VK Multimedia Information Systems

Mathias Lux, mlux@itec.uni-klu.ac.at

Dienstags, 16.oo Uhr s.t., E.1.42

Agenda

- Topics & Goals
- Modalities & Examination
- Schedule

- What is **Information**?
- What are **Information Systems**?
- The Information Overload
- Current state in MuMe Consumption

C.V.

- Technische Mathematik an der TU Graz
- Doktoratsstudium Telematik

- 98-01 Entwicklung von Web-Applikationen
- 01-06 Know-Center in Graz (KPlus)
- 05-06 Ass. am KMI / TU Graz
- 06- … Ass. am ITEC / Uni Klagenfurt

Course Topics

Multimedia Databases

Multimedia Management

Social Media Sharing

Video Analysis

Metadata

Data Mining

Digital Audio

Information Retrieval

Social Networks

Image Processing

Retrieval Evaluation

Goals I

Basic (and a little more) understanding of

- Multimedia Retrieval
- Multimedia Analysis
  - Images in the spatial domain
  - Audio & video processing
- Multimedia Databases & Metadata

Goals II

- Overview on State of the Art
  - o Who is who in research
  - o What to read if I want to know more?
  - o Available tools in development
  - o De facto & de jure standards

Goals III

- Providing a solid base for
    - Further research,
    - Consulting and
    - Practical development
- in the area of
    - Multimedia information
    - Multimedia information systems
- And: Hands on experience!

Modalities

Multimedia Information Systems ist eine *„prüfungsimmanente Lehrveranstaltung"*

The grade comes from

- Some few mandatory exercises / readings
- Ongoing collaboration
- A mid term and a final project

Modalities

Mid term projects are

- The same for everyone
- A simple VIR system + evaluation

Final projects are

- Practical implementations of MMIS
- Research work & studies

Projects topics will be

- … assigned after Easter holidays
- … assigned to groups or individual students

Team Work

- Preferably teams of **2** students
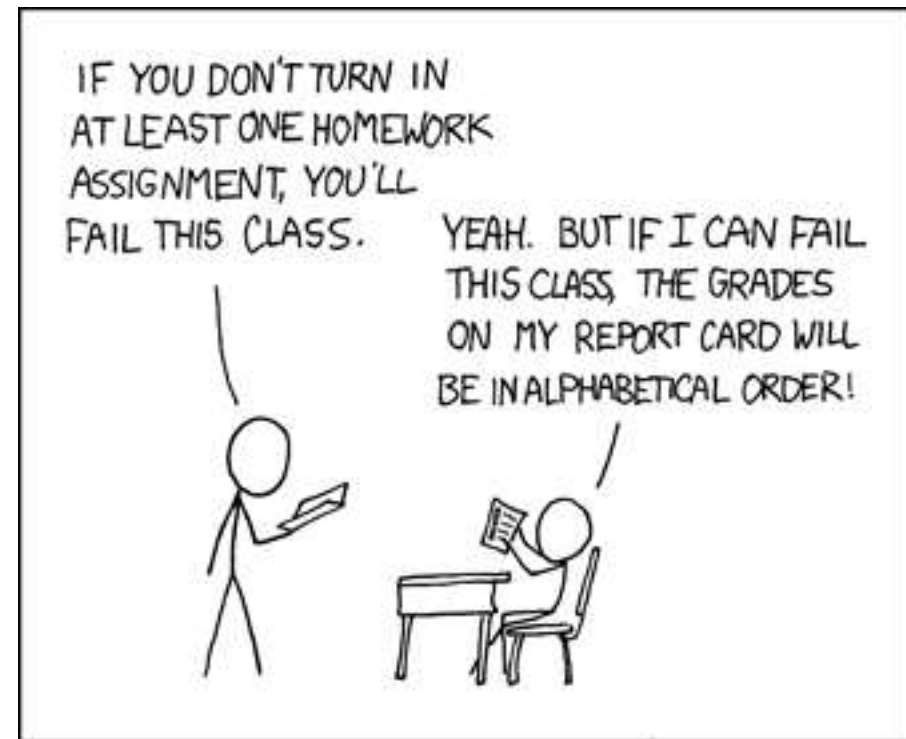- TEAM = „Toll-Ein-Anderer-Machts" ?

Grades

Grade is derived from:

http://www.xkcd.com

- 1/4 Exercises
  - Pen & paper
  - Readings
- 1/2 Project
  - ¼ each
  - Implementation
- 1/4 Presentation

Schedule

- Introduction, Motivation, Information Theory & Systems
- Information Retrieval
- Web based IR, PageRank, HITS
- Network Analysis & Social Networks (Guest Lecture)
- Multimedia Metadata
- Image Analysis and Content Based Image Retrieval
- Audio & Sound Analysis
- Video Information Systems
- Multimedia Databases

Questions?

Any questions regarding organizational issues left?

Agenda

- What is **Information**?
- What are **Information Systems**?
- The Information Overload
- Current state in MuMe Consumption

What is Information?

Definition von Aamondt und Nygard (1995) :

- Data
- Information
- Knowledge

Aamodt, A. & Nygard, M. "Different roles and mutual dependencies of data, information, and knowledge - an AI perspective on their integration" *Data Knowl. Eng., Elsevier Science Publishers B. V.*, **1995**, *16*, 191-222

Data are syntactic entities

- Patterns without meaning
- Input to an interpretation process

Example:
- Bits & Bytes of a JPEG encoded image

Information is interpreted data

- Information is data with meaning
- Output from interpretation
- Input to knowledge based process

Example:
- Decoded (and displayed) JPEG image

Knowledge is learned information

- Incorporated in an agents (software / human) reasoning resources
- Ready for active use
- Output of learning process

Example:
- There is a dog shown on the JPEG image

What is Information?

