VK Multimedia Information Systems

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

Dienstags, 16.oo Uhr, E.2.69

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
  - Who is who in research
  - What to read if I want to know more?
  - Available tools in development
  - 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 final project

Modalities

Final projects are

- Practical Implementations
- 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“ ?

Projects

- Combination with “Softwarepraktikum” possible

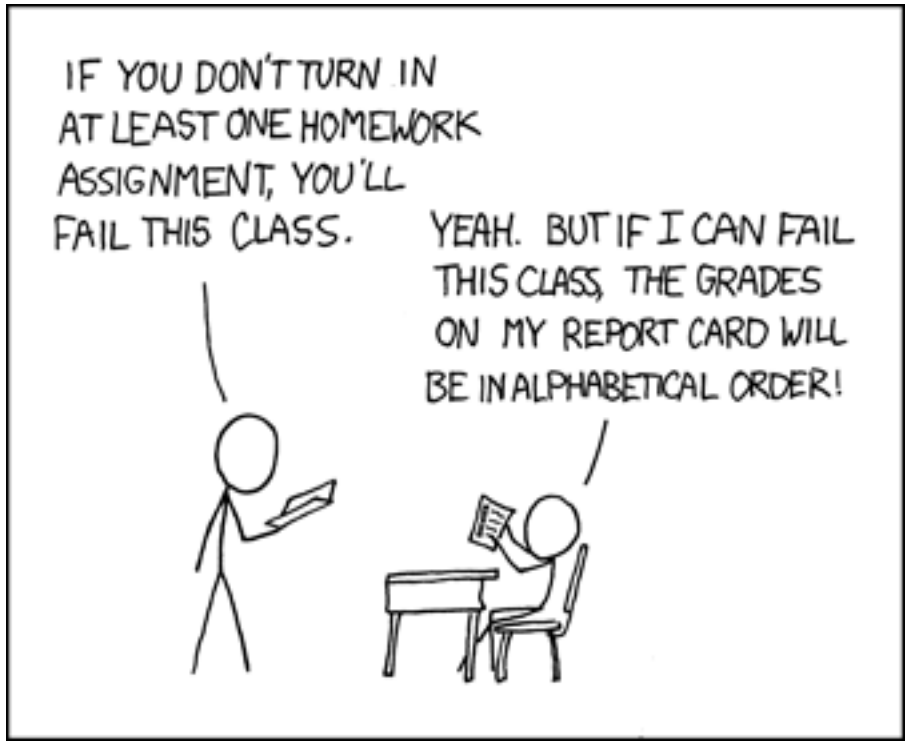
- Topics:
  - Topics on the course home page
  - Bring along you own Topic

Grades

Grade is derived from:

- 1/3 Exercises
  - Pen & paper
  - Readings
- 1/3 Project
  - Implementation
  - Documentation
- 1/3 Presentation

http://www.xkcd.com



Schedule

- 04.03. – Introduction
- 11.03. – Information Retrieval (IR 1)
- 18.03. – IR 2 + IR Evaluation
- 01.04. – Multimedia Metadata
- 08.04. – Network Analysis & SNA (Guest)
- 15.04. – Image Analysis
- 22.04. – Video Analysis I
- 29.04. – Video Analysis II

Schedule

- 06.05. – Audio Analysis & Music Retrieval
- 20.05. – Multimedia Databases
- 27.05. – Multimedia IS: Applications
- 03.06. – Project Presentations I
- *10.06. – entfällt*
- *17.06. – entfällt*
- 24.06. – Project Presentations II

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
Aamondt und Nygard (1995)

Data are syntactic entities

- Patterns without meaning
- Input to an interpretation process

Example:

- Bits & Bytes of a JPEG encoded image

Information

Aamondt und Nygard (1995)

Information is interpreted data

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

Example:

- Decoded (and displayed) JPEG image

Knowledge
Aamondt und Nygard (1995)

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):

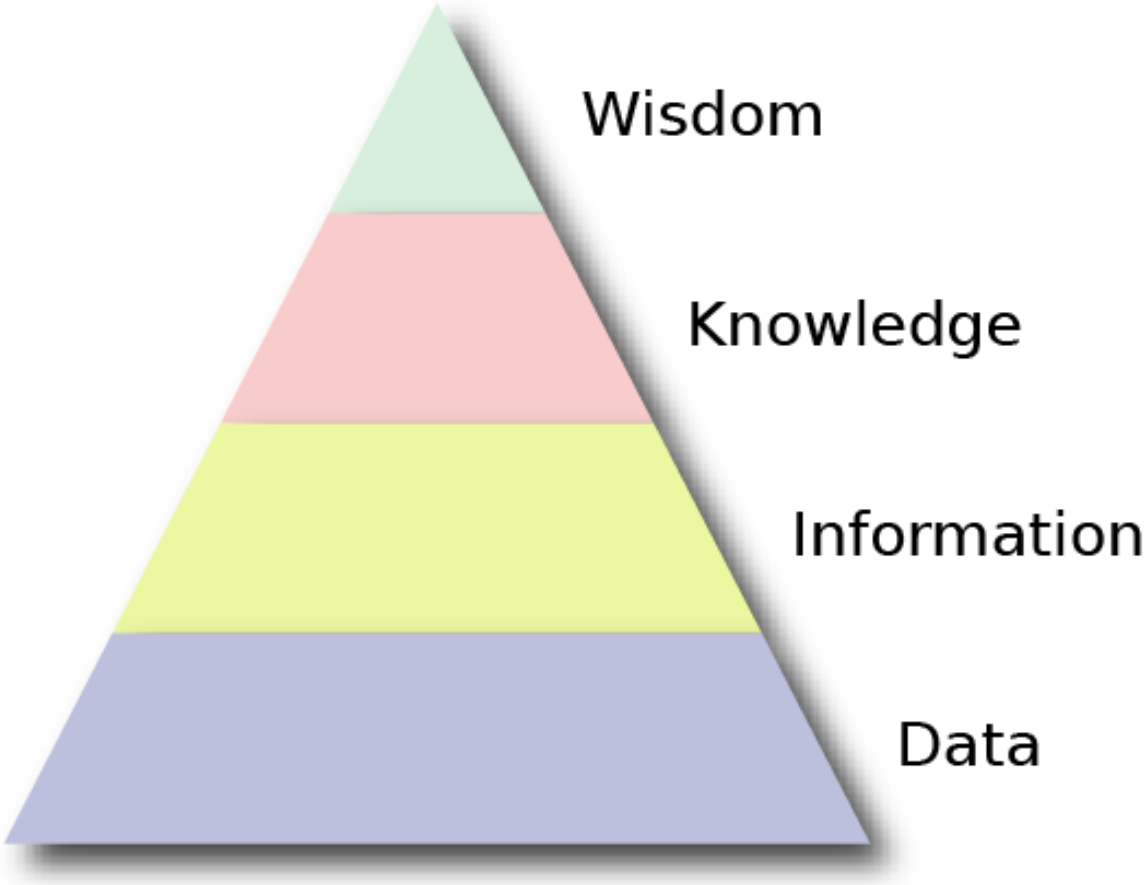
- Data
- Information
- Knowledge
- Wisdom

