

## VK Multimedia Information Systems

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

#### Dienstags, 16.00 Uhr c.t., E.2.69



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 2.0 License. See http://creativecommons.org/licenses/by-nc-sa/2.0/at/

Mu 🕑 Department for Information Technology, Klagenfurt University, Austria

# Präsentationen 03.06.

http://www.uni-klu.ac.at

- Video Summary Pörtsch
- Flickr Related Tags Pitman
- MP3 LSA Waltl & Jonke
- YouTube Usage Bartha



# Präsentationen 24.06.

http://www.uni-klu.ac.at

- LSH Kofler & Pasterk
- Image Features Finke & Unterzaucher
- Seam Carving Wanschou
- Retrieval von Intentionen Doliner & Irrasch



## Contents



#### MPEG-7

- MPEG-21
- Metadata Generation & Annotation
- Social Software & Metadata





ITEC, Klagenfurt University, Austria – Multimedia Information Systems

## MPEG-7



- ISO/IEC Standard: Multimedia Content
  Description Interface
- Moving Pictures Expert Group
  - Specification goes on ...
- It's based on XML (Schema)
  - Binary representations possible (BiM)
- Allows differing granularity of descriptions
  - Extensive to very simple

# **MPEG-7 History**



- Call for Proposals: October 1998
- Evaluation: February 1999
- First version of Working Draft (WD): December 1999
- Committee Draft (CD): October 2000
- Final Committee Draft (FCD): February 2001
- Final Draft International Standard (FDIS): July 2001
- International Standard (IS): September 2001

## **MPEG-7 Basics**

- Descriptors
  - Syntax and semantics of exactly one (low or high level) elementary feature
  - Also base data types are defined
- Description Schemes
  - Defines structures within a framework
- Description Definition Language (DDL)
  - Extension of XML Schemes
- Coding Schemes
  - Create and interpret descriptions in BiM



@klu 🕖 ITEC, Klagenfurt University, Austria – Multimedia Information Systems

## **MPEG-7 Parts**



#### 1. MPEG-7 Systems

- Tools needed to prepare MPEG-7 descriptions for efficient transport and storage and the terminal architecture.
- 2. MPEG-7 Description Definition Language
  - Language for defining the syntax of the MPEG-7 Description Tools and for defining new Description Schemes.
- 3. MPEG-7 Visual
  - Description Tools dealing with (only) visual descriptions.
- 4. MPEG-7 Audio
  - Description Tools dealing with (only) audio descriptions.
- 5. MPEG-7 Multimedia Description Schemes
  - Description Tools dealing with generic features and multimedia descriptions.

## **MPEG-7 Parts**



- 6. MPEG-7 Reference Software
  - Implementation of relevant parts of the MPEG-7 Standard with normative status.
- 7. MPEG-7 Conformance Testing
  - Guidelines and procedures for testing conformance of MPEG-7 implementations
- 8. MPEG-7 Extraction and Use of Descriptions
  - Informative material about the extraction and use of some of the Description Tools.
- 9. MPEG-7 Profiles and levels
  - Provides guidelines and standard profiles.
- 10. MPEG-7 Schema Definition
  - Specifies the schema using the Description Definition Language

# **Scope of MPEG-7**



from: http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm

uni@klu 🕐 ITEC, Klagenfurt University, Austria – Multimedia Information Systems

## **Basic Elements**

Basic elements are fundamental constructs and used throughout the whole MPEG-7 description

- Basic datatypes
  - Time and date, relative and absolute
  - Numeric datatypes like matrices and vectors
- Links & Media Localization
  - Interconnections and content linking

## **Navigation & Access**



- Descriptors for Browsing & Retrieval
  - Summaries
  - Partitions (time, space & frequency)
  - Decompositions (time, space & frequency)
  - Variations



## **User Interaction**

http://www.uni-klu.ac.at

- Pertaining consumption of AV data
  - user preferences
  - usage history
- Meant to facilitate personalization
  - Matching User Interaction DS with content description
  - Is research topic @ ITEC

