Modern Construction of Industrial Strength Compilers

1

Computer Science Summer in Russia (CSSR)

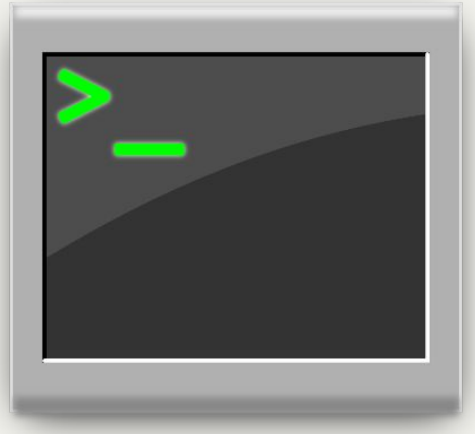
Dr. Vadim Zaytsev aka @grammarware

raincode

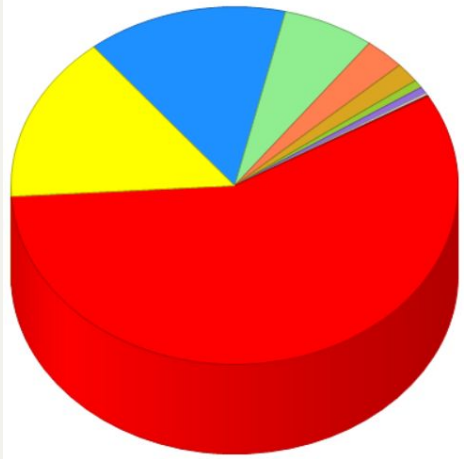
LABS

compiler experts

Introduction



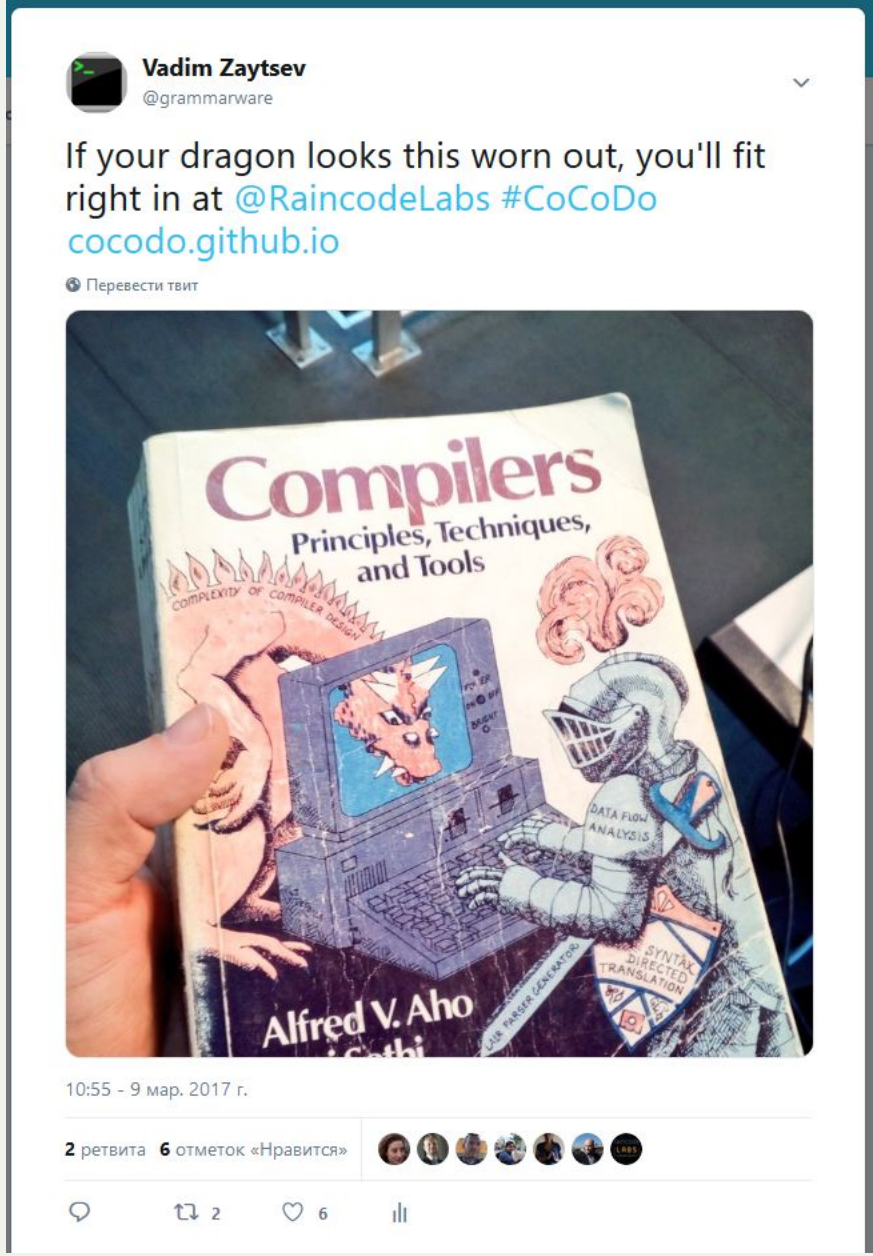
- I am @grammarware
- MSc (ru; nl), PhD (nl/de)
- Ex-researcher (UKL/de, CWI/nl)
- Ex-lecturer (UvA/nl)
- Now @ Raincode [Labs]/be
- Chief Science Officer
- Writes compilers for a living
  - Sometimes publishes and lectures



raincode LABS
compiler experts

Raincode Labs

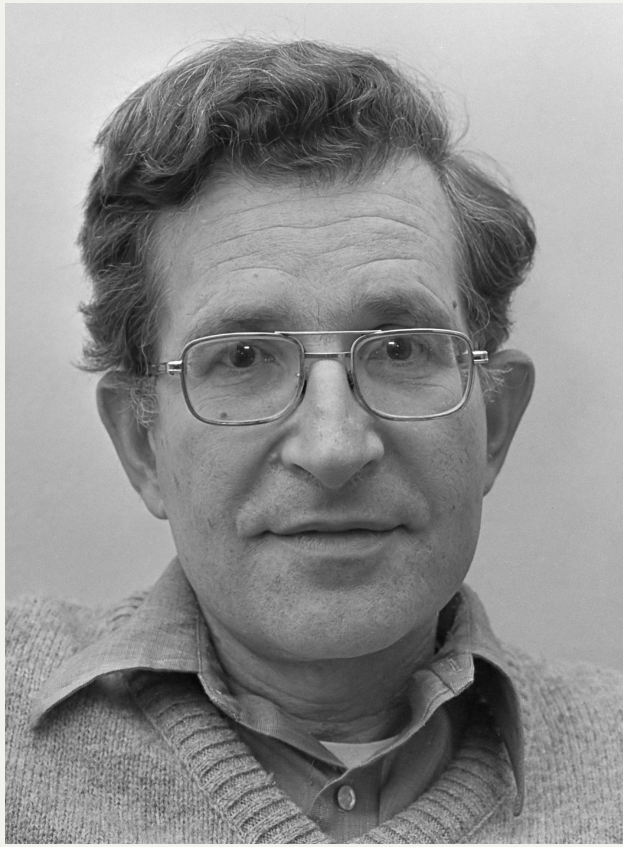
- We are compiler experts
- Typical context
  - COBOL + PL/I + 4GL(s) + HLASM
- Typical customer
  - banking, insurance, logistics
- Typical solutions
  - bankruptcy
  - new platform
  - new compiler
  - new language



Let's revisit FLT/CC/SLE/SE/PLT

Chomsky hierarchy: languages & automata

- Finite languages
  - countries, languages, postcodes, timestamps
  - constant time algorithms: fetch
- Regular languages
  - PCRE, strings, numbers
  - linear algorithms: search
  - catastrophic backtracking; efficient minimisation
- Context-free languages
  - GPLs, DSLs, 4GLs, ...
  - pushdown automata
  - algorithms: parsing and unparsing nested things
- Context-sensitive languages
  - framework hacks
- Recursively enumerable languages



