Parsing (in a Broad Sense)

Software Evolution — L3P1

Dr. Vadim Zaytsev aka @grammarware, February 2021

UNIVERSITY OF TWENTE.
Formal Methods & Tools

Parsing in a Narrow Sense: ↑

- Reduce the input back to the start symbol
- Follow the input
- Recognise terminals
- Replace terminals by nonterminals
- Replace terminals and nonterminals by left-hand side of rule
- LR, LR(0), LR(1), LR(k), LALR, SLR, GLR, SGLR, CYK, …

Parser Generators: ↑

- LALR(1)
  - Beaver
  - YACC/byacc/bison
  - Eli
  - Irony
  - SableCC
  - yecc
- GLR
  - bison
  - DMS
  - GDK
  - Tom
- SGLR
  - ASF+SDF Meta-Env
  - Spoofax Stratego/XT

Parsing in a Narrow Sense: ↓

- Imitate the production process by rederivation
- Follow the grammar
- Each nonterminal is a goal
- Replace each goal by subgoals (= elements of its rule)
- Parse tree is built from top to bottom

- LL, LL(1), LL(k), LL(*), GLL, DCG,
  recursive descent, Packrat, Earley

Parser Generators: ↓

- LL(k)
  - JavaCC
- LL(*)
  - ANTLR3
- ALL(*)
  - ANTLR4
- Earley
  - Marpa
  - ModelCC
- GLL
  - Rascal – SGTDBF
  - gll-combinators (Scala)
- Packrat
  - Rats!
  - OMeta
  - PetitParser
- Others
  - TXL

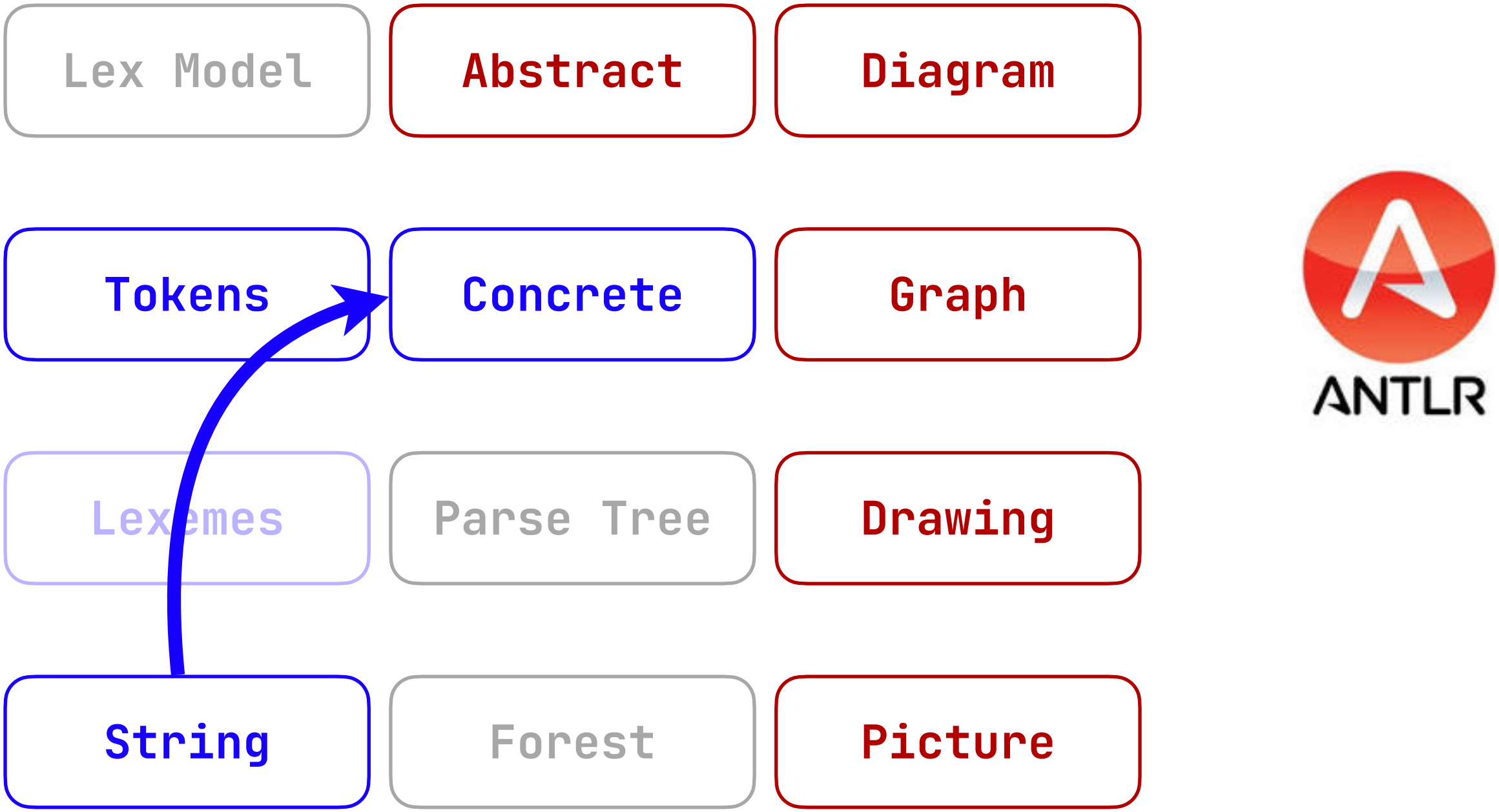
Parsing in a Broad Sense

- Parsing in recognising structure
  - text → tree
  - text → graph
  - tokens → hierarchy
  - forest disambiguation
  - image recognition
- Unparsing is representing structure
  - model → picture
  - tree → text
  - (re)formatting
  - serialisation

