

Domain-Specific Languages

Tijs van der Storm



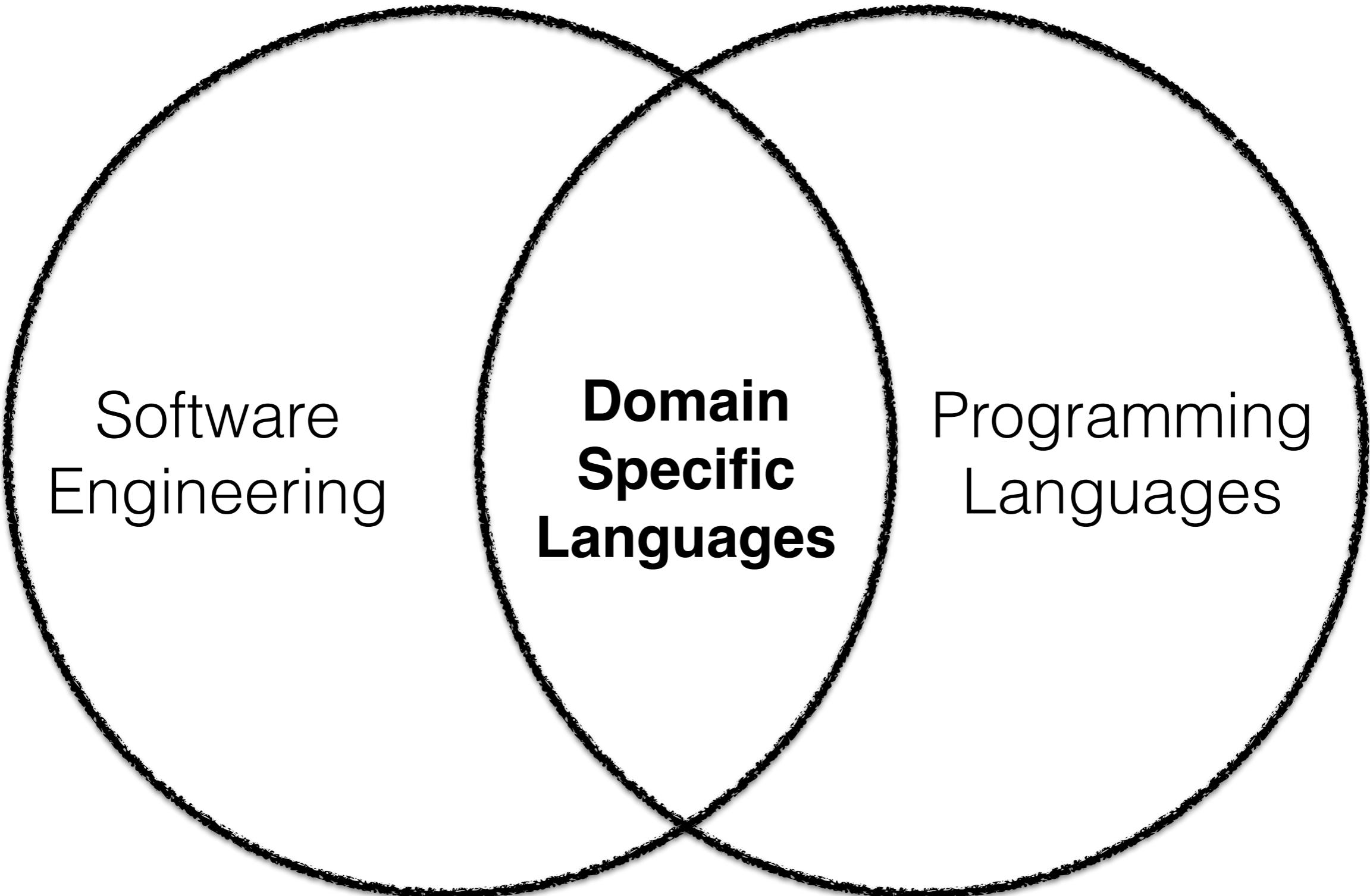
Centrum Wiskunde & Informatica



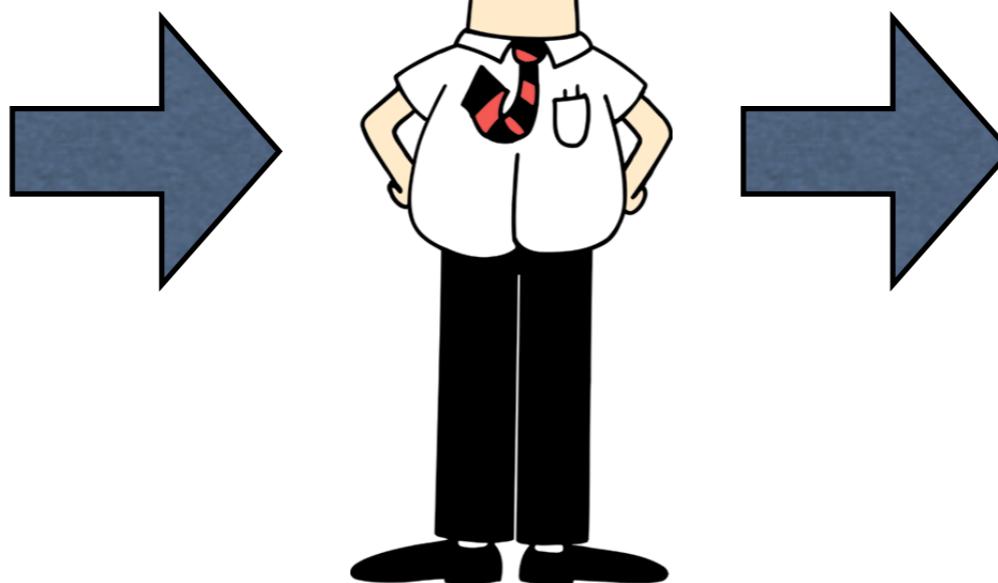
university of
groningen

Domain-specific languages

- Formal representations of domain knowledge
- AKA: Executable specification languages
- Write down “what”, generate the “how”
- Tailored to particular problem domain (e.g. Tax administration, digital forensics, etc.)
- Easier to validate by domain experts



Programming



```
.text:131411EF ; FUNCTION CHUNK AT .text:13141239 SIZE 000
.text:131411EF ; FUNCTION CHUNK AT .vmp0:13143000 SIZE 000
.text:131411EF ; FUNCTION CHUNK AT .vmp0:13143522 SIZE 000
.text:131411EF ; FUNCTION CHUNK AT .vmp0:13145C99 SIZE 000
.text:131411EF 68 F1 4F 5B FF
.text:131411F4 E9 A0 4A 00 00
.text:131411F4
.text:131411F4
.text:131411F4
.text:131411F4 64 69 31
.text:131411FC 60 16 42 75 9A C4+
.text:131411FC CD CA FA 7B 9D 7B+
.text:131411FC D5 D4 48 2B 85 4B+
.text:13141238
.text:13141238 58
.text:13141239
.text:13141239
.text:13141239 52
.text:1314123A 55
