Praktikum Datamining und Sequenzanalyse

Einführung

Marvin Meusel – marvin.meusel@uni-jena.de
Bertram Vogel – bertram.vogel@uni-jena.de

Organisatorisches

Vorträge

Zu bestimmten Terminen im SR 3423

freies Arbeiten

Montags und Freitags im Linuxpool

`bio.informatik.uni-jena.de/lehre/`

Ziele

- Teamarbeit
- Implementierung von Algorithmen
- Unterschied zwischen Theorie und Praxis
- Arbeiten unter Linux

Themen

Exakte Suche

naive Suche, KMP, Boyer-Moore

Alignments

globales/lokales Alignment, Alignments mit linearem Speicher

Clustering, phylogenetische Bäume

UPGMA, WPGMA

Aufgaben

- Programmieren
  - „schön programmieren“ - OOP
  - Benutzerinterface
  - Dokumentation
  - Tests
- Auswertung
- Vorträge

Tools

- Objektorientierte Programmierung mit **`Java`**
- Dokumentation mit **`Javadoc`**
- Arbeiten in einer **`Linux`** Umgebung
- Versionskontrolle mit **`git`**
- Projektmanagement mit **`Gradle`**
- Unit-Test mit **`JUnit`**
- Benutzerinterface per **`GUI`** oder **`CLI`**

Auswertung und Vortrag

Auswertung

- Aufgabenzettel bearbeiten
- Zeitmessungen
- auf mögliche Fehler eingehen

Vortrag

- Jede Gruppe trägt einmal vor (Dauer etwas 45 min)
  - Vorstellung und Diskussion der Ergebnisse
  - Vorführen der Benutzerschnittstelle
  - Details zur Implementierung

Gruppen

- 2-3 Personen pro Gruppe
- jede Gruppe bekommt eine eigene Linux-Gruppe biodmXY
- jede Gruppe bekommt ein gemeinsames git-Repository /home/stud/bipr/biodmXY

Linux

Linux bedeutet Vielfalt

verschiedene Distributionen

Mint	Ubuntu	Debian	Mageia
Fedora	openSUSE	Arch	CentOS
Zorin	FreeBSD	Knoppix	Red Hat

...

Desktopumgebungen

KDE	Gnome	LXDE	Xfce
MATE	Unity	Cinnamon	

Verzeichnisstruktur

UNIX, Linux, ZETA

- / [Wurzelverzeichnis]
  - bin/
  - lib/
  - etc/
  - dev/
    - sda1 [Gerätedatei]
    - sda2 [Gerätedatei]
  - home/
    - benutzername/
    - anwendername/
      - Office Dokument.odt
      - Foto.png
      - Verzeichnis/
  - media/
    - disk1/ [Einhängpunkt für Datenträger]
    - usbdisk1/ [Einhängpunkt für Datenträger]
      - tabelle.ods
      - Bild.jpg
      - Text.txt
      - Unterordner/
  - usr/

Mac OSX

- / [Wurzelverzeichnis]
  - bin/ [versteckt]
  - System/
  - etc/ [versteckt]
  - dev/ [versteckt]
    - disk [Gerätedatei]
    - disk2 [Gerätedatei]
  - Users/
    - benutzername/
    - anwendername/
      - Office Dokument.odt
      - Foto.png
      - Verzeichnis/
  - Volumes/
    - macosx/ [Einhängpunkt für Datenträger]
    - usbmedium/ [Einhängpunkt für Datenträger]
      - tabelle.ods
      - Bild.jpg
      - Text.txt
      - Unterordner/
  - Applications/
  - usr/ [versteckt]

Windows NT/2000/XP

- C:\ [Laufwerk]
  - Programme\
  - Dokumente und Einstellungen\
    - Adminstrator\
    - Benutzername\
      - Office Dokument.odt
      - Foto.png
      - Verzeichnis\
  - Windows\
    - System32
- E:\ [Laufwerk]
- F:\ [Laufwerk, USB-Medium]
  - tabelle.ods
  - Bild.jpg
  - Text.txt
  - Unterordner\

