MUCKE

Multimedia and User Credibility Knowledge Extraction

http://ifs.tuwien.ac.at/~mucke/

Mihai Lupu

Vienna University of Technology

lupu@ifs.tuwien.ac.at

Tea

- Bil
- “A
- Vie
- Ce
  Fra

- Adrian Iftene (bureaucrat, administrator) (Created on November 16, 2012 at 10:57)
- Adrian Popescu (bureaucrat, administrator) (Created on November 16, 2012 at 10:57)
- Ahmet.Iscen (Created on April 11, 2013 at 10:20)
- Alexandra.Siriteanu (Created on April 11, 2013 at 10:20)
- Alexandru.Ginsca (Created on April 11, 2013 at 10:20)
- Andrei Panu (Created on January 29, 2014 at 08:29)
- Anil.Armagan (Created on April 11, 2013 at 10:20)
- ~~Claudia.Gheorghiu (Created on April 11, 2013 at 10:20)~~
- Cristina.Serban (Created on April 11, 2013 at 10:20)
- Diana.Trandabat (Created on April 11, 2013 at 10:20)
- Eren.Gölge (Created on April 11, 2013 at 10:20)
- Hanbury (bureaucrat, administrator) (Created on November 12, 2012 at 11:00)
- ~~Hichem.Bannour (Created on April 11, 2013 at 10:20)~~
- Ilker.Sarac (Created on April 11, 2013 at 10:20)
- Kkontokotsios (Created on February 6, 2014 at 16:18)
- Lenuta.Alboaie (Created on April 11, 2013 at 10:18)
- MUCKEguest1 (Created on April 8, 2013 at 08:07)
- Mihaela.Breaban (Created on April 11, 2013 at 10:20)
- Mihai Lupu (bureaucrat, administrator) (Created on November 12, 2012 at 10:05)
- Mircea.petic (Created on October 3, 2013 at 08:08)
- Pinar Duygulu (bureaucrat, administrator) (Created on November 16, 2012 at 10:58)
- Ralf Bierig (bureaucrat, administrator) (Created on April 8, 2013 at 08:36)

CEA : LVIC - Laboratory for Vision and Content Engineering

- ~ 60 persons in all, with 25 people working on multimedia
  - 30 ongoing projects for the multimedia theme
    - USEMP, Periplus, Egonomy, DataScale, ePoolice
    - Large number of direct collaborations with industrial partners
  - ~35 publications/year
- Objective – understand and describe multimedia documents (text, image, video)
  - Information retrieval over multimedia collections
  - Document filtering using domain related criteria
  - Document summarization and presentation
- Application domains
  - Electronic content Management
  - Cultural heritage and tourism applications
  - Collaborative filtering for product and service proposal
  - Technological watch
- Participation to/organization of evaluation campaigns

BILKENT University

- the first private, nonprofit university in Turkey
- founded on October 20, 1984
- “Bilkent” = an acronym of "bilim kenti": Turkish for "city of learning and science.”

- Computer Engineering Department
  - 22 faculty members
  - algorithms, artificial intelligence, bioinformatics, computer architecture, computer graphics, computer networks, computer vision, cryptography, data mining, database systems, information retrieval, machine learning, parallel and distributed systems, performance evaluation, scientific computing, and software engineering.

“Al. I. Cuza” University

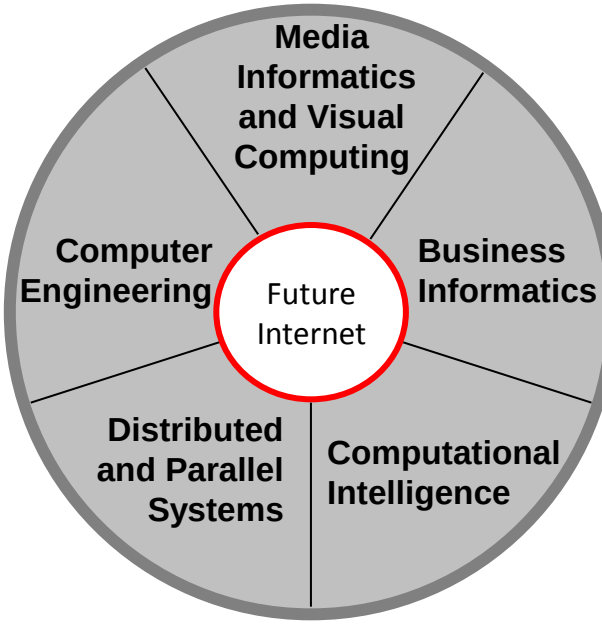
- Computer Science Department
  - 22 years Faculty of Computer Science
  - ~ 1400 students (1150 Bachelor, 200 Master,50 PhD Students)
  - ~ 40 Professors (9 Full Professors)
- Research Projects
  - Natural Language Processing – Dan Cristea
  - Software Engineering – Dorel Lucanu

- NLP
  - METANET4U
  - ATLAS
  - LT4eL
  - ELIAS
  - eDTLR
  - CLEF, TAC, RTE campaigns