# **Content Organization**



- Organization & modelling of collections
  - Audio-visual content, segments, events, and/or objects
    - E.g. pictures, scenes, music files, etc.
  - Allows collection description as a whole
    - E.g. "Pictures of my holiday in Ebonia"



# **Content Management**



- Creation & Classification
  - Title, textual annotation, creators, creation locations, and dates.
  - Categories such as genre, subject, purpose or language.
  - Review and guidance information: Age classification, parental guidance, and subjective review.
  - Related material information.
- Media coding, storage & file formats
  - Media profiles & master media
- Content Usage
  - Usage rights, usage record, and financial information

#### **Content Description: Structural vs. Conceptual Aspects**

http://www.uni-klu.ac.at

- Program DS (in sense of TV program)
- Analogy to
  - Table of content Region tree (linear partitioning)
  - Index Object tree (non-linear structure)



## **Content Description: Structural Aspects**



http://www.uni-klu.ac.at

- Divide a video stream into physical and logical video segments
- The higher the level of a physical video unit, the more semantic information is necessary
- Logical units are based on semantic content



### **Region and Object Trees**

http://www.uni-klu.ac.at

#### Region Tree







ITEC, Klagenfurt University, Austria – Multimedia Information Systems

## **Content Description: Semantic Aspects**



#### Low Level Features

- Extraction from Content
- Descriptors for
  - Shape, color, texture (visual)
  - Timbre, rhythm (audio)
- High Level Features
  - Annotation
  - So called semantic descriptors
    - Textual information
    - Conceptual information

# MPEG-7 High Level Descriptors



- Textual Descriptions
  - Text to describe temporal / spatial regions
- The W's
  - Structured way of textual descriptions
    - Who, Where, What Object, When, Why, How & Where
- Instead of textual descriptions
  - Controlled Terms
    - Dictionaries, Taxonomies, Classifications Schemes
  - Semantic Description Scheme

# MPEG-7 Semantic Description Scheme

http://www.upi.klu.ac.at



uni@klu

ITEC, Klagenfurt University, Austria – Multimedia Information Systems

# Actual Description in MPEG-7



Narrative World

## Contents



• MPEG-7

#### MPEG-21

- Metadata Generation & Annotation
- Social Software & Metadata





ITEC, Klagenfurt University, Austria – Multimedia Information Systems

# MPEG-21 – motivation and scope



http://www.uni-klu.ac.at



uni@klu 💓 ITEC, Klagenfurt University, Austria – Multimedia Information Systems

# **MPEG-21 Objectives**

http://www.uni-klu.ac.at

MPEG-21's goal is to create an *interoperable and integrated multimedia framework* in three steps:

- Develop "big picture": understand how the components of the framework are related and identify where gaps in the framework exist
- 2. Fill the gaps: develop new standard specifications where needed
- Integrate: achieve the integration of standards to support harmonized technologies for the management of multimedia content

# **MPEG-21 Digital Item**

http://www.uni-klu.ac.at

- A **Digital Item (DI)** is a structured digital object with a standard representation, identification, and metadata within the MPEG-21 framework
- Digital Items are "the content"
- DIs consist of
  - Resources (individual assets, distributed content),
  - Metadata (data about or pertaining the DI) and
  - Structure (relationships between parts of the DI)

# **Digital Item - Example**

http://www.uni-klu.ac.at



## The DI is the fundamental unit for distribution and transaction within the MPEG-21 framework.

uni@klu 💓 ITEC, Klagenfurt University, Austria – Multimedia Information Systems

# MPEG-21 User and User Interaction



- Any entity that interacts in the MPEG-21 environment or makes use of a Digital Item
- Users include individuals, organisations, corporations, consortia, governments, other standards bodies, etc.
- Roles including creators, consumers, rights holders, content providers, distributors, etc.
- Each User will assume specific rights and responsibilities according to their interaction with other users