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

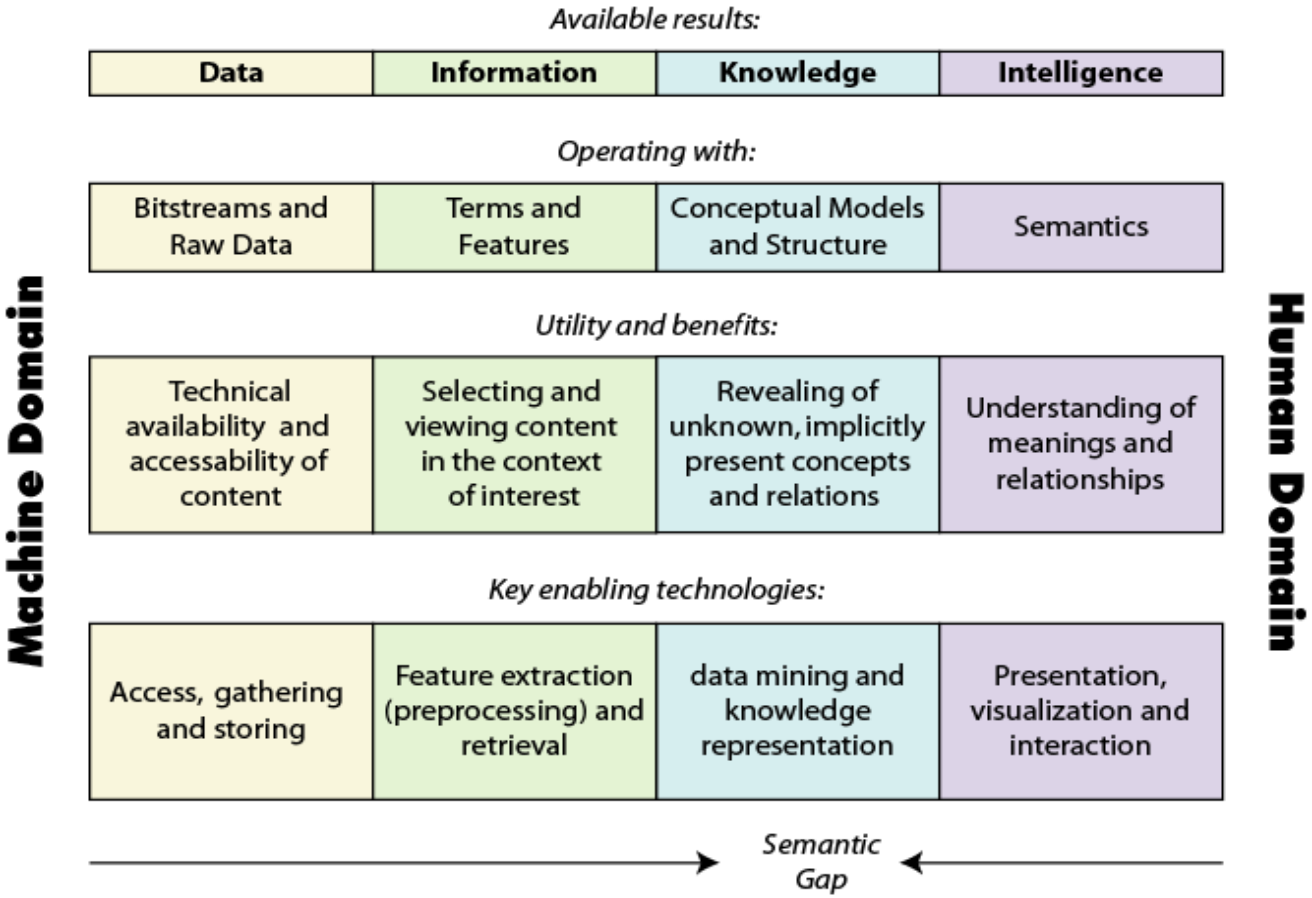


The DIKW Hierarchy

Definition of the DIKW levels:

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

Modified DIKW (IBM)



