

Automatische Erzeugung und Bewertung von Aufgaben zu Algorithmen und Datenstrukturen

Johannes Waldmann¹

Abstract: Students' understanding of data structures and algorithms can be enhanced and tested without sidetracking them by idiosyncrasies of concrete general-purpose programming languages. In the Leipzig autotool system for E-Learning and E-Assessment, I use restricted modelling languages with uniform and trivial syntax, and problem-specific semantics. The system automatically creates instances of exercises, and grades solutions semantically.

Abstract: Ich zeige, wie das Verständnis von Algorithmen gefördert und überprüft werden kann, ohne dass Studenten durch Eigenheiten konkreter Programmiersprachen abgelenkt werden. Ich verwende stattdessen stark eingeschränkte Modellierungssprachen mit unformer und trivialer Syntax sowie aufgabenspezifischer Semantik. Aufgaben-Erzeugung sowie -Bewertung werden im E-Learning/E-Assessment-System autotool technisch realisiert, das seit ca. 2000 an der Universität Leipzig, der HTWK Leipzig, sowie gelegentlich an weiteren Hochschulen eingesetzt wird.

1 Motivation

Ein klassischer Bestandteil der Lehre von Algorithmen und Datenstrukturen [CLR09] ist das Thema „balancierte Suchbäume“, mit dem Beispiel AVL-Bäume. Der Inhalt sind Sätze über Algorithmen: Balance erzwingt logarithmische Höhe, Rotation erhält die Suchbaum-Eigenschaft und repariert die Balance lokal; sowie die zugrundeliegenden Beweismethoden (Induktion über die Höhe) sowie Programmierkonzepte (Rekursion, Iteration). Das Verständnis dieser Ideen kann dadurch unterstützt werden, dass Studenten die Algorithmen ausführen (auf dem Papier, an der Tafel) und implementieren (als Programmtext).

Ich beschreibe hier eine weitere Möglichkeit von unterstützenden Aufgaben sowie deren Automatisierung innerhalb des Systems autotool[RW02, Wa17]. Studenten sollen eine Folge s von Einfüge- und Lösch-Operationen finden, die aus einem vorgegebenen AVL-Baum t_1 einen vorgegebenen Baum t_2 erzeugt. Für die Folge s sind die Länge sowie einige Elemente vorgegeben und die restlichen zu finden. Die Vorgabe ist ein Folge mit Lücken $s?$. Die Aufgabenstellung besteht dann aus $(t_1, s?, t_2)$.

Studenten tragen einen Lösungsversuch (eine Folge s) in einem Web-Formular ein. Nach Absenden führt das autotool-System die Operationsfolge s aus und gibt die Folge der dabei erzeugten Bäume aus. Schließlich wird das Endresultat mit der Vorgabe verglichen. Eine Lösung ist jede Folge s , die zu $s?$ passt und t_1 in t_2 überführt. Korrektheit ist also

¹ HTWK Leipzig, Fakultät IMN, Postfach 301166, 04251 Leipzig, <https://www.imn.htwk-leipzig.de/~waldmann/>

nicht definiert als Übereinstimmung mit einer Musterlösung, sondern als das Erfüllen der Spezifikation. Das hat auch den Vorteil, dass die Ablehnung einer falschen Lösung nachvollziehbar begründet werden kann, aber in der Begründung keine richtige Lösung verraten wird. Deswegen ist es unschädlich, und im Gegenteil sogar erwünscht, dass Studenten mehrere Lösungsversuche ausführen.

Der abschließende Test auf Korrektheit ist gar nicht der wesentliche Nutzen dieser Aufgabe. Vielmehr geht es um die ausführliche Beschreibung der Wirkung der einzelnen Operationen in der Ausgabe. Dabei können die Studenten sehen, wie die korrekte Implementierung einzelner Operationen wirkt. Die Studenten können dabei auch selbständig und frei experimentieren und z.B. beobachten, was passiert, wenn im nächsten Schritt die Wurzel gelöscht wird u.ä., indem sie eine geeignete Folge s eingeben und absenden. Diese löst dann zwar die vorgegebene Aufgabe nicht, aber die Anzahl der Fehlversuche spielt bei der Bewertung keine Rolle.

Für Aufgaben des beschriebenen Typs können automatisch Instanzen erzeugt werden: das System würfelt einen Baum t_1 sowie eine Operationsfolge s_0 und bestimmt daraus eine Folge $s_?$ durch Ersetzen von Elementen an zufällig gewählten Positionen durch Markierungen ?. Der Aufgabensteller konfigurierte den Würfel durch Vorgabe der Größe der Größe von t_1 , der Länge von $s_?$ sowie der Anzahl der Fragezeichen in $s_?$.

