Einführung in die Informationsverarbeitung

Øyvind Eide

Woche 9
Der Informationsbegriff
Datenstrukturen

oeide@uni-koeln.de
http://idh.uni-koeln.de

Was ist ein Informationsmodell?

●● + ● = ● ● ●

●● + ● ● = ● ● ● ●

●●● + ●● = 

Herzliche Glückwünsche, Sie lesen Maya.

Was ist ein Informationsmodell?

III == 3 == γ‘ == ●●●

Vier Repräsentationen eines Konzepts.

Der Informationsbegriff:

Häufige Missverständnisse 1

Universität zu Köln
Digital Humanities – Historisch-Kulturwissenschaftliche Informationsverarbeitung
Prof. Dr. Eide

Kurz über echte Informationen und Bytes.

Warum Byte-Zählen kann Wahnsinn sein.

Information

1. Es ist schön in Köln zu Ihnen zu sprechen.
2. Es ist schön in Kln zu Ihnen zu sprechen.
3. Es ist schön in Kön zu Ihnen zu sprechen.
4. Es ist schn in Kön zu Ihnen zu sprchen.
5. S'is schö in Kölle zu Ihnen zu sprchen.

Information

Luat enier sidtue an eienr elgnhcsien uvrsnäiett, ist es eagl in wcheler rhnfgeeloie die bstuchbaen in eniem wrot snid. Das eniizg whictgie ist, dsas der etrse und der lztete bstuchbae am rtigeichn paltz snid. Der rset knan tatol deiuranchnedr sien und man knan es ienrmomch onhe porbelm lseen. Das legit daarn, dsas wir nhcit jeedn bstuchbaen aeilln lseen, srednon das wrot als gzanes.


contex.com

Information

Claude Shannon: "A Mathematical Theory of Communication", Bell System Technical Journal, 1948.

Enthält eine quantitative Definition von Information.

Zweck: Wie kann ein Signal zwischen einem Sender und einem Empfänger mit dem geringstmöglichen Aufwand "korrekt" übertragen werden.

Information

Shannon hat also eine technische Definition von "Information", die die Bedeutungsebene völlig ausklammert. Er betont, dass "die semantischen Aspekte der Kommunikation für die ingenieurwissenschaftliche Seite irrelevant sind".

Information

Nahezu alle mir bekannten Lehrbücher der Informatik beginnen mit Shannons Definition der Information.

Räumen aber ein, dass die ingenieurwissenschaftliche Definition von Information defizitär sei. Insbesondere:

Information

(1) Bildet sie nur einen Teil des intellektuellen Umfanges des Konzepts "Information„ (syntaktisch, semantisch, pragmatisch) ab.

(2) Gibt es dagegen keine operable Definition die dieses Konzept in seiner vollen Breite abdeckt.

Aber: Die vorhandenen Konzepte reichen aus, um Information auf einem Digitalrechner so darzustellen, dass man sie sinnvoll verarbeiten kann.

Begriffe rund um die “Information”

Universität zu Köln
Digital Humanities – Historisch-Kulturwissenschaftliche Informationsverarbeitung
Prof. Dr. Eide

Informationsebenen

Syntax: Beziehungen der „Zeichen“ untereinander.

Semantik: Beziehungen zwischen „Zeichen“ und „Gegenständen“.

Pragmatik: Beziehungen zwischen „Zeichen“ und ihren „Benutzern“.

Information und Wissen

Daten sind speicherbare Angaben - 22°C.

Information stellt Daten in einen Kontext:

"In diesem Hörsaal herrscht eine Temperatur von 22°C".

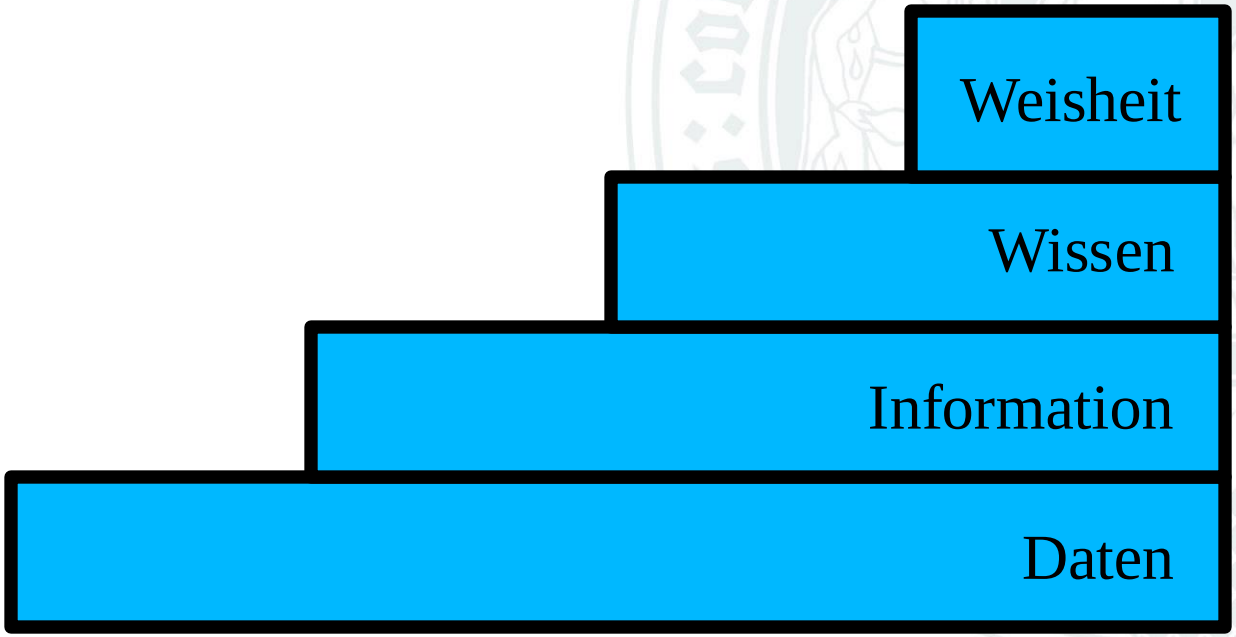
Dieser Kontext ist jedoch noch fest (und für alle Informationsempfänger identisch).