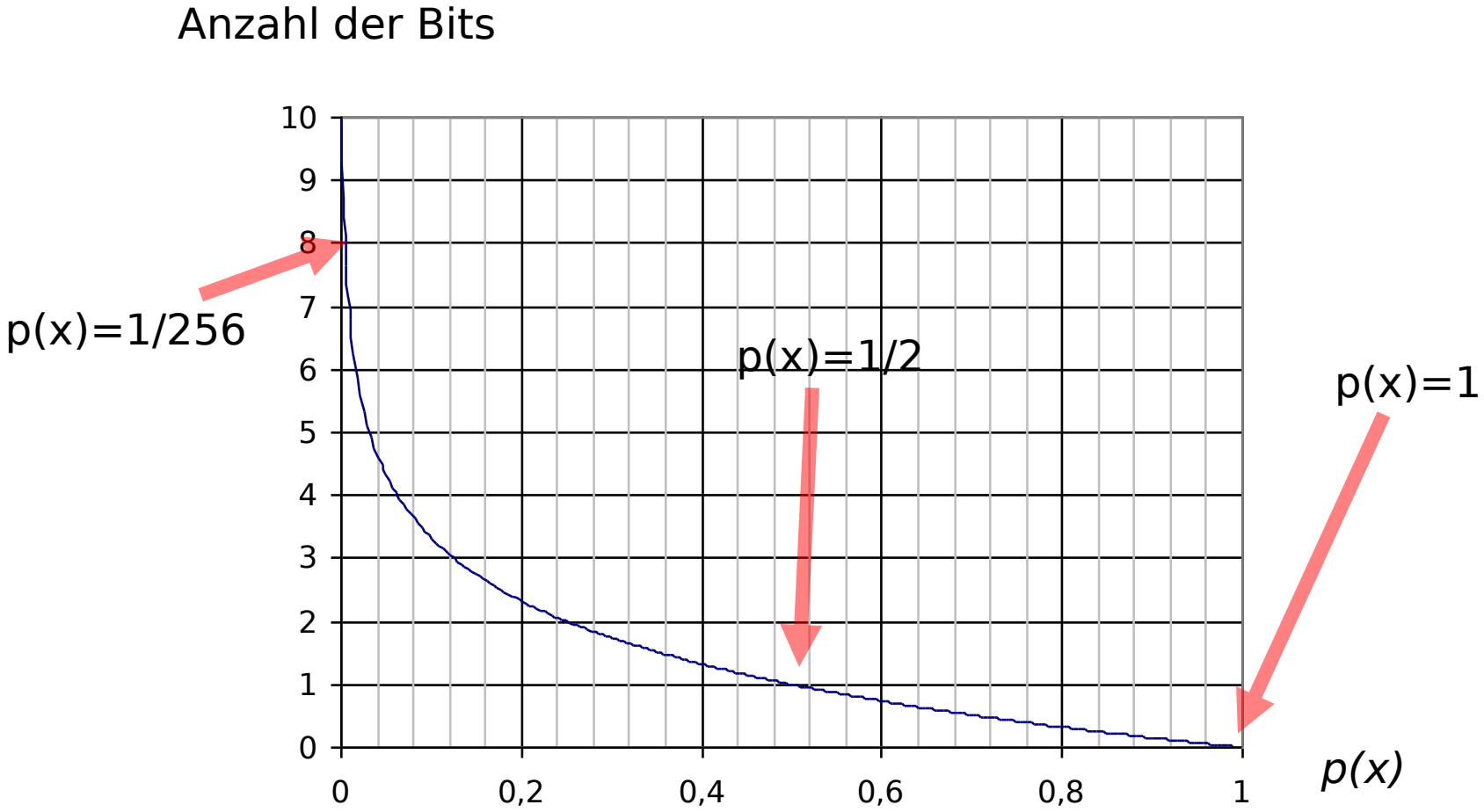
What is Information?

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)

$$I(A_n) = \log_2\left(\frac{1}{P(A_n)}\right) = -\log_2(P(A_n))$$

Shannon's Information Theory



Grice's Maxims of Conversation

- 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

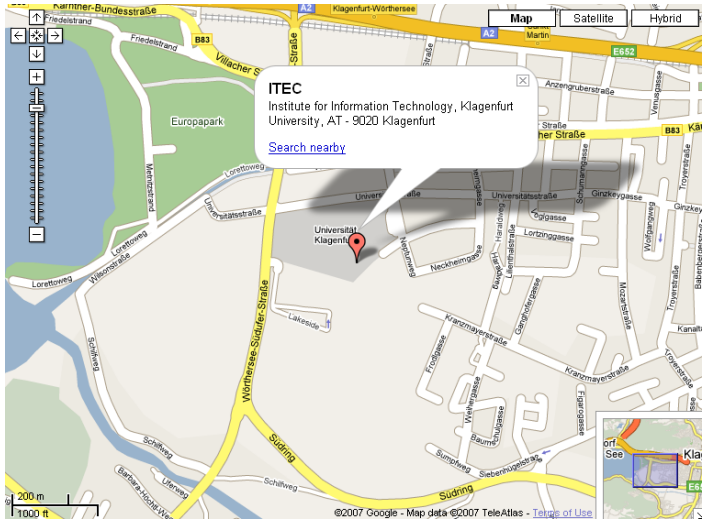
MEMEX

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

Agenda

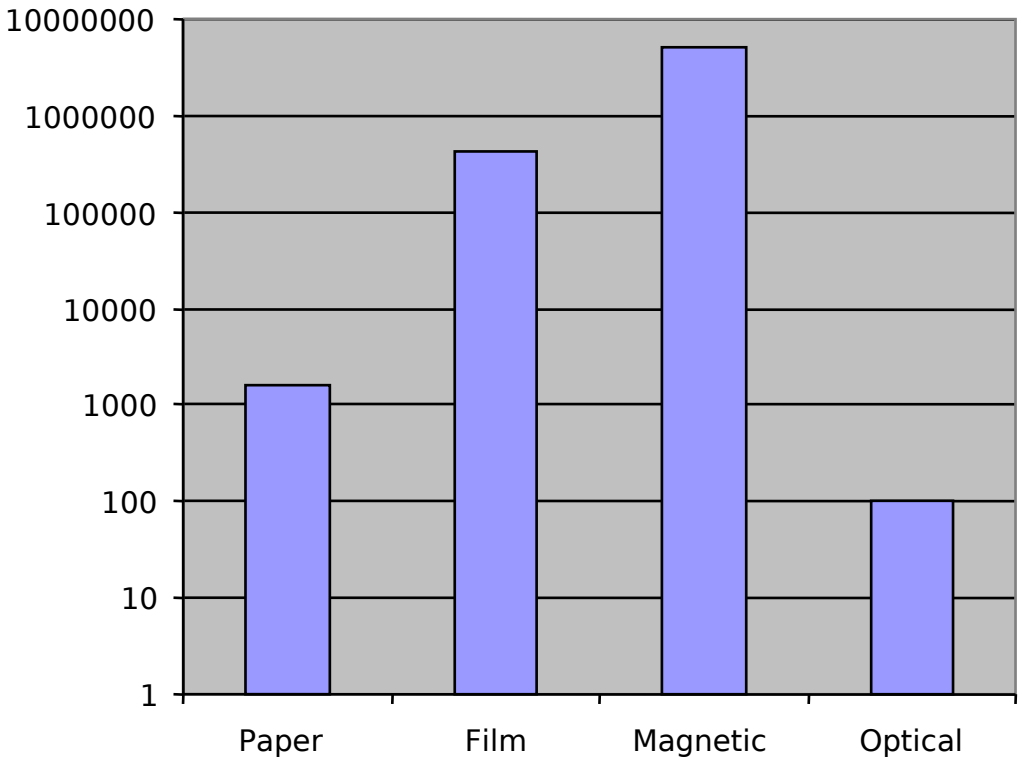
- What is **Information**?
- What are **Information Systems**?
- Information Overload
- Current state in MuMe Consumption

