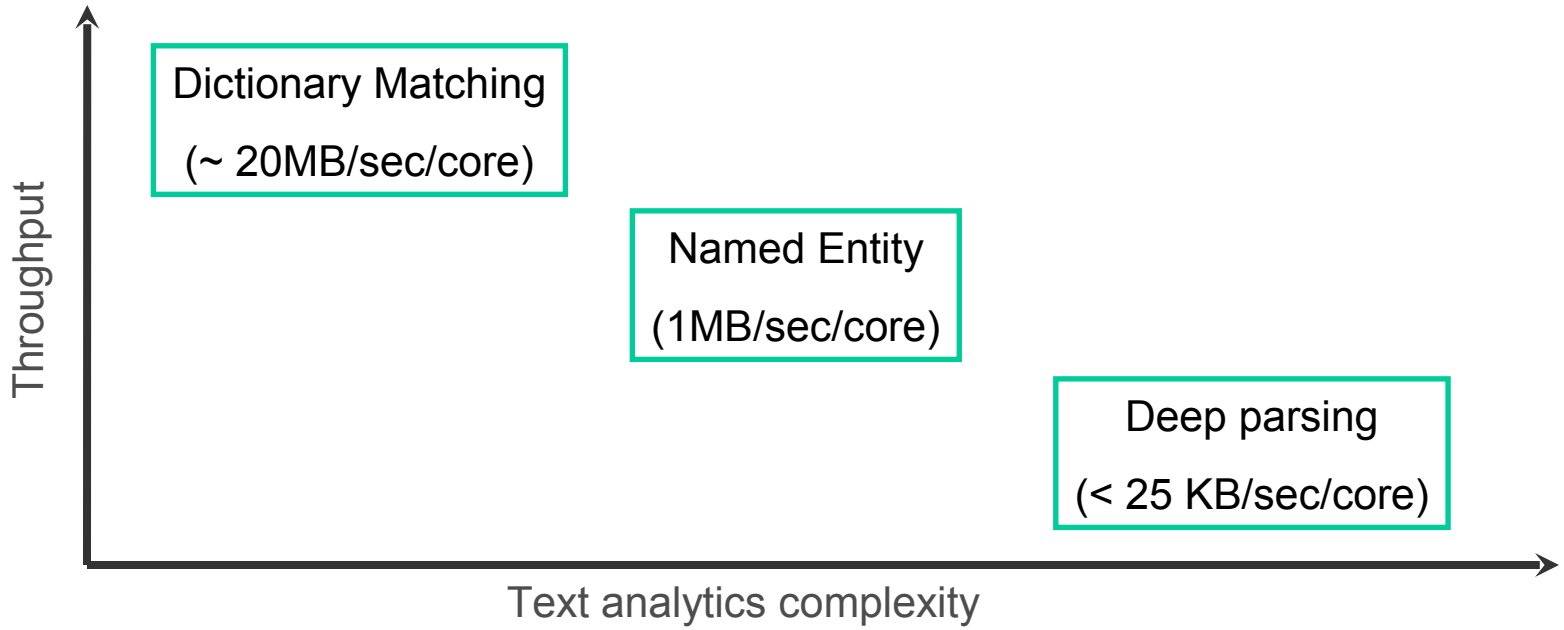
– Social Media Data
  - 1+ TB Twitter data per day
  - 400+ TB per year

– Financial Data
  - Regulatory filings can be in tens of millions and several TBs

– Machine data
  - 1GB of app server logs per day
  - A medium-size data center has tens of thousands of servers → Tens of Terabytes of system logs per day

Source: UCSC Lecture on Information Extraction by F. Reiss, L. Chiticariu, Y. Li, 2014
Big Data image by Camelia Boban; Social Media image by Yoel Ben-Avraham

Spectrum of throughput



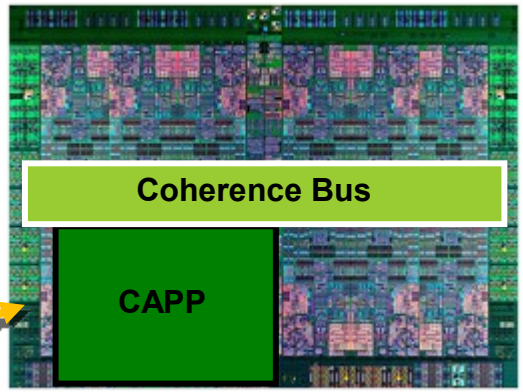
- **The more complex the task, the slower the runtime performance**
- **But the higher the information accuracy**

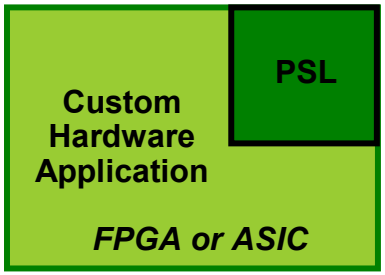
Source: UCSC Lecture on Information Extraction by F. Reiss, L. Chiticariu, Y. Li, 2014

Coherent Accelerator Processor Interface (CAPI)

Virtual Addressing
- Accelerator can work with same memory addresses that the processors use
- Pointers de-referenced same as the host application
- Removes OS & device driver overhead

Hardware Managed Cache Coherence
- Enables the accelerator to participate in "Locks" as a normal thread
- Lowers Latency over IO communication model





PCIe Gen 3
Transport for encapsulated messages

Processor Service Layer (PSL)
- Present robust, durable interfaces to applications
- Offload complexity / content from CAPP

Customizable Hardware Application Accelerator
- Specific system SW, middleware, or user application
- Written to durable interface provided by PSL

Hardware-accelerated text analytics

IBM InfoSphere BigInsights

AQL Query

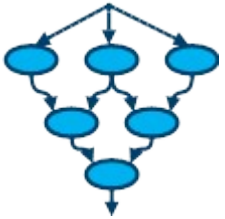
IBM InfoSphere Streams

SystemT Text Analytics Compiler & Runtime

POWER8

CAPI

FPGA



Outline

- Introduction
- **Text analytics use cases**
- SystemT text analytics software
- HW-accelerated SystemT
- HW-accelerated regex matching
- Conclusions

Text analytics - Example

For years, Microsoft Corporation CEO Bill Gates was against open source. But today he appears to have changed his mind. “We can be open source. We love the concept of shared source,” said Bill Veghte, a Microsoft VP. “That is a super-important shift for us in terms of code access.”

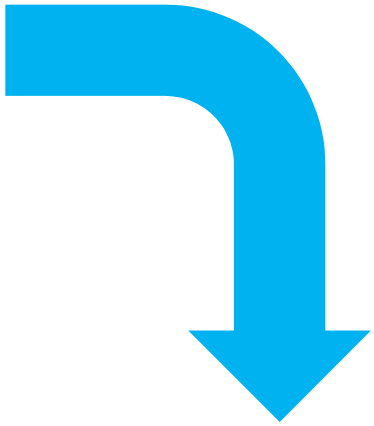
Richard Stallman, founder of the Free Software Foundation, countered saying …

Source: Example by Cohen, 2003

Text analytics - Example

For years, Microsoft Corporation CEO Bill Gates was against open source. But today he appears to have changed his mind. “We can be open source. We love the concept of shared source,” said Bill Veghte, a Microsoft VP. “That is a super-important shift for us in terms of code access.”

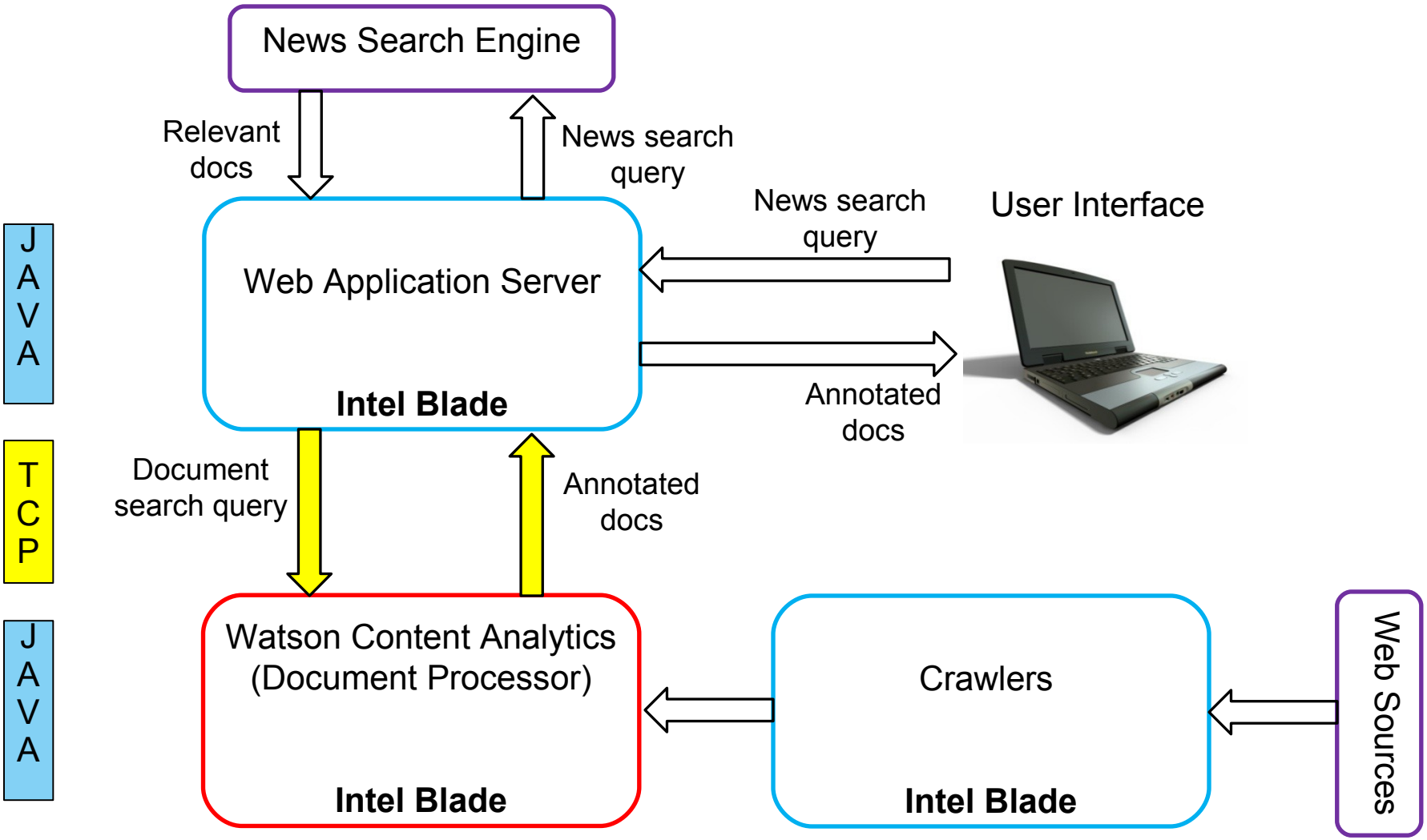
Richard Stallman, founder of the Free Software Foundation, countered saying …



