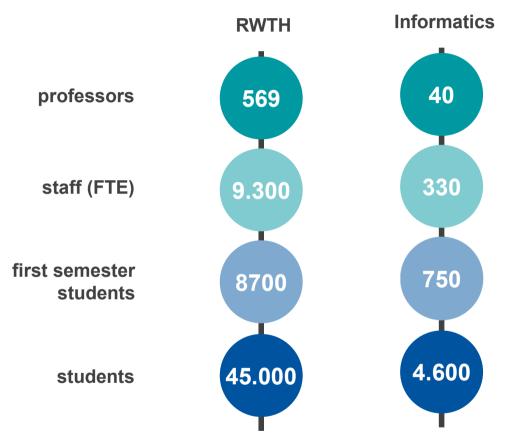


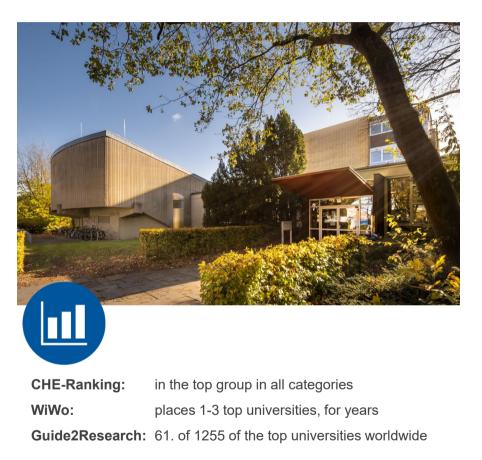
Addressing the "Engineering" in "Software Language Engineering"

Bernhard Rumpe Software Engineering RWTH Aachen http://www.se-rwth.de/



RWTH Informatics: Facts and Figures





State: 06/2025

2

Software Engineering | RWTH Aachen



Software Language

A software language is a human readable and computer processable language addressing a particular problem.

A modeling language is a software language used for modeling software or systems.

- Software languages facilitate
 - automated tool-based analysis, synthesis and code generation on models
 - re-use of models

- engineering productivity
- adaptivity, for example with Models@Runtime
- Any form of automation needs a precisely and explicitly defined language.

- Examples:
- UML:
- Java:
- XML:
- a general-purpose modeling language
- a general-purpose programming language
- a format for structured data











GUI Modeling in MontiGem: For Information Systems, Digital Twins, IOT-Services

€ Finanzen	< PROF Einstellung Mein Benutze	gen → Profil		admin STESTDB 9-09-12T14:19) 12:09:2019 Instanz-Verwaltung	2 String username; 3 Optional <string> encodedPassword;</string>	Data structure
Personal Einstellungen Mein Profil Benutzer/Rollen Daten Import		E-Mail Adresse Kürzel Registrierungsdatum	macoco@se.rwth-aachen.de N.N. 29.11.2018	?	<pre>1 datatable "meinBenutzerInfoTabelle" { 2 columns < uit { 3 row "Benutzername" , <username "e-mail="" "kürzel"="" "registrierungsdatum"="" "tim-kennung"="" (editable)="" ,="" 4="" 5="" 6="" 7="" 8="" 9="" <email="" <initials="" <tim="" adresse"="" date(<registrationdate)="" pre="" row="" }="" }<=""></username></pre>	User interface
Drucken The second sec		Min. 5 Zeichen Neues Passwort Min 5 Zeichen	Passwort	Ändern	<pre>1 context User inv isPasswordValid: 2 password.length() < 5; 3 shortError: "Min. 5 Zeichen"; 4 error: "Das Passwort muss aus mindestens 5 Zeichen 5 bestehen, hat aber nur " + 6 passwort.length() + " Zeichen. ";</pre>	Constraints

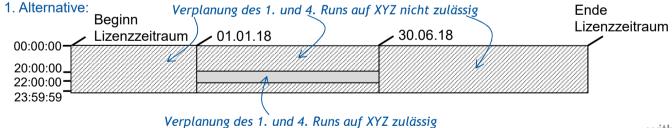


Description Language for Planning TV Broadcasting

- Planning audio-visual offers such as TV program and video-on-demand
- Restrictions in licensing contracts
- Risks: Misinterpretation and resulting planning errors
- DSL for
 - Verification of plans
 - Calculation of allowed planning periods

1 Alternative 1: 2 Der 1. & 4. Run innerhalb von 01.01.2018 bis 30.06.2018 turnus immer von 20:00 bis 3 22:00 auf Nutzer XYZ 4 Alternative 2: 5 Alle Run innerhalb von 01.01.2018 bis 30.06.2018 turnus ohne auf Nutzer XYZ



