TINF 13 AI-BI

Duale Hochschule Baden-Württemberg Mannheim



Studienarbeit Modul T3201 DBHW Mannheim

Implementierung von SetlX-Programmen in Python

Name: Johann Boschmann Name: Joseph Palackal

Matrikelnummer: XXXXXXX Matrikelnummer: 5839999

Kurs: TINF13AIBI

Studiengangsleiter: Prof. Dr. A. Müller Betreuer: Prof. Dr. K. Stroetmann

Datum:

Zeitraum: XX.XX. – XX.XX.XXXX Unterschrift

Betreuer:

Inhaltsverzeichnis

1.	Einle	eitun	g	. 3
2.	War	um F	- Python?	. 3
			n reinem Python //TODO: besseren Titel finden	
4.	Pyth	non N	Modul lecture	. 4
4	l.1.	Sets		. 5
4	1.2.	Mat	ches	. 5
4	1.3.	Übe	rsetzung komplexerer Programme	. 6
	4.3.	1.	Schiebepuzzle	. 6
	4.3.3.		Watson	10
			Wolf Ziege Kohl	10
			8 Damen Problem	10

Abbildungsverzeichnis

Abbildung 1: if-Abfrage in SetIX	Fehler! Textmarke nicht definiert.
Abbildung 2: if-Abfrage in Python	Fehler! Textmarke nicht definiert.
Abbildung 3: Fehler bei Mengen in Mengen	5
Abbildung 4: find path im Schiebepuzzle	7
Abbildung 5: next states im Schiebepuzzle	8
Abbildung 6: mov dir im Schiebepuzzle	
Abbildung 7: find blank im Schiebepuzzle	g

1. Einleitung

In der Vorlesung "Grundlagen und Logik" des Moduls Theoretische Informatik I führt der Dozent Prof. Dr. Karl Stroetmann die Programmiersprache SetlX ein. Diese Programmiersprache wirbt damit, dass die Verwendung von Mengen und Listen sehr gut unterstützt wird. Außerdem können Ausdrücke aus der Mengenlehre, so wie andere mathematischen Ausdrücke in einer Syntax, die sehr ähnlich zur mathematischen Notation ist, implementiert werden.¹ Da diese Vorlesung bereits im ersten Semester stattfindet und die Studenten parallel dazu eine Vorlesung aus dem Modul Mathematik I besuchen, können die Studenten Themen wie beispielsweise die Mengenlehre schneller kennen lernen. Die komplementäre Auseinandersetzung mit ähnlichen bis gleichen Themen in beiden Vorlesungen ermöglicht das gleichzeitige Lernen für zwei Vorlesungen.

Ein weiterer Vorteil für die Studenten ist, dass die Syntax von SetX, zusätzlich zum sehr mathematischen Stil, auch starke Ähnlichkeiten zur Programmiersprache C aufweist. Selbst für die Studenten, die zuvor keinen Kontakt mit C hatten ist das ein großer Vorteil, da im ersten Semester parallel zur theoretischen Informatik Vorlesung auch eine Vorlesung mit dem Titel "Programmieren in C" besucht werden muss. So muss kein starkes Umdenken stattfinden, wenn von SetlX zu C und auch umgekehrt gewechselt wird.

- SetIX für mathematische Ausdrücke gut
- Viel mit Sets gearbeitet

2. Warum Python?

Viele Informatik-Kurse oder Vorlesungen für Anfänger benutzen die Programmiersprache Python als erste Programmiersprache. Von den 39 besten Einführungskursen für Informatik in den USA verwenden 27 Python als erste Programmiersprache.² Mit 69% ist Python somit mit einer eindeutigen Mehrheit deutlich die meist verwendete Programmiersprache unter diesen Kursen. Auch die Online-Lernplattform Udacity verwendet für den Kurs "Intro to Computer Science" Python als Sprache, um die Themen der theoretischen Informatik zu erläutern.