Information und Wissen

Wissen ist das Ergebnis von Erkenntnisprozessen. Es bezieht die praktische Anwendung der Daten und Informationen ein. Es muss nicht "absolut wahr" sein, sondern adäquates Handeln zu ermöglichen.

Z.B. die Entscheidung einen Pulli (nicht) auszuziehen, um sich angenehm zu fühlen, ohne sich zu erkälten.

„Ladder of Knowledge“



Häufige Missverständnisse 2
… und was heisst “Information darstellen?”

Wahrheit

George Bool (1815-1864), "The Laws of Thought". Formales Rechnen mit logischen Ausdrücken, die nur die Zahlenwerte "1" (wahr) und "0" (falsch) kennen.

Aussage: Computer kennen nur wahr oder falsch, keine "Zwischentöne".

Wahrheit

Umgesetzt in Programmiersprachen als:

```
if (aussage == 1) nimm an, dass Aussage wahr.
else nimm an, dass Aussage falsch.
```


Wahrheit

… aber auch als:

```
if (aussage == 0) nimm an, dass Aussage falsch.
else nimm an, dass Aussage wahr.
```

Wahrheit

Z.B. in folgendem Programm:

```
int aussage;
aussage=bewerteZustandDerWelt();
if (aussage == 0) nimm an, dass Aussage falsch.
else nimm an, das Aussage wahr.
```


Wahrheit

Z.B. in folgendem Programm:

```
int aussage;
```

"int" kennzeichnet eine ganze Zahl, die die Zahlenwerte von 0 bis 2^{32}-1 annehmen kann.

```
aussage=bewerteZustandDerWelt();
if (aussage == 0) nimm an, dass Aussage falsch.
else nimm an, das Aussage wahr.
```

Wahrheit

Z.B. in folgendem Programm:

```
int aussage;
```

"int" kennzeichnet eine ganze Zahl, die die Zahlenwerte von 0 bis 2^{32}-1 annehmen kann.

```
aussage=bewerteZustandDerWelt();
if (aussage == 0) nimm an, dass Aussage falsch.
else nimm an, das Aussage wahr.
```

Es gibt also genau eine Art "falsch" auszudrücken, aber 4.294.967.294 Varianten von "wahr"?

Wahrheit

Unabhängig von dem, was George Bool gesagt hat, gilt also in Wirklichkeit nicht

"0" = falsch.
"1" = wahr.

Wahrheit

... sondern in Wirklichkeit:

"0" = falsch.

"1" bis "4.294.967.295" = wahr.

Wahrheit

dann könnte aber genau so gut gelten:

"0" = falsch.
"1" bis "4.294.967.294" = unklar.
"4.294.967.295" = wahr.

Wahrheit

oder genauso willkürlich:

"0" bis "255" = falsch.
"256" bis "4.294.967.039" = unklar.
"4.294.967.040" bis "4.294.967.295" = wahr.

Wahrheit

oder auch:

"0" = sicher falsch.
"20.000" = ziemlich sicher falsch.
"100.000" = ziemlich sicher falsch, aber mit fünfmal größerer Wahrscheinlichkeit richtig als "20.000".
"4.000.000.000" = ziemlich sicher richtig.
"4.294.967.295" = sicher wahr.

Wahrheit – Begriffe

Binäre Logik
Kennt nur "wahr" und "falsch".

n-äre oder mehrwertige Logik
Kennt n Wahrheitswerte.

Kontinuierliche Wahrheitsfunktionen
Nehmen unendlich viele Wahrheitswerte an.

Häufige Missverständnisse 3
… oder: Die Mystik des Digitalen

Universität zu Köln
Digital Humanities – Historisch-Kulturwissenschaftliche Informationsverarbeitung
Prof. Dr. Eide

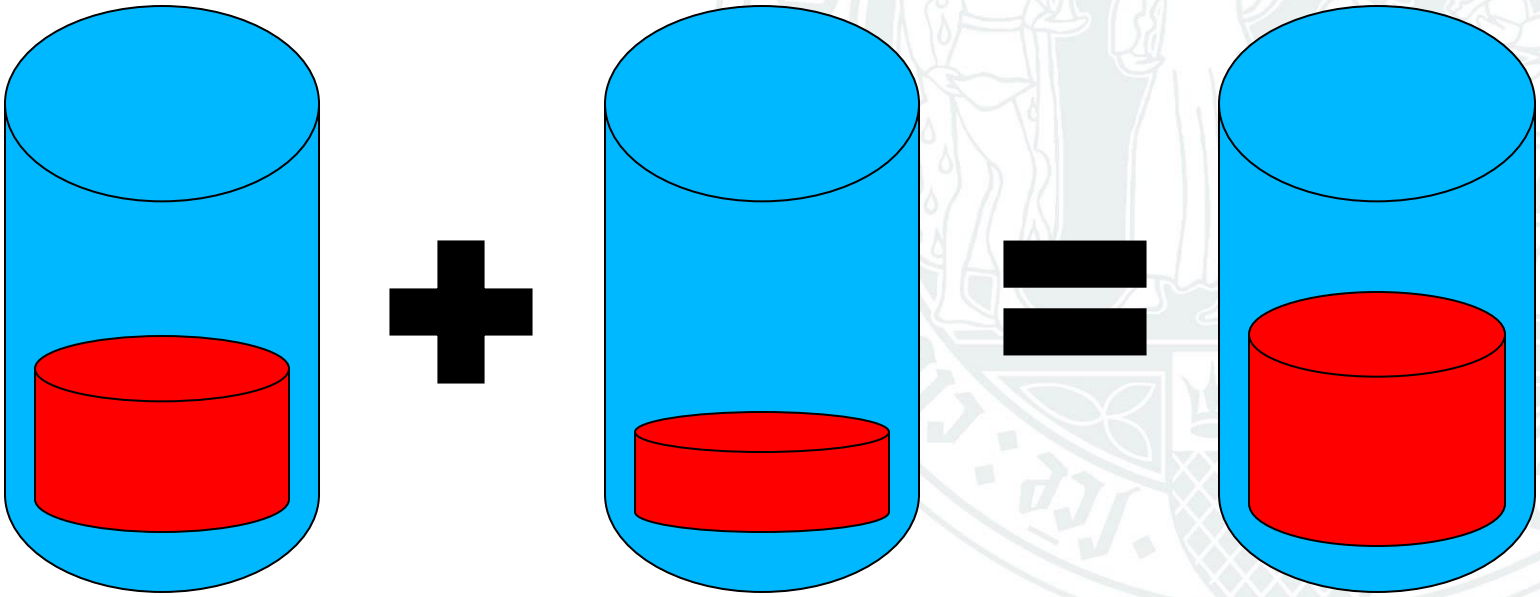
Digital - Analog

Das korrekte Begriffspaar in der Informatik:

„diskret“ v. „kontinuierlich“.