.text:1314123B 53
.text:1314123C 51
.text:1314123D 9C
.text:1314123E 57
.text:1314123F 50
.text:13141240 56
.text:13141241 51
.text:13141242 68 00 00 00 00
.text:13141247 8B 74 24 28
.text:13141248 BF F9 11 14 13
.text:13141250
.text:13141250 89 F3
.text:13141252 03 34 24
.text:13141255
.text:13141255
.text:13141255 ; FUNCTION CHUNK AT .text:13141255 SIZE 000
.start push 0FF5B4FF1h
.start jmp loc_13145C99
.start endp
; -----
byte_131411F9 db 64h, 69h, 31h ; D
byte_131411F9 dd 75421660h, 0CACDC49Ah, 7
byte_131411F9 dd 6EEE5440h, 10D69610h, 0F
byte_131411F9 dd 0EB6323E4h, 82B58465h, 0
; -----
.pop eax
; START OF FUNCTION CHUNK FOR start
loc_13141239: push edx
loc_13141239: push ebp
loc_13141239: push ebx
loc_13141239: push ecx
loc_13141239: push edi
loc_13141239: push eax
loc_13141239: push esi
loc_13141239: push ecx
loc_13141239: push 0
loc_13141239: mov esi, [esp+2Ch+var_4]
loc_13141239: mov edi, offset byte_131411F9
; -----
loc_13141250: mov ebx, esi
loc_13141250: add esi, [esp+2Ch+var_2]
loc_13141250: ; S
loc_13141255: ; C
loc_13141255: ; S
```