Information Overload

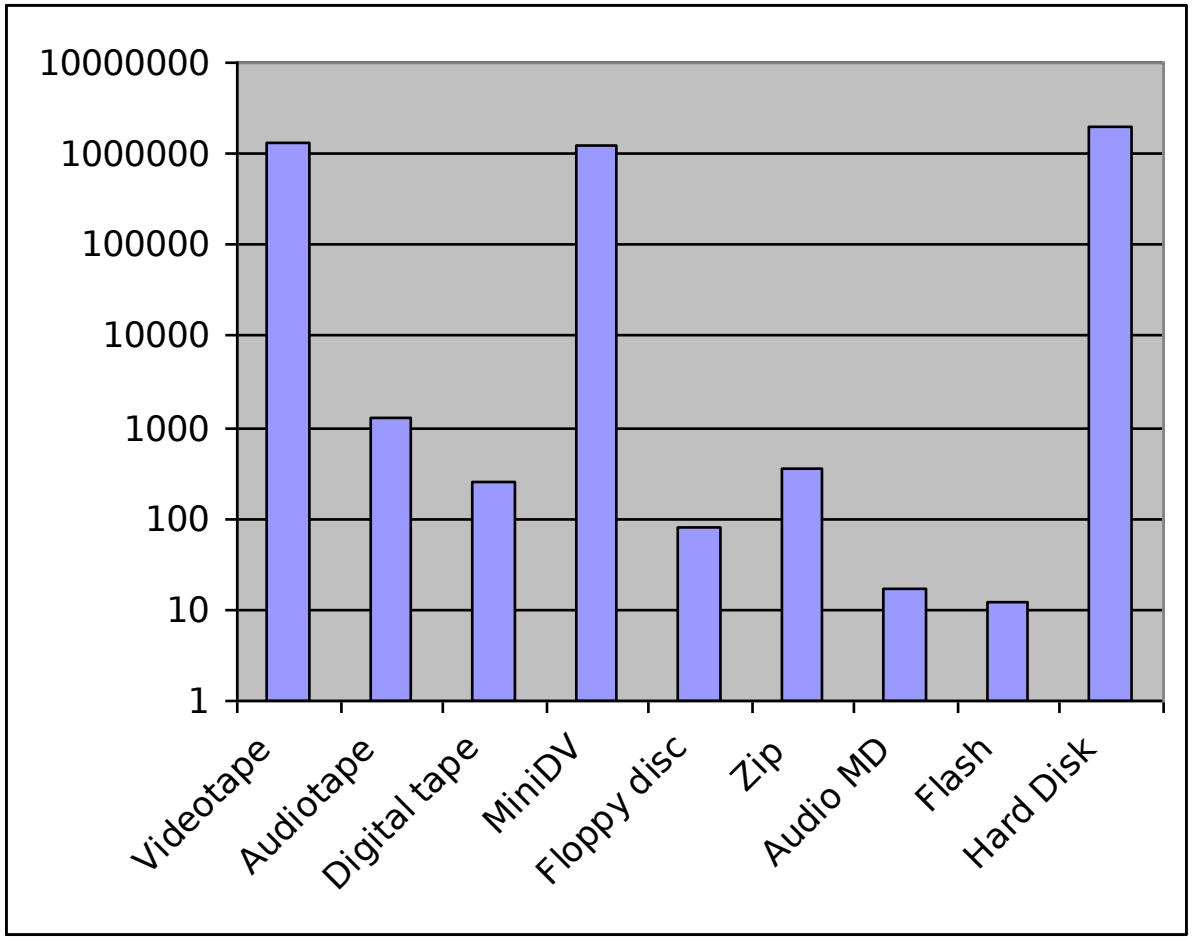
- 5 Exabytes of new information in 2002.
  - 92% of the new information was stored on magnetic media, mostly in hard disks.
  - That’s 800 MB per person on the globe
  - That’s 37.000 times the LoC
  - That’s 30% more than in 1999

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]

5 Exabytes of new information in 2002



Magnetic Storage Chart



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

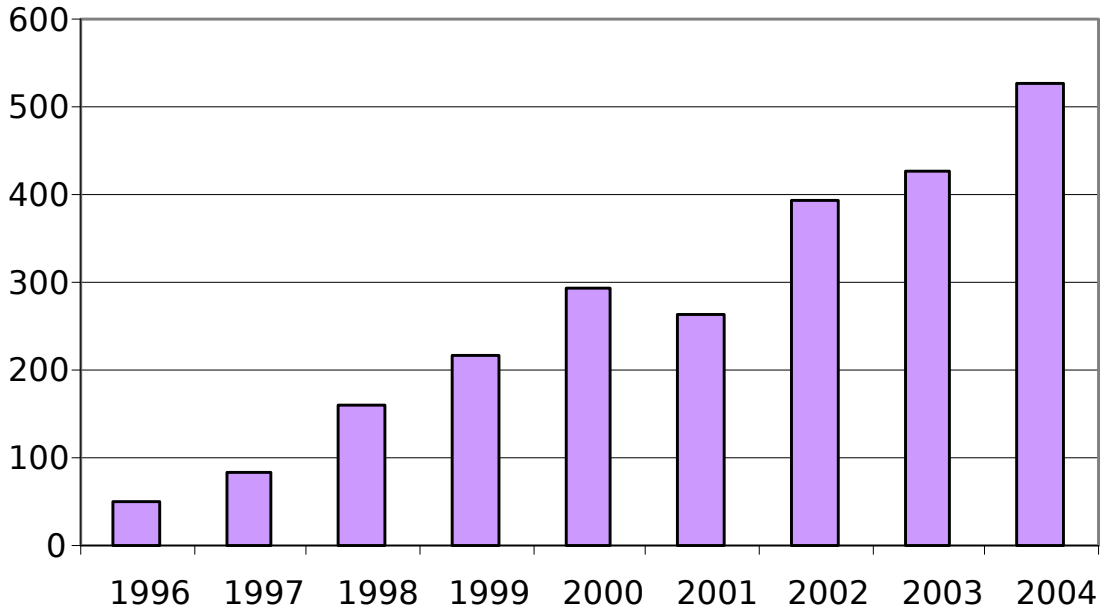
Information Usage

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

Interest in Multimedia Retrieval

- Documents in IEEE Online Database:
  - `((image retrieval) or (multimedia retrieval) or (video retrieval)) <in> metadata`