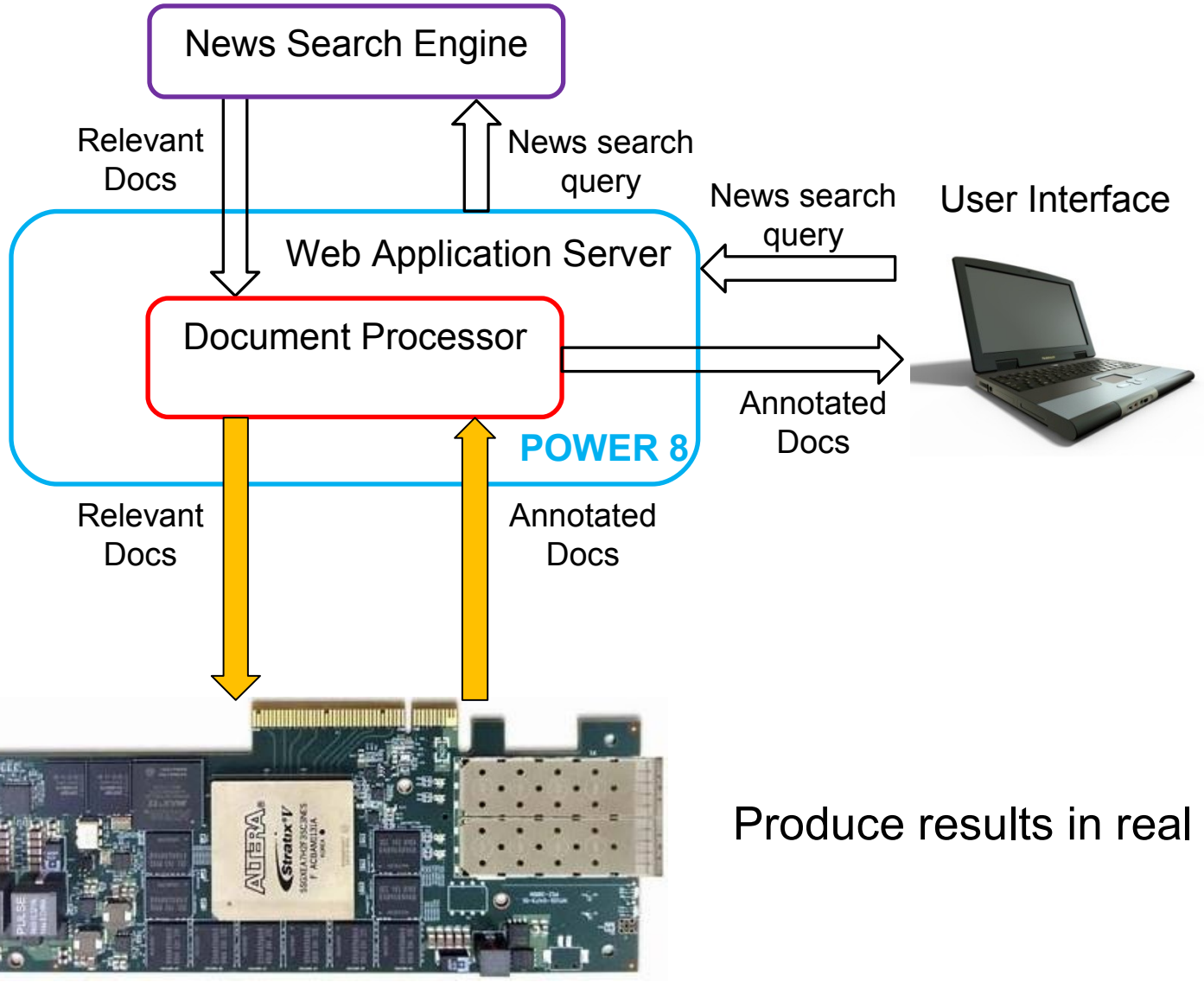
Name	Title	Organization
Bill Gates	CEO	Microsoft
Bill Veghte	VP	Microsoft
Richard	Founder	Free

Source: Example by Cohen, 2003

Text analytics in IBM Crystal+: news search services



Acceleration of Crystal+ news search services

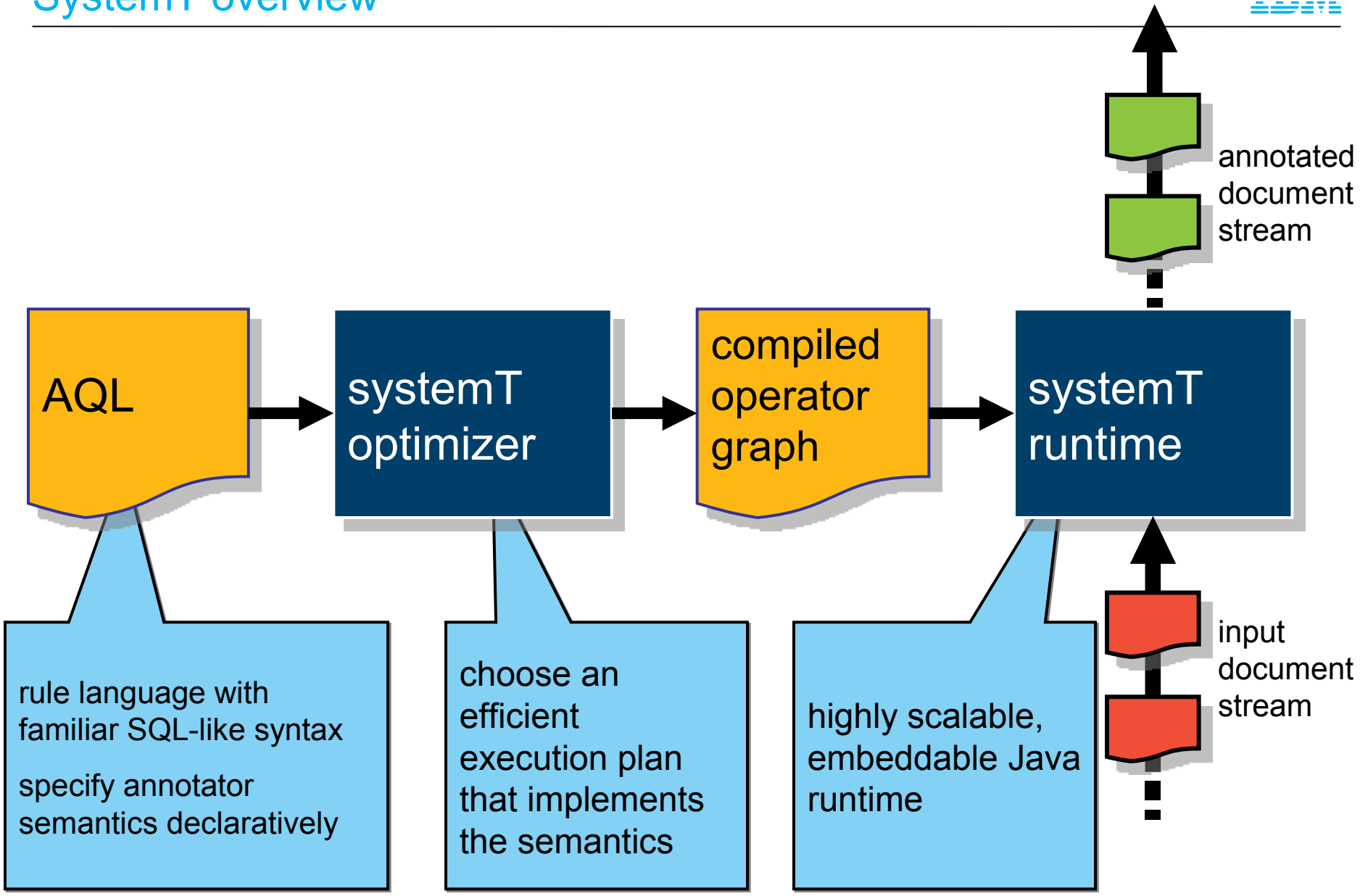


Produce results in real time!

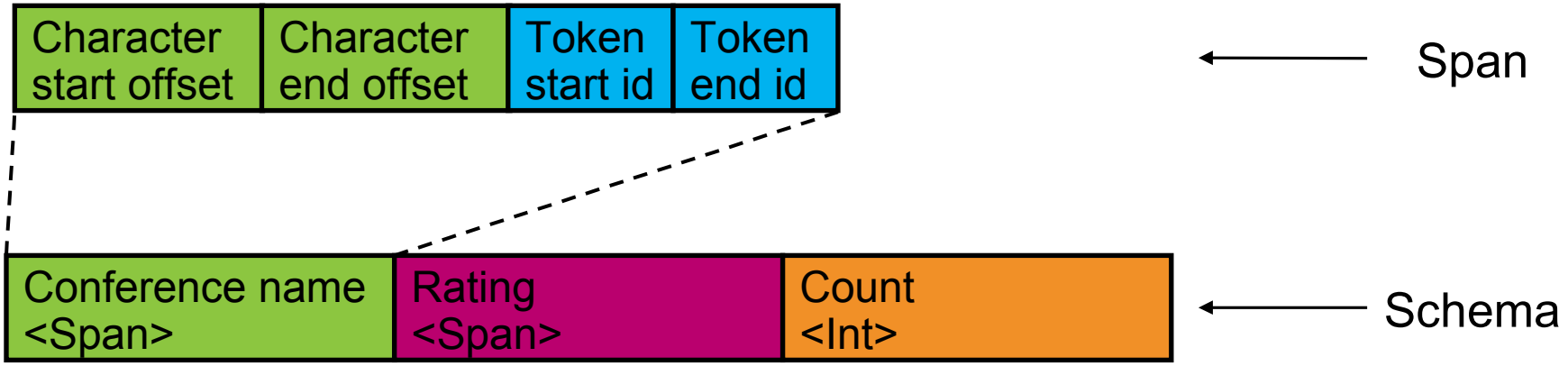
Outline

- Introduction
- Text analytics use cases
- **SystemT text analytics software**
- HW-accelerated SystemT
- HW-accelerated regex matching
- Conclusions