Domain
Knowledge

Hard labour

Code



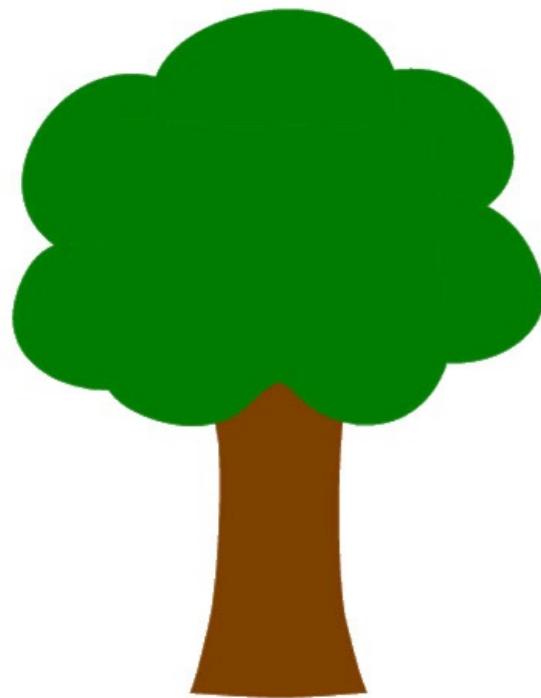
Change

???