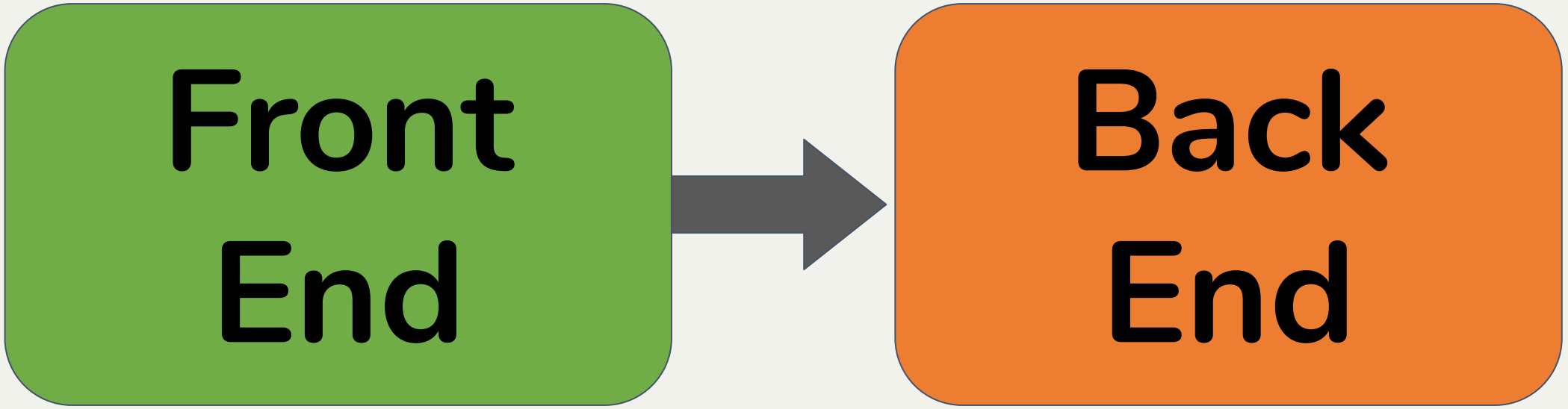
404 Not found in Chomsky

- abductive logic grammar
- abstract categorial grammar
- adaptive star grammar
- adaptor grammar
- affix grammar
- ambiguous grammar
- analytic grammar
- array grammar
- assumption grammar
- attribute grammar (AG)
- augmented transition network (ATN) grammar
- bilexical [context-free] grammar
- binary feature grammar
- Boolean grammar
- bridge grammar
- cascaded augmented transition network grammar
- categorial grammar
- character level grammars
- Chomsky grammar
- classical categorial grammar
- combinatory grammar
- compilable grammar
- computational grammar
- conditional adaptive star grammar
- conjunctive grammar
- constrained W-grammars
- constraint handling grammar (CHR)
- construction grammar
- context-free grammar (CFG)
- context-free graph grammar (CFGG)
- context-sensitive grammar
- continuous grammar
- cooperating array grammar system
- cooperating puzzle grammar system
- cooperating string grammar system
- cover grammar
- cyclic grammar
- data-dependent grammars
- decorated attribute grammar
- definite clause grammars (DCG)
- description tree grammars
- description tree substitution grammars
- deterministic graph grammars
- Dik functional grammar
- distributional lattice grammars
- e-grammars
- environment grammars
- extended affix grammars
- extended attribute grammars
- extended context free grammar
- finite-state grammar
- functional discourse grammar
- general combinatory grammar
- generalised phrase structure grammar
- grammar in a broad sense
- grammar in a reduced form
- grammar with a two-sided context
- grid grammar
- head automaton grammar
- head grammar
- head-driven phrase structure grammar
- history based grammar
- hyperedge replacement graph grammar
- indexed grammar
- infinitely ambiguous grammar
- information extraction grammar
- inversion transduction grammar
- island grammar
- k-valued grammar
- Kay functional grammar
- keyword grammar
- least cardinality grammar
- least-valued grammar
- lexical functional grammar
- lexical interpretive grammar
- lexicalised back-off grammar
- lexicalised context-free grammar
- lexicalised grammar
- lexicalised probabilistic grammar
- lexicalised tree-adjoining grammar
- linear graph grammar (LGG)
- link grammar
- linked grammar
- logic grammar
- Markovian context-free grammar
- matrix grammar
- metamorphosis grammar
- minimal grammar
- minimalist grammar
- monotonic grammar
- Montague universal grammar
- motion picture grammar
- multi-amalgamated triple graph grammar
- multi-box grammar
- multicomponent grammar
- multiple context-free grammar
- narrative grammar
- node replacement graph grammar
- non-e-grammars
- notional grammar
- object grammar
- optimal grammar
- pair grammar
- pattern grammar
- Peirce grammar
- permissive grammar
- phrase-structure grammar
- picture grammar
- picture processing grammar
- Pitman-Yor adaptor grammar (PYAG)
- pregroup grammar
- probabilistic constrained W-grammars (PCWG)
- probabilistic context-free grammar (PCFG)
- probabilistic dependency grammar
- probabilistic feature grammar
- probabilistic tree-adjoining grammar (PTAG)
- production grammar
- proper grammar
- puzzle grammar
- range concatenation grammar
- recursive grammars/languages???
- reflective grammar
- regular grammar
- regular right part grammar
- rigid grammar
- S-grammar
- self-embedding grammar
- sentence-generating grammar
- simple chain grammar
- simple literal movement grammar (sLMG)
- simple transduction grammar
- skeleton grammar
- stochastic attribute-value grammar
- stochastic context-free grammar
- stochastic head-driven phrase structure grammar
- stochastic lexicalised context-free grammar
- stochastic lexicalised tree-adjoining grammar
- stochastic tree-substitution grammar (STSG)
- string adjunct grammar
- synchronous grammar
- tear-insert-fold grammar
- tile grammar
- transformational grammar
- transition network grammar
- tree adjoining grammar (TAG)
- tree adjunct grammar (TAG)
- tree description grammar
- triple graph grammar (TGG)
- two level grammar
- two-level meta-controlled substitution grammar (2MSG)
- two-level probabilistic grammars
- type 0 grammar
- type 1 context-sensitive grammar
- type 1 grammar
- type 1 monotonic grammar
- type 2 grammar
- type 3 grammar
- type grammar
- unification grammar
- unrestricted grammar
- van Wijngaarden grammar (WG)
- vW-grammar
- W-grammar
- Łukaszewicz functional grammar
- Łukaszewicz universal grammar