Digital - Analog

1,00134… + 0,48723… = 1,48857…



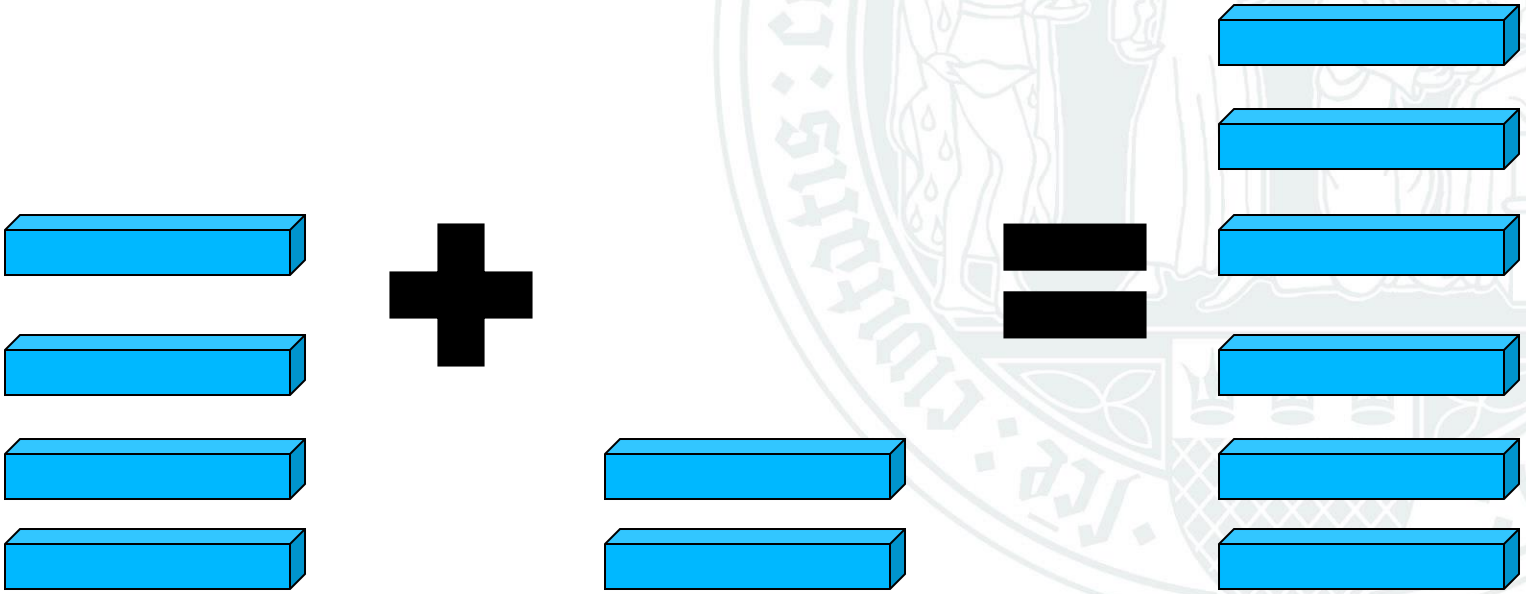
Eine Addition kontinuierlicher Zahlenwerte.

Digital - Analog

Nach diesem Prinzip – heute nur mehr für Spezialanwendungen – verwendete Computer, die z.B. Spannungen oder Stromstärken „addieren“ heißen „Analogrechner“.