Domain-specific
Language

Automation

Code



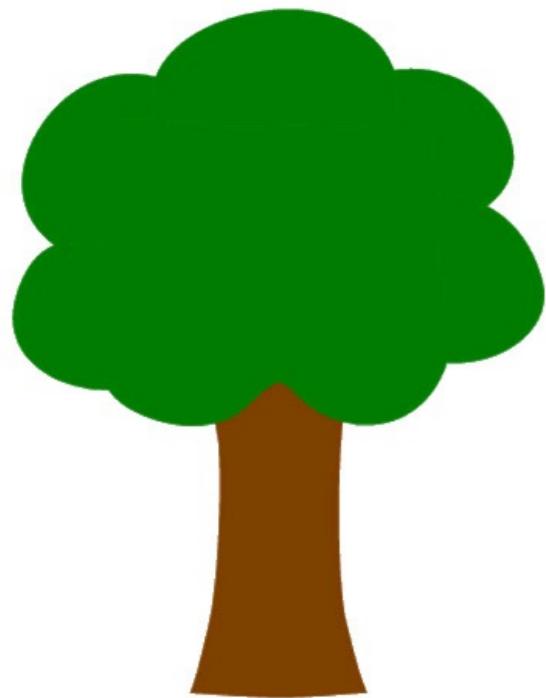
“What”
formalized
requirements

“How”
formalized design
and architecture

Domain-specific
Language

Automation

Code

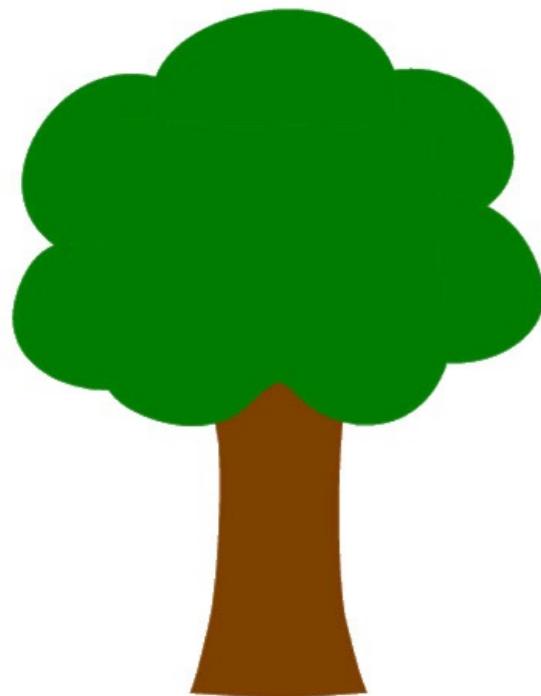


Change

Domain-specific
Language

Automation

Code



Programming is lossy

- encoding
- obfuscating
- encrypting
- dispersing
- tangling
- distorting



Time consuming



SCRIPTORIUM MONK AT WORK. (From *Lacroix*.)

