Technische Universität Braunschweig

ISF

SAT Encodings of the At-Most-*k* Constraint

A Case Study on Configuring University Courses

Paul Maximilian Bittner, Thomas Thüm, Ina Schaefer

Tailoring Computer Science to your Taste

ZEUGNIS | CERTIFICATE

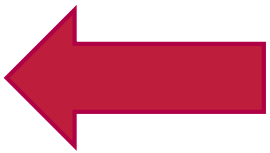
Master of Science

Herr | Mr.
Max Mustermann

geboren am | born on
01. Januar 2019 in Musterhausen

bestand die Masterprüfung im Studiengang | successfully completed the master degree in

Informatik | Computer Science
Studienrichtung: Medizinische Informatik | Field of Study: Medical Informatics
am 01. Januar 2019



mit der Gesamtnote | with an overall grade of

sehr gut (1,4) | excellent (1.4)

Choose (n)one of

- Big Data Management
- Automotive Informatics
- Hardware and Software System Design and Analysis
- Industrial Data Science
- IT-Security
- Medical Informatics
- Networked Systems
- Robotics
- Visual Computing

Specification on University's Website

Studierende
Master-Studiengang Informatik
Das Studium
Studienrichtungen
Big Data Management
Fahrzeuginformatik
Hardware-/Softwaresystement und -analyse
Industrial Data Science
IT-Sicherheit
Medizinische Informatik
Networked Systems

Studienrichtung "Big Data Management"

Motivation

„Big Data“ ist eine der zentralen Herausforderungen der Informations- und Wissensgesellschaft. Immer mehr verschiedene Datenarten werden immer schneller in immer größeren Mengen erhoben, charakterisiert durch die berühmten 3 Vs – Volume, Velocity und Variety. Das hat natürlich einerseits große Auswirkungen darauf wie wir mit diesen Daten umgehen, also auf das klassische Data Management. Aber es hat eben auch Auswirkungen auf die Methoden, aus diesen Daten Wissen zu generieren und dadurch letztendlich Nutzen aus ihnen zu ziehen. Zentrale textbasierte Datenquellen wie das World Wide Web spielen dabei heute eine ebenso große Rolle wie weitgehend automatisiert aufgenommene Daten von Sensorik, z.B. in der Umwelt (Klima-/Wetterdaten) oder industriellen Prozessen (Industrie 4.0).

Qualifikationsziele

Die Studienrichtung Big Data Management beschäftigt sich mit der Vermittlung und kompetenten Anwendung von informatischen Methoden zur Beherrschung der Datenflut im

Specification on University's Website

Studierende
Master-Studiengang Informatik
Das Studium
Studienrichtungen
Big Data Management
Fahrzeuginformatik
Hardware-/Softwaresystement und -analyse
Industrial Data Science
IT-Sicherheit
Medizinische Informatik
Networked Systems

Pflichtveranstaltungen (insgesamt 40 LP)

- Relationale Datenbanksysteme II (5 LP, INF-IS-49)
- Data Warehousing & Data-Mining-Techniken (5 LP, INF-IS-54)
- Masterarbeit (30 LP) – im Bereich Informationssysteme

Wahlpflichtveranstaltungen (insgesamt 30 LP)

- Projektarbeit (15 LP) - im Bereich Informationssysteme
- Seminar (5 LP) - im Bereich Informationssysteme
- Distributed Data Management (5 LP, INF-IS-48)
- Wissensbasierte Systeme und deduktive Datenbanksysteme (5 LP, INF-IS-51)
- Multimedia-Datenbanken (5 LP, INF-IS-52)
- Information Retrieval & Web Search (5 LP, INF-IS-53)
- Ausgewählte Themen der Informationssysteme (5 LP, INF-IS-55)

Example: Big Data Management

Compulsory (40 CP)

- Relational Databases II (5 CP)
- Data Warehousing & Data-Mining-Techniques (5 CP)
- Master Thesis (30 CP) – Topic: Information Systems

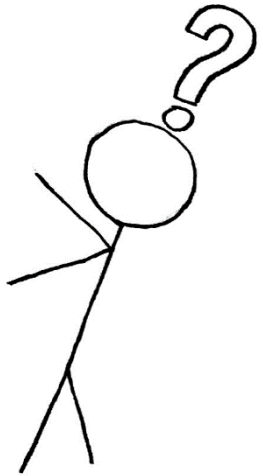
Compulsory Elective (at least 30 CP)

- Project Thesis (15 CP) – Topic: Information Systems
- Seminar (5 CP) – Topic: Information Systems
- …

Problem: Inconsistency

Compulsory (55 CP)

- one Lab from NetSys-Courses (5 CP)
- one Seminar from NetSys-Courses (5 CP)
- desired: Project Thesis (15 CP) – in one field of the branch
- Master Thesis (30 CP)



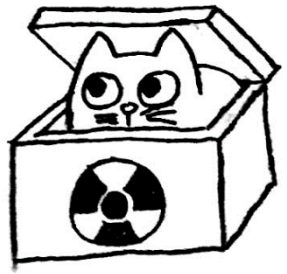
Problem: Ambiguity

Compulsory (55 CP)

- one Lab from NetSys-Courses (5 CP)
- one Seminar from NetSys-Courses (5 CP)
- desired: Project Thesis (15 CP) – in one field of the branch
- Master Thesis (30 CP)

Compulsory Elective (at least 15 CP)

- Project Thesis (15 CP) – in one field of the branch
- Seminar (5 CP) – in one field of the branch
- …