- Sehr verbreitet
 - o Erarbeiten warum

3. Skripte in reinem Python //TODO: besseren Titel finden

Trotz dessen, dass die Erstellung des Python Moduls als Hauptbestandteil dieser Arbeit gesehen wird, ist es durchaus möglich einige SetlX Programme ohne zusätzliche, nicht enthaltene Module anzufertigen. Diese Skripte wurden als erstes angefertigt, um feststellen zu können, ob es möglich ist

¹ Vgl. SetIX Tutorial

² Vgl. Guo, Philip

Python Syntax zu verwenden, ohne die Eleganz des Codes zu verlieren. Einer der Ziele der Übersetzung ist die Eleganz der Programme beizubehalten.

Das erste Codebeispiel aus dem Logik-Skript befasst sich mit der Berechnung einer Summe der Zahlen von 1 bis zur eingegebenen Zahl. Dieses Programm lässt sich auch nahezu eins-zu-eins so in Python abbilden. Das originale SetlX Programm verwendet hierfür eine Menge, die die Zahlen von 1 bis zur eingegebenen Zahl n enthält. Daraufhin wird die Summe aller in der Menge enthaltenen Zahlen mit dem "+/"-Operator ermittelt und ausgegeben.

```
n := read("Type a natural number and press return: ");
s := +/ { 1 .. n };
print("The sum 1 + 2 + ... + ", n, " is equal to ", s, ".");
```

In Python wurde fast dasselbe Verhalten nachgebildet. Jedoch wurde anstatt eine Menge anzufertigen eine Range der Zahlen von 0 bis n angelegt. Die Summe wird über die in Python bereits integrierte Funktion sum () berechnet und daraufhin ausgegeben.

```
n = int(input('Type a natural number and press return: '))
s = sum(range(n + 1))
print('The sum 1 + 2 + ... + ', n, ' is equal to ', s, '.')
```

Allgemein kann gesagt werden, dass ein SetlX-Programm ohne spezielle Funktionen oder Strukturen, die nicht in Python wiedergefunden werden, meist eine große Ähnlichkeit mit der Python-Implementierung hat.

4. Python Modul lecture

Das Verhalten der Mengen in SetlX ist in einigen Bereichen anders als bei den Mengen in Python. Da zusätzlich auch gewisse Funktionen von den Python-Mengen nicht unterstützt werden, wurde das Python Modul lecture im Rahmen dieser Studienarbeit angefertigt.

Ein großes Problem, das sehr früh erkannt wurde, ist, dass die Mengen in Python nur hashbare Werte in einer normalen Menge hinterlegt werden dürfen. Somit ist es beispielsweise verboten Mengen in Mengen zu hinterlegen. Diese Strukturen finden in den SetlX Programmen, die in diversen Vorlesungen gezeigt werden, öfter Anwendung, jedoch gibt es in Python den Fehler, der in Abbildung 1 zu sehen ist. Zwar können anstelle der "normalen" sets frozensets verwendet werden, allerdings können die Elemente in der Menge nicht geändert werden, weil die Mengen, wie der Name es bereits impliziert, eingefroren sind. Eine weitere Möglichkeit wäre die Verwendung von Listen, anstelle von Mengen. Prinzipiell ist das in einigen Python-Übersetzungen der SetlX Programme möglich und wurde so auch teilweise umgesetzt. Es werden andere Datentypen verwendet um die Informationen zu hinterlegen, meist Listen statt Mengen, da für die Ausführung der SetlX-Programme keine besonderen Eigenschaften der Mengen verwendet werden. Ein großes Problem an dieser Lösung ist allerdings, dass Listen nun mal keine Mengen sind und sobald Mengeneigenschaften oder Mengenoperatoren, die nicht für Listen gelten, verwendet werden, Listen eher ungeeignet sind. Der Workaround, besondere Funktionen für die Listen zu schreiben, um das Verhalten von Mengen zu imitieren, wurde auch als Ansatz bedacht, allerdings nach einigen kleinen Beispielübersetzungen wieder verworfen. Eine wichtige Anforderung, die Erhaltung der Eleganz konnte nicht immer erfüllt werden. Somit war dies keine Lösung, die so für alles verwendet werden kann. Die Funktionen, die implementiert werden Kommentiert [J1]: Es geht nicht um Funktionen die Python Mengen nicht unterstützen sondern um die Tatsache das in Python Mengen alle Elemente per Hash "sortiert" etc. Somit können nur Hashbare Element, also Elemente die nicht weiter veränderbar sind in Python Mengen eingefügt werden. Wie das auch im weiteren Verlauf beschrieben ist. Also muss hier definitiv die Einleitung nochmal überarbeitet werden.