Durch den Würfel erhält jeder Student eine andere Aufgaben-Instanz, damit sind Plagiate von Lösungen unmöglich.

Alle hier beschriebenen Aufgaben können unter dieser Adresse ausprobiert werden: <https://autotool.imn.htwk-leipzig.de/cgi-bin/Trial.cgi?lecture=240>.

Ich danke Herrn Mario Wenzel sowie den anonymen Gutachtern für Diskussionen und Hinweise zu der Vorabversion dieses Berichtes.

2 Aufgaben zu Suchbäumen

Suchbäume $S(K)$ mit Schlüsseln einer Menge K realisieren den abstrakten Datentyp Menge mit folgender Schnittstelle. Die Semantik $\llbracket t \rrbracket$ eines Baumes ist die Menge seiner Schlüssel.

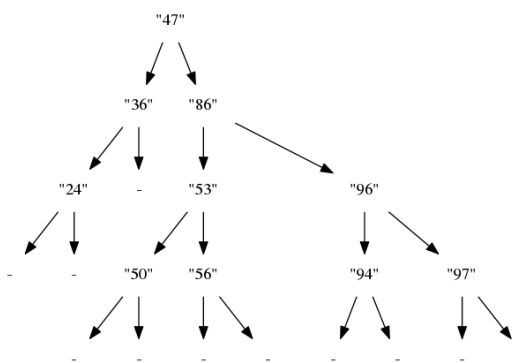
- $\text{empty} : S(K)$ mit $\llbracket \text{empty} \rrbracket = \emptyset$,
- $\text{null} : S(K) \rightarrow \mathbb{B}$ mit $\llbracket \text{null}(t) \rrbracket = (\emptyset = \llbracket t \rrbracket)$
- $\text{contains} : K \times S(K) \rightarrow \mathbb{B}$ mit $\llbracket \text{contains}(x, t) \rrbracket = (x \in \llbracket t \rrbracket)$
- $\text{insert} : K \times S(K) \rightarrow S(K)$ mit $\llbracket \text{insert}(x, t) \rrbracket = \{x\} \cup \llbracket t \rrbracket$
- $\text{delete} : K \times S(K) \rightarrow S(K)$ mit $\llbracket \text{delete}(x, t) \rrbracket = \llbracket t \rrbracket \setminus \{x\}$

Für die Aufgaben verwenden wir Zahlen als Schlüssel. Jeder implementierende Typ $S(K)$ besitzt eine Funktion zur grafischen Darstellung. Wir verwenden `dot` [Gal17].

Derzeit existieren diese Implementierungen: nicht balancierte binäre Bäume, AVL-Bäume, Rot-Schwarz-Bäume, B-Bäume.

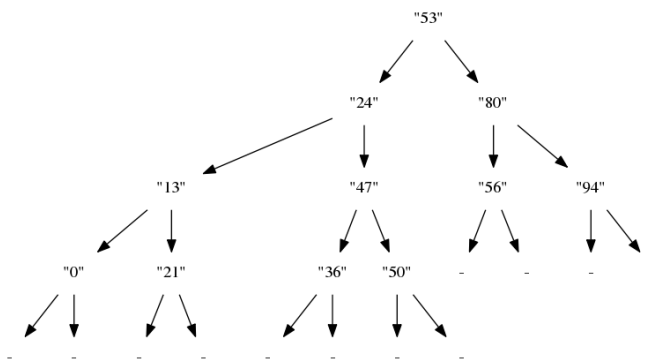
Ein Beispiel für eine Aufgabenstellung für AVL-Bäume ist:

Auf den Baum



sollen diese Operationen angewendet werden
(wobei Sie Any geeignet ersetzen sollen):

[Any , Any , Any , Delete 86 , Any , Any , Insert 13 , Insert 21]
so dass dieser Baum entsteht:



Eine korrekte Lösung ist

[Insert 53, Insert 80, Delete 97, Delete 86
, Delete 96, Insert 0, Insert 13, Insert 21]

Für die Erzeugung zufälliger Aufgabeninstanzen werden, wie in der Einleitung beschrieben, t_1 und s_0 ausgewürfelt und daraus t_2 bestimmt. Damit interessante, d.h., nicht trivial lösbare, Instanzen entstehen, werden jeweils mehrere Kandidaten-Instanzen bestimmt und dann diejenige verwendet, die den Hamming- oder Editier-Abstand von t_1 zu t_2 maximiert. Die Wurzeln von t_1 und t_2 sollen sich unterscheiden und die Wurzeln der nächsten Teilbäume möglichst auch, usw., damit die Aufgabe nicht durch Betrachtung unabhängiger Teilbäume einfach lösbar ist. Unterschiedliche Wurzeln erreicht man bei nicht balancierten Bäumen durch delete-Operationen, bei anderen auch als Folge von Balance-Reparaturen.