Problem: Optional Subjects being Mandatory

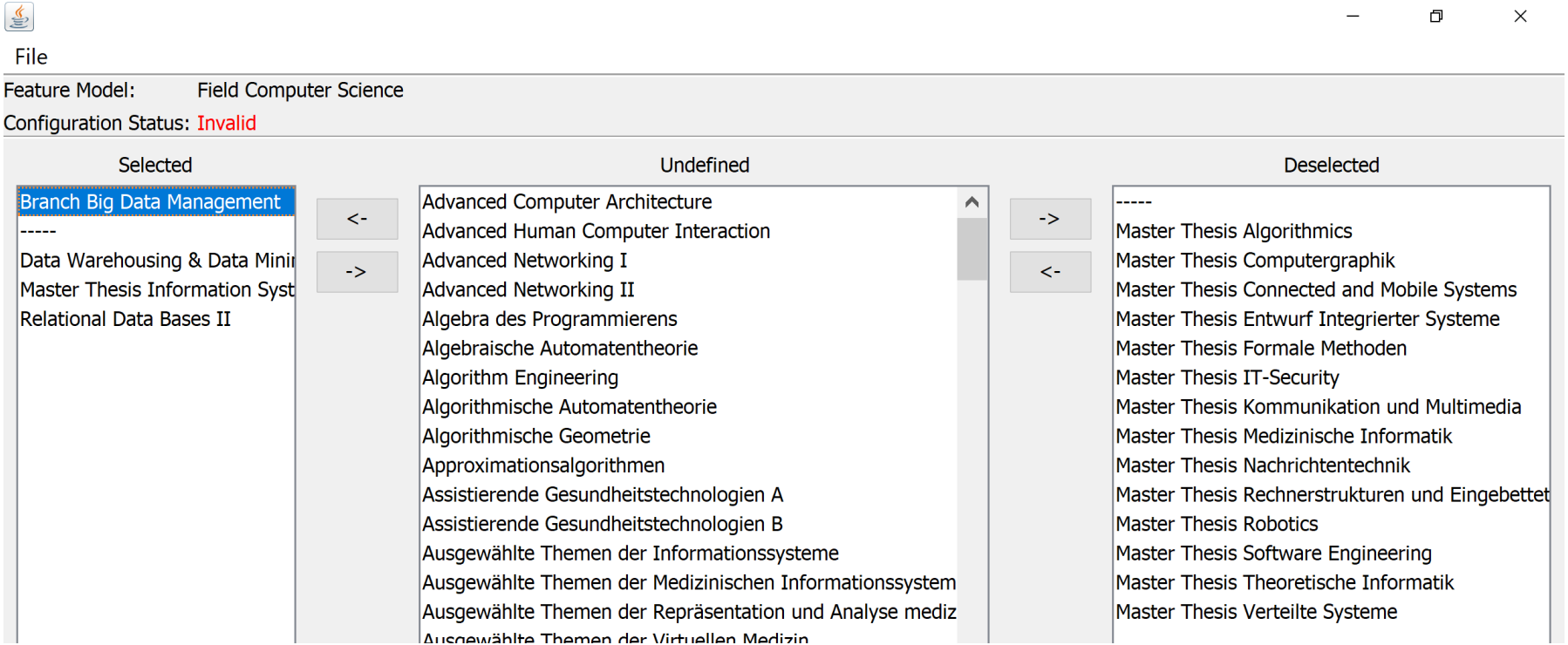
Compulsory Elective (at least 35 CP)

- Advanced IT-Security (5 CP)
- Machine Learning for IT-Security (5 CP)
- Lab on IT-Security (5 CP)
- Lab on Intelligent System Security (5 CP)
- Management von Information Security (5 CP)
- Operating Systems Security (5 CP)
- Project Thesis (15 CP)

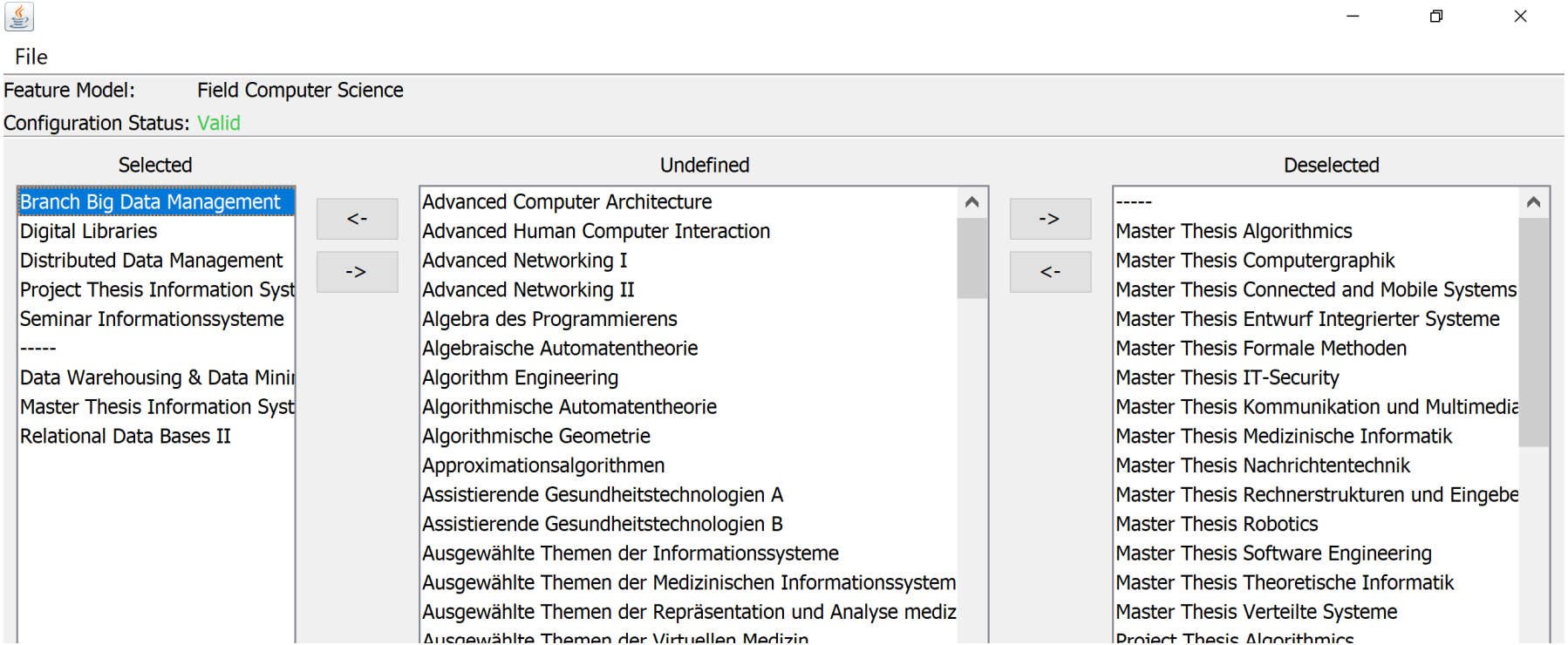
only 30 CP total

Problem: Partial Configurations

Solution: BroT (Branch of Study Tool)



