„Untersuchung zu den Begriffen Gültigkeit und Beweisbarkeit in einer mathematischen Theorie"

Reto Hablützel, Max Schrimpf

Inhaltsverzeichnis

1 Einleitung

1.1 Aufgabenstellung

> Es soll augehend von einem gegebenen einfachen Axiomensystem (z.B "Gruppe") ein einfacher Satz einereits bewiesen werden und anderseits seine "Gültigkeit" illustriert werden. Ein Beweis soll entlang von gegebenen formalisierten Schlussregeln erfolgen, von der Gültigkeit soll man sich mittels "durchrechnen" überzeugen, bei unendlichen Strukturen mit einer zufallsgenerierten Auswahl von Fällen.

Ausgehend von dieser vorgegebenen Aufgabenstellung soll in der vorliegenden Projektarbeit eine Implementation mit der Fähigkeit die Gültigkeit eines Axiomsystemes in den Grundsätzen zu bewisen, erarbeitet werden. Durch das enorme Wachstum der Ausgangswerte sind dabei schon relativ kleine Systeme mit bedeutendem Rechenaufwand verbunden, weshalb ein sinnvoller Ansatz zum Umhang mit rechenintensiven Aufgaben gefunden werden muss.

Die genaue Schwerpunktsetzung bei der Umsetzung und das Vorgehen war dabei durch die offene Aufgabenstellung dem Projektteam überlassen. Vorgegeben war mur die Implementation eines automatisierten Gültigkeitsbeweises des Satzes zur Rechtsauslöschung in Gruppen.

1.2 Ausgangslage

Wir konnten uns im letzten Semester als Vertiefungsarbeit im Fach „Algorithmen und Datenstrukturen" mit dem Thema MapReduce beschäftigen. Dazu hatten wir, anhand des Entwrufsmuster (1.2.1 MapReduce), unser eigenes Framework für parallelisierte MapReduce Berechnungen implementiert.

Bei der Analyse der vorliegenden Problemstellung kam uns schnell die Idee , dass es möglich wäre die im letzten Semester gewonnenen Erkenntnisse sowie die bestehende Software zur Umsetzung einer Lösung zu Nutzen.

Aufgrund der beschränkt zur Verfügung stehenden Zeit, mussten wir es jedoch bei der damaligen Ausführung auf einem einzelnen Rechner mit mehreren Threads belassen, anstatt ein ganzes Computernetz verwenden zu können. Ausserdem ergaben sich weitere technische Limitierungen. So konnte unser Framework zum Beispiel nicht mit dem Szenario umgehen, dass ein Thread abnormal terminiert. Die Handhabung solcher Ereignisse ist, insbesondere im Ausblick auf eine echt verteilte Berechnung interessant, da in einem Netzwerk immer wieder eine einzelne Maschine aussteigt.

1.2.1 MapReduce

MapReduce ist ein Entwurfsmuster, welches die parallele Verarbeitung grosser Datenmengen vereinfachen soll und erstmals von Google beschrieben [DG04] sowie in einem Framework implementiert wurde. Google ist mit dem Problem grosser Datenmengen zum Beispiel beim erstellen der Google Suche konfrontiert, hierbei muss quasi das gesammte Internet indexiert werden.

Um ein Programm so zu entwickeln, dass es auf einer verteilten Architektur gut skaliert, ist jedoch ein enormer Aufwand, welcher nichts mit dem eigentlich zu lösenden Problem zu tun hat, nötig. Daher wurde das Framework so konzipiert, dass es Benutzern erlaubt ihr Problem in zwei nur Phasen zu definieren

ohne sich Gedanken um die anschliessende Verteilung und Berechnung machen zu müssen. So kümmert sich das Framework zum Beispiel darum, was passiert, wenn eine Maschine im ausfällt oder wie die einzelnen Arbeittsschritte auf die verschiedenen Maschinen aufgeteilt werden.

Die zwei Phasen des Entwurfsmusters sind, wie der Name berieits suggeriert, Map und Reduce (mit einer eine optionalen, zwischenliegenden, für das Verständnis irrelevanten, Combiner Phase). In diesen Phasen werden im Wesentlichen die beiden Funktionen Map und Reduce aufgerufen, welche hier kurz erklärt werden sollen.

$$map^1 : (a \to b) \to A \to B$$

Die Map Funktion bildet eine Menge von Elementen elementweise auf eine neue Menge ab. Dazu muss ihr zwei Parameter übergeben werden – eine Funktion, welche eine Element aus der Menge A auf ein Element in der Menge B abbildet und eine Menge von Elementen A. Daraufhin wird die Funktion sukzessive auf die Elemente der Liste A angewendet und somit die Menge B konstruiert, welche das Resultat ist.

$$reduce : (b \to c \to c) \to c \to B \to c$$

Die Reduce Funktion bildet eine Menge B auf einen einzelnen Wert ab c. Dafür erhält sie als ersten Parameter eine Funktion, mit einen Wert aus der Menge B sowie einen Akkumulator, welche einen Wert b zurückgibt. Als zweiten Parameter erhält die Reduce Funktion einen initialen Wert für den Akkumulator und als dritten und letzten Parameter die Menge B. Dann wird die übergebene Funktion mit jedem Element aus der Menge B und dem Akkumulator aufgerufen. Dabei ist das Resulat der Funktion jeweils der neue Wert für den Akkumulator. Sobald die ganze Liste traversiert wurde, ist der Wert vom Akkumulator das Resultat.

Der Benutzer des MapReduce Frameworks muss also für eine auszuführende Berechnung nur noch eine Funktion schrieben, welche ein Element aus der Menge A auf ein Element in der Menge B abbildet (Map Phase), eine Funktion schreiben, welche für ein Element aus der Menge B, kombiniert mit einem Akkumulator, auf ein Element aus der Menge C abbildet (Reduce Phase) und initial die Menge$A$ bereitstellen.

Für das MapReduce Framework ist es daraufhin möglich zum Beispiel die erste Funktion mit einem Element aus der Menge A an eine berechnende Maschine senden. Auch ordnet das Framework in mit einem sogenannten Shuffle nach der Map Phase die Elemente so an, dass sie von der Reduce Phase verarbeitet werden können. Die Anordnung der Elemente der Menge B muss dabei zu beginn der Reduce Phase nicht zwingend ihrer Anordnung nach der Map Phase entsprechen.