Digital - Analog

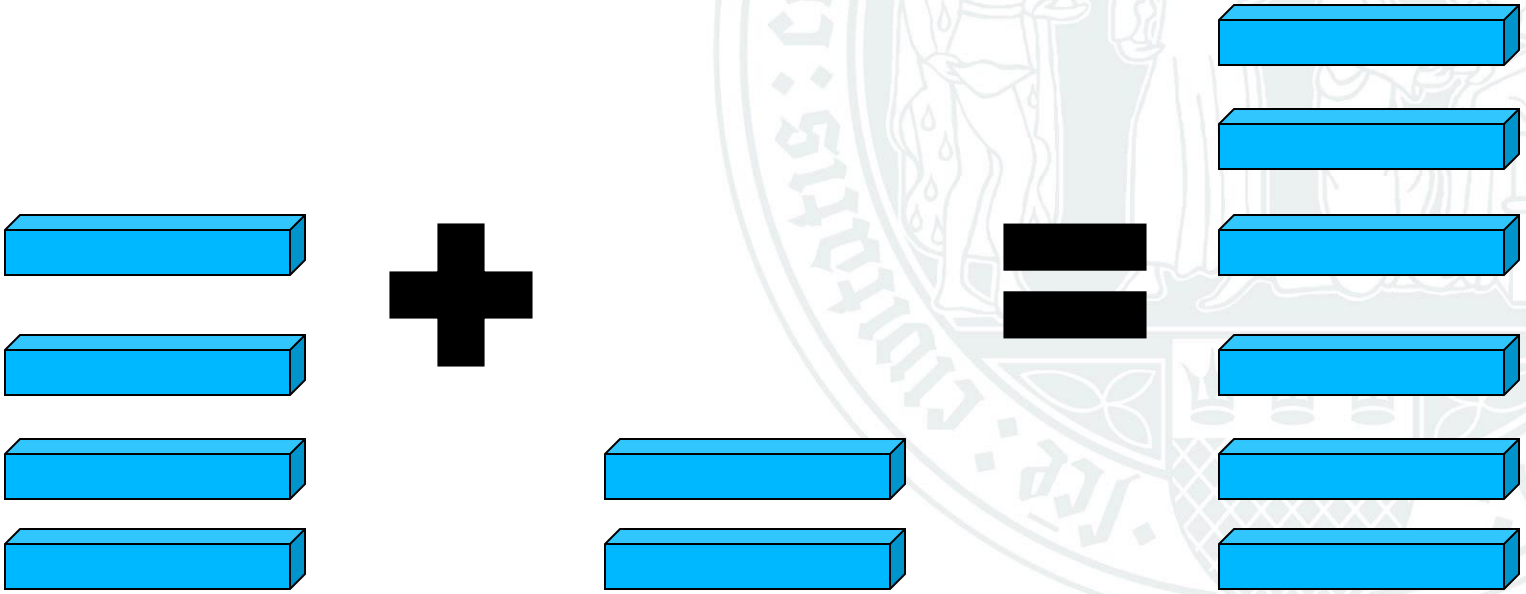
4,00000 + 2,00000 = 6,00000



Eine Addition diskreter Zahlenwerte.

Digital - Analog

0,00004 + 0,00002 = 0,00006



Ebenfalls eine Addition diskreter Zahlenwerte.

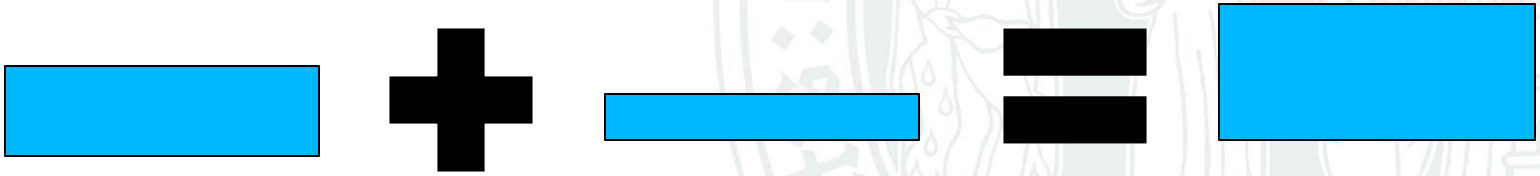
Digital - Analog

0,00004 + 0,00002 = 0,00006

Auch eine Addition diskreter Zahlenwerte.

Digital - Analog

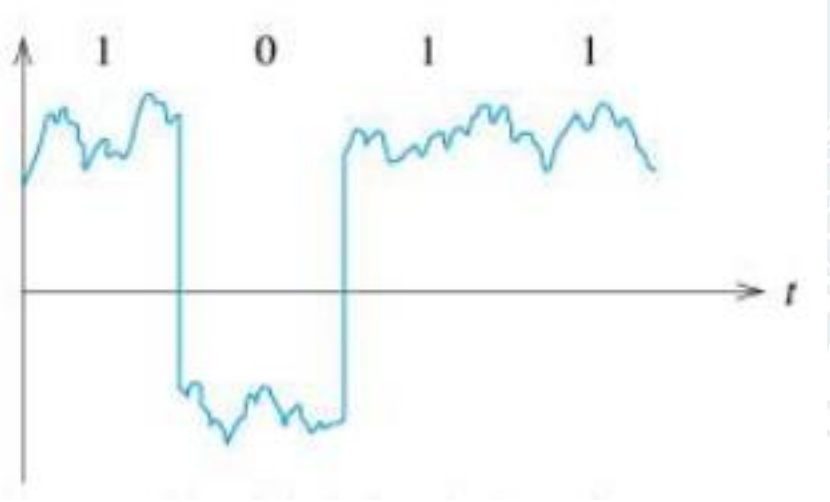
0,00004 + 0,00002 = 0,00006



Kontinuierlich oder diskret?

Digital - Analog

Ob eine Darstellung „analog“ oder „digital“ ist, hängt in erster Linie von der Meßgenauigkeit ab.



"Imagen 4" by Mcanet - Own work. Licensed under CC BY 3.0 via Commons - https://commons.wikimedia.org/wiki/File:Imagen_4.png#/media/File:Imagen_4.png

„Arten“ von Information

1. "Selbstabbildende Information".
   Es kann "gerechnet" werden.
   Bilder.

2. "Kodierte Information".
   Zeichenketten und Teilketten können verglichen werden.
   Texte.

3. "Symbolische Information".
   Terme können verglichen werden.
   Terminologien, "Ontologien" u.ä.

Datenstrukturen
Grundbegriffe

Universität zu Köln
Digital Humanities – Historisch-Kulturwissenschaftliche Informationsverarbeitung
Prof. Dr. Eide

Datentyp

- Zahlen
- Bilder
- Zeichenketten
- Geburtstage
- Briefe

Operationen

- Addieren
- Komprimieren
- Vergleichen
- Abstand berechnen
- Beziehen

Datenstrukturen

- Zahl {Darstellung, Addieren, …}
- Bild {Darstellung, Komprimieren, …}
- Text {Darstellung, Vergleichen, … }
- Zeit {Darstellung, Abstand berechnen, … }
- Brief {Darstellung, Beziehen, …}

Datenstruktur =

{Datentyp, Legale Operationen }

“Datentyp” und “Datenstruktur” oft aber auch synonym!

Basisdatenstrukturen

- Boolean / Logischer Wert
- Integer
- [Rationale Zahlen]
- Realzahlen
- Zeichen
- Zeichenketten

Datenstrukturen und Hardware

Datenstrukturen geben Regeln wieder, wie ein bestimmter Speicherbereich interpretiert wird.

ASCII Zeichen 'a' = 97; 'A' = 65.

oder

'Pixel' 97 ist eineinhalb mal heller als 'Pixel' 65.

Datenstrukturen und Hardware

Festlegungen sind „willkürlich“.

