

INFO221v11

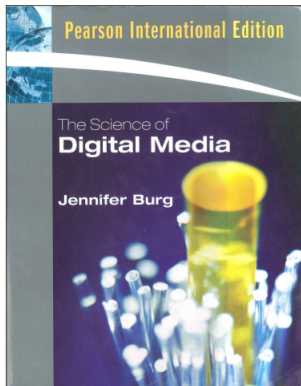
“Hei!”

Truls Pedersen
Institutt for informasjons- og medievitenskap
Universitetet i Bergen

INFO221v12

Multimediabehandling vår 2012

Foreleser:	Truls Pedersen (tpe056@uib.no, rom 541)
Stud.ass:	Øyvind Døskeland (oyvind.doskeland@gmail.com) Snorre Davøen (sda087@student.uib.no)
Forelesning:	15 stk; onsdager kl. 10:15 Bjørn Chr. hus, Rom 135 (01.02.12: 555)
Gruppe/lab:	U. Pihl (205 & 215)
Oblig. oppg:	Tre stk. (hhv. 15%, 15%, 30%)
Eksamen:	Skriftlig (40%)



Hele* boken.

Pensum

Information Retrieval 1

The screenshot shows a web browser window displaying the 'Multimedia Information Retrieval Systems' book frame. The browser's address bar shows the URL 'www.nordbotten.com/ADM/ADM_book/MIRS-frame.htm'. The page has a header with the title 'Multimedia Information Retrieval Systems' and the author 'Joan C. Nordbotten'. Below the header is a navigation bar with links: TOC, IR, Data modelling, MDR Design, SQL3, Image Retrieval, and Glossary. The main content area is divided into two columns. The left column contains the 'Preface' section, which discusses the second version of the book, its development as a textbook, and the technical advances in data collection and dissemination. The right column contains the 'Contents' section, which lists the book's structure: Preface, Using this book, 1. Information Retrieval (with sub-sections: Concepts and Definitions, Data representations, Data collections and databases, Media Applications), and 2. Data Management Systems (with sub-sections: Data Management Functions, Information & data retrieval, The Evolution of Data Management Systems).

Multimedia Information Retrieval Systems
Joan C. Nordbotten

TOC TOF IR DMS Data modelling Media modelling MDR Design indexing SQL3 Text Retrieval Image Retrieval MMIRS Glossary Bibliography

Preface

This second version of Multimedia Information Retrieval Systems extends the previous with material on the management of multiple, multiple media database systems. MDRS: This Web-book (in its various versions) has been developed for use as a textbook for graduate level courses in Advanced Data Management that have been held since 2000 at the Dept. of Information and Media Science, University of Bergen, Norway and at the Dept. of Information and Computer Sciences, University of Hawaii at Manoa, USA.

A number of technical advances have made possible the collection and dissemination of increasing quantities of semi- and unstructured data representing various media objects such as: text, images, maps, audio (speech & music) streams and combined media as video (film).

First, the increased capacity of data storage devices has made feasible development of large collections (databases) representing single media data, such as text, images, audio, video, or combinations thereof.

Second, the development and general availability of digitalization devices has made scanning of existing material, as well as capture of new media data, feasible for a wide range of applications serving both individuals and organizations.

Third, the increased bandwidth in communication networks supports transportation of large quantities of data within reasonable time frames.

Finally, the expansion of the Internet has made feasible dissemination of large quantities of multiple media data to worldwide audiences for business, education and entertainment.

Most data collections are still administered by data management systems specialized for the particular type of data in the collection. For example:

- *Relational database management systems, R-DBMS*, have been constructed to manage the structured data typically found in administrative systems. These have been extended to allow storage of large media objects and their related descriptive data.
- *Information retrieval systems, IRS*, are used for management of unstructured and semi-structured representations of full-text documents and support similarity search within text

Contents

Preface
Using this book

1. Information Retrieval

1. Concepts and Definitions
 1. Data representations
 - Media data
 - Digital media
 2. Data collections and databases
 - Media databases
 - Web databases
2. Media Applications
 1. Administrative applications with Media Data
 2. Text-based applications
 3. Image-based applications
 4. Streamed data applications
 5. Multiple Media applications