[1] Schreibweise: a ist ein Element aus einer Menge A

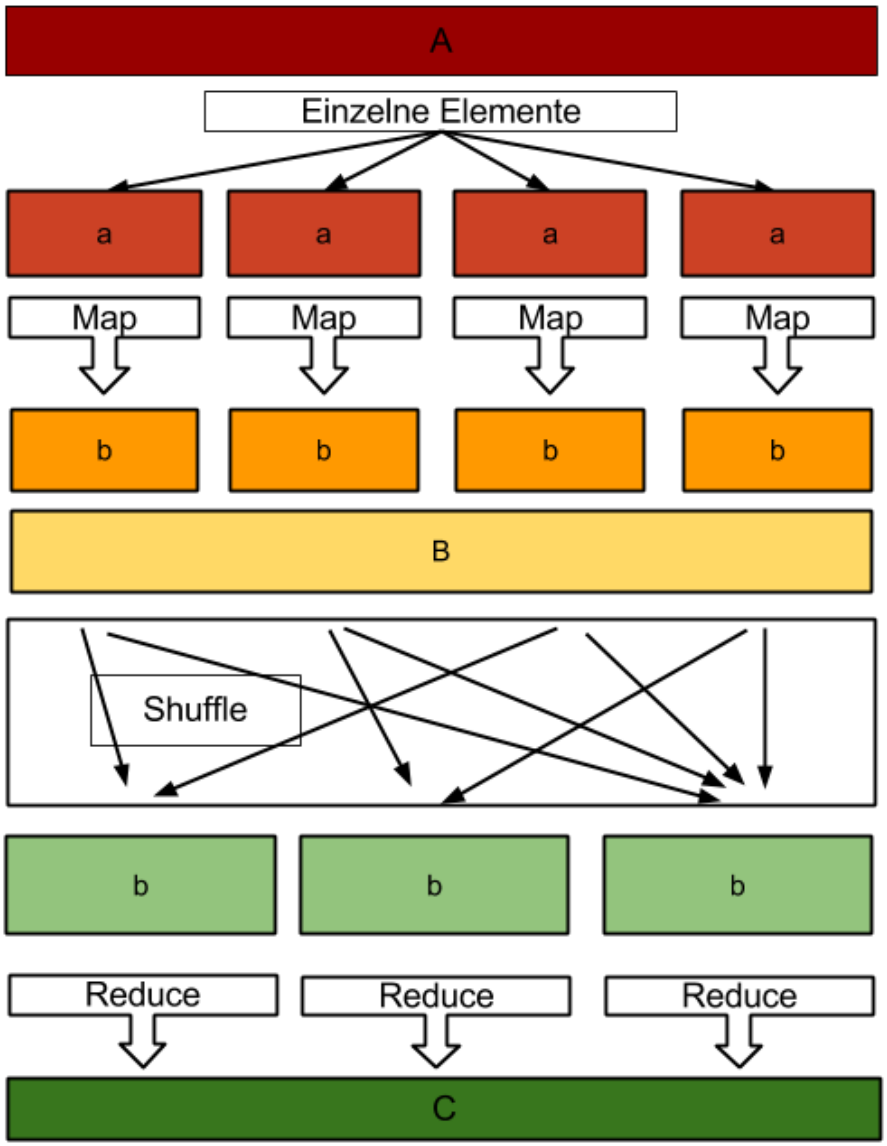

Abbildung 1: Übersicht über die Phasen des MapReduce

Wie die Grafik illustriert, gibt es in der Map Phase keine Datenabhängigkeiten der einzelnen Elemente (1:1 Mapping), weshalb jede Funktionsanwendung sehr gut auf verschiedenen Maschinen ausgeführt werden kann. Die Map Phase gilt als abgeschlossen, wenn sämtliche Elemente aus der Menge A auf ein Element in der Menge B abgebildet wurden. Nun kann das MapReduce Framework auf einer Maschine die Reduce-Phase durchführen. Hierbei ist wichtig zu bemerken, dass die Reduce Funktion nicht äquivalent parallelisiert werden kann, da es Datenabhängigkeiten gibt. Eine parallele Verarbeitung ist nur möglich, wenn die Reduce Funktion assoziativ ist. In diesem Fall könnte die Reduce-Phase in einer Art Baum-Struktur ausgeführt werden – auf solche für das verständ-

nis nicht förderlichen Details soll jedoch im folgenden nicht weiter eingegangen werden.

Hier ein Beispiel umd die Idee der zwei Phasen des MapReduce zu verdeutlichen. Wir definieren ein initiale Menge an Elementen

$$A = \{1, 2, 3, 4\}$$

eine Funktion für die Map Phase

$$f_m\, x = x^2$$

und eine Funktion für die Reduce Phase

$$f_r\, x\, a = x + a.$$

Nach der Map Phase entsteht die Menge

$$B = \{1, 4, 9, 16\}.$$

Welche in der der Reduce Phase aufaddiert[2] wird und als Endresultat die einelementige Menge

$$C = \{30\}$$

ergibt

1.3 Abgeleitete Aufgabenstellung

Basierend auf unserem Wissen aus den Algebra-Vorlesungen, unserem Interesse für verteilte Berechnungen und dem bestehenden MapReduce Framework, einigten wir uns mit Herrn Heuberger dahingehend, dass wir in einem ersten Teil die Prüfung eines Satzes für ein bestimmtes Axiomsystem als MapReduce Problem formulieren und dieses dann mit unserem Framework ausführen sollten.

Anschliessend daran wurde uns die Option gewährt selbst eine vertiefende Weiterführung der Projektarbeit mit den in Teil 1 erworbenen Kenntnissen wählen. Hierzu mehr im Kapitel „Möglichkeiten der Vertiefung der Aufgabenstellung".

[2] Der initiale Akkumulator ist in diesem Fall das erste Element aus der Menge B oder 0.

