Modelli Semantici, Collaborativi e Sociali: problematiche, architetture ICT, campi applicativi





seminario per il Corso di Dottorato

Prof. Paolo Nesi

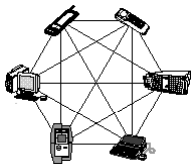
Department of Systems and Informatics
University of Florence
Via S. Marta 3, 50139, Firenze, Italy
tel: +39-055-4796523, fax: +39-055-4796363

Lab: DISIT, Sistemi Distribuiti e Tecnologie Internet

nesi@dsi.unifi.it, nesi@computer.org

http://www.disit.dsi.unifi.it/

Struttura del Seminario



- Sistemi Distribuiti
- Sistemi Cooperativi, CSCW
- Sistemi collaborativi
- Social Networks in general
- Semantics and Social Networks
- Semantic processing
- Suggestions
- Architecture of a Social Network

Sistemi Distribuiti

- Un Sistema distribuito è composto da componenti/ strumenti SW messi in relazione tramite una rete di computer, Che comunicano fra di loro tramite messaggi
  - Messaggi portano: controlli, dati
- Esempi di sistemi distribuiti sono:
  - Internet, intranet, mobile and ubiquitous computing
- Tecnologie gestire la
  - Concorrenza, fra processi distribuiti
  - Sincronizzazione temporale: clock comune, assoluto, precisione
  - Fault (fallimenti) in sistemi distribuiti, architetture fault tolerant
- Sistemi tipicamente eterogenei
  - Diversi per: Sistema operativo, interfaccia di comunicazione, potenza, CPU, etc.

DEVICE	Laptop	PDA	Handset	
NETWORK	WLAN	GSM	GPRS	UMTS
PROTOCOL	SMS	EMS	MMS	I-mode	WAP
LANGUAGE	WML	XML	HTML	
INTERACTION	One way down: Alert, Download	Two ways: Near real time browsing, Real time browsing		
CONSULTATION MODE	Location based	Non-Location based		
SUPPORT	Text	Image	Video	Software	Audio
APPLICATION	Gaming	News	Financial info	Travel	Edutainment
INDUSTRY PROVIDER	Public inst.	Newspapers	....	Software devel.	

Source: Andersen

DS Application areas 1/2

- **Content and resource sharing**
  - Network-wide file/document sharing (e.g. Mangosoft, napster, eDonkey, Gnutella, Freenet)
  - Distributed databases: Mariposa
  - knowledge management (e.g. NextPage)
  - Resource sharing: seti@home, Popular power, mojo natio
  - Cascaded content distribution
  - Edge services
  - P2P search and discovery (e.g. www.fedstats.gov)
  - Network bandwidth sharing
- **Distributed computation (GRID)**
  - Internet-based (e.g. United devices, entropia)
  - Intranet-based (www.datasynapse.com, NetBatch of Intel)
  - Web testing (e.g., United devices)
  - Esempio: gridella, etc….


DS Application areas 2/2

- **collaborations → CSCW (Computer Support Cooperative Work)**
  - On-demand, multi-institutional virtual organizations
  - Marketplace (e.g. www.firstpeer.com)
  - Peer communities of common interests
  - Online development projects (e.g. www.oculustech.com)
  - Online games
  - Remote maintenance
  - Examples: Groovem Buzpad, WuWu
  - E-commerce: ebay, B2B market, etc.
- **Social Networks**
  - **PC and mobiles, CSCW**


Struttura del Seminario

- Sistemi Distribuiti
- Sistemi Cooperativi, CSCW
- Sistemi collaborativi
- Social Networks in general
- Semantics and Social Networks
- Semantic processing
- Suggestions
- Architecture of a Social Network

Concepts of C-S-C-W

- Computer Supported Cooperative Work
- **Computer**: Computer has the potential to improve the technology of cooperative work
- **Supported**: the support is provided by the computer at the cooperative work, new forms of cooperative work
- **Cooperative**: the execution of task, division and organisation of work, ne forms of cooperation
- **Work**: what is cooperative, the task to be executed in cooperative manner

Perché CSCW, pros

- Incremento della produttività
- Riduzione di tempi
  - Tempi di modifica e integrazione dei dai
  - Tempi di convergenza ad un comprensione comune…
- Riduzione dei costi
  - Costo di comunicazione e' minore del costo di viaggio
  - Costo del controllo e monitoraggio e' minore se effettuato sul supporto SW per il CSCW rispetto a chiedere alle persone o analizzare il loro lavoro tramite documenti
- Incremento della qualità
- Piu' divertimento ed interesse, piu' motivazioni
- Crescita culturale e professionale delle persone
  - Soddisfazione, piu' motivazioni

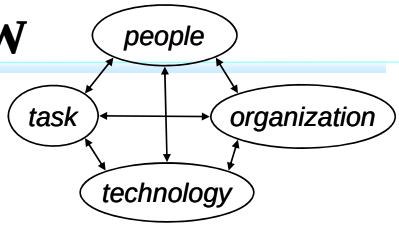


Examples of CSCW Applications

- Email
- NewGroups
- Mailing Lists
- Web Pages
- Common Calendar
- Wiki Portals
- White and life boards
- Virtual/remote meetings
- Workflow tools
- Multiplayer game
- Decision Support Systems
- Chat lines
- Cooperative Editors (real time and for development)
- Distributed database, connected archives, P2P
- Social networks



Discipline of CSCW



- **Analyze for:**
  - **Task/work:**
    - Actions, processes,
    - dependecies, parallelisms
  - **People/users:**
    - How they interact
    - Hierarchy among them
    - user interface
    - Omogenei e non
  - **Organisation/information**
    - Data
    - Flow of data
    - Granularity needed
  - **Technology/tools**
    - Sinc/async, granularity possible, real-time or not
    - Etc.

CSCW, Tipologie di massima

- **Asincrone, Asynchronous**
  - collaborazione non in tempo reale (real-time)
    - Reply, forwarding, distribution list
    - Org by topic, linking
    - Usually text, images, etc.
  - Per esempio:
    - mailing
    - Versioning del testo, integrazione delle versioni, etc.
- **Sincrone, Synchronous**
  - Real-time
  - Tutti vogliono vedere la stessa versione aggiornata allo stesso tempo
  - Editing cooperativo, video conferencing, media spaces, virtual reality, audio conference

CSCW, other Applications

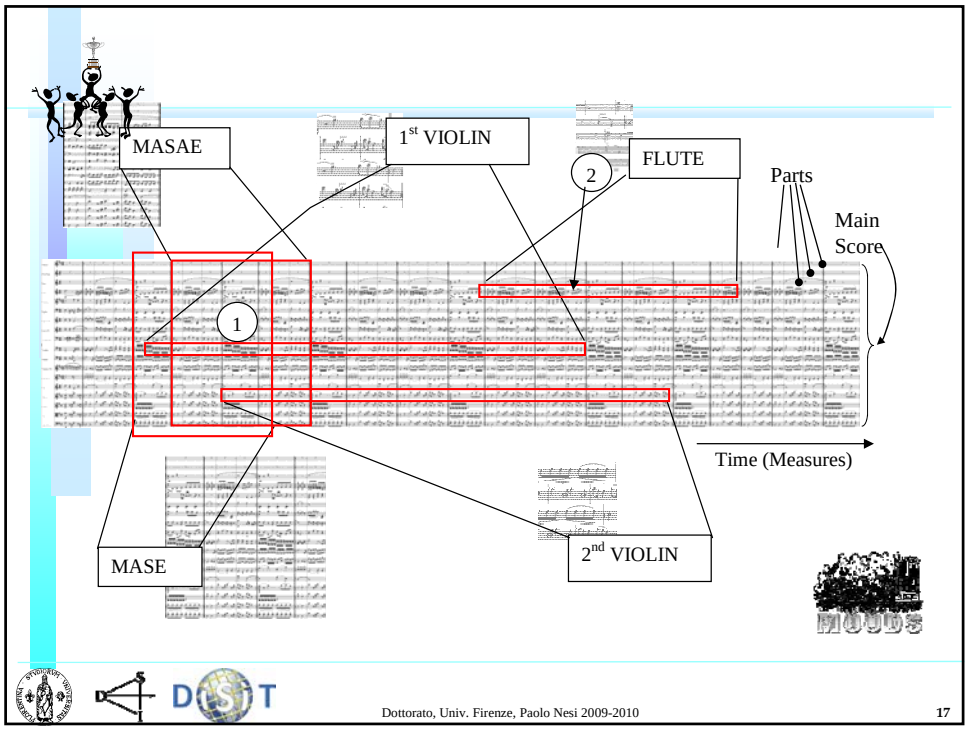
- Multiple-players Games
  - See example on Microsoft XP
  - Sincrono bidirezionale
  - Messaggi real-time, sincroni
  - Discovery di altri potenziali utenti tramite un server centrale
- Decision Support Systems
  - Collaborative environment to produce data for decision and reach a consensus
  - Asincrone e sincrone, n:m, bidirezionale

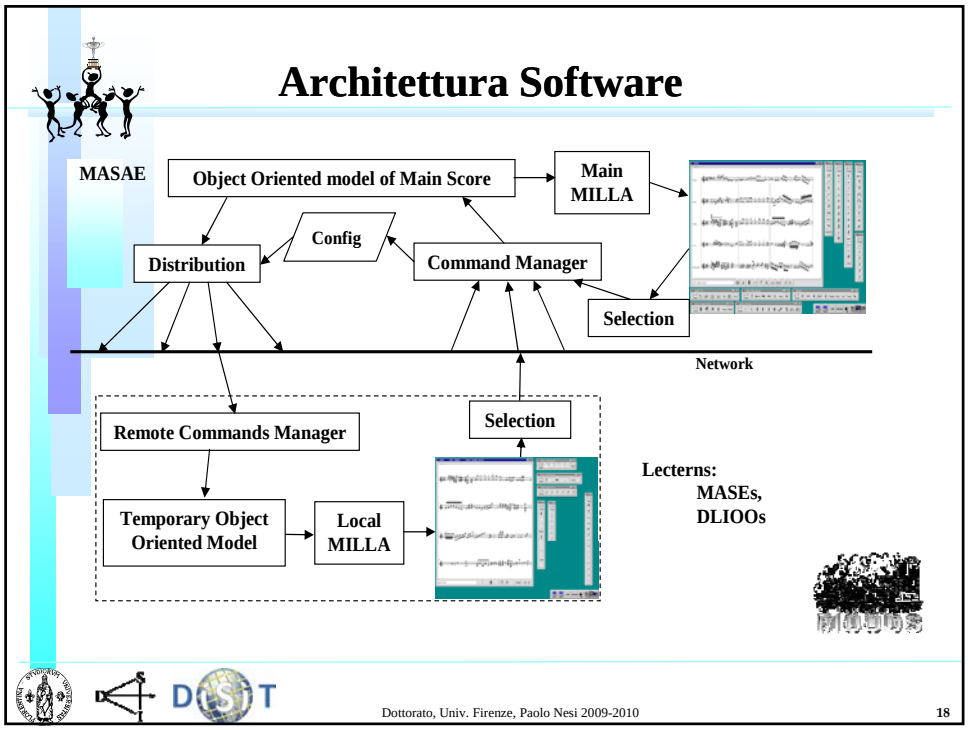
Space and time taxonomy (Borghoff-98)

Space/time	Same time (sync)	Diff time (async) predictable	Diff time (async) Unpredictable
Same place	Face to face meeting, games, class rooms	Shift work	Blackboard, posti it note
Different place (predictable)	Video conference, chat	Email, RCS, netnews	Joint editing of documents
Different place (unpredictable)	Mobile phone conference	Non real time computer conference	Workflow management, letter

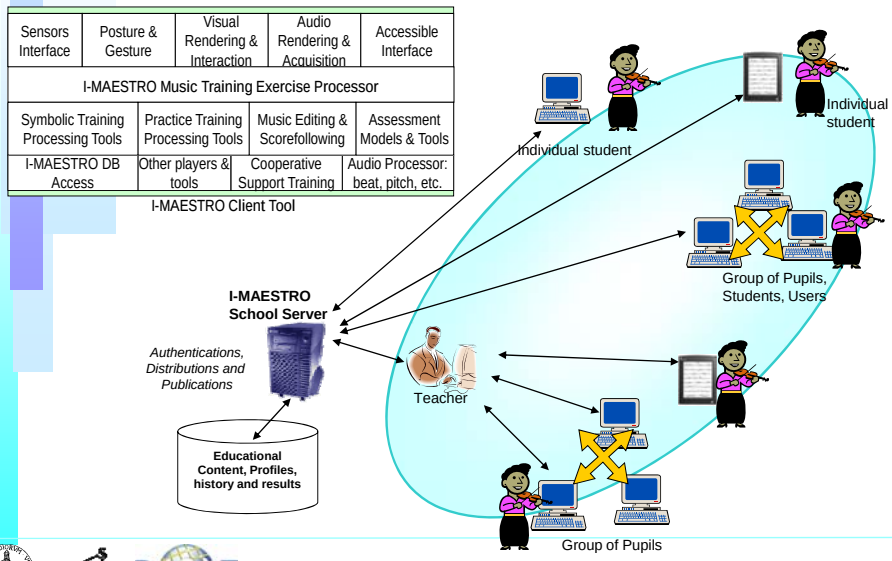
CSCW, User Interface

- Protocollo di Visualizzazione
  - What You See Is What I See
  - What You See Is What I May See
- In caso di conflitto
  - Vince chi arriva prima
  - Vince chi ha la priorità
  - Si cancella il comando a tutti e due
- Controllo accessi
  - Controllo azioni
  - Controllo modifiche
  - Undo delle azioni


MASAE
1st VIOLIN
FLUTE
2
Parts
Main Score
1
Time (Measures)
MASE
2nd VIOLIN
Dottorato, Univ. Firenze, Paolo Nesi 2009-2010
17

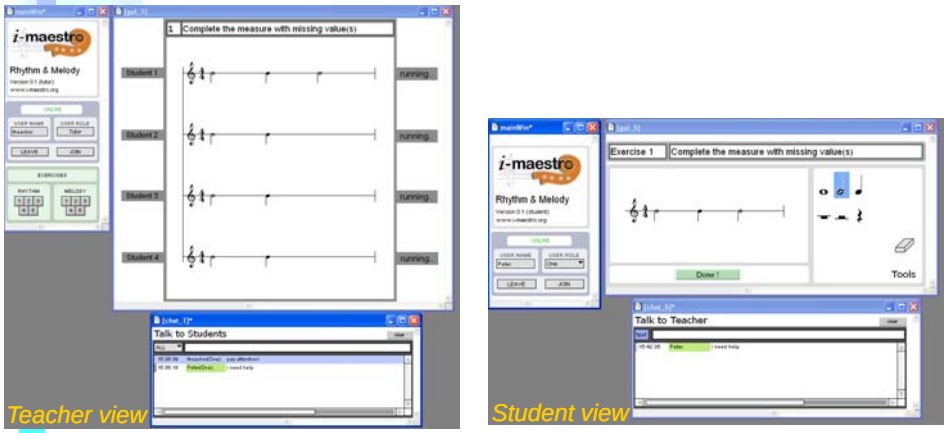

Architettura Software
MASAE
Object Oriented model of Main Score
Main MILLA
Config
Distribution
Command Manager
Selection
Network
Remote Commands Manager
Selection
Temporary Object Oriented Model
Local MILLA
Lecterns:
MASEs,
DLIOOs
Dottorato, Univ. Firenze, Paolo Nesi 2009-2010
18

I-MAESTRO Overview



Max/MSP Cooperative exercise

Example of cooperative exercises for theory training: students have to aswer to a number of questions on music theory



Comparison of Collaborative solutions

	Objectives	Interaction	Observation	Assessment
	Single/ Common	**Live/ recorded**	**Yes/ No**	**Single/ Common**
Competitive	Common	Live	Yes	Common
Collaborative	Single, separate/identical	Live	Yes	Single, separate/identical
Simulative Competitive	Common	Recorded	Yes	Common
Simulative Collaborative	Single	Recorded	Yes	Single, separate/identical

Struttura del Seminario

- Sistemi Distribuiti
- Sistemi Cooperativi, CSCW
- Sistemi collaborativi
- Social Networks in general
- Semantics and Social Networks
- Semantic processing
- Suggestions
- Architecture of a Social Network

Introduction to Social Networks

- With the *users demand* in collaborating and sharing information Social Networks have been created
- **Social Networks** (according to OECD, Organisation for Economic Co-operation and Development) are web portals that **allow users to**:
  - provide and share User Generated Content
  - valorize their creative effort: the content should be originally produced by the users -- e.g., take a picture, compose a set of images, sync. images and audio, etc.
  - users produce content by using non professional solutions and techniques
- Other solutions using UGC are Blogs, Wiki, Forum, etc.


Forrester Trend and Evolution (2009)