„Filesystem“ von Thomei08 (Stefan Hagen, Bichwil) – de.wikipedia.org

Besonderheiten der Shell

Hilfe zu Befehlen

`info, man <Befehl>`	Hilfe zu einem Befehl
`<Befehl> --help`	kurze Hilfe

Completion

Name-Completion mit Tab-Taste

Befehle

`ls`	Anzeige des Verzeichnisinhalts
`cd`	Wechseln des Verzeichnisses
`mkdir`	Erzeugen eines Verzeichnisses
`rmdir`	Löschen eines leeren Verzeichnisses
`rm`	Löschen einer Datei
`rm -r`	Löscht rekursiv einen nicht leeren Ordners
`mv`	Verschieben
`cp`	Kopieren
`cat / less`	Anzeigen von Dateien
`grep`	Suchen in einer Datei
`find`	Suchen einer Datei
`df -h`	Anzeigen der Festplattenbelegung

Benutzerrechte

- Jede Datei und jedes Verzeichnis ist einem Eigentümer und einer Gruppe zugeordnet.
- verschiedene Rechte für Eigentümer, Gruppe und andere
- Anzeigen der Rechte mit z.B. `ls -la`

`chmod`	Setzen der Dateirechte
`chown`	Ändern des Eigentümers
`chgrp`	Ändern der Gruppe
umask	Setzen der Standardrechte für neue Dateien

Java und OOP

Bitte nicht so!

```
public class Class1{

    public static void save(String path, double score){...}
    public static double[][] load(String path){...}
    public static double algo1(double[][] data){...}
    public static double algo2(double[][] data){...}

    public static void main(String[] args){
        double[][] data = load(args[0]);
        double score1 = algo1(data);
        double score2 = algo2(data);
        save(args[1],score1);
        save(args[2],score2);
    }

}
```