- multilingualism, services, resources

TU Wien - Informatik

- Informatics Dept.
  - 7 Institutes
  - 19 Full Professors (+ 1 to be appointed)
  - 32 Associate Professors
  - Postdoctoral Researchers
  - Research Assistants (incl. external funding)
  - Technical and administrative Personnel (incl. external funding)
  - ~7.500 Students
- Information Management and Preservation Lab
  - Data Mining and Machine Learning
  - Information Retrieval
  - Digital Preservation
  - Led by Prof. Andreas Rauber
  - 20 people (of which 19 funded by external funds)



Project status

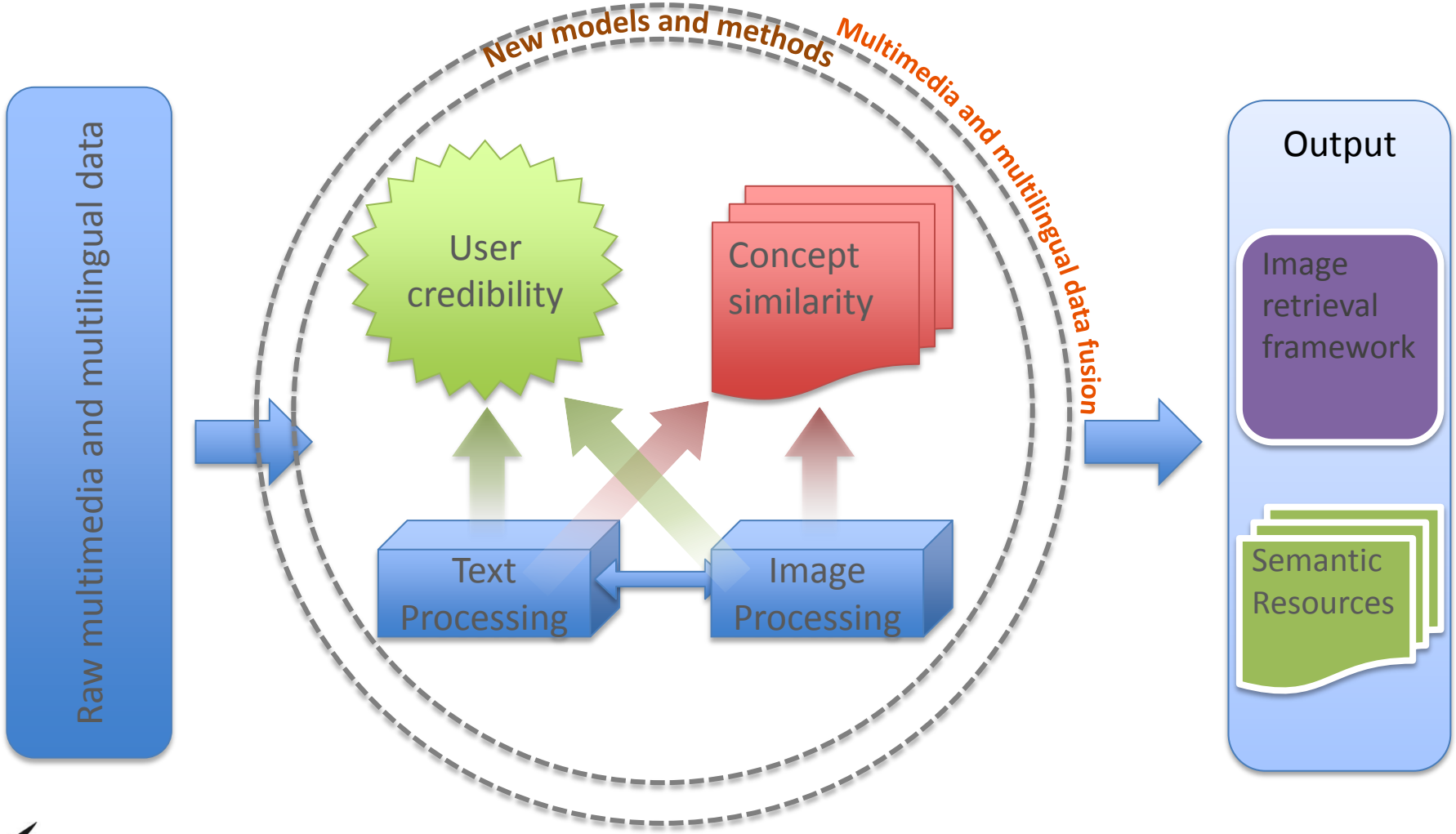
Start date: Oct 1st, 2012

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16	M17	M18	M19	M20	M21	M22	M23	M24	M25	M26	M27	M28	M29	M30	M31	M32	M33	M34	M35	M36	Total
WP 1 - Data Collection	2.8	2.8	2.8	2.8	2.8	2.8	1.8	1.8	1.8	1.8	1.8	1.8																									**28**
T1.1 Assessment and Collection of Existing Resources	1.0	1.0	1.0	1.0	1.0	1.0																															6
T1.2 New Data Collection	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8																									22
WP 2 - Multimedia Processing and Fusion		2.8	2.8	2.8	4.1	4.1	4.1	4.1	4.1	4.1	4.1	4.1	4.1	4.1	4.1	4.1	4.1	4.1	1.4	1.4	1.4																**70**
T2.1 Text processing		1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2																			20
T2.2 Image processing		1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6																			27
T2.3 Multimedia fusion					1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4																23
WP 3 - User Credibility for Multimedia Streams						1.4	1.4	1.4	1.4	1.4	3.6	3.6	3.6	3.6	3.6	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.2													**45**
T3.1 Credibility Model Definition						1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4																						14
T3.2 User Context Mining											1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0													14
T3.3 User content mining and fusion											1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2													17
WP 4 - Multimedia Concept Similarity Models													2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	**60**
T4.1 Textual Concept Similarity													1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	28
T4.2 Multimedia Concept Similarity													1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	32
WP5 - User-Centred Image Retrieval																								2.7	2.7	2.7	5.4	5.4	5.4	5.4	4.1	4.1	4.1	2.7	2.7	2.7	**50**
T5.1 - Image Search Results Diversification																								1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4				14
T5.2 - User-Centred Image Retrieval Model																								1.3	1.3	1.3	1.3	1.3	1.3	1.3							9
T5.3 - Retrieval Prototype Integration																											2.7	2.7	2.7	2.7	2.7	2.7	2.7	2.7	2.7	2.7	27
WP6 - Evaluation and Resource Sharing						0.8	2.3	2.3	2.3	0.8	0.8	1.3	1.3	1.3	1.3	1.3	1.3	1.3	0.8	0.8	0.8	1.3	1.3	1.3	2.8	2.8	2.8	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	**44**
T6.1 - Algorithms and Prototype Evaluation												0.5	0.5	0.5	0.5	0.5	0.5	0.5				0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	11
T6.2 - Shared Task for User Credibility Estimation						0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	24
T6.3 - Resource Sharing							1.5	1.5	1.5																1.5	1.5	1.5										9
WP7 - Project Management	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	**5**
Total	3.0	5.7	5.7	5.7	7.1	9.3	9.8	9.8	9.8	8.3	10.5	11.0	11.6	11.6	11.6	10.2	10.2	10.2	7.0	7.0	7.0	6.1	6.1	8.8	8.1	8.1	10.8	9.3	9.3	9.3	8.0	8.0	8.0	6.6	6.6	6.6	**302**
																			D2.2						D3.3												D6.1
							D1.1			D6.3			D1.2			D3.1			D2.1			D2.3			D3.2			D4.2			D5.2			D5.1			D5.3
																									D4.1												D6.2