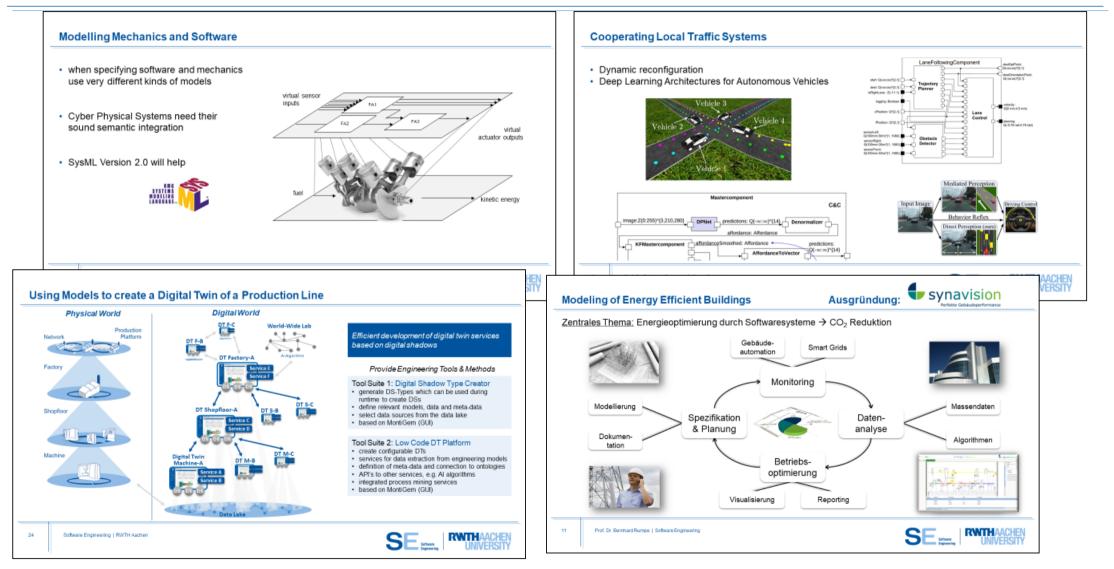


with I. Drave, K. Hölldobler, O. Kautz, J. Michael

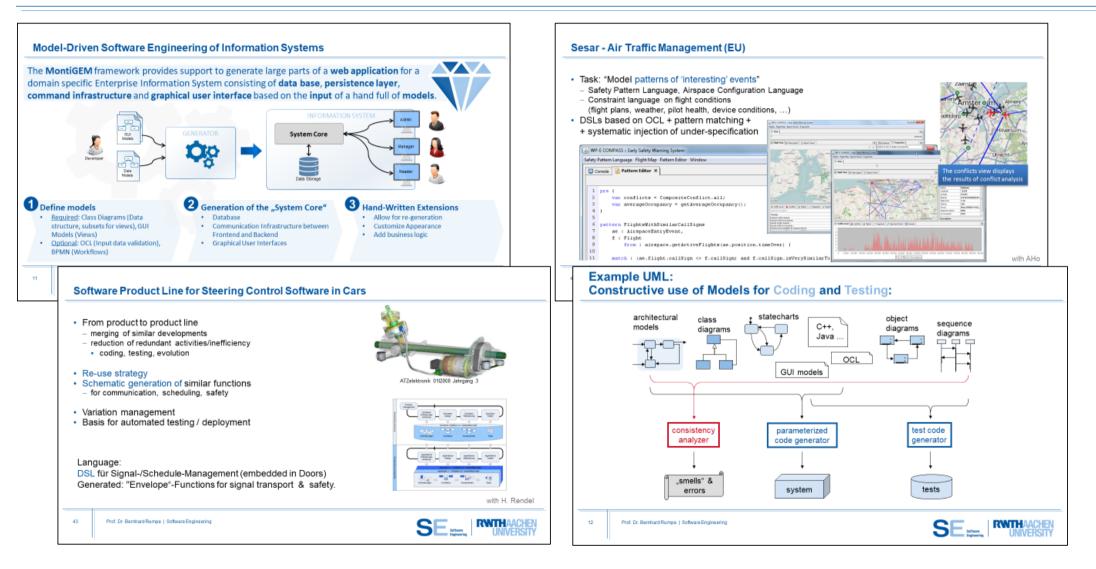


Software Engineering | RWTH Aachen

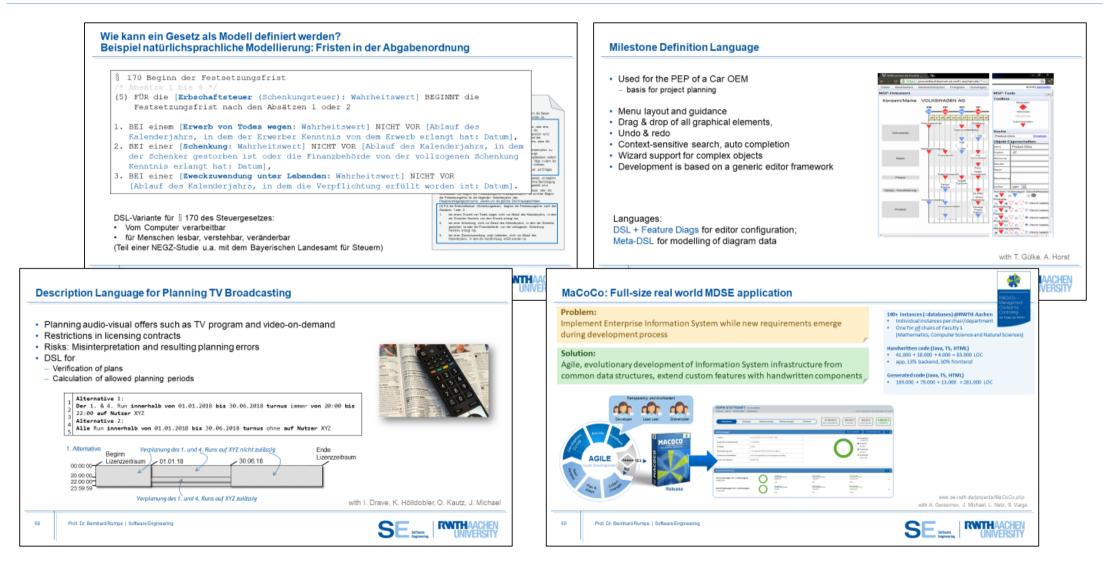
Some Examples For Usage of DSLs



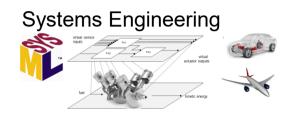
Some Examples For Usage of DSLs



Some Examples For Usage of Our DSLs



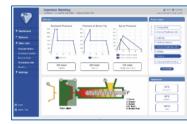
Software Engineering: Our Mission: Improving Software and Systems Development



Contracts, Regulations,

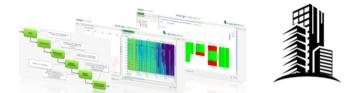


Digital Twin Cockpits



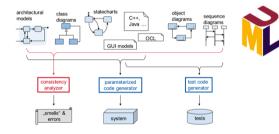
Languages, methods, concepts, tools and infrastructures for

- better and faster agile development,
- resulting in high quality products.



Energy Efficiency e.g. in Buildings

Software, Tests, Deployment, Architecture, Design, Agility, Process management





Information Systems: Management Cockpits







Software Languages, DSLs, LowCode

onti

A basic question:

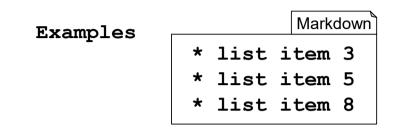
How to **engineer** a DSL?