2. Data Management Systems

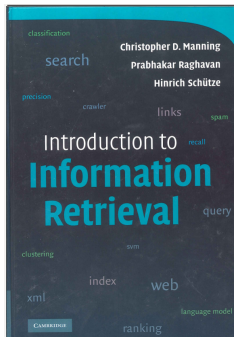
1. Data Management Functions
 1. Database design & creation
 2. Information & data retrieval
2. The Evolution of Data Management Systems
 1. Administrative database management systems
 2. Text-based Information Retrieval Systems
 3. Image retrieval systems

www.nordbotten.com/ADM/ADM_book/MIRS-frame.htm

Kapitler : 1 - 4, 6

Pensum

Information Retrieval 2



Kapitler: 1 - 4, 6 - 9

Multimediabehandling og gjenfinning

Motivasjon



...8A 2B 23 29 CA D9 83 23 0D 92 B9 FD 89...

Det er ikke lett å ha en vag idé om et stykke musikk og søke gjennom mp3 filer som “ligner”.

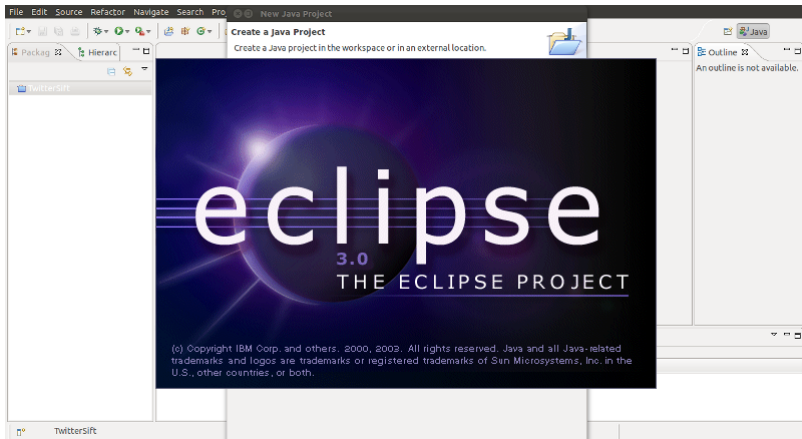
Hva kan vi fra før?

- ▶ programmere
- ▶ databaser
- ▶ enkel matematikk
 - ▶ og vektorer

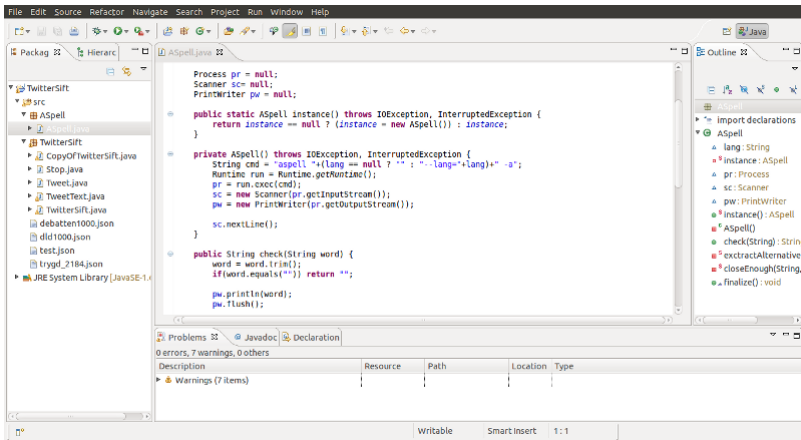
Programmering

```
public static void main(String[] args) {  
    int w = 10, h = 6, n = w*h;  
  
    for (int i = 0; i < n; i++) {  
        out.print("(" + i%w + ", " + i/w + (i%w == w-1 ? ") \n" : ") ");  
    }  
}
```