Kommentiert [J2]: Man kann in Python Mengen in Mengen legen myset.add(frozenset(myset)). Wichtig ist das anschließend die Elemente der Menge nicht verändert werden können dürfen. Kern ist nur die Tatsache mit finalen Elementen. Und ich finde es sollte darauf zu allererst eingegangen werden. Dann braucht man die Set-in-Set-Geschichte auch nur am Rande erwähnen.

Kommentiert [J3]: Allgemein fehlt ein wenig die Begründung warum die Elemente in Python-Sets nicht verändert werden dürfen (weil bei einer Änderung der Hash sich verändert und somit das Element nicht mehr in der Menge gefunden werden könnte, da die Menge das Element unter dem alten Hash gesichert hat)

Kommentiert [J4]: Das lieber zu einem späteren Zeitpunkt erwähnen und vollständig erklären. Z.B. als ersten Ansatz zur Umsetzung und mit Code-Beispiel andernfalls ist nicht klar worüber du genau sprichst.

mussten waren das Entfernen von Duplikaten aus einer Liste, so wie die Ermittlung der Differenz zweier Listen und die Ermittlung der Potenzmenge (wobei jedoch eigentlich die "Potenzliste" ermittelt wurde).

```
>Traceback (most recent call last):
> File "<stdin>", line 1, in <module>
>TypeError: unhashable type: 'set'
```

Abbildung 1: Fehler bei Mengen in Mengen

//TODO: Effizient von Sets vs Lists

Allerdings sind die Mengen nicht der einzige Grund, warum das Modul benötigt wird. Ein in SetlX sehr hilfreiches Konstrukt, namens match, wird in Python nicht wiedergefunden. Dessen Syntax ist an das sehr bekannte switch-case-Syntax angelehnt, welches auch nicht in Python enthalten ist. Matches können vier verschiedene Datentypen verwendet werden: Strings, Listen, Mengen und Terme. Das Matchen von Strings, Listen und Mengen kann für das Erkennen des ersten Zeichens und dem Rest oder auch das Herauspicken von Paaren verwendet werden. In dem SetlX-Tutorial wird das Matching zur Generierung des Inversen oder das Erstellen einer sortierten Liste aus einer Menge verwendet.³

Die interessanteste Anwendung von Matches ist jedoch, wenn Terme verwendet werden. Die "[...]Art von Matchen [in SetlX] ist ähnlich zum Matching das in den Programmiersprachen Prolog und ML gegeben ist."⁴ Dieses Matching wird auch in einigen Programmen der Logik-Vorlesung von Herrn Prof. Dr. Stroetmann verwendet. Deshalb ist es wichtig, dass diese Funktion auch in einer Python Version der Programme möglich ist.

4.1. Sets

In den Vorlesungs-Programmen von Herrn Prof. Dr. Stroetmann werden häufig Mengen, sowie Ausdrücke aus der Mengenlehre sehr ähnlich zur mathematischen Darstellung verwendet. Neben den Mengenoperationen werden zusätzlich diverse Eigenschaften von Mengen implementiert. Beispielsweise wird genutzt, dass Mengen keine Duplikate enthalten und die beliebige Reihenfolge wird verwendet. In SetlX wird eine Sortierung der Elemente durchgeführt, wodurch Vorteile in der Programmierung entstehen.

Um die Mengen, wie sie in den SetlX-Programmen verwendet werden, auch in Python verwenden zu können wurden eigene Mengen implementiert, die alle notwendigen Aufgaben erfüllen können.

- Implementierung erläutern

4.2. Matches

- Implementierung erläutern

Kommentiert [J5]: Abbildung ist bisschen schlecht Platziert, da Referenz und Abbildung weit voneinander entfernt sind

³ Vgl. SetIX Tutorial S. 39 ff

⁴ SetIX Tutorial S. 42

4.3. Übersetzung komplexerer Programme

Es wurden zwar einige SetlX-Programme in Python-Skripte übersetzt, allerdings werden in dieser Arbeit hauptsächlich Programme, die die Eleganz der Programmiersprache SetlX verdeutlichen, genauer betrachtet.