SystemT overview

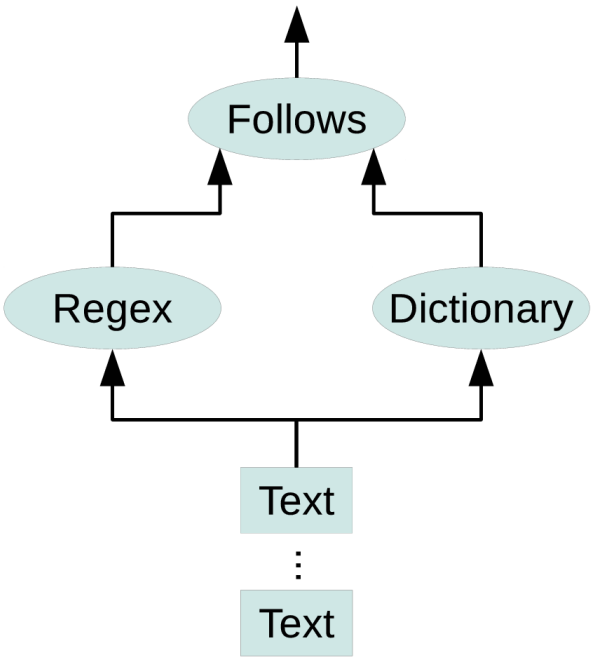


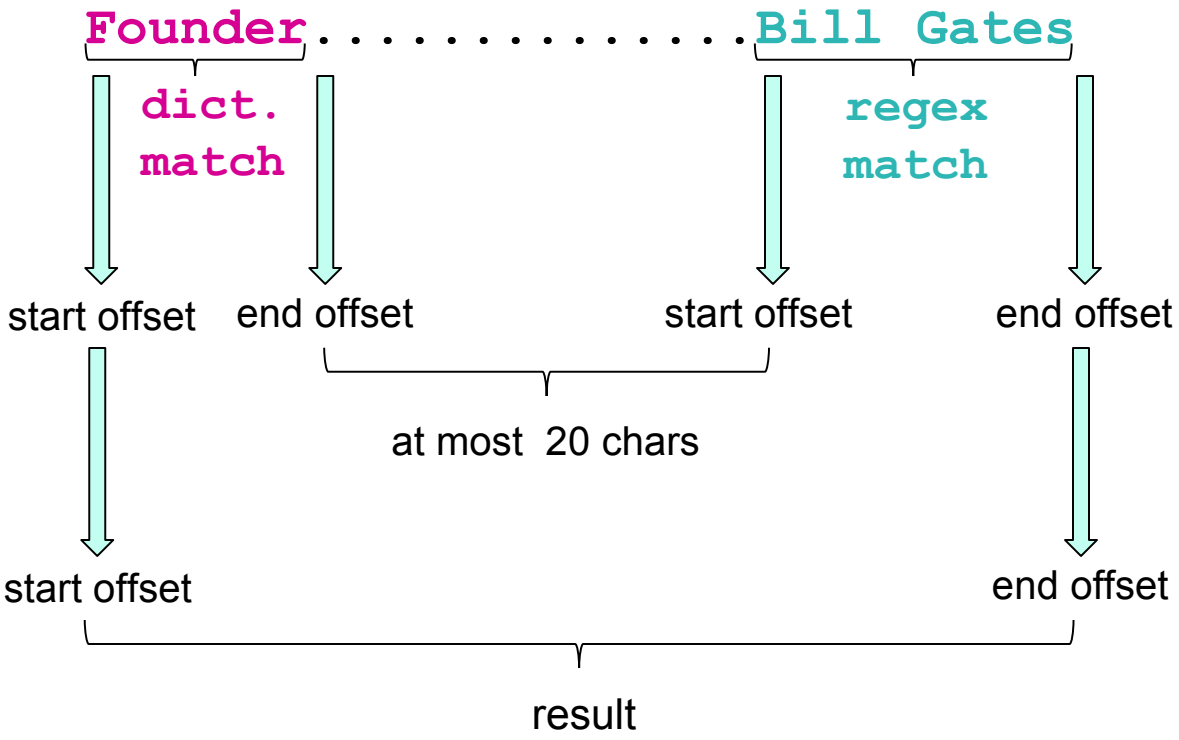
Basic data structures



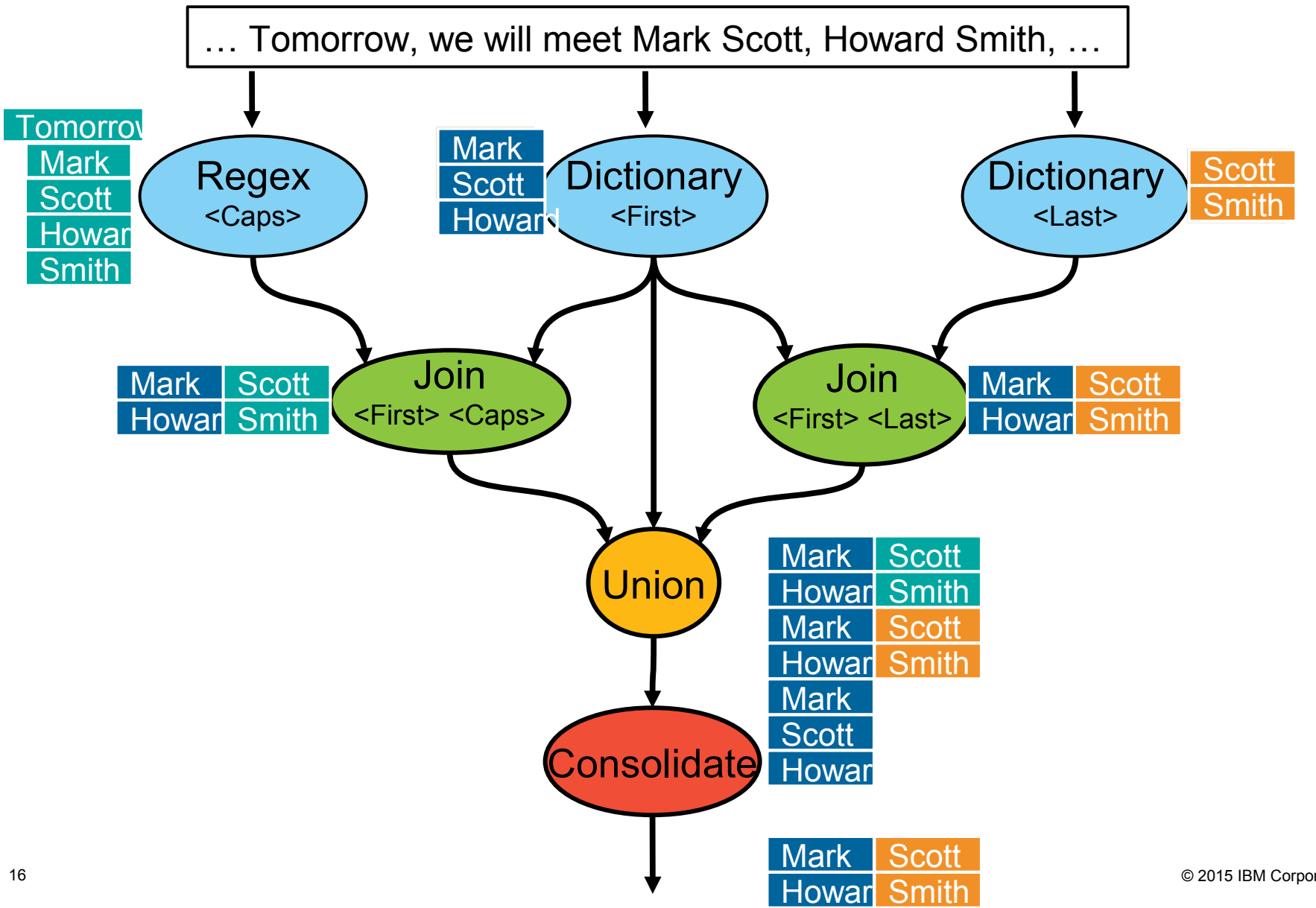
A simple SystemT information extraction rule

- Find the names (**regex**) that are at most 20 chars after a title (**dict.**)

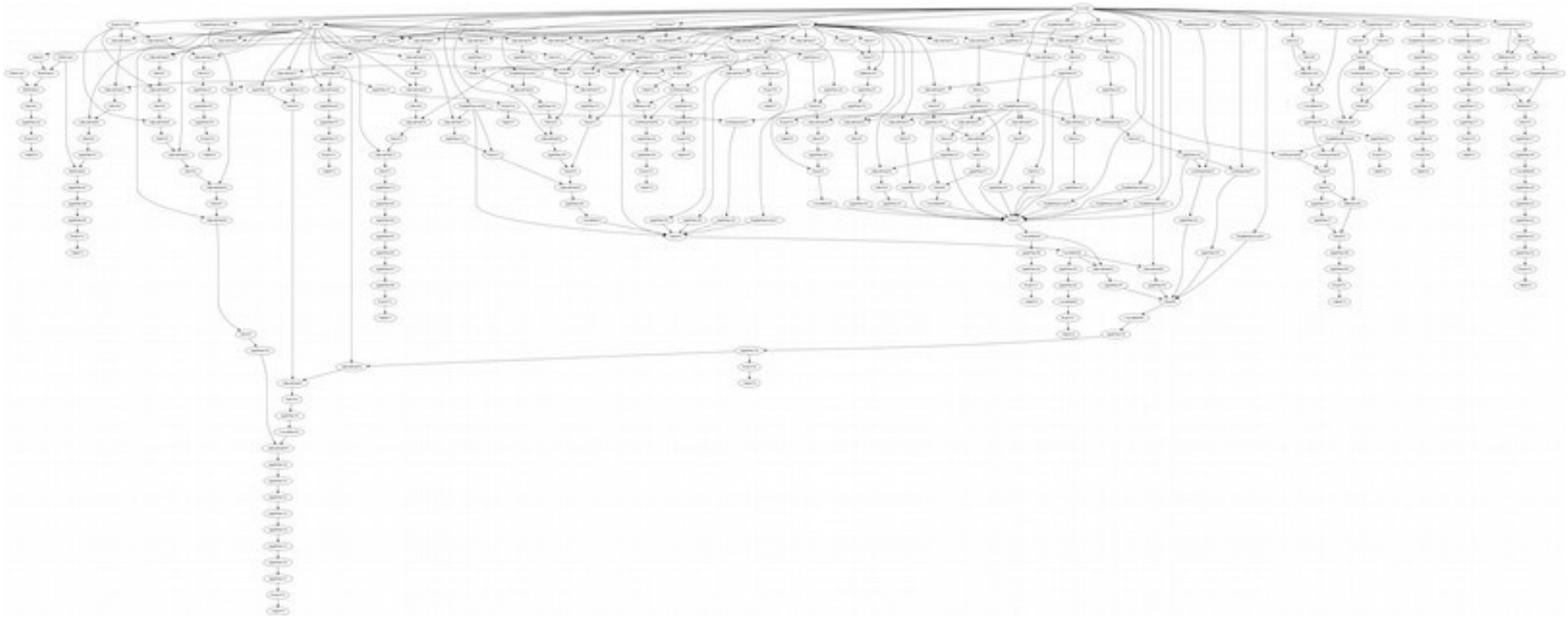




Annotation Operator Graph (AOG)