Programming



Programming



Databaser



Databaser

```
> CREATE DATABASE t1f;
```

Databaser

```
> CREATE DATABASE tlf;  
> USE tlf;  
> CREATE TABLE liste (navn VARCHAR(255),  
                        nummer INT,  
                        PRIMARY KEY (navn,nummer));
```

Databaser

```
> CREATE DATABASE tlf;  
> USE tlf;  
> CREATE TABLE liste (navn VARCHAR(255),  
                        nummer INT,  
                        PRIMARY KEY (navn,nummer));  
> INSERT INTO liste (navn) VALUES ('Truls');
```

Databaser

```
> CREATE DATABASE tlf;  
> USE tlf;  
> CREATE TABLE liste (navn VARCHAR(255),  
                        nummer INT,  
                        PRIMARY KEY (navn,nummer));  
> INSERT INTO liste (navn) VALUES ('Truls');  
> INSERT INTO liste (navn) VALUES ('Truls');
```

Databaser

```
> CREATE DATABASE tlf;  
> USE tlf;  
> CREATE TABLE liste (navn VARCHAR(255),  
                        nummer INT,  
                        PRIMARY KEY (navn,nummer));  
> INSERT INTO liste (navn) VALUES ('Truls');  
> INSERT INTO liste (navn) VALUES ('Truls');  
> INSERT INTO liste (navn,nummer)  
    VALUES ('Truls',55589124);
```


Databaser

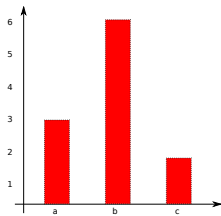
```
> CREATE DATABASE tlf;  
> USE tlf;  
> CREATE TABLE liste (navn VARCHAR(255),  
                        nummer INT,  
                        PRIMARY KEY (navn,nummer));  
> INSERT INTO liste (navn) VALUES ('Truls');  
> INSERT INTO liste (navn) VALUES ('Truls');  
> INSERT INTO liste (navn,nummer)  
                        VALUES ('Truls',55589124);  
> CREATE INDEX hvemringer ON liste (nummer);
```

Databaser

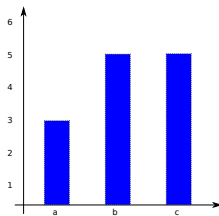
```
> CREATE DATABASE tlf;
> USE tlf;
> CREATE TABLE liste (navn VARCHAR(255),
                        nummer INT,
                        PRIMARY KEY (navn,nummer));
> INSERT INTO liste (navn) VALUES ('Truls');
> INSERT INTO liste (navn) VALUES ('Truls');
> INSERT INTO liste (navn,nummer)
    VALUES ('Truls',55589124);
> CREATE INDEX hvemringer ON liste (nummer);
> SELECT (navn) FROM liste
    WHERE nummer=55589124;
```

Matematikk

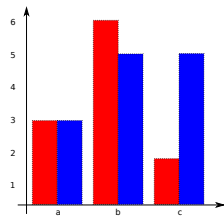
Vektorer



Lettøl

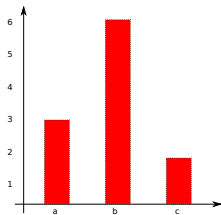


Is

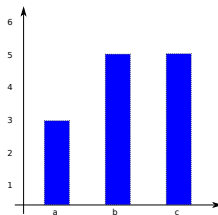


Matematikk

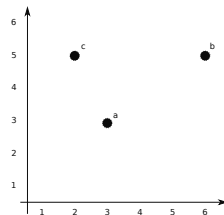
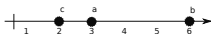
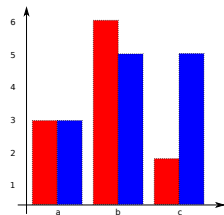
Vektorer



Lettøl

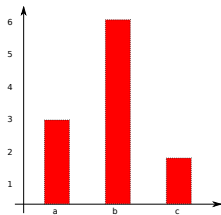


Is

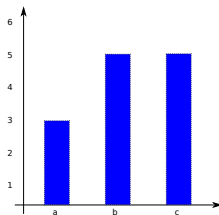


Matematikk

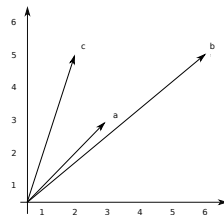
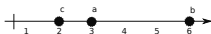
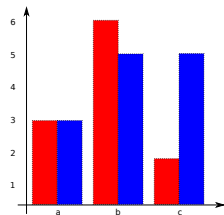
Vektorer