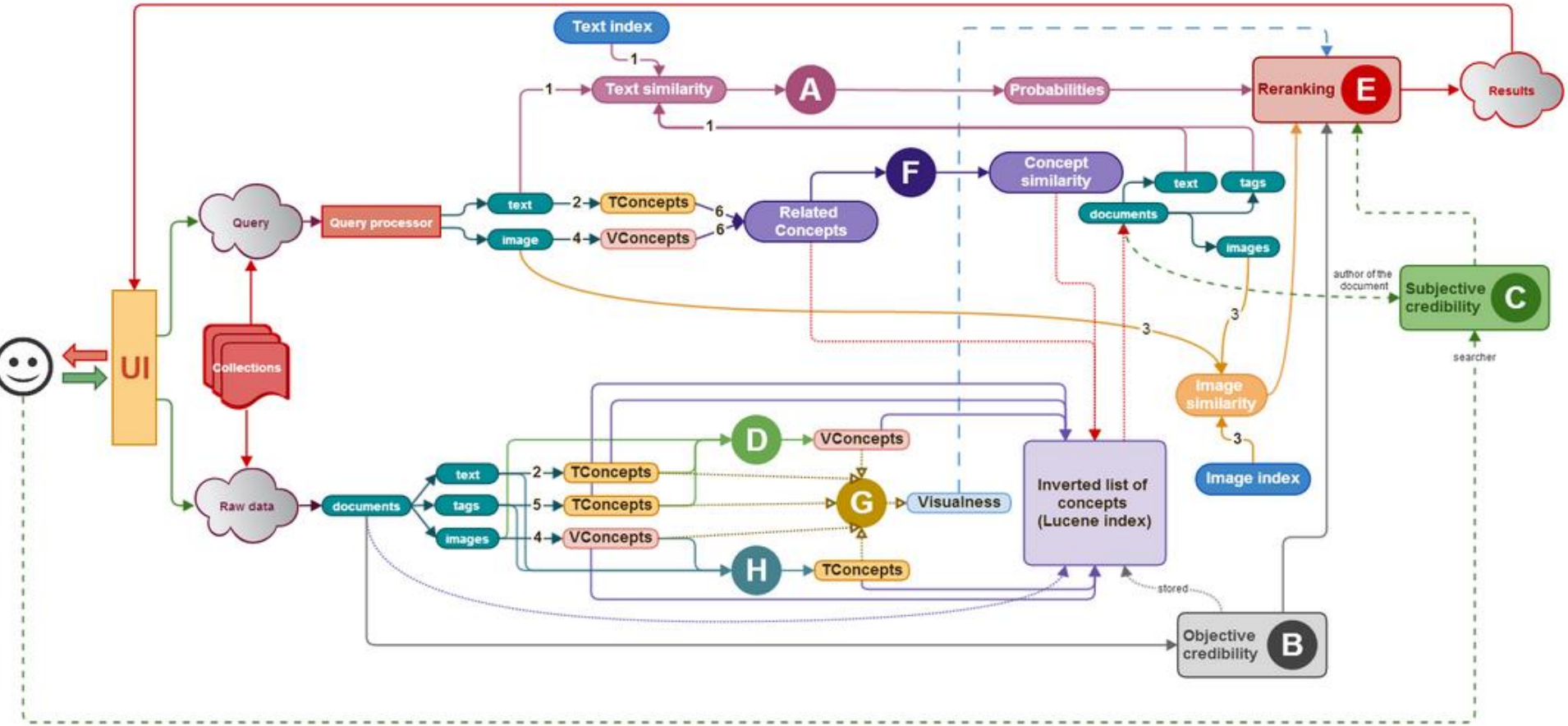
Scientific background
Objectives

- Can we extract, from text processing alone, an understanding of how likely it is that the top N returned results are useful for the user? Is this **likelihood of relevance** improved by NLP methods?
- Can we extract, from image processing alone, an understanding of how likely it is that the top N returned results are useful for the user? Is this **likelihood of relevance** improved by semantic annotations? Is this limited by domain?
- Are the likelihoods above **comparable** and can they be integrated in a coherent framework?
- How to model the semantic **entities** extracted from text and image data in order to compare them? Do we have to use a pre-existing semantic resources or is text enough to extract semantic entities and link them to images?
- Can the above likelihoods be improved by considering **data apparently outside the immediate relevance context**? In particular, can user performance in other contexts be used as a factor in the fusion of modalities?
- What is **user credibility** and how is it perceived and used by the users? How can this perception be modelled formally in order to obtain automatic credibility estimations?
- Can we develop a better system for multimedia access taking advantage of the **social** network relations (not limited to actual 'friends-of-friends' connections, but rather in a more general Web 3.0 sense) at a deeper level than simply filtering results based on graph links

Scientific Background



MUCKE Framework



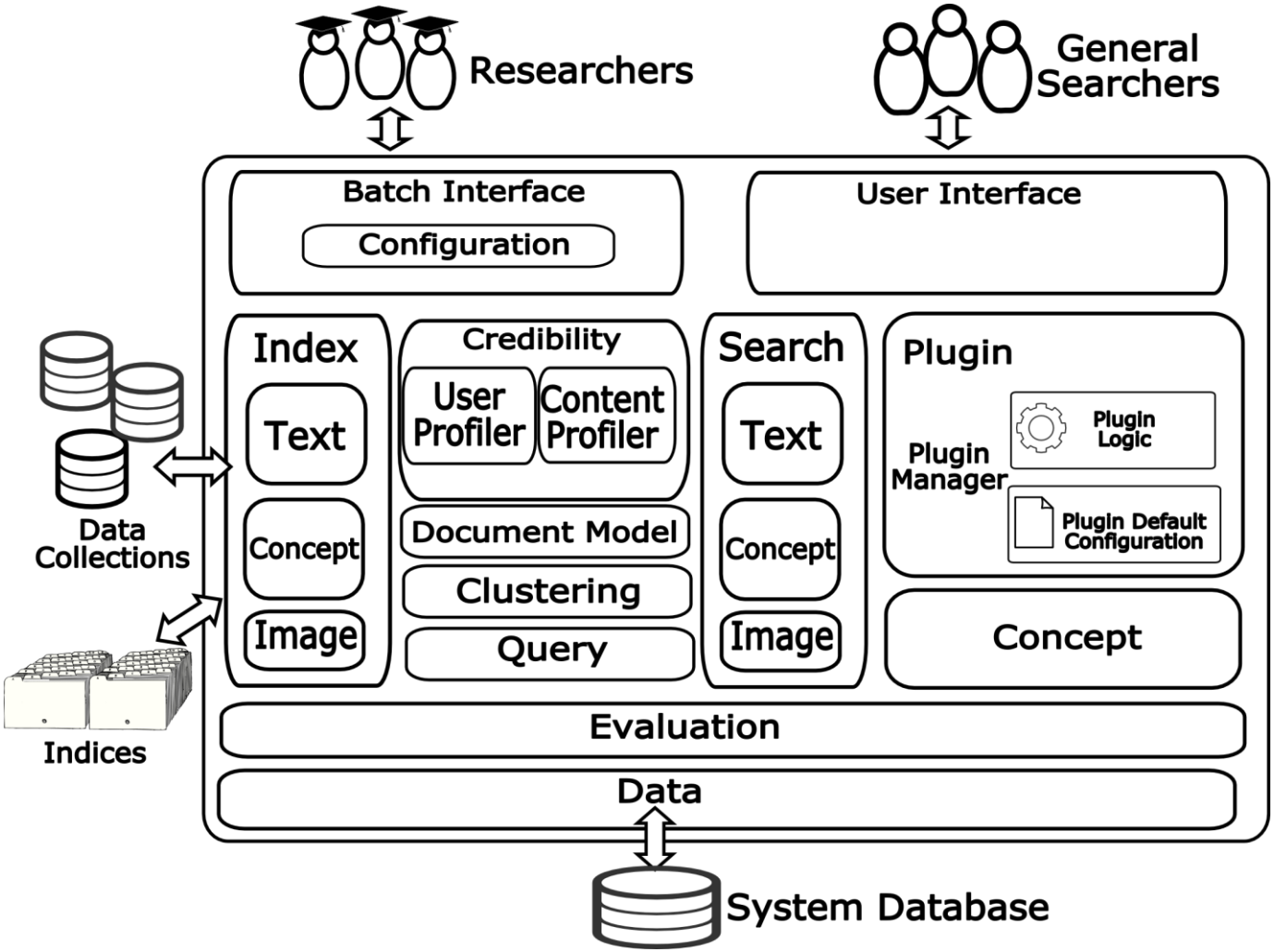
Tools at hand

1. Text -> Vectors -> Similarity
2. Text -> NLP/NER -> T-Concepts
3. Image -> Low-level features -> Similarity
4. Image -> Object Descriptions -> T-Concepts
5. Tags -> Disambiguate -> T-Concepts
6. Concepts -> Concepts[]