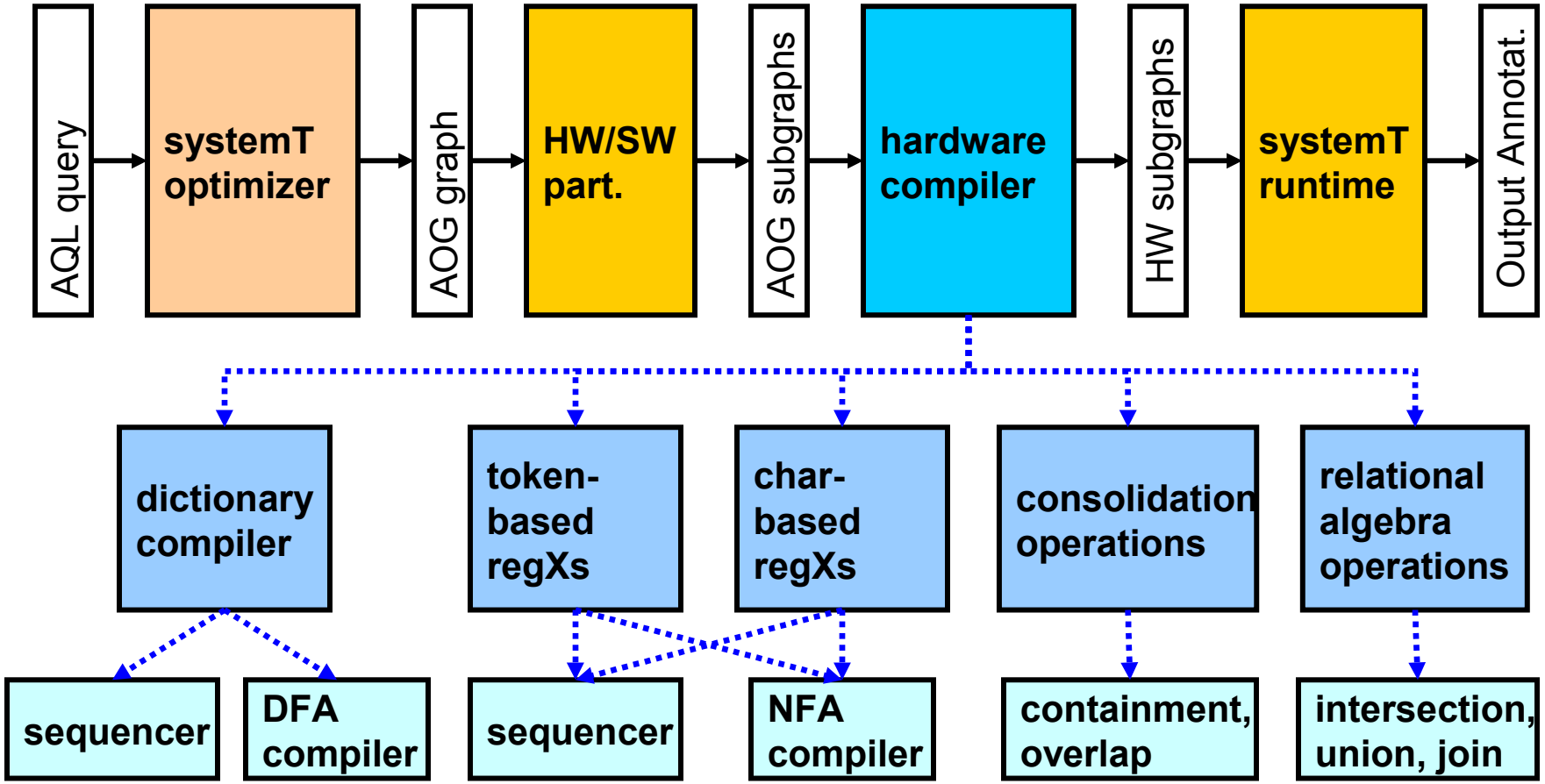
AOG of a real-life SystemT IE query



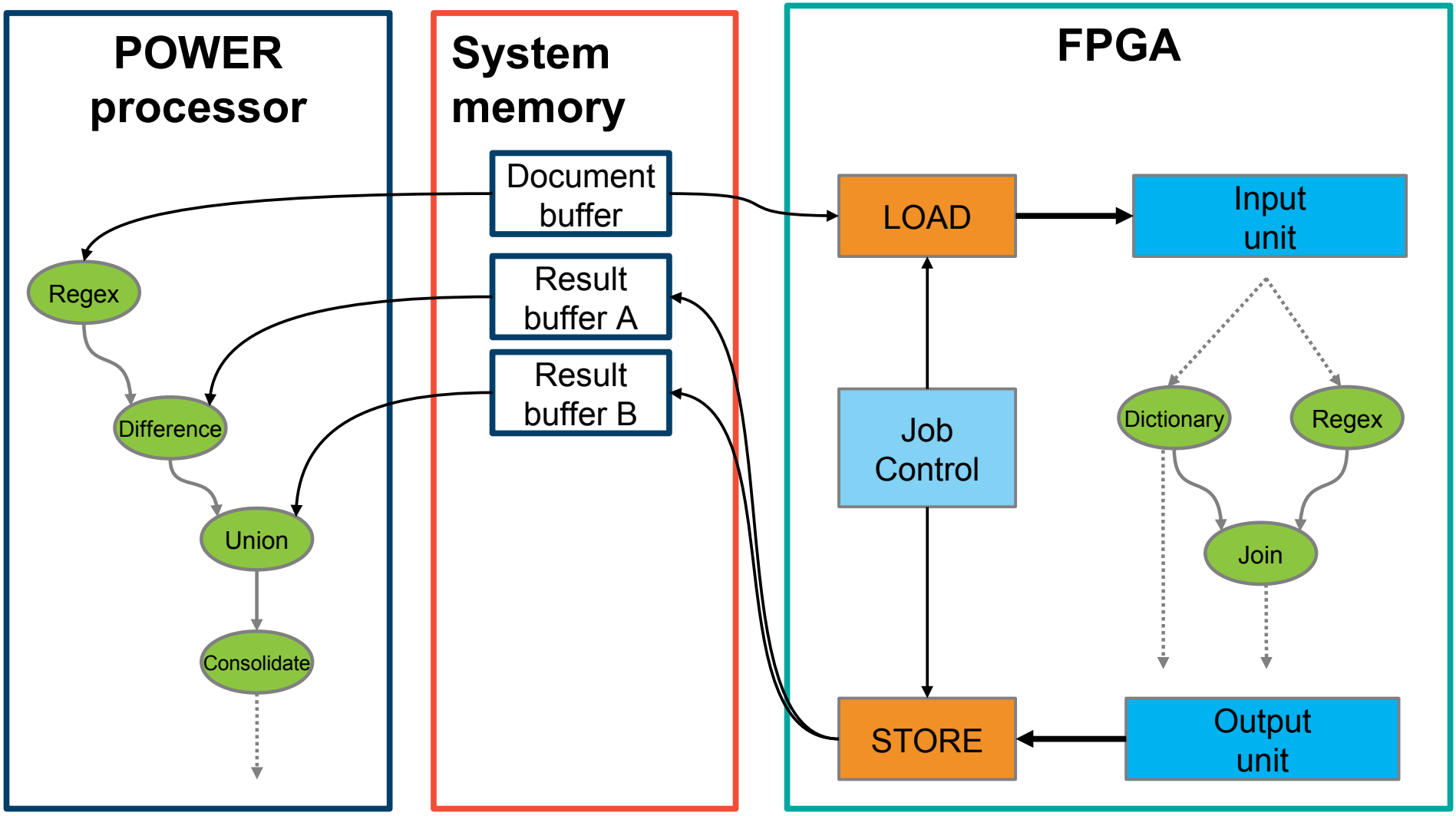
Outline

- Introduction
- Text analytics use cases
- SystemT text analytics software
- **HW-accelerated SystemT**
- HW-accelerated regex matching
- Conclusions