Change is error-prone



Cognitive distance



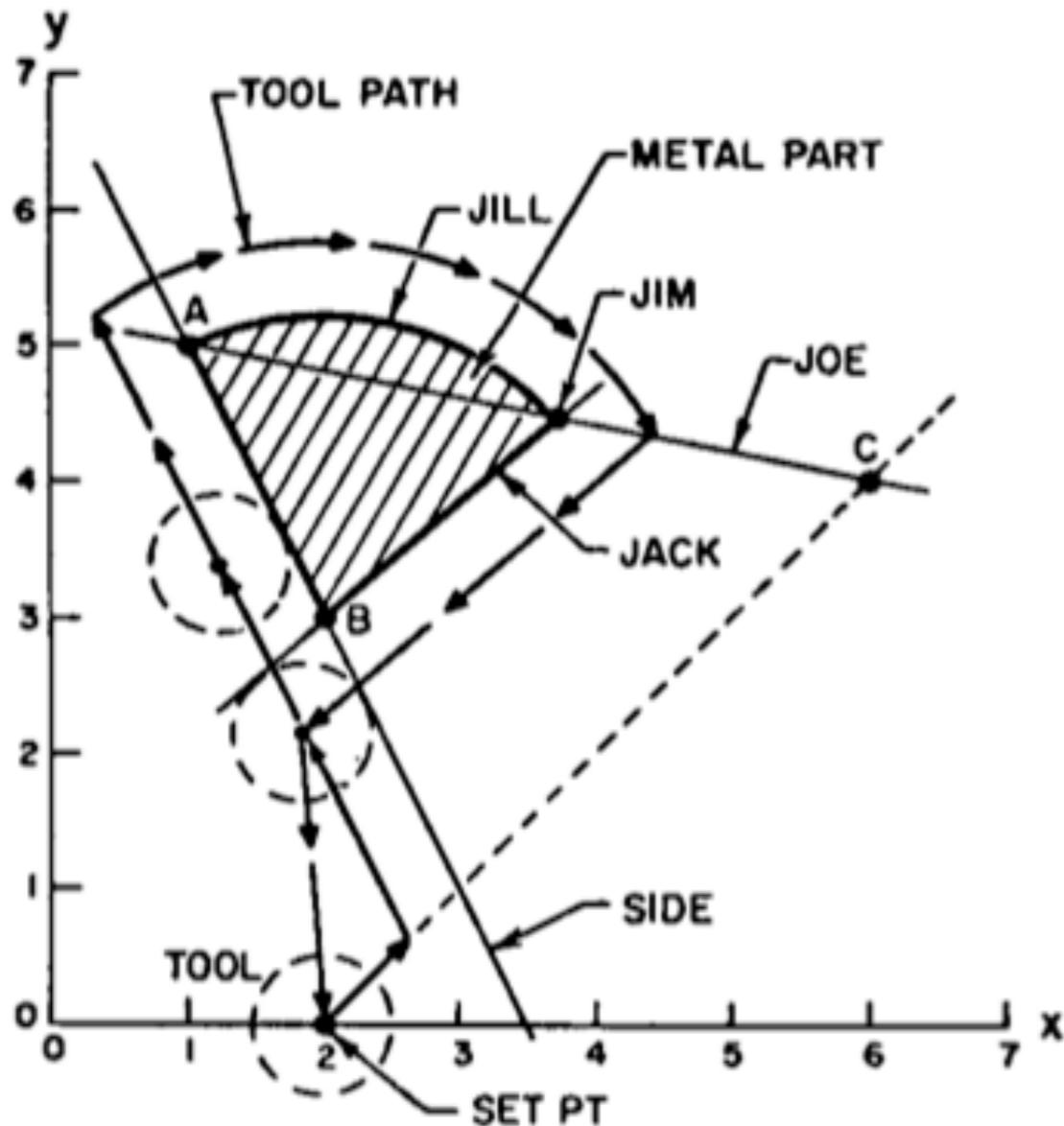
General purpose
languages



Domain-specific
languages



APT: robotic control



A • POINT / 1, 5
B • POINT / 2, 3
C • POINT / 6, 4
TL DIA / +1.0, INCH
FEDRAT / 30, IPM
SET PT • FROM, POINT / 2, 0
IN DIR, POINT / C
SIDE • GO TO, LINE / THRU, A, AND, B
WITH, TL LFT, GO LFT, ALONG / SIDE
JILL = GO RGT, ALONG, CIRCLE / WITH, CTR AT, B, THRU, A
JOE = LINE / THRU, A, AND, C
JIM = POINT / X LARGE, INT OF, JOE, WITH, JILL
JACK • LINE / THRU, JIM, AND, B
GO RGT, ALONG / JACK, UNTIL, TOOL, PAST, SIDE
GO TO / SET PT
STOP, END, FINI

from
the '50s (!)

Risla: financial products

product LOAN

declaration

contract data

PAMOUNT : amount
STARTDATE : date
MATURDATE : date
INTRATE : int-rate
RDMLIST ::= [] : cashflow-list

%% Principal Amount
%% Starting date
%% Maturity data
%% Interest rate
%% List of redemptions.

information

PAF : cashflow-list
IAF : cashflow-list

%% Principal Amount Flow
%% Interest Amount Flow

registration

%% Register one redemption.
RDM(AMOUNT : amount, DATE : date)

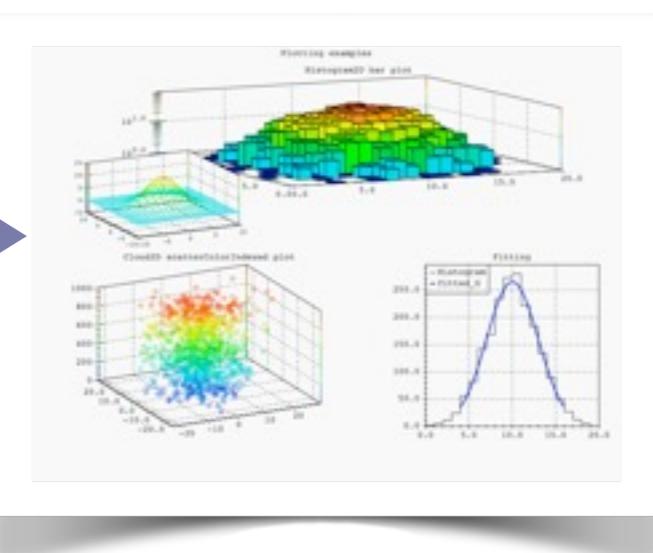
Developed
at CWI

Case 1: digital forensics



Acquisition

Recovery



Analysis

File carving tools

- Carving: recovering files without using file system
- Hard to change because of tangling
- Hard to change because highly optimized
- Complicating forces:
 - high variability in the domain (formats, devices)
 - deadline pressure during pre-charge detainment

Derric

format jpeg

extension jpeg jpg jfif

unit byte

size 1

sign false

type integer

endian big

strings ascii

sequence

SOI

([APP0JFIF APP0JFXX?] [APP1 APP2?])