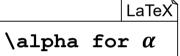
Software Engineering | RWTH Aachen



Steps for Designing a Language

- 1. Understand the goal of the DSL: What to be described? 2 Write down examples of the DSL 3. Identify reusable language components from a language library Build the language from components 4. using extension and adaptation as glue for abstract and concrete syntax, symbols: names, kinds and visibility, context conditions 5. What kinds of tools users need? editor, wizard,
 - guidelines, metrics/smell analyzer,
 - code generator, interpreter,
 - diffing of models, etc.





	Java
int a = 0, b = 1;	
while (a <= 100) {	
System.out.println(a + " -> " + (a *	a));
int next = $a + b$; $a = b$; $b = next$;	
}	



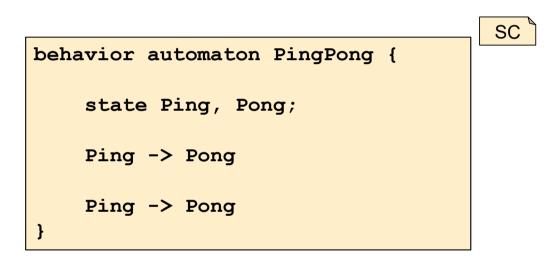
Language Extension - Starter

• Lets start with one language L1

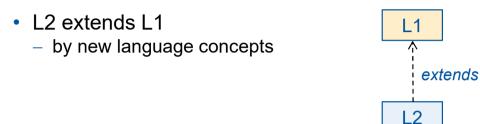
L1

- The automaton has
 - 2 states and
 - 2 transitions
 - describing a ping pong game

• Automaton language L1:

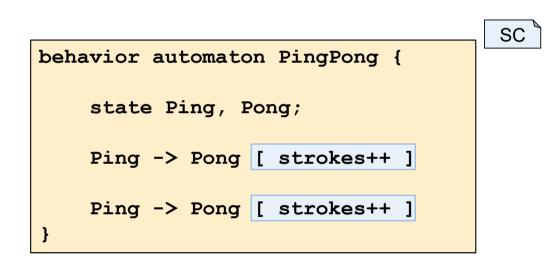


Language Extension



- One model contains language concepts of both languages
- Either L1 or L2 becomes the master language and the other the multiply embedded sub-language
- Semantics, code generation is often defined together, but ideally reuse L1-semantics, generators, etc. should be possible

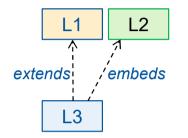
Automaton language L1 is extended by actions in L2:
 Actions are embedded at multiple places:





Language Embedding

 A new language L3 embeds model concepts from L2 in the language L2



- Models have parts conforming to sublanguages
- Languages L1 and L2 were independently developed
- Enables reuse and extension of languages
- Allows to define language components

 E.g. expressions, literals, type definitions.

• Automaton language L1 and action language L2 are combined to a language embedding the actions into the automaton:

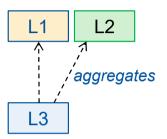
	SC			
behavior automaton PingPong {				
state Ping, Pong;				
37 37				
Ping -> Pong [strokes++]				
Ping -> Pong [strokes++]				
}				

• "Glue" can be added, e.g. the square brackets



Language Aggregation

An aggregated language
 L3 combines L1, L2, and more …



- Models are independent artefacts

 they can be edited, reused, etc. individually
- Models are only semantically composed
 there is no model belonging "only" to L3
- Models syntactically refer to each other
 "Symbols" are imported / exported

• Two models:

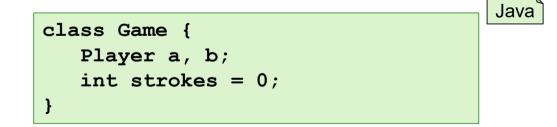
}

 An automaton and a java class sharing symbols (e.g. strokes)



```
state Ping, Pong;
```

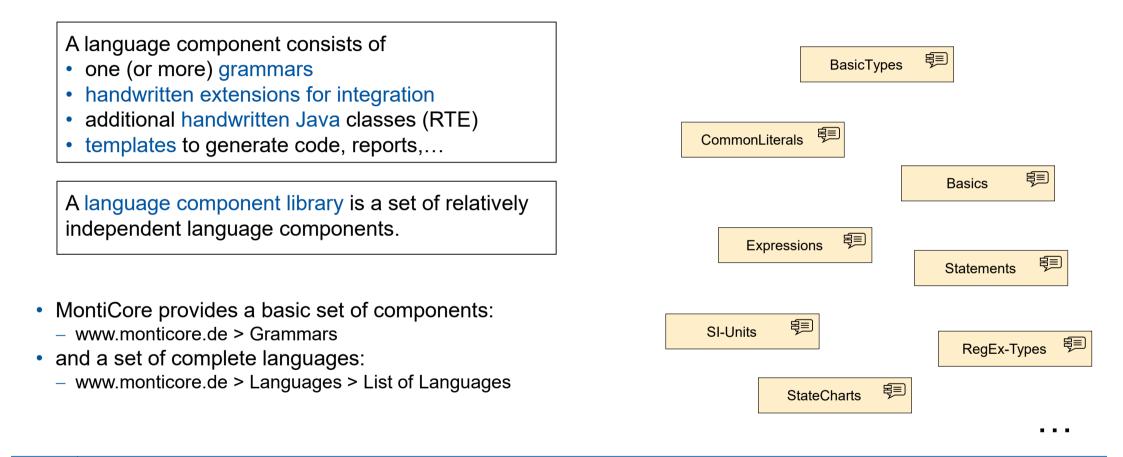
Ping -> Pong [strokes++];





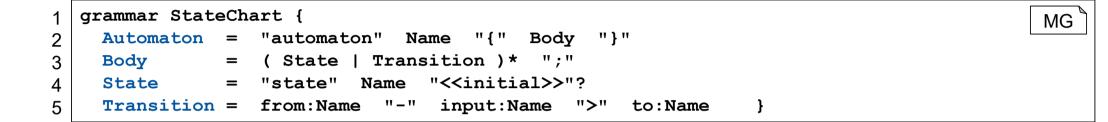
SC

Language Component Library builds on Composition





In MontiCore: Grammars define a DSL



MontiCore Grammar (MC) has

"",

...?

(...),

....