2 Teil 1 Gültigkeit des Satzes zur Rechtsauslöschung in Gruppen

2.1 Mathematische Grundlagen

2.1.1 Gruppe

Eine Gruppe ist eine algebraische Struktur bestehend aus einer Menge S und einer Verknüpfung $\diamond$, welche die folgenden Axiome erfüllt:

- Assoziativität: $\forall a, b, c.\ (a \diamond b) \diamond c\ = a \diamond (b \diamond c)$
- Neutrales Element: $\exists e.\ \forall a.\ (a \diamond e = e \diamond a = a)$
- Inverses Element[3]: $\forall a.\ \exists b.\ (a \diamond b = e)$

Es gibt verschiedene Strukturen, die diese Eigenschaften haben. Ein Beispiel davon ist die Restklasse 3 mit einer Operation definiert als: $x \diamond y = (x + y)\,\%\,3$, also die gewöhnliche Addition in dieser Restklasse. Für diese Struktur kann man nun beweisen, dass sie eine Gruppe ist, indem man sich überzeugt, dass die Addition assoziativ ist, dass ein neutrales Element existiert (nämlich 0) und dass jedes Element ein inverses Element hat, mit welchem es auf die 0 abbildet (z.B. $1 \diamond 2 = 0$).

Es gibt weit mehr Strukturen, die als Gruppe eingestuft werden können. Es ist zum Beispiel durch stupides ausprobieren möglich diese zu finden. Diese mechanische Arbeit eignet sich bestens für einen Computer, da sie strikt Regelgesteuert abläuft. Genau dies ist der erste Schritt unserer Arbeit. Natürlich gibt es unendlich viele Möglichkeiten für mögliche Gruppen. Daher ist es klar, dass wir unsere Berechnunfen auf ein klar abgegrenztes Gebiet einschränken müssen, doch bereits bei wenigen Elementen ist eine grosse Rechenleistung erforderlich, die sequenziell de facto undenkbar ist. Genau hier kommt unser MapReduce Framework zur paralellen Berechnung ins Spiel.

2.1.2 Satz

> Ein Satz oder Theorem ist in der Mathematik eine widerspruchsfreie logische Aussage, die mittels eines Beweises als wahr erkannt, das heißt, aus Axiomen und bereits bekannten Sätzen hergeleitet werden kann.[Sat]

Nachdem wir mittels dem MapReduce Framework irgendwelche Strukturen gefunden habe, welche (evt. per Zufall) die Axiome einer Gruppe erfüllen, können wir in einem zweiten Schritt noch prüfen, ob für diese auch bestimmte Sätze zutreffen. Im Gegensatz zu den Gruppen-Axiomen gibt es keine definierte Menge an Sätzen, welche für eine Gruppe gültig sind. Es werden ständig neue Sätze gefunden und ausserdem existieren Sätze, für die noch nicht allgemeingültig bewiesen werden konnte, dass sie für sämtliche Gruppen gelten - es wurde aber auch noch kein Gegenbespiel gefunden. Trotzdem wir den vorhandenen Mitteln keine allgemeingültigen Beweise führen können, ist es mit genügend Rechenleistung ein Ansatz nach Gegenbespielen, also bestimmte Konstellationen von Mengen

[3] Hierbei ist e das neutrale Element

und Verknüpfungen auf diesen, welche zwar eine Gruppe sind, aber einen bestimmten Satz nicht erfüllen, zu suchen. Ein Beispiel eines Satzes, ist der Satz der Rechtsauslöschung:

$$\forall a, b, c.\, b \diamond a = c \diamond a \Rightarrow lb = c$$

Sämtliche Strukturen, welche die Axiome der Gruppe erfüllen, erfüllen auch diesen Satz. Den Beweis dazu findet man online[Can].

2.2 Generieren von Gruppen mittels Computer

Es gibt unendlich viele Mengen und darauf mögliche Verknüpfungen. In unserer Implementation kann bei Programmstart angegeben werden nach Mengen mit wie vielen Elementen gesucht werden soll[4]. Um Gruppen zu finden, haben wir einen Algorithmus implementiert, welcher alle möglichen zweistelligen Verknüpfungen auf einer endlichen Menge generiert. Grafisch können diese Verknüpfungen als Wertetabellen interpreiert werden. Hier drei Beispiele für mögliche Verknüpfungen mit zwei Elementen:

$\diamond$	0	1
0	0	0
1	1	1

Tabelle 1: Intuitive Addition der Restklasse 2

$\diamond$	0	1
0	0	1
1	1	0

$\diamond$	0	1
0	1	1
1	1	1

Tabelle 2: Weitere mögliche Wertetabellen der Verknüpfung

Insgesamt gibt es 16 mögliche Permutationen für Wertetabellen von zweistellige Verknüpfungen bei einer Menge aus 2 Elementen. Für eine Menge aus 3 Elementen sind es Bereits 19 683. Allgemein fomuliert besitzt eine Menge mit n-Elementen n^{n*n} mögliche Wertetabellen für zweistellige Verknüpfungen.

n	Mögliche Wertetabellen für zweistellige Verknüpfungen
1	1
2	16
3	19 683
4	4 294 967 296
5	298 023 223 876 953 125
6	10 314 424 798 490 535 546 171 949 056
7	256 923 577 521 058 878 088 611 477 224 235 621 321 607

Tabelle 3: Mögliche Permutationen einer zweistelligen Verknüpfung in einem n-Elementigen System

[4] Aus technischen Gründen war es jedoch nötig die mögliche Anzahl an Elementen auf 9 zu beschränken.

Weiterhin gibt es in jeder Gruppe ein für jedes Element ein inverses Element bezüglich der zweistelligen Verknüpfung. Um alle möglichen inversen Elemente zu finden, haben wir für diese einstellige Verknüpfung ebenfalls Wertetabellen generiert. Folgend zwei Beispiele für Verknüpfungen auf der Restklasse 2.

a	0	1
a^{-1}	1	0

a	0	1
a^{-1}	0	0

Tabelle 4: Beispiele für einstellige Verknüpfungen mit zwei Elementen

Für eine Gruppe aus zwei Elementen gibt es also 4 mögliche Wertetabellen und allgemein gibt es 2^n Möglichkeiten für einstellige Verknüpgungen aus Gruppen mit n-Elementen. Auch diese Tabellen können wir mit unserem Algorithmus simulieren.

Das neutrale Element ist anschliessend einfach zu finden, da es schlicht das Ergebnis der zweistelligen Verknüpfung eines Elementes mit seinem Inversen ist.

Schlussendlich suchen wir Gruppen indem wir eine zweistellige Verknüpfung, eine einstellige Verknüpfung und ein neutrales Element zusammen auf die Axiome überprüfen. Dazu prüfen wir jede Kombination dieser drei Komponenten in der Map Phase. Die genaue Formulierung dieses Problems als Map Funktion ist im nächsten Abschnitt beschrieben.

