

Laborumgebung für algorithmische Lösungsverfahren

R. Schuler, W. Lindner, M. Arnold, H.A. Kestler

Abt. Theoretische Informatik, Neuroinformatik
und Innere Medizin I

Laborumgebung zur Simulation und Veranschaulichung von Methoden und Verfahren der Themenkreise:

- ▶ Algorithmen, z.B. Netzwerke, maximaler Fluß
- ▶ Kryptographie, z.B. Euklid, Miller-Rabin Algorithmus
- ▶ Bioinformatik, z.B. Phylogenetische Bäume, hierarchische Clusteralgorithmen

Zur vorlesungsbegleitenden Unterstützung von

- ▶ Übungen
- ▶ Selbststudium
- ▶ Beispielen in der Vorlesung

Aussagekräftige Beispiele zu Algorithmen und Verfahren

- ▶ sind zeitaufwendig von Hand durchzurechnen
- ▶ haben in der Regel viele Verfahrensschritte und Parameter
- ▶ sollen realistische Problemgrößen darstellen

- ▶ Steht für **V**isual **I**nteractive **T**yped **A**pplicative **L**anguage

Vital

- ▶ Worksheetartig
- ▶ Programmierumgebung (Untermenge von Haskell)
- ▶ Nahe an der mathematischen Formulierung
- ▶ Integrierte Graphikprimitive
- ▶ Internetfähig
- ▶ Frei verfügbar

Vital Beispiel

The screenshot shows a software interface with a menu bar (File, Edit, View, Options, Window, Help) and a toolbar with buttons for New, Open, Save, Copy, Paste, Undo, Back, Origin, PV, Print, Stop, Eval, and Anim. Below the toolbar is an 'Item' field containing the code `[fib i | i <- [1 ..]]`. To the right of the 'Item' field are buttons for 'Done', 'Cancel', and 'Delete item'. Below the 'Item' field is a 'Style' field with a checkmark icon. The main area is a yellow code editor containing the following code:

```
□ fib i = case i of
  {
    1 -> 1;
    2 -> 1;
    otherwise -> fib (i-1) + fib (i-2)
  }
```

Below the code editor is a red square bullet point followed by the code `[fib i | i <- [1 ..]]` and a list of Fibonacci numbers: `[1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, 377, 610, 987, 1597, 258]`. At the bottom left, there is a status bar with the text 'Save current module'.

Beispiele der Implementierung in Vital

1. Kryptographie: **Euklidischer Algorithmus**
2. Bioinformatik: **UPGMA**, hierarchisches Clustern von Expressions- oder Aberrationsprofilen
3. Statistische Lerntheorie: VC Dimension

Euklidischer Algorithmus

Zur Berechnung des **ggT** zweier Zahlen r_0 und r_1 :

- ▶ Berechne alle positiven Reste $r_i > 0$ der Divisionen von r_{i-2} durch r_{i-1} für $i = 2, \dots$

$$r_i = r_{i-2} \bmod r_{i-1}$$

- ▶ Der ggT ist der letzte positive Rest r_n

- ▶ **Phylogenetische Rekonstruktion**
- ▶ Hypothesengenerierung auf Basis von Genexpressionsdaten
 - ▶ Identifikation von Gruppen co-regulierter Gene
 - ▶ Identifikation von Tumoruntergruppen

Hierarchisches Clustern

Agglomerativer Basisalgorithmus

1. Initial bildet jedes Objekt (z.B. Sequenz, Profil) ein eigenes Cluster.
2. Man berechnet die (mittlere) paarweise Distanz aller Cluster.
3. Danach faßt man die beiden Cluster zusammen, die die kleinste Distanz zueinander haben.
4. Die beiden letzten Schritte werden wiederholt, bis alle Objekte in einem Cluster sind.

Vital Beispiel Clustering

Anwendungen Orte System Vital

File Edit View Options Window Help

New Open Save Copy Paste Undo Back Origin PV Print Stop Eval Anim

Item: --
Beispiel: UPGMA

Style: [comment1] Done Cancel Delete item

M_Cluster_Ex2

- Beispiel: UPGMA
- M_Tree
- M_BuildTree
- M_Utl
- M_Aggl
- import Styles
- import M_Tree
- import M_BuildTree
- import M_Utl
- import M_Aggl

d = [[['S', 'c', 'h', 'i', 'm', 'p', 'a', 'n', 's', 'e'], ['Z', 'w', 'e', 'r', 'g', 'c', 'h', 'i', 'm', 'p', 'a', 'n', 's', 'e'], ['G', 'o', 'r', 'i', 'l', 'a'], ['M', 'e', 'n', 's', 'c', 'h'], ['O', 'r', 'a', 'n', 'g', 'l', 'i', 't', 'a', 'n']]]

"Schimpanse"	0.0	0.0118	0.0427	0.0182	0.0953
"Zwergschimpanse"		0.0	0.0416	0.0327	0.0916
"Gorilla"			0.0	0.0371	0.0965
"Mensch"				0.0	0.0928
"Orang Utan"					0.0

printPhyloTree d 5 4

```
graph TD
    Root((0.0940)) --- OrangUt[Orang Utan]
    Root --- Node1((0.0404))
    Node1 --- Gorilla[Gorilla]
    Node1 --- Node2((0.0354))
    Node2 --- Mensch[Mensch]
    Node2 --- Node3((0.0117))
    Node3 --- Schimpanse[Schimpanse]
    Node3 --- Zwergschimpanse[Zwergschimpanse]
```

Switch to module "M_Cluster_Ex2"

Posteingang für Michael.Arnold@in... [michael@ubuntu: ~] Vital Adobe Reader - vital270106.pdf

Zusammenfassung

- ▶ Beschränkte Ausführungsgeschwindigkeit (Haskell interpretiert in Java in VM)
- ▶ Lazy evaluation, d.h. 'unendliche' Datenstrukturen möglich
- ▶ Beliebig großes Worksheet
- ▶ Kostenfrei
- ▶ Integration der vorgestellten Laborumgebung in das nächste Vital Release

Zugang: <http://www.informatik.uni-ulm.de/ni/staff/HKestler/vital/>