Wie zuvor beschrieben, ermöglicht SetlX dem Programmierer in einem sehr mathematischen Stil zu programmieren. Somit können Personen, die erste Berührungen mit der Mengenlehre oder von der Mathematik kommen, sowie Studenten, die mathematische Konstrukte verstehen und anwenden müssen, beim Programmieren diese Erfahrungen sammeln.

4.3.1. Schiebepuzzle

Das Schiebepuzzle ist eine Aufgabe die den Studenten mit Lücken als Aufgabe gegeben wird, um Vorlesungsinhalte direkt anwenden zu können. Mit diesem Programm sollen die Studenten eine für Menschen nicht triviale Lösung zu einem Schiebepuzzle berechnen lassen. Aufgrund der Berechnung aller möglichen Pfade, das Puzzle zu lösen, lässt sich das Programm nicht so schnell wie die meisten anderen SetlX-Programme durchführen. Die genaue Zeit, wie lange das Programm für die Berechnung gebraucht hat, wird in der Kommandozeile ausgegeben. Somit wird den Studenten klar, dass selbst der Rechner diese Berechnungen nicht sofort liefern kann. Um einen Vergleich der Performance von SetlX zu Python zu haben, wird auch die Berechnungszeit der Python-Implementierung aufgeführt. Die Eigenschaften des Rechners, mit dem die Berechnungen durchgeführt wurden sind:

Prozessor: Intel i7 6700hq 2,6-3,1 GHz

Hauptspeicher: 8 GB DDR4

Das SetlX-Programm lief in 17,4 Sekunden, während das Python-Skript 46,6 Sekunden für die Berechnung benötigte. Auffällig ist, dass in SetlX die Ausführung über doppelt so schnell wie bei der Python-Implementierung ist. Grund hierfür ist, dass die virtuelle Maschine, in der Java ausgeführt wird, etwas effizienter als die virtuelle Maschine von Python ist. SetlX ist eine auf Java basierende Programmiersprache. Es wird deshalb die Effizienz von Java zu Python verglichen. Außer dem Unterschied bei den virtuellen Maschinen, unterstützt die Programmiersprache Java zusätzlich eine statische Typisierung. Python hingegen unterstützt, wie es für Skriptsprachen üblich ist, eine dynamische Typisierung, die etwas ineffizienter ist. Eine statische Typisierung ist effizienter da der Rechenaufwand für eine Typüberprüfung wegfällt. Eine unterschiedliche Implementierung der für diese Arbeit entwickelten Mengen und der Mengen, die in SetlX verwendet werden, kann nicht die Ursache für diese Abweichungen in der Performanz sein. Die Sets des Python-Moduls "lecture" wurden basierend auf die Implementierung, wie die in SetlX verwendet wird, umgesetzt.

Sowohl das SetlX-Programm, wie auch die Übersetzung in Python definieren zu Beginn die Funktion, mit der aus einem State (einem derzeitigen Zustand des Puzzles, abgelegt in einer Liste) ein String erzeugt werden kann, um eine bessere Visualisierung zu ermöglichen. Bei der Übersetzung ist in dieser Methode nichts großartig Interessantes zu sehen, da in den meisten Zeilen fast eins-zu-eins dasselbe steht. Allerdings ist zu beachten, dass die for-Schleifen in SetlX über Listen von 1-3 iterieren, während in Python dafür eine Range mit den Werten 0-2 verwendet wird. Es wird allerdings die selbe Ausführung erreicht, da Listen-Indizes in SetlX bei 1 anfangen, während Python die 0 als Index verwendet, um das erste Element aufzurufen.

Kommentiert [J6]: Vielleicht erst auf die Implementierungen eingehen dann auf die Laufzeit aufmerksam machen. So fällt am Ende die Begründung leichter aus

Kommentiert [J7]: Vielleicht SetlX im Vorfeld vorstellen, weil sonst musst du das erstens bei jeder Implementierung mit Zeitmessung erwähnen, zweitens ist es an dieser Stelle seltsam das erst jetzt erwähnt wird das SetlX auf Java basiert. Ggf einen Grundlegenden Sprachenvergleich vornehmen?