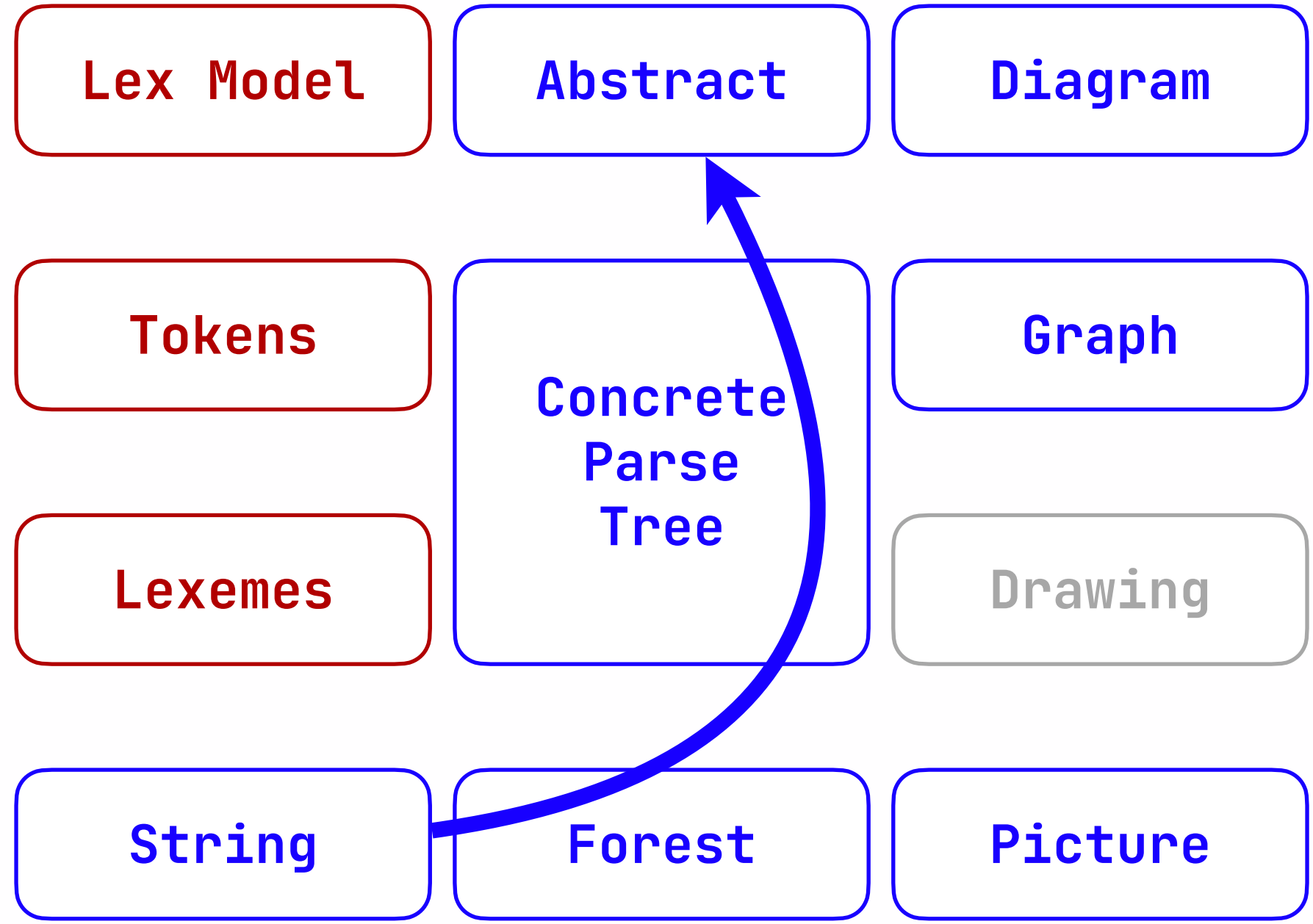
V. Zaytsev, A. H. Bagge, *Parsing in a Broad Sense*, MoDELS 2014.

Lex Model	Abstract	Diagram
Tokens	Concrete	Graph
Lexemes	Parse Tree	Drawing
String	Forest	Picture

V. Zaytsev, A. H. Bagge, *Parsing in a Broad Sense*, MoDELS 2014.



V. Zaytsev, A. H. Bagge, *Parsing in a Broad Sense*, MoDELS 2014.



V. Zaytsev, A. H. Bagge, *Parsing in a Broad Sense*, MoDELS 2014.

Lex Model	Abstract	Diagram
Tokens	Concrete	Graph
Lexemes	Parse Tree	Drawing
String	Forest	Picture

V. Zaytsev, A. H. Bagge, *Parsing in a Broad Sense*, MoDELS 2014.

Conclusion

- Parsing in a narrow sense
  - Top-down: predict-match; backtracking; memoisation
  - Bottom-up: shift-reduce; linear; hard to debug
- Parsing in a broad sense
  - helps to compare technologies
- Q&A Sessions @ Canvas
  - ⇒ v.zaytsev@utwente.nl
  - ⇒ https://discord.gg/n7VQAPNBPD

Fenton, Bieman, *Software Metrics: A Rigorous and Practical Approach*, 2015.