!(SOI APP0JFIF APP0JFXX EOI)*

EOI

- Separate file format description from carving algorithms
- File formats can be understood by forensic experts
- Generate high performance code.

Lessons learned

- Derric carvers perform at least as good as existing state-of-the-art carvers
 - Accuracy and run-time performance
- Derric file carver amounts to ± 1000 SLOC
 - Including: specification, code generator, runtime
- Compare: PhotoRec, Revivelt are huge C programs with only builtin file format support.

Case 2: tax administration

Gegevensspecificatie

Behorende bij de specificatie
van de berichtstructuur
DA-2015

Versie 3.0 van 17/8/2015

Versie 3.0
Definitief
17/8/2015

1147 page
PDF
document

The running program

Aangifte inkomstenbelasting 2010 – Persoonlijke gegevens

Persoonlijke gegevens

Naam: Bla
Telefoonnummer: 323
Burgerservicenummer/sofinummer: 1430.95.067
Geborendatum: 11-02-1979
Nummer belastingconsulent:

Hebt u van ons bericht ontvangen om aangifte te doen? Ja Nee
Wilt u een rekeningnummer opgeven of wijzigen? Ja Nee

Uw persoonlijke situatie in 2010: Een deel van 2010 getrouwd
Periode dat u getrouwde was in 2010: 01-02 03-05
Woonde u voor of na deze periode samen met uw echtgenoot? Ja Nee
Willen u en uw echtgenoot heel 2010 als fiscale partners worden beschouwd? Ja Nee
Woonde u buiten de periode dat u getrouwde was nog met iemand anders samen? Bijvoorbeeld met uw kind van 27 jaar of ouder? Ja Nee

IB 602E - ZZ01FOL2A

Akkoord

Stoppen Instellingen Rekenmachine Help Printen Open bestand

Consequences

- Enormous amount of work: programmers manually implement ± 1500 data items + interaction logic
- Gargantuan testing effort: lots of data, lot of choices in the UI, — no automated support.
- Validation: how to know if the code is correctly implementing the spec?
- Every year, everything changes again, but unit of change \neq unit of code.

Using a DSL

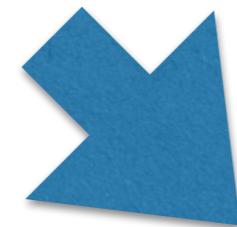
extract

Gegevensspecificatie

Behorende bij de specificatie
van de berichtstructuur
DA-2011

Versie 2.3 van 29/9/2011

Versie 2.3
Definitief
29/9/2011



formal
specification



generate



Aangifte inkomstenbelasting 2010 – Persoonlijke gegevens

Persoonlijke gegevens

Naam: Bla
Telefoonnummer: 323
Burgerservicenummer/sofnummer: 1430.95.067
Geboortedatum: 11-02-1979
Nummer belastingconsulent:

Hebt u ons bericht ontvangen om aangifte te doen? Ja Nee
Wilt u een rekeningnummer opgeven of wijzigen? Ja Nee
Uw persoonlijke situatie in 2010
Periodie dat u getrouwd was in 2010: 01-02 03-05
Woonde u voor of na deze periode samen met uw echtgenoot? Ja Nee
Willen u en uw echtgenoot heel 2010 als fiscale partners worden beschouwd? Ja Nee
Woonde u buiten de periode dat u getrouwde was nog met iemand anders samen? Bijvoorbeeld met uw kind van 27 jaar of ouder? Ja Nee