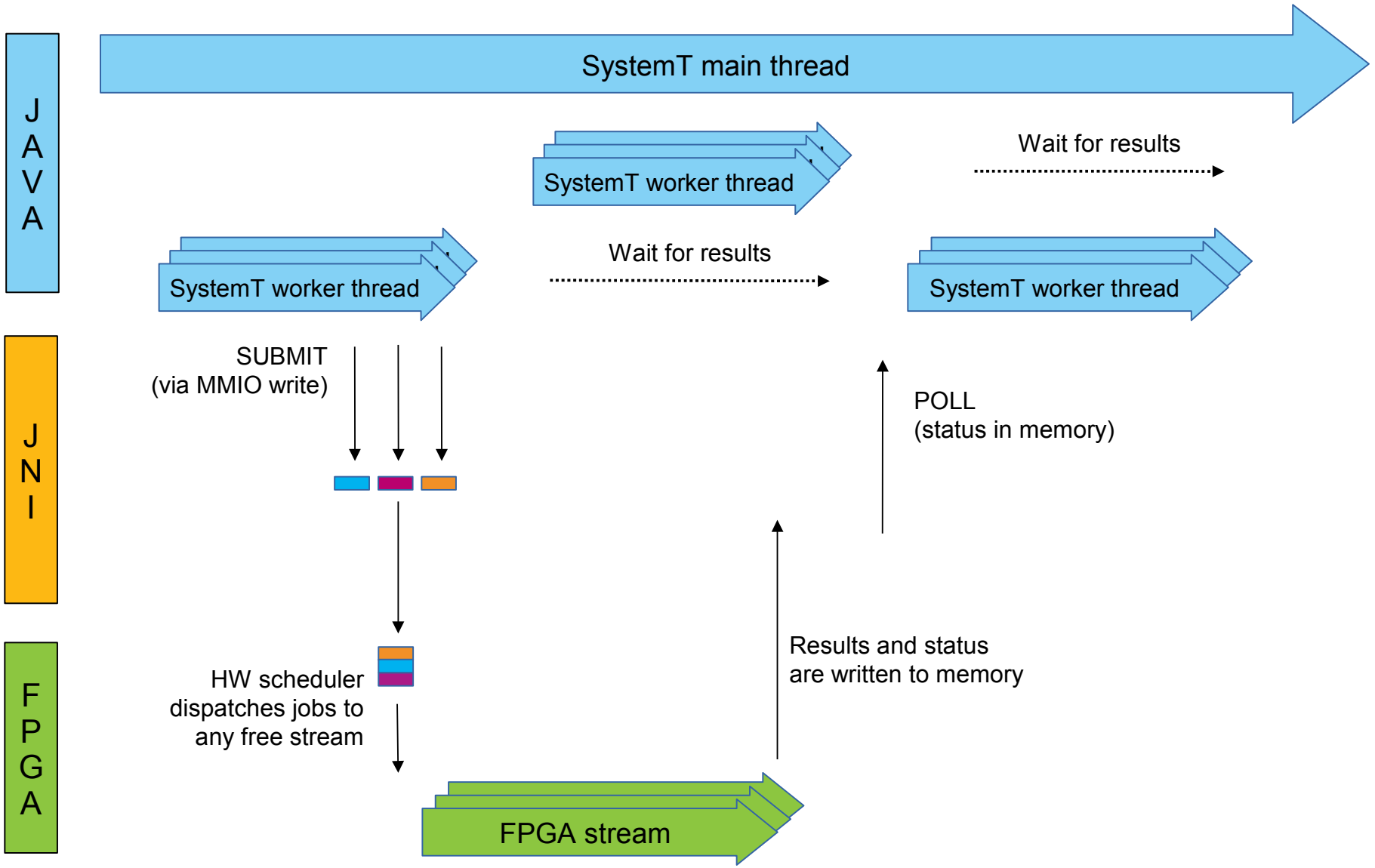
SystemT acceleration: Building blocks



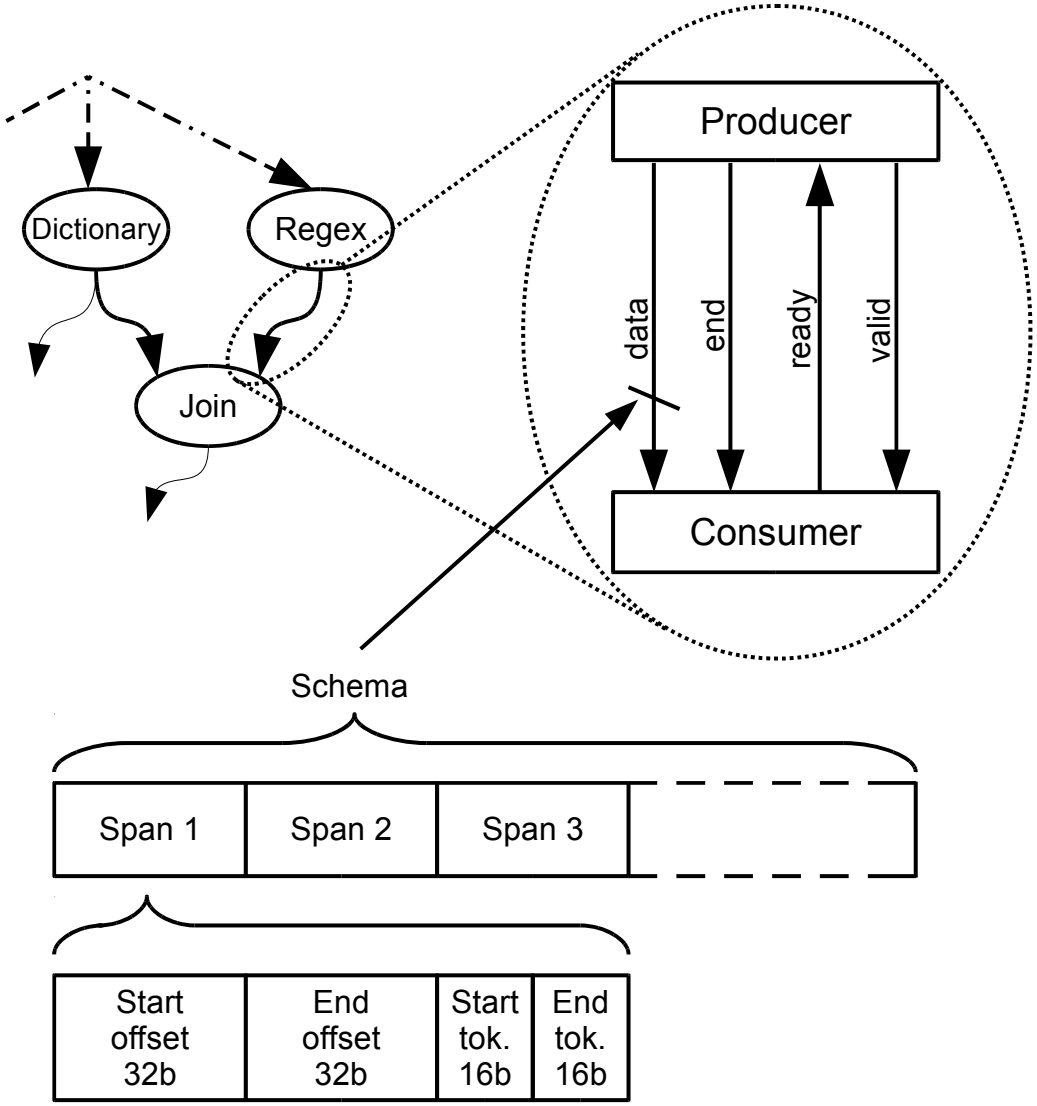
SystemT acceleration: Partitioned system



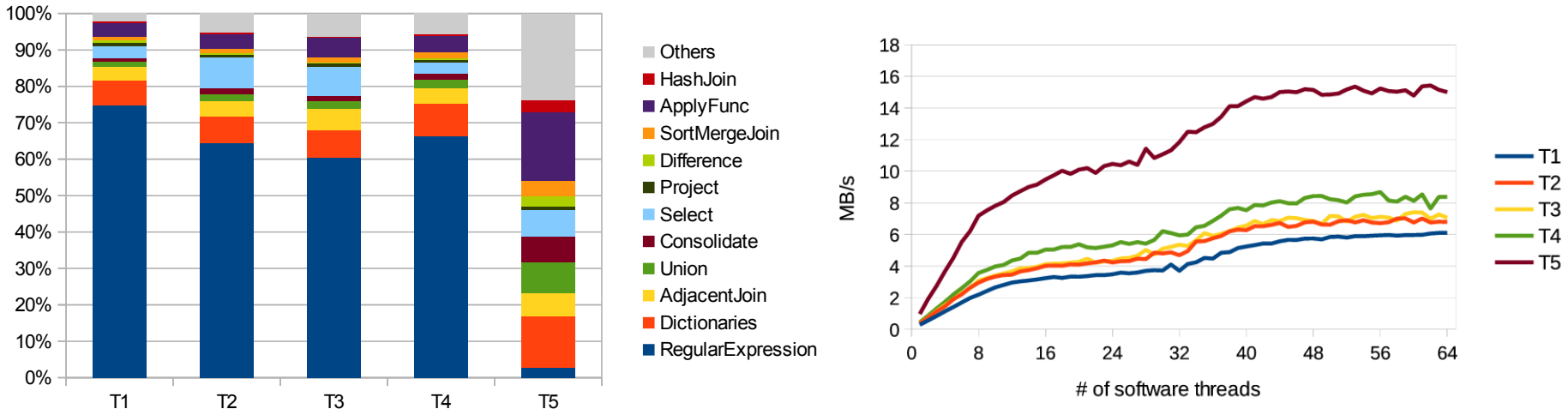
Multithreaded communication scheme



Elastic hardware interface



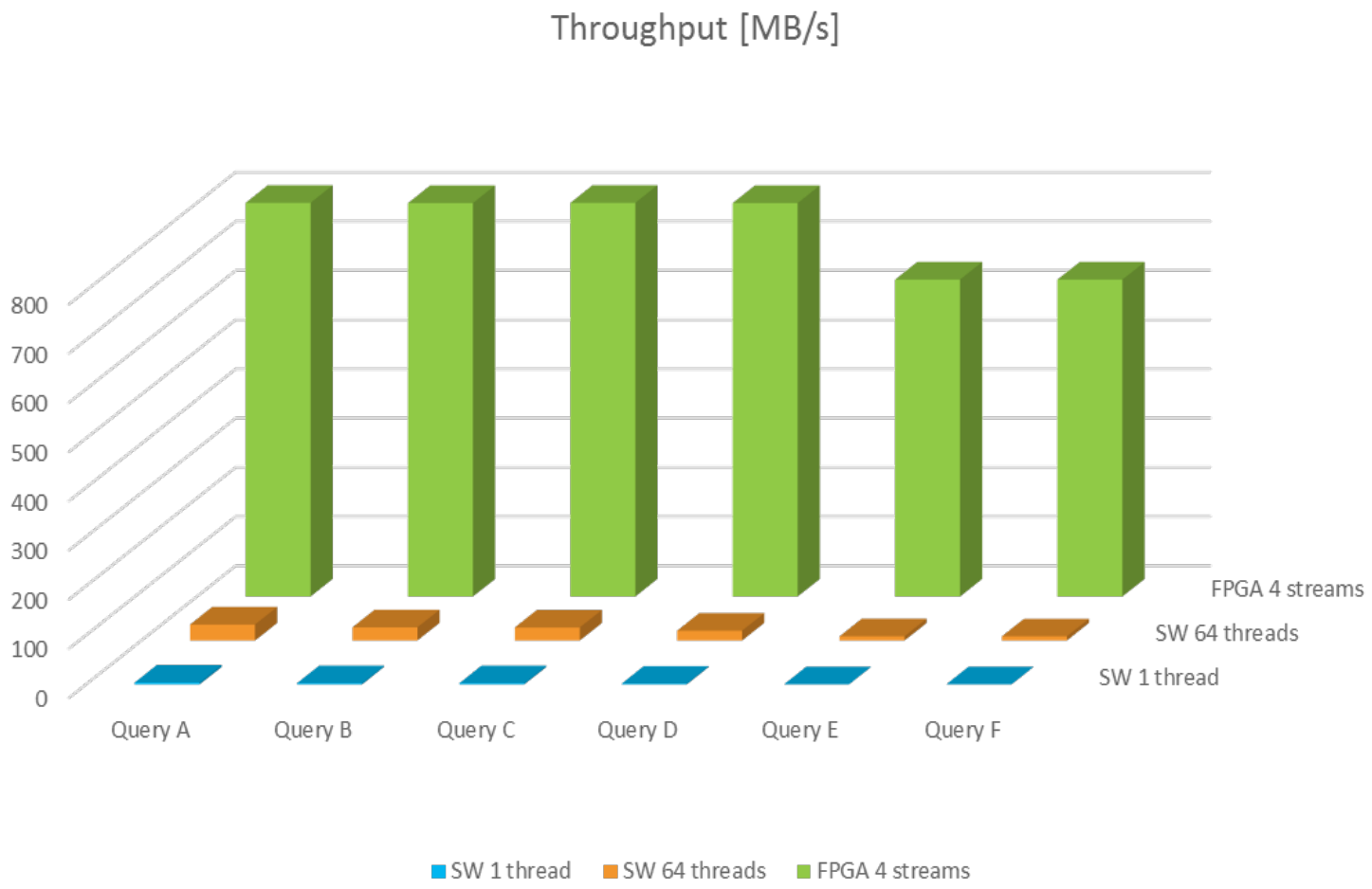
IEEE MICRO paper: Software profiling results (no acceleration)



- Measurements on a two socket POWER7 server with 8 cores per CPU @3.55GHz

R. Polig, K. Atasu, C. Hagleitner, L. Chiticariu, F. R. Reiss, H. Zhu, H. P. Hofstee: Giving text analytics a boost. IEEE MICRO Special Issue on Big Data, 2014.