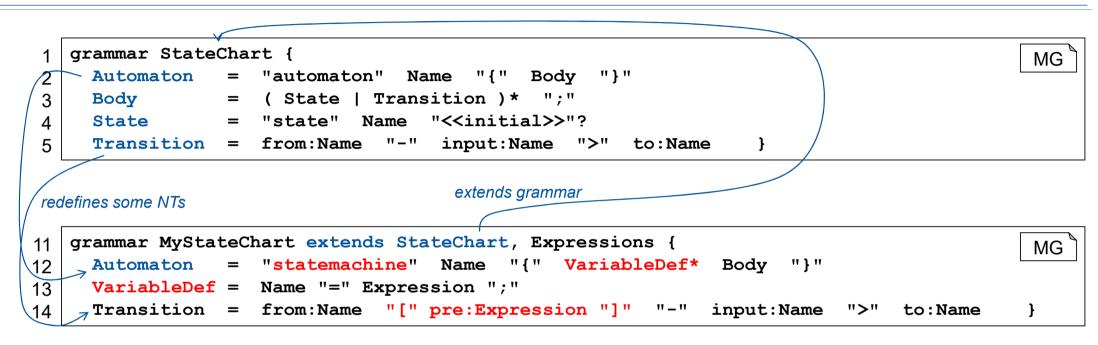
 \dots^* , \dots^+ , (body | delimiter)*

- nonterminals,
- keywords
- grouping
- alternatives
- iteration
- optionals
- token, etc.

- MontiCore generates from the grammar:
 - Abstract Syntax Tree Classes
 - Parser + Tree Builder
 - Visitors
 - Pretty printing
- And furthermore:
 - Symbol Tables
 - Typecheck
 - Context Condition Infrastructure
 - Transformation engine, ...



In MontiCore: Grammars extend other Grammars

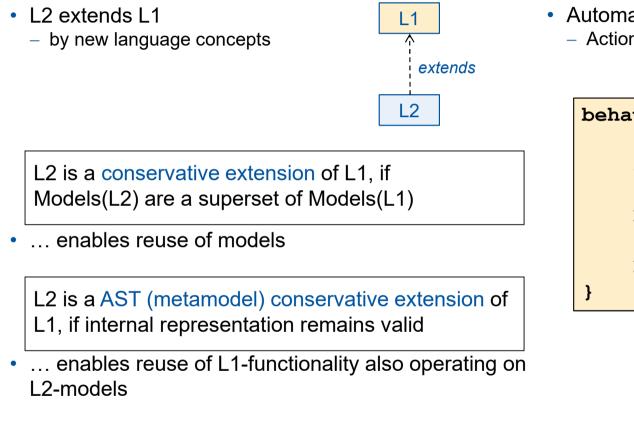


- MontiCore allows to extend grammars
 - and overwrite existing nonterminals (like in OOP)

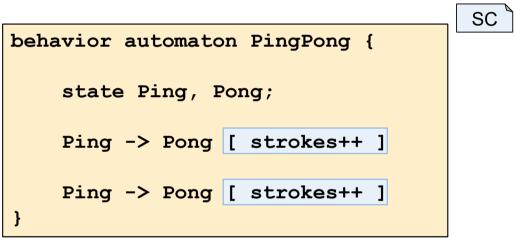
- MontiCore generates only the delta and thus allows
 - black-box reuse of language components
 - (parser is an exception)



Conservative Language Extension

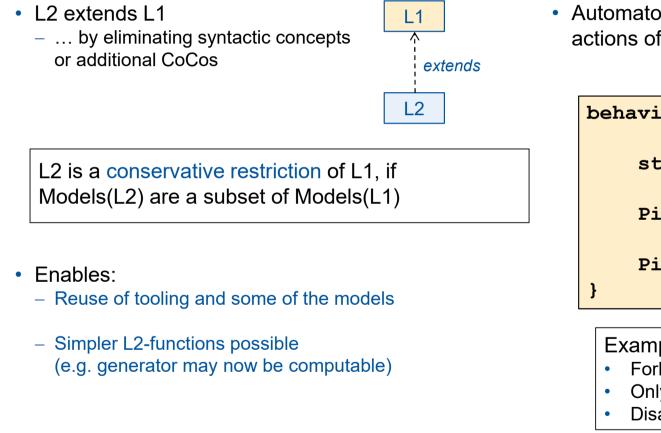


Automaton language L1 is extended by actions in L2:
 Actions are embedded at multiple places:

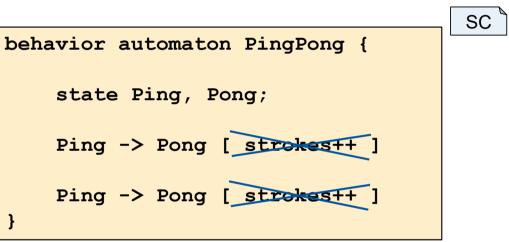




Language Restriction



 Automaton language L2 does not allow the Java actions of L2 anymore:



Examples:

- Forbid actions, forbid hierarchy of states
- Only basic types (integer ...) used
- Disallow certain names,



Software Engineering | RWTH Aachen

MontiCore Language Workbench

- Definition of modular language components
- Quick definition of domain specific languages (DSLs)
- Library of existing languages ٠
- Code generation

21

- Assistance for analysis ٠
- Assistance for transformations •
- Pretty printing, editors (graphical + textual)
- Namespaces/scopes, typing (fits GPL, UML) ٠
- Variability in syntax, context conditions, generation, semantics

with NJ, Alu, MSh, FDr, AHe, FDr, DS, KH, AW, PN, AR, HK, SV, HG, MS, AHo, IW, AHa, AP, ML, GV, MB et.al.