Definition of Zeleny (1987):

- Data
- Information
- Knowledge
- Wisdom

Zeleny, M. "Management Support Systems: Towards Integrated Knowledge Management" Human Systems Management, 1987, 7, 59-70
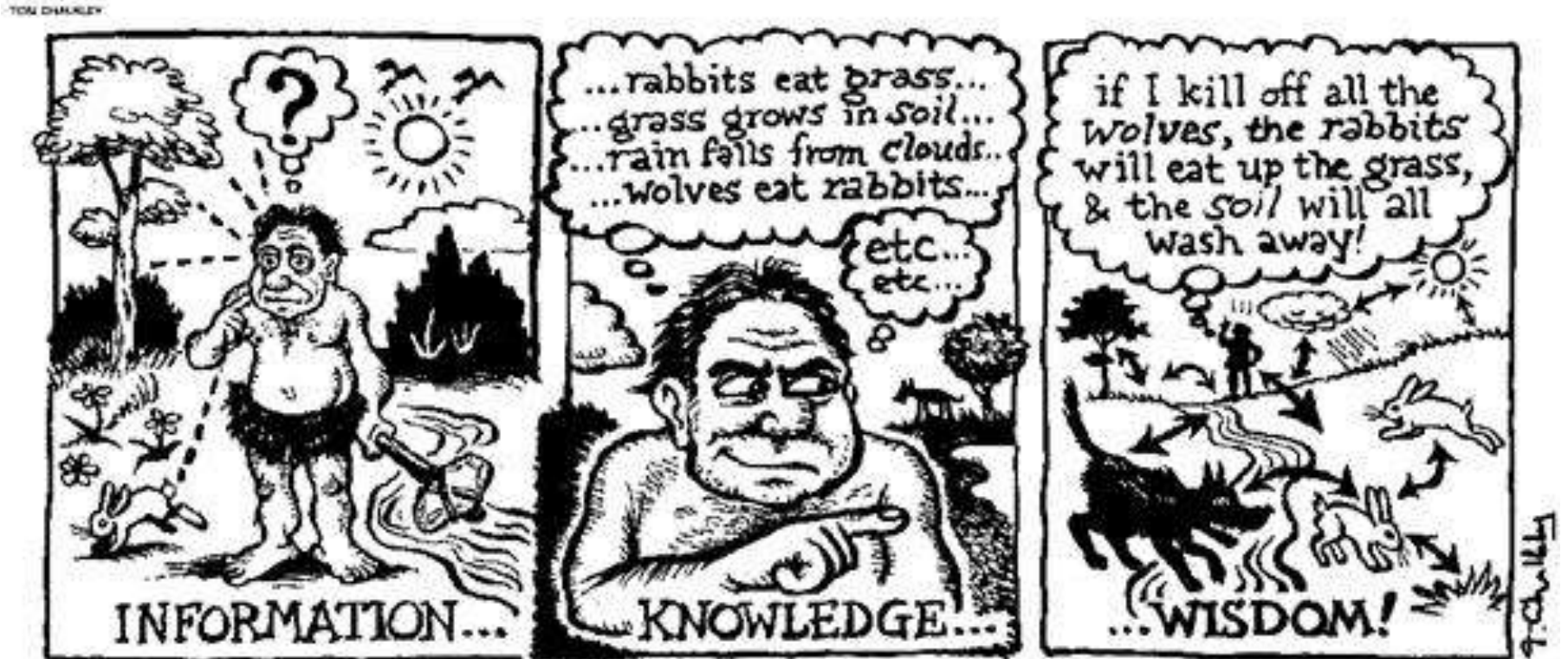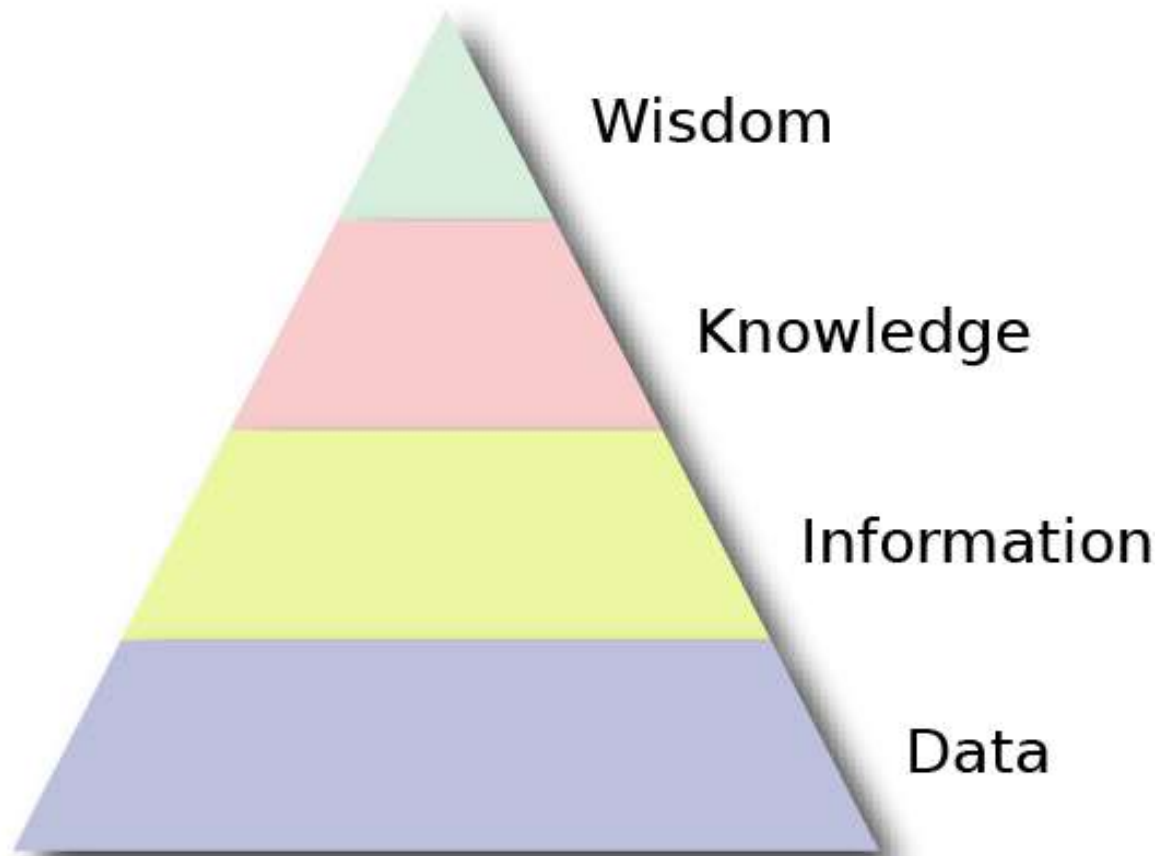
What is Information?