404 Not found in Chomsky

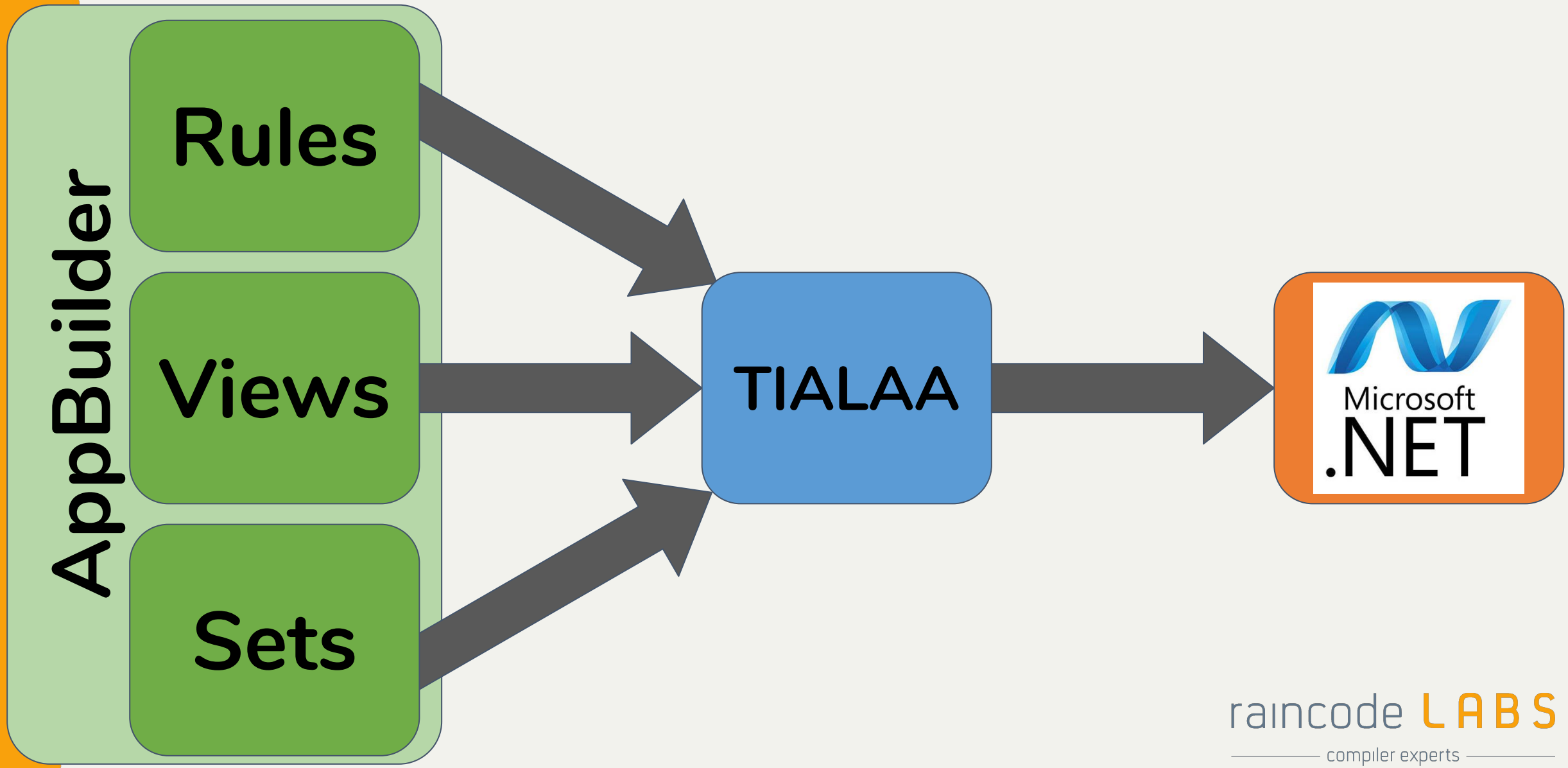
- Extended context-free grammars
- Parsing expression grammars
- Boolean grammars
- Data dependent grammars
- Attribute grammars
- Triple graph grammars