Solution: BroT (Branch of Study Tool)



Implementation Idea: Use Feature Models

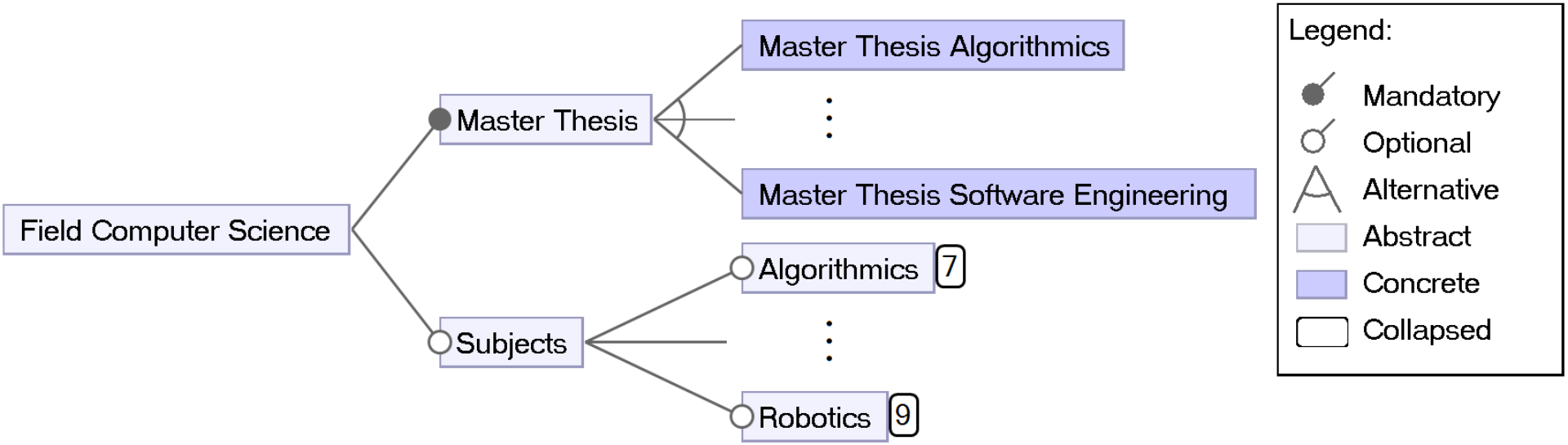
What is a Feature Model?

- Variability model for configurable systems

  (e.g., Software Product Lines)

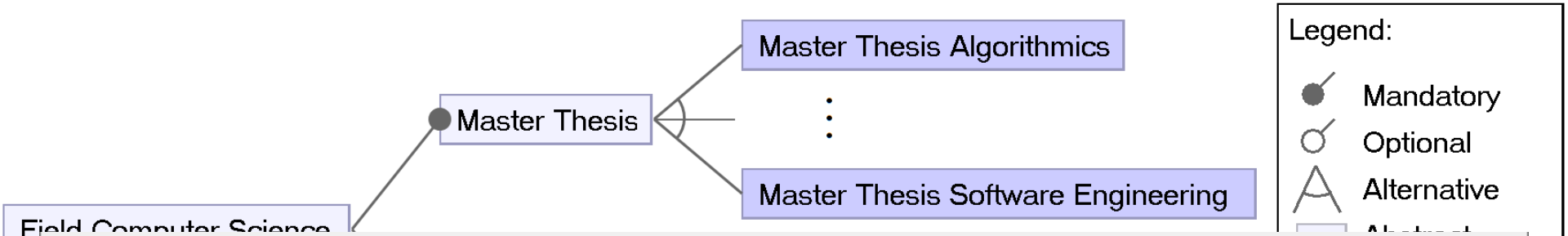
- Tree of features describing valid feature configurations

Example



+ optional boolean cross-tree constraints

Example



Selected		Undefined		Deselected
Master Thesis Algorithmics	<-	Lab Reaktive Systeme	->	-----
-----	->	Lab VLSI-Design II	<-	Master Thesis Computergraphik
		Logik in der Informatik		Master Thesis Connected and Mobile Systems
		Machine Learning for Pervasive Computing		Master Thesis Entwurf Integrierter Systeme
		Management von Informationssicherheit		Master Thesis Formale Methoden
		Maschinelles Lernen in der IT-Security		Master Thesis IT-Security
		Mathematische Methoden der Algorithmics		Master Thesis Information Systems
		MedizinRobotics		Master Thesis Kommunikation und Multimedia
		Medizinische Informationssysteme B		Master Thesis Medizinische Informatik
		Mensch-Maschine-Interaktion		Master Thesis Nachrichtentechnik
		Mobile Computing Lab		Master Thesis Rechnerstrukturen und Eingebettete Systeme
		Mobilkommunikation		Master Thesis Robotics
		Modellbasierte Softwareentwicklung		Master Thesis Software Engineering
		Multimedia-Datenbanken		Master Thesis Theoretische Informatik
		Mustererkennung		Master Thesis Verteilte Systeme
		Nebenläufigkeitstheorie		

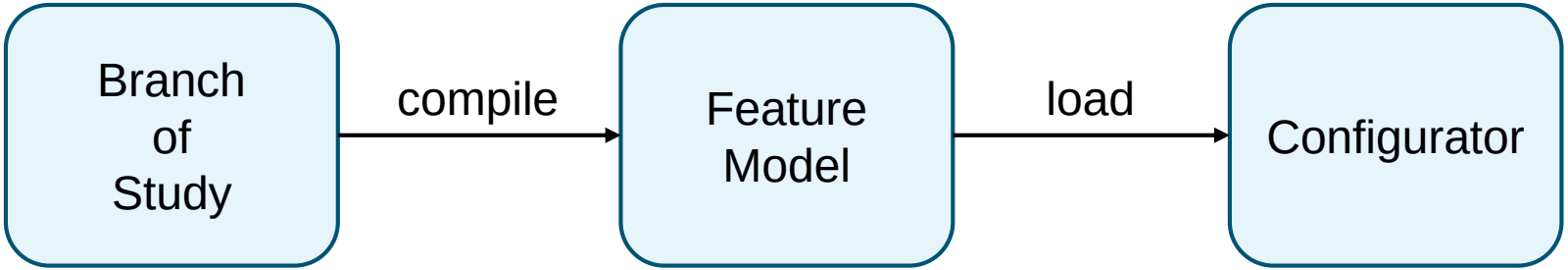
Why Feature Models?