Image originally published in the December 1982 issue of THE FUTURIST, taken from http://www-personal.si.umich.edu/~nsharma/dikw_origin.htm
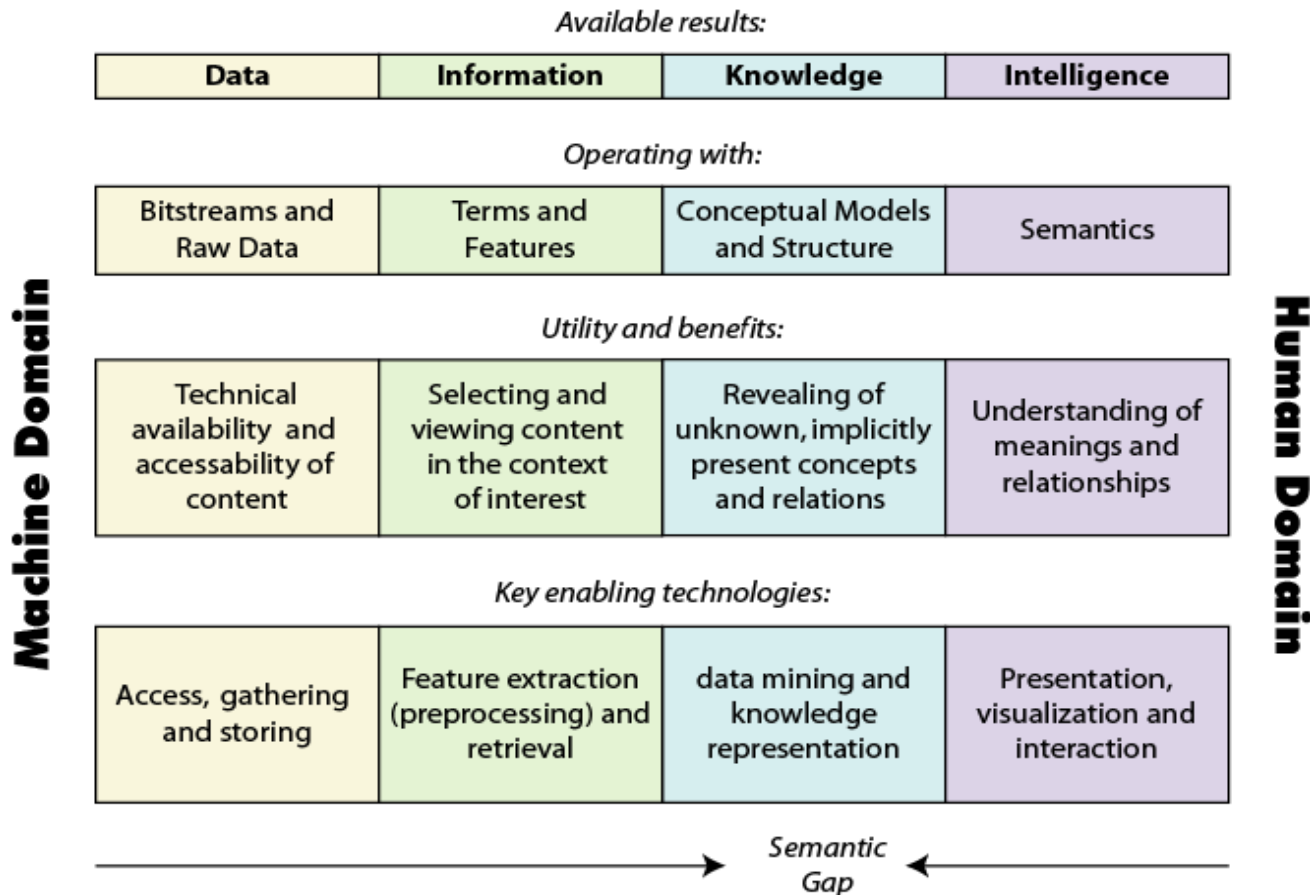
The DIKW Hierarchy

Definition of the DIKW levels:

Data	Know nothing
Information	Know what
Knowledge	Know how
Wisdom	Know why

Modified DIKW (IBM)

Available results:

Data	Information	Knowledge	Intelligence

Operating with:

Bitstreams and Raw Data	Terms and Features	Conceptual Models and Structure	Semantics

Machine Domain

Human Domain

Utility and benefits:

Technical availability and accessability of content	Selecting and viewing content in the context of interest	Revealing of unknown, implicitly present concepts and relations	Understanding of meanings and relationships

Key enabling technologies:

Access, gathering and storing	Feature extraction (preprocessing) and retrieval	data mining and knowledge representation	Presentation, visualization and interaction

Semantic Gap

What is Information?