Compiler pipeline

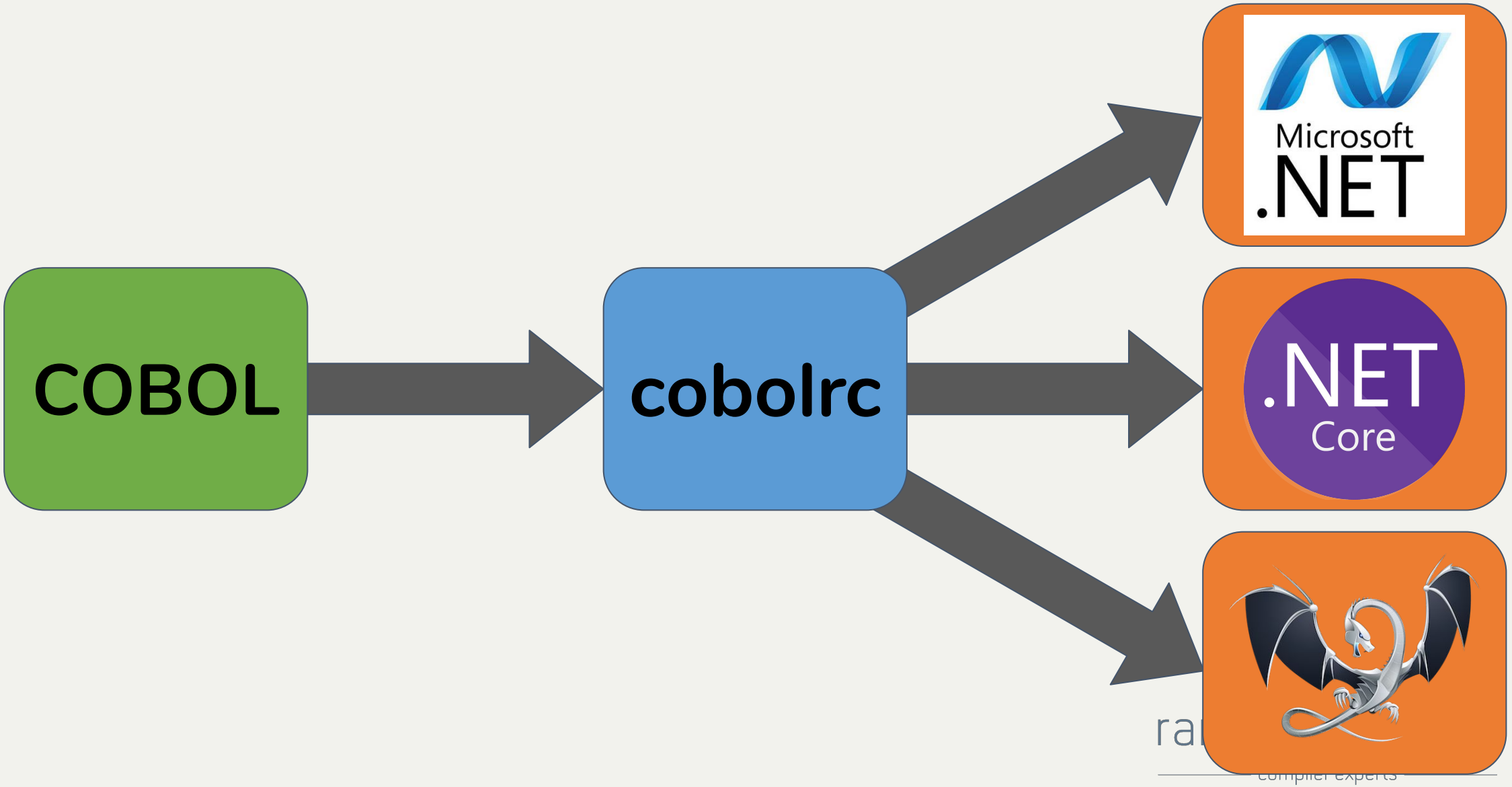


Compiler pipeline

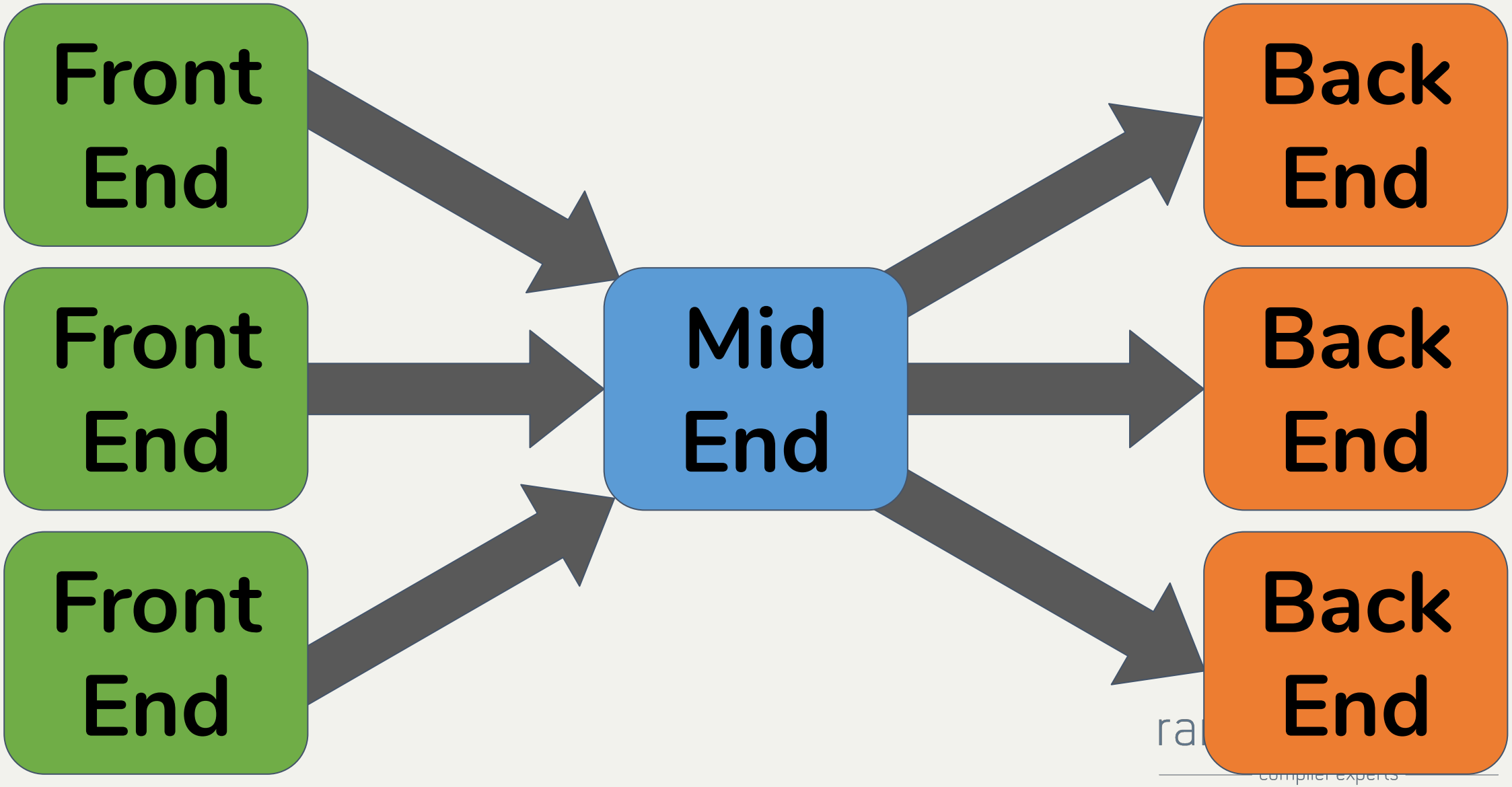
Compiler pipeline [multi-language]



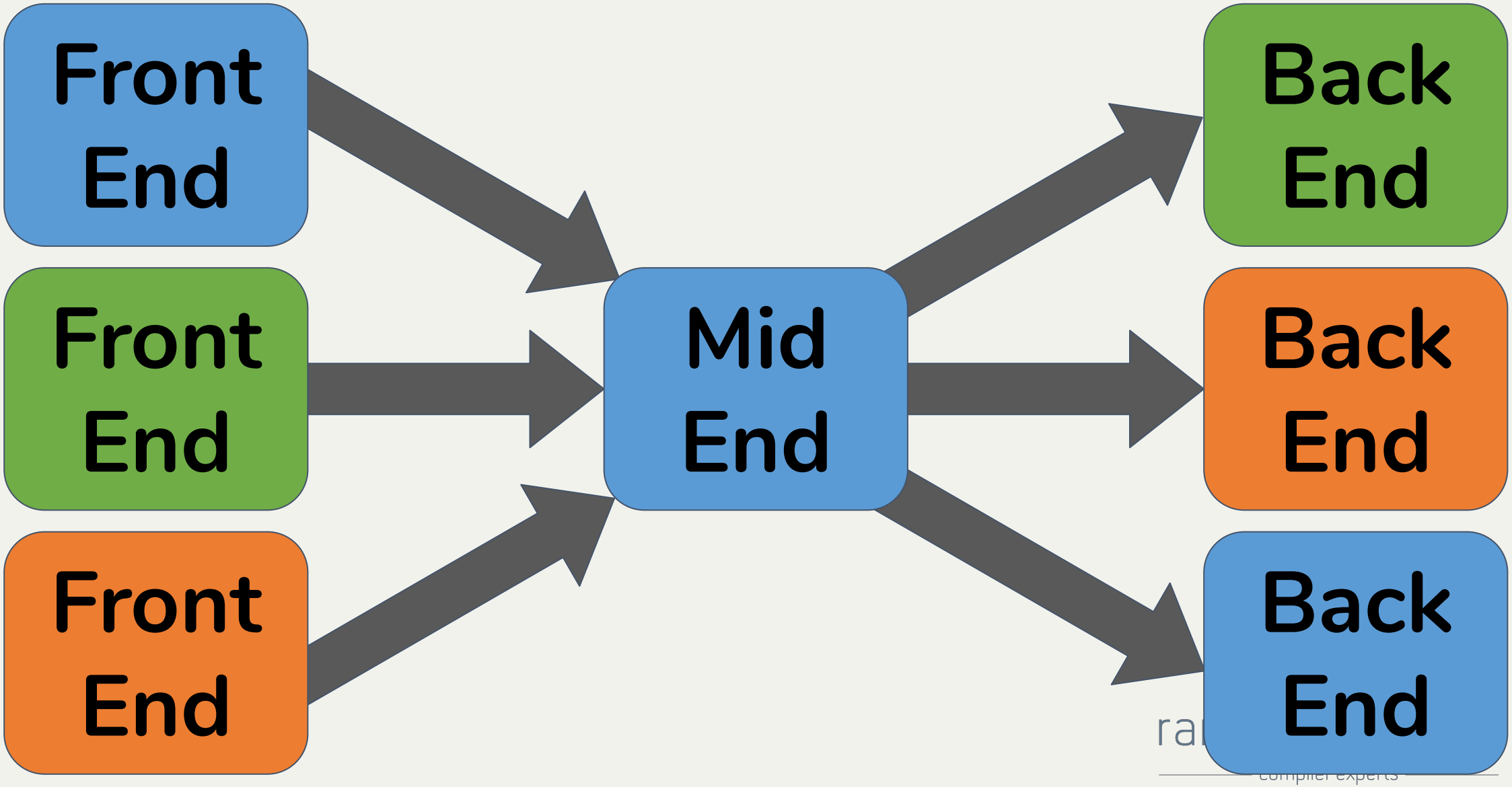
Compiler pipeline [multi-target]



Compiler pipeline [ideal?]



Compiler pipeline [actual!]



Grammarware

- Parser
- Compiler
- Interpreter
- Prettyprinter
- Scanner
- Browser
- Static checker
- Struct.editor
- IDE
- DSL
- API
- Preprocessor
- Postprocessor
- Validator
- Model checker
- Refactorer
- Code slicer
- Decompiler
- Modelling workbench
- Language workbench
- Rev.eng. tool
- Benchmark
- Recommender
- Renovation tool