2.2.1 Generierung der Wertetabellen

Bereits für Mengen mit drei Elementen gibt es, wie oben gezeitg, bei der zweistelligen Verknüpfung tausende von Möglichkeiten. Da wir die Prüfung der Axiome, spätestens für die Berechnung von Gruppen aus einer vier-Elementigen Menge, verteilen müssen, gilt es einerseits einen Algorithmus finden, der sich auf verschiedene Rechner aufteilen lässt, sodass jeder Rechner eine Teilmenge der Wertetabellen überprüfen kann, und andererseits diese Teilmengen so abzubilden, dass sie nicht zu viel Speicher verbrauchen. Auf einem Rechner alle Wertetabellen zu generieren und diese dann an die verschiedenen Rechner über das Netz zu schicken ist also nicht praktikabel.

Nehmen wir als Beispiel die Menge mit zwei Elementen. Bei genauerer Betrachtung einer Wertetablle sieht man, dass diese anstatt zweidimensional auch eindimensional interpretiert werden könnte

$\diamond$	0	1
0	a	b
1	c	d

Tabelle 5: Logischer Aufbau einer zweistelligen Verknüpfung für zwei Elemente

eindimensional sähe die abgebildete Wertetabelle folgendermassen aus:

$$\diamond = \{a, b, c, d\}$$

Für eine Menge mit zwei Elementen kann man sich leicht überzeugen, dass sämtliche Kombinationen die Binärwerte der Zahlen zwischen 0 und 15 sind: z.B. $0 = \{0, 0, 0, 0\}$(alles bildet auf 0 ab) oder $6 = \{0, 1, 1, 0\}$ (die intuitive

Addition der Restklasse 2). In der zweidimensionalen Wertetabelle konnten wir die Funktionswerte mithilfe der x und y Indizes ablesen. Diese x und y Werte können mit der folgenden Funktion auf den Index in der Binärrepräsentation abgebildet werden: $f_{trans2}(x, y) = y * 2 + x$

Wäre man als Beispiel in der sechsten Wertetabelle an der Abbildung $1 \diamond 0$ interessiert, würde das $x = 1$, $y = 0$ entsprechen. Transformiert auf die eindimensionale Repräsentation ergibt dies die Binärrepräsentation 0110 und für den Index $f_{trans2}(1, 0) = 1$. Also das zweite Element der Binärrepräsentation, was eine eins ist - wir in der Wertetabelle.

2.2.2 Implementation von Quantoren in Java

Wie im Kapitel „Mathematische Grundlagen" beschrieben, werden Axiome einer Gruppe und Sätze einer beliebigen Struktur mittels Quantoren definiert. Als Beispiel soll das Axiom 'Es gibt ein neutrales Element' der Struktur Gruppe näher erläutert werden: $\exists e. \forall a. (a \diamond e = e \diamond a = a)$. Dieses Axiom verwendet zwei Quantoren: Einen Existenzquantor ($\exists$) und einen Allquantor ($\forall$). Mit dem Existenzquantor wird ausgesagt, dass es in einer bestimmten Struktur ein e gibt, für das das folgende Prädikat wahr ist, während mit dem Allquantor gesagt wird, dass das Prädikat für alle Elemente dieser Struktur gelten muss. In unserem Beispiel 'Es gibt ein neutrales Element' tauchen diese beiden Quantoren direkt nacheinander auf, wodurch dieses Axiom in Prosa folgendermassen formuliert werden kann: In einer Gruppe gibt es ein Element e, welches, wenn es mit irgendeinem Element a aus dieser Struktur verknüpft wird, immer auf das Element a abbildet. Dabei ist es gleichgültig, ob a mit e verknüpft wird oder e mit a - es muss immer auf a abbilden. Genau dann, wenn es ein solches Element e gibt, dann ist e das neutrale Element dieser Struktur und das Axiom ist erfüllt.

Da es in Java keine Quantoren gibt, haben wir uns Schleifen zu Hilfe genommen um die Axiome zu implementieren. Mittels der folgenden Funktion wird für ein bestimmtes Element e geprüft, ob es das neutrale Element zweistellligen Verknüpfung für eine Gruppe mit n-Elementen ist.

```
1 boolean isNeutral(int numEle, int aPerm, int e) {
2   for (int a = 0; a < numEle; a++) {
3     int b = map2d(a, e, aPerm, numEle);
4     if (a != b) { return false; }
5     int c = map2d(e, a, aPerm, numEle);
6     if (a != c) { return false; }
7   }
8   return true;
9 }
```

Abbildung 2: Beispielimplementation von Quantoren

Diese Funktion besitzt drei Parameter: Der erste ist ein Wert vom Typ Integer, also eine Ganzzahl, welcher die Anzahl an Elementen der zu prüfenden Gruppe repräsentiert.

Der zweite Parameter ist der Identifikator für die zweistellige Wertetabelle, welcher auch als Ganzzahl modelliert ist.

Der letzte Parameter ist schliesslich das vermeindlich neutrale Element, für welches diese Funktion prüft, ob es wirklich das neutrale Element ist.

In der zweiten Zeile wird durch alle Elemente der Menge iteriert, wodurch die Variable a den Wert jedes Elementes einmal annimmt (beginnend bei 0). In der dritten Zeile wird die Funktion map2d aufgerufen, welche das Element a (also jedes Element der Menge einmal) mit dem neutralen Element e verknüpft. Diese Funktion map2d funktioniert nach unserem eigents erfundenen Algorithmus, den wir im Kapitel „Generieren von Gruppen mittels Computer" beschrieben haben. Der Rückgabewert dieser Funktion wird der Variable b zugewiesen um in der nächsten Zeile mit der Variable a verglichen zu werden. Wenn diese Bedingung zutrifft, bedeutet das, dass $a \diamond e = e$ gilt, also ist das e bereits linksneutral. Wenn es nicht linksneutral ist, wird die Funktion sofort abgebrochen, da das Prädikat einen Allquantor beinhaltet - es muss also für alle Elemente gelten. Auf den folgenden Zeilen fünf und sechs wird die Variable a nochmals mit dem vermeindlich neutralen Element verknüft - diesmal aber um die Rechtsneutralität zu überprüfen (deshalb sind hier die Parameter a und e für die Funktion map2d vertauscht). Wenn sämtliche Elemente dieser Gruppe links- und rechtsneutral sind, wird die Schleife verlassen und die Funktion erreicht die Zeile Nr.8, in der der boolsche Wert 'wahr' zurückgegeben wird. Dies bedeutet, dass das Element e tatsächlich das neutrale Element in dieser Struktur ist.