Wish list of components

- A Text similarity -> Probabilities
- B (producer) Data -> Credibility (Double)
- C (producer) Data, Consumer -> Credibility
- D T-Concepts, Image -> V-Concepts
- E Query -> Results -> Clusters
- F Concept, Concept -> Similarity
- G Concept -> Visualness
- H V-Concepts, Text -> T-Concepts

MUCKE Framework



Open framework

PUBLIC mucke / mucke

Unwatch 6 | Star 1 | Fork 0

branch: master | mucke / MUCKE / | History

1) Some bugfixes while fully enabling the properties file for all four ...

biergi authored 6 days ago — latest commit 0fff62b8eb

..		
.settings	Query Processor initial check in -> Strategy pattern	6 months ago
WebContent	Query Processor initial check in -> Strategy pattern	6 months ago
bin	1) Some bugfixes while fully enabling the properties file for all four	6 days ago
doc	This commit contains the following improvements:	18 days ago
lib	This commit contains the following improvements:	18 days ago
src	1) Some bugfixes while fully enabling the properties file for all four	6 days ago
.classpath	1) Some bugfixes while fully enabling the properties file for all four	6 days ago
.project	Query Processor initial check in -> Strategy pattern	6 months ago
javadoc.xml	This commit contains the following improvements:	18 days ago

© 2014 GitHub, Inc. Terms Privacy Security Contact

Status API Training Shop Blog About

Workplan

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16	M17	M18
WP 1 - Data Collection	2.8	2.8	2.8	2.8	2.8	2.8	1.8	1.8	1.8	1.8	1.8	1.8						
T1.1 Assessment and Collection of Existing Resources	1.0	1.0	1.0	1.0	1.0	1.0												
T1.2 New Data Collection	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8						
WP 2 - Multimedia Processing and Fusion		2.8	2.8	2.8	4.1	4.1	4.1	4.1	4.1	4.1	4.1	4.1	4.1	4.1	4.1	4.1	4.1	4.1
T2.1 Text processing		1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2
T2.2 Image processing		1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6
T2.3 Multimedia fusion					1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4
WP 3 - User Credibility for Multimedia Streams						1.4	1.4	1.4	1.4	1.4	3.6	3.6	3.6	3.6	3.6	2.2	2.2	2.2
T3.1 Credibility Model Definition						1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4			
T3.2 User Context Mining											1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
T3.3 User content mining and fusion											1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2
WP 4 - Multimedia Concept Similarity Models													2.5	2.5	2.5	2.5	2.5	2.5
T4.1 Textual Concept Similarity													1.2	1.2	1.2	1.2	1.2	1.2
T4.2 Multimedia Concept Similarity													1.3	1.3	1.3	1.3	1.3	1.3
WP5 - User-Centred Image Retrieval																		
T5.1 - Image Search Results Diversification																		
T5.2 - User-Centred Image Retrieval Model																		
T5.3 - Retrieval Prototype Integration																		
WP6 - Evaluation and Resource Sharing						0.8	2.3	2.3	2.3	0.8	0.8	1.3	1.3	1.3	1.3	1.3	1.3	1.3
T6.1 - Algorithms and Prototype Evaluation												0.5	0.5	0.5	0.5	0.5	0.5	0.5
T6.2 - Shared Task for User Credibility Estimation						0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
T6.3 - Resource Sharing							1.5	1.5	1.5									
WP7 - Project Management	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Total	3.0	5.7	5.7	5.7	7.1	9.3	9.8	9.8	9.8	8.3	10.5	11.0	11.6	11.6	11.6	10.2	10.2	10.2
							D1.1			D6.3			D1.2			D3.1		