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
- Online Digital Video
  - Video in web sites, no streaming servers

Digital Imaging Devices (global)

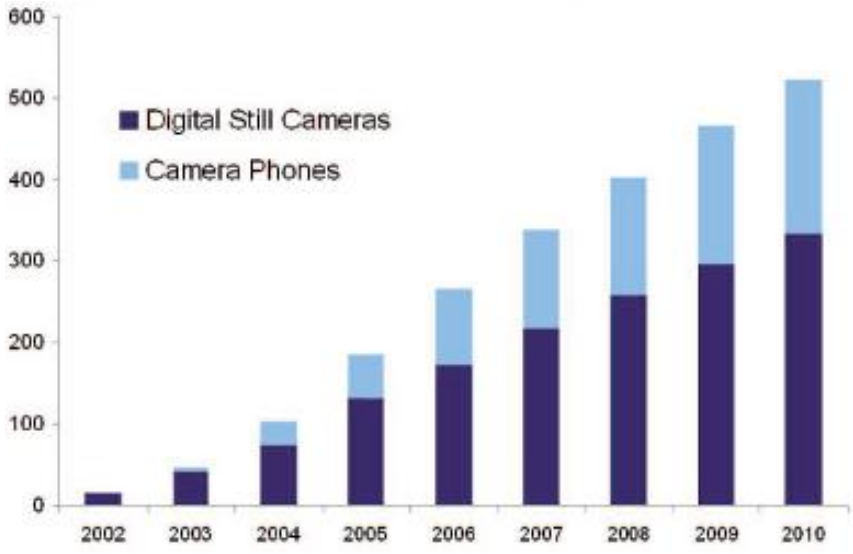
- First question: How many devices exist?

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)

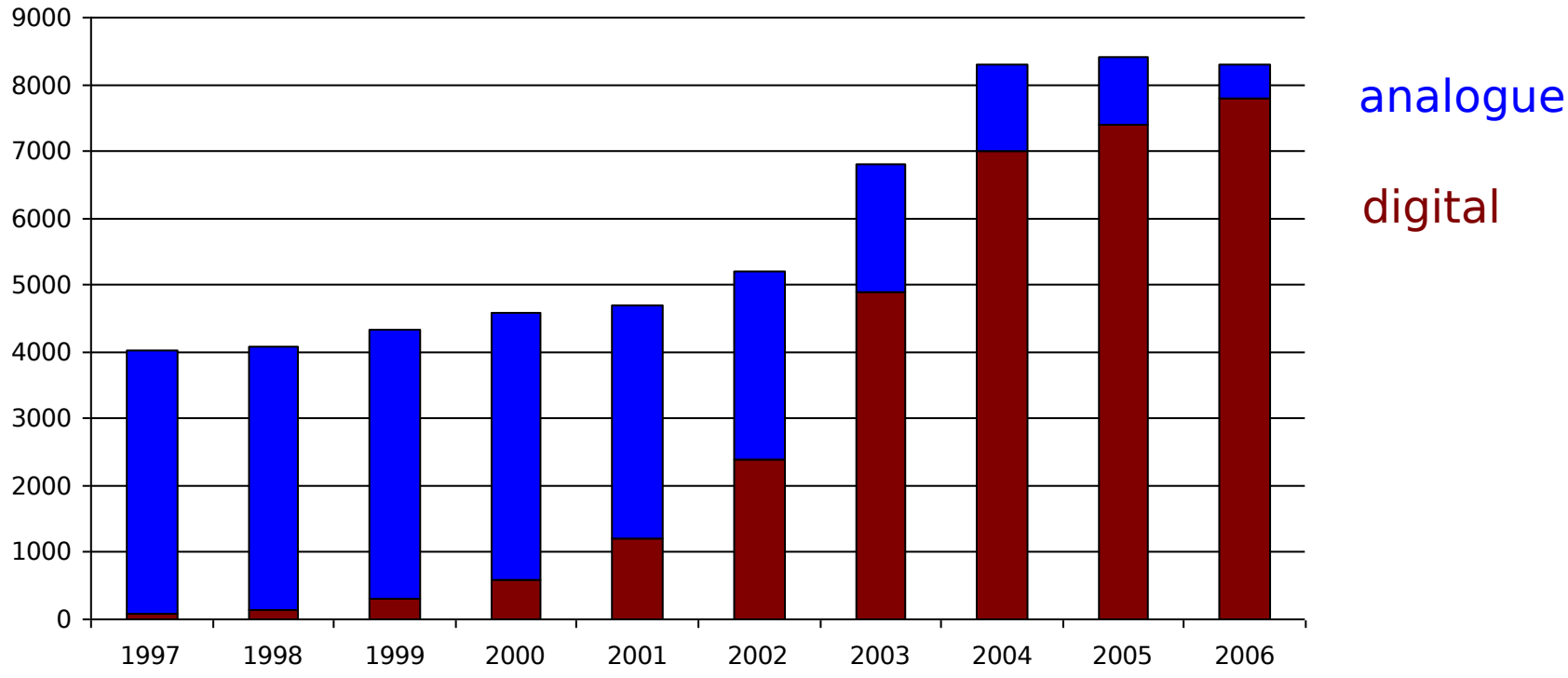
- Estimate 2006
  - > 150 billion photos from cameras
  - > 100 billion photos from camera phones
- Forecast 2010
  - > 500 billion photos
  - + increased resolution



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

Digital Imaging Devices (Germany)