Groß- / Klein v. Umlaute

Zeichen

a	097	A	65
b	098	B	66
c	099	C	67
d	100	D	68
e	101	E	69
f	102	F	70
g	103	G	71
h	104	H	72
i	105	I	73
j	106	J	74
k	107	K	75
…	…	…	…

Zeichen

a	01100001	A	01000001
b	01100010	B	01000010
c	01100011	C	01000011
d	01100100	D	01000100
e	01100101	E	01000101
f	01100110	F	01000110
g	01100111	G	01000111
h	01101000	H	01001000
i	01101001	I	01001001
j	01101010	J	01001010
k	01101011	K	01001011
…	…	…	…

Zeichen

a	01100001	A	01000001
b	01100010	B	01000010
c	01100011	C	01000011
d	01100100	D	01000100
e	01100101	E	01000101
f	01100110	F	01000110
g	01100111	G	01000111
h	01101000	H	01001000
i	01101001	I	01001001
j	01101010	J	01001010
k	01101011	K	01001011
…	…	…	…

Zeichen

Festlegungen sind „willkürlich“.

lower(x) = upper(x) | '00100000'

= schnellste verfügbare Operation des Rechners!

Merke

Darstellung von Datenstrukturen sind „willkürlich“.

… können den Aufwand für eine Anwendung aber entscheidend beeinflussen!

Datenstruktur „Zeiger“

Diagrammatische Darstellung:

A „zeigt auf“ B

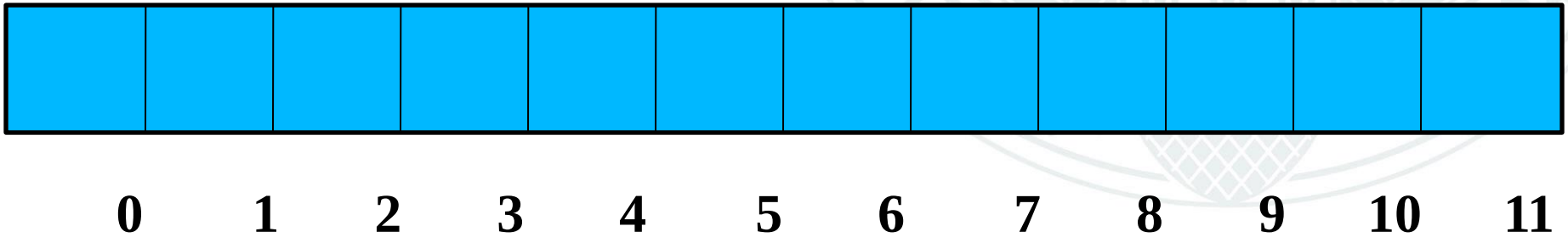
Datenstruktur „Zeiger“

Diagrammatische Darstellung:

„Zeiger“: Ein Speicherinhalt eines Rechners verweist auf einen anderen.