## **Seven Architectural "Elements"**





# **Roles of the Architectural Elements**

http://www.uni-klu.ac.at



uni@klu 🔍

ITEC, Klagenfurt University, Austria – Multimedia Information Systems

## Contents



- MPEG-7
- MPEG-21

#### Metadata Generation & Annotation

Social Software & Metadata





ITEC, Klagenfurt University, Austria – Multimedia Information Systems

## **Metadata Generation & Annotation**



- Process of creating data about data
- Content has to be known
  - Watch & understand video / image collection
  - Listen and assess audio
- Metadata standard has to be known
  - What are the possible fields
  - What are the used classification systems

# **Evaluation (1/2)**

http://www.uni-klu.ac.at

- Goal: Identify the opinion of users on manual semantic annotation
- 5 Users with following (median) background:
  - 17 years of computer experience
  - Using a computer 50 h / week
  - 2 years experience with digital photo cameras
  - 4 years experience with imaging software

# **Evaluation (2/2)**



- 2 Tasks were given:
  - Annotate a photo with a given description and an extensive prior introduction to semantic photo annotation with Caliph
    - video was shown,
    - concept was explained and
    - questions were answered
  - Annotate a photo fully on your own
  - After Tasks:
    - Questionnaire with several subjective questions
    - Evaluation Scale from: -3 (strongly disagree) to 3 (strongly agree)

# **Evaluation Results: General Questions**



- The concept of meta data is very new to me: -2.6
- It was easy to understand the concept of semantic meta data while using Caliph: 1.8
- The visualization of the semantic meta data within Caliph is easy to understand and interpret: 2.2
- The annotation of images with textual descriptions can be done fast and easily: 1.4
- The annotation of images with semantic meta data can be done fast and easily: 1.2
- I can see an obvious benefit by using semantic meta data for image (multimedia) annotation: 1.4

Scale: (disagree) -3 to 3 (agree)

## **Evaluation Results: Scenario based questions**

http://www.uni-klu.ac.at

- The complexity of semantic annotation is too high to be useful for organizing photos.
- I would find it easy to annotate a large set digital photos (e.g. 100+).
- I would recommend Caliph or a similar tool to annotate digital photos.
- I can see an obvious benefit by using semantic meta data for the organization of photos.

| Personal | Newspaper |
|----------|-----------|
| -0.6     | -1.8      |
| -0.6     | -0.2      |
| 0.8      | 1.4       |
| 1.4      | 2.2       |

Scale: (disagree) -3 to 3 (agree)
## **Evaluation Results: Annotation performance**

25 20 15 min. ■ Test 1 ■ Test 2 10 5 0 User 1 User 2 User 5 User 3 User 4

uni@klu 🕑

ITEC, Klagenfurt University, Austria – Multimedia Information Systems

## **Evaluation Results: Annotation performance**

- Median times for annotation:
  - 15.4 minutes for the 1<sup>st</sup> test and
  - 6 minutes for the 2<sup>nd</sup> test
- Median time in a self test with 17 photos:
  - 1 minute and 53 seconds per photo
- That results in an approximate time of 10 h 27 min. for annotation of a set of 333 photos

http://www.uni-klu.ac.at

#### Structured text annotation field "Who":

- 1. Vedran, Wolfgang, Armin
- 2. Wolf, Armin, Vedran
- 3. Wolfgang Kienreich, Vedran Sabol, Armin Ulbrich
- 4. wolfgang, armin, vedran
- 5. W.Kienreich, A.Ulbrich, V.Sabol



#### • Free text annotation:

- 1. Stadthalle, Graz, Austria I-Know '04 Knowledge Managment Conference
- 2. The three are sitting ...
- Wolfgang Kienreich, Armin Ulbrich und Vedran Sabol (v.l.n.r.) sprechen miteinander auf der I-Know.Wolfgang Kienreich, Vedran Sabol, Armin Ulbrich are at I-Know, Graz for Talking
- 4. Stadthalle, Graz, Austria I-Know '04 Knowledge Managment Conference
- 5. Wolfgang, Armin and Vedran talking to each other on I-Know 04 at Stadthalle Graz.

http://www.uni-klu.ac.at



uni@klu 🕑

ITEC, Klagenfurt University, Austria – Multimedia Information Systems

http://www.uni-klu.ac.at



ITEC, Klagenfurt University, Austria – Multimedia Information Systems

http://www.uni-klu.ac.at



uni@klu 💟 ITEC, Klagenfurt University, Austria – Multimedia Information Systems

## **Lessons Learned**



- Users like the graphical annotations editor
- Users see semantic annotation in a professional (business) environment
- Semantic annotation is very time consuming
- The MPEG-7 nomenclature is not intuitive
  - Semantic agent / place / object & relations
  - Creator of image / description / quality rating
- Tagging with central tag repository ...