2.3 Map und Reduce Funktionen für die Aufgabenstellung

2.3.1 Prüfen der Axiome als Map-Funktion

Das CPU-intensive überprüfen der Strukturen, welche möglicherweise die Axiome erfüllen, wurde als MAP-Funktion definiert. Dabei werden den einzelnen rechnenden Einheiten, im folgenden Worker genannt, da auch schon ein Rechner mehrere rechnende Einheiten besitzen kann, jeweils ein Bereich an Permutationen der Additionstabelle und ein n im Sinne der derzeit vorhandenen Anzahl an verschiedenen Elementen als Input übergeben. Die Worker führen dann die Überprüfung der Axiome durch, indem sie jeweils für alle Additionstabellen in ihrem Bereich alle möglichen Inversentabellen mit allen möglichen neutralen Elementen prüfen.

Hierbei gibt es Vereinfachungen, die inhaltlich korrekt sind jedoch die benötigte Rechenleistungen massiv reduzieren:

- Sowohl die Additions- wie auch die Permutationstabelle werden nicht immer vollständig vor der Prüfung der Axiome aufgestellt, sondern durch die oben beschriebene Funktion (2.2 Generieren von Gruppen mittels Computer) implementiert, mit der zu einer gegebenen Permutationsnummer eines beliebigen n das Ergebnis für ein bestimmtes Element zurückgibt. Das heisst der Rechenaufwand wird substantiell reduziert da bei einer Prüfung der Axiome im Normalfall nur ein quasi vernachlässigbarer Teil der Additionstabelle und, abhängig von Erfolg oder Misserfolg der Überprüfungen, auch nur ein minimer Teil der Inversentabelle benötigt wird.

- Für eine gegebene Additions- und Inversen-Tabellenkombination wird jeweils nur das neutrale Element, das für den ersten Eintrag in der Inversentabelle berechnet wurde, geprüft. Da per Definition nur ein neutrales Element für

eine spezifische Kombination aus Tabellen existiert, muss es auch für alle weiteren Einträge an Inversentabellen gelten. Sobald ein Eintrag gefunden wurde für den ein anderes neutrales Element gelten würde, wird die Berechnung für die vorliegenden Tabellen verworfen Auch diese Annahme reduziert substantiell die durchgeführten, nicht zielführenden Rechnungen.

Die Ausgabe der Map-Funktion sind schlussendlich alle Gruppen auf die die gegebenen Axiome zutreffen. Eine Gruppe wird ausreichend durch 3 Zahlen beschrieben:

- Die Nummer der Additionstabelle. Aus Ihr lassen sich mit der oben beschriebenen Funktion alle Additionen, die in der Gruppe möglich sind, herleiten.
- Die Nummer der Inversentabelle. Aus Ihr lassen sich mit der oben beschriebenen Funktion die Inversen zu allen Elementen der Gruppe herleiten
- Das neutrale Element der Gruppe

Diese effiziente Speicherung ermöglicht das Vorhalten von enorm vielen Gruppen ohne einen hohen Speicherbedarf.

2.3.2 Prüfen der Sätze als Reduce-Funktion

In der Reduce-Phase wird für die, relativ zur Menge der möglichen Gruppen, enorm kleine Menge an effektiven Gruppen der Satz jeweils geprüft. Als Output gibt sie die Gruppen für die der Satz nicht zutrifft. Da anzunehmen ist, dass somit der Output immer leer ist, kann er auch als booleanscher Wahrheitswert interpretiert werden:

- Wenn der Output leer ist, ist der Satz für die geprüften Gruppen wahr.
- Wenn der Output nicht leer ist, ist der Satz für die geprüften Gruppen falsch (was einer mathematischen Sensation gleich käme).

3 Weiteres Vorgehen

3.1 Möglichkeiten der Vertiefung der Aufgabenstellung

Wie bereits geschildert, wurde uns die genaue Vertiefung der Aufgabenstellung selbst überlassen. Für uns kamen drei mögliche Schwerpunkte in Frage auf die im folgenden näher eingegangen werden soll.

3.1.1 Erweiterung des MapReduce Frameworks

Bei diesem Schwerpunkt wäre die Zielsetzung eine möglichst grosse Kapazität um Gültigkeiten zu beweisen mit dem MapReduce Framework und verteiltem Rechnen zu erzielen. Die enormen Datenmengen, die bereits bei kleinen Eingabewerten entstehen sind auf einem einzelnen Rechner spätestens ab Gruppen mit 4 Elementen quasi nicht zu bewältigen dabei ist primär die Rechenzeit relevant. Es müssten daher Möglichkeiten gefunden werden, um eine Menge an Inputwerten zu anderen Computern zu senden, dort berechnen zu lassen und für das weitere Vorgehen relevante Daten zu returnieren.

Anforderungen bei diesem Schwerpunkt wären primär die Handhabung von verteilten Systemen, Design der Applikation und eine performante Umsetzung.

3.1.2 Freie Eingabe von auf Gültigkeit zu prüfenden Annahmen

Hierbei läge das Hauptaugenmerk darauf eine Eingabesprache für ein automatisiertes Gültigkeitsbeweis-Programm zu erstellen. Es müsste eine Möglichkeit gefunden werden dem Programm mathematische Strukturen mit bestimmten Eigenschaften, darauf gültige Axiome und daraus resultierend zu prüfende Sätze zu übermitteln. Dabei müsste, gleich ob interaktiv oder durch die Auswertung von Eingabedateien, eine Eingabesyntax definiert werden und, was wohl die wirkliche Problemstellung des Schwerpunktes wäre, eine Auflösung der Syntax und eine darauf basierende dynamische Umsezuung in ausführbaren Code.

Anforderungen bei diesem Schwerpunkt lägen primär in der theoretischen Informatik beim Thema Compilerbau sowie der dynamischen Generierung einer Interpretation der Mathematischen Objekte, ihrem Einbinden und dem Ausführen in einer Applikation zur schlussendlichen Gültigkeitsprüfung.