3 Aufgaben zu heap-geordneten Bäumen

Heaps $H(K)$ mit Schlüsseln (Prioritäten) einer total geordneten Menge K realisieren den abstrakten Datentyp Prioritätswarteschlange unter anderem den folgenden Operationen. Die Semantik $\llbracket h \rrbracket$ eines Heaps h ist die (Multi)menge der Schlüssel. Die insert-Operation liefert einen Verweis $p \in \text{Ref}(K)$, der bei decrease verwendet wird. In der grafischen Darstellung werden für jeden Knoten der Schlüssel sowie der Name des Verweises auf diesen Knoten gezeichnet.

Derzeit wird die Schnittstelle implementiert durch fast vollständig balancierte binäre Heaps. Dabei sind alle Blätter in Tiefe h oder, rechts davon, in Tiefe $h - 1$. Solche Heaps können implizit durch Arrays repräsentiert werden, was in dieser Aufgabe jedoch keine Rolle spielt.

Grundsätzlich treffen Bemerkungen zu Suchbäumen aus Abschnitt 2 zu. Es zeigte sich, dass Aufgaben zu Heaps schwieriger waren, erkennbar an einer höheren Anzahl von Fehlversuchen. Das kann daran liegen, dass die decrease-Operation eine direkte Zuordnung zwischen den Schlüsseln in der Eingabe t_1 und denen in der Ausgabe t_2 erschwert.

4 Aufgaben zu Hashtabellen

Eine Hashtabelle $T(K)$ repräsentiert den abstrakten Datentyp Menge, die Implementierung verwendet direkten Zugriff mittels Hashfunktionen. Gegenüber den vorher beschriebenen Aufgaben zu Bäumen enthalten Aufgaben zu Hashing nicht nur eine Beschreibung der vorgegebenen Start- und Endzustände, hier: Belegung der Hashtabelle, sondern zusätzlich eine Spezifikation des verwendeten Hash-Verfahrens sowie seiner Parameter, insbesondere seiner Hashfunktion(en). Form und Koeffizienten von Hashfunktionen können teilweise versteckt werden. Damit wird die Angabe der Hashfunktion zu einem Teil der Aufgabe.

Derzeit gibt es Aufgaben zu diesen Hashverfahren:

- Hashing mit Verkettung: jede Zelle der Hashtabelle enthält eine Liste von Elementen.
- offenes Hashing mit linearem Sondieren: jede Zelle enthält kein oder ein Element, die Sondierungsreihenfolge für x mit $\text{hash}(x) = p$ ist $p, p + 1, p + 2, \dots$ (modulo Tabellengröße).

- doppeltes Hashing: jede Zelle enthält kein oder ein Element, die Sondierungsreihenfolge für x mit $\text{hash}_1(x) = p$ und $\text{hash}_2(x) = q$ ist $p, p + q, p + 2q, \dots$

Die Hashfunktion kann vorgegeben werden als lineare Funktion mit unbekanntem Koeffizienten $h(x) = ? + ? \cdot x$. Bei Hashing mit Verkettung sind die Koeffizienten leicht zu rekonstruieren, da man Paare $(x, h(x))$ direkt aus der Tabelle ablesen kann. Bei offenem Hashing ist das schwieriger, da nicht klar ist, nach wieviel Sondierungs-Schritten ein Element eingefügt wurde. Das ist gut so, denn dann wird genau das beim Bearbeiten der Aufgabe geübt. Eine Beispiel-Aufgabenstellung für doppeltes Hashing ist

```
Replace each hole ( _ ) in the configuration
Config { size = 10 , hash1 = \ x -> _ + _ * x , hash2 = \ x -> _ }
and in the sequence of operations
[ Insert _ , Insert _ , Insert _
, Insert 40 , Insert _ , Insert _ , Insert 22 ]
with a numerical expression
such that the sequence of operations transforms
Table (listToFM [ (2,12),(9,41)])
to Table (listToFM
[ (0,11),(1,25),(2,12),(3,89),(4,22),(6,40),(7,97),(8,64),(9,41) ])
```