Workplan

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16	M17	M18
WP 1 - Data Collection	2.8	2.8	2.8	2.8	2.8	2.8	1.8	1.8	1.8	1.8	1.8	1.8						
T1.1 Assessment and Collection of Existing Resources	1.0	1.0	1.0	1.0	1.0	1.0												
T1.2 New Data Collection	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8						
WP 2 - Multimedia Processing and Fusion		2.8	2.8	2.8	4.1	4.1	4.1	4.1	4.1	4.1	4.1	4.1	4.1	4.1	4.1	4.1	4.1	4.1
T2.1 Text processing		1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2
T2.2 Image processing		1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6
T2.3 Multimedia fusion					1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4
WP 3 - User Credibility for Multimedia Streams						1.4	1.4	1.4	1.4	1.4	3.6	3.6	3.6	3.6	3.6	2.2	2.2	2.2
T3.1 Credibility Model Definition						1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4			
T3.2 User Context Mining											1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
T3.3 User content mining and fusion											1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2
WP 4 - Multimedia Concept Similarity Models													2.5	2.5	2.5	2.5	2.5	2.5
T4.1 Textual Concept Similarity													1.2	1.2	1.2	1.2	1.2	1.2
T4.2 Multimedia Concept Similarity													1.3	1.3	1.3	1.3	1.3	1.3
WP5 - User-Centred Image Retrieval																		
T5.1 - Image Search Results Diversification																		
T5.2 - User-Centred Image Retrieval Model																		
T5.3 - Retrieval Prototype Integration																		
WP6 - Evaluation and Resource Sharing						0.8	2.3	2.3	2.3	0.8	0.8	1.3	1.3	1.3	1.3	1.3	1.3	1.3
T6.1 - Algorithms and Prototype Evaluation												0.5	0.5	0.5	0.5	0.5	0.5	0.5
T6.2 - Shared Task for User Credibility Estimation						0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
T6.3 - Resource Sharing							1.5	1.5	1.5									
WP7 - Project Management	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Total	3.0	5.7	5.7	5.7	7.1	9.3	9.8	9.8	9.8	8.3	10.5	11.0	11.6	11.6	11.6	10.2	10.2	10.2
							D1.1			D6.3			D1.2			D3.1		

Completed tasks

- Assessment and Collection of Existing Resources
  - Deliverable 1.1. *Report on Data Collections*
    - existing data collections, characteristics, APIs
- New Data Collection
  - Deliverable 1.2 *New Data Collected and Associated Report*
    - CEA provided hooks to the Flickr API, TUW the download tasks distribution mechanism, all downloaded data
    - UAIC received all data during S2 and then sent it to CEA
    - 78million images + metadata collected (9TB), 60k wikipedia concepts
- Resource Sharing
  - Deliverable 6.3 *Report on Resource Sharing Framework*
    - *UAIC coordinated the collection of available resources from each partner*
- Credibility Model Def
  - Deliverable 3.1 *Cre*

	expertise	trustworthiness	quality	reliability
User (creator)	●●●	●●●	●○○	●○○
Data object	●○○	●○○	●●●	○○○
IR engine	●●○	●●○	●○○	●●●
HCI	○○○	●●○	●●●	●●●

Workplan

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16	M17	M18
WP 1 - Data Collection	2.8	2.8	2.8	2.8	2.8	2.8	1.8	1.8	1.8	1.8	1.8	1.8						
T1.1 Assessment and Collection of Existing Resources	1.0	1.0	1.0	1.0	1.0	1.0												
T1.2 New Data Collection	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8						
WP 2 - Multimedia Processing and Fusion		2.8	2.8	2.8	4.1	4.1	4.1	4.1	4.1	4.1	4.1	4.1	4.1	4.1	4.1	4.1	4.1	4.1
T2.1 Text processing		1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2
T2.2 Image processing		1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6
T2.3 Multimedia fusion					1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4
WP 3 - User Credibility for Multimedia Streams						1.4	1.4	1.4	1.4	1.4	3.6	3.6	3.6	3.6	3.6	2.2	2.2	2.2
T3.1 Credibility Model Definition						1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4			
T3.2 User Context Mining											1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
T3.3 User content mining and fusion											1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2
WP 4 - Multimedia Concept Similarity Models													2.5	2.5	2.5	2.5	2.5	2.5
T4.1 Textual Concept Similarity													1.2	1.2	1.2	1.2	1.2	1.2
T4.2 Multimedia Concept Similarity													1.3	1.3	1.3	1.3	1.3	1.3
WP5 - User-Centred Image Retrieval																		
T5.1 - Image Search Results Diversification																		
T5.2 - User-Centred Image Retrieval Model																		
T5.3 - Retrieval Prototype Integration																		
WP6 - Evaluation and Resource Sharing						0.8	2.3	2.3	2.3	0.8	0.8	1.3	1.3	1.3	1.3	1.3	1.3	1.3
T6.1 - Algorithms and Prototype Evaluation												0.5	0.5	0.5	0.5	0.5	0.5	0.5
T6.2 - Shared Task for User Credibility Estimation						0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
T6.3 - Resource Sharing							1.5	1.5	1.5									
WP7 - Project Management	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Total	3.0	5.7	5.7	5.7	7.1	9.3	9.8	9.8	9.8	8.3	10.5	11.0	11.6	11.6	11.6	10.2	10.2	10.2
							D1.1			D6.3			D1.2			D3.1		

Current Tasks

- Credibility Estimation
  - Evaluation campaign
- Text / Image processing
- Multimedia Processing and Fusion

Credibility Estimation for Multimedia

- Credibility model defined
  - Combination of contextual factors and content analysis
  - Cast as a machine learning problem
- Context:
  - user's social graph analysis,
  - statistics of contributions to the social network (number of photos, vocabulary etc.)
  - opinion mining
- Content:
  - Coherence of textual annotations
  - Image content classification using ImageNet concepts: i.e. given an image-tag association, how illustrative of the tag is the image?
- Encouraging preliminary results
  - a theoretical 50% improvement in image retrieval using user credibility