Shannon's Information Theory

- Problem: Communication over a noisy channel

- Fundamental finding:
  - Information Content (measured in bits) of an Event (e.g. letter) depends on the entropy (probability of occurrence)
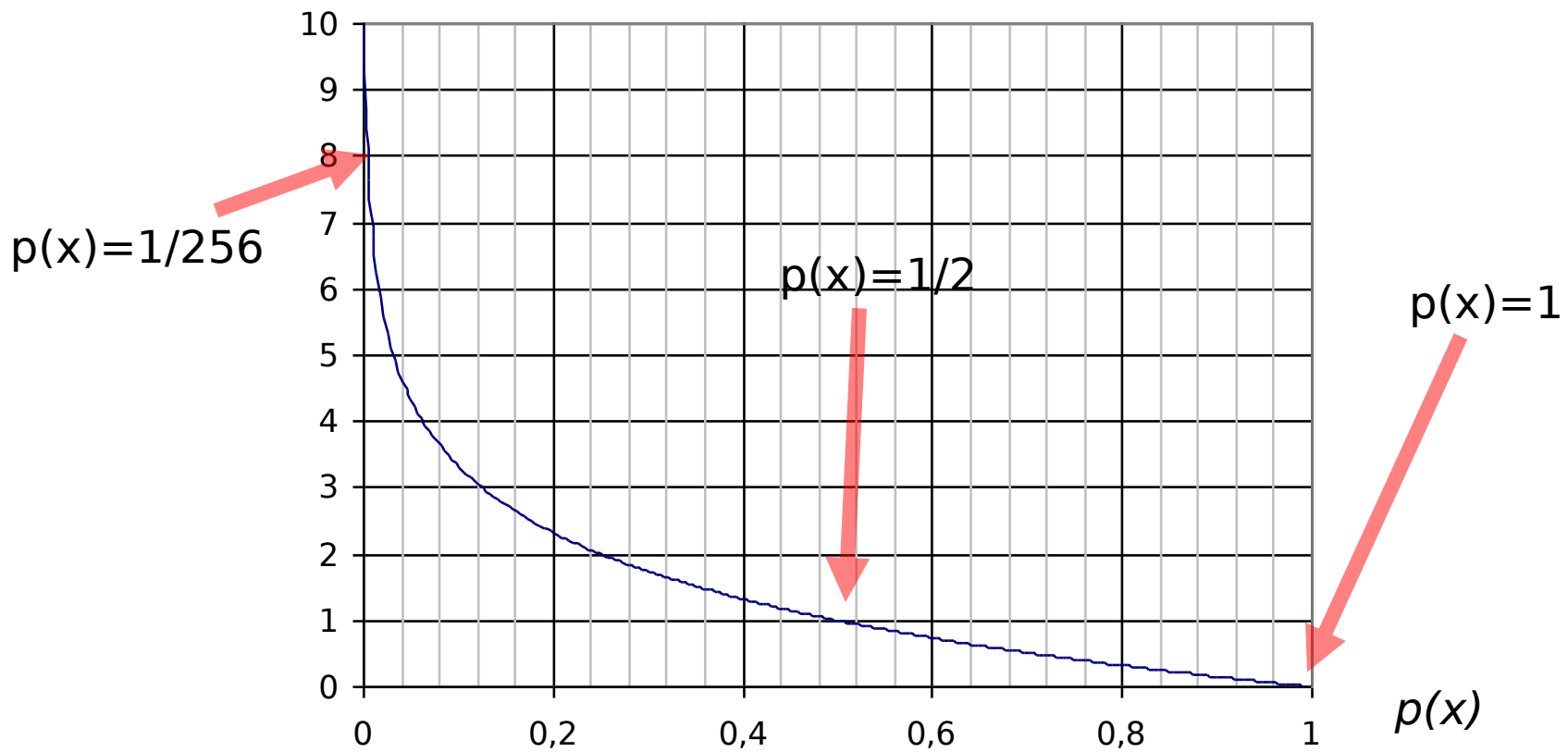
$$I(A_n) = \log_2 \left(\frac{1}{P(A_n)} \right) = -\log_2(P(A_n))$$

Anzahl der Bits

p(x)=1/256

p(x)=1/2

p(x)=1

p(x)

- As informative as required
- As correct as possible
- Relevant to the aims of the conversation
- Contribution should be clear, unambiguous and concise

Haupmann, A. G. & Witbrock, M. J. "Story Segmentation and Detection of Commercials in Broadcast News Video" ADL '98: Proceedings of the Advances in Digital Libraries Conference, IEEE Computer Society, 1998

Agenda

- What is **Information**?
- What are **Information Systems**?
- Information Overload
- Current state in MuMe Consumption

What are Information Systems?

- Systems for handling information
  - Collect, Store & Organize
  - Process, Disseminate & Transmit
- Three main parts in these systems
  - People,
  - Machines &
  - Methods

MEMEX

Memory **Ex**tender – Vannevar Bush

- Published in 1945 (Atlantic Monthly)
- An electromechanical device for
  - Viewing books and films
  - Adding information and comments
  - Interlinking information
  - Browsing links
- MEMEX is an early hypertext system.

Geographic IS

- Focus on **Spatially Referenced Data**
  - Coordinates, Height
  - Distance, Inclusion, Neighbouring
  - Hierarchical organisation

taken from Google Maps

Multimedia IS

- Focus on Multimedia Data & Metadata
  - Storage, Transmission
  - Search & Retrieval
  - Organization & Dissemination
- Media types
  - Textual / Visual / Auditive / Haptic / Olfactory
  - Rastered or Rendered / Modelled
    - Midi vs. MP3
    - VRML vs. PNG
    - LASER vs. MPEG-2

Demo: PSP

Agenda

- What is **Information**?
- What are **Information Systems**?
- Information Overload
- Current state in MuMe Consumption

Information Overload

- 5 Exabytes of new information in 2002.
    - 92% of the new information was stored on magnetic media, mostly in hard disks.
    - That's 800 MB per person on the globe
    - That's 37.000 times the LoC
    - That's 30% more than in 1999

Lyman, Peter and Hal R. Varian, "How Much Information", 2003. Retrieved from http://www.sims.berkeley.edu/how-much-info-2003 on [2007-02-07]

Information Overload

18 Exabytes of new information in information flow in 2002

- 98% of new information generated by phone calls

- Most radio and TV broadcast content is **not new** information.

  - ~ 70 out of 320 million h of radio is new, that's 3.500 TB

  - ~ 31 out of 123 million h of TV are new, that's 70.000 TB

Information Overload

Information Flow (ctd.)

- IM: 5 billion messages / day or 274 TB p.a.

- Email: 400.000 TB p.a.

- P2P: growing, but not yet estimated,
  - significant traffic has been observed on different backbones
  - Ranges from 20%-60% are mentioned

Information Usage

An average American adult:

- Telephone - 16.17 hours a month
- Radio - 90 hours a month
- TV - 131 hours a month
- 53% of the U.S. uses the Internet
  - ~ 25.5 h / month at home
  - ~ 74.5 h / month at work

Agenda

- What is **Information**?
- What are **Information Systems**?
- Information Overload
- Current state in MuMe Consumption

Current State in MuMe Consumption

- Digital Photography
  - Still images
- Digital Video in General
  - Streaming and download

- First question: How many devices exist?