Een deel van 2010 getrouwd? Ja Nee
Een deel van 2010 samen wonende? Ja Nee
Een deel van 2010 samen gehuwd? Ja Nee

IB 602E - Z201FOL2A

Stopper Instellingen Rekenmachine Help Printer Open bestand Akkoord

gegeven periode samenwonen in belastingjaar [117315]: DDMMEEJJ {
alias EPerSamenwonen
minimaal voorkomen 0
maximaal voorkomen 1
gebruikt in
 117308
pagina 73
definitie {
 De periode in het belastingjaar waarin de
 aangever en de partner onafgebroken een
 gezamenlijke huishouding hebben gevoerd
 en op hetzelfde woonadres stonden ingeschreven.
}
bron {
 Wet IB2001, art 1.2
}
conditie
[117315.B]<periode samenwonen in belastingjaar.Begindatum>
 <= [117315.E]<periode samenwonen in belastingjaar.Einddatum>
}

EasierTax

- ▼ DA-2015
- ▼ 3.0
 - ▼ Inkomensheffingen, Zorgverzekeringswet, Ve
 Zorgverzekeringswet
 - ▼ Inkomstenbelasting en premies volksverzek
 ▼ Inkomensopstelling kwalificerende buiten
 - ▼ Inkomensopstelling kwalificerende buiten
 Vermeerdering-vermindering inkome
 - ▼ Grondslag inkomsten verhouding Ned
 Inkomen belast Nederland samenst
 Inkomen in en buiten Nederland sam
 - ▼ Inkomensopstelling partner kwalificerer
 Vermeerdering-vermindering inkome
 - ▼ Grondslag inkomsten verhouding Ned
 Inkomen in en buiten Nederland pa
 Inkomen belast Nederland partner :
- ▼ Persoonsgebonden aftrek
 - ▼ Totaal aftrekbare giften
 - ▼ Andere giften
 - ▼ Andere giften aan culturele instellir
 Andere giften culturele instelling
 - ▼ Overige andere giften
 Overige andere giften spec
 - ▼ Andere giften aan steunstichtingen
 Andere giften steunstichtingen S
 - ▼ Periodieke giften
 - ▼ Overige periodieke giften
 Overige periodieke giften spec

Giften aftrekbaar aangever migratiejaar Nederl periode

Het deel van de aftrekbare giften van de aangever dat betrekking heeft op de periode wonen Nederland in het migratiejaar.

Bronnen: [Wet IB 2001, art 6.32](#)

n..13

Andere giften culturele instelling

De giften die betaald, verrekend of ter beschikking zijn gesteld aan een culturele instelling.

Bronnen: [Wet IB 2001, art 6.32](#) [Wet IB 2001, art 6.39a](#)

n..13

Omschrijving andere giften culturele instelling

De naam van de culturele instelling of omschrijving waaraan de gift is gedaan.

Bronnen: [Wet IB 2001, art 6.33](#) [Algemene Wet Inzake Rijksbelastingen, artikel 5b](#)

an..70

+ -

Verhoging aftrek andere giften culturele instellingen

De verhoging van de aftrek van andere giften gedaan aan culturele instellingen.

Bronnen: [Wet IB 2001, art 6.39a, lid 1](#)

n..13

Andere giften culturele instellingen totaal

Het totaal van de giften aan culturele instellingen.

Bronnen: [Wet IB 2001, art 6.32](#) [Wet IB 2001, art 6.39a](#)

n..13

Aftrekbare andere giften

Het bedrag van de andere giften aan instellingen of aan steunstichtingen SBBI



Lessons we are learning

- Formal specification of data items and rules
- Enables:
 - consistency checking
 - generation of code
 - impact analysis upon new versions
 - ...