Evaluation Task

- MediaEval 2014 - *Retrieving Diverse Social Images task*
  - *1 May: Development data release / 2 June: Test data release / 9 September: Run submission*
  - in addition to relevance, we provide user credibility estimations
  - additional dataset used to train the credibility descriptors (credibility set, 300 locations, 1,000 users, with at least 50 images per user)
- MUCKE datasets
  - credibility role in image retrieval
    - topic dependent: 160 topics (90 training, 70 test)
    - per train topic: concept, *image*, relevance,
    - per test topic: concept, *image*, ??
    - where *image* has user credibility estimation/features
  - direct assessment of credibility
    - topic independent, set of 1000 users, 50 images / user
    - data: user context & content features

Text processing

- Focused on Explicit Semantic Analysis
  - Mapping of words/tags into a conceptual space defined by Wikipedia/other resources
- Classical version implemented at M8
  - 10 languages including English, French, German, Romanian
  - Tested during the CLEF CHIC text retrieval campaign
  - 2nd/7 participants
- Ongoing work on an improved version
  - Including multiword detection and concept disambiguation
- Combination of Language Models and User Models
  - the Geographic domain
  - MediaEval Placing Task 2013
  - 1st/7 participants
- The obtained resources will be publicly released

Image processing

- Benchmarking of different SoTA features in Image Retrieval & Classification
- Joint participation of BILKENT and CEA at MediaEval Diverse Images 2013
  - 3rd/11 participants
- Extraction of compact semantic features based on ImageNet
  - Dimension reduction by 100 with classification accuracy loss of ~7%
- Use of features derived with deep learning architectures seems very promising
  - MAP 0.77 on PascalVOC 2007