Device	# in 2006
digital cameras	$400 * 10^6$
camera phones	$600 * 10^6$

Source: IDC Study "Expanding Digital Universe" http://www.emc.com/about/destination/digital_universe/
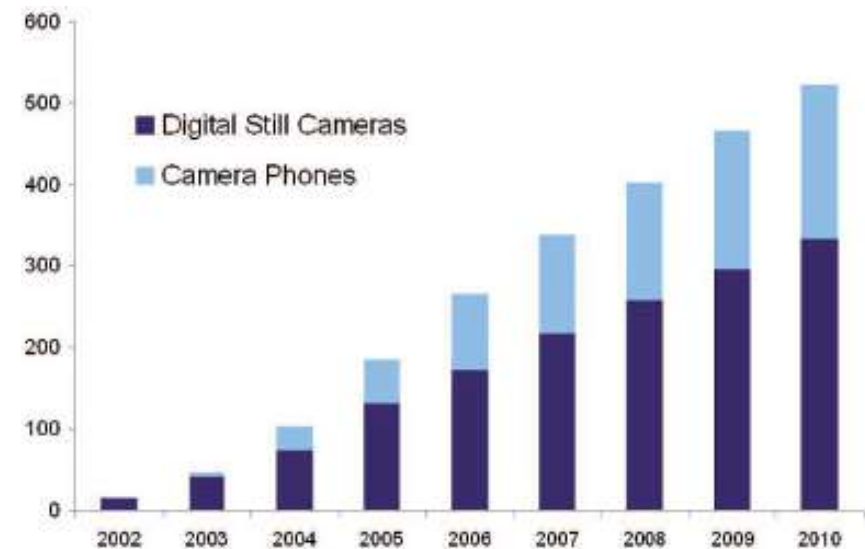
Number of Digital Photos (global)

- Estimate 2006
  - > 150 billion photos from cameras
  - > 100 billion photos from camera phones

- Forecast 2010
  - > 500 billion photos
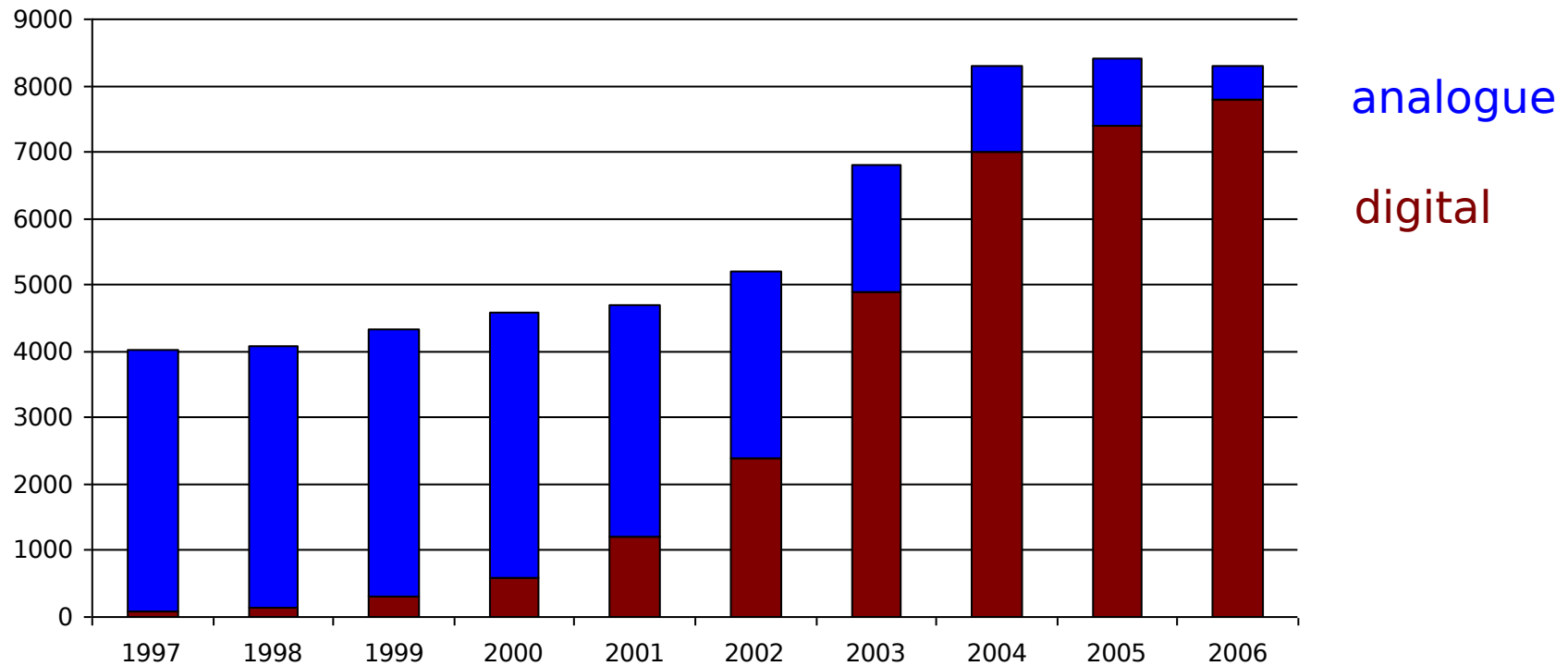 - + increased resolution

Source: IDC Study "Expanding Digital Universe" http://www.emc.com/about/destination/digital_universe/

Digital Imaging Devices (Germany)
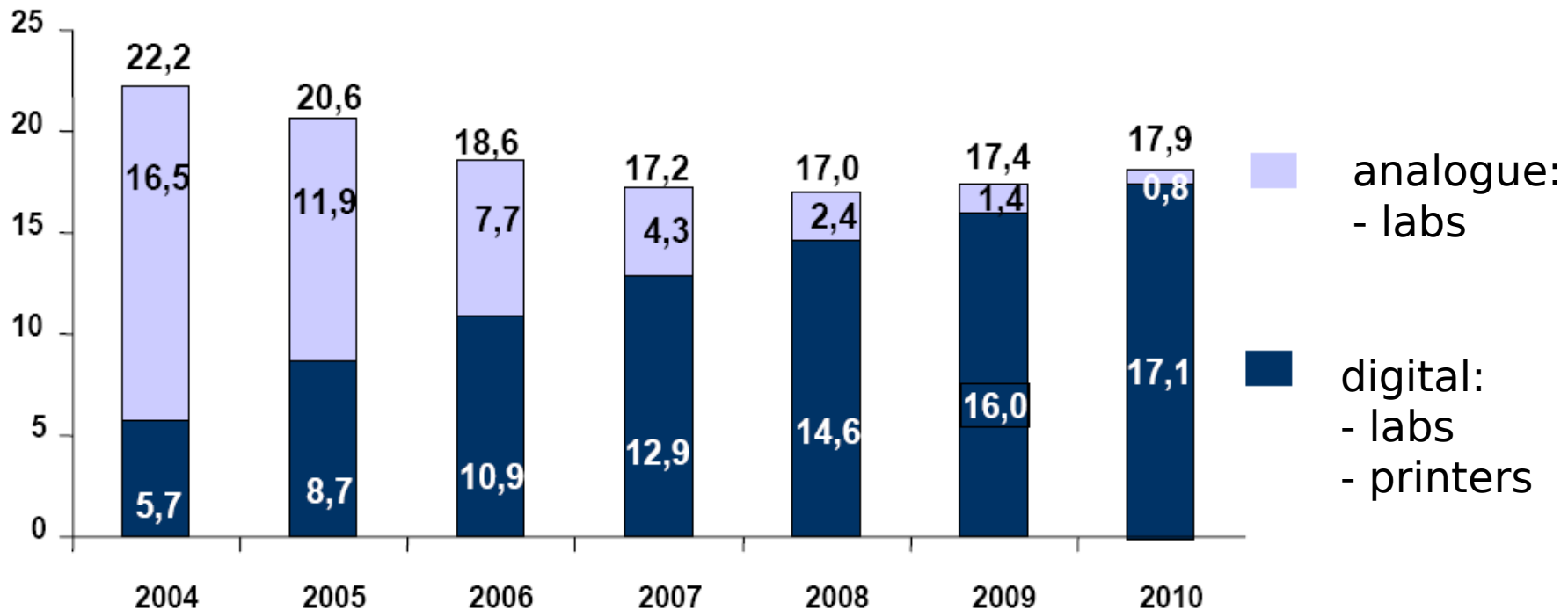
Still image cameras sold in Germany (thousands)