OOP

Prinzipien

- Abstraktion
- Datenkapselung
- Polymorphie
- Vererbung

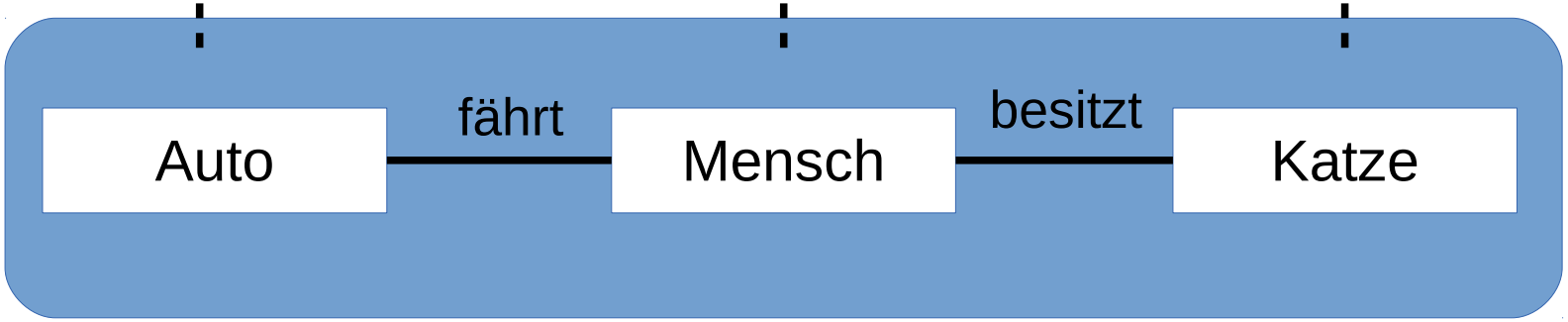
Die OOP Welt

- Klasse
- Objekt
- Attribute
- Methoden

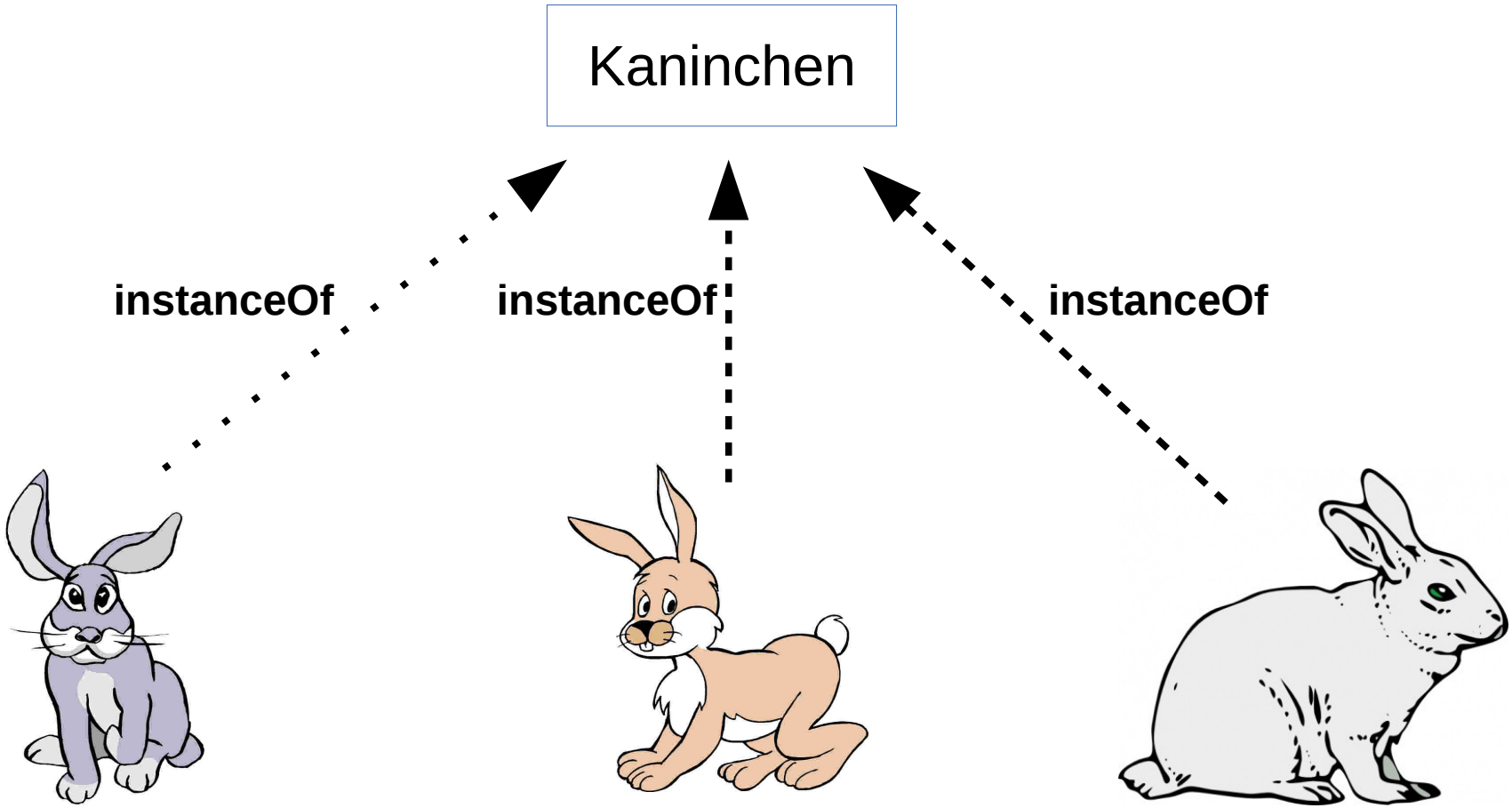
Grundlage

Realität

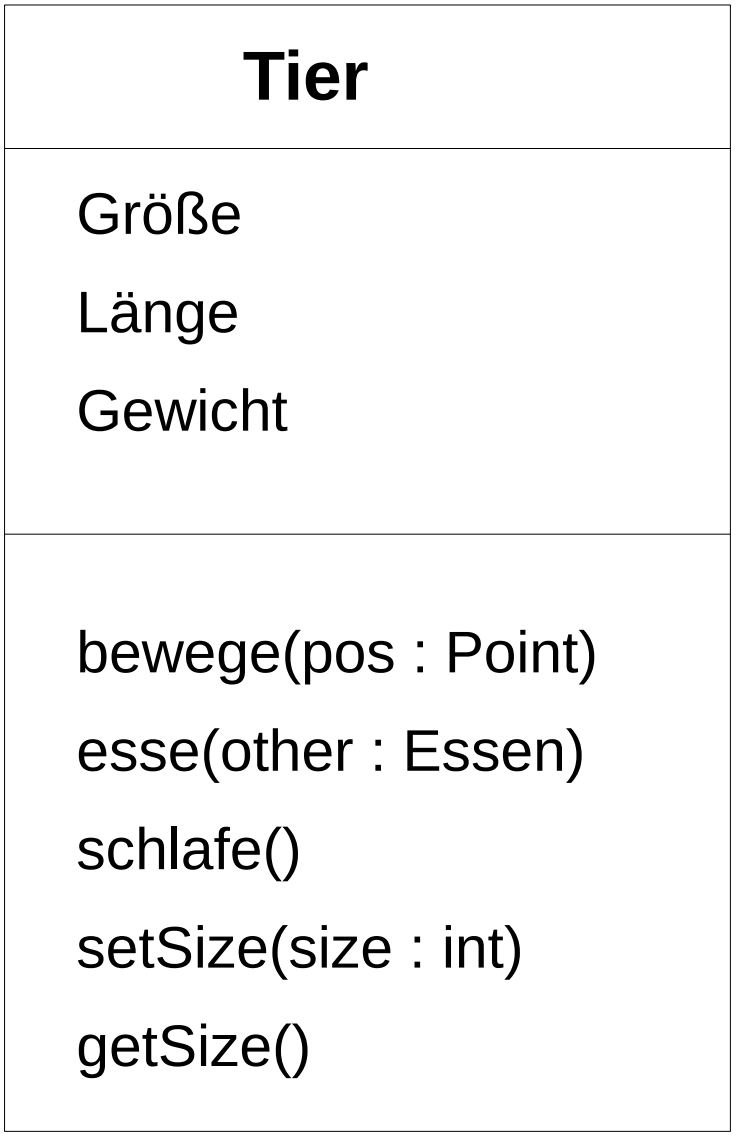
Modell



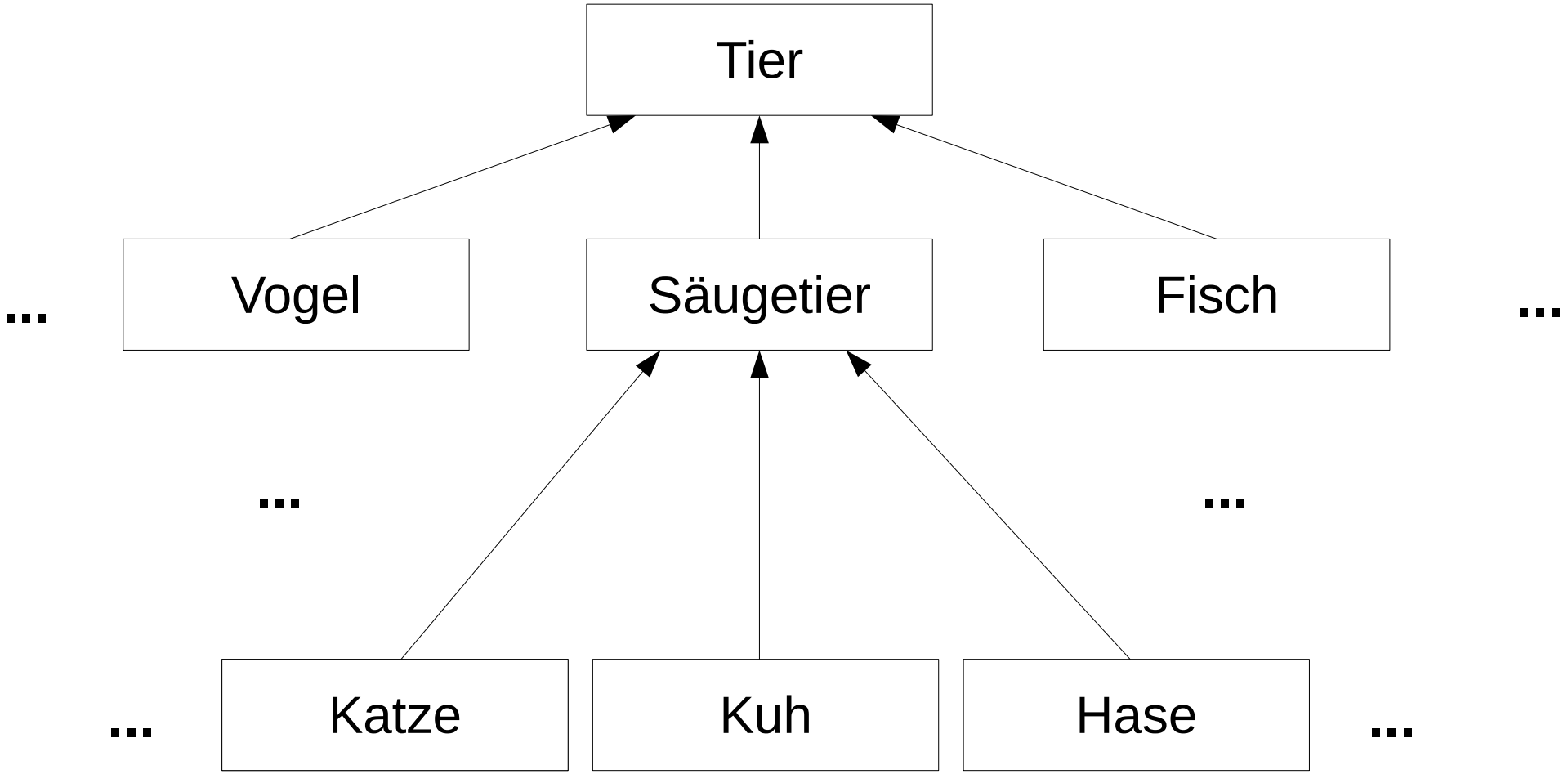
Klassen und Objekte



Attribute und Methoden



Vererbung



Abstrakte Klassen und Interfaces

```
public abstract class Tier{
    public abstract void move();
}

public interface CanFly{
    public void fly();
}

public class Bird extends Tier implements CanFly{
    public void move(){