Because the configurator comes for free [4, 6, 7, 8, 9, 10]

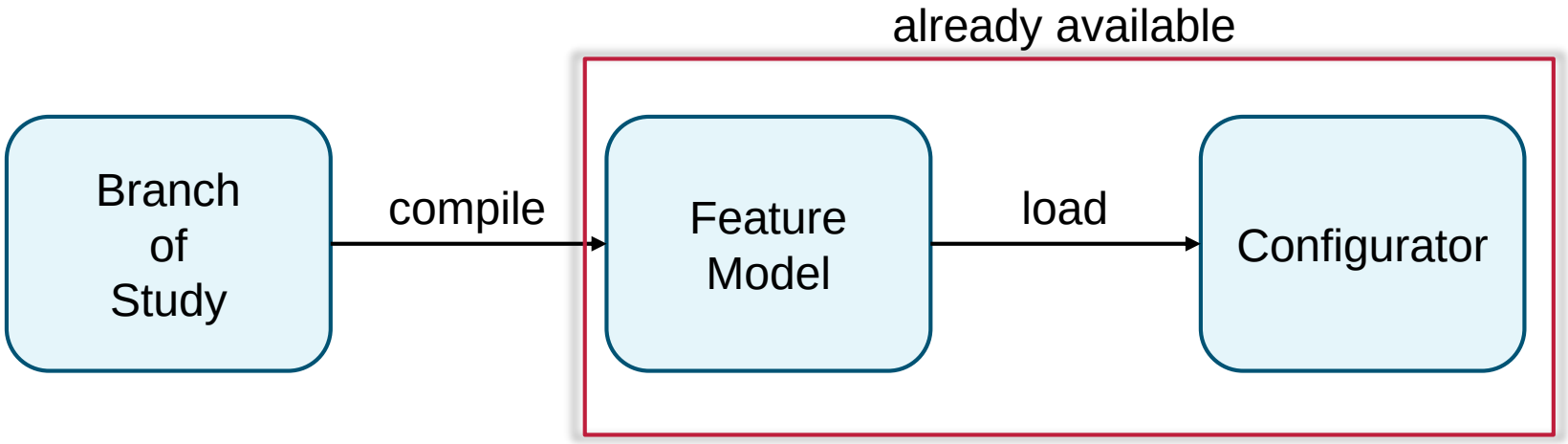
- decision propagation
- differentiation between manually and automatically chosen subjects
- (explanations on automatic subject selection.) [11]

⇒ Creating solely a feature model suffices!

Tool Design



Tool Design


already available
Branch of Study
compile
Feature Model
load
Configurator

Formal Specification with Domain Specific Language

```
Field “Computer Science”
  Branch “IT-Security”
    Compulsory
      “Seminar IT-Security”
      “Master Thesis IT-Security”

    CompulsoryElective 35 CP
      “Advanced IT-Security”
      “Machine Learning for IT-Security”
      “Lab on IT-Security”
      “Lab in Intelligent System Security”
      “Project Thesis IT-Security”
      “Management of Information Security”
      “Operating Systems Security”
```

Formal Specification with Domain Specific Language

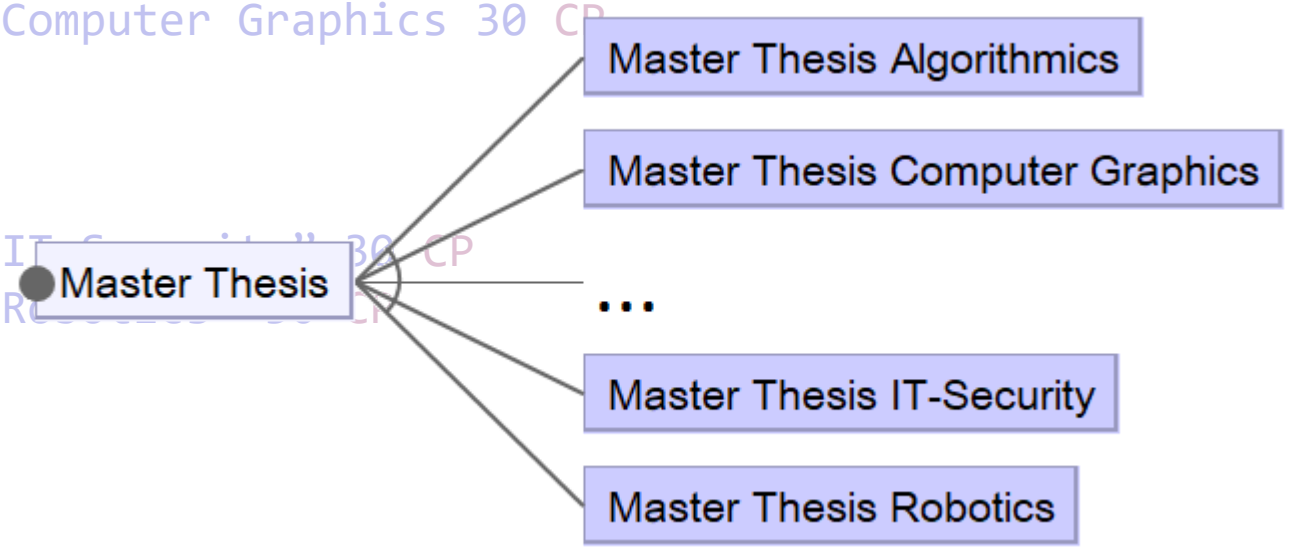
```
Category “Master Thesis” [1; 1] {
  ...
}
Category “Project Thesis” [0; 1] {
  ...
}
Category “Subject” {
  Category “Algorithmics” {
    “Computational Geometry” 5 CP
    “Lab on Algorithmics”    5 CP
    “Algorithm Engineering”  5 CP
    ...
  }
  ...
}
```