analogue

digital

Source: Cewe Factbook, http://www.cewecolor.de

Photo prints market (Western Europe)

- Photo prints forecast (in billions)

analogue:
- labs

digital:
- labs
- printers

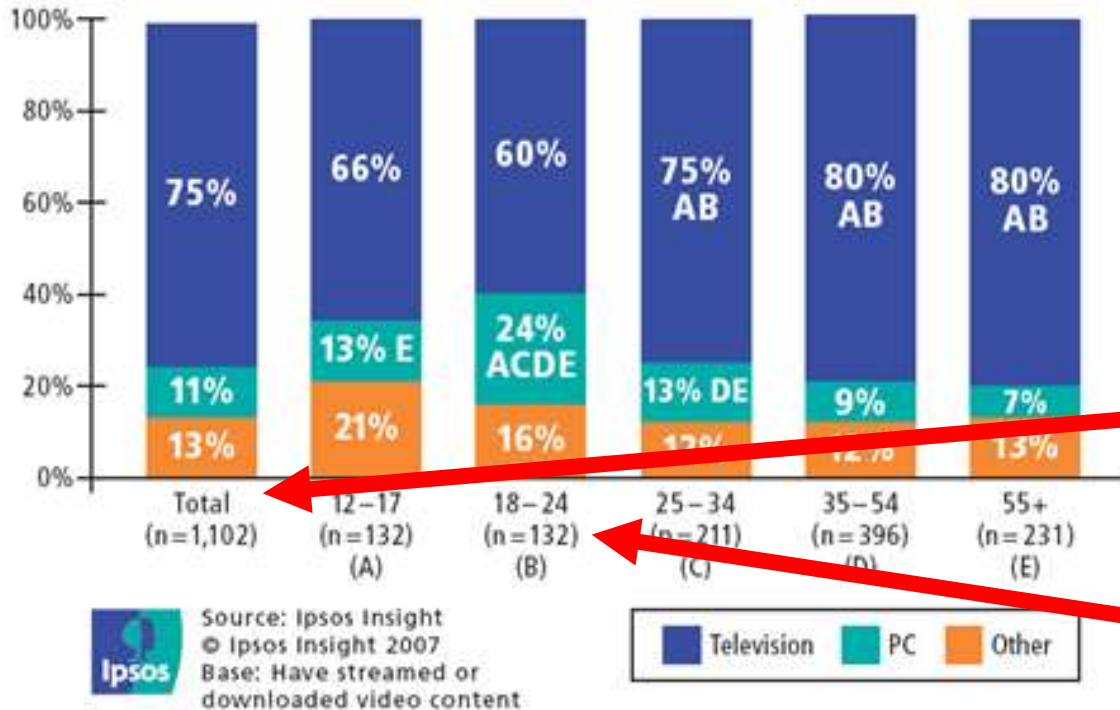Source: Cewe Factbook, http://www.cewecolor.de

Content

- What is **Information**?
- What are **Information Systems**?
- Information Overload
- Current state in MuMe Consumption
  - Digital Photography
  - Digital Video in General

Where video content is watched? (US, 2007)

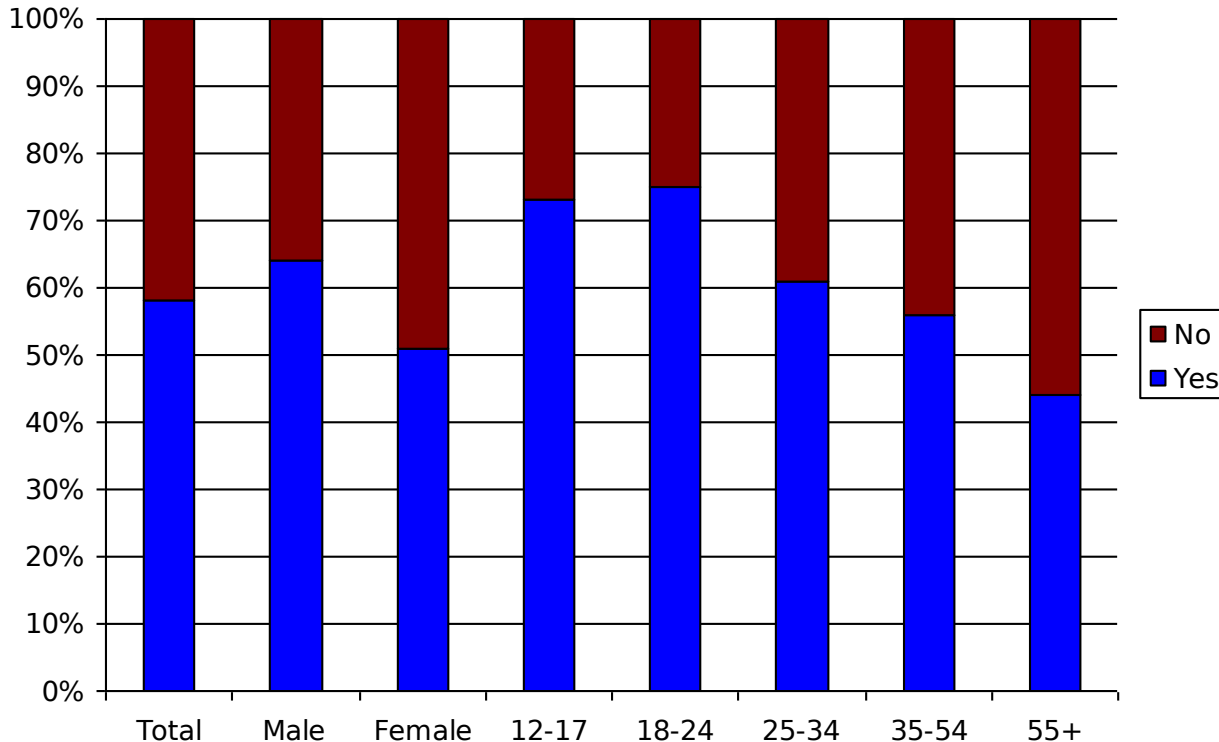- Among people having downloaded / streamed content