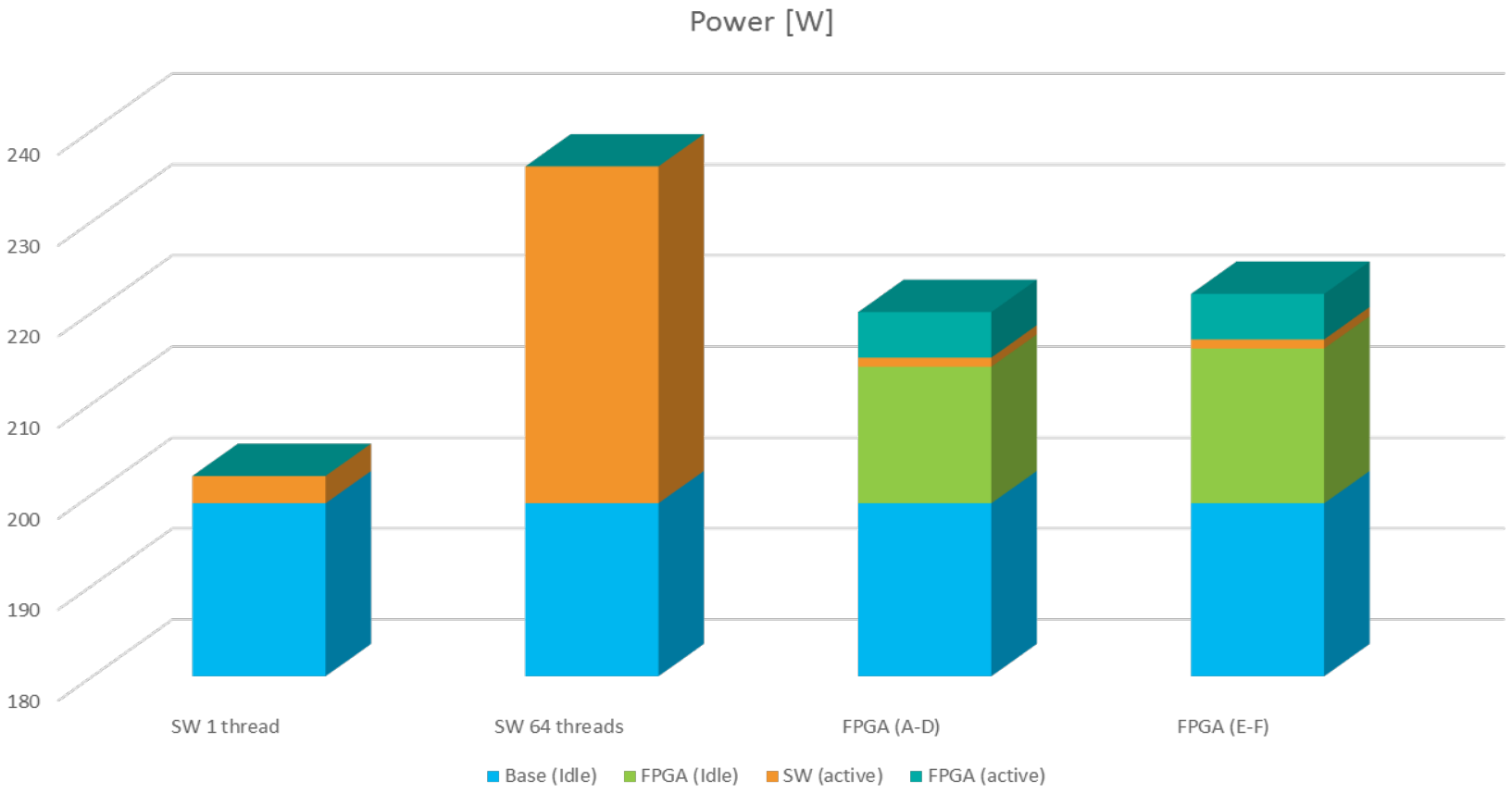
FPL 2014 paper: Measured throughput rates



- The evaluated queries are completely offloaded to FPGA logic (no partitioning)

R. Polig, K. Atasu, H. Giefers, L.Chiticariu: Compiling Text Analytics Queries to FPGAs. FPL 2014

FPL 2014 paper: Measured power consumption

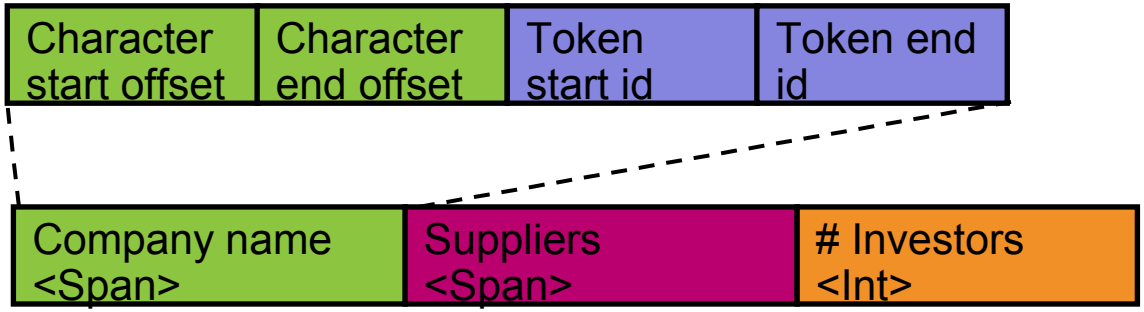


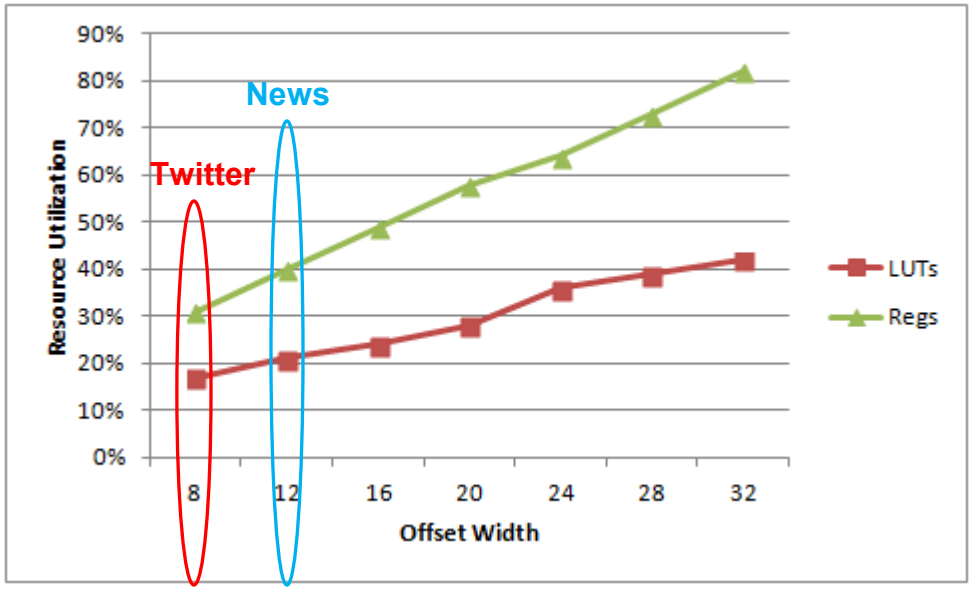
- The evaluated queries are completely offloaded to FPGA logic (no partitioning)

R. Polig, K. Atasu, H. Giefers, L.Chiticariu: Compiling Text Analytics Queries to FPGAs. FPL 2014

FPGA resource utilization: Effect of reducing the offset width

- Basic data structures: spans and schemas





- Scalability is limited by BRAM & register usage

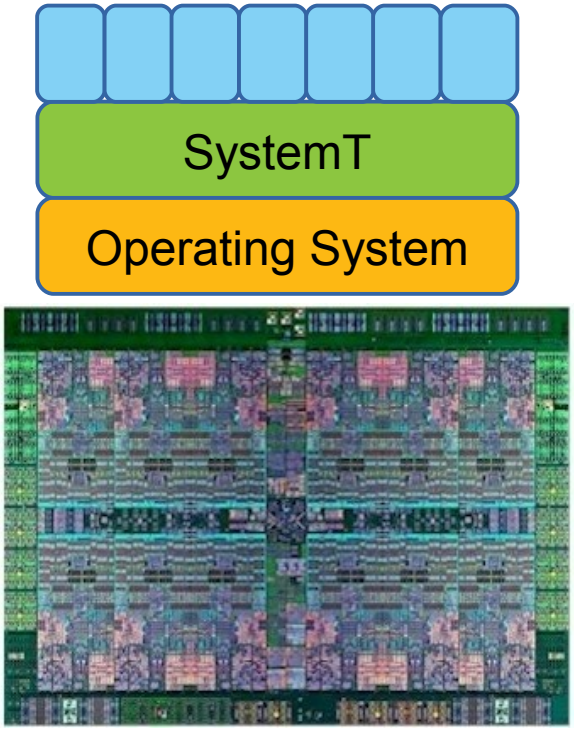
- 4 pipelines: 4 bytes/cycle

- 200 MHz → 800 MB/s

Live demo of Crystal+ acceleration on POWER8

- Higher SW performance
- CAPI system
- Virtual addressing from user FPGA
- Multiple cards per CPU

Stratix V GX A7



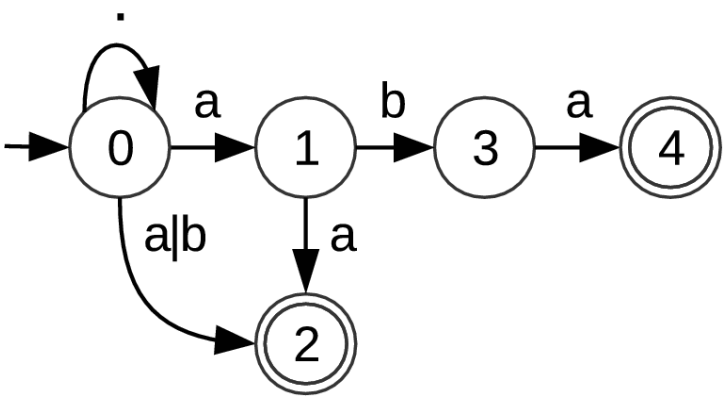
POWER 8

Outline

- Introduction
- Text analytics use cases
- SystemT text analytics software
- Hardware-accelerated SystemT
- **HW-accelerated regex matching**
- Conclusions

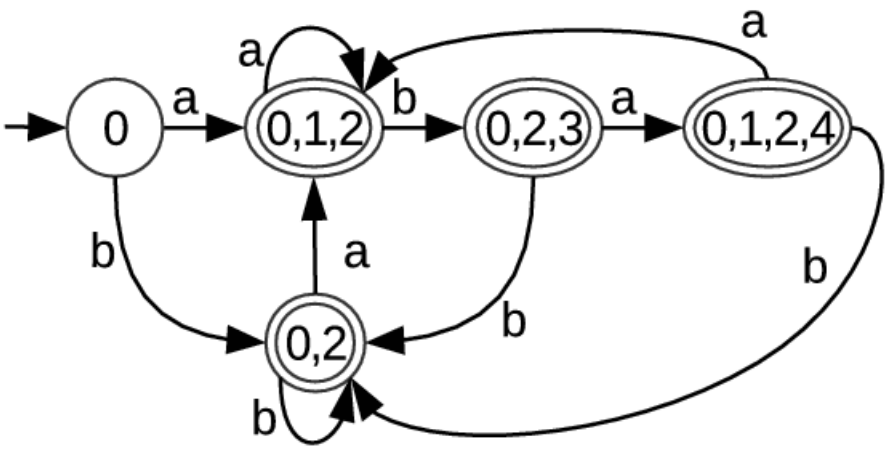
Regex matching: Background

- Consider the regex .*(a|b|aa|aba)