### Demo



uni@klu ITEC, Klagenfurt University, Austria – Multimedia Information Systems

## Contents



- MPEG-7
- MPEG-21
- Metadata Generation & Annotation
- Social Software & Metadata



ITEC, Klagenfurt University, Austria – Multimedia Information Systems

## **Social Software**

http://www.uni-klu.ac.at

- Social Software
  - Integration of the User
  - Common in the Web 2.0
  - User participate
- Social aspects
  - U. connect to users -> Social Networking
  - U. connect to information -> 43things.com
  - U. connect to resources -> social bookmarking
  - U. connect to media -> social media sharing

## **Example: Social Bookmarking**



Social Bookmarking defined:

- Bookmarking Resources
- Providing a "stream of bookmarks"
- Eventually additional support for
  - Tagging (keywords)
  - Caching (Saving the state of the bookmark)
  - Organization & Collaboration (Groups)



## **Example: del.icio.us**

| Elle       Edit       View       Higtory       Bookmarks       Tools       Help://del.icio.us/         Image: Complex of the complex of   | earch<br>ut   help<br>nore               |  |
|---|--|--|
| Image: Construct of the second of | earch<br>ut   help                       |  |
| del.icio.us        hotlist     what's hot right now on del.icio.us     see also: popular   recent     interesting   Backgrounds save this   rist posted by waphle   rist posted by waphle   css   webdesign   table   <   | earch<br>ut   help<br>nore               |  |
| hotlist what's hot right now on del.icio.us tags to watch   Matter                                      | ut   help<br>nore                        |  |
| hotlist       what's hot right now on del.icio.us       tags to watch       mage         See also: popular   recent       see also: popular   recent       interesting       interesting         Image: See also: popular   recent       180 people       10 Ways Women Judge You       Seed: Who Wants to Be a Cognitive         Image: Seed: Who Wants to Be a Cognitive       Seed: Who Wants to Be a Cognitive       Seed: Who Wants to Be a Cognitive         Image: Seed: Who Tik Routers and Wireless save this       163 people       Image   | nore                                     |  |
| A List Apart: Articles: Super-Easy Blendy       180 people       10 Ways Women Judge You         Backgrounds save this       Seed: Who Wants to Be a Cognitive         first posted by waphle       css webdesign       tutorial       images       tags         Mikro Tik Routers and Wireless       save this       163 people       linux  |  |  |
| A List Apart: Articles: Super-Easy Blendy       180 people       10 Ways Women Judge You         Backgrounds save this       Seed: Who Wants to Be a Cognitive         first posted by waphle       css       webdesign       tutorial       images       tags         MikroTik Routers and Wireless       save this       163       people       linux   |  |  |
| Backgrounds save this       Seed: Who Wants to Be a Cognitive         first posted by waphle       css       webdesign       design       tutorial       images       tags         MikroTik Routers and Wireless       save this       163       people       linux   |  |  |
| Extremely Useful Macbook Freeware  MikroTik Routers and Wireless save this  IG3 people  linux   |  |  |
| MikroTik Routers and Wireless save this 163 people  | Extremely Useful Macbook Freeware        |  |
| linux   |  |  |
| first posted by martinoso network tools freeware monitor tags   |  |  |
| software Top 10 Free Linux Games  | -1.1                                     |  |
| games/movie console - Software Reality  | 310<br>'                                 |  |
| Tross Knot save this 177 people Mount your widows Partitions and make   | Mount your widows Partitions and make it |  |
| first posted by pemdp howto tie ties knot tips tags read/writable in ubuntu Debian Admin  |  |  |
| a search  |  |  |
| Articles save this 146 people Like.com Visual Search  | Like.com Visual Search                   |  |
| first posted by dcogs mac howto mediacenter tv hometheater tags Ms. Dewey   |  |  |
| LISZEN: Search 500+ Library Blogs   | LISZEN: Search 500+ Library Blogs        |  |
| 10 Things I Wish I Had Never 176 people div   |  |  |
| Believed-steve-olson.com save this O'Reilly: Hacking iPod and iTunes  |  |  |
| first posted by solson lifehacks life advice productivity business tags Make Home Page  |  |  |
| TNPI - Homemade (Do It Yourself) .mac   | using 🚽                                  |  |