Lettøl



Is



Matematikk

Vektorer

$$\begin{array}{ll} \vec{a} = (3, 3) & |\vec{a}|^2 = 3^2 + 3^2 \\ \vec{b} = (6, 5) & |\vec{b}|^2 = 6^2 + 5^2 \\ \vec{c} = (2, 5) & |\vec{c}|^2 = 2^2 + 5^2 \end{array}$$

Generelt:

$$\begin{aligned} \vec{v} &= (x_1, x_2, \dots, x_n) \\ |\vec{v}|^2 &= x_1^2 + x_2^2 + \dots + x_n^2 \end{aligned}$$

Matematikk

Vektorer

$$\vec{v} = (x_1, x_2, \dots, x_n)$$

$$2\vec{v} = (2x_1, 2x_2, \dots, 2x_n)$$

$$(x_1 + x_1, x_2 + x_2, \dots, x_n + x_n) = \vec{v} + \vec{v}$$

Matematikk

Vektorer

$$\vec{v} = (x_1, x_2, \dots, x_n)$$

$$2\vec{v} = (2x_1, 2x_2, \dots, 2x_n)$$

$$(x_1 + x_1, x_2 + x_2, \dots, x_n + x_n) = \vec{v} + \vec{v}$$

$$\frac{\vec{v}}{3} = \left(\frac{x_1}{3}, \frac{x_2}{3}, \dots, \frac{x_n}{3}\right)$$

Matematikk

Vektorer

$$\vec{v} = (x_1, x_2, \dots, x_n)$$

$$2\vec{v} = (2x_1, 2x_2, \dots, 2x_n)$$

$$(x_1 + x_1, x_2 + x_2, \dots, x_n + x_n) = \vec{v} + \vec{v}$$

$$\frac{\vec{v}}{3} = \left(\frac{x_1}{3}, \frac{x_2}{3}, \dots, \frac{x_n}{3}\right)$$

Normalisering:

$$\hat{v} = \frac{\vec{v}}{|\vec{v}|}$$

Matematikk

Vektorer - indreprodukt

Hvis $\vec{v} = (x_1, x_2, x_3)$ og $\vec{u} = (y_1, y_2, y_3)$, så er

$$\vec{v} \cdot \vec{u} = x_1 y_1 + x_2 y_2 + x_3 y_3$$

Matematikk

Vektorer - indreprodukt

Hvis $\vec{v} = (x_1, x_2, x_3)$ og $\vec{u} = (y_1, y_2, y_3)$, så er

$$\vec{v} \cdot \vec{u} = x_1 y_1 + x_2 y_2 + x_3 y_3$$

Og generelt, hvis $\vec{v} = (x_1, x_2, \dots, x_n)$ og $\vec{u} = (y_1, y_2, \dots, y_n)$, så

$$\vec{v} \cdot \vec{u} = \sum_{i=1}^n x_i y_i$$

Matematikk

Vektorer - indreprodukt

Hvis $\vec{v} = (x_1, x_2, x_3)$ og $\vec{u} = (y_1, y_2, y_3)$, så er

$$\vec{v} \cdot \vec{u} = x_1 y_1 + x_2 y_2 + x_3 y_3$$

Og generelt, hvis $\vec{v} = (x_1, x_2, \dots, x_n)$ og $\vec{u} = (y_1, y_2, \dots, y_n)$, så

$$\vec{v} \cdot \vec{u} = \sum_{i=1}^n x_i y_i$$

Greit å huske

$$\vec{v} \cdot \vec{v} = |\vec{v}|^2$$

Matematikk

$$\cos \Theta = \frac{\vec{v} \cdot \vec{w}}{|\vec{v}| |\vec{w}|}$$

