WeKnowIt

Making the Collective Intelligence of Social Media Searchable

Yiannis Kompatsiaris

CERTH – ITI

CIVR 2010, July 2010

http://www.weknowit.eu

SEVENTH FRAMEWORK PROGRAMME

co-funded by the European Union

Contents

- Introduction
- Social Media - Collective Intelligence
- WeKnowIt approach
- Community detection in Social Media
- Demos
- Conclusions - Issues

Web 2.0 content

flickr

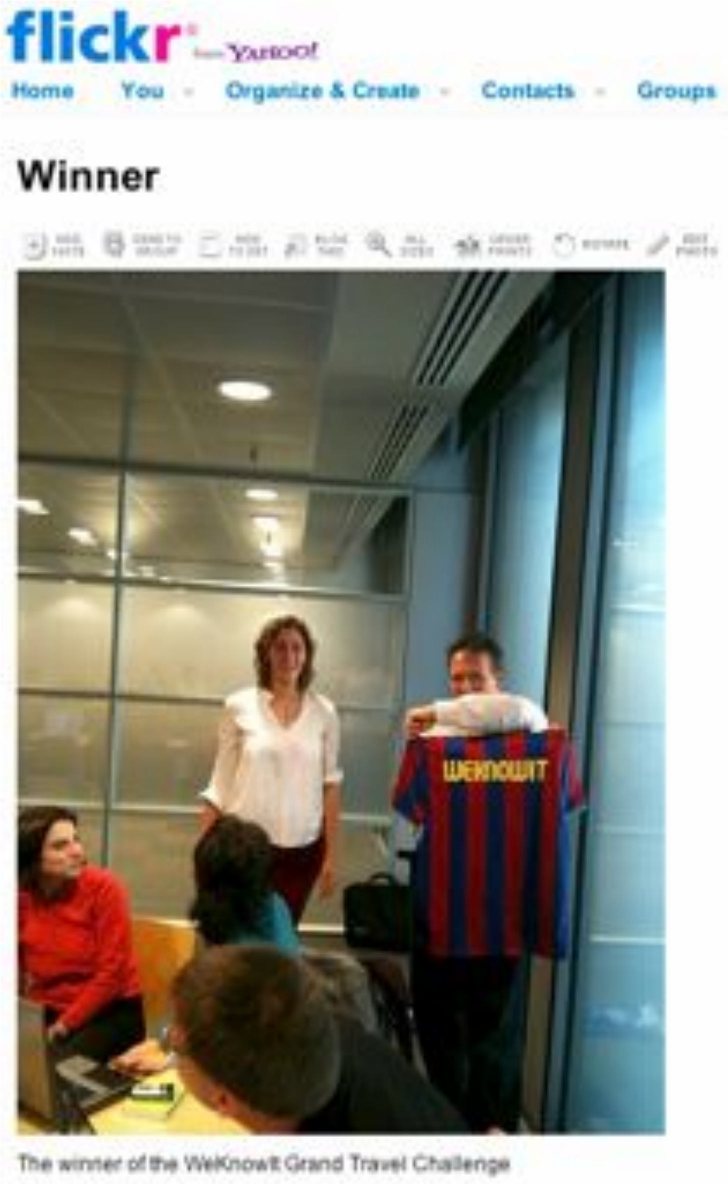
- 3,190 uploads in the last minute
- 3.2 million things geotagged this month
- 4,754,012,299 photos (2 July 2010)

YouTube

- 24h of video content uploaded every minute
- 2 billion movies watched every day

facebook

- More than 400 million active users
- More than 200 million users log on at least once each day
- 2.5 billion photos uploaded each month

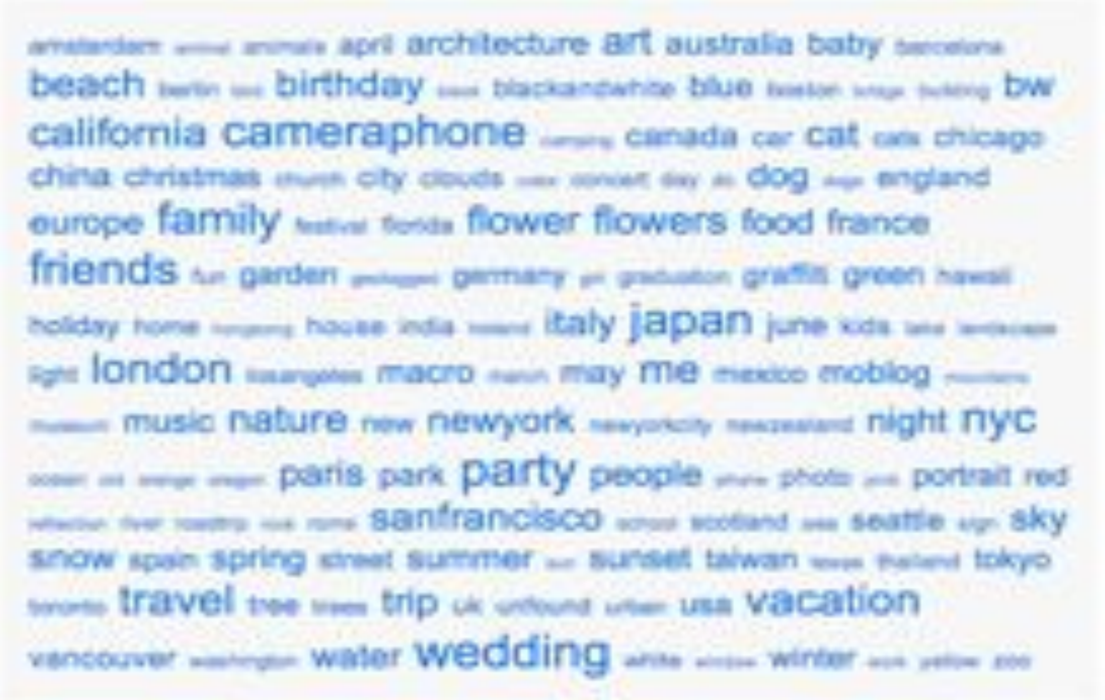


Winner

The winner of the WeKnowIt Grand Travel Challenge

Tags, content everywhere

Upload, tag, share, search



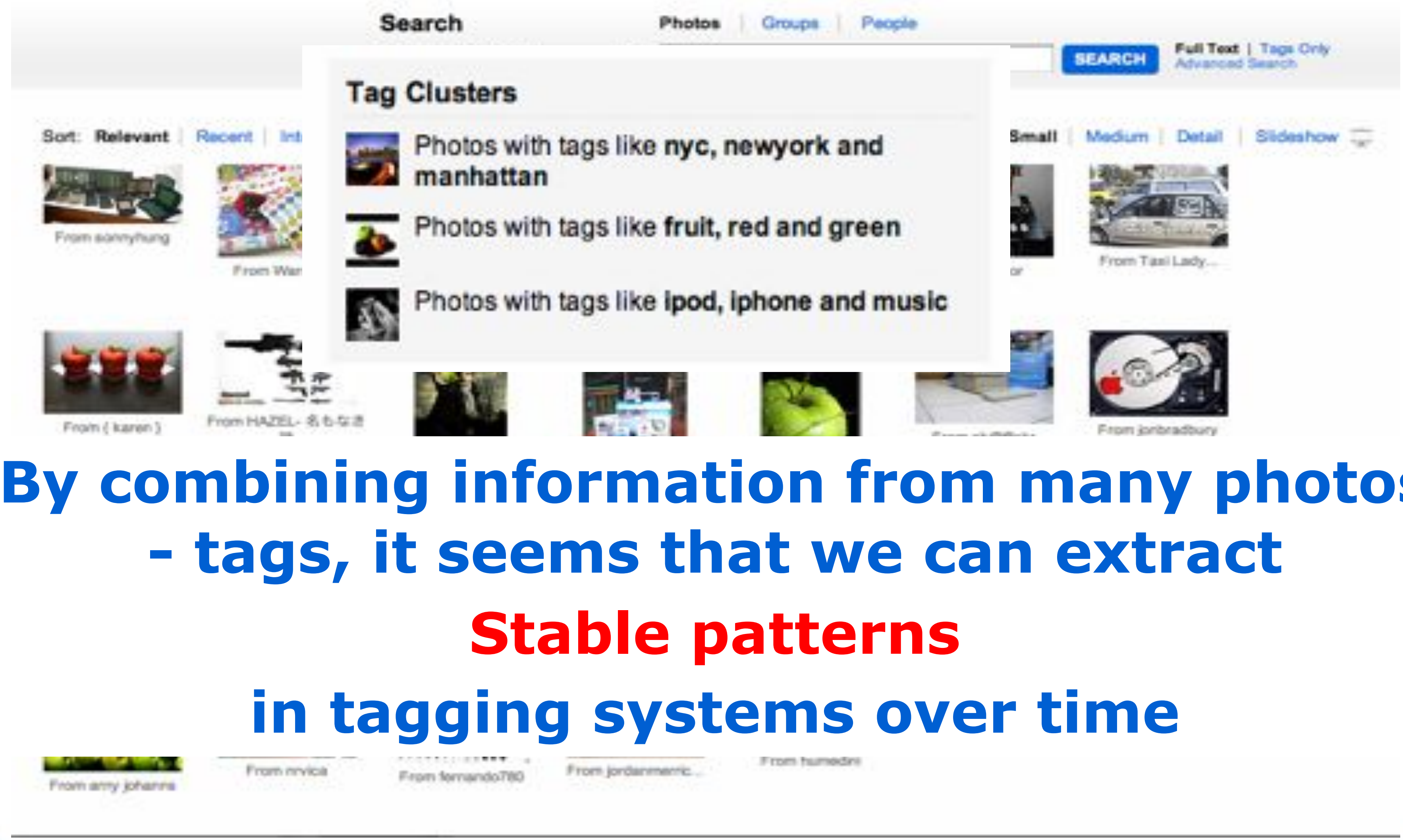
amsterdam animal animals april architecture art australia baby barcelona
beach berlin bird birthday black blackandwhite blue boston bridge building bw
california cameraphone camping canada car cat cats chicago
china christmas church city clouds color concert day de dog dogs england
europe family festival florida flower flowers food france
friends fun garden geotagged germany girl graduation graffiti green hawaii
holiday home hongkong house india ireland italy japan june kids lake landscape
light london losangeles macro march may me mexico moblog mountains
museum music nature new newyork newyorkcity newzealand night nyc
ocean old orange oregon paris park party people phone photo pink portrait red
reflection river roadtrip rock rome sanfrancisco school scotland sea seattle sign sky
snow spain spring street summer sun sunset taiwan texas thailand tokyo
toronto travel tree trees trip uk unfound urban usa vacation
vancouver washington water wedding white window winter work yellow zoo

weknowit

SEVENTH FRAMEWORK PROGRAMME

co-funded by the European Union

Can we do more things?



Tag Clusters

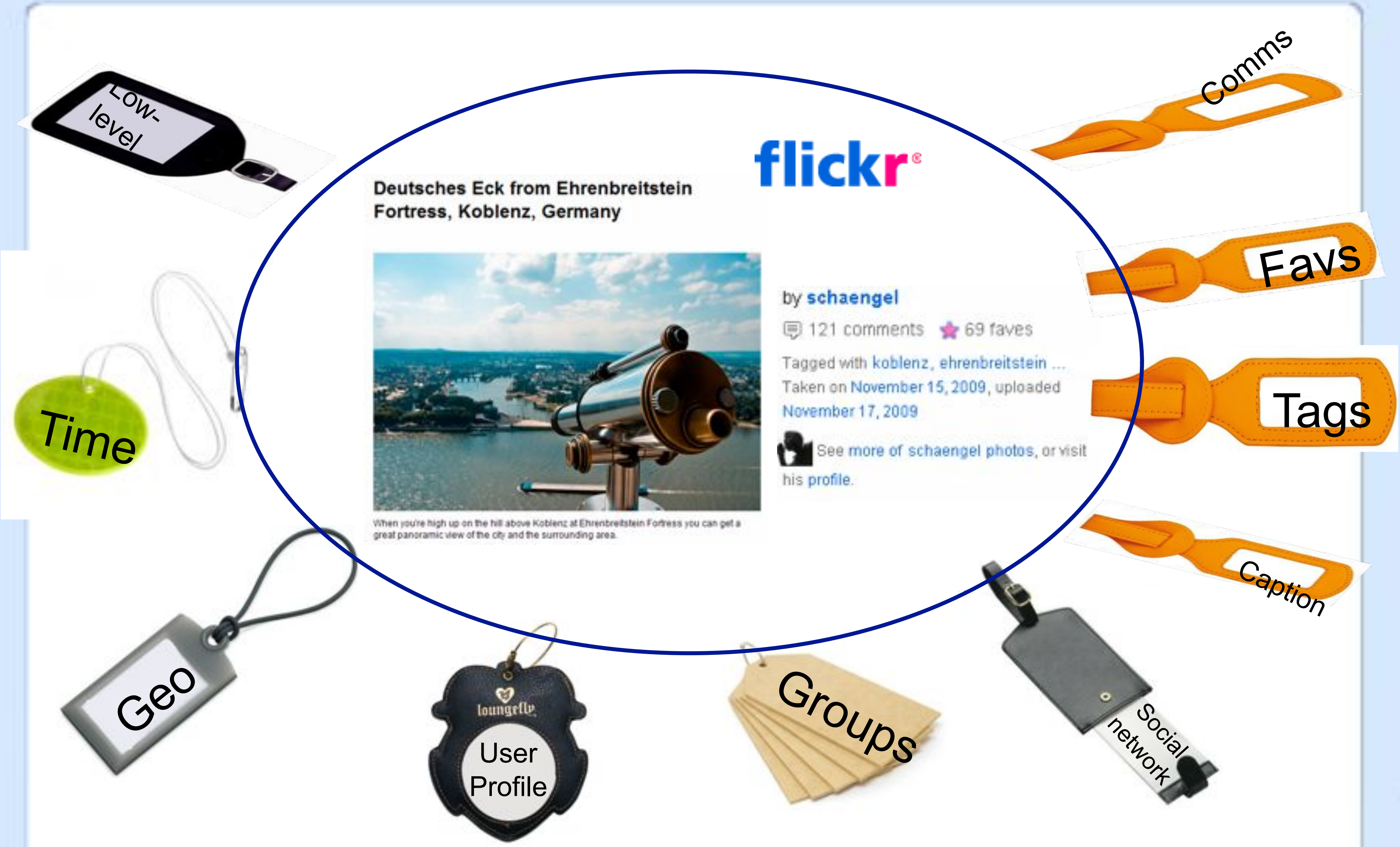
- Photos with tags like **nyc, newyork and manhattan**
- Photos with tags like **fruit, red and green**
- Photos with tags like **ipod, iphone and music**

By combining information from many photos - tags, it seems that we can extract

Stable patterns

in tagging systems over time

SEVENTH FRAMEWORK PROGRAMME
co-funded by the European Union



flickr

Deutsches Eck from Ehrenbreitstein Fortress, Koblenz, Germany

When you're high up on the hill above Koblenz at Ehrenbreitstein Fortress you can get a great panoramic view of the city and the surrounding area.

by schaengel

121 comments ★ 69 faves

Tagged with koblenz, ehrenbreitstein ...