Monti









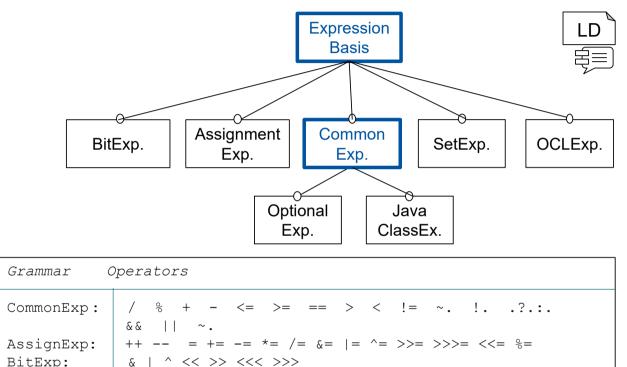
Oliver Kautz

http://www.se-rwth.de http://www.monticore.de

MontiCore's Expressions as Language Components

- 9 Grammars; 2⁴ * (2² +1) = 80 variants
- Explanation to be found at:
 - MC/.../Grammars.md
 - Details: Expressions.md
- ~100 nonterminals, most relevant: Expression
- Type checks implemented as visitors
- After parsing a small predefined transformation is applied
- Example:

```
!(i+1 \ge 3 * foo(1, 0xFE, x))
```



this .[.] (.). super .instanceof.

forall exists any let.in. .@pre

?: ?<= ?>= ?< ?> ?== ?!= ?~~

.isin. .in. union intersect setand setor

JavaClass:

Option.Exp:

SetExp:

OCLExp:



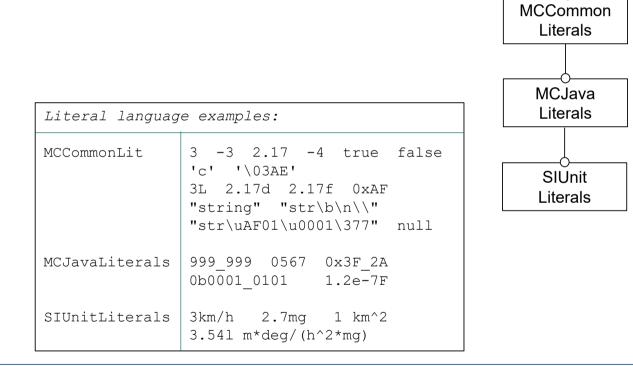
 $\{.|.\}$

?!~

MontiCore's Literals as Language Components

- 4 Grammars; 4 variants
- Explanation to be found at:
 - MC/.../Grammars.md
 - Details: Literals.md
- Used for primitives in Expression grammars
- ~95 nonterminals, most relevant:
 Literal implements Expression
- Type checks implemented as visitors
- Example for an expression:

```
i + 1mm >= 3km/h * foo("!")
```





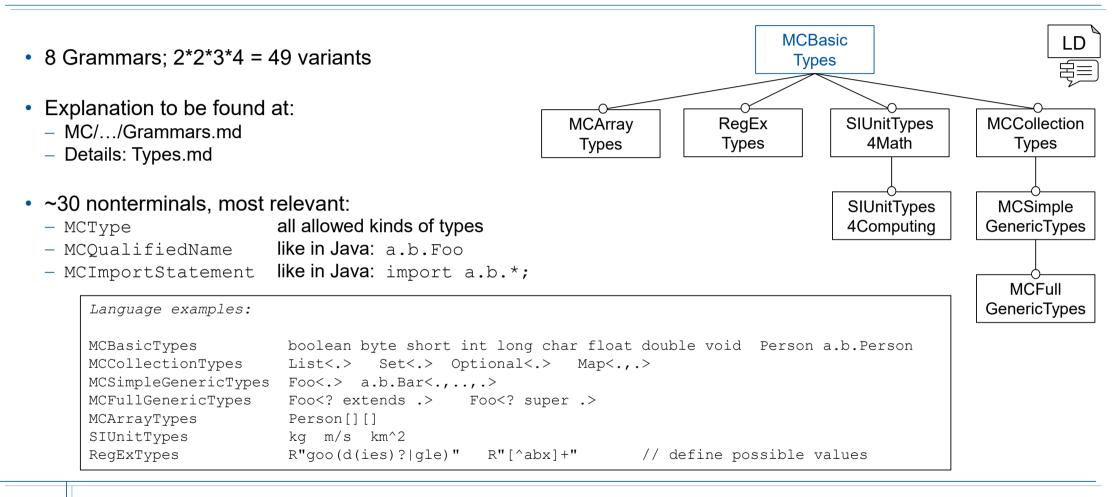
LD

<u>书</u>

MCLiterals

Basis

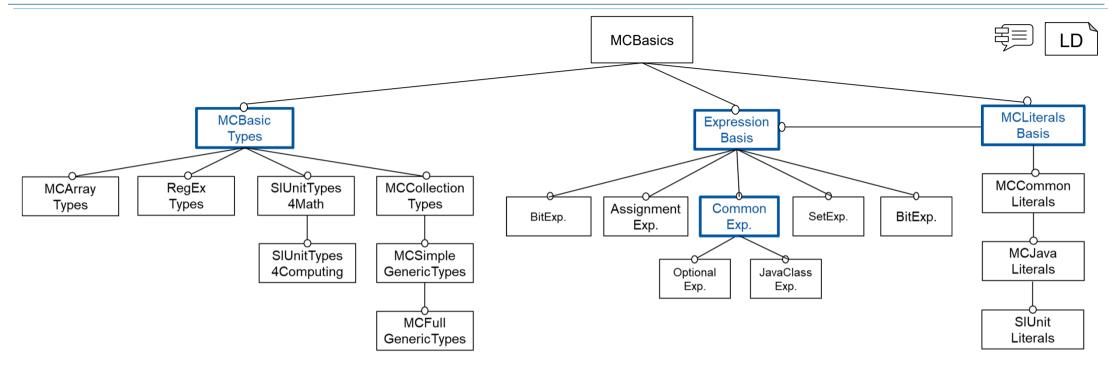
Types as Language Components



Software Engineering | RWTH Aachen



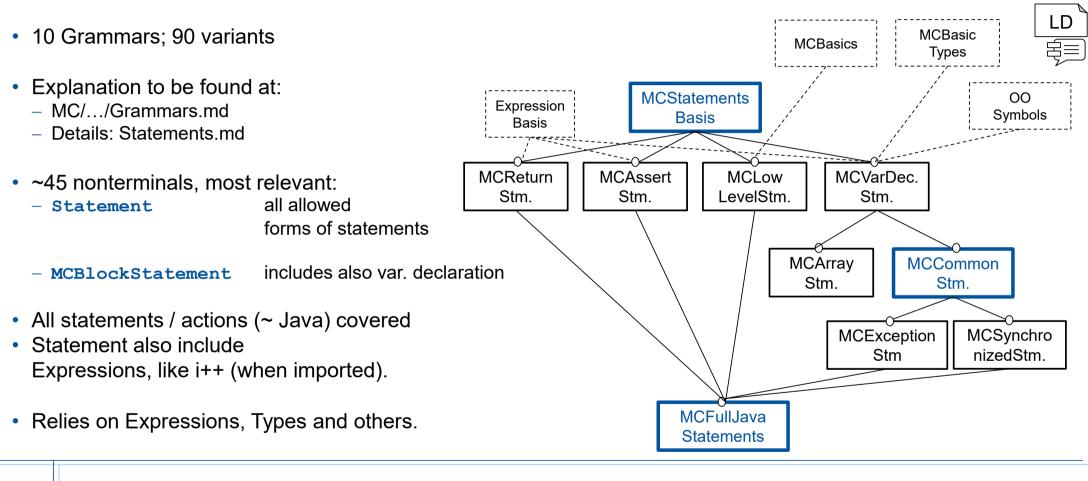
Summary: Component Hierarchy for Expression, Type and Literal