Datenstruktur im Speicher

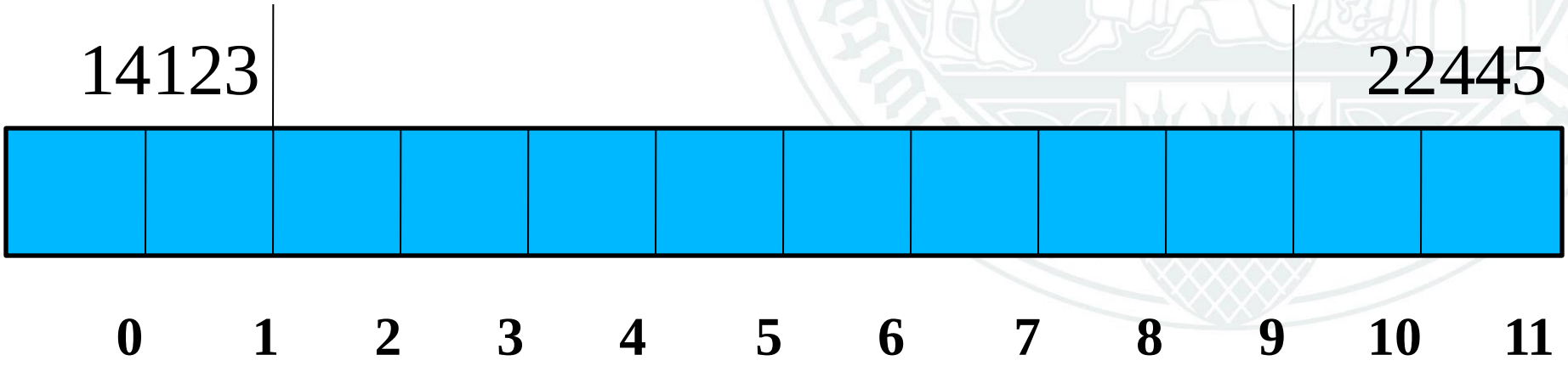
Speicher als „karierte Zeile“



Datenstruktur im Speicher

Zahl „14123“ in Bytes 0 bis 1

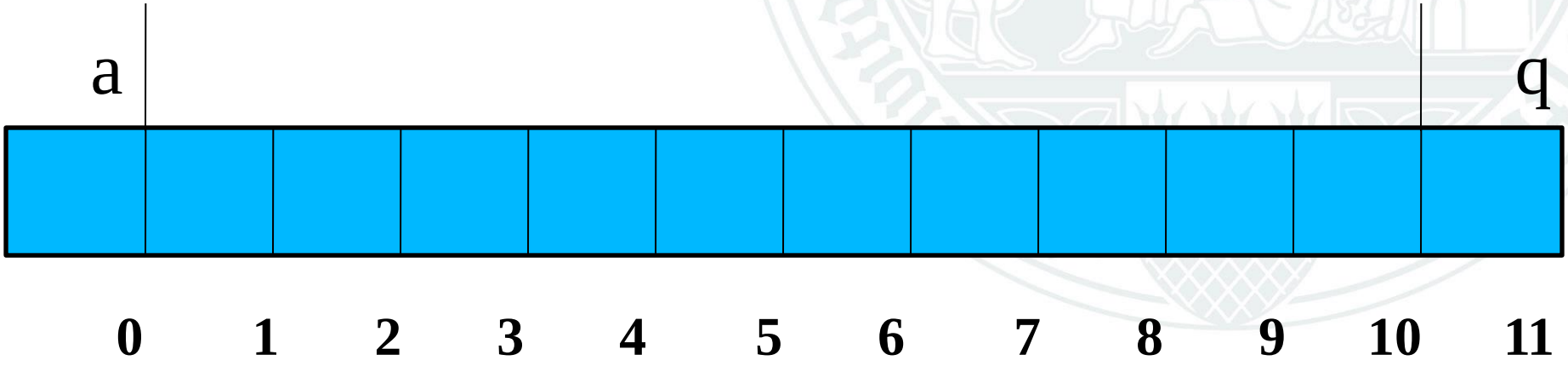
Zahl „22445“ in Bytes 10 bis 11



Datenstruktur im Speicher

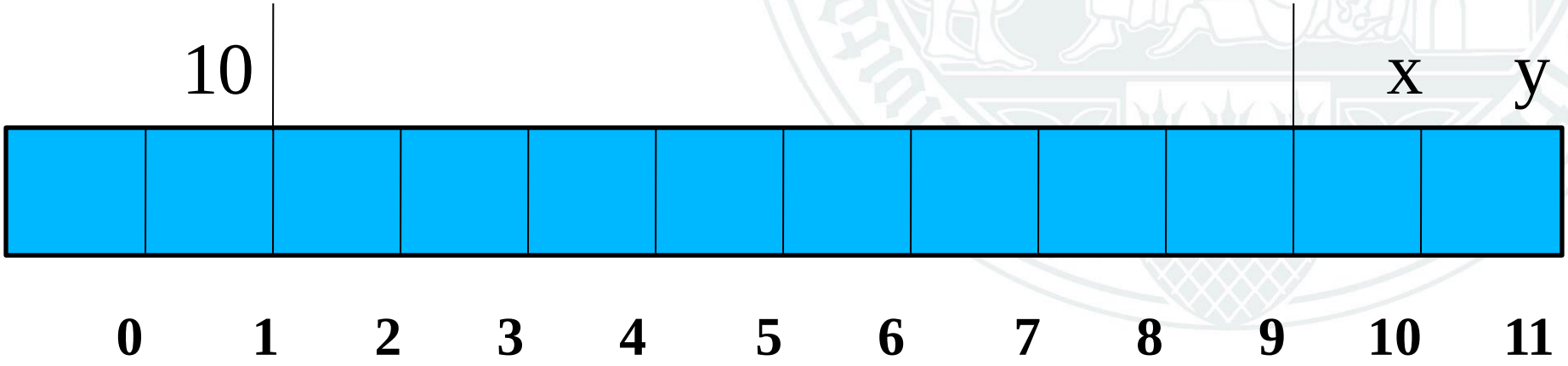
Zeichen „a“ in Byte 0

Zeichen „q“ in Byte 11



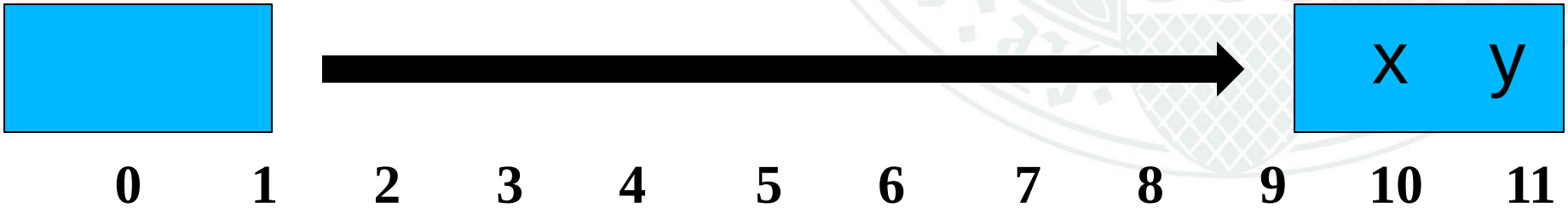
Datenstruktur im Speicher

Zeiger in Bytes 0 bis 1 verweist auf Speicherblock, enthaltend „xy“, beginnend in Byte 10



Zeiger graphisch

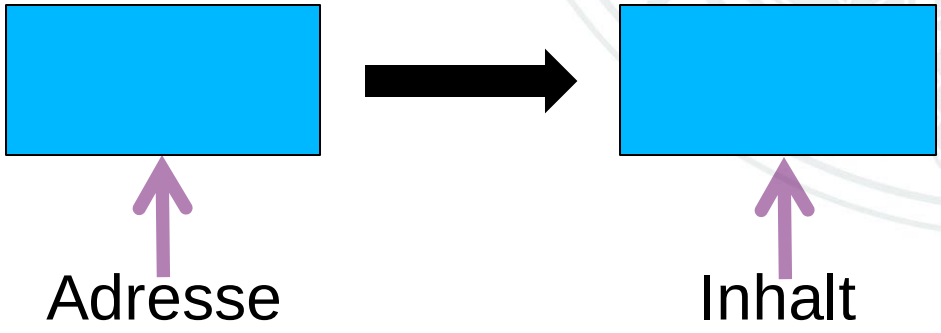
Zeiger in Bytes 0 bis 1 verweist auf Speicherblock, enthaltend „xy“, beginnend in Byte 10