Eine Lösung ist

```
Solution { config = Config { size = 10 , hash1 = \ x -> 6 + 3 * x
, hash2 = \ x -> 1 }
, ops = [ Insert 25 , Insert 11 , Insert 89 , Insert 40
, Insert 97 , Insert 64 , Insert 22 ] }
```

Ein Ausschnitt aus der Antwort des System ist

```
Table (listToFM [ ( 1, 25) , ( 2, 12) , ( 9, 41) ])
execute Insert 11
insert(x,S)    x = 11
hash1(x) = 9 (mod 10)    hash2(x) = 1 (mod 10)
probe at address 9      contains 41
probe at address 0      is empty, insert x here
result: Table (listToFM [ ( 0, 11) , ( 1, 25) , ( 2, 12) , ( 9, 41) ])
```

Die Aufgaben zum doppelten Hashing erwiesen sich beim Einsatz im Sommersemester 2017 tatsächlich als schwierig, mit einer hohen Anzahl von Fehlversuchen.

5 Algorithmen auf Graphen

Die Breitensuche wird durch folgenden Aufgabentyp getestet: eine Aufgaben-Instanz besteht aus einem Muster G_M für die Adjazenz-Listen-Darstellung eines Graphen, einem Startknoten s und einem Muster T_M für einen Baum. Eine Lösung ist eine Adjazenz-Listen-Darstellung G eines Graphen. Die Lösung ist korrekt, wenn G zu G_M paßt und der Breitensuchbaum $\text{BFS}(G)$, der von autotool berechnet wird, zu T_M paßt.

Eine Beispiel-Aufgabenstellung ist

```
Replace the holes (_) in the adjacency list of the graph G
Adjacency_List [ ( _, [ _, 4, 7] ), ( _, [ 5, _] ), ( 4, [ _, _] )
, ( _, [ _, _, 4, 5] ), ( _, [ 4, _, 1, 3] ), ( 3, _ ), ( 6, [ _, 1, _] ) ]
so that the breadth first search tree of G, starting at 3,
matches the pattern
Branch 3 [ Branch _ [ Branch 6 _
, Branch 5 [ Branch 2 [ ] , Branch _ [ ] ] ] ]
```

Die Konfiguration des Aufgabengenerators ist

```
Config
{ graph_generator = Graph_Generator
{ directed = True, vertices = 7, out_degree_range = ( 1, 4 ) }
, handle_graph = Show 0.7, handle_bfs_tree = Show 0.7, candidates = 100 }
```

Es werden 100 Graphen G erzeugt und für jeden $T = \text{BFS}(G)$ bestimmt. In G und T werden zufällig ausgewählte Teilstrukturen versteckt, d.h., durch $_$ ersetzt, so dass der Anteil der sichtbaren Funktionssymbole möglichst nahe bei 70 Prozent liegt.

Eine ähnliche Aufgabe behandelt Tiefensuch-Wälder.

6 Technische Details

Das System autotool besteht aus einem zustandslosen *Semantik-Service* zur Generierung von Aufgabeninstanzen und zur Bewertung von Lösungsversuchen, implementiert als XML-RPC-Service mit haxr, einer *Persistenz-Schicht* zur Speicherung von Aufgaben, Einschreibungen, Punkten u.ä., implementiert als mysql (mariadb) mit persistent [Sn17] einer *Web-Oberfläche* [Si16], implementiert mit yesod [Sn12]. Damit können Studenten Aufgaben bearbeiten und Bewertungen betrachten sowie Dozenten die Aufgaben konfigurieren und testen sowie Aufgaben und Punkte verwalten.

Die hier beschriebenen Aufgaben sind als Module im Semantik-Service realisiert. Die gemeinsame Modulschnittstelle ist durch diese mehrparametrische Relation (Typklasse) für p (Aufgabentyp), i (Typ der Instanz), b (Typ der Einsendung) beschrieben:

- `describe`: $p \times i \rightarrow \text{Text}$ beschreibt Aufgabenstellung,
- `initial`: $p \times i \rightarrow b$ berechnet einen syntaktisch korrekten, aber semantisch falschen Lösungsvorschlag,
- `eval`: $p \times i \times b \rightarrow \mathbb{B} \times \text{Text}$ bewertet die Einsendung, liefert Resultat richtig/falsch und Begründung

Das Würfeln von Aufgabeninstanzen nach einer Konfiguration c wird beschrieben durch

- `generate`: $p \times c \times \text{Seed} \rightarrow i$

wobei `Seed` aus der Matrikelnummer (oder einem anderen Kennzeichen des Studenten) bestimmt wird und der Initialisierung eines deterministischen Pseudozufallsgenerators dient.