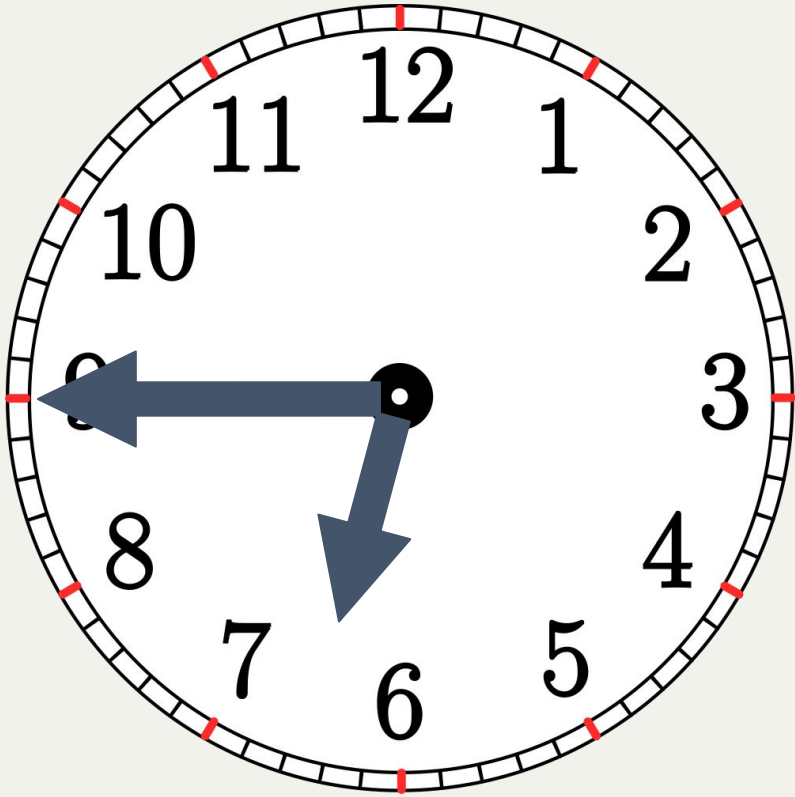
Grammarware is a technological space

Grammarware is a technological space

There are others:

- UMLware — Mainstream software modeling
- Modelware — Model Driven Engineering
- Javaware — Mainstream OO programming
- XMLware — Interoperability in data exchange
- SQLware — Mainstream database management
- Haskellware — Strongly typed functional programming
- Scriptware — Scripting and frameworks in Python or JS
- RDFware — Semantic Web and Linked Data
- JSONware — Modern interoperability
- COBOLware — Legacy programming technology