Compilation to Feature Model: *Subjects* ⟼ *Features*

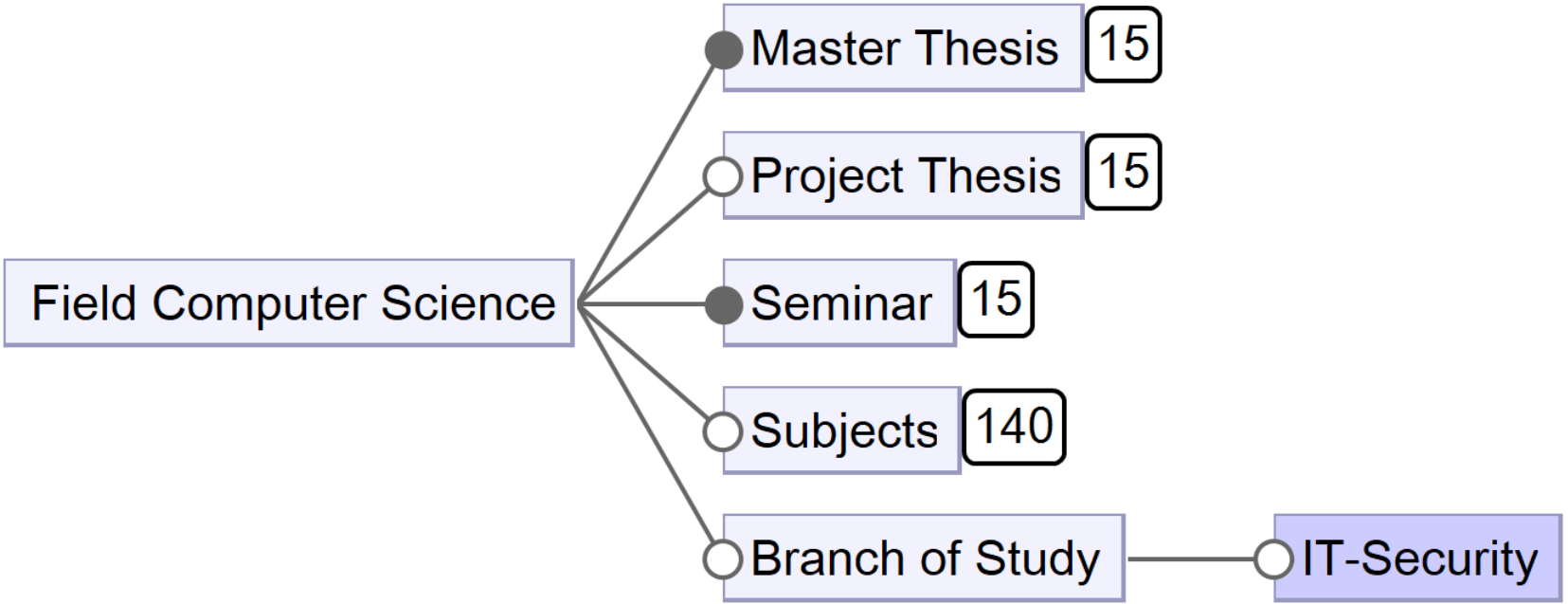
```
Category “Master Thesis” [1; 1] {
  “Master Thesis Algorithmics” 30 CP
  “Master Thesis Computer Graphics 30 CP

  ...

  “Master Thesis IT-Security” 30 CP
  “Master Thesis Robotics” 30 CP
}
```

Compilation to Feature Model: *Subjects* ⟼ *Features*

```
Category “Master Thesis” [1; 1] {
  “Master Thesis Algorithmics” 30 CP
  “Master Thesis Computer Graphics 30 CP

  ...

  “Master Thesis IT-S...” 30 CP
  “Master Thesis R...” 30 CP
}
```



Compilation to Feature Model: *Subjects* ⟼ *Features*



Compilation to Feature Model: *Branches* ⟼ *Constraints*

```
Compulsory
  “Seminar IT-Security”
  “Master Thesis IT-Security”
```

$$\text{IT-Security} \Rightarrow \text{Seminar IT-Security} \wedge \text{Master Thesis IT-Security}$$

Compilation to Feature Model: *Branches* ⟼ *Constraints*

```
CompulsoryElective 20 CP
  “Advanced IT-Security”
  “Machine Learning for IT-Security”
  “Lab on IT-Security”
  “Lab on Intelligent System Security”
  “Project Thesis IT-Security”
  “Management of Information Security”
  “Operating Systems Security”
```

IT-Security ⇒ choose at least 4

How to express $atleast_k$?

With $atleast_k(S) \equiv atmost_{|S|-k}(\neg S)$

- a standard problem.
- well-researched.
- a metric for benchmarks.

How to express $atmost_k$?

Not more than k variables can be true

⇒ build all subsets of size $k + 1$

⇒ at least one variable has to be false in those subsets

$$atmost_1(\{A, B, C\}) \equiv (\neg A \vee \neg B) \wedge (\neg A \vee \neg C) \wedge (\neg B \vee \neg C)$$

Problem: Grows with $\binom{n}{k+1}$, i.e., really fast!

1 GB constraint for Automotive Informatics (XML)!

How to express $atmost_k$ efficiently?

Frisch and Giannaros present a number of encodings [1]:

Binomial	naïve approach

How to express $atmost_k$ efficiently?