3.1.3 Automatisierte Beweisbarkeit

In diesem Fall wäre das Ziel ein Framework zu erstellen welches mit einem gegebenen Set aus mathematischen Operationen (zugelassenen Schlussregeln) automatisch Sätze aus Eingaben (Sätzen und Axiomen) herleitet. Die Operationen, die das System vornehmen könnten entsprächen den im Kalkül zugelassenen Schlussregeln. Ziel wäre es somit automatisiert mehr oder weniger sinnvolle Aussagen mathematisch formell korrekt herzuleiten. Dabei müssten Kriterien definiert werden mit denen die Unterscheidung von formal korrekten aber irrelevanten und formal korrekten, relevanten Aussagen möglich macht.

Anforderungen bei diesem Schwerpunkt wäre wieder die Interpretation der in einer gegebenen Syntax vorliegenden Eingaben jedoch des weiteren das Erkennen von neuen relevanten Schlussfolgerungen.

3.2 Entscheidung

Der auf Anhieb spannendste Schwerpunkt stellte nach der ersten Bertrachtung die „Freie Eingabe von auf Gültigkeit zu prüfenden Annahmen" dar. Die in dem Schwerpunkt vorkommenden Themen wurden weitestgehend im bisherigen Studium noch nicht behandelt und böten somit eine gute Möglichkeit neues kennenzulernen.

Die Automatisierte Beweisbarkeit erschien bereits bei der ersten Betrachtung als eine eher unwahrscheinliche Vertiefung, da der schlussendliche Nutzen massiv in Frage gestellt werdern kann. Es wäre zu befürchten, dass eine enorme Menge an Arbeitszeit und Herzblut investiert würde um Schlussendlich eine Applikation zu erstellen welche Mathematisch völlig irrelevante Sätze herleiten würde.

Eine Erweiterung des MapReduce Framework erschien bereits nach der ersten Analyse als überschaubares Problem im Sinne der Beherrschbarkeit der Problemstellung. Zwar ist der mit der Vertiefung verbundene Aufand alles andere als als klein zu erachten, jedoch sollten keine vollkommen unvorhergesehenen Fragestellungen, zu denen ohne eine sehr grosse Menge an Vertiefung keine Lösung gefunden werden könnte, auftreten.

Das Kriterium der Beherrschbarkeit gab schlussendlich auch den Ausschlag anstatt der Vertiefung Nr. 2 eine Erweiterung des Frameworkes anzustreben. Die Gefahr durch die enorme Komplexität am Ende nur eine mehr prototypisch als sinnvolle Implementation eines Programmes zu erstellen schreckte uns ab. Desweiteren stellte die Idee mit der gegebenen Infrastruktur der Fachhochschule eventuell eine grosse Menge an PCs für einen Feldversuch zu besitzen einen grossen Ansporn dar.

4 Teil 2 MapReduce Framework zur verteilten Berechnung

4.1 Technologie

In Absprache mit unserem betreuenden Dozenten Herrn Albert Heuberger soll an dieser Stelle nur sehr Abstrakt auf die Implementation eingegangen werden. Um einen groben Überblick zu gewähren, sind die nun folgende Buzzwords leider unumgänglich.

Als Programmiersprache wählten wir, aufgrund der für den Aufbau eines verteilten Rechnernetzes sehr vorteilhaften Betriebssystem Unabhängigkeit und unserer bereits vorhandenen Kenntnisse Java aus. Git [GIT] wurde als Versionsverwaltungstool eingesetzt und mit den verschiedenen Bitbucket Projekten verknüpft. Für Build Management wurde Maven [MVN] eingesetzt, zum Tracking der testcoverage Cobertura [Cob], für Dependency Injection Guice [GUI]und für das essentielle Logging die Standardmässige Java Logging API.

Auf ein GUI wurde mangels Notwendigkeit weitestgehend verzichtet. Die Applikation und das Framework lassen sich über Property-Files steuern und geben, neben einem Log mit variablem Detailierungsgerad, HTML-Files mit den gefundenen Gruppen zurück.

Das Framework besitzt Lines of code, die Applikation . Es wurden insgesamt Testklassen mit Tests im Rahmen des Projektes genutzt.

4.2 Master als zentrale Kontrollinstanz

Der Master ist ein Server, welcher die Aufgaben (Tasks) und Worker verwaltet. Ein Task ist eine abstrakte Einheit, welche von einem Worker erfüllt werden kann, hier also entweder ein Teil einer Map- oder einer Reduce-Berechnung. Im klassischen MapReduce, wie es von Google beschrieben wurde, ist ein Worker eine Maschine (PC). Somit verteilt eine Hauptmaschine (Master) die verschiedenen Aufgaben an verschiedene Rechner in einem Netzwerk.

Davon abweichend ist in unserer Implementation ein Worker nicht zwingend eine andere Maschine, sondern schlicht eine abstrakte Einheit, die Arbeit ausführen kann. In der ersten Version unseres Frameworks gab es, mit lokalen Threads auf Maschine des Master, nur eine Art von Workern.

Im Rahmen dieser Projektarbeit wurde die Möglichkeit geschaffen via Plugin irgendeine Art von Worker dem Master zu übergeben. Ein Plugin ist Stück Code, welches eine Worker-Technologie zur Verfügung stellt. Somit gibt es also zum Beispiel ein Plugin, welches Threads als Worker zur Verfügung stellt und ein Plugin, welches andere Rechner über eine Netzwerk-Verbindung zur Verfügung stellt. Dies wurde soweit abstrahiert, dass in Zukunft auch weitere Arten von Workern erstellt werden könnten. Denkbar wären hier zum Beispiel CUDA (NVidia Grafikkarten) oder Android Smartphones.

Die konkrete Verwaltung der Worker wird anschliessend vom sogenannten Pool übernommen, dieser kennt den Status seiner Worker (z.B. beschäftigt, frei, unauffindbar, etc) und weiss, welche Tasks gerade zu erledigen sind.

4.3 Ablauf

Wie im Kapitel „MapReduce" beschrieben muss der Benutzer nur eine Map- und eine Reduce Funktion zur Nutzung des Frameworks erstellen.

In unserer Implementation geschieht dies durch die beiden Interfaces *MapInstruction* und *ReduceInstruction* die in einer MapReduceTask zusammengefasst und anschliessend mit einer Menge an Input dem Framework übergeben werden.

Daraus kann dann der Master die Map- und Reduce Instruktionen extrahieren und erstellt zunächst aus je einem Teil vom Input und einer Instanz der MapInstruction einen MapTask. Folglich entstehen für eine MapReduce Berechnung viele MapTasks, welche dem Pool zur Verarbeitung übergeben werden. Hier wird vorausgesetzt, dass dem Pool durch ein oder mehrere Plugins bereits Worker zur Verfügung gestellt worden sind, da die Tasks sonst niemals ausgeführt werden könnten.