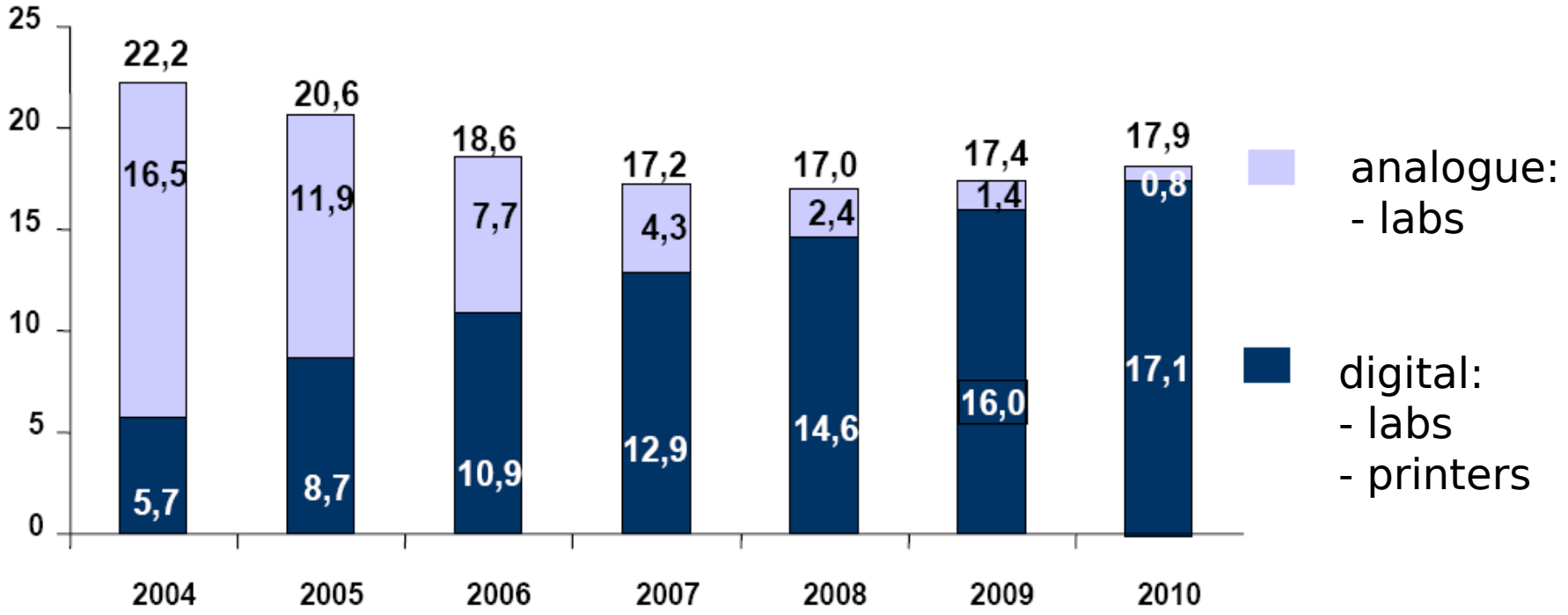
Still image cameras sold in Germany (thousands)



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

Photo prints market (Western Europe)

- Photo prints forecast (in billions)