Frisch and Giannaros present a number of encodings [1]:

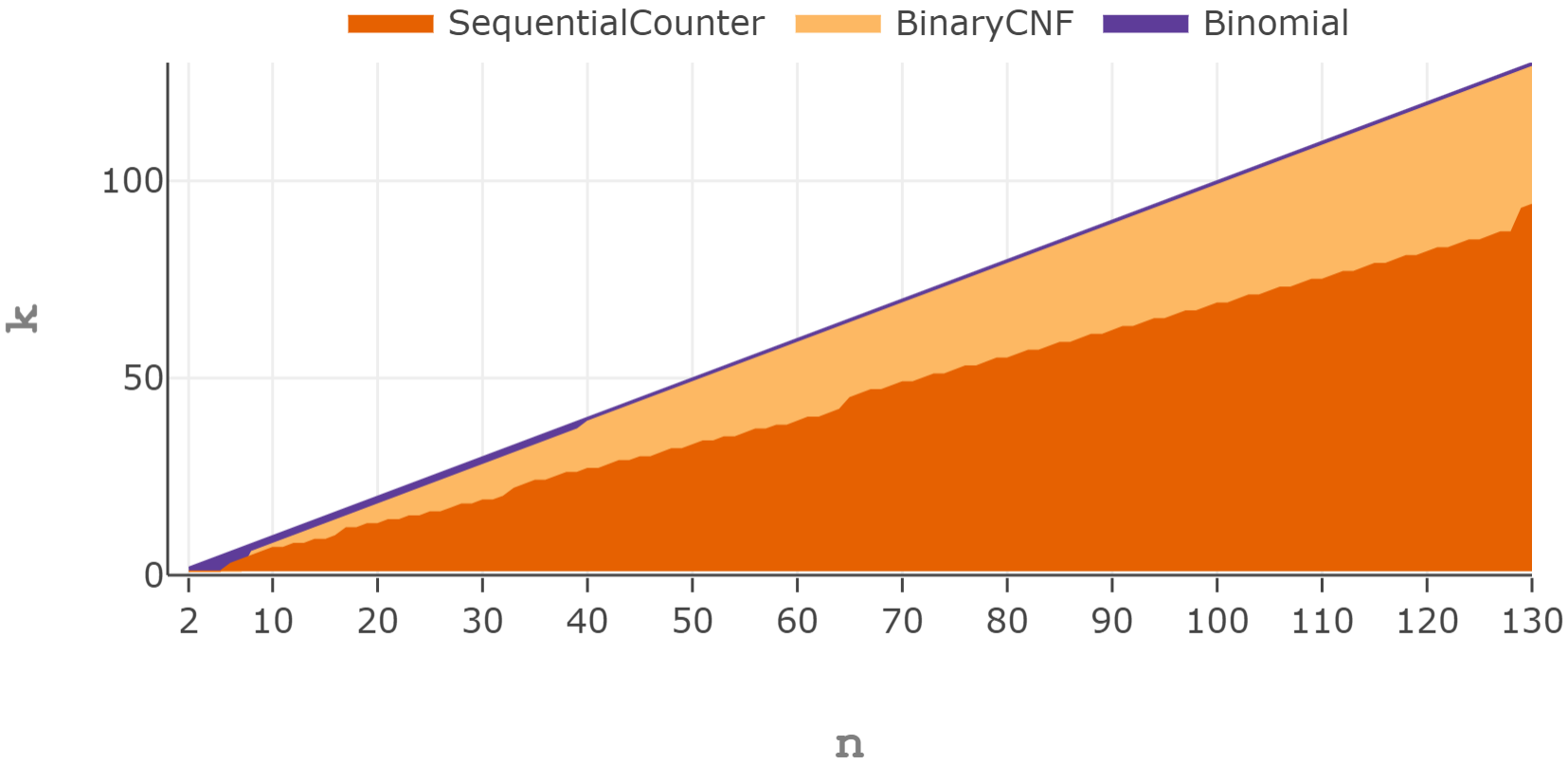
Binomial	naïve approach
Binary	
Commander	introduce new variables
Sequential Counter	

When to use which encoding?

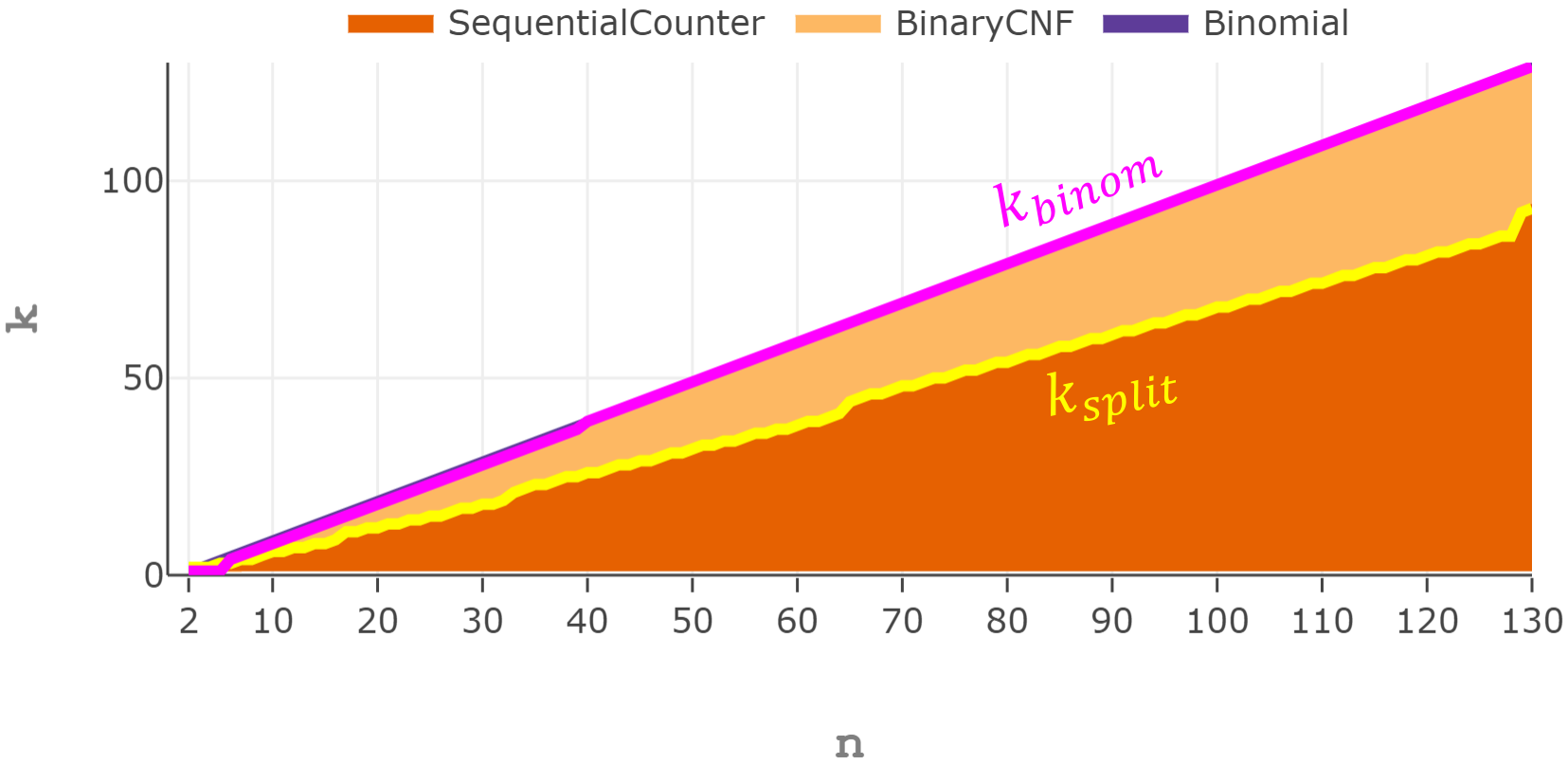
Which encoding produces the smallest formula,