- 11% in total watched on PC

- 24% in age 18-24

Source: Ipsos Insight's 2007 MOTION Study - http://www.ipsosinsight.com/pressrelease.aspx?id=3500

Ever Streamed a File Off of the Internet? (US, 2007)

- 75% of the 12-24 years old.

- More than half in total.

Source: Ipsos Insight's 2007 MOTION Study - http://www.ipsosinsight.com/pressrelease.aspx?id=3500
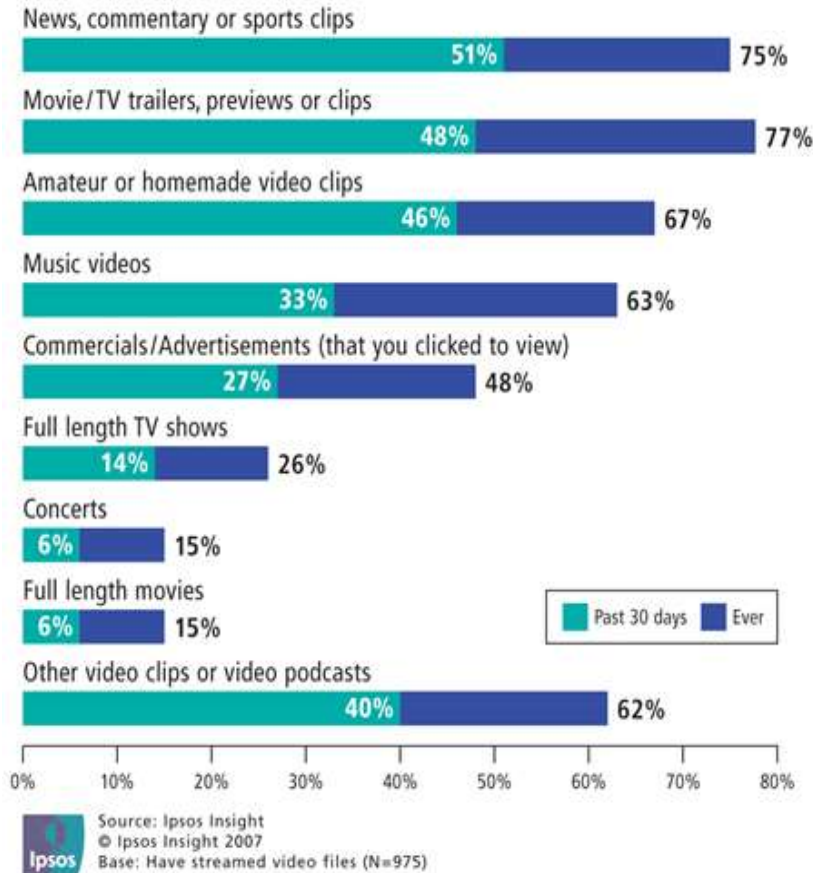
Short vs. Long Clips (US, 2007)

- Short clips are preferred
  - 3/4 of streamers have streamed short news or sports clips
  - 2/3 of streamers have streamed amateur or homemade clips
- Also due to YouTube …
  - 40% of the streamers use YouTube
  - Common restrictions in video size & length

Source: Ipsos Insight's 2007 MOTION Study - http://www.ipsosinsight.com/pressrelease.aspx?id=3500

Types of Digital Video Files Streamed (US, 2007)

News, commentary or sports clips — 51% (Past 30 days), 75% (Ever)
Movie/TV trailers, previews or clips — 48%, 77%
Amateur or homemade video clips — 46%, 67%
Music videos — 33%, 63%
Commercials/Advertisements (that you clicked to view) — 27%, 48%
Full length TV shows — 14%, 26%
Concerts — 6%, 15%
Full length movies — 6%, 15%
Other video clips or video podcasts — 40%, 62%

Past 30 days / Ever

Source: Ipsos Insight
© Ipsos Insight 2007
Base: Have streamed video files (N=975)

- Concerts, Movies: 15%
- TV Shows: 26%

But:

- 43% express some level of interest in downloading full length movies
- 38% express interest in full length TV show downloads

This also means:

- They haven't even tried.

Source: Ipsos Insight's 2007 MOTION Study - http://www.ipsosinsight.com/pressrelease.aspx?id=3500

Most Common Barrieres for payed DLs (US, 2007)

- Unwillingness to pay for this content,
- Difficulty or inability to
  - burn these files onto DVD
  - watch this content on their living room TV.