        //do something
    }
    public void fly(){
        //do something
    }
}
```

Polymorphismus dynamisch/statisch

```
public class Clock{

   public void setTime(long ns){...}
   public void setTime(int h, int m, int s, int ms){...}

}

public class BetterClock extends Clock{

   @Override
   public void setTime(long ns){...}

}
```

Sichtbarkeit

Modifier	Class	Package	Subclass	World
public	✓	✓	✓	✓
protected	✓	✓	✓	✗
default	✓	✓	✗	✗
private	✓	✗	✗	✗

Hinweise

Zeit messen

via command line time

```
#$ time sleep 10
real 0m10.116s
user 0m0.001s
sys 0m0.007s
```

via Java

- System.currentTimeMillis()
- System.nanoTime()

```
long time = System.nanoTime();
//do something
long duration = System.nanoTime() - time;
```

Zeit messen

Vorgehen

- mehrfach messen
- Minimum aller Messungen verwenden

Testumgebung beschreiben

- Betriebssystem
- Systemspeicher
- CPU
- Java VM version
- Heapspace

IO

```
try(BufferedReader reader =
          new BufferedReader(new FileReader(path))){
   String temp = null;
   while((temp = reader.readLine()) != null){
      //werte aus
   }
}catch(IOException e){
   // ...
}
```

Links

distrowatch.com

galileo-press.de/openbook

ubuntuusers.de

wiki.ubuntuusers.de/Shell/Befehlsübersicht