## **Example: del.icio.us**



## **Example: del.icio.us**



- User Interface
  - Clean and easy2use
  - Powerful tools (bookmarklets & plugins)
- Additional Features
  - Thumbnails
  - Social Networking



## del.icio.us



- User intentions are unclear:
  - Self-organization or group organization
  - Participation / Being part of it
- Explicitly Generated
  - Bookmarking & Tagging
  - Tag Bundles
- Implicitly Generated
  - Time, Interestingness, The "Seen Web"
  - User Profile, Social Network

del.icio.us

social bookmarking

## **Examples: Social Media Sharing**

- Flickr.com, Bubbleshare.com, Zooomr.com, ...
  - Sharing images & annotations
- YouTube.com, Google Video, VideoEgg.com. ...
  - Sharing videos & annotations
- Pandora, Last.fm
  - Sharing music & flavors











## **Example: Google Video**

http://www.uni-klu.ac.at



## Metadata in Social Software?



#### Bottom up

- In contrast to controlled vocabularies
- In contrast to quality ensured content creation processes
- Superimposed structure
  - Instead of using predefined hierarchies
  - Through heavy use of linking / interrelation
- Huge and fuzzy
  - Unimaginable mass of links & tags
  - Lots of redundant information
- Spammed
  - Just starting ...

## **Folksonomies**

http://www.uni-klu.ac.at

- Definition & Description
- Why do tagging systems work?



## Folksonomies



Network of Tags, Users and URLs

- Users describe resources
- By using (multiple) tags

#### Examples:

Social bookmarking, media sharing, etc.

## Folksonomies: The Structure



User *tags* resource (URL)

- 1+ words or phrases (bonn, "mathias lux")
- No controlled vocabulary, taxonomy
- No quality control
- No constraints (language, length, number)

## **Folksonomies: Structure**

http://www.uni-klu.ac.at

- Tag to URL is a n:m relation
- Superimposed structure through bidirectional links
- Structure is called "folksonomy"





ITEC, Klagenfurt University, Austria – Multimedia Information Systems

## Folksonomy Example: Flickr





klu 🕐 ITEC, Klagenfurt University, Austria – Multimedia Information Systems

60

## Folksonomy Example: Technorati





## **Folksonomy Example: 43things**







ITEC, Klagenfurt University, Austria – Multimedia Information Systems uni@klu

## **Types of Folksonomies**

http://www.uni-klu.ac.at

- Narrow folksonomies
  - Each one tags her/his own resources
  - All above examples are narrow f.
- Broad folksonomies
  - Each tags whatever s/he wants
  - Example: Social bookmarking
- Difference
  - Narrow folksonomies are more sparse

# Why do tagging systems work?



This was topic of a panel at CHI 2006, following conclusions were drawn:

- Tagging has a benefit for the user
  - Similar to bookmarking, integrated apps
  - Benefit of accessibility from everywhere in the internet
- Tagging allows social interaction
  - Connecting a user to a community trough tags
  - People can subscribe your stream

# Why do tagging systems work? (2)



- Tags are useful for retrieval
  - Synonyms and typos vanish in the mass of tags
  - Communities can retrieve "their" stuff (e.g. by special tag)
- Tagging Systems have a low participation barrier
  - Apps are easy to use, intuitive, responsive
  - Free text is used to do the tagging
  - Requires no previous considerations & training