Shorter programs

- Commonality is in the language implementation
- Only specify relevant variation points
- Remove repetitive patterns / “boilerplate” which are error-prone
- “Higher level of abstraction”

Fewer degrees of freedom

- Less room to make mistakes
- Guaranteed absence of certain problems:
 - termination, resource usage, etc.
- Better checking / domain-specific error messages

Notation close to problem domain

- Facilitates validation / stake holder communication
- Towards “formalized” requirements

Untangling concerns

- Coupling/scattering issue resolved in code generator
- Think: security, optimization, contracts, persistence, etc.

Domain-specific AVOPT

- Limited expressivity + encoded domain knowledge:
 - better error messages
 - better performance
 - parallelization

Design reuse

- Code generator / interpreter encodes design knowledge
- Leveraged across multiple systems of a family.
- Change the design without changing application logic (“what”)

DSLs

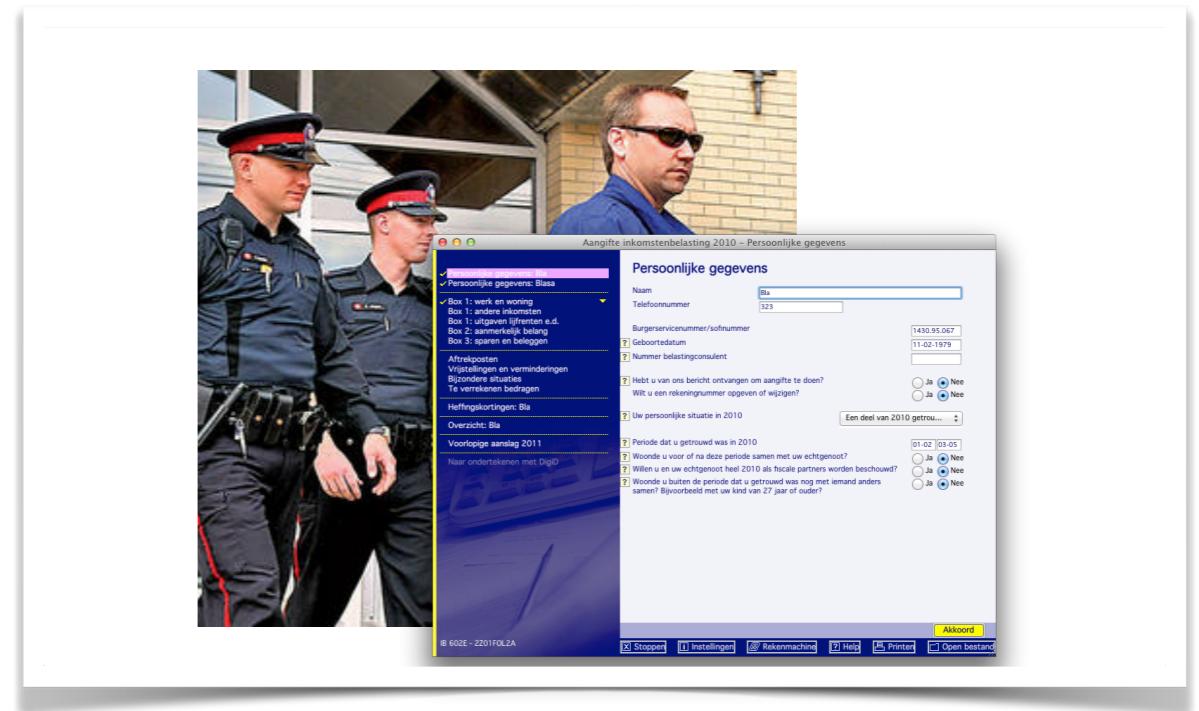
- “A programming language is low level when its programs require attention to the irrelevant.”
(Perlis, Epigrams in Programming)
- Ensure quality by language design & engineering
- Use the best language for the job



Little languages vs
general purpose



Raise level of abstraction;
generate the implementation



Examples: forensics,
tax administration

*Find the right language
to express the problem,
instead of adapting the
process to the solution*