i.e., lowest amount of literals for given *n* and *k*?

Encoding Producing Lowest Amount of Literals for $atmost_k$



Analytically Derive Intersections



Our Novel Meta-Encoding

$$selective(n,k) = \begin{cases} binomial & k_{binom}(n) \leq k, \\ binary & k_{split}(n) < k < k_{binom}(n), \\ seq.counter & \text{otherwise.} \end{cases}$$

Problem: Not all Subjects are Equally Awarding

CompulsoryElective 20 CP
- “Advanced IT-Security”
- “Machine Learning for IT-Security”
- “Lab on IT-Security”
- “Lab on Intelligent System Security”
- “Project Thesis IT-Security”
- “Management of Information Security”
- “Operating Systems Security”

IT-Security ⇒ choose at least 4

Problem: Not all Subjects are Equally Awarding

```
CompulsoryElective 20 CP
  “Advanced IT-Security”
  “Machine Learning for IT-Security”
  “Lab on IT-Security”
  “Lab on Intelligent System Security”
  “Project Thesis IT-Security”
  “Management of Information Security”
  “Operating Systems Security”
```

5 CP

15 CP

5 CP

IT-Security ⇒ choose at least 4

Problem: Not all Subjects are Equally Awarding

CompulsoryElective 20 CP
- “Advanced IT-Security”
- “Machine Learning for IT-Security”
- “Lab on IT-Security”
- “Lab on Intelligent System Security”
- “Project Thesis IT-Security”
- “Management of Information Security”
- “Operating Systems Security”

5 CP

15 CP

5 CP

IT-Security ⇒ choos[illegible]east 4

Problem: Not all Subjects are Equally Awarding

“Advanced IT-Security”
“Machine Learning for IT-Security”
“Lab on IT-Security”
“Lab on Intelligent System Security”
“Project Thesis IT-Security”
“Management of Information Security”
“Operating Systems Security”

Problem: Not all Subjects are Equally Awarding

“Project Thesis IT-Security”

“Advanced IT-Security”
“Machine Learning for IT-Security”
“Lab on IT-Security”
“Lab on Intelligent System Security”
“Management of Information Security”
“Operating Systems Security”

Problem: Not all Subjects are Equally Awarding

"Project Thesis IT-Security"

"Advanced IT-Security"
"Machine Learning for IT-Security"
"Lab on IT-Security"
"Lab on Intelligent System Security"
"Management of Information Security"
"Operating Systems Security"

Choose how many you want from these …

Problem: Not all Subjects are Equally Awarding

“Project Thesis IT-Security”

“Advanced IT-Security”
“Machine Learning for IT-Security”
“Lab on IT-Security”
“Lab on Intelligent System Security”
“Management of Information Security”
“Operating Systems Security”

Choose how many you want from these …

… and …

... acquire remaining CP from rest.

Problem: Not all Subjects are Equally Awarding

"Project Thesis IT-Security"

"Advanced IT-Security"
"Machine Learning for IT-Security"
"Lab on IT-Security"
"Lab on Intelligent System Security"
"Management of Information Security"
"Operating Systems Security"

$$\bigvee_{k=0}^{|BlackSheeps|} exactly_k(BlackSheeps)$$ … and …

... acquire remaining CP from rest.

Problem: Not all Subjects are Equally Awarding

"Project Thesis IT-Security"

"Advanced IT-Security"
"Machine Learning for IT-Security"
"Lab on IT-Security"
"Lab on Intelligent System Security"
"Management of Information Security"
"Operating Systems Security"

$$\bigvee_{k=0}^{|BlackSheeps|} exactly_k(BlackSheeps) \quad \wedge \quad resolve(Rest, RemainingCredits)$$

Problem: Not all Subjects are Equally Awarding

"Project Thesis IT-Security"

"Advanced IT-Security"
"Machine Learning for IT-Security"
"Lab on IT-Security"
"Lab on Intelligent System Security"
"Management of Information Security"
"Operating Systems Security"

$$\bigvee_{k=0}^{|BlackSheeps|} exactly_k(BlackSheeps) \quad \wedge \quad resolve(Rest, RemainingCredits)$$

$\rightarrow atleast,$ iff CP are equal

Problem: Not all Subjects are Equally Awarding

"Project Thesis IT-Security"

"Advanced IT-Security"
"Machine Learning for IT-Security"
"Lab on IT-Security"
"Lab on Intelligent System Security"
"Management of Information Security"
"Operating Systems Security"

$$\bigvee_{k=0}^{|BlackSheeps|} exactly_k(BlackSheeps) \quad \wedge \quad resolve(Rest, RemainingCredits)$$

- $atleast$, iff CP are equal
- recursively, else

Evaluation

Generate compulsory elective constraints of all branches:

	Literals	Variables
Binomial	1,177,570	179
Binary		
Sequential Counter		
Selective (ours)		

Feature Model Size

Evaluation

Generate compulsory elective constraints of all branches:

	Literals	Variables
Binomial	1,177,570	179
Binary	16,223	3,131
Sequential Counter	22,182	4,727
Selective (ours)		