Kommentiert [J8]: Kurze Quellenangabe oder ähnliches wäre zu dem Inhalt passend.

```
findPath := procedure(start, goal, nextStates) {
            := 1;
   count
             := { [start] };
   paths
            := { start };
    states
    explored := {};
    while (states != explored) {
       print("iteration number $count$");
        count += 1;
        explored := states;
                 := \{ 1 + [s] \}
        paths
                    : 1 in paths, s in nextStates(1[-1])
                    | !(s in states)
                    };
                 += { p[-1] : p in paths };
        print("number of states: $#states$");
        if (goal in states) {
           return arb({ l : l in paths | l[-1] == goal });
    }
};
```

Abbildung 2: findPath im Schiebepuzzle (SetlX)

```
def find_path(start, goal, next_states):
   count iteration = 1
                  = 0
   count_states
                   = Set([start])
   paths
   states
                   = Set(start)
   while len(states) != count states:
       count states = len(states)
       print('Iteration number %s' % count_iteration)
       count iteration += 1
       paths = Set(x + [s])
                   for x in paths for s in next_states(x[-1])
                    if not s in states)
       states += Set(p[-1] for p in paths)
       print('Number of states: %s' % len(states))
       if goal in states:
            return Set(1 for 1 in paths if 1[-1] == goal).arb()
```

Abbildung 3: find_path im Schiebepuzzle (Python)

Die Funktion, mit der letztendlich auch der Pfad vom Start-Zustand zum Ziel-Zustand ermittelt wird, ist in Python die find_path Methode. Die Parameter, die übergeben werden, sind identisch zu der SetlX-Implementierung. Die am Anfang der Funktion definierten Variablen weichen vom SetlX-Code um eine Variable ab. Die Variablen paths und states sind in beiden Versionen zu finden und der Integer count_iteration ist in SetlX als count zu finden. Allerdings werden zur Prüfung ob neue Zustände hinzukommen unterschiedliche Ansätze verwendet. In der Python-Implementierung keine Menge mit allen entdeckten Zuständen, sondern die Prüfung der Anzahl der verschiedenen Zustände verwendet. Somit vergleicht die äußere while-Schleife zwei Integer, während die Vorlage zwei Mengen vergleicht. Abgesehen davon ist der Ablauf sehr ähnlich. Es wird die derzeitige Iteration angegeben,

daraufhin die neuen Pfade anhand von next_states ermittelt und dann alle derzeitigen Zustände der Pfade ermittelt und die Anzahl ausgegeben. Zuletzt wird noch im Falle, dass das Ziel bereits erreicht wurde, ein beliebiger Lösungspfad zurückgegeben.

Abbildung 4: nextStates im Schiebepuzzle (SetIX)

Abbildung 5: next_states im Schiebepuzzle (Python)

In find_path muss, für die Ermittlung der möglichen Pfade, die Funktion next_states aufgerufen werden. Diese gibt eine Liste zurück, die alle erreichbaren Zustände vom Zustand state, der als Parameter übergeben wird, enthält. In SetlX wird diese Liste als Menge zurückgegeben, allerdings werden Listen in diesem Fall bevorzugt, da sie in Python geläufiger verwendet werden und keine Mengeneigenschaften in dieser Situation benötigt werden. Die Richtungen werden allerdings nicht wie im SetlX-Programm als Menge von Listen, sondern als Liste von Tupeln definiert, da Tupel in Python, so wie Listen in SetlX, unzipped werden können. Eine Änderung, die an der Vorlage unternommen wurde ist, dass die Bewegung von einer Variable move auf zwei Variablen dx und dy aufgeteilt wurden.

```
moveDir := procedure(state, row, col, dir) {
    [dx, dy] := dir;
    nextState := state;
    nextState[row ][col ] := state[row + dx][col + dy];
    nextState[row + dx][col + dy] := 0;
    return nextState;
};
```