NFA

- Can be transformed into NFA/DFA



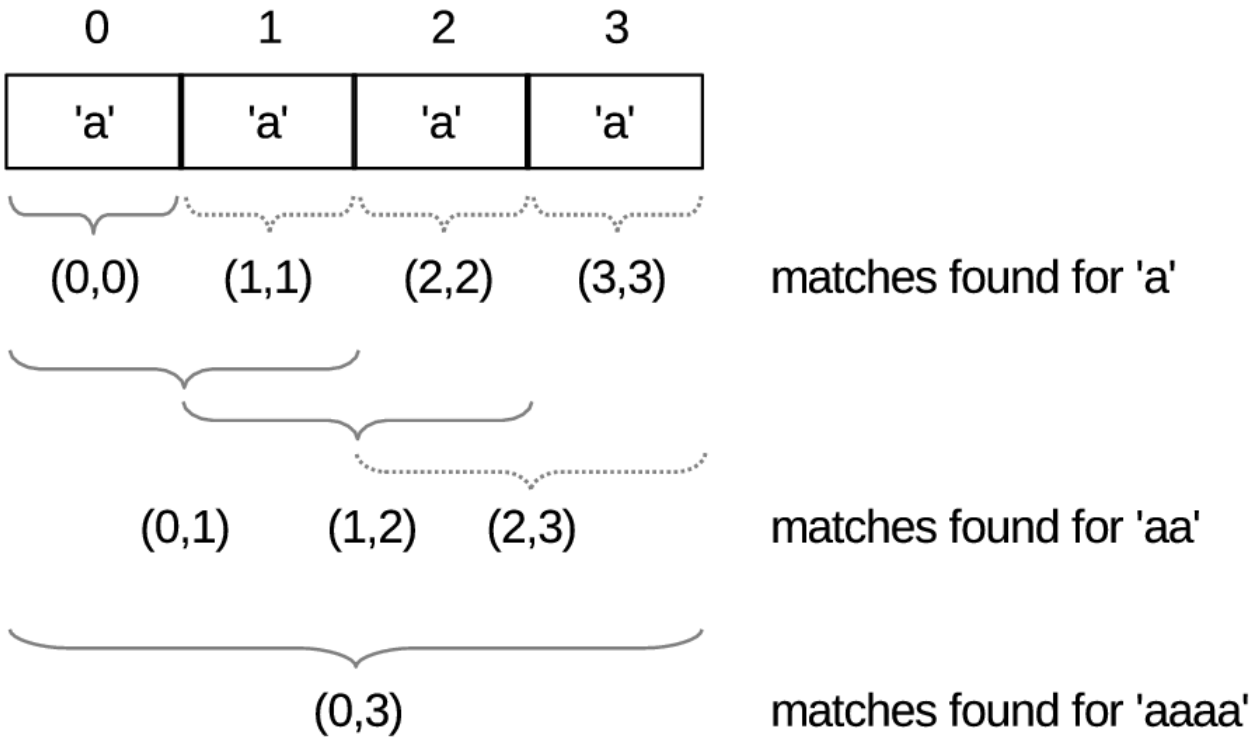
DFA

Traditional architectures do not support start offset reporting & leftmost matching:

- Reconfigurable NFAs (Sidhu FCCM 2001, Bispo FPT 2006 , Yang ANCS 2008)
- Programmable DFAs (Smith SIGCOMM 2008, Van Lunteren MICRO 2012)

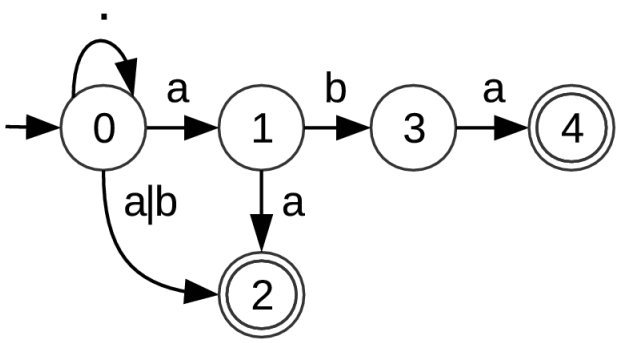
Finding leftmost regular expression matches

- Assume that we are searching for the regex .*(a|aa|aaaa) in the input string “aaaa”
- Find the regex match with the smallest start offset value at each end offset position
- The leftmost maches are marked using solid lines

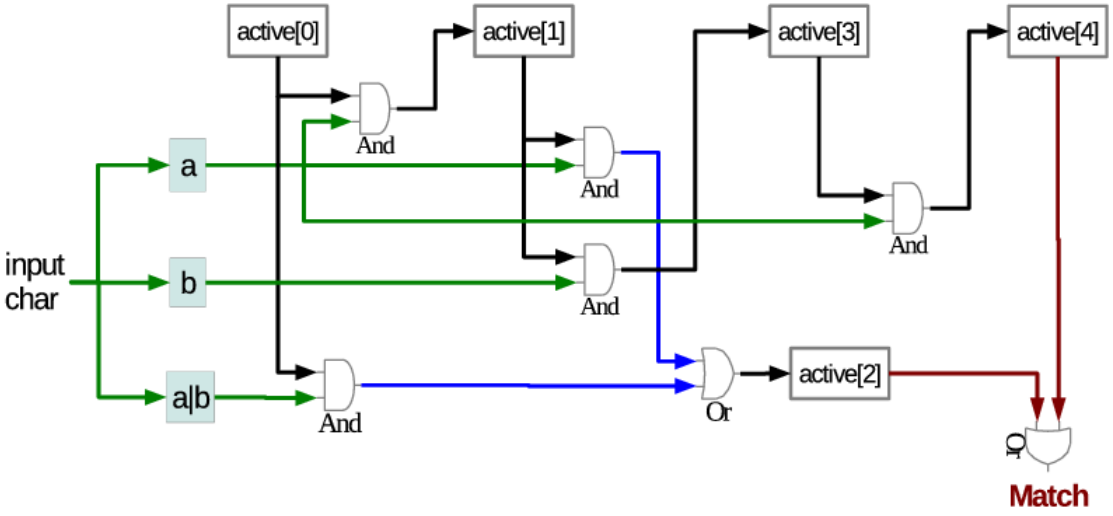


Contributions of this work

1. Extending Sidhu and Prasanna's NFA architecture to support start offset reporting
2. A graph coloring based register clustering method to minimize the register usage
3. An efficient leftmost match computation method without using offset comparisons



NFA

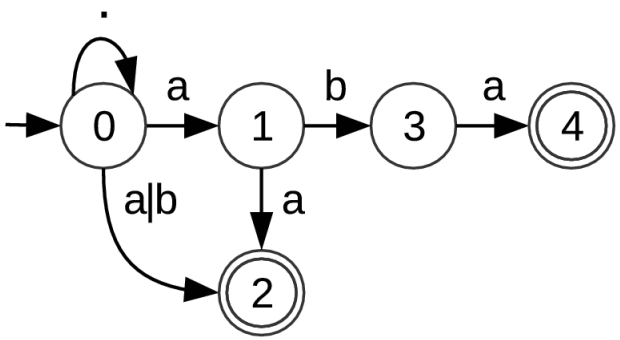


Sidhu and Prasanna's NFA Architecture

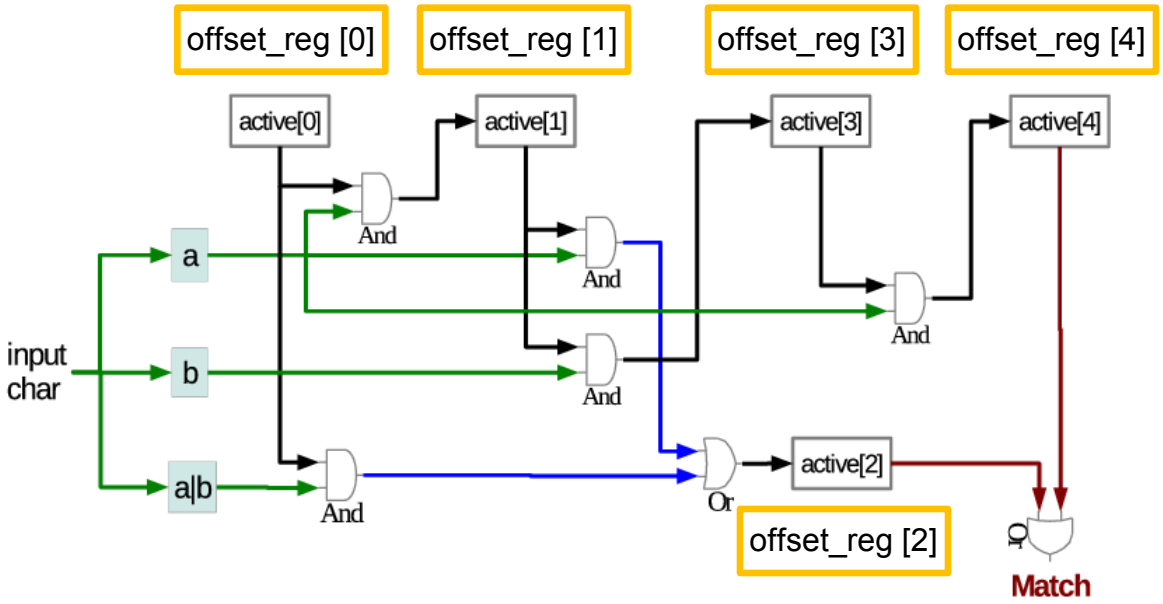
Kubilay Atasu: Resource-efficient regular expression matching architecture for text analytics. ASAP 2014.