Der Student gibt eine textuelle Repräsentation der Lösung (vom Typ b) ein. Das Texteingabefeld wird mit `initial(p, i)` initialisiert. Mit Absicht gibt es keine grafische Eingabemöglichkeit, um Studenten an die wissenschaftliche Notation (durch Terme in einer geeigneten Signatur) zu gewöhnen. Bei den allermeisten Aufgaben ist die konkrete Syntax für b die Daten-Syntax von Haskell [Sim10], mit den Ausdrucksmitteln für primitive Daten: `123`, `False`, für Tupel: `(124, False)`, für (typreine) Listen: `[(123, False), (456, True)]`, Konstruktoren algebraischer Datentypen, mit Argumenten: `Just (123, False)`, Records mit benannten Komponenten: `Config{size=10}`. Pretty-Printer und Parser werden automatisch generiert. Gelegentlich wird eine aufgabenspezifische konkrete Syntax verwendet, z.B. für arithmetische Ausdrücke in Hashfunktionen.

Die Eingabe-Parser [LM16] erzeugen nützliche Fehlermeldungen. Die fehlerhafte Stelle wird markiert und mögliche Fortsetzungen werden ausgegeben. Jedes Eingabefeld (für den Studenten: das für den Lösungsversuch, für den Dozenten: das für die Konfiguration der Instanz oder des Instanzengenerators) ist typisiert und wird automatisch mit einem Verweis auf die API-Dokumentation des jeweiligen Typs versehen.

Die hier gezeigten Aufgabe verwenden das Prinzip, dass Teile einer Datenstruktur (Bsp: Koeffizienten der Hashfunktion, Elemente der Liste von Operationen) in der Aufgabenstellung versteckt werden und zur Lösung zu ergänzen sind. Ein Typ p dient in diesem Sinne als Vorlagentyp, wenn für ihn diese Eigenschaften deklariert werden:

- der zugehörige Basistyp: `type (Base p)`
- der Mustervergleich `match`: $p \times \text{Base}(p) \rightarrow \mathbb{B} \times \text{Text}$
- die Injektion `inject`: $\text{Base}(p) \rightarrow p$
- das pseudozufällige Verstecken von Bestandteilen `obfuscate`: $\text{Base}(p) \times \text{Seed} \rightarrow p$.

7 Diskussion

Der autotool-Semantik-Service wird ebenfalls vom autolat-Plugin [FS10] für OpenO-LAT [Fr17] unterstützt und an der Universität Leipzig angewendet. Ich möchte autotool gern in weitere Systeme integrieren. Eine Verwendung der QTI-Schnittstelle [IM15] ist naheliegend, scheint aber derzeit nicht möglich, da im Standard weder das externe Generieren von Aufgaben noch das externe Bewerten von Einsendungen vorgesehen sind, sondern höchsten Zeichenkettenvergleiche, die von QTI-Player selbst ausgeführt werden. Das ist für eine semantische Bewertung mit sinnvoller Rückmeldung völlig unzureichend.

Die hier beschriebenen Aufgaben habe ich für die Vorlesung Algorithmen und Datenstrukturen für Informatik-Bachelor-Studenten im 2. Semester, die ich vertretungsweise übernommen habe, überarbeitet bzw. neu entwickelt. Dabei haben ca. 80 Studenten über das Sommersemester 2017 hinweg jede Woche 1 bis 2 autotool-Aufgaben bearbeitet, mit ca. 10 Einsendungen pro Aufgabe. Für einzelne Aufgaben habe ich deutlich mehr Einsendungen registriert. Das kann an der Schwierigkeit der Aufgabe liegen, aber auch daran, dass ich explizit zum Durchprobieren verschiedener, auch “falscher” Einsendungen ermuntert hatte.

Ich habe damit die eigentlich vorgesehenen Java-Programmier-Übungen ersetzt. Die autotool-Aufgaben schienen mir dem didaktischen Ziel des Verständnisses der Algorithmen nützlicher. Der Aufwand zum Lösen und Korrigieren von Java-Programmen ist für Student und Dozent deutlich höher, so dass Programmieraufgaben üblicherweise nur dreimal während des Semesters gestellt wurden. Tatsächlich habe ich wohl die Arbeit verschoben: vom (z.T. händischen) Korrigieren der Java-Aufgaben zum Programmieren der autotool-Aufgaben — allerdings in Haskell, das sehr konzise Quelltexte ermöglicht und sehr hohe statische Sicherheit garantiert und somit den Entwicklungsaufwand verringert.