Abbildung 6: movDir im Schiebepuzzle (SetlX)

```
def move_dir(state, row, col, direction):
    (dx, dy) = direction
    next_state = [list(x) for x in state]
    next_state[row][col] = next_state[row + dx][col + dy]
    next_state[row + dx][col + dy] = 0
    return next_state
```

Abbildung 7: mov dirim Schiebepuzzle (Python)

Die Methode mov_dir erhält dieselben Parameter wie in der Vorlage und gibt den nächsten Zustand zurück, nachdem vom derzeitigen Zustand aus eine Bewegung in entweder die x- oder die y-Achse erfolgt. Einer der Unterschiede zur ursprünglichen Umsetzung ist hier, wie zuvor bereits erwähnt wurde, dass die Richtung als Tupel und nicht als Liste gewertet wird.

Ein erheblicher Unterschied zwischen den beiden Implementierungen ist das initiale Setzen der Variable next_state. Während in SetIX der Parameter state dafür verwendet wird und eine einfache Zuweisung erfolgt, muss in Python die list() Funktion auf alle Listen in state verwendet werden. Grund dafür ist, dass ansonsten die Referenzen übergeben werden und somit dann die Werte von state ebenfalls geändert werden, wenn sie in next_state bearbeitet werden. Durch die list() Funktion wird eine Kopie erzeugt und somit sind next_state und state zwei unabhängige Listen. SetIX erkennt an dieser Stelle intern, ob ein Objekt geändert wurde und erstellt gegebenenfalls eine Kopie. Deshalb wird dieser Aufruf nur in der Python-Version benötigt.

```
findBlank := procedure(state) {
    for (row in [1 .. 3]) {
        for (col in [1 .. 3]) {
            if (state[row][col] == 0) {
                return [row, col];
            }
        }
    }
};
```

Abbildung 8: findBlank im Schiebepuzzle (SetlX)

```
def find_blank(state):
    for row in range(3):
        for col in range(3):
        if state[row][col] == 0:
            return (row, col)
```

Abbildung 9: find blank im Schiebepuzzle

Damit in next_states eine Zahl "bewegt" werden kann, muss über find_blank das freie Feld gefunden werden. In der Vorlage wird in findBlank eine Menge erzeugt, die alle Reihen-Zeilen-Kombinationen enthält und aus dieser einer der Werte, an denen der übergebene Zustand die 0 enthält, zurückgegeben. In dieser Arbeit wird nur ein Tupel mit einem Reihen- und einem Zeilenwert, an denen der Zustand die Null enthält zurückgegeben. Zu Vergleichszwecken ist die Funktion, so wie sie in Python geschrieben wurde, auch in SetlX implementiert. In SetlX merkt man keine Unterschiede in der Performance. Diese Methode zeigt wie ähnlich der Code in SetlX und Python sein können.

Nachdem die Funktionen alle definiert sind ist der Ablauf komplett identisch zur Vorlage. Es wird die Zeitmessung begonnen, der Start- und End-Zustand definiert, daraufhin der Pfad ermittelt und die Zeitmessung beendet. Abschließend werden der Lösungspfad und die Zeitmessungsergebnisse ausgegeben.

Im Allgemeinen ist der Code in beiden Programmiersprachen sehr ähnlich und der Ablauf, so wie die Syntax, teilweise sogar identisch. Einige Ausdrücke sehen auf Grund der Programmiersprache unterschiedlich aus, erfüllen aber denselben Zweck. Bei der Zeitmessung beispielsweise wird in SetlX nur die Methode now() aufgerufen, während in Python timeit.default_timer() aufgerufen wird und sogar ein Import dafür notwendig ist. Andere Abweichungen sind unterschiedliche Datentypen, die in Python gewählt wurden. Diese werden verwendet um die Laufzeit etwas zu verbessern, weil die Implementierung, die für Mengen verwendet werden muss, nicht so effizient wie die SetlX-Implementierung ist. An einigen Stellen sind aber auch keine Mengen notwendig und werden deshalb durch Listen ersetzt.

4.3.2. Watson

4.3.3. Wolf Ziege Kohl

4.3.4. 8 Damen Problem