Extending Sidhu & Prasanna's architecture

- Add a start offset register to each NFA state
- offset_reg[0] = value of current offset position
- DRAWBACK: redundant start offset registers

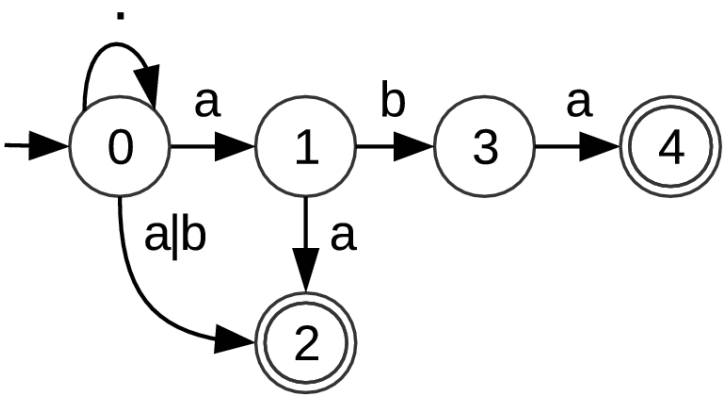


NFA

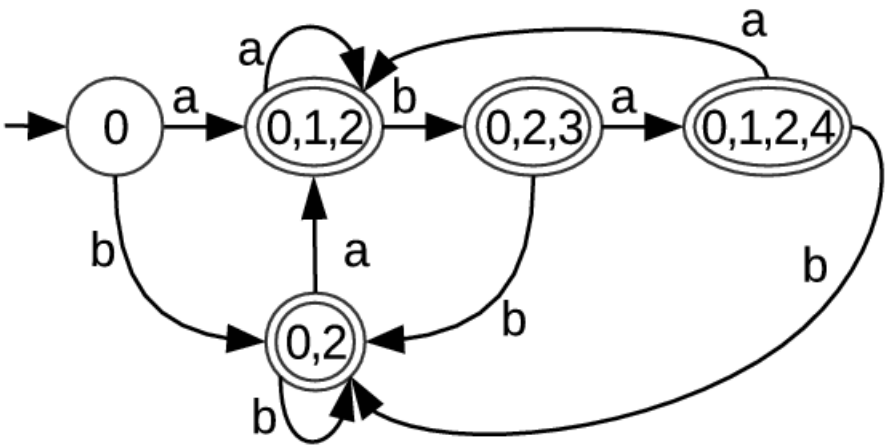


Baseline Architecture

Clustering offset registers

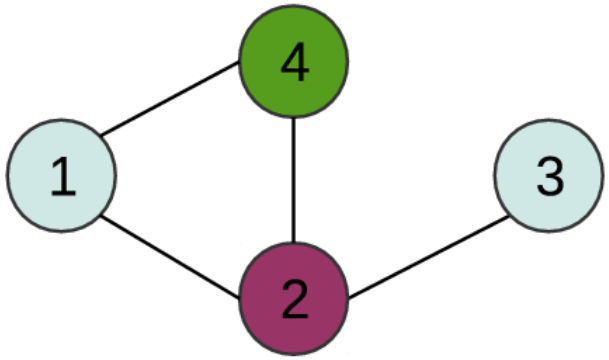


NFA



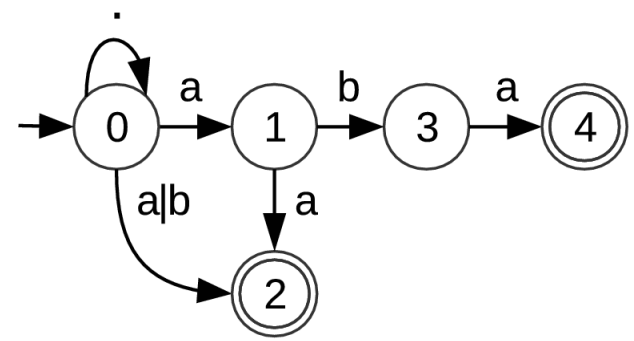
DFA

- Build a conflict graph and apply graph coloring
- States with the same color can share registers

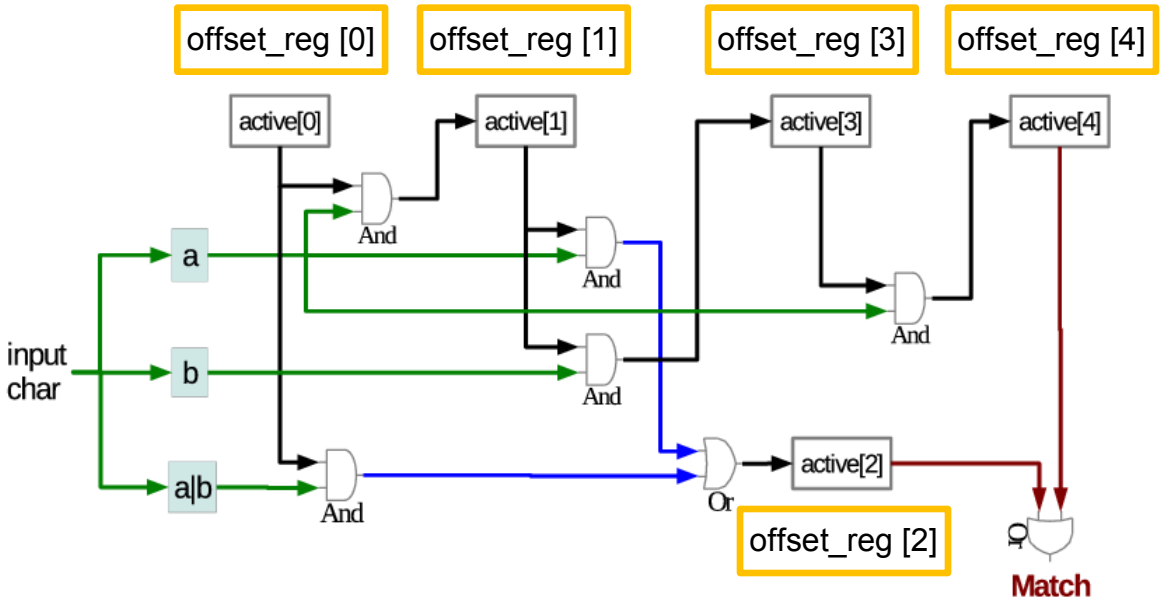


Leftmost match computation

- Assume that state 0 and state 1 are active and the current input is “a”
- We have to compute offset_reg[2] = MIN(offset_reg[0], offset_reg[1])



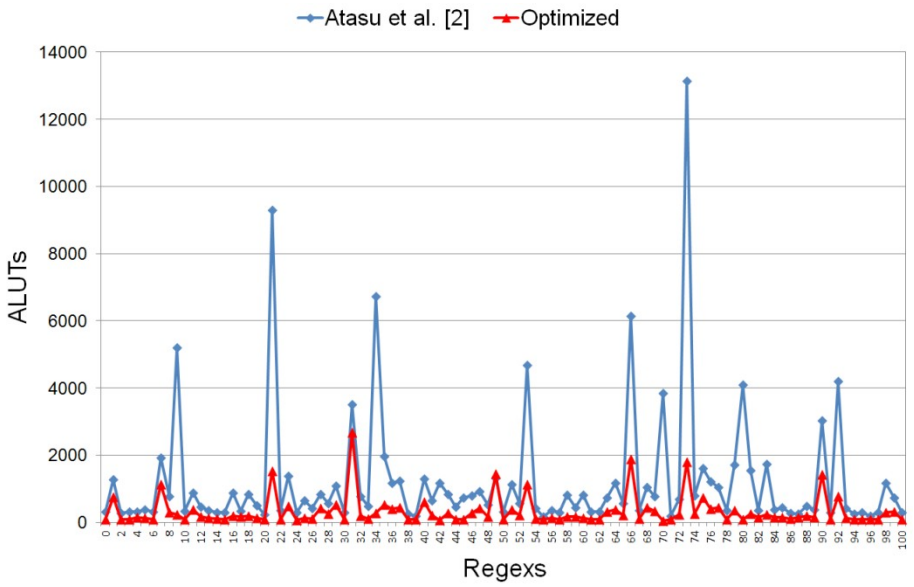
NFA

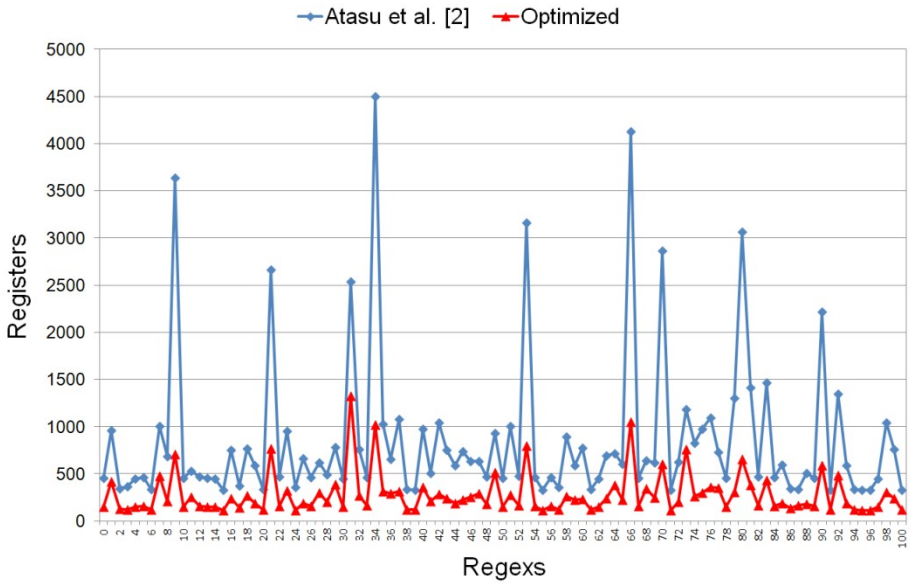


Baseline Architecture

Experiments (L7 filter regexs)

- Altera Stratix IV GX530KH40C2, Altera Quartus II V11 tools
- 32-bit start offset registers, 250 MHz target clock frequency
- NFA representation: Follow Automata with character classes
- Scalability: 1000 regexs with start offset reporting on FPGAs





Kubilay Atasu: Resource-efficient regular expression matching architecture for text analytics. ASAP 2014.

Outline

- Introduction
- Text analytics use cases
- SystemT text analytics software
- Hardware-accelerated SystemT
- HW-accelerated regex matching
- **Conclusions**

Conclusions

- A prototype system that accelerates execution of text analytics queries by
  - utilizing POWER processors and an Altera Stratix FPGAs
  - defining a flexible HW/SW interface with multi-threading support
  - automating generation of query-specific hardware accelerators
- Up to 79x higher document processing throughput vs multi-threaded SW
  - up to 85x better system energy efficiency vs. POWER 7 processor
- Scalable regular expression accelerator that supports advanced features
- Live demonstration of Crystal+ news search acceleration on POWER8
- Ongoing work: programmable overlay architectures to avoid re-synthesis
  - enables support for interactive and complex user queries

QUESTIONS?

Related Publications

- Architecture & Hardware Compiler:
  - R. Polig, K. Atasu, C. Hagleitner, L. Chiticariu, F. R. Reiss, H. Zhu, H. P. Hofstee: Hardware-accelerated text analytics. *Hot Chips 2014.*
  - R. Polig, K. Atasu, C. Hagleitner, L. Chiticariu, F. R. Reiss, H. Zhu, H. P. Hofstee: Giving text analytics a boost. *IEEE MICRO Special Issue on Big Data, 2014.*
  - *R. Polig, K. Atasu, H. Giefers, L.Chiticariu: Compiling Text Analytics Queries to FPGAs. FPL 2014.*
- Regular Expression Matching:
  - Kubilay Atasu: Resource-efficient regular expression matching architecture for text analytics. *ASAP 2014.*
  - Kubilay Atasu, Raphael Polig, Christoph Hagleitner, Frederick R. Reiss: Hardware-accelerated regular expression matching for high-throughput text analytics. *FPL 2013.*
  - *Kubilay Atasu, Raphael Polig, Jonathan Rohrer, Christoph Hagleitner: Exploring the design space of programmable regular expression matching accelerators, Journal of Systems Architecture, 2013.*
- Dictionary Matching:
  - Raphael Polig, Kubilay Atasu, Christoph Hagleitner: Token-based dictionary pattern matching for text analytics. *FPL 2013.*
  - Kanak Agarwal, Raphael Polig: A high-speed and large-scale dictionary matching engine for Information Extraction systems. *ASAP 2013.*