Evaluation

Generate compulsory elective constraints of all branches:

	Literals	Variables
Binomial	1,177,570	179
Binary	16,223	3,131
Sequential Counter	22,182	4,727
Selective (ours)	**14,803**	**2,654**

Contributions

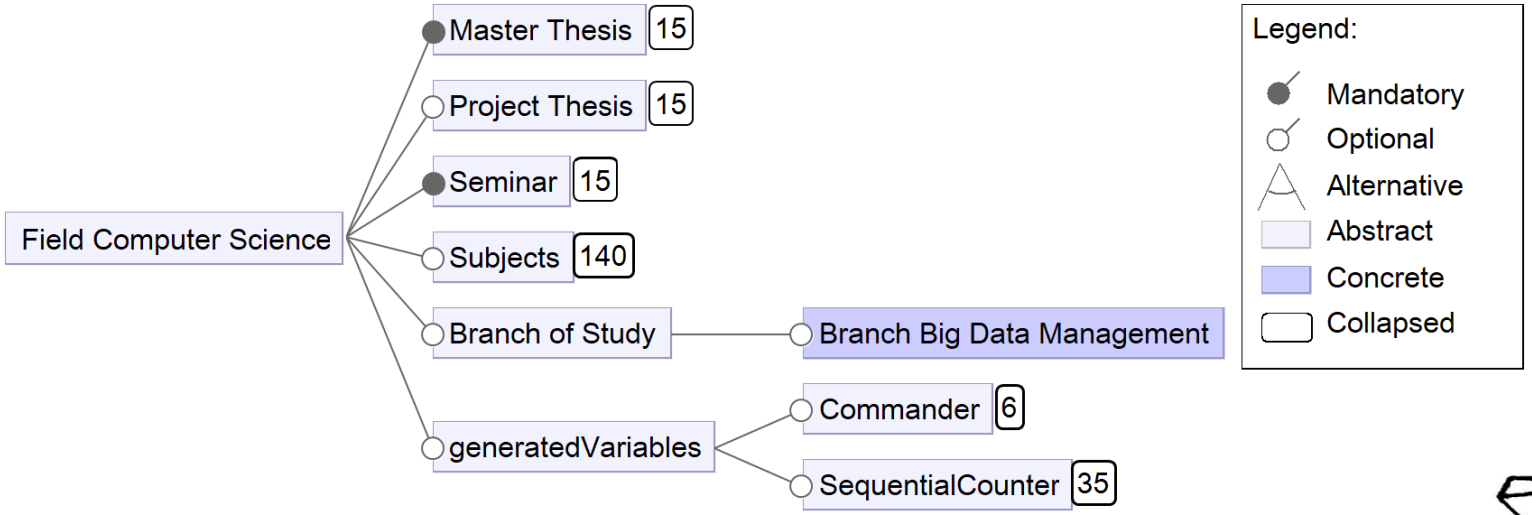
- New Meta-Encoding for At-Most-*k* constraint
- Method for generating propositional formulas requiring a sum of weighted variables to be reached
- Branch of study Tool (BroT) with DSL and compiler to feature model + source code and evaluation data at

https://github.com/PaulAtTUBS/BroT

Limitations

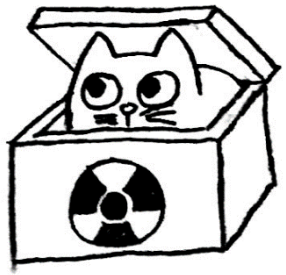
- Compulsory elective constraints not in CNF
- Only 6/9 branch models usable
- Single model with all branches still too large

Thanks for your attention!



"Branch Big Data Management" ⇒ "Relational Data Bases II" ∧ "Data Warehousing

"Branch Big Data Management" ⇒ ¬"Project Thesis Information Systems" ∧ (¬"Digit



References

(1) Alan M Frisch and Paul A Giannaros. 2010. SAT encodings of the at-most-k constraint. someold, somenew, somefast, someslow.InProc. of the Tenth Int. Workshop of Constraint Modelling and Reformulation.

(2) Will Klieber and Gihwon Kwon. 2007. Efficient CNF encoding for selecting 1 from n objects. In Proc. International Workshop on Constraints in Formal Verification.

(3) Carsten Sinz. 2005. Towards an optimal CNF encoding of boolean cardinality constraints. In International conference on principles and practice of constraint programming. Springer, 827–831.

(4) Thomas Thüm, Christian Kästner, Fabian Benduhn, Jens Meinicke, Gunter Saake, and Thomas Leich. 2014. FeatureIDE: An extensible framework for feature-oriented software development. Science of Computer Programming 79 (2014), 70–85.

(5) Olivier Bailleux and Yacine Boufkhad. 2003. Efficient CNF encoding of Boolean cardinality constraints. In International conference on principles and practice of constraint programming. Springer, 108–122.

References

(6) Benavides, D., Segura, S., Ruiz-Cortes, A.: Automated Analysis of Feature Models 20 Years Later: A Literature Review. Information Systems 35(6), 615{708 (2010)

(7) Krieter, S., Th um, T., Schulze, S., Schr oter, R., Saake, G.: Propagating Conguration Decisions with Modal Implication Graphs. pp. 898{909 (May 2018). https://doi.org/10.1145/3180155.3180159

(8) Mannion, M.: Using First-Order Logic for Product Line Model Validation. pp. 176{187 (2002) Meinicke, J., Th um, T., Schr oter, R., Benduhn, F., Leich, T., Saake, G.: Mastering Software Variability with FeatureIDE (2017)

(9) Mendonca, M.: Ecient Reasoning Techniques for Large Scale Feature Models. Ph.D. thesis, University of Waterloo, Canada (2009)

(10) Thüm, T., Apel, S., Kästner, C., Schaefer, I., Saake, G.: Analysis Strategies for Software Product Lines: A Classication and Survey. pp. 57{58. Gesellschaft für Informatik (GI), Bonn, Germany (Mar 2015)

(11) Günther, T.: Explaining Satisability Queries for Software Product Lines. Master's thesis, Braunschweig (2017). https://doi.org/10.24355/dbbs.084-201711171100, https://publikationsserver.tu-braunschweig.de/receive/dbbs mods 00065308