Zeiger graphisch

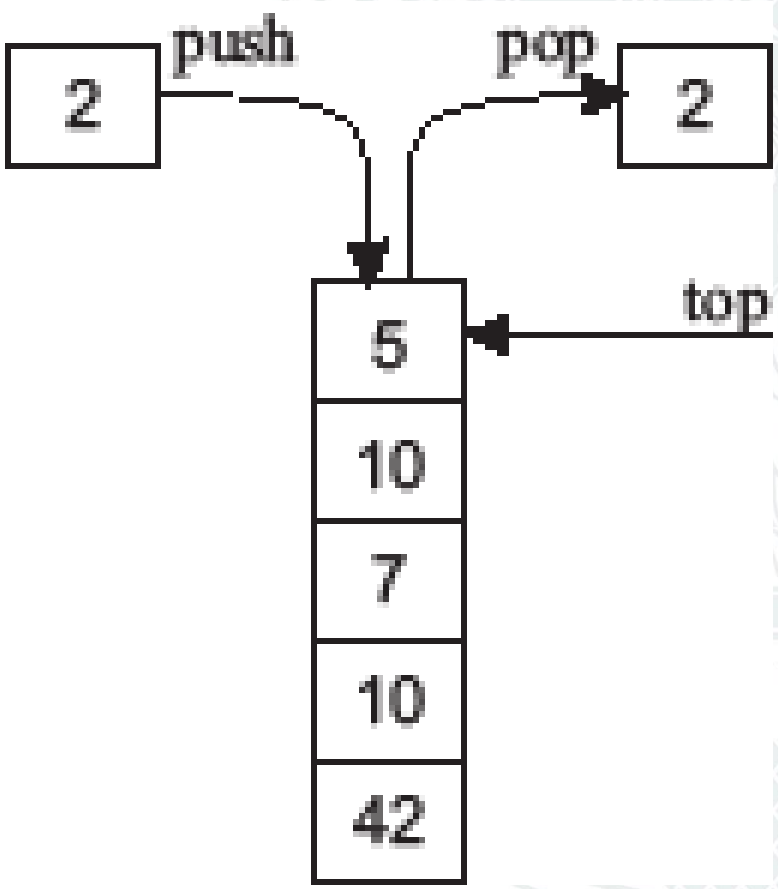
Zeiger verweist von einem Datenblock auf einen anderen.

Inhalt im Datenblock: Kontext-basierte Interpretation



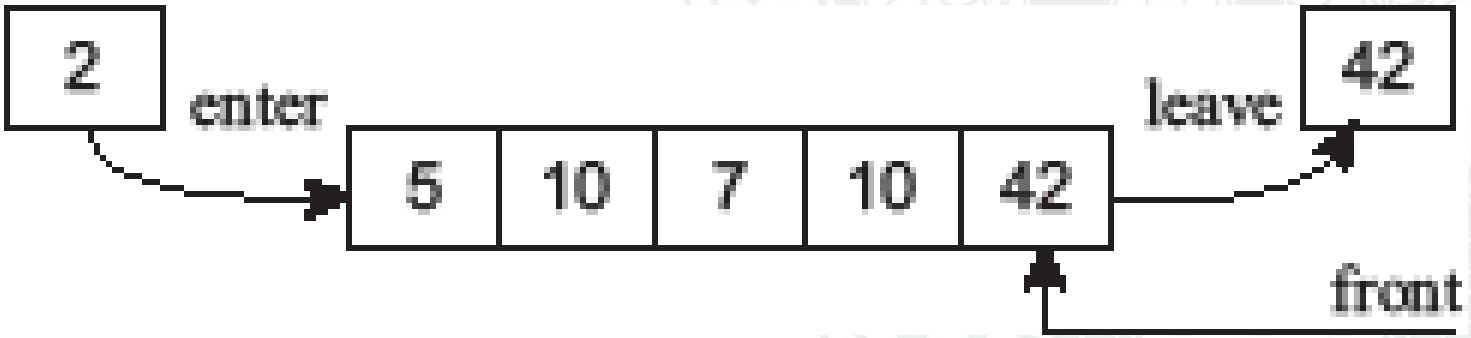
Technische Datenstrukturen

Stacks



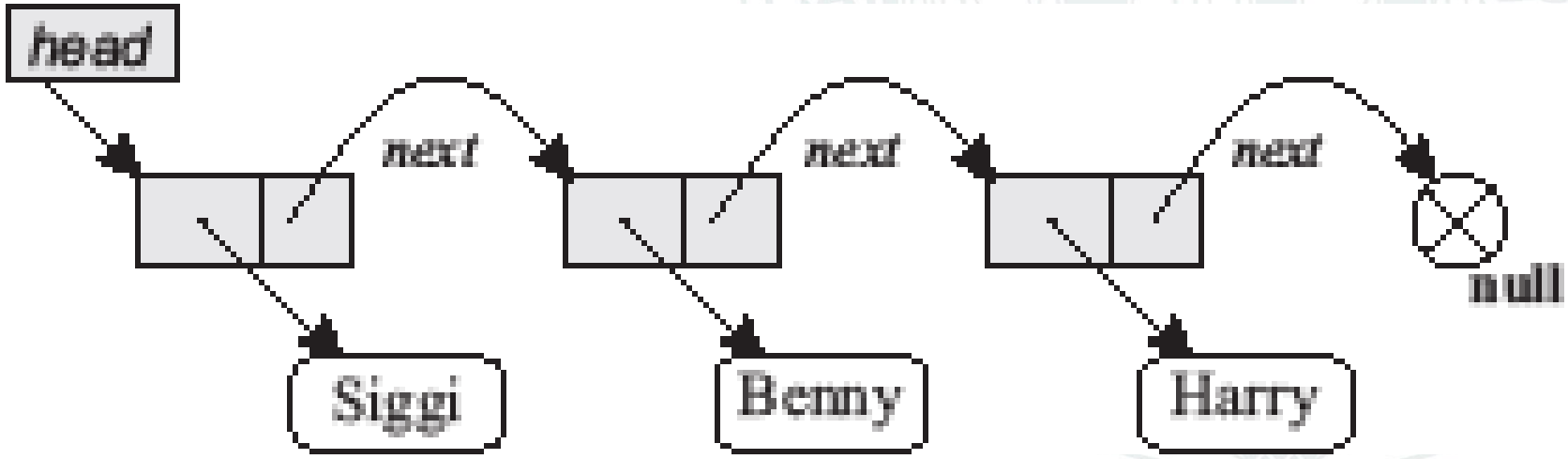
Auch bekannt als: „LIFO“ – Last In, First Out