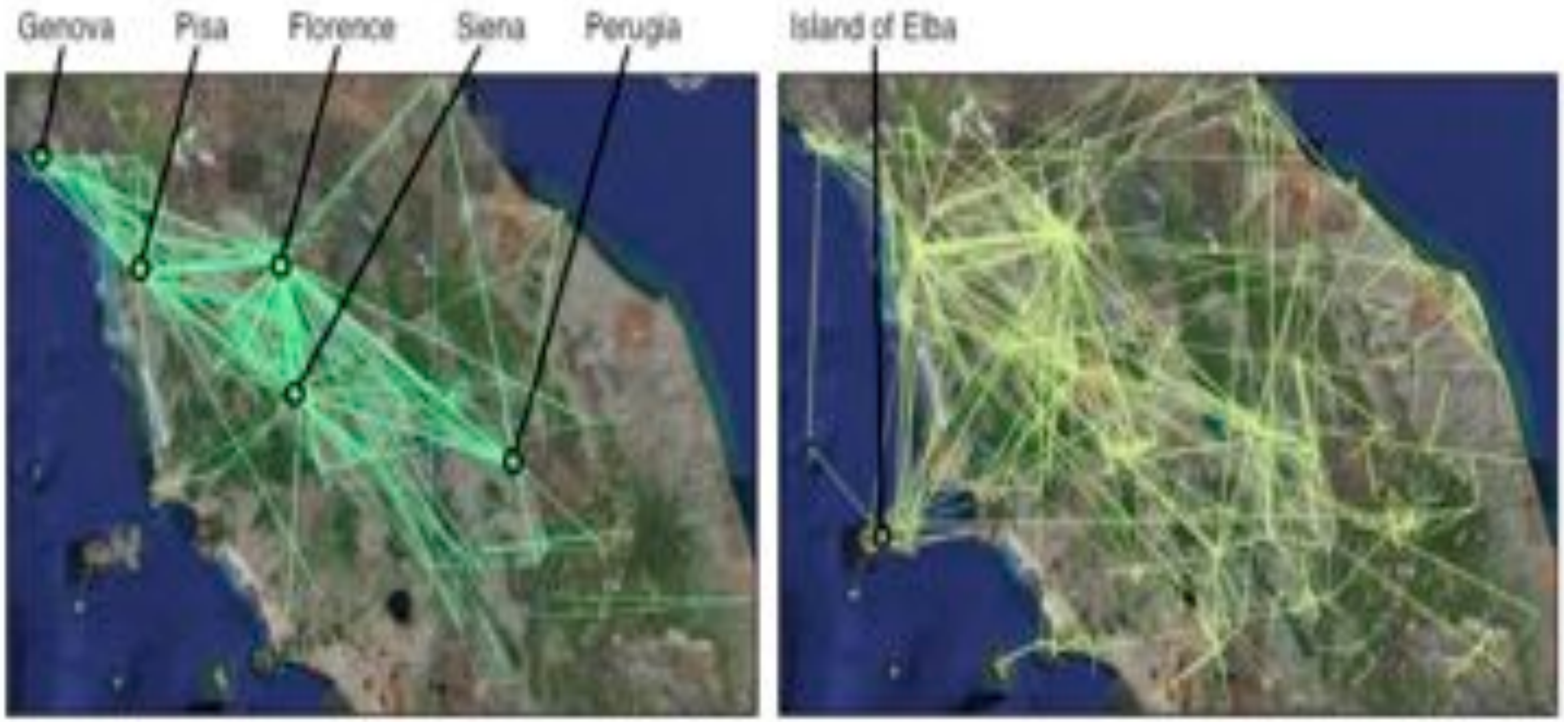
Taken on November 15, 2009, uploaded November 17, 2009

See more of schaengel photos, or visit his profile.

weknowit

SEVENTH FRAMEWORK PROGRAMME
co-funded by the European Union

… and more: Travel trends using flickr



Trace Flickr users from a chronologically ordered set of geographically referenced photos

Who are the Italians and who are the Americans?

MIT SENSEABLE CITY LAB, "The World's eyes"

SEVENTH FRAMEWORK PROGRAMME
co-funded by the European Union

Collective Intelligence!

Mass user-generated content Web 2.0

Little understanding

Organisations – Processes

No benefits from community and mass content

Analysis techniques: Content, Social, Mass

Loose interaction

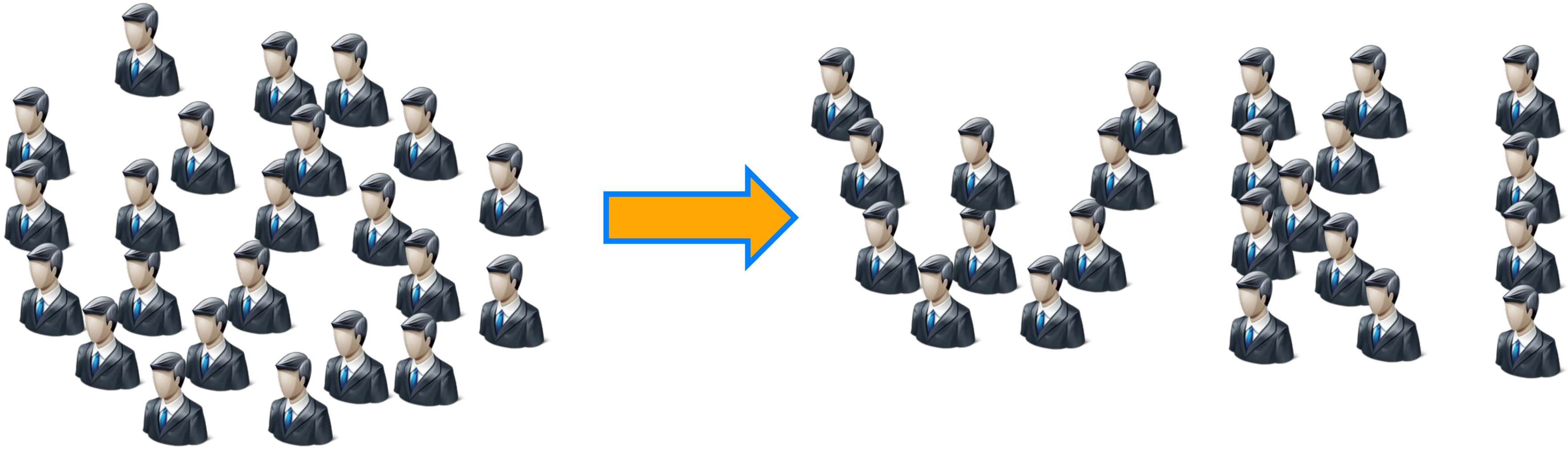
Users & Devices

Need for new services

SEVENTH FRAMEWORK PROGRAMME
co-funded by the European Union

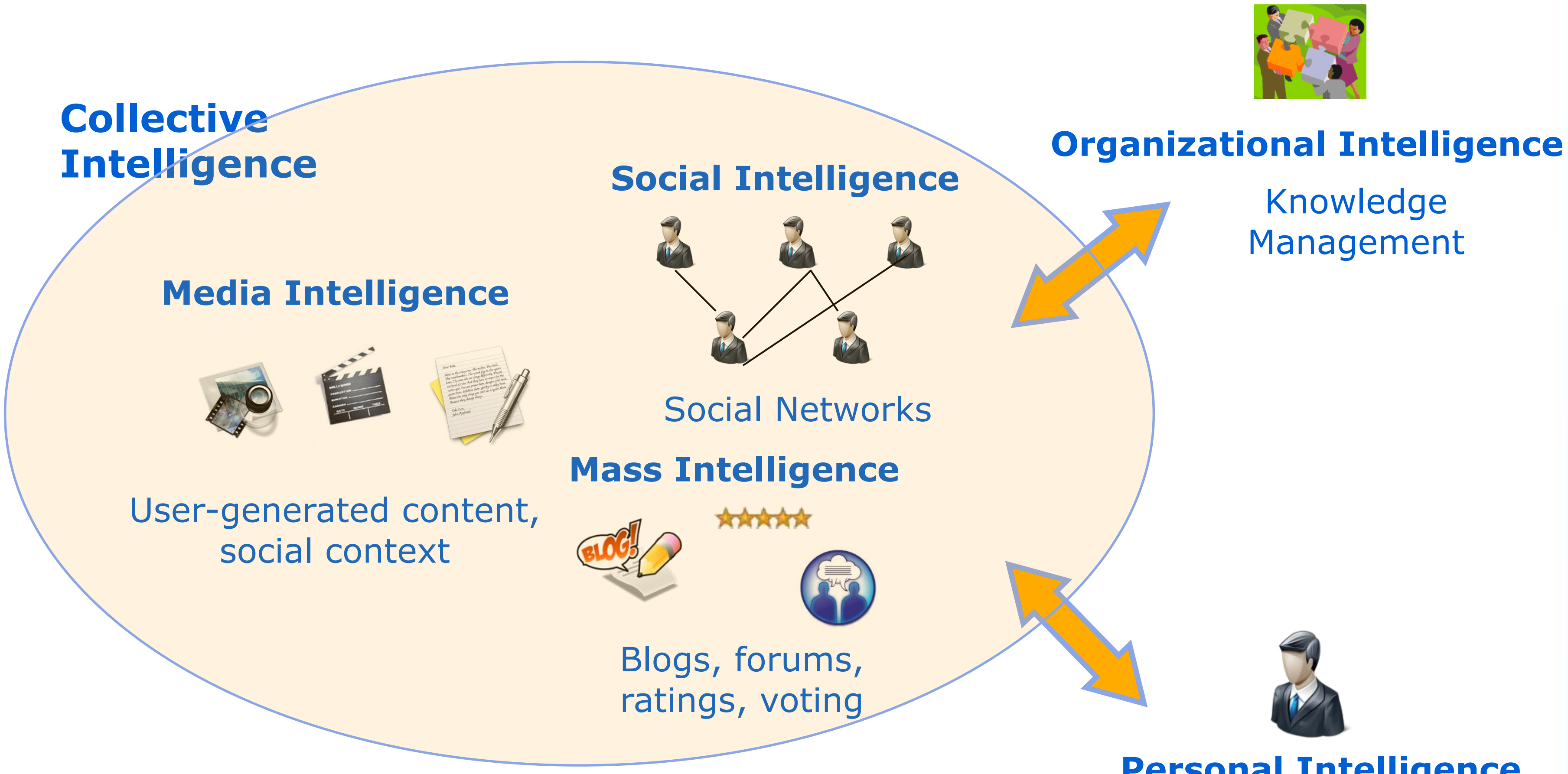
Defining Collective Intelligence

Collective Intelligence is the Intelligence which emerges from the collaboration, competition and coordination among individuals.



...an Intelligence greater than the sum of the individuals' intelligence

WeKnowIt and CI



Collective Intelligence

Media Intelligence

User-generated content, social context

Social Intelligence

Social Networks

Mass Intelligence

Blogs, forums, ratings, voting

Organizational Intelligence

Knowledge Management

Personal Intelligence

Upload and Access

Emergency Response

Personal Intelligence

>> Login, Upload

>> Tag recommendation,

>> Spam detection

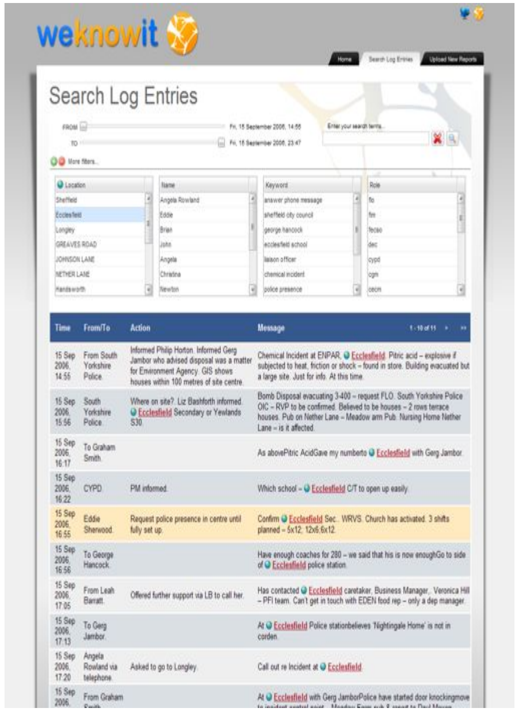
Media Intelligence

Picture arrives at emergency response

>> Automatic localisation of photo

Organisational Intelligence

>> Log Merger



Search Log Entries

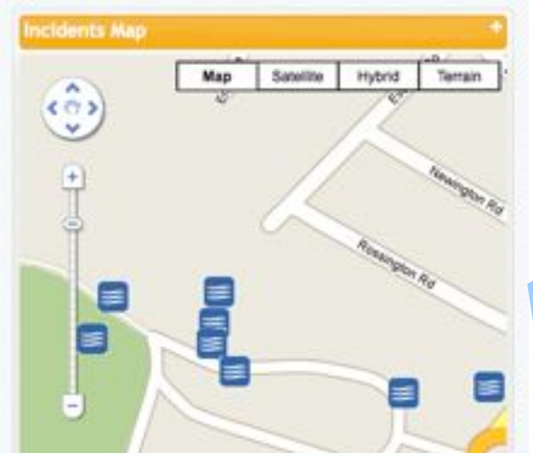
FROM Fri, 15 September 2006, 14:55

TO Fri, 15 September 2006, 23:47

Location	Name	Keyword	Role
Sheffield	Angela Rowland	answer phone message	flo
Ecclesfield	Eddie	sheffield city council	fm
Longley	Brian	george hancock	fecso
GREAVES ROAD	John	ecclesfield school	dec
JOHNSON LANE	Angela	liaison officer	cypd
NETHER LANE	Christina	chemical incident	cgn
Handsworth	Newton	police presence	cecm

Time	From/To	Action	Message 1 - 10 of 11 › ››
15 Sep 2006, 14:55	From South Yorkshire Police	Informed Philip Horton. Informed Gerg Jambor who advised disposal was a matter for Environment Agency. GIS shows houses within 100 metres of site centre.	Chemical Incident at ENPAR, Ecclesfield. Pitric acid – explosive if subjected to heat, friction or shock – found in store. Building evacuated but a large site. Just for info. At this time.
15 Sep 2006, 15:56	South Yorkshire Police	Where on site? Liz Bashforth informed. Ecclesfield Secondary or Yewlands S30.	Bomb Disposal evacuating 3-400 – request FLO. South Yorkshire Police OIC – RVP to be confirmed. Believed to be houses – 2 rows terrace houses. Pub on Nether Lane – Meadow arm Pub. Nursing Home Nether Lane – is it affected.
15 Sep 2006, 16:17	To Graham Smith.		As abovePitric AcidGave my numberto Ecclesfield with Gerg Jambor.
15 Sep 2006, 16:22	CYPD	PM informed.	Which school – Ecclesfield C/T to open up easily.
15 Sep 2006, 16:55	Eddie Sherwood.	Request police presence in centre until fully set up.	Confirm Ecclesfield Sec.. WRVS. Church has activated. 3 shifts planned – 5x12; 12x6;6x12.
15 Sep 2006, 16:56	To George Hancock.		Have enough coaches for 280 – we said that his is now enoughGo to side of Ecclesfield police station.
15 Sep 2006, 17:05	From Leah Barratt.	Offered further support via LB to call her.	Has contacted Ecclesfield caretaker, Business Manager,. Veronica Hill – PFI team. Can't get in touch with EDEN food rep – only a dep manager.
15 Sep 2006, 17:13	To Gerg Jambor.		At Ecclesfield Police stationbelieves 'Nightingale Home' is not in corden.
15 Sep 2006, 17:20	Angela Rowland via telephone.	Asked to go to Longley.	Call out re Incident at Ecclesfield.
15 Sep 2006,	From Graham Smith		At Ecclesfield with Gerg JamborPolice have started door knockingmove

Access



Mass Intelligence

Many contributors

>> Clustering

>> Tag Harmonization

>> Enrichment from add. sources

Social Intelligence

>> ER Alert Service

weknowit

SEVENTH FRAMEWORK PROGRAMME

co-funded by the European Union

Travel prototype

Mass Intelligence

>> Media Collection: flickr, query logs,

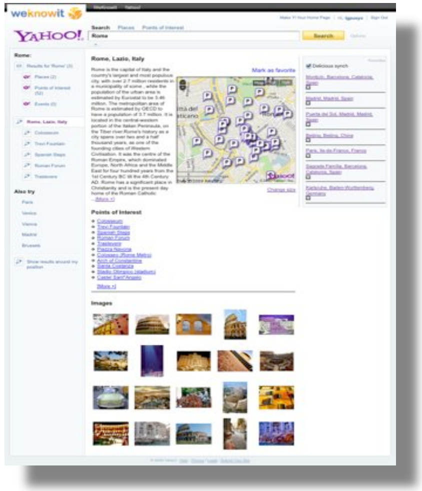
>> Automatic generation of ranked facet lists of POIs

Media Intelligence

>> Hybrid Clustering

>> Image Localisation

>> Tag suggestions



Travel Preparation

Personal Intelligence

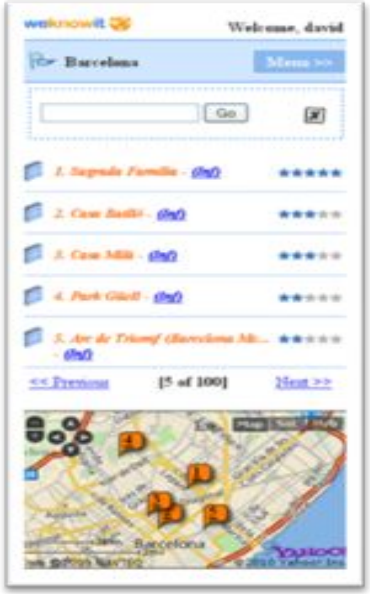
Profile of contributor

>> Recommendations

Social Intelligence

Can your social network be of help?

>> Friends position, alert


weknowit Welcome, david

Barcelona Menu >>

Go

1. Sagrada Familia - (Inf) ★★★★★

2. Casa Batlló - (Inf) ★★★☆☆

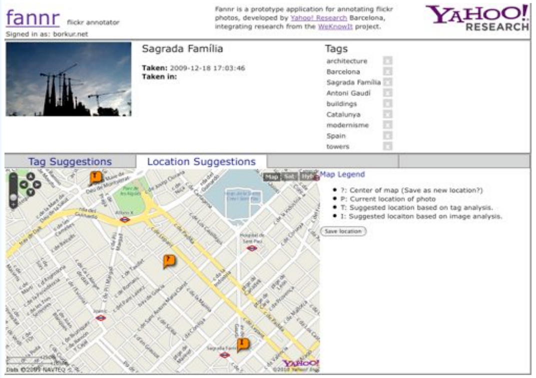
3. Casa Milà - (Inf) ★★★☆☆

4. Park Güell - (Inf) ★★☆☆☆

5. Arc de Triomf (Barcelona Mi... - (Inf) ★★☆☆☆

<< Previous [5 of 100] Next >>

Mobile Guidance


fannr flickr annotator

Signed in as: borkur.net

Fannr is a prototype application for annotating flickr photos, developed by Yahoo! Research Barcelona, integrating research from the WeKnowIt project.

YAHOO! RESEARCH

Sagrada Família

Taken: 2009-12-18 17:03:46
Taken in:

Tags

architecture
Barcelona
Sagrada Família
Antoni Gaudí
buildings
Catalunya
modernisme
Spain
towers

Tag Suggestions | Location Suggestions

Map Legend

- ?: Center of map (Save as new location?)
- P: Current location of photo
- T: Suggested location based on tag analysis.
- I: Suggested locaiton based on image analysis.

Save location

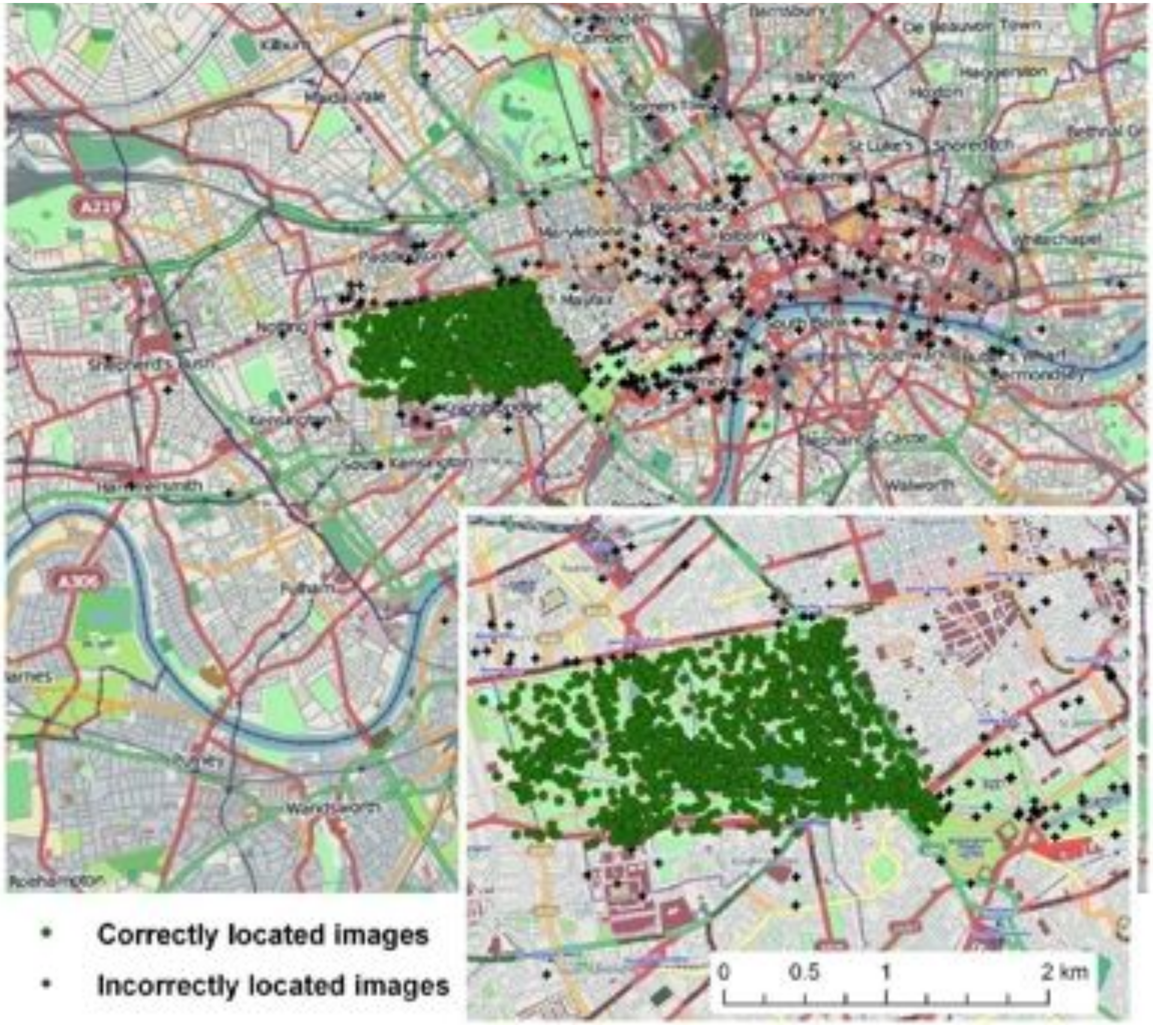
Post Travel

weknowit

SEVENTH FRAMEWORK PROGRAMME

co-funded by the European Union

Relevant Activities (*ER*)



Automatically describe city cores

Distinction between administrative and vernacular uses of place names

Potential for confusion in the dispatch of emergency services

Livia Hollenstein and Ross S. Purves, "Exploring place through user-generated content: using Flickr to describe city cores", JOURNAL OF SPATIAL INFORMATION SCIENCE

Relevant activities

MIT Center for Collective Intelligence

http://cci.mit.edu/index.html

The Climate Collaboratorium

Collective prediction ...*accurate predictions about future events such as product sales, political events, and outcomes of medical treatments*....

Collective intelligence in healthcare

Studying collective intelligence in today's organizations

Relevant activities

collective intellect

Live Traffic Information

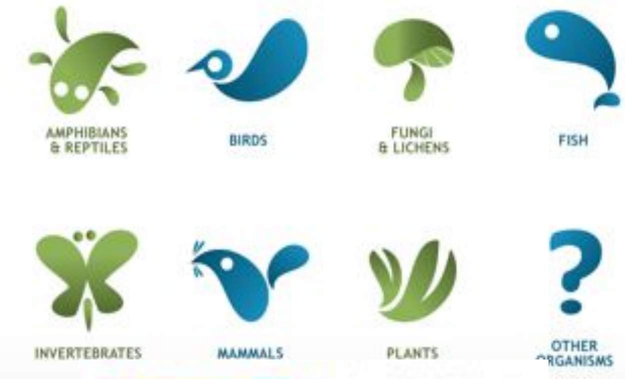
http://traffic.berkeley.edu

FixMyStreet

Mobnotes beta™
The World, Around You!

Boston Citizens Connect

iSpot
your place to share nature



DOPPLR

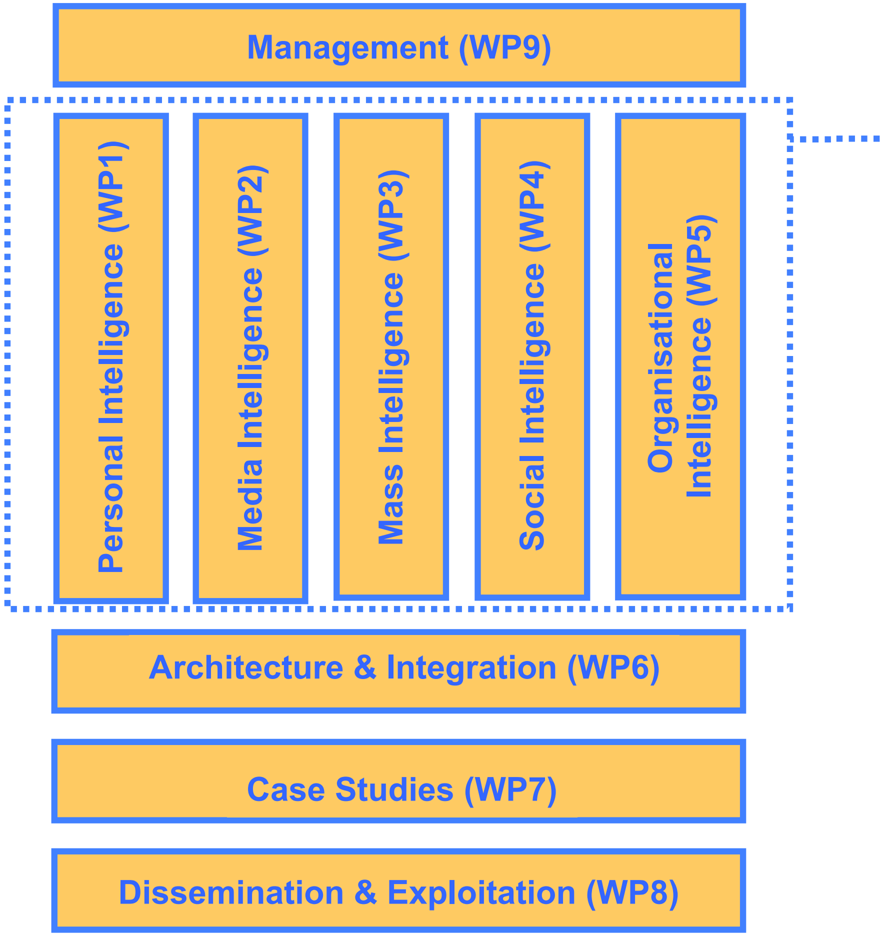
Dopplr helps you share your ... travel .... and exchange tips ... presents this **collective intelligence** - the travel patterns and advice ... as the Social Atlas.

Relevant activities

- Most applications are still harnessing Collective Intelligence
  - Upload applications
- Emphasis is on visualization of results
- Few focus on analysis
- No fusion of modalities – sources
- Unlimited applications
    - Draught detection, through estimation of green levels in flickr photos for fire protection in Spain (MIT Senseable lab)
    - Hollywood stock Exchange – HP Labs

SEVENTH FRAMEWORK PROGRAMME
co-funded by the European Union

Project work Overview



Management (WP9)

Personal Intelligence (WP1)

Media Intelligence (WP2)

Mass Intelligence (WP3)

Social Intelligence (WP4)

Organisational Intelligence (WP5)

Collective Intelligence

Architecture & Integration (WP6)

Case Studies (WP7)

Dissemination & Exploitation (WP8)

Research:	**WP1 – WP5**
Development:	**WP6**
Application:	**WP7**
Dissemination & Exploitation:	**WP8**
Management:	**WP9**

weknowit

SEVENTH FRAMEWORK PROGRAMME

co-funded by the European Union

Content in WeKnowIt

offline → model creation, training

Non-Web 2.0 training data

Standard annotated corpora used for training.

- **Single-modality:** text (Brown corpus), speech (TIMIT database), image (Corel database)
- **Single-source:** prepared by a single person/organization
- **Consistent quality:** absence of spam, malicious or erroneous data
- **Small-moderate volume:** Manually produced

Massive Web 2.0

Massive user generated content and feedback from Web 2.0 applications

- **Multi-modality:** e.g. image + tags, image + geo-location + time
- **Multi-source:** may be generated by different applications, user communities, e.g. delicious, StumbleUpon and reddit are all social bookmarking sites
- **Inconsistent quality:** noise, spam, ambiguity
- **Huge volume:** Massively produced and disseminated

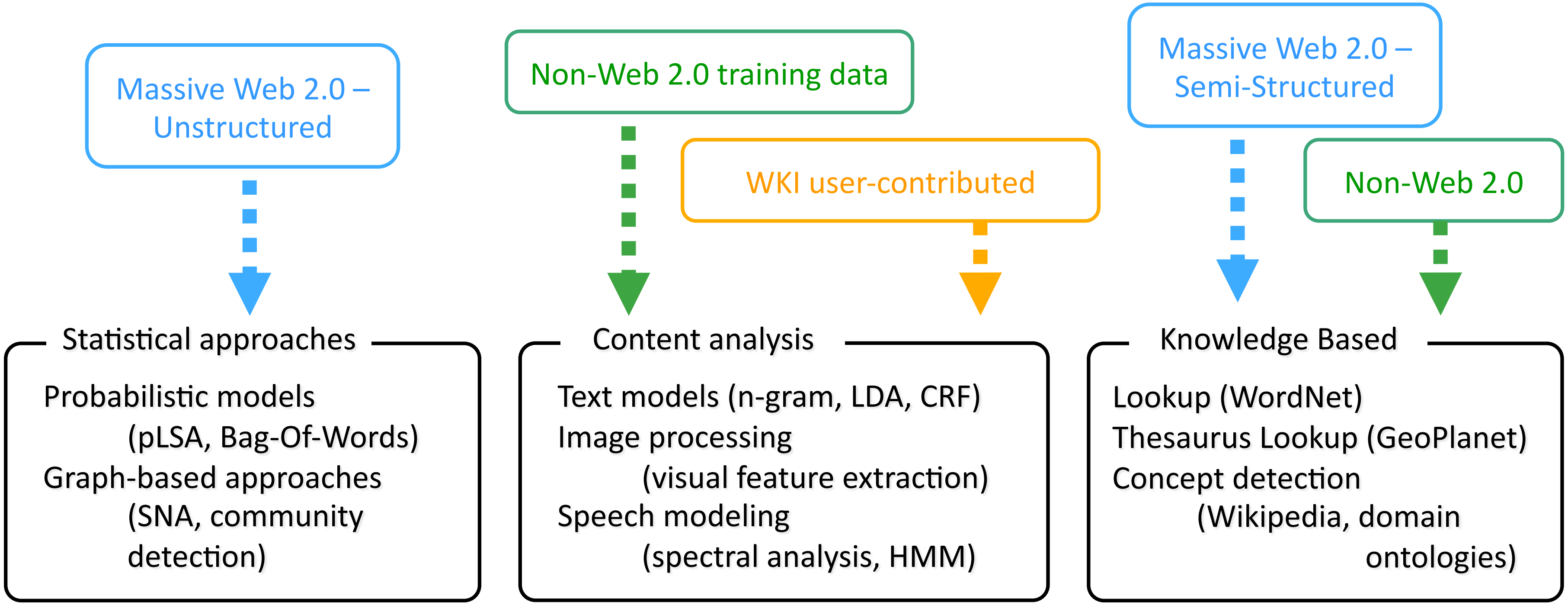
online → user profiling, method invocation

WKI user-contributed

Online content and user actions by WeKnowIt users. It is mainly used for triggering WeKnowIt services and for providing context to them, e.g. user profile, input content to be used as example for querying, etc.

Technical approach

Variety of approaches depending on content-metadata input.



Massive Web 2.0 – Unstructured

Non-Web 2.0 training data

WKI user-contributed

Massive Web 2.0 – Semi-Structured

Non-Web 2.0

Statistical approaches

Probabilistic models
(pLSA, Bag-Of-Words)
Graph-based approaches
(SNA, community detection)

Content analysis

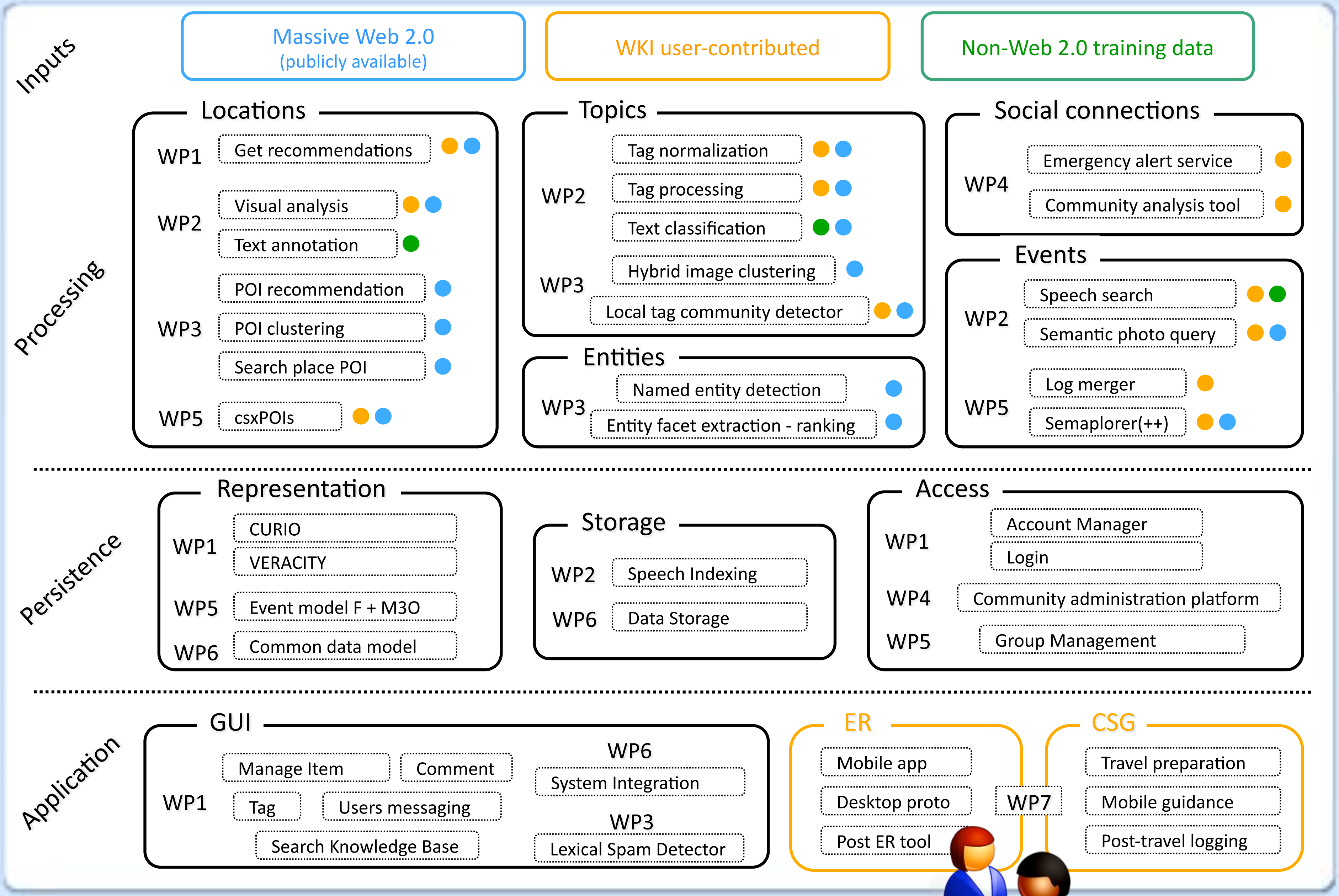
Text models (n-gram, LDA, CRF)
Image processing
(visual feature extraction)
Speech modeling
(spectral analysis, HMM)

Knowledge Based

Lookup (WordNet)
Thesaurus Lookup (GeoPlanet)
Concept detection
(Wikipedia, domain ontologies)

Massive → Collective Intelligence

Multi-Modal (Fusion) →Combined CI



Inputs

- 🔵 Massive Web 2.0 (publicly available)
- 🟠 WKI user-contributed
- 🟢 Non-Web 2.0 training data

Processing

Locations

- WP1
  - Get recommendations 🟠🔵
- WP2
  - Visual analysis 🟠🔵
  - Text annotation 🟢
- WP3
  - POI recommendation 🔵
  - POI clustering 🔵
  - Search place POI 🔵
- WP5
  - csxPOIs 🟠🔵

Topics

- WP2
  - Tag normalization 🟠🔵
  - Tag processing 🟠🔵
  - Text classification 🟢🔵
- WP3
  - Hybrid image clustering 🔵
  - Local tag community detector 🟠🔵

Entities

- WP3
  - Named entity detection 🔵
  - Entity facet extraction - ranking 🔵

Social connections

- WP4
  - Emergency alert service 🟠
  - Community analysis tool 🟠

Events

- WP2
  - Speech search 🟠🟢
  - Semantic photo query 🟠🔵
- WP5
  - Log merger 🟠
  - Semaplorer(++) 🟠🔵

Persistence

Representation

- WP1
  - CURIO
  - VERACITY
- WP5
  - Event model F + M3O
- WP6
  - Common data model

Storage

- WP2
  - Speech Indexing
- WP6
  - Data Storage

Access

- WP1
  - Account Manager
  - Login
- WP4
  - Community administration platform
- WP5
  - Group Management

Application

GUI

- WP1
  - Manage Item
  - Comment
  - Tag
  - Users messaging
  - Search Knowledge Base
- WP6
  - System Integration
- WP3
  - Lexical Spam Detector

ER

- Mobile app
- Desktop proto
- Post ER tool

WP7

CSG

- Travel preparation
- Mobile guidance
- Post-travel logging

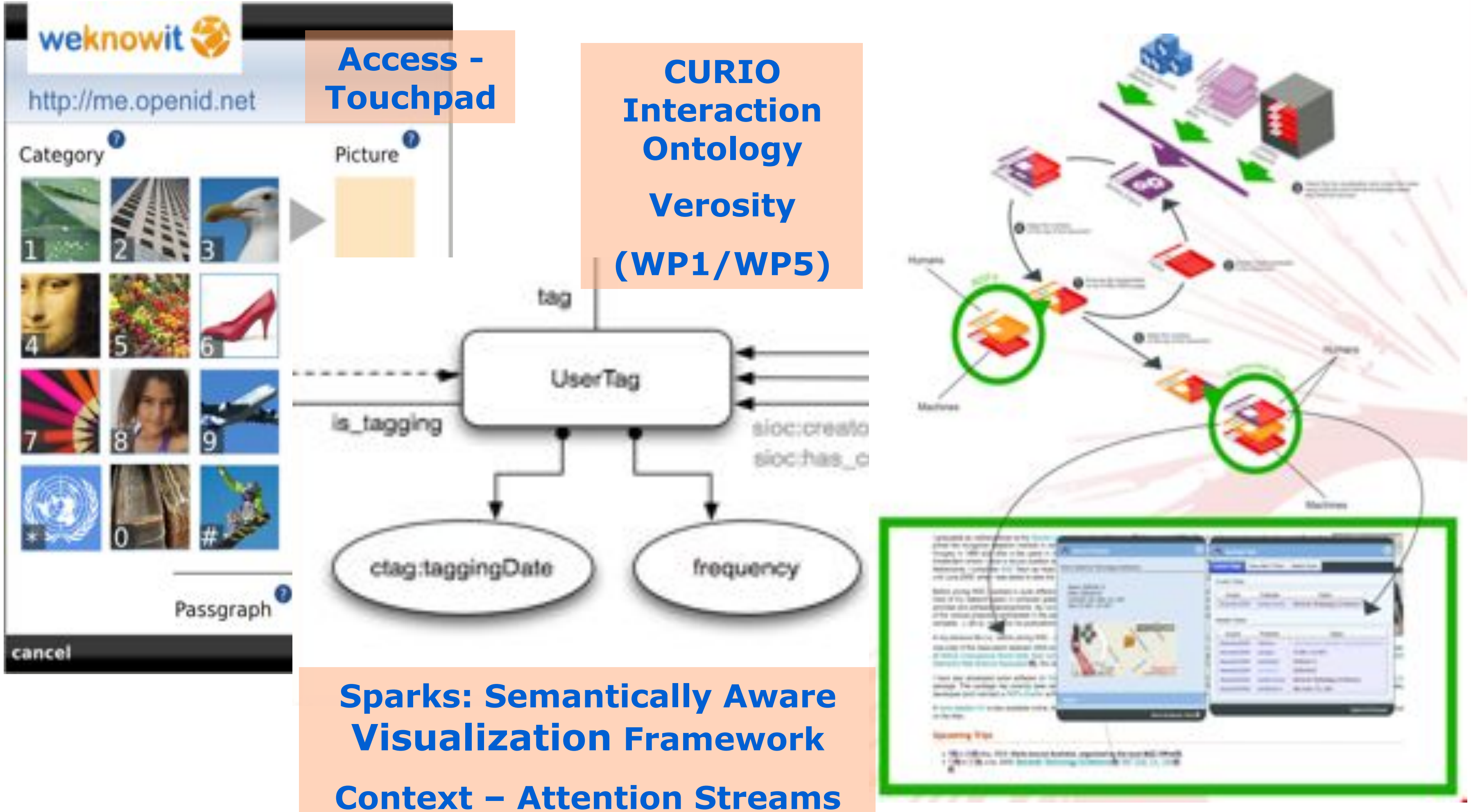
weknowit

Technologies Overview

SEVENTH FRAMEWORK PROGRAMME

co-funded by the European Union

WP1: Personal Intelligence



Access - Touchpad

CURIO Interaction Ontology

Verosity

(WP1/WP5)

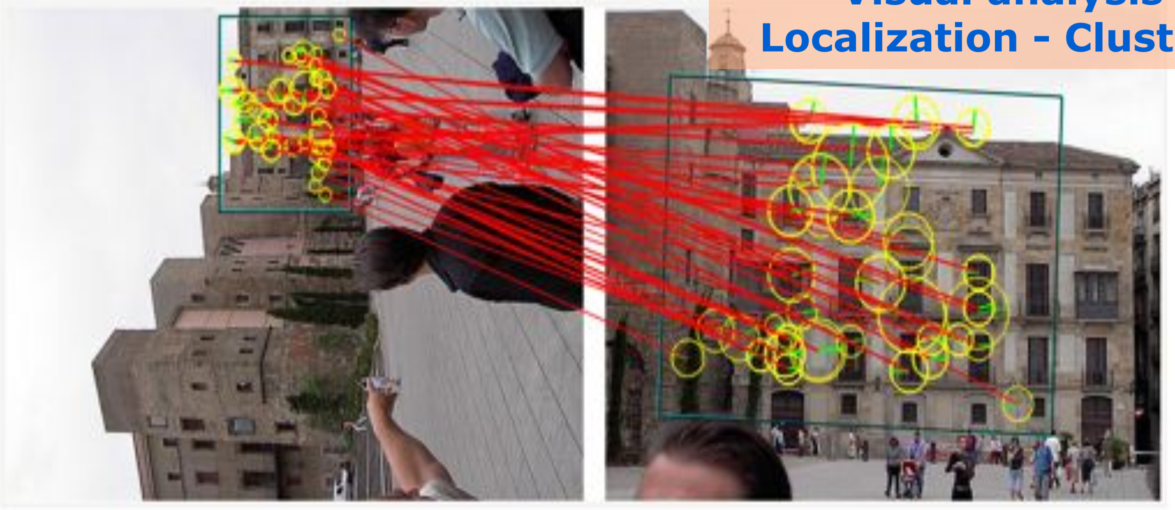
Sparks: Semantically Aware Visualization Framework

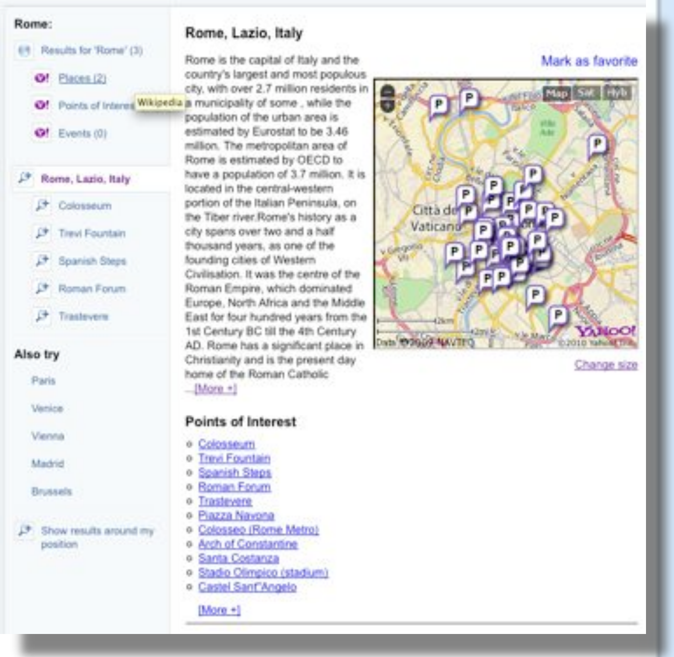
Context – Attention Streams

co-funded by the European Union

WP2: Media Intelligence

Visual analysis –
Localization - Clustering





Rome:

Results for 'Rome' (3)

- Places (2)
- Points of Interest
- Events (0)

Rome, Lazio, Italy

- Colosseum
- Trevi Fountain
- Spanish Steps
- Roman Forum
- Trastevere

Also try

- Paris
- Venice
- Vienna
- Madrid
- Brussels

Show results around my position

Rome, Lazio, Italy

Mark as favorite

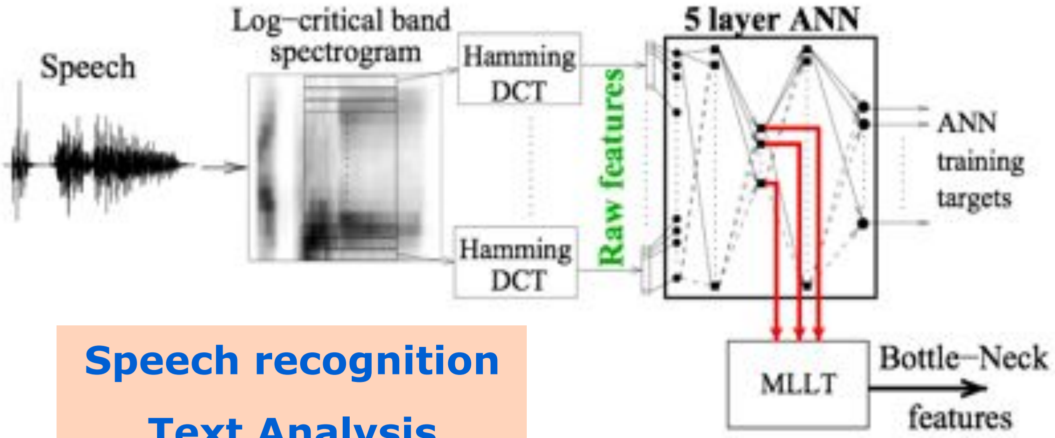
Rome is the capital of Italy and the country's largest and most populous city, with over 2.7 million residents in a municipality of some , while the population of the urban area is estimated by Eurostat to be 3.46 million. The metropolitan area of Rome is estimated by OECD to have a population of 3.7 million. It is located in the central-western portion of the Italian Peninsula, on the Tiber river.Rome's history as a city spans over two and a half thousand years, as one of the founding cities of Western Civilisation. It was the centre of the Roman Empire, which dominated Europe, North Africa and the Middle East for four hundred years from the 1st Century BC till the 4th Century AD. Rome has a significant place in Christianity and is the present day home of the Roman Catholic ...[More +]

Change size

Points of Interest

- Colosseum
- Trevi Fountain
- Spanish Steps
- Roman Forum
- Trastevere
- Piazza Navona
- Colosseo (Rome Metro)
- Arch of Constantine
- Santa Costanza
- Stadio Olimpico (stadium)
- Castel Sant'Angelo

[More +]



Ranked entities

Speech recognition

Text Analysis

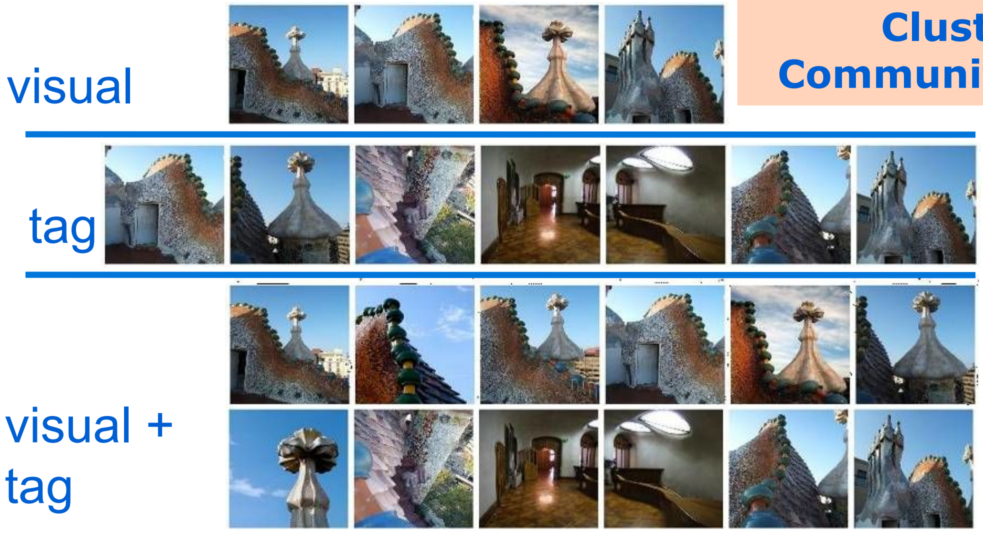
Combined CI Services

weknowit

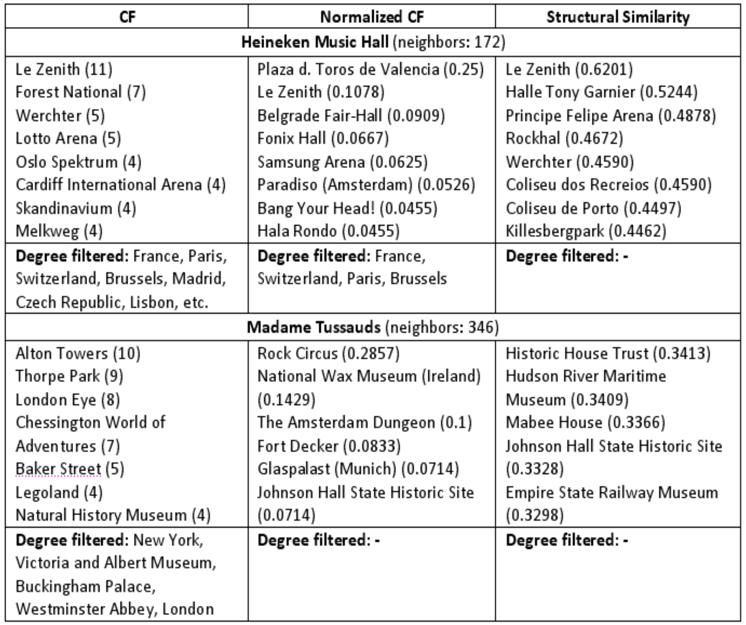
SEVENTH FRAMEWORK PROGRAMME

co-funded by the European Union

WP3: Mass Intelligence

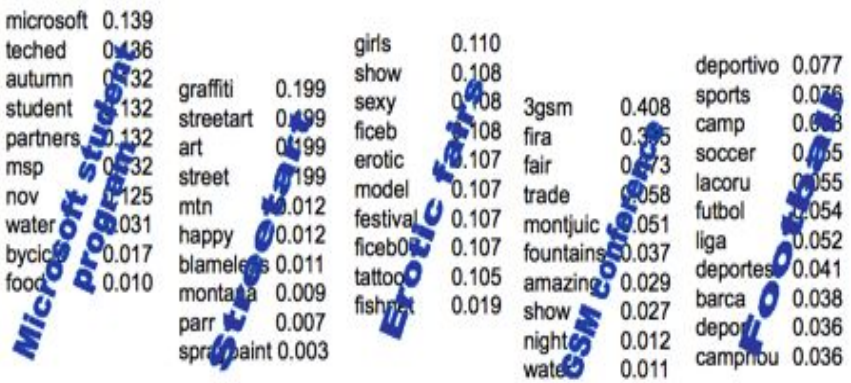


Clustering – Community Detection



CF	Normalized CF	Structural Similarity
Heineken Music Hall (neighbors: 172)		
Le Zenith (11) Forest National (7) Werchter (5) Lotto Arena (5) Oslo Spektrum (4) Cardiff International Arena (4) Skandinavium (4) Melkweg (4)	Plaza d. Toros de Valencia (0.25) Le Zenith (0.1078) Belgrade Fair-Hall (0.0909) Fonix Hall (0.0667) Samsung Arena (0.0625) Paradiso (Amsterdam) (0.0526) Bang Your Head! (0.0455) Hala Rondo (0.0455)	Le Zenith (0.6201) Halle Tony Garnier (0.5244) Principe Felipe Arena (0.4878) Rockhal (0.4672) Werchter (0.4590) Coliseu dos Recreios (0.4590) Coliseu de Porto (0.4497) Killesbergpark (0.4462)
Degree filtered: France, Paris, Switzerland, Brussels, Madrid, Czech Republic, Lisbon, etc.	**Degree filtered:** France, Switzerland, Paris, Brussels	**Degree filtered:** -
Madame Tussauds (neighbors: 346)		
Alton Towers (10) Thorpe Park (9) London Eye (8) Chessington World of Adventures (7) Baker Street (5) Legoland (4) Natural History Museum (4)	Rock Circus (0.2857) National Wax Museum (Ireland) (0.1429) The Amsterdam Dungeon (0.1) Fort Decker (0.0833) Glaspalast (Munich) (0.0714) Johnson Hall State Historic Site (0.0714)	Historic House Trust (0.3413) Hudson River Maritime Museum (0.3409) Mabee House (0.3366) Johnson Hall State Historic Site (0.3328) Empire State Railway Museum (0.3298)
Degree filtered: New York, Victoria and Albert Museum, Buckingham Palace, Westminster Abbey, London	**Degree filtered:** -	**Degree filtered:** -

PoI recommendation


Microsoft student program
Street art
Erotic fairs
GSM conference
Football

microsoft	0.139
teched	0.136
autumn	0.132
student	0.132
partners	0.132
msp	0.132
nov	0.125
water	0.031
bycic[illegible]	0.017
food	0.010

graffiti	0.199
streetart	0.199
art	0.199
street	0.199
mtn	0.012
happy	0.012
blameless	0.011
montana	0.009
parr	0.007
spraypaint	0.003

girls	0.110
show	0.108
sexy	0.108
ficeb	0.108
erotic	0.107
model	0.107
festival	0.107
ficeb08	0.107
tattoo	0.105
fishnet	0.019

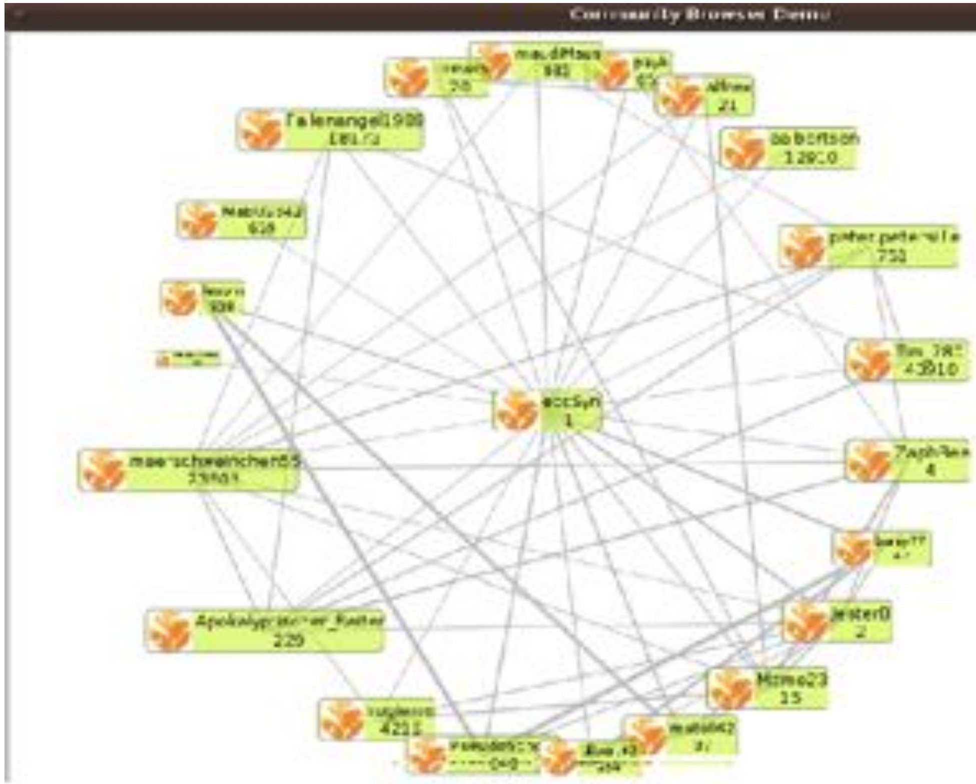
3gsm	0.408
fira	0.3[illegible]5
fair	0.[illegible]73
trade	0.058
montjuic	0.051
fountains	0.037
amazing	0.029
show	0.027
night	0.012
water	0.011

deportivo	0.077
sports	0.076
camp	0.0[illegible]3
soccer	0.[illegible]55
lacoru	0.055
futbol	0.054
liga	0.052
deportes	0.041
barca	0.038
depor	0.036
campnou	0.036

Event detection

SEVENTH FRAMEWORK PROGRAMME
co-funded by the European Union

WP4: Social Intelligence

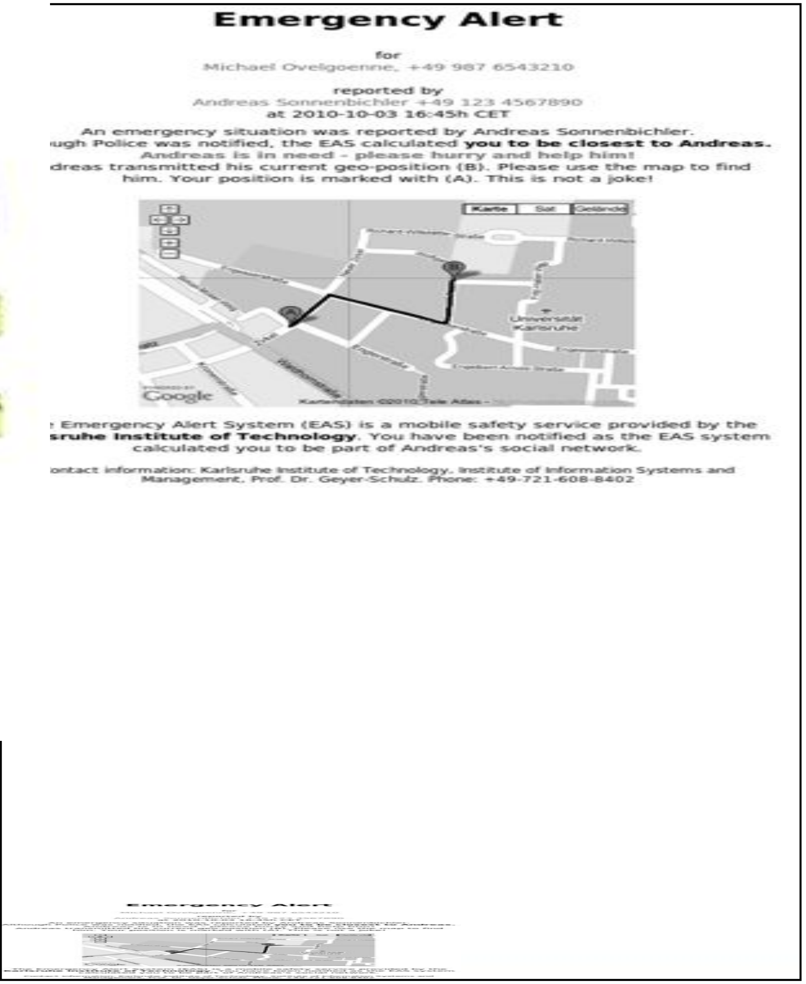


Community administration platform

Community analysis tool

Community browser

Emergency Alert Service



Emergency Alert

for
Michael Ovelgoenne, +49 987 6543210

reported by
Andreas Sonnenbichler +49 123 4567890
at 2010-10-03 16:45h CET

An emergency situation was reported by Andreas Sonnenbichler.
ugh Police was notified, the EAS calculated **you to be closest to Andreas.**
Andreas is in need - please hurry and help him!
dreas transmitted his current geo-position (B). Please use the map to find him. Your position is marked with (A). This is not a joke!

: Emergency Alert System (EAS) is a mobile safety service provided by the **sruhe Institute of Technology**. You have been notified as the EAS system calculated you to be part of Andreas's social network.

ontact information: Karlsruhe Institute of Technology, Institute of Information Systems and Management, Prof. Dr. Geyer-Schulz. Phone: +49-721-608-8402

weknowit

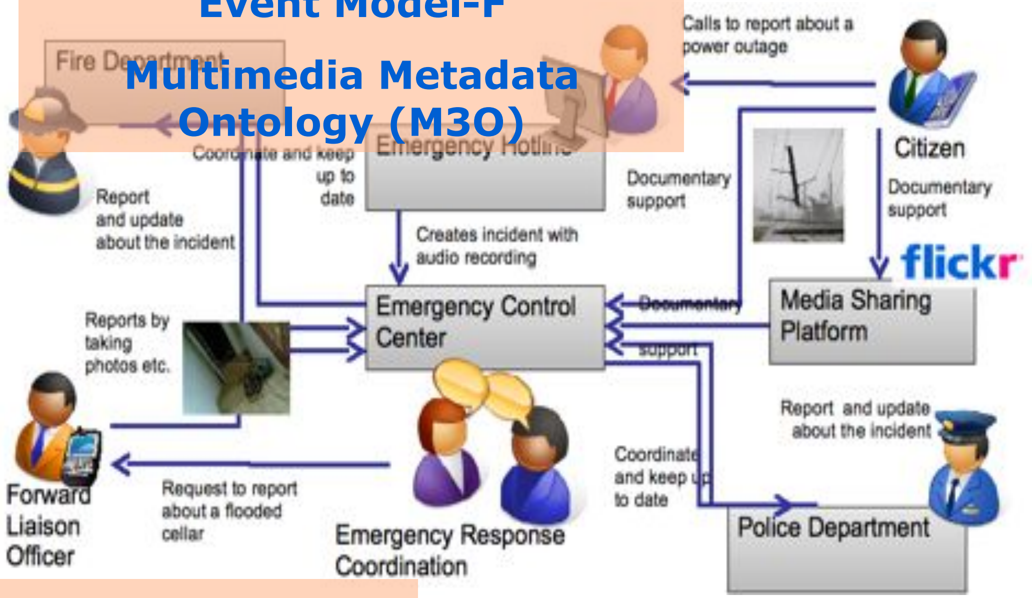
SEVENTH FRAMEWORK PROGRAMME

co-funded by the European Union

WP5: Organisational Intelligence

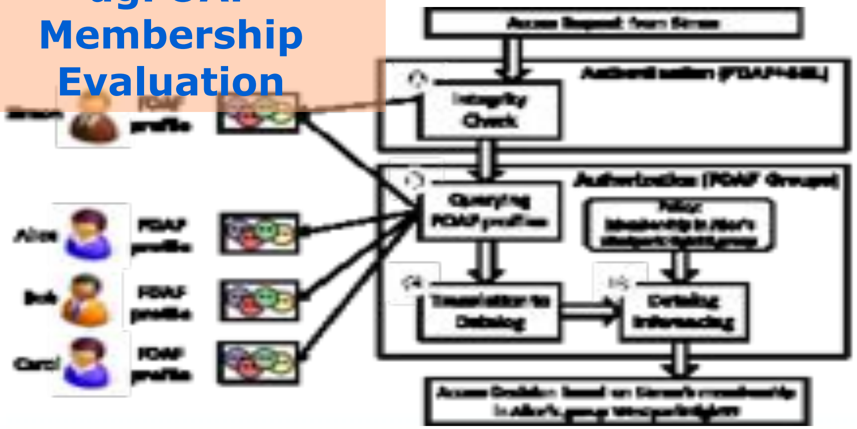
Event Model-F

Multimedia Metadata Ontology (M3O)

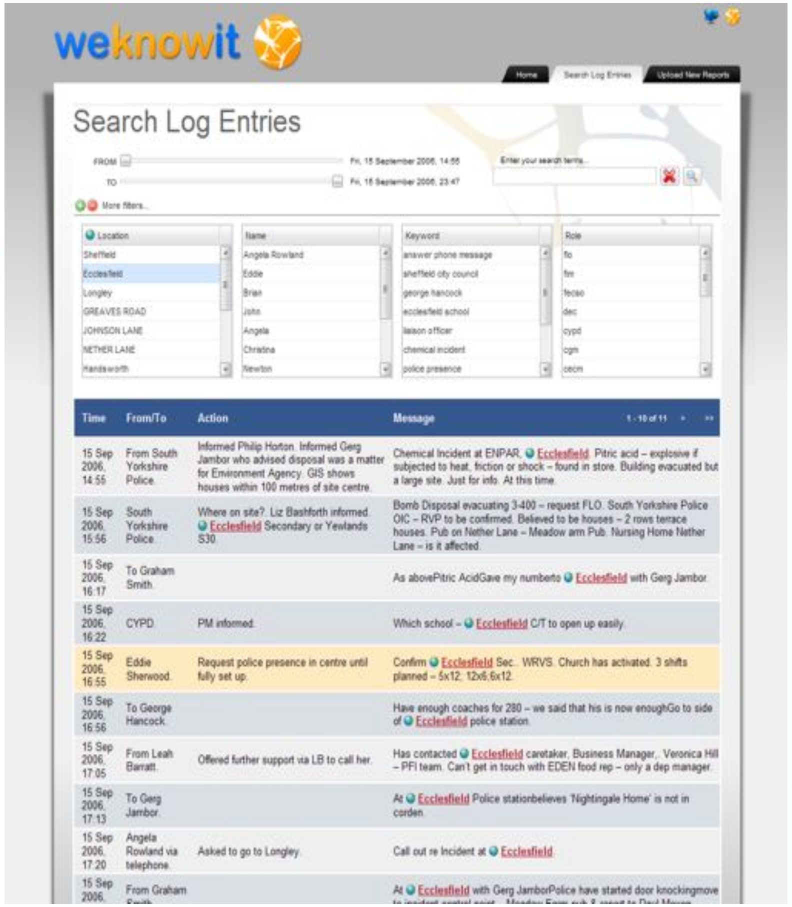


Fire Department
Emergency Hotline
Calls to report about a power outage
Citizen
Coordinate and keep up to date
Report and update about the incident
Documentary support
Documentary support
Creates incident with audio recording
flickr
Emergency Control Center
Documentary support
Media Sharing Platform
Reports by taking photos etc.
Report and update about the incident
Coordinate and keep up to date
Forward Liaison Officer
Request to report about a flooded cellar
Emergency Response Coordination
Police Department

dgFOAF Membership Evaluation



Access Request from Simon
Authentication (FOAF+SSL)
Integrity Check
Simon
FOAF profile
Authorization (FOAF Groups)
Querying FOAF profiles
Policy: Membership in Alice's student/sport group
Alice
FOAF profile
Bob
FOAF profile
Carol
FOAF profile
Translation to Datalog
Datalog Inferencing
Access Decision based on Simon's membership in Alice's group



Search Log Entries

FROM Fri, 15 September 2006, 14:55

TO Fri, 15 September 2006, 23:47

Enter your search terms...

More filters...

Location	Name	Keyword	Role
Sheffield	Angela Rowland	answer phone message	flo
Ecclesfield	Eddie	sheffield city council	fm
Longley	Brian	george hancock	fecso
GREAVES ROAD	John	ecclesfield school	dec
JOHNSON LANE	Angela	liaison officer	cypd
NETHER LANE	Christina	chemical incident	ogm
Handsworth	Newton	police presence	cecm

1 - 10 of 11

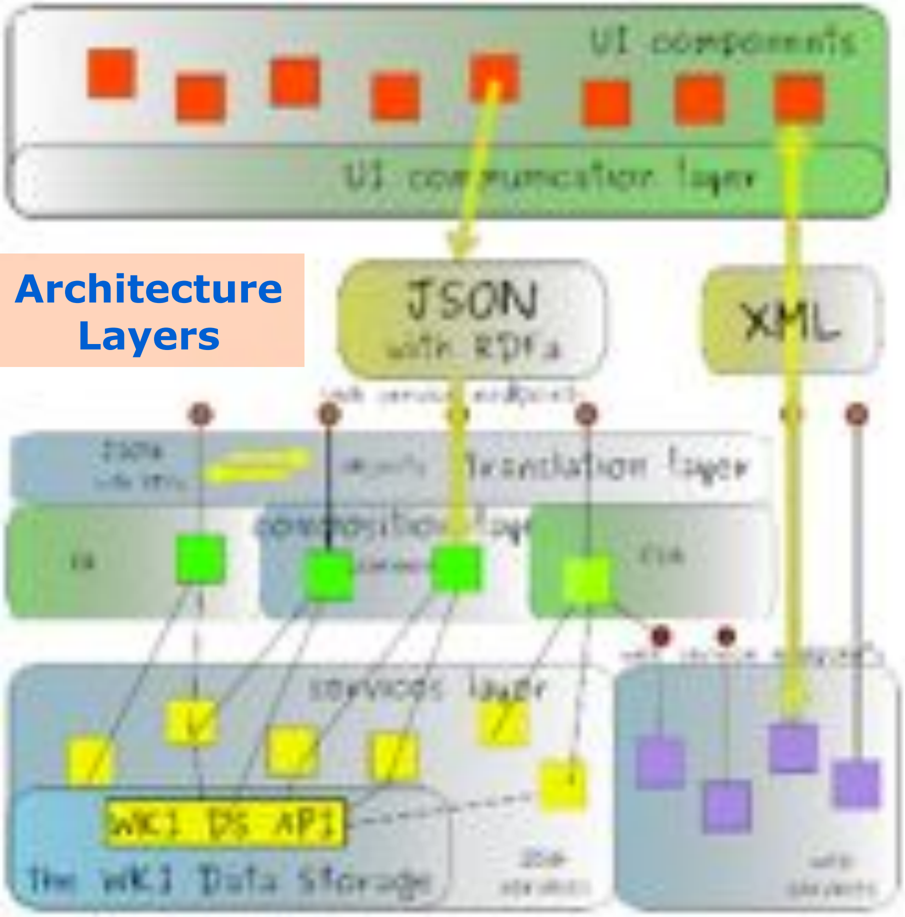
Time	From/To	Action	Message
15 Sep 2006, 14:55	From South Yorkshire Police	Informed Philip Horton. Informed Gerg Jambor who advised disposal was a matter for Environment Agency. GIS shows houses within 100 metres of site centre.	Chemical Incident at ENPAR, Ecclesfield. Pitric acid – explosive if subjected to heat, friction or shock – found in store. Building evacuated but a large site. Just for info. At this time.
15 Sep 2006, 15:56	South Yorkshire Police	Where on site? Liz Bashforth informed. Ecclesfield Secondary or Yewlands S30.	Bomb Disposal evacuating 3-400 – request FLO. South Yorkshire Police OIC – RVP to be confirmed. Believed to be houses – 2 rows terrace houses. Pub on Nether Lane – Meadow arm Pub. Nursing Home Nether Lane – is it affected.
15 Sep 2006, 16:17	To Graham Smith.		As abovePitric AcidGave my numberto Ecclesfield with Gerg Jambor.
15 Sep 2006, 16:22	CYPD	PM informed.	Which school – Ecclesfield C/T to open up easily.
15 Sep 2006, 16:55	Eddie Sherwood.	Request police presence in centre until fully set up.	Confirm Ecclesfield Sec.. WRVS. Church has activated. 3 shifts planned – 5x12; 12x6;6x12.
15 Sep 2006, 16:56	To George Hancock.		Have enough coaches for 280 – we said that his is now enoughGo to side of Ecclesfield police station.
15 Sep 2006, 17:05	From Leah Barratt.	Offered further support via LB to call her.	Has contacted Ecclesfield caretaker, Business Manager,. Veronica Hill – PFI team. Can't get in touch with EDEN food rep – only a dep manager.
15 Sep 2006, 17:13	To Gerg Jambor.		At Ecclesfield Police stationbelieves 'Nightingale Home' is not in cordon.
15 Sep 2006, 17:20	Angela Rowland via telephone.	Asked to go to Longley.	Call out re Incident at Ecclesfield.
15 Sep 2006,	From Graham		At Ecclesfield with Gerg JamborPolice have started door knockingmove

Log Merger

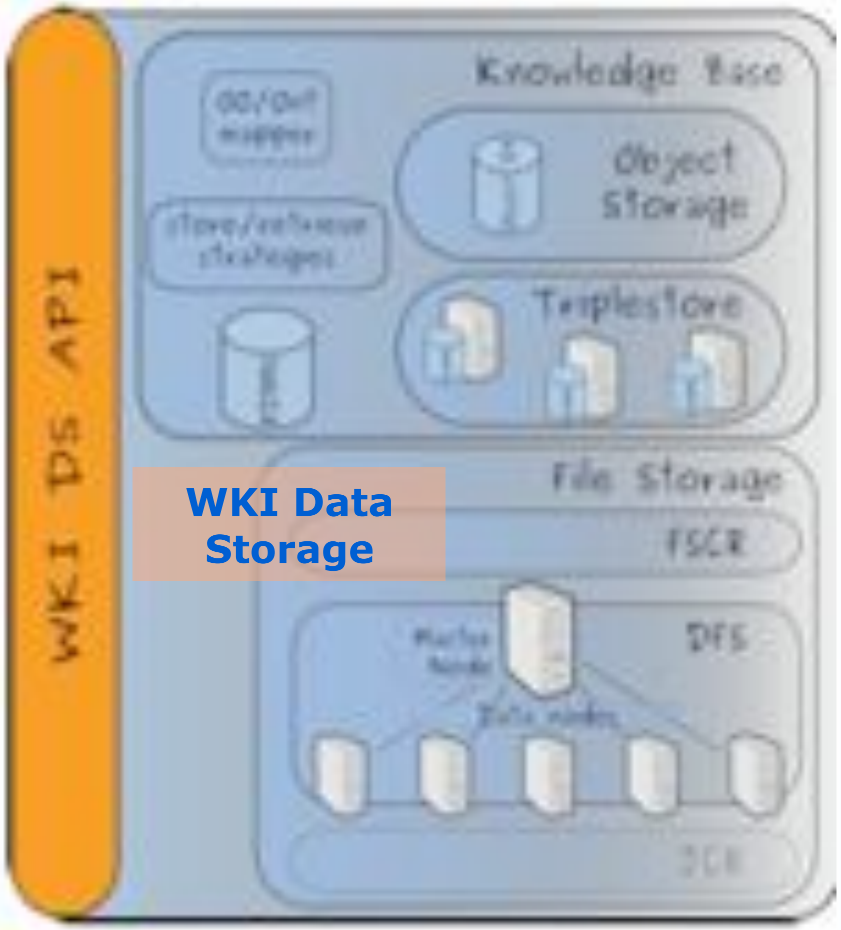
SEVENTH FRAMEWORK PROGRAMME

co-funded by the European Union

WP6: Architecture & Integration



Architecture Layers



WKI Data Storage

30 services integrated

weknowit

SEVENTH FRAMEWORK PROGRAMME
co-funded by the European Union

WeKnowIt Community Detection

Challenges in Social Media network mining

No prior assumptions about structure:

- Complex & evolving structure
- No possibility for knowing structural features (e.g. number of clusters on a graph) in advance

→ Unsupervised

Scale

- Tens of millions of active users frequently contributing loads of content links + metadata (tags, comments, ratings)

→ Efficient - scalable

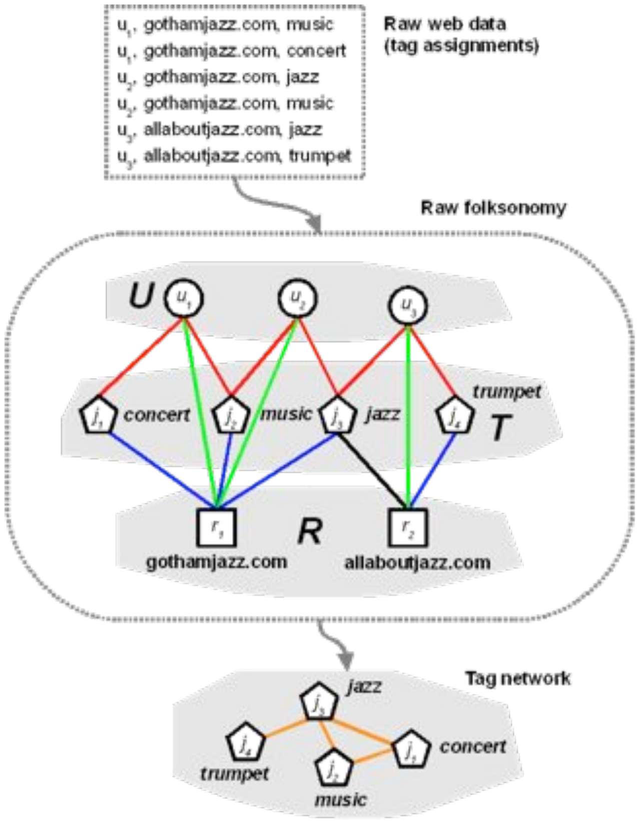
Quality

- Spam is very common. Only a portion of user contributions is worth further analysis.

→ Noise resilient

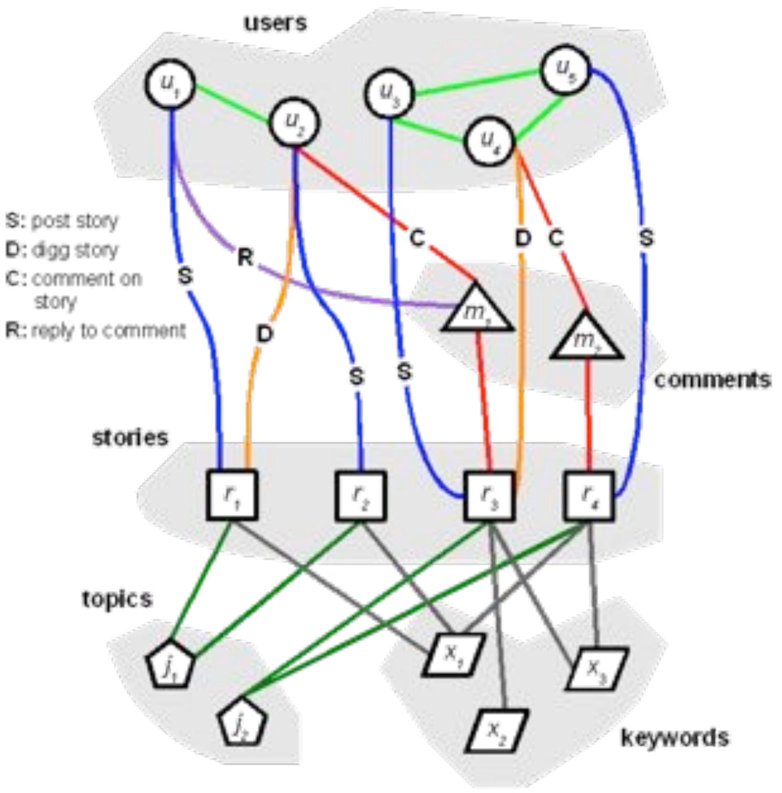
Examples of Social Media networks

Folksonomy (Delicious)



Mika, P. (2005) Ontologies Are Us: A Unified Model of Social Networks and Semantics. Proceedings of the 4th International Semantic Web Conference (ISWC 2005), Springer Berlin / Heidelberg, pp. 522-536

MetaGraph (Digg)

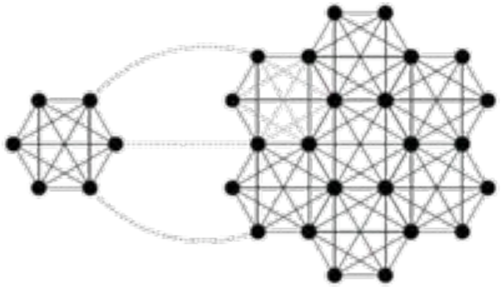


Lin, Y., Sun, J., Castro, P., Konuru, R., Sundaram, H., and Kelliher, A. (2009) MetaFac: community discovery via relational hypergraph factorization. Proceedings of KDD '09, ACM, pp. 527-536

What is a community in a network?

Group of vertices that are more densely connected to each other than to the rest of the network.

Multiple definitions to quantify communities:



Fortunato S. (2010) Community detection in graphs. Physics Reports486: 75-174

Global: N-cut, conductance, modularity

Local: Local modularity, (μ,ε)-cores

Ad hoc: Label propagation, dynamic synchronization

Related to clustering, but: (a) not necessary to know number of communities, (b) computationally more efficient

In WeKnowIt, we focus on local definitions, because of the properties of Social Media networks: efficiency-scalability and noise resilience.

Approach illustration

Two-step process:

- 1st step:

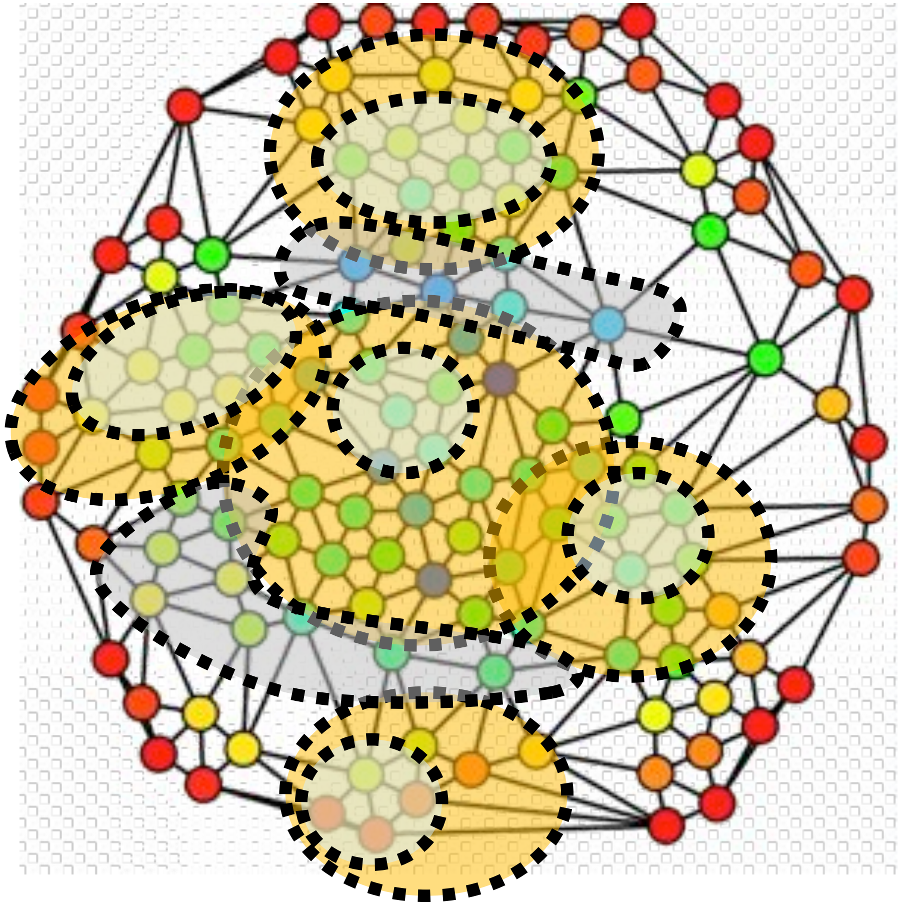
  (μ, ε) – core detection

- 2nd step:

  Local expansion

- 3rd step:

  Characterization of remaining vertices as *hubs* or *outliers*



Hybrid Photo Clustering

Goal:

- Group large photo collections into clusters based on how much they are related to each other
- Assist browsing and navigation by means of a map-based application
- Detect landmark and event clusters.

Combine both visual features *and* tags

- Two kinds of similarity (visual and tag networks) are complementary to each other
- Many times one photo has missing tags or is hard to interpret visually
- Graph-based approach - superimpose visual and tag graphs
- Use photo cluster features for classification to landmarks/events

Results

- Higher quality clusters by use of both visual and tag similarity instead of only each one of them.
- Clusters can be used for landmark and event detection.
- Integrated in CSG prototype and ClustTour stand-alone demo.

Overview of approach

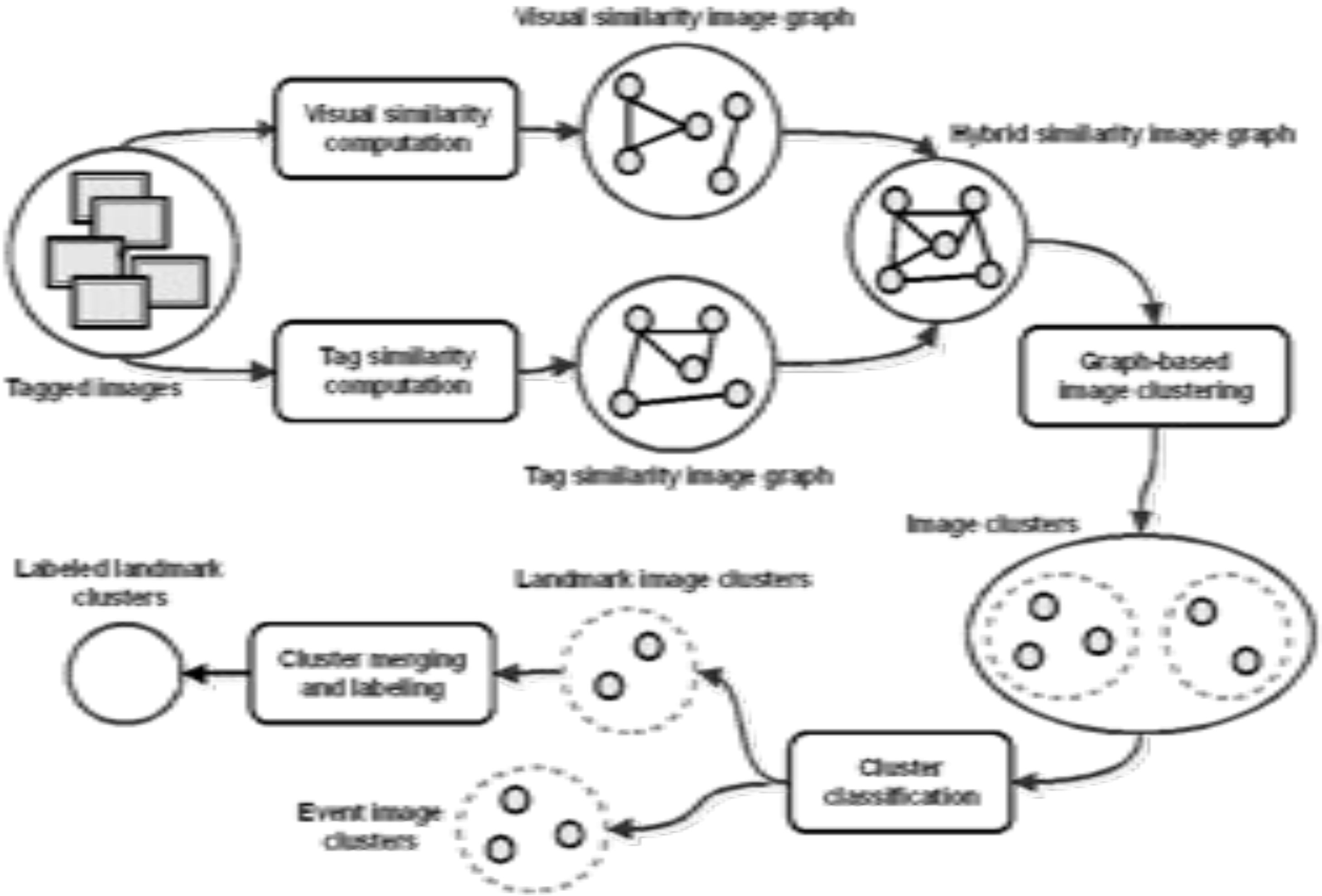
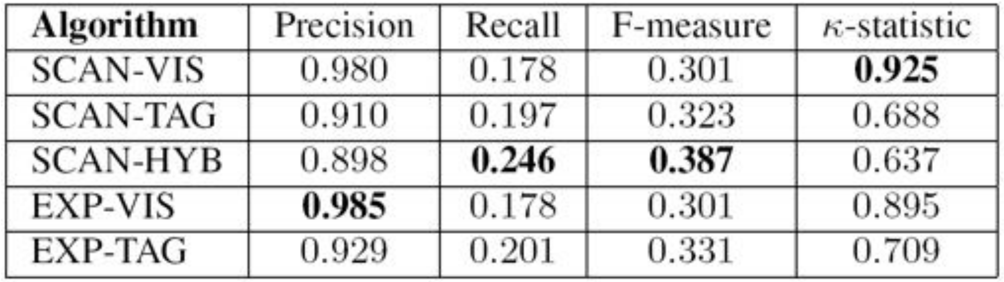

Visual similarity image graph
Visual similarity computation
Hybrid similarity image graph
Tagged images
Tag similarity computation
Tag similarity image graph
Graph-based image clustering
Image clusters
Labeled landmark clusters
Landmark image clusters
Cluster merging and labeling
Cluster classification
Event image clusters

Photo clustering results (1)

User study (involving 20 people)

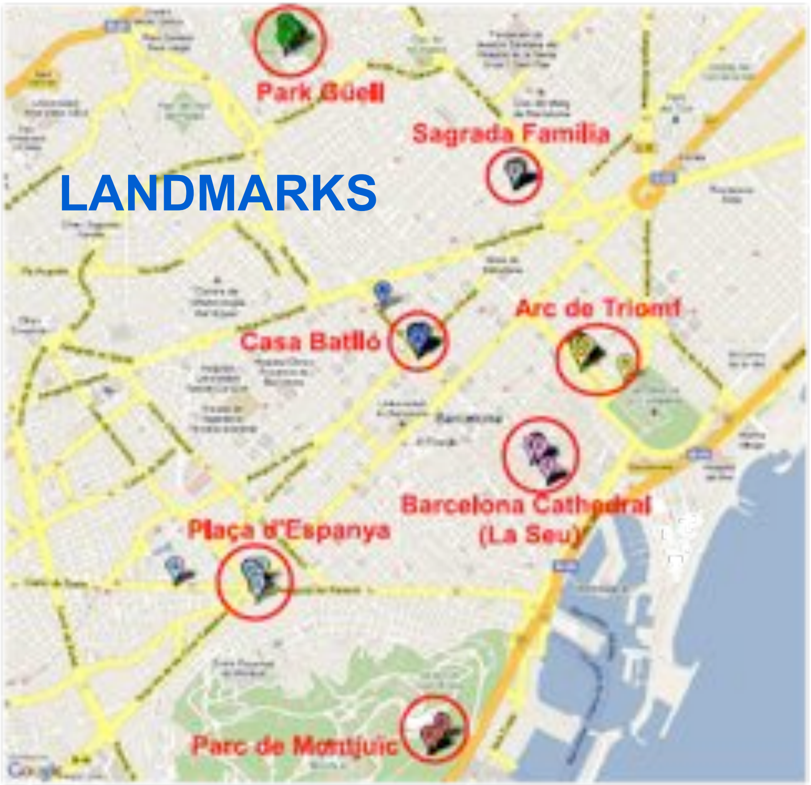
- Users where shown photo clusters and they were asked to judge how relevant the photos of each cluster were related to each other
- Each cluster was produced by different notion of similarity (tag-only, visual-only, hybrid). Obviously, users were not aware of this information
- Hybrid clusters were found to be of superior quality (highest F-measure)



Algorithm	Precision	Recall	F-measure	κ-statistic
SCAN-VIS	0.980	0.178	0.301	**0.925**
SCAN-TAG	0.910	0.197	0.323	0.688
SCAN-HYB	0.898	**0.246**	**0.387**	0.637
EXP-VIS	**0.985**	0.178	0.301	0.895
EXP-TAG	0.929	0.201	0.331	0.709

Photo clustering results (2)

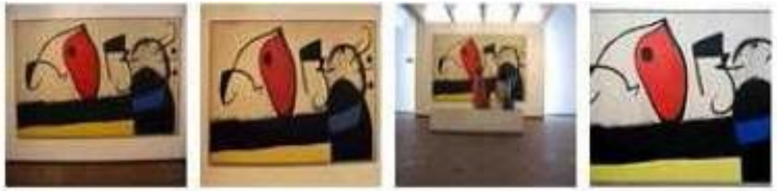
Geographic localization of results was also found to be very high. Most clusters correspond to landmarks or events.



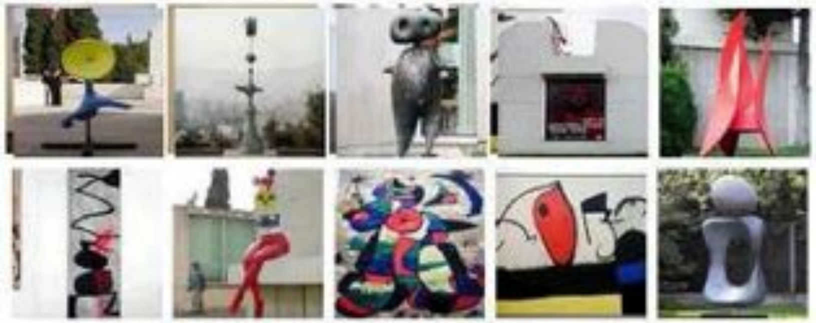
EVENTS

Sample results:
[Visual] vs. [Tag] vs. [Visual + Tag]

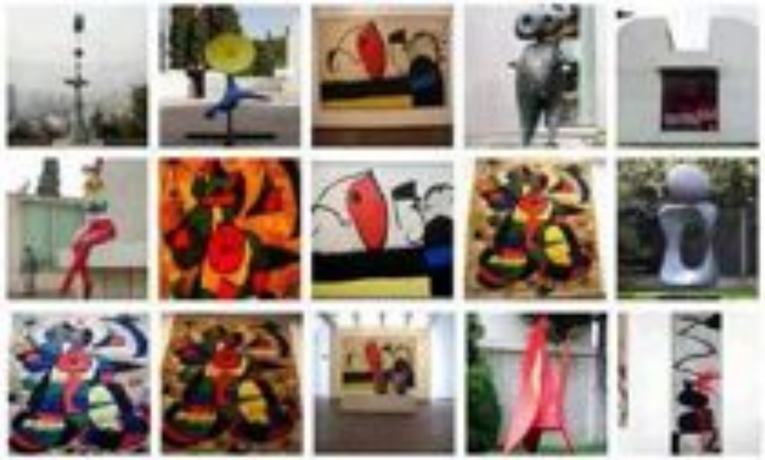
VISUAL



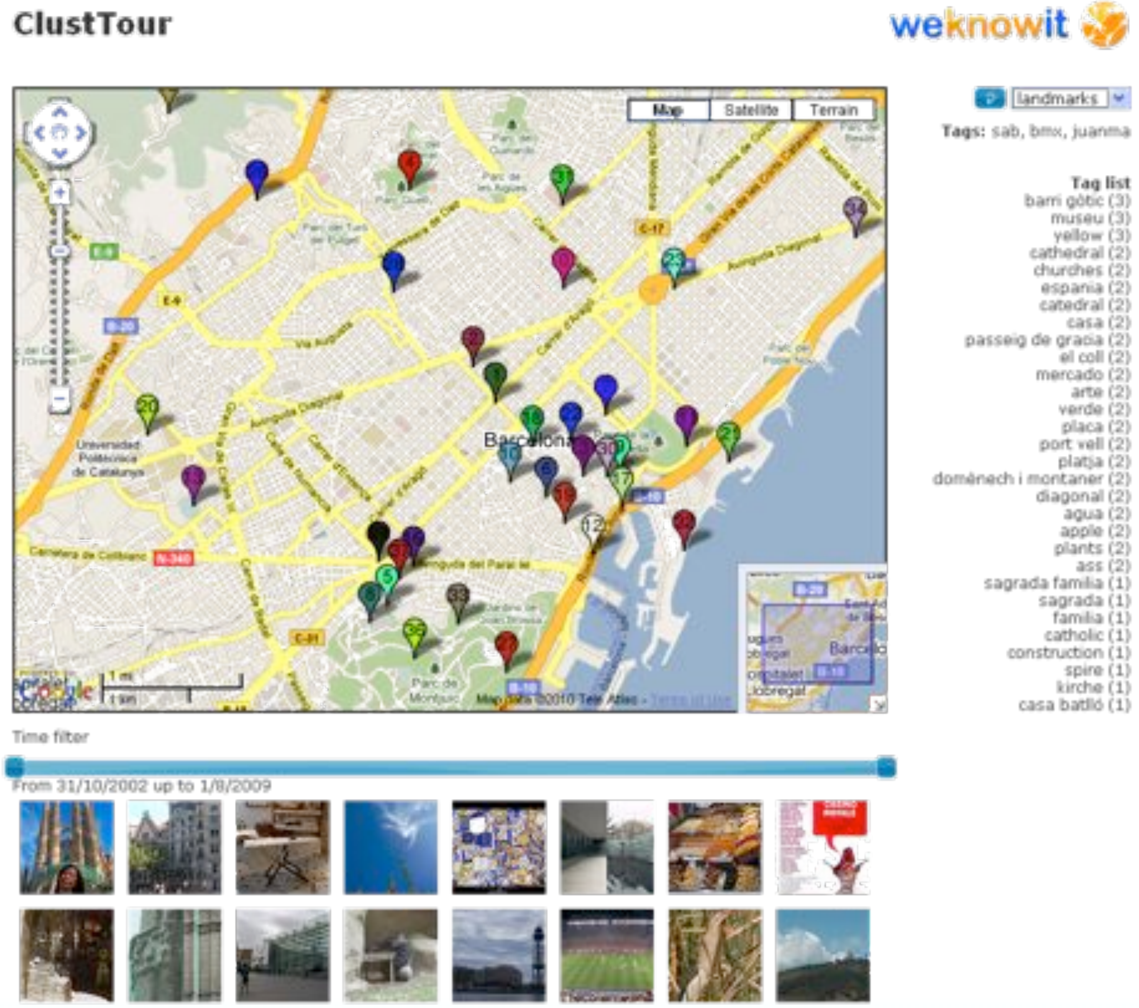
TAG



HYBRID



ClustTour demo: City exploration by means of photo clusters



ClustTour

landmarks

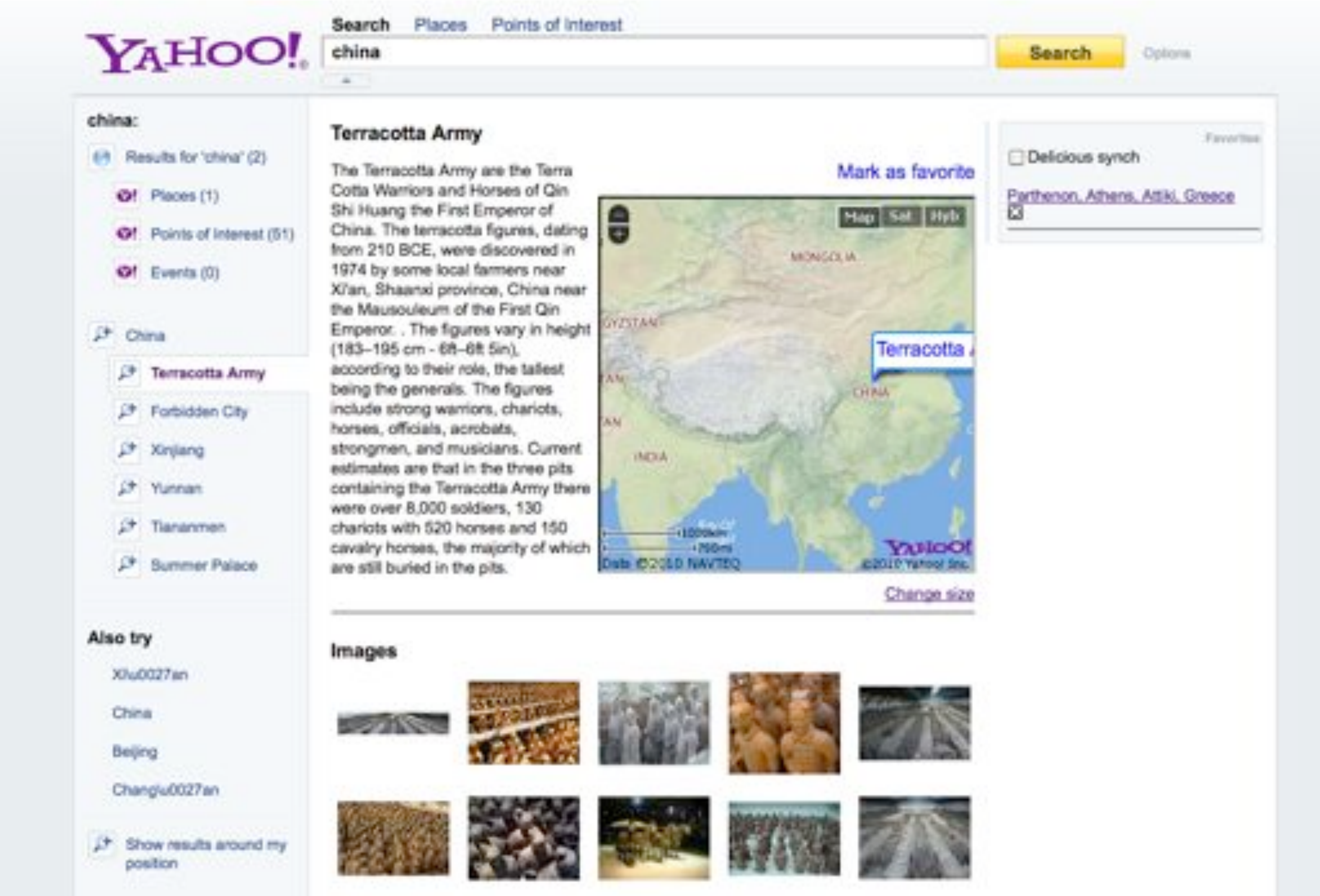
Tags: sab, bmx, juanma

Tag list
barri gòtic (3)
museu (3)
yellow (3)
cathedral (2)
churches (2)
espania (2)
catedral (2)
casa (2)
passeig de gracia (2)
el coll (2)
mercado (2)
arte (2)
verde (2)
placa (2)
port vell (2)
platja (2)
domènech i montaner (2)
diagonal (2)
agua (2)
apple (2)
plants (2)
ass (2)
sagrada familia (1)
sagrada (1)
familia (1)
catholic (1)
construction (1)
spire (1)
kirche (1)
casa batlló (1)

Time filter

From 31/10/2002 up to 1/8/2009

Travel demo



china

china:

Results for 'china' (2)

Places (1)

Points of interest (51)

Events (0)

China

Terracotta Army

Forbidden City

Xinjiang

Yunnan

Tiananmen

Summer Palace

Also try

Xi\u0027an

China

Beijing

Changl\u0027an

Show results around my position

Terracotta Army

Mark as favorite

The Terracotta Army are the Terra Cotta Warriors and Horses of Qin Shi Huang the First Emperor of China. The terracotta figures, dating from 210 BCE, were discovered in 1974 by some local farmers near Xi'an, Shaanxi province, China near the Mausoleum of the First Qin Emperor. . The figures vary in height (183–195 cm - 6ft–6ft 5in), according to their role, the tallest being the generals. The figures include strong warriors, chariots, horses, officials, acrobats, strongmen, and musicians. Current estimates are that in the three pits containing the Terracotta Army there were over 8,000 soldiers, 130 chariots with 520 horses and 150 cavalry horses, the majority of which are still buried in the pits.

Change size

Images

□ Delicious synch

Parthenon, Athens, Attiki, Greece
☒

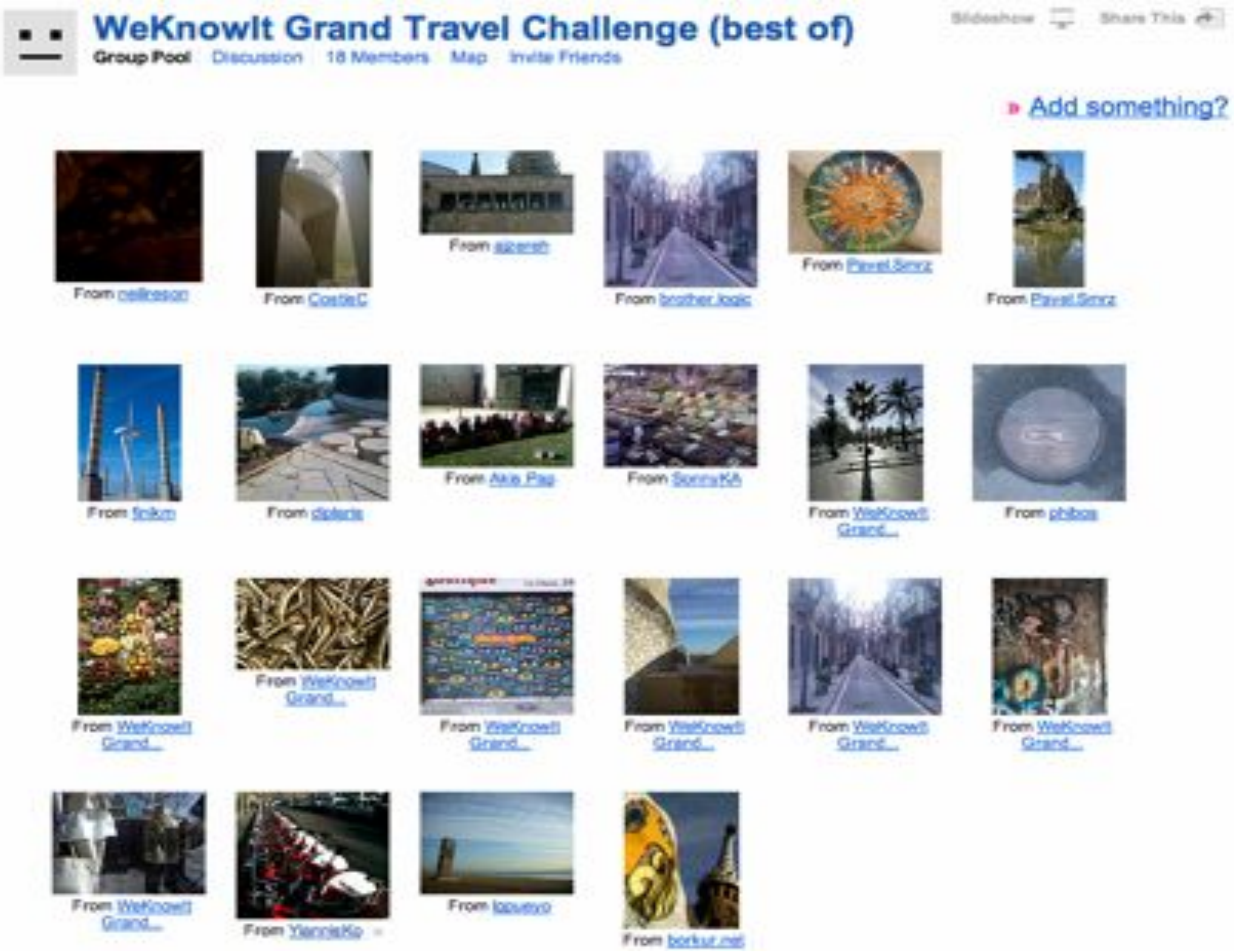
WKI Grand Travel Challenge

Barcelona, January 21st 2010

SEVENTH FRAMEWORK PROGRAMME
co-funded by the European Union

WKI Grand Travel Challenge

Barcelona, January 21st 2010

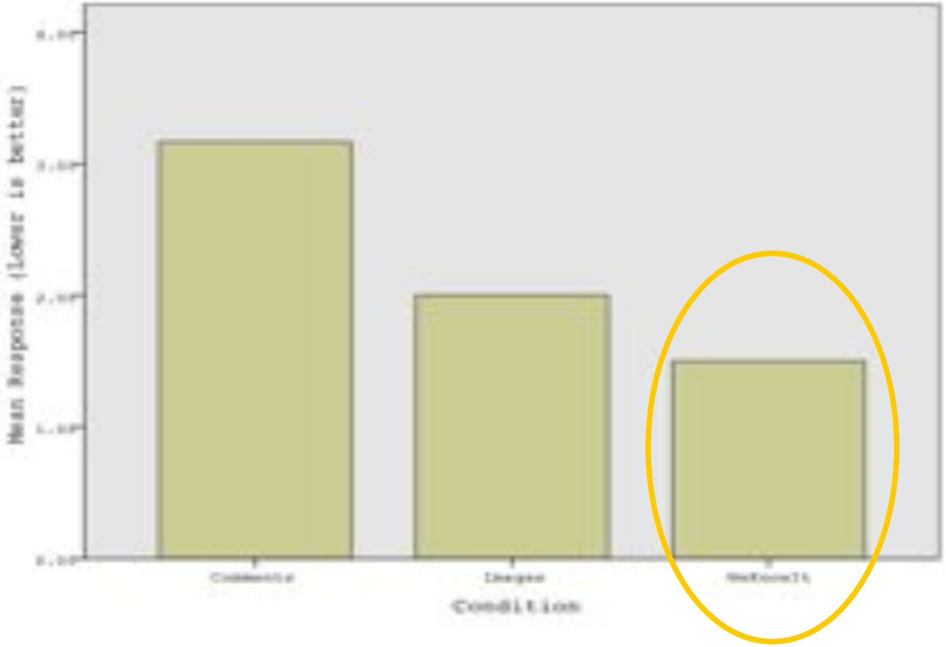


WeKnowIt Grand Travel Challenge (best of)

Evaluation results

- 5 different evaluation activities
  - 1 ER specific
  - 2 CSG mobile
  - 2 CGS desktop
- 59 users
- 81 system tests
  - ER personell
  - Citizens
  - Users
  - TID personell
  - WKI consortium

"I found it easy to process all the information available to me"



Experts

"I felt I understood what was going on quickly"

co-funded by the European Union

Content - Emergency Response

- Text
    - Sheffield floods forum messages and posts (thousands)
    - Police & fire service logs
    - ABC news tagged articles (~7000 metadata files)
- Image
    - Flickr images + metadata
        - 136 related to June 2007 Sheffield floods
        - ~27K geo-coded photos around the area of Sheffield
            - 1400 ER images (after WKI clustering)
- Speech
    - 1000+ emergency phone calls on Sheffield flooding event
    - 1000+ voice-tagging events by at least 10 users
    - Fused text+speech dataset

SEVENTH FRAMEWORK PROGRAMME
co-funded by the European Union

Content - Consumer Social Group

- Text
  - Flickr – metadata from geotagged London images (4300+ files)
  - Wikipedia processing
  - GeoPlanet processing
- Image
  - VIRaL
    (1.2 million geotagged images - 22 European cities)
  - Barcelona meeting dataset - geotagged and tagged (647 images, 1669 tags)
  - 1000 restaurant images
- Social networks
  - Barcelona meeting network of contacts (14 users)

SEVENTH FRAMEWORK PROGRAMME

co-funded by the European Union

Research Fields and Issues

- Statistical analysis, machine learning, data mining, pattern recognition, social network analysis
- Clustering
- Graph theory
- Image, text, video analysis
- Information extraction
- Fusion techniques
- Trust, security, privacy
- Performance, scalability
  - speed, storage, power, grids, clouds

Conclusions

- Collective Intelligence can be extracted by social media
- New applications and services can be developed
- Fusion of multimodal – multisource info remains a challenge
- Scalability, quality, coverage are important issues

weknowit

SEVENTH FRAMEWORK PROGRAMME

co-funded by the European Union