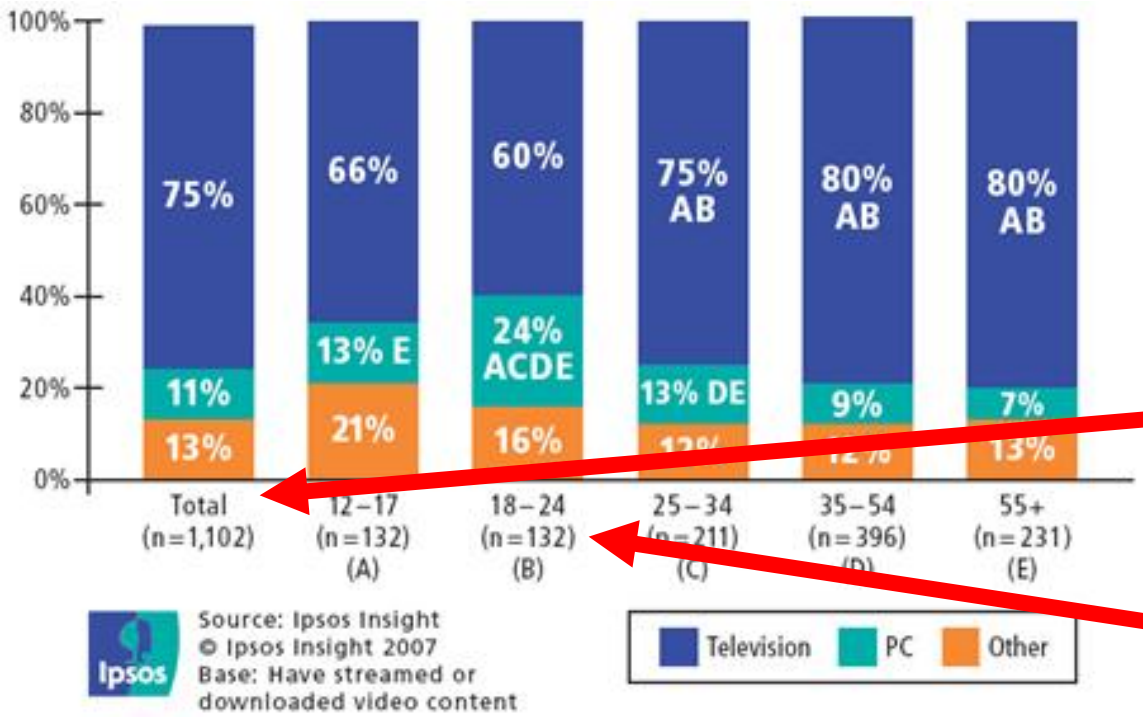


Source: Cewe Factbook, http://www.cewecolor.de

Content

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

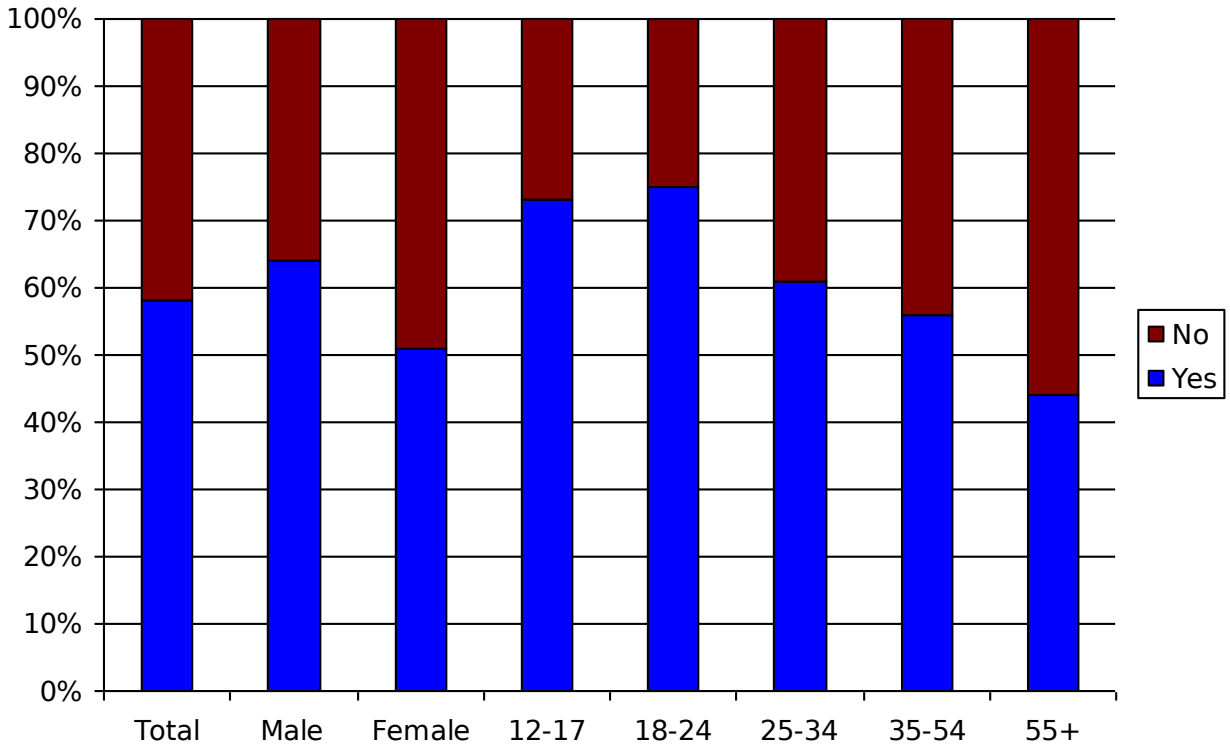
Where video content is watched? (US, 2007)



- 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.

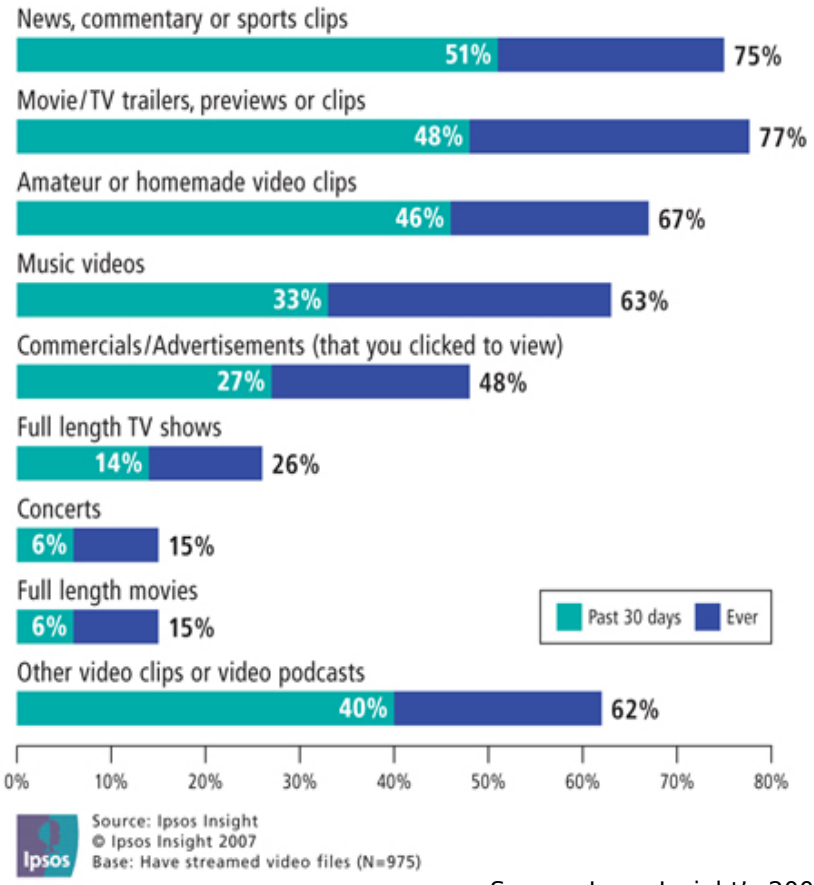
Source: Ipsos Insight's 2007 MOTION Study - http://www.ipsosinsight.com/pressrelease.aspx?id=3500

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)



- 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 (US, 2007)