## **Folksonomy Analysis**

http://www.uni-klu.ac.at

Some scientific background ...

image from http://www.squaredot.com/geek.html



## **Unified Model for Social Networks & Semantics**



Mika P. (2004) "Ontologies are us: A unified model of social networks and semantics"

- Ontologies contain instances I and concepts C
- Ontologies are formal specifications
  - Which are stripped from their original social context of creation
  - Which are static and may get outdated

# Where do semantics emerge from?



A third set besides C and I is needed

- Agents A are those who specify
- Agent defines
  - which Concept C is
  - assigned to Instance I
- ⇒ A tripartite model can be identified

## A tripartite model



- 3 partitions: *A*, *C* & *I*
- Hyperedges connect exactly one a ∈ A with one c ∈ C and i ∈ I
- One edge denotes that a user assigns a concept to a resource.

But tripartite graphs are rather hard to understand and to work with!



/ ITEC, Klagenfurt University, Austria – Multimedia Information Systems

## **Simplifying the tripartite Model**



Similar to the introduced structure of folksonomies:

- An instance is connected to a concept
  - like a tag to a resource
- The edge is labeled by the user or
- Weighted by the number of assignments

## A bipartite Model ...

http://www.uni-klu.ac.at

- A graph connecting
- Instances i to
- Concepts c
- We call this IC-Graph
- The weights can be expressed in an association matrix

|    | <b>c1</b> | <b>c2</b> | <b>c3</b> | ••• |
|----|-----------|-----------|-----------|-----|
| i1 | 1         | 5         | 0         |     |
| i2 | 0         | 3         | 0         |     |
| i3 | 4         | 2         | 2         |     |
|    |           |           |           |     |



## **The Association Matrix**

- This matrix connects two different sets
- Folding allows to transform the Matrix to a one mode network
- Just like the co-occurence matrix in text retrieval:

$$M_c = M_{IC} \cdot M'_{IC}$$

 Result is a matrix connecting concepts to concepts
## **Example: Concepts**

| i1      | ح المالح | 5 pda | cellphone | o wlan    | r network |          |    |     |     |                                       |
|---------|----------|-------|-----------|-----------|-----------|----------|----|-----|-----|---------------------------------------|
| i2      | 7        | 1     | 1         | 1         | 2         |          |    | e   |     |                                       |
| i3      | 0        | 4     | 5         | 0         | 0         | ute      |    | IOL |     | - Y                                   |
| i4      | 8        | 0     | 0         | 0         | 6         | d        |    | hd  | С   | N N N N N N N N N N N N N N N N N N N |
| i5      | 3        | 3     | 0         | 4         | 0         |          | qa | ell | /la | et                                    |
|         |          |       |           |           |           | <u> </u> | 0  | U   | 5   | <u> </u>                              |
|         |          |       | CC        | computer  |           | 111      | 62 | 20  | 62  | 60                                    |
| [       |          |       | ро        | da        |           | 62       | 56 | 9   | 68  | 28                                    |
|         |          |       | Ce        | cellphone |           | 20       | 9  | 41  | 0   | 12                                    |
| wlan    |          |       | 62        | 68        | 0         | 100      | 24 |     |     |                                       |
| network |          |       | 60        | 28        | 12        | 24       | 34 |     |     |                                       |

uni@klu [ ITEC, Klagenfurt University, Austria – Multimedia Information Systems

# **The Association Matrix**



 Also instance based co-occurrence can be calculated

$$M_{I} = M_{IC}' \cdot M_{IC}$$

- Based on the co-occurrence clustering algorithms can be applied:
  - Instance Clustering
  - Concept Clustering

## **Other Association Matrices**



- Based on the AC-Graph
  - Bipartite agent2concept graph
  - Instances are used as weights
- Based on the AI-Graph
  - Bipartite agent2instance Graph
  - concepts are used as weights
- Based on A[C|I]-Graph the social network between agents can be analyzed

#### **Application to Folksonomies**



- Concepts, agents and instances in Folksonomies:
  - Tags are concepts
  - Agents are users
  - Resources are instances
- Tags are error prone, but semantics can eventually emerge (see P. Mika for the example del.icio.us)

## **Problems of the approach**

- Community based concepts & associations
- Tags have typos, synonyms
- Tags have different intentions
  - Abstract semantics (funny, sad, friendship)
  - Media description (pdf, online, word, image)
  - Rights and authors (persons names)
  - Organizational (2read, todo, marker)
  - etc.