Nun kann der Pool einen auszuführenden MapTask einem verfügbaren Worker übergeben. Wenn dieser Worker eine andere Maschine im Netz abstrahiert, würde das bedeuten, dass dieser MapTask auf einem anderen Rechner ausgeführt wird.

Zur Speicherung von Zwischenergebnissen werden den einzelnen Tasks Kontexte übergeben. Somit ist es aus Sicht des Tasks nicht transparent, ob die berechneten Zwischenergebnisse im Arbeitsspeicher, auf der Festplatte oder an einem völlig andern Ort gespeichert werden. Diese weitere Abstraktion hilft auch enorm bei der Abstraktion des Workers.

Sobald die Berechnung (z.B. auf einem anderen Rechner) ausgeführt wurde, meldet sich der Worker wieder beim Pool, dass er wieder bereit ist, Aufgaben anzunehmen. Also könnte eine ganze Berechnung entweder von einem einzelnen Worker in mehreren Schritten ausgeführt werden oder auf mehrere Worker aufgeteilt, je nach dem wie schnell ein einzelner ist.

Gleichzeitig prüft der Master kontinuierlich, ob alle MapTasks ausgeführt worden sind. Sobald alle MapTasks für eine Berechnung ausgeführt worden sind, beginnt der Master alle Resultate zu sammeln und schliesst somit die Map-Phase ab.

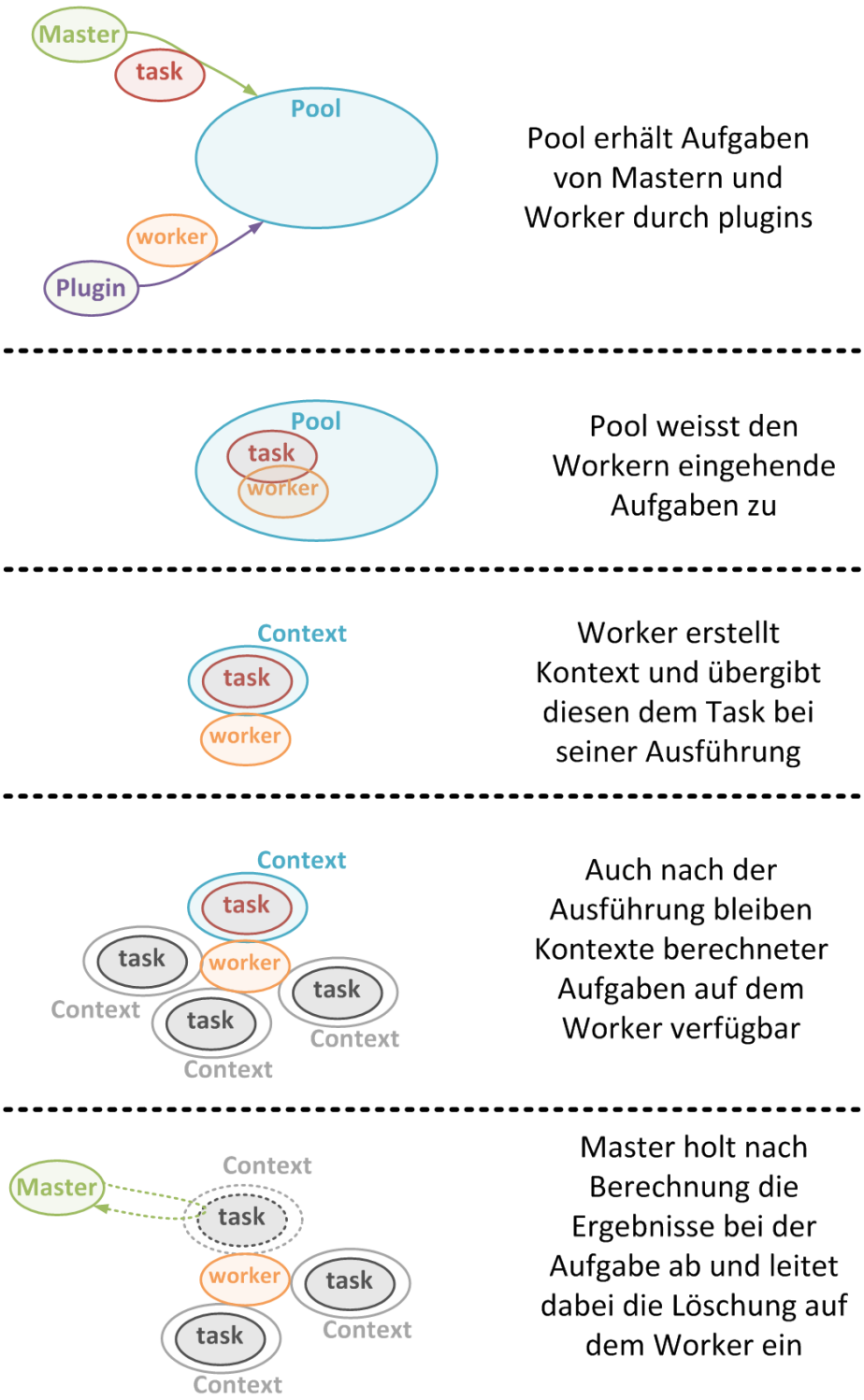

Abbildung 3: Interaktion von Task, Pool, Worker und Kontext

Darauf folgt die Shuffle-Phase, die die Resultate der Map-Phase nach Schlüssel gruppiert und somit auf die Reduce-Phase vorbereitet. Da vor der Shuffle-Phase alle MapTasks ausgeführt werden sein müssen, gilt dies als Bottleneck von MapReduce. Nach der Shuffle-Phase beginnt der Master damit die einzelnen ReduceTasks, bestehend aus einer Instanz der ReduceInstruction und einem Teil der Shuffle Resultate, zu erstellen und diese wieder dem Pool zu übergeben. Dann beginnt der Pool wieder diese ReduceTasks den verfügbaren Worker zuzuteilen und der Master wartet wiederum auf deren Fertigstellung. Sobald auch diese fertig sind, ist die MapReduce Berechnung abgeschlossen.