Short break

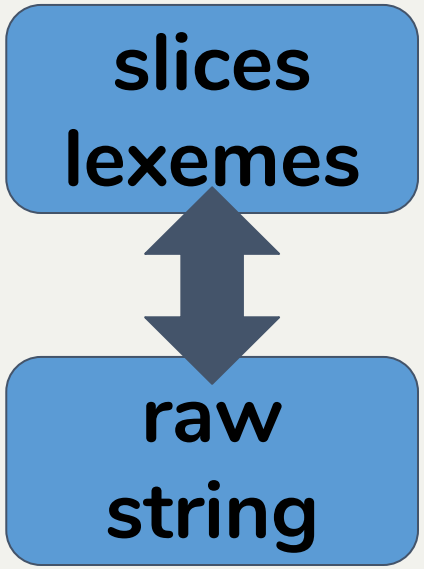


Parsing in a broad sense

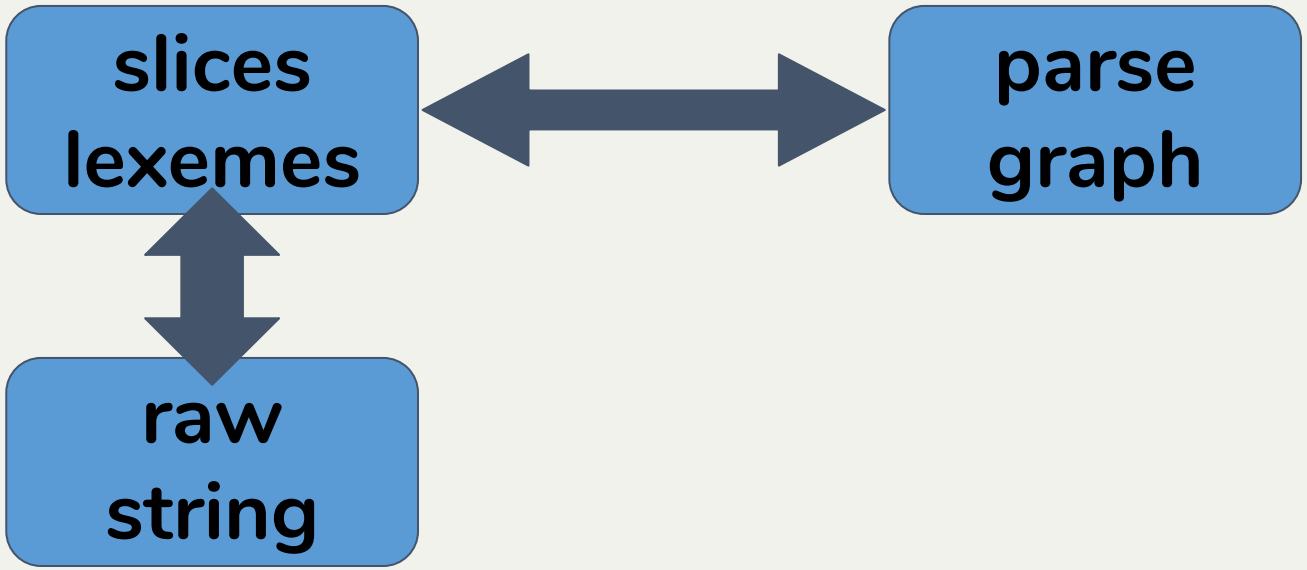
Parsing in a Broad Sense

raw string

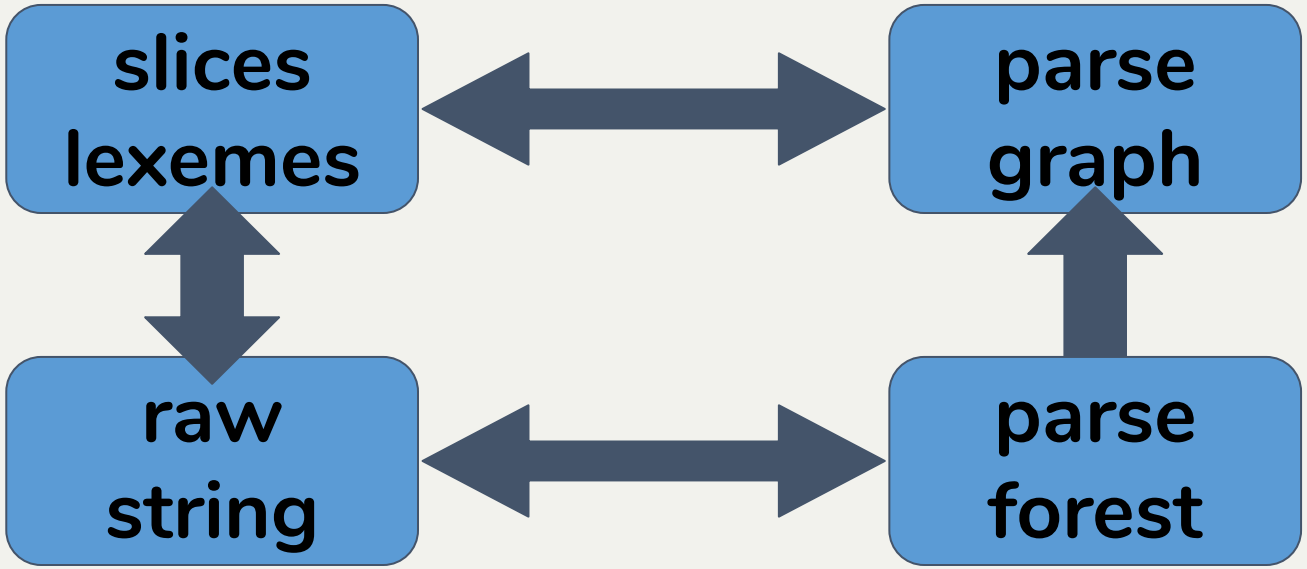
Parsing in a Broad Sense



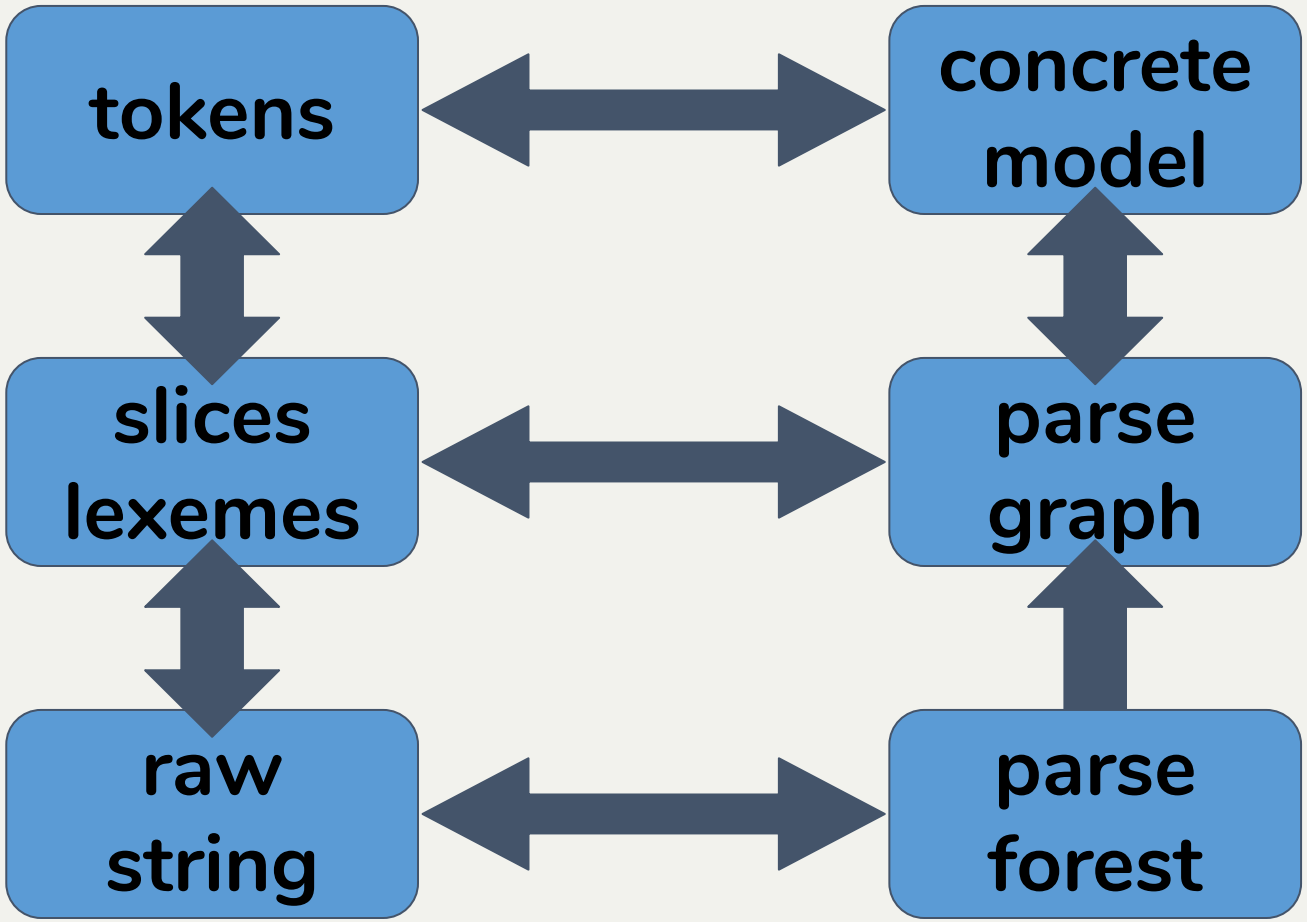
Parsing in a Broad Sense



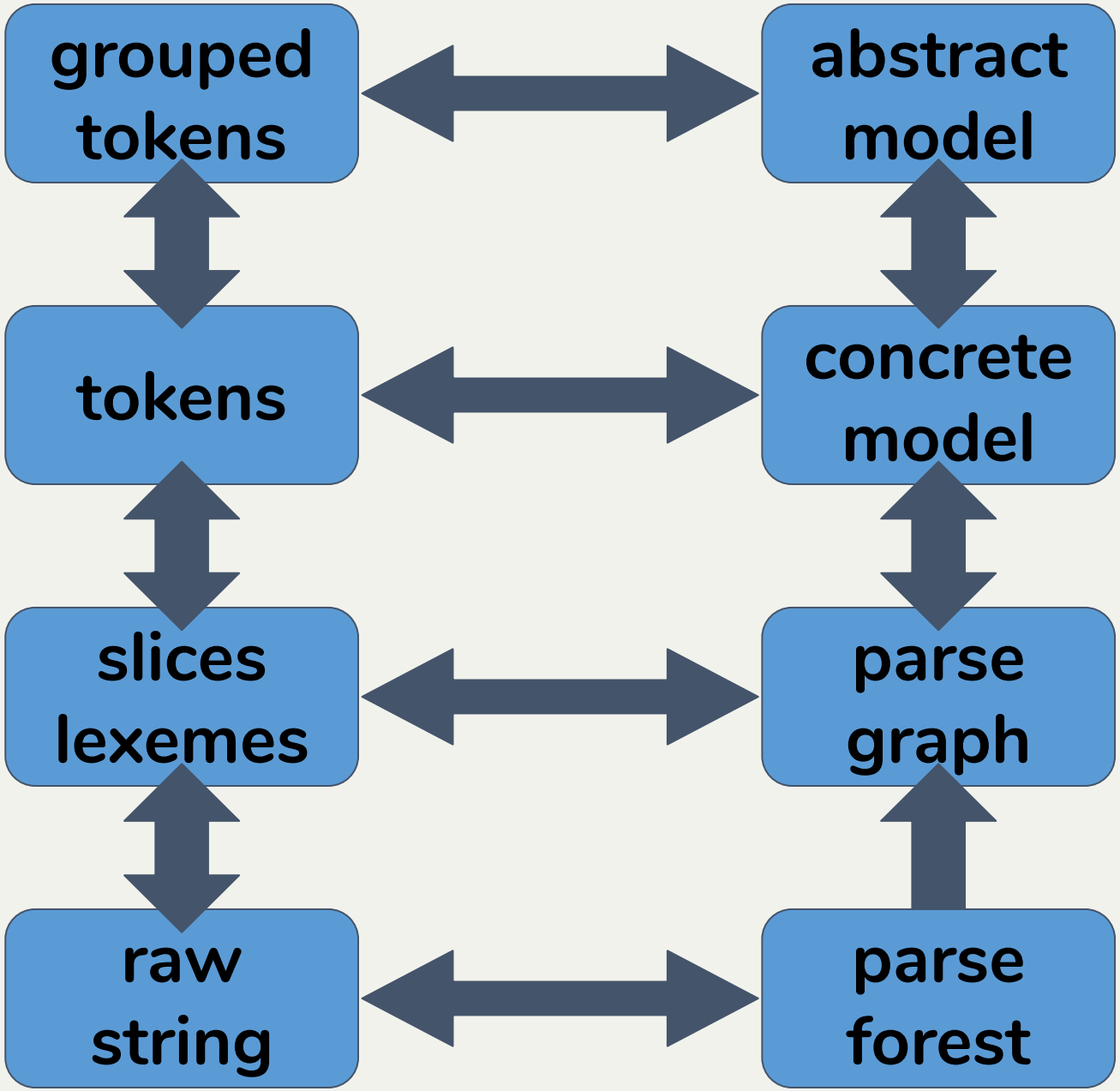