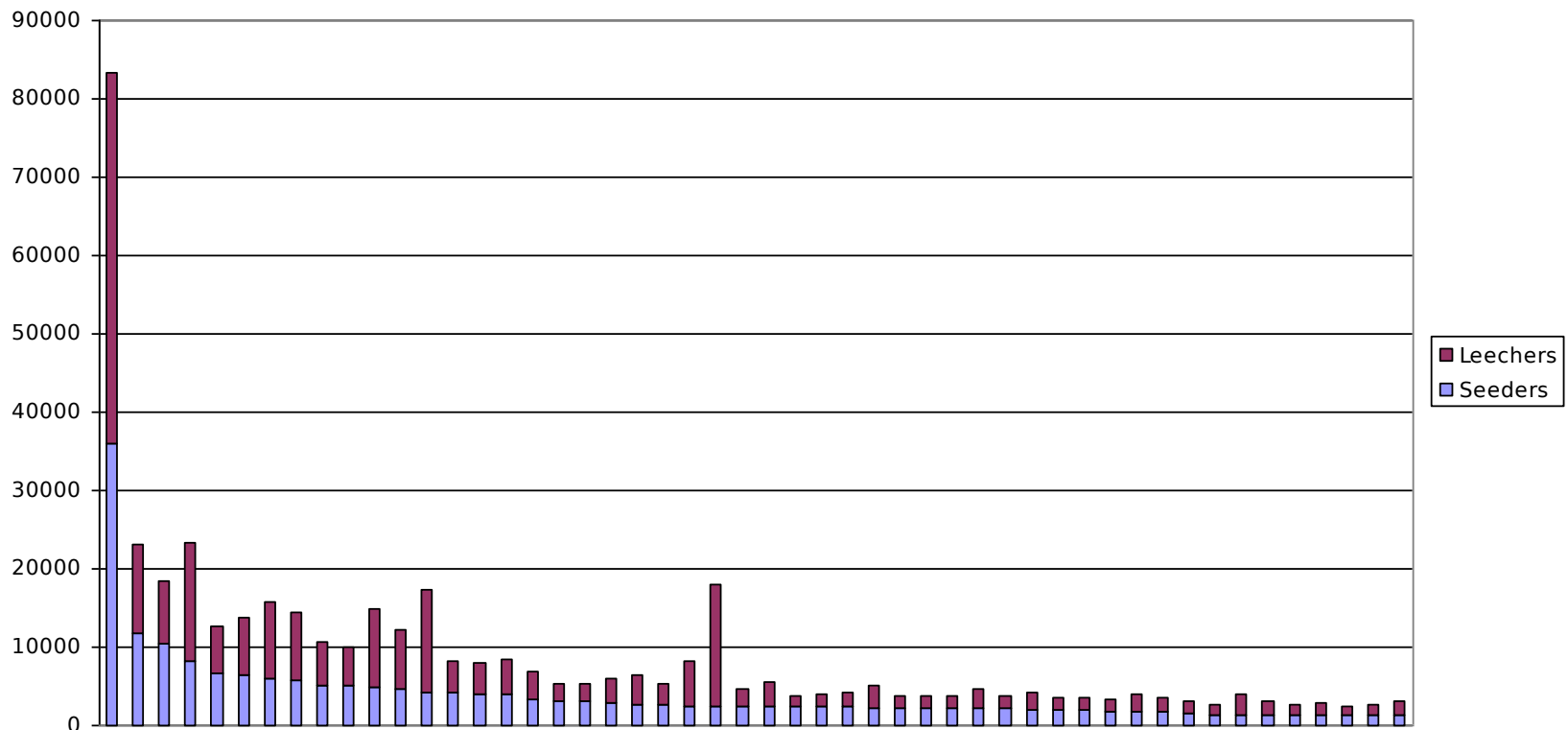But there are also benefits in the future:
  - 24/7 access
  - reasonable fees or free (ad supported)

Seeders & Leechers of Top 50 *aXXo* releases on isohunt.com

Bittorrent Movie Download
(aXXo releases, 24.10. 2007, isohunt.com)

Age (weeks)	Torrent file	Size (MB)	Seeders	Leechers
4.4w	Transformers[2007]DvDrip[Eng]-aXXo	900.23 MB	36059	47194
8w	Blades.Of.Glory[2007]DvDrip.AC3[Eng]-aXXo	699.76 MB	11740	11419
16.2w	300[2006]DvDrip[Eng]-aXXo	701.06 MB	10540	7921
5.2w	Evening[2007]DvDrip[Eng]-aXXo	701.17 MB	8214	15068
9.6w	Reign.Over.Me[2007]DvDrip[Eng]-aXXo	701.54 MB	6610	6027
7.2w	Meet.The.Robinsons[2007]DvDrip.AC3[Eng]-aXXo	699.56 MB	6529	7329
3.6d	I.Now.Pronounce.You.Chuck.And.Larry[2007]DvDrip[Eng]-aXXo	700.94 MB	6087	9600
1.1w	Pirates.Of.The.Caribbean-At.World's.End[2007]DvDrip[Eng]-aXXo	900.29 MB	5802	8642
14w	Mr.Brooks[2007]DvDrip[Eng]-aXXo	701.3 MB	5109	5656
7.6w	Grindhouse-Death.Proof[2007][Unrated.Editon]DvDrip[Eng]-aXXo	700.98 MB	5034	4861
2d	Harry.Potter.And.The.Order.Of.The.Phoenix[2007]DvDrip[Eng]-aXXo	800.08 MB	4907	10025
3.6d	License.To.Wed[2007]DvDrip.AC3[Eng]-aXXo	699.91 MB	4595	7654
6.6w	Lucky.You[2007]DvDrip[Eng]-aXXo	701.43 MB	4235	13146
7.5w	This.Is.England[2006]DvDrip[Eng]-aXXo	701.15 MB	4139	3999
12.6w	Perfect.Stranger[2007]DvDrip[Eng]-aXXo	701.65 MB	4007	4078
22w	Ghost.Rider[2007]DvDrip[Eng]-aXXo	700.86 MB	3975	4438
7.2w	First.Born[2007]DvDrip[Eng]-aXXo	701.31 MB	3279	3571
3.4w	Grindhouse-Planet.Terror[2007][Unrated.Edition]DvDrip[Eng]-aXXo	701.59 MB	3170	2238
8.7w	28.Days.Later[2002]DvDrip[Eng]-aXXo	700.98 MB	3143	2288
16.7w	The.Hills.Have.Eyes-2[2007]DvDrip.AC3[Eng]-aXXo	699.9 MB	2899	3117

Readings

C.J. van Rijsbergen: *Information Retrieval – Introduction*, London, Butterworth, 1979

Available on homepage: mmis08 / multimedia

Your task:

- Read + answer questions
- Send me an **email** with the answers until next course.

Questions:

- What is the difference between Data Retrieval and Information Retrieval?
- What does "relevance" mean in the context of Information Retrieval?

Thanks ...

... for your attention!