## **Problems of the approach**

- Computational problems
  - Big matrix multiplications are hard to compute
- Narrow folksonomies restrict tagging to the originating user:
  - Flickr tags can only be assigned by the uploader
  - YouTube has the same restriction

#### Folksonomy Analysis Example





## **Tag Gathering: del.icio.us**

- Based on RSS feeds of del.icio.us
  - Read main feed
  - Get entries for each user
- Avoid spamming
  - Use entries on URIs with a min. of 2 users
- Write to relational database
  - In this case MySQL 5.1

## **Tag Database**





| entry2tag                  |   |  |  |  |  |  |
|----------------------------|---|--|--|--|--|--|
| PK<br>PK,I2,I1<br>PK,I3,I1 | <u>usernameID</u><br><u>urIID</u><br><u>tagID</u> |  |  |  |  |  |
|                            |   |  |  |  |  |  |

| t  | ags  | urls       |           | usernames  |           |
|----|------|------------|-----------|------------|-----------|
| PK | id   | PK         | <u>id</u> | PK         | <u>id</u> |
| 11 | name | <b>I</b> 1 | name      | <b>I</b> 1 | name      |

uni@klu 🕑

ITEC, Klagenfurt University, Austria – Multimedia Information Systems

## **Tag database issues**

http://www.uni-klu.ac.at

- Group by & Having, Indexes
- Memory (temp tables)
- MySQL is just like Oracle:
  - tune it or leave it.
- Sample statement Top tags:

SELECT COUNT(e.tagid), t.name, t.id FROM entry2tag e, tags t WHERE t.id = e.tagid GROUP BY e.tagid ORDER BY 1 DESC

# **Tag similarity**



- Tags are assigned to resources
- Tags describe same URIs-> Similarity
  - E.g. Javascript & Ajax
  - E.g. Windows & Software
  - E.g. Linux & Kernel
- Tags never describe same URIs-> Dissimilarity
  - E.g. Free & Shop
  - E.g. Usability & SAP

## **Tag Merging: Objectives**

- Main problems within del.icio.us (and possibly in many folksonomies due to their nature)
  - Synonyms
  - Basic level variation
- Encounter these problems by "merging" synonyms
  - Different spellings: e.g. eLearning & e-Learning
  - Typos & plurals

## **Tag Networks: Objectives**

- What is the conceptual structure within a community?
- Which tags are similar / interconnected?
- Direction of the connection?
- Probability of transition for network edges?
- Network Analysis?
  - Hubs, central authorities
  - Clusters

## **Tag Centrality: Objectives**

- Which are the most prominent nodes?
- Based on different measures?
  - In degree
  - In Betweenness
  - PageRank / HITS
- The removal of central nodes would hit the connectivity hard!

#### Tag Clustering: Objectives

- What are interesting conceptual clusters?
  - {design, webdesign, graphics}
  - {html, xhtml, css}
  - {ajax, javascript, prototype, script.aculo.us}
- What is a meaningful disambiguation of a topic / tag?
  Clusters of tag programming
  - 1. <u>systems+unix</u> (3,42)
  - 2. developer+development (2,49)
  - 3. webdevelopment+javascript+webdev (2,34)
  - 4. ebook+books+book (2,19)
  - 5. Coding (2,19)
  - 6. programacao+ruby (2,14)
  - 7. script+ajax (1,78)
  - 8. <u>DotNet+.NET</u> (1,65)





#### ... for your attention



ITEC, Klagenfurt University, Austria – Multimedia Information Systems