Parsing in a Broad Sense



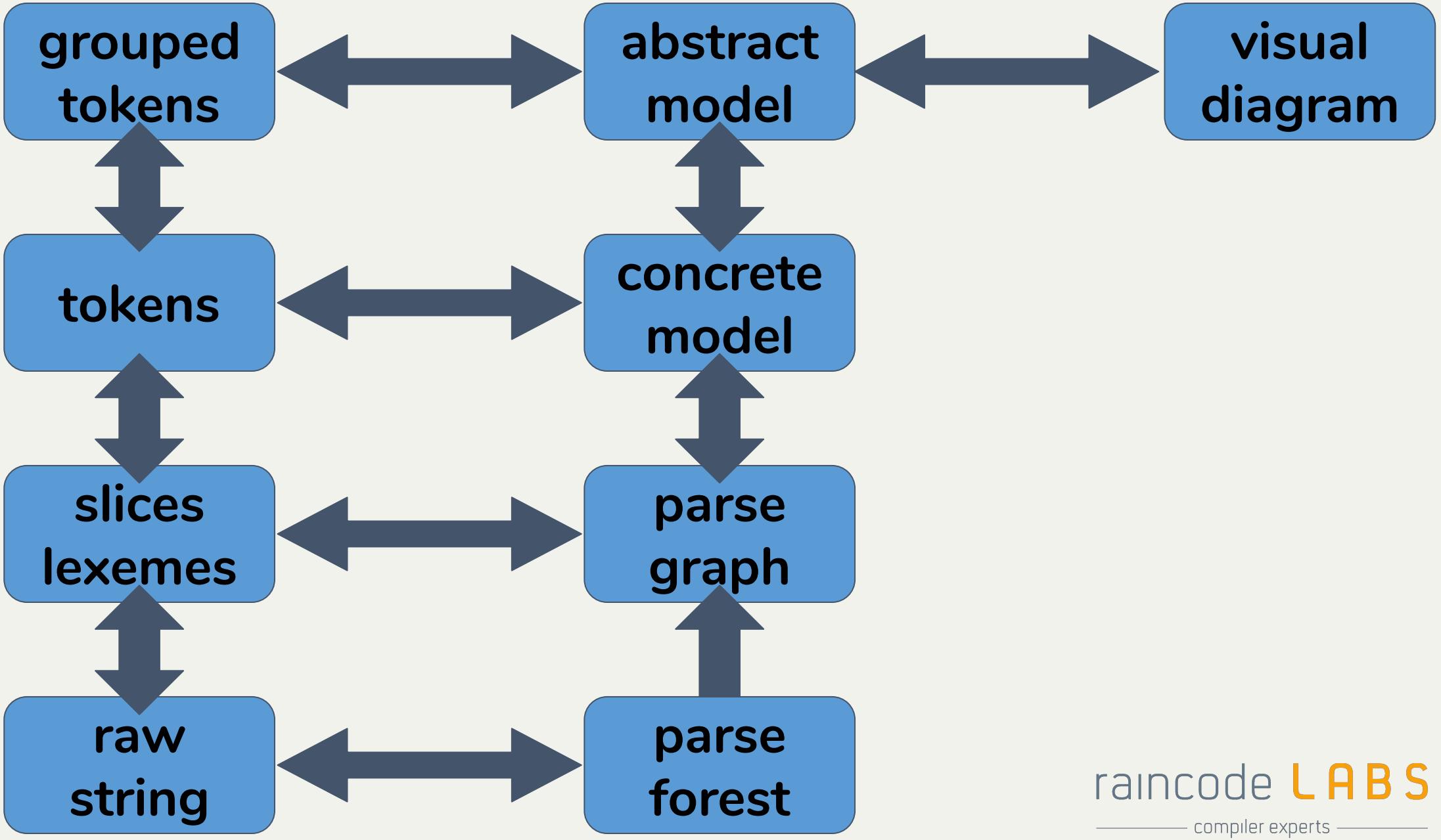
Parsing in a Broad Sense



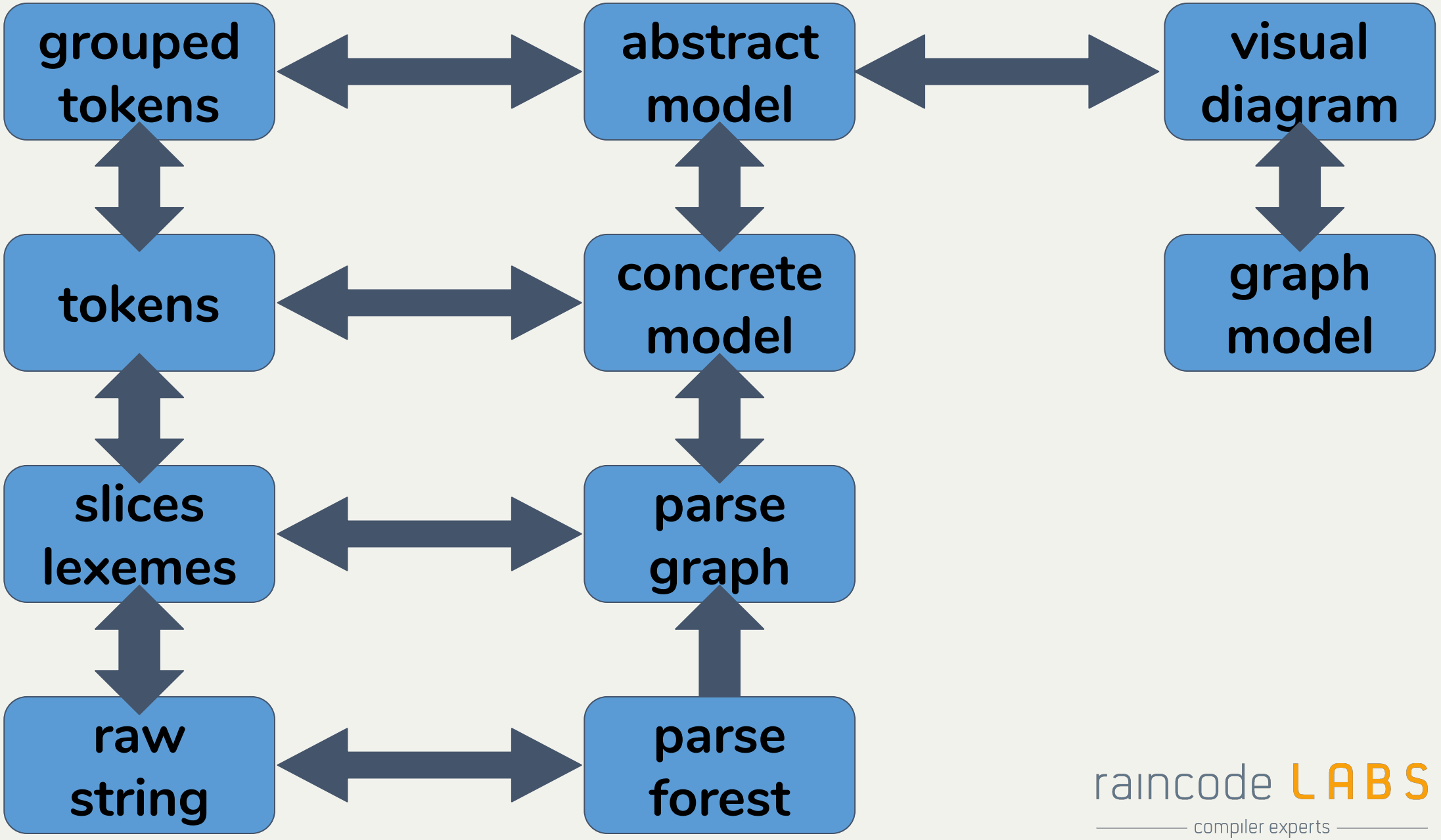
Parsing in a Broad Sense



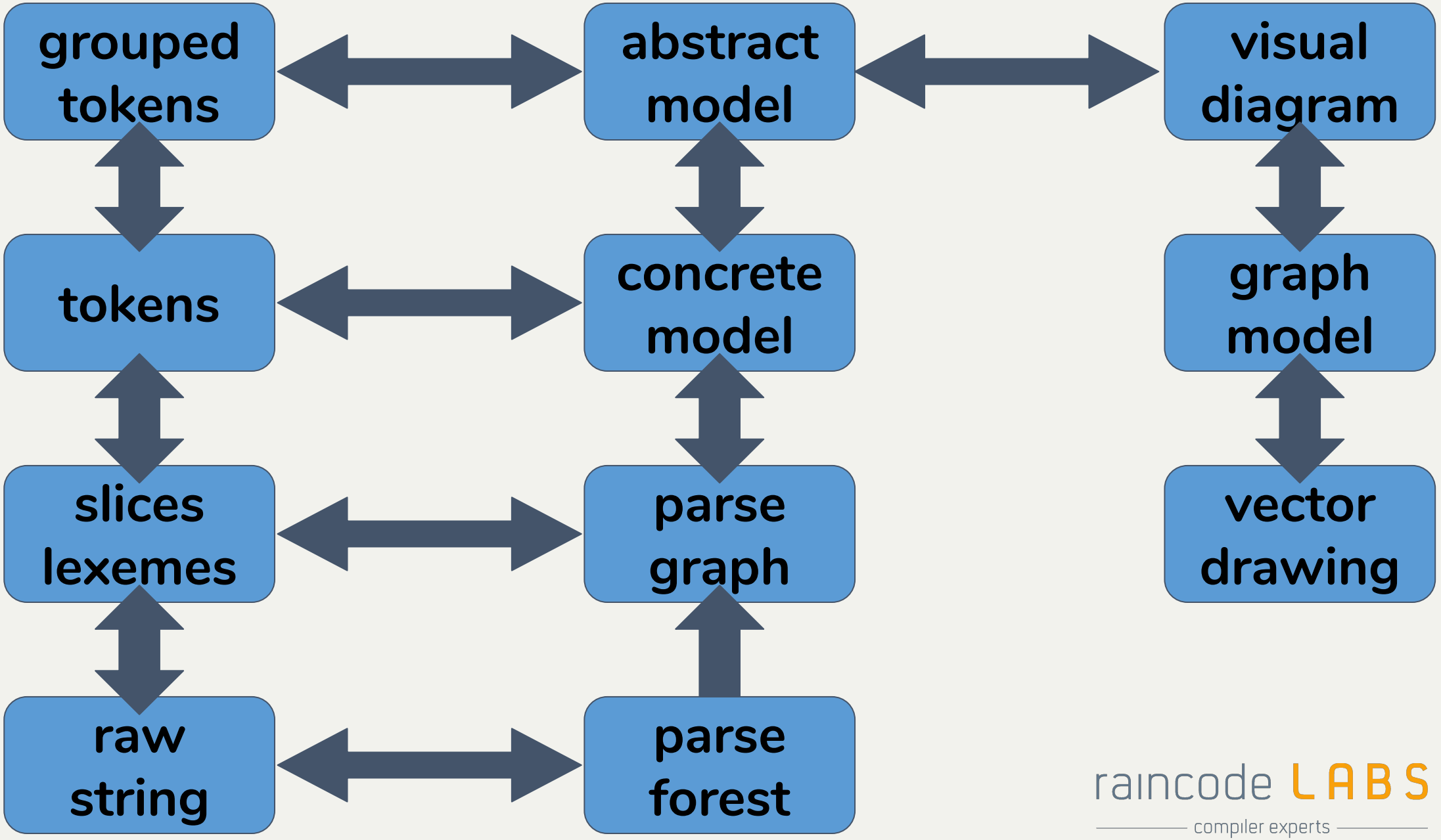
Parsing in a Broad Sense



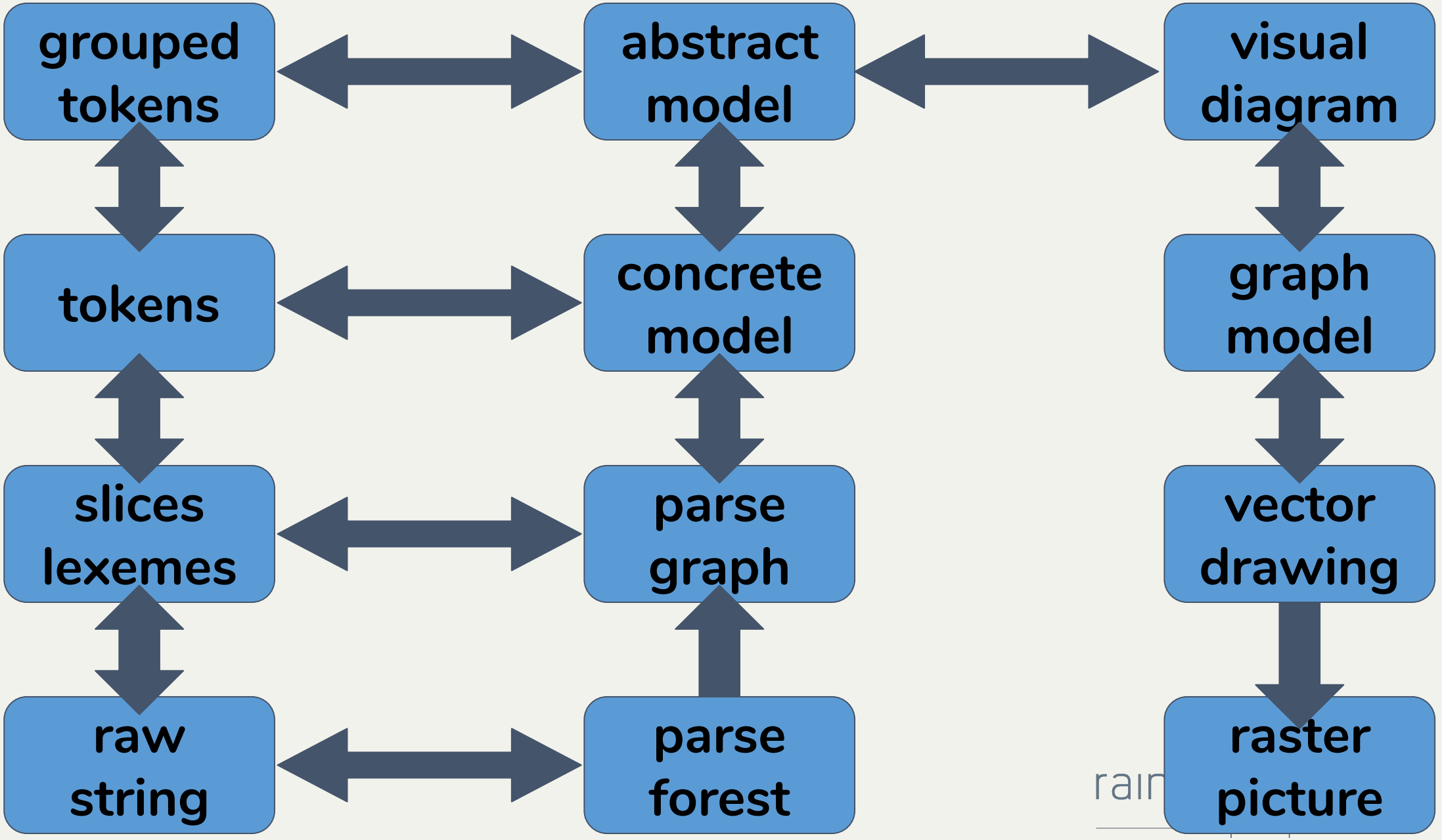
Parsing in a Broad Sense



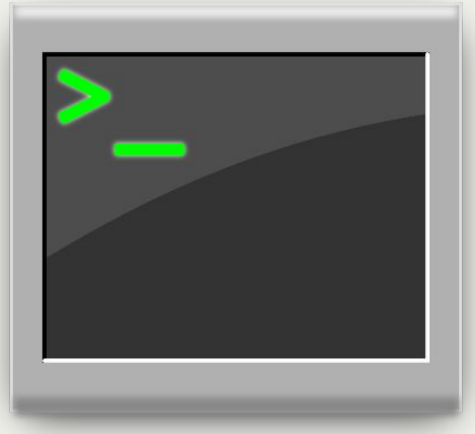
Parsing in a Broad Sense



Parsing in a Broad Sense



Conclusion



- Chomsky hierarchy
- Compiler pipeline and variations
- Grammarware as a technological space
- Parsing in a broad sense
- Follow @grammarware
- Next lecture: parsing in a narrow sense
- Extras: http://grammarware.github.io/cssr/