Overall Grammar Hierarchy for Expression, Type and Literals 15925 Variants possible

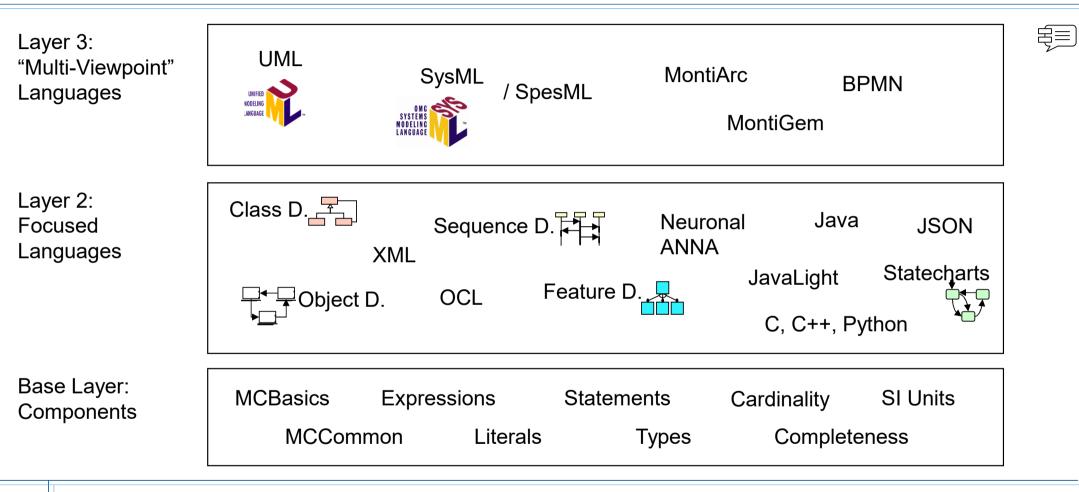


Statements as Language Components



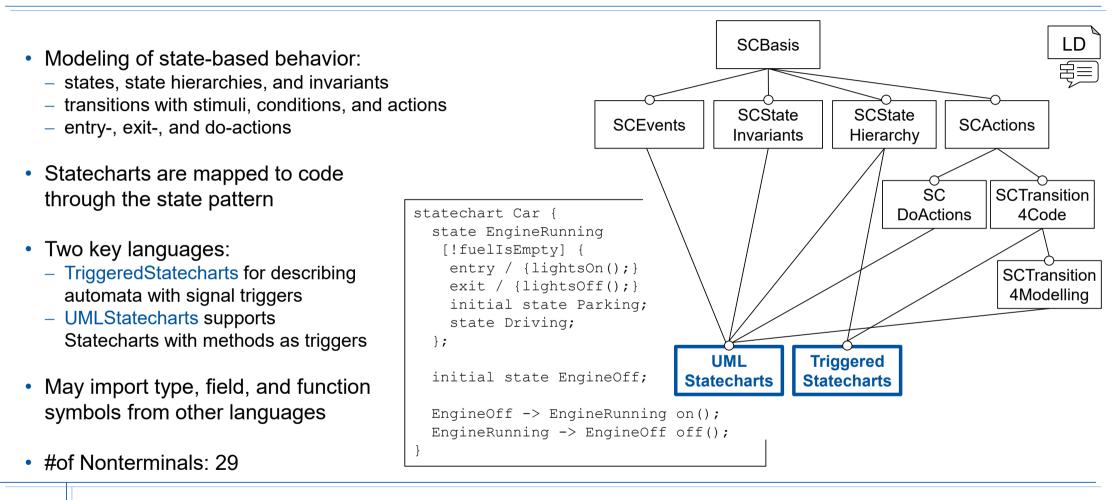
Software

MontiCore Libraries of Reusable Language Components Build a Language Zoo:





Statecharts





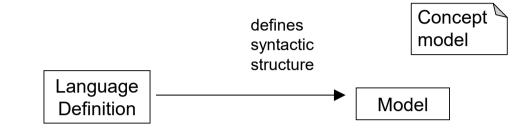
Is the role of libraries in modelling underestimated?

Software Engineering | RWTH Aachen



Role of Libraries in Language Definitions

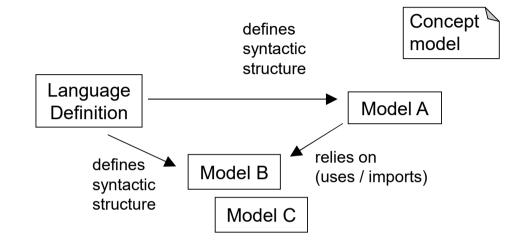
- Language ~ set of wellformed models.
- Wellformedness is defined in two levels of constraints:
 - 1. Basic definition (context free syntax)
 - 2. Context conditions (aka constraints)





Role of Libraries in Language Definitions

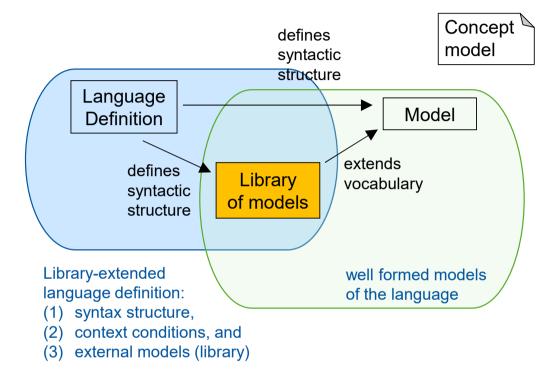
- A library ~ set of models that can be "imported".
- A model library
 extends the language vocabulary
- We distinguish:
 - Language structure (such as given by a grammar)
 - Language vocabulary:
 - Usually lightweight extensions of a language that can be defined within the language itself