Die Studenten haben in dieser Vorlesung keine Programme selbst geschrieben und (außer Pseudocode im Skript und autotool-Quelltexten) kaum welche gesehen — das finde ich nicht so tragisch, zumal es parallel eine Vorlesung zur anwendungsorientierten Java-Programmierung gibt und vorher eine zu Grundlagen der Programmierung.

Durch die autotool-Aufgaben haben sich die Studenten auch ohne Programmieren mit den Algorithmen beschäftigt, indem sie ihre Ausführung durch das autotool beobachtet und durch geeignete Eingaben zielgerichtet gesteuert haben. Für die Übungen habe ich damit Zeit gewonnen für die Diskussion von Aufgaben zu Eigenschaften von Algorithmen. (Die Prüfungszulassung ergibt sich aus beiden Anteilen.) Eine wissenschaftliche fundierte Einschätzung des tatsächlichen Nutzens (und Aufwandes) von autotool-Aufgaben scheint mir hier nicht möglich, da zu viele weitere Parameter variieren: nicht nur Änderung der Übungs-Aufgaben, sondern auch im Vortragsstil, im Skript, in Klausuraufgaben.

Berichten von Kollegen sowie eigene Erfahrung mit autotool in anderen Vorlesungen [Wa15, Wa14] bestätigen, dass Studenten das System schnell verstehen und gern benutzen und besonders die sofortigen, ausführlichen und objektiven Antworten begrüßen.

Literaturverzeichnis

- [CLR09] Cormen, Thomas H.; Leiserson, Charles E.; Rivest, Ronald: Introduction to Algorithms. MIT Press, 2009.
- [Fr17] Frentix GmbH: , LMS OpenOLAT. <https://www.openolat.com/>, 2017.
- [FS10] Felgenhauer, Bertram; Schöhlhorn, Klemens: , Autolat: autotool integration for OpenOLAT. <https://github.com/klemens/openolat/wiki>, 2010.
- [Ga17] Gansner, Emden et al.: , Graphviz - Graph Visualization Software. <http://www.graphviz.org/>, 2017.
- [IM15] IMS Global Learning Consortium: , IMS Question and Test Interoperability v2.2 Final Specification. <https://www.imsglobal.org/question/index.html>, 2015.
- [LM16] Leijen, Daan; Martini, Paolo: , parsec: Monadic Parser Combinators. <https://hackage.haskell.org/package/parsec>, 2016.
- [RW02] Rahn, Mirko; Waldmann, Johannes: The Leipzig autotool System for Grading Student Homework. In: Intl. Workshop on Functional and Declarative Programming in Education (FDPE 2002). Technical Report 0210, University of Kiel, 2002.
- [Si16] Siegburg, Marcellus: REST-orientierte Refaktorisierung des E-Learning-Systems Autotool. Masterarbeit, HTWK Leipzig, 2016. <http://www-ps.informatik.uni-kiel.de/~msi/thesis/thesis.pdf>.
- [Sim10] Haskell 2010 Language Report. <https://www.haskell.org/onlinereport/haskell2010/>, 2010.
- [Sn12] Snoyman, Michael: , Developing Web Applications with Haskell and Yesod, 2012.
- [Sn17] Snoyman, Michael: , persistent: Type-safe, multi-backend data serialization. <https://hackage.haskell.org/package/persistent>, 2017.
- [Wa14] Waldmann, Johannes: Automated Exercises for Constraint Programming. In: 28th Workshop on (Constraint) Logic Programming (WLP 2014), Wittenberg, 2014. <http://www.imn.htwk-leipzig.de/~waldmann/talk/14/wlp/auto/>.
- [Wa15] Waldmann, Johannes: Automatisierte Bewertung und Erzeugung von Übungsaufgaben zu Prinzipien von Programmiersprachen. In: 18. Kolloquium Programmiersprachen und Grundlagen der Programmierung KPS 2015, Pörschach, 2015. <http://www.imn.htwk-leipzig.de/~waldmann/talk/15/kps/paper/>.
- [Wa17] Waldmann, Johannes: , Leipzig autotool. <https://gitlab.imn.htwk-leipzig.de/autotool/all0>, 2017.