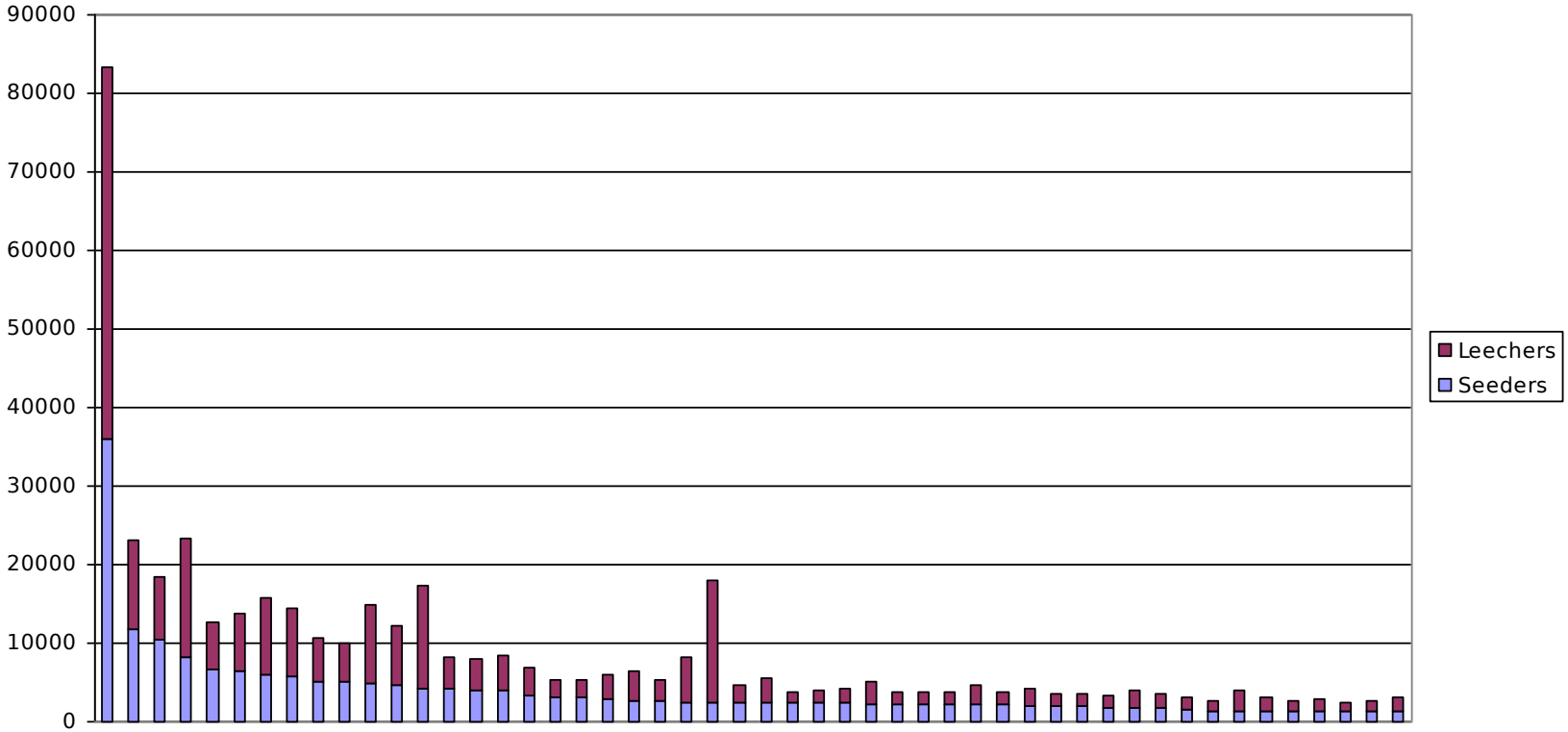
- Unwillingness to pay for this content,
- Difficulty or inability to
  - burn these files onto DVD
  - watch this content on their living room TV.

But there are also benefits in the future:

- 24/7 access
- reasonable fees or free (ad supported)

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

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

Project Topic

- Implement (enhance) a Bittorrent client
  - That logs activity on given torrents
  - Files stats on clients, seeders and leechers
- Analyze data
  - Find power law distribution and
  - Up & Downs (popularity)

Conclusion - Bittorrent

- Movie download is rather easy
  - Implicit quality control (release groups, tracker quality, stats, comments)
- Only very few users seed, therefore
  - Seeds result in multiple number of downloads
    - Not all leechers end up as seeders
  - Exact stats are not available

Content

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

3 Hours on Average (US, July 2007)

In July 2007:

- **74.2%** of U.S. internet users watched online videos
- For **3 hours** on average
- With an average video duration of **2.7 minutes**, on average **68 videos**.
- **36.7%** of U.S. internet users watched online videos on YouTube

Source: Comscore Press Release, http://www.comscore.com/press/release.asp?press=1678

Online Video Market Shares (US, July 2007)

	Views	
	millions	(%)
Google Sites	2,454	27.0%
Yahoo! Sites	390	4.3%
Fox Interactive Media	298	3.3%
Viacom Digital	281	3.1%
Disney Online	182	2.0%
Time Warner Network	181	2.0%
Microsoft Sites	149	1.6%
ESPN	75	0.8%
Veoh.com	53	0.6%
Comcast Corporation	51	0.6%

- Nearly all of Google's views come from YouTube

Source: Comscore Press Release, http://www.comscore.com/press/release.asp?press=1678

Unique Online Video Viewers (US, July 2007)

	Unique Viewers (thousands)	U.S. Internet Reach
Total Internet	*133.646*	*74.2%*
Google Sites	67.782	37.6%
Fox Interactive Media	35.834	19.9%
Yahoo! Sites	35.325	19.6%
Time Warner Network	26.571	14.8%
Viacom Digital	22.652	12.6%
Microsoft Sites	18.847	10.5%
Disney Online	13.907	7.7%
ESPN	7.733	4.3%
MLB.com	7.671	4.3%
Photobucket.com LLC	6.684	3.7%

- 134 Mio. U.S. Americans viewed online videos in July 2007

Source: Comscore Press Release, http://www.comscore.com/press/release.asp?press=1678

When do users watch? (US, January 2007)

Video Consumption Analysis by Daypart (comScore, January 2007, U.S. Internet Users)			
Daypart Time Segments	**Daypart Hours as a Percent of Total Week**	**Daypart Share of Weekly Video Consumption**	**Daypart Index***
Weekday Total	**71.4%**	**74.7%**	**105**
Monday - Friday, 1:00 A.M. - 7:00 A.M.	17.9%	5.9%	33
Monday - Friday, 7:00 A.M. - 10:00 A.M.	8.9%	6.9%	78
Monday - Friday, 10:00 A.M. - 5:00 P.M.	20.8%	30.5%	146
Monday - Friday, 5:00 P.M. - 8:00 P.M.	8.9%	14.3%	160
Monday - Friday, 8:00 P.M. - 11:00 P.M.	8.9%	11.8%	132
Monday - Friday, 11:00 P.M. - 1:00 A.M.	6.0%	5.4%	90
Weekend Total	**28.6%**	**25.3%**	**88**
Saturday - Sunday, 1:00 A.M. - 8:00 A.M.	8.3%	2.8%	34
Saturday - Sunday, 8:00 A.M. - 1:00 P.M.	6.0%	5.4%	91
Saturday - Sunday, 1:00 P.M. - 7:00 P.M.	7.1%	8.5%	119
Saturday - Sunday, 7:00P.M. - 11:00 P.M.	4.8%	6.2%	131
Saturday - Sunday, 11:00P.M. - 1:00 A.M.	2.4%	2.3%	96

Daypart Index = (Daypart Share of Weekly Video Consumption / Daypart Hours as a percent of Total Week) x 100

Source: Comscore Press Release, http://www.comscore.com/press/release.asp?press=1264

Current Software

- Joost (former: The Venice Project)
  - U.S. based, free
  - 2/3 p2p, AVC, client software
- Azureus Vuze
  - U.S. based, partially free
  - HD, bittorrent, ~streaming, Java software
- Babelgum
  - EU based, free
  - streaming + p2p, QT client software

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!