4.4 Ablauf bei verteilter Berechnung

Für das Ausführen von verteilen Berechnungen haben wir das Framework SIMON [SIM] verwendet. Mit Hilfe dieses Frameworks können Methoden auf einer entfernten Instanz irgendwo im Netzwerk gemacht werden. Damit sich der Master nicht um die Komplexitäten im Zusammenhang mit Remote Procedure Call (RPC) kümmern muss, wird ein einzelner Client, also eine entfernte Instanz, auf dem Server genauso wie ein gewöhnlicher Thread als Worker im Pool registriert. Die konkrete Implementation dieses Workers ruft aber, anstatt den Task direkt auszuführen, via Remote Procedure Call [RPC] eine Methode auf dem Client auf und übermittelt im die Instruktion mit zugehörigen Eingabewerten.

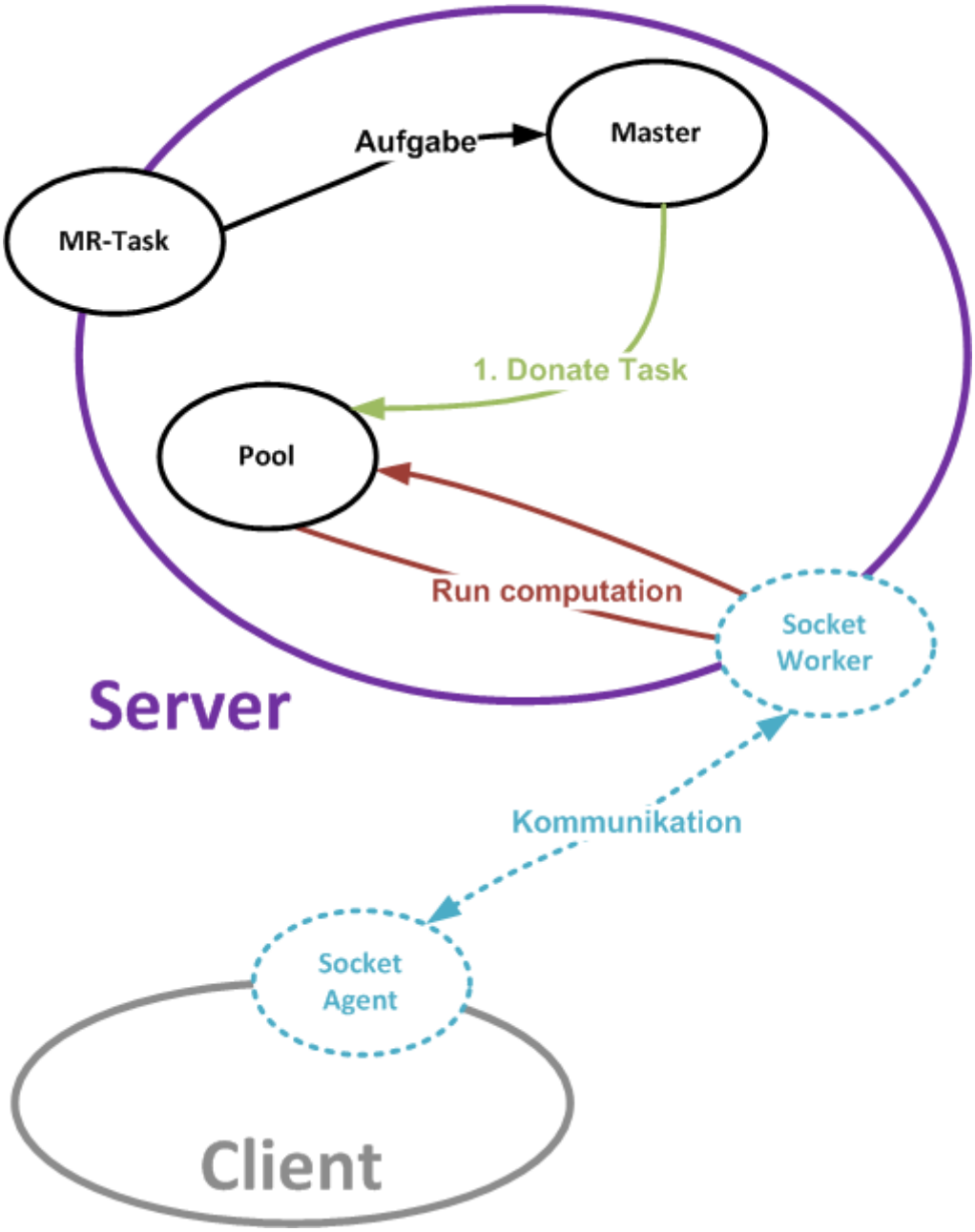

Abbildung 4: Interaktion von Client und Server

Auf dem Client wartet ein sogenannter SocketAgent ständig darauf, dass er Tasks vom Server zugeordnet bekommt. Sobald ein Task angekommen ist, wird dieser ausgeführt und die Resultate werden wieder an den SocketWorker geschickt, welcher auf dem Server ist. Der SocketWorker kümmert sich dann um das korrekte Speichern der Resultate und meldet dem Pool, dass er wieder Arbeit aufnehmen kann.

5 Ergebnisse

5.1 Anwendung der Applikation

5.2 Durchführung

5.3 Auswertung der gefundenen Gruppen

5.4 Schlusswort

6 Apendix

Abbildungsverzeichnis

Tabellenverzeichnis

Literatur

[Can] *Rechtsausloeschung im Proofwiki.* `http://www.proofwiki.org/wiki/Cancellation_Laws`. – Stand: 30.05.2013

[Cob] *Cobertura Hompage.* `http://cobertura.sourceforge.net/`

[DG04] Dean, Jeffrey ; Ghemawat, Sanjay: MapReduce: Simplified Data Processing on Large Clusters. In: *OSDI* (2004)

[GIT] *Git Homepage.* `http://git-scm.com/`

[GUI] *google-guice Hompage.* `http://code.google.com/p/google-guice/`

[MVN] *Maven Hompage.* `http://maven.apache.org/`

[RPC] *Remote Procedure Call auf Wikipedia.* `http://de.wikipedia.org/wiki/Remote_Procedure_Call`. – Stand: 30.05.2013

[Sat] *Satz auf Wikipedia.* `http://de.wikipedia.org/wiki/Satz_(Mathematik)`. – Stand 30.05.2013

[SIM] *Simon Hompage.* `http://dev.root1.de/projects/simon`