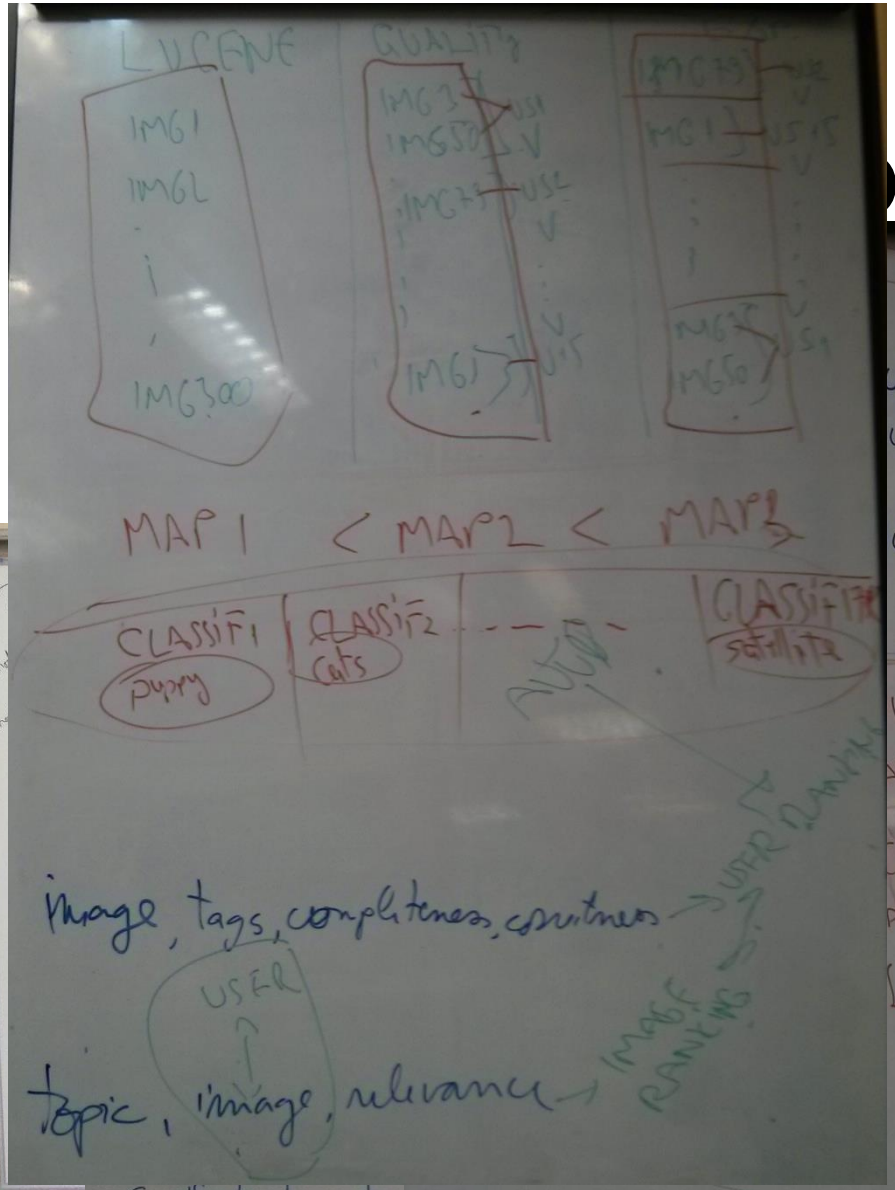
Multimedia Fusion

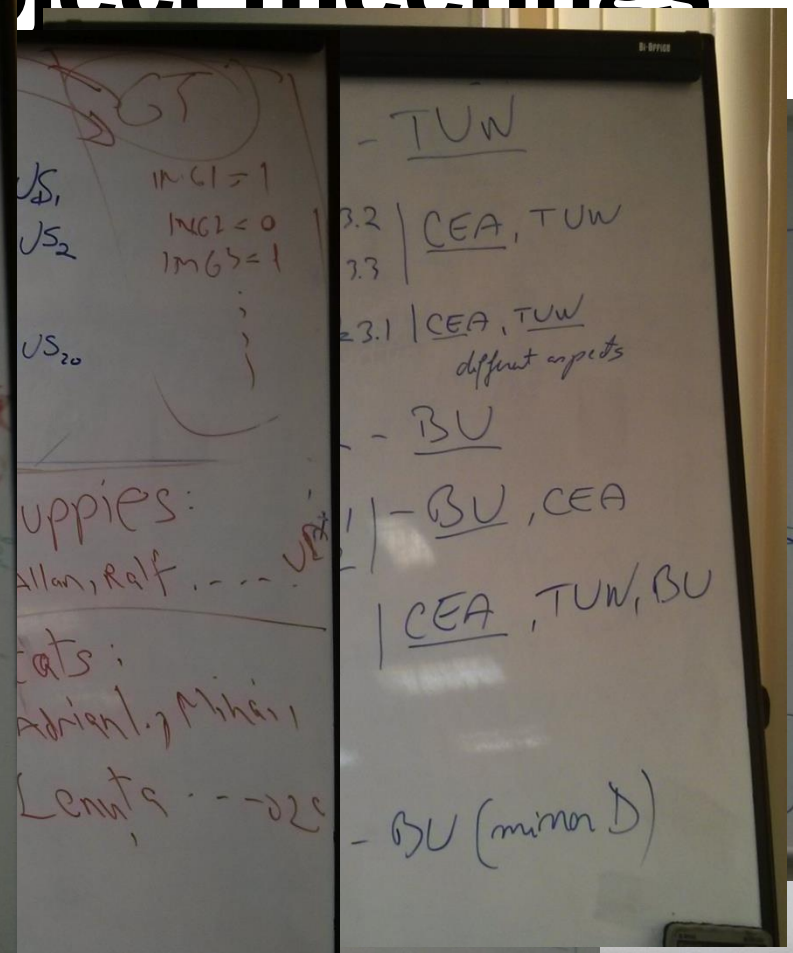
- Exploration of both early and late fusion techniques
  - Results indicate that the latter type is more promising
  - Applied late fusion for diversification at MediaEval Diverse Images 2013
- ongoing work focuses on the Concept Index and Credibility integration

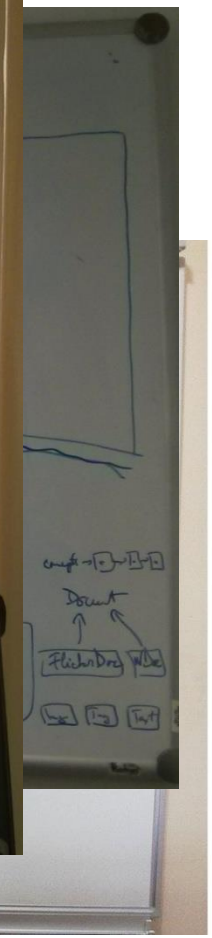
Problems / Issues

- Delays in national financing
  - Mitigated
- Staffing problems at CEA
  - Post-doc left before the term of the contract
  - Mitigated through the implication of a PhD student
- Differences between national and CHIST-ERA legal responsibilities
  - Consortium Agreement
    - Austria (FWF) grants to individual.
    - Others, EU, grant to institute

ject meetings







Student Exchanges

- June 2013
  - UAIC – TUW
    - framework definition
- February and March 2014
  - UAIC – CEA:
    - Alexandra Siriteanu (2 weeks) –MsC thesis on image retrieval result diversification
    - Cristina Serban 1 week – MsC thesis on trust in social networks

Communication

- Website http://ifs.tuwien.ac.at/~mucke
- Publications
  - 9 papers accepted so far
- Evaluation tasks @ MediaEval 2013, 2014
- Exchanges
  - TUWien – NII researcher exchange on credibility in information retrieval (June-July 2013)

Financial reporting

N°	Partner	Person.months	Total costs	Percentage of requested budget
1	TUW	16	61,588 €	15%
2	CEA LIST	21.63[1]	42,288 €	15.60%
3	Bilkent University	27[2]	44,690 €	42%
4	“Al. I. Cuza” University	27	102,678 €	37.92%

[1] Including 5.63 PMs of post-doc financed by ANR and 16 months which are not financed by ANR: 12 PMs permanent staff and 4 PMs doctoral student.
[2] Estimated at the time of writing of this report.

Summary

- For the task of multimedia retrieval, MUCKE introduces new concepts and model - merge topical relevance and domain specific user credibility
- Using the yet untapped data in multimedia retrieval, social networks, and by creating semantic descriptors of groups and using them to calibrate probabilities of semantic tags applied to individual data → more relevant results
- Our transition from scores to probabilities allows the systems to be aware of low levels of confidence in their results
- A mixture fusion approach (applied late, but based on early processing), based on moving from ranking scores to probability values which can be applied to merge any type of data

Thank you

MUCKE

Multimedia and User Credibility Knowledge Extraction

http://ifs.tuwien.ac.at/~mucke/