Role of Libraries in Language Definitions

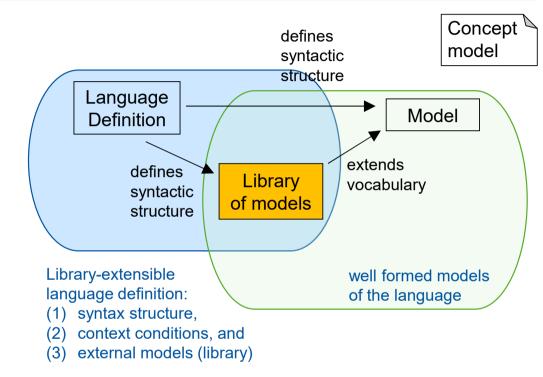
- Examples for vocabulary definitions:
 - Java allows to define classes, methods, variables
 - CD's allow to define classes,
 - Statecharts: states, ...
 - Scala, C++ allow infix operations: .>>.
 - Natural language has glossaries
- Good languages allow
 - introduce new symbols,
 - define "meaning" using the language itself,
- Model libraries
 - 1) modularize/decompose models to allow reuse
 - 2) allow lightweight language extension





Extensible Language?

- A language is defined in three stages:
 - 1. syntactic structure
 - 2. context conditions
 - \rightarrow well-formed models:
 - 3. predefined, external models (libraries)
 - → library-extensible language definition
- Consequence:
 - SLE needs to engineer extensible languages:
 - Definition of symbols, their meaning, and their use
 - Import / include / rely on other models





In MontiCore: Grammars define where symbols are defined and used

Old:		
1	grammar StateChart {	MG
2	Automaton = "automaton" Name "{" Body "}"	
3	Body = (State Transition)* ";"	
4	<pre>State = "state" Name "<<initial>>"?</initial></pre>	
5	Transition = from:Name "-" input:Name ">" to:Name	}

Symbol enhanced grammar:

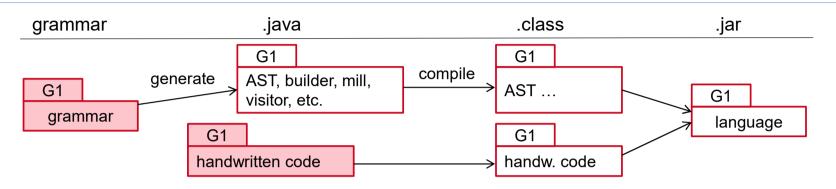
34

11	grammar StateChart {						
12	symbol scope Auto	omaton =	"automaton" Name "{" Body "}"	MG			
13	Body		(State Transition)* ";"				
14	symbol Stat	:e =	"state" Name "< <initial>>"?</initial>				
15	Tran	sition =	<pre>from:Name@State "-" input:Name ">" to:Name@State</pre>	}			

"symbol" : here symbols are defined
 "Name@State" : here State-symbols are used
 MontiCore generates the full infrastructure for symbol management including typecheck, load/store symbol tables. (and its also language compositional!)



Generation Chain (in MontiCore)

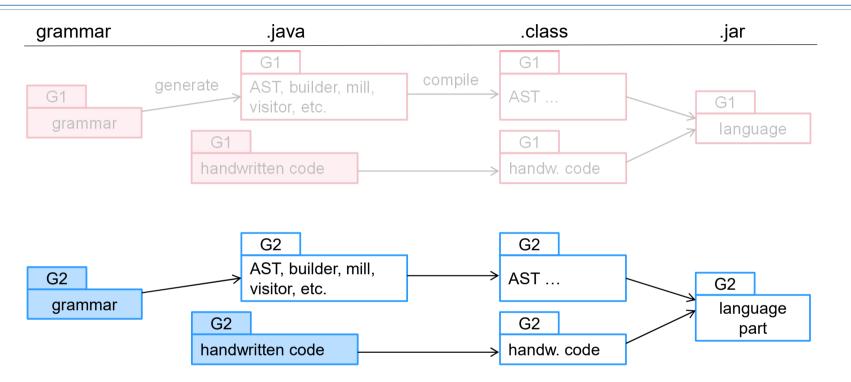


- MontiCore generates the full infrastructure for
 - syntax tree,
 - parsing,
 - traversing/visitors,
 - symbol management,
 - typecheck,
 - load/store symbol tables
 - syntax tree transformations
 - prettyprinting

- MontiCore automatically integrates handwritten code via Design Patterns
- (and NOT via copy/paste into generated code)



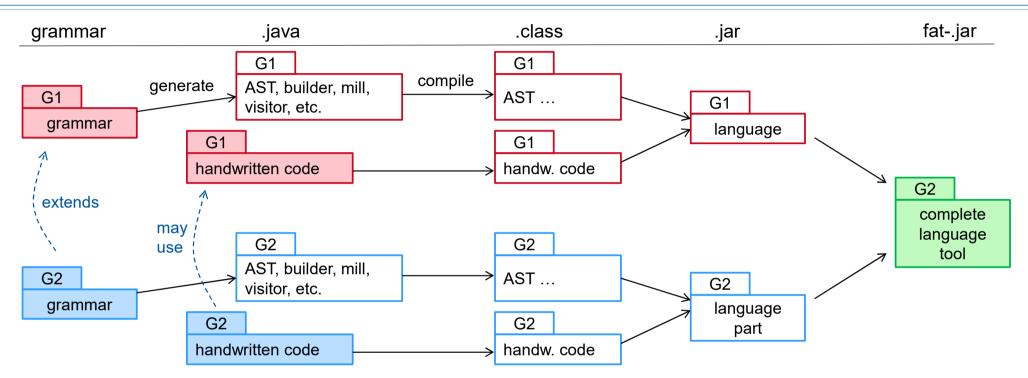
Modularity in the Generation Chain (in MontiCore)





Software Engineering | RWTH Aachen

Compositionality in the Generation Chain (in MontiCore)



- Generated parts are compositional:
 - late binding of pre-generated/pre-compiled code

and reuse of handwritten code is without adaptation



Question:

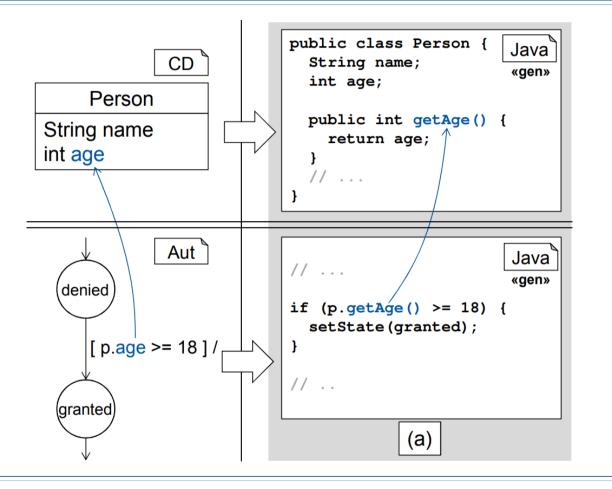
Composition of backends: generators resp. their artifacts?

Software Engineering | RWTH Aachen



Generated Artifacts: They use, extend, configure each other: An Example

- Generator 1 creates class "Person"
 and functions for attribute "age"
- Generator 2 needs to know,
 - how to assess "age"
 - Here using "getAge()"
 - and thus imports Person
- Gen.2 knows the result of Gen.1
- Gen.1 is independent of Gen.2

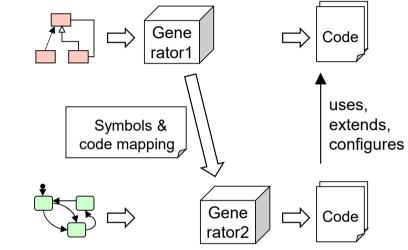




Modularity of Code Generation: Collaboration between Generators

- Main Solution: Modular mapping of models,
 - i.e. process only one model at a time: completely decoupled generators
 - code uses various design patterns
 - let generators "communicate" via stored symbol table
- Stored symbol table contains information about the mapping of a symbol to code (typically CRUD using Freemarker templates)
- Advantage:
 - Independent run; incremental re-run; efficient (if no circular dependency)
- MontiCore doesn't force to compose generator backends, but keeps them independent & communicating







Question:

What can be done with such an infrastructure?

Software Engineering | RWTH Aachen

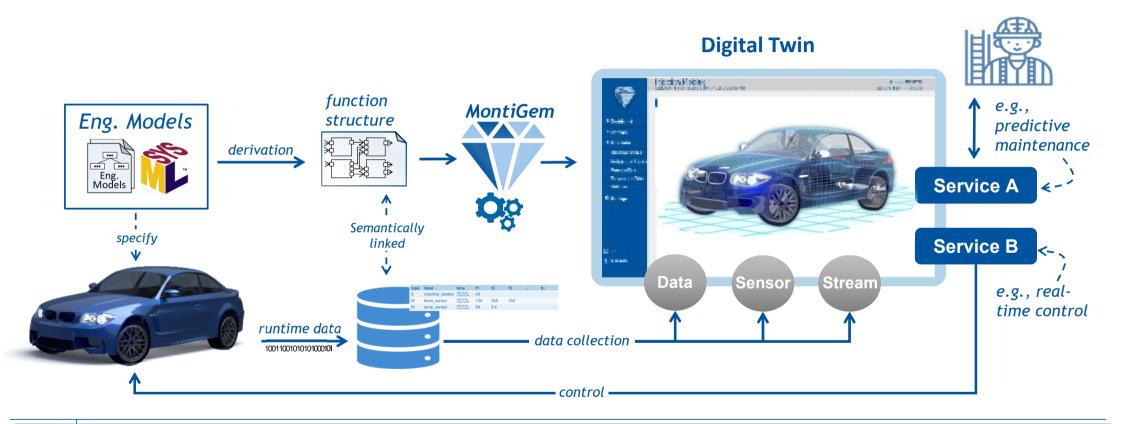


GUI Modeling in MontiGem: For Information Systems, Digital Twins, IOT-Services

€ Finanzen	_	en → Profil		admin STESTDB 9-09-12T14:19) 12.09.2019 Instanz-Verwaltung	1 2 3 4 5 6 7 8 9	<pre>class User { CD4A String username; Optional<string> encodedPassword; ZonedDateTime registrationDate Optional<string> initials; String email; boolean authentifiziert; Optional<string> timID; }</string></string></string></pre>	Data structure
** Personal * * Einstellungen * Mein Profil * Benutzer/Rollen * Daten Import *		E-Mail Adresse Kürzel Registrierungsdatum Altes Passwort	macoco@se.rwth-aachen.de N.N. 29.11.2018	?	1 2 3 4 5 6 7 8 9	<pre>datatable "meinBenutzerInfoTabelle" { GUI-DSL columns < uit { row "Benutzername" , <username "e-mail="" "kürzel"="" "registrierungsdatum"="" "tim-kennung"="" (editable)="" ,="" <email="" <initials="" <tim="" adresse"="" date(<registrationdate)="" pre="" row="" }="" }<=""></username></pre>	User interface
Drucken THILFE ABMELDEN V1.13.1 Impressum Datenschutz		Min. 5 Zeichen Min. 5 Zeichen Min 5 Zeichen	Passwort	Ändern	1 2 3 4 5 6	<pre>context User inv isPasswordValid: password.length() < 5; shortError: "Min. 5 Zeichen"; error: "Das Passwort muss aus mindestens 5 Zeichen bestehen, hat aber nur " + passwort.length() + " Zeichen. ";</pre>	Constraints



Transformation from the Engineering Models to the Digital Twin





Software Engineering | RWTH Aachen

Summary: Software Languages

- ... are essential for the progress of digitalization.
- Engineering a language means:
 - language composition
 - ... refinement
 - ... embedding
 - ... aggregation
 - ... extension
 - ... derivation
 - reusing language components
 - libraries of language components
 - tools
 - And language variants + tools are almost for free
- Ludwig Wittgenstein (philosopher):
 "The limits of my language are the limits of my world."