Queues

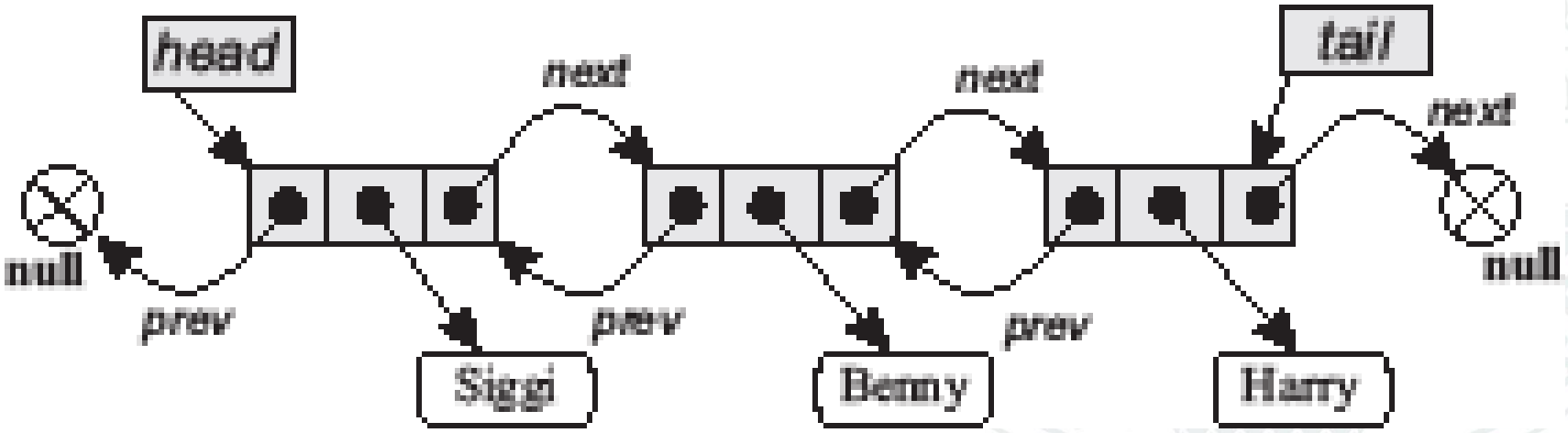


Auch bekannt als: „FIFO“ – First In, First Out

Einfach Verknüpfte Listen



Doppelt Verknüpfte Listen



Inhaltliche Datenstrukturen

Datentypen allgemein

<Datentyp> ::=
ein 3-Tupel (oder Tripel) **{ E, I, O }**
wobei
E ::= Externe Darstellung
I ::= Interne Darstellung
O ::= Menge auf I definierter Operationen

(Notation: " ::= " = "definiert als ")

Datentyp Zeit allgemein

E Regel für "4.6.2007"

I Zeit ist ein Vektor von Tagen seit einem willkürlichen Tag 0, wobei ein beliebiger Tag i als Offset t vom Ursprung definiert ist.

O

t-less(i,j) ===> **Boolean**

t-less(4.6.2007,5.6.2007) ===> **True**

t-subtract(i,j) ===> **Ganze Zahl**

t-subtract(5.6.2007,4.6.2007) ===> 1

Datentyp „Historische Zeit“ I

E Regel für "pri non jun 2007"

I Zeit ist ein Vektor von Tagen seit einem willkürlichen Tag 0, wobei ein beliebiger Tag i als Offset t vom Ursprung definiert ist.

O

t-less(i,j) ===> **Boolean**

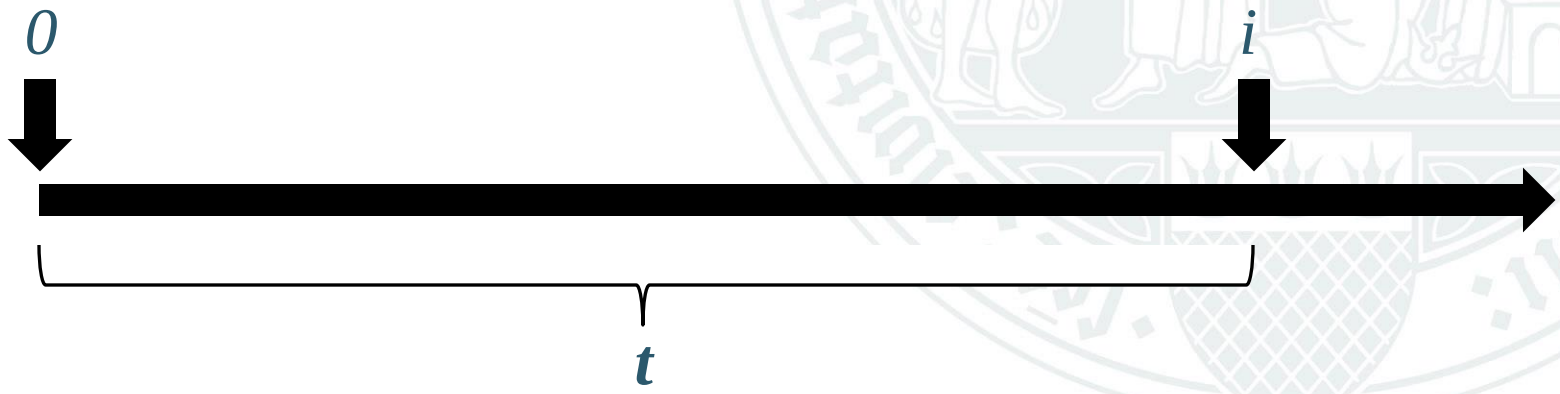
t-less(pri non jun 2007,non jun 2007) ===> **True**

t-subtract(i,j) ===> **Ganze Zahl**

t-subtract(non jun 2007,pri non jun 2007) ===> 1

Datentyp „Historische Zeit“ I

I Zeit ist ein Vektor von Tagen seit einem willkürlichen Tag 0, wobei ein beliebiger Tag *i* als Offset **t** vom Ursprung definiert ist.



Datentyp „Historische Zeit“ II

E Regel für "6 Tammuz 5763"
I Zeit ist ein Vektor von Tagen seit einem willkürlichen Tag 0, wobei ein beliebiger Tag i als Offset t vom Ursprung definiert ist.

O
t-less(i,j) ===> **Boolean**
t-less(6 Tammuz 5763,7 Tammuz 5763) ===> **True**

t-subtract(i,j) ===> **Ganze Zahl**
t-subtract(7 Tammuz 5763,6 Tammuz 5763) ===> 1

Datentyp „Historische Zeit“ II

I Zeit ist ein Vektor von Tagen seit einem willkürlichen Tag 0, wobei ein beliebiger Tag *i* als Offset **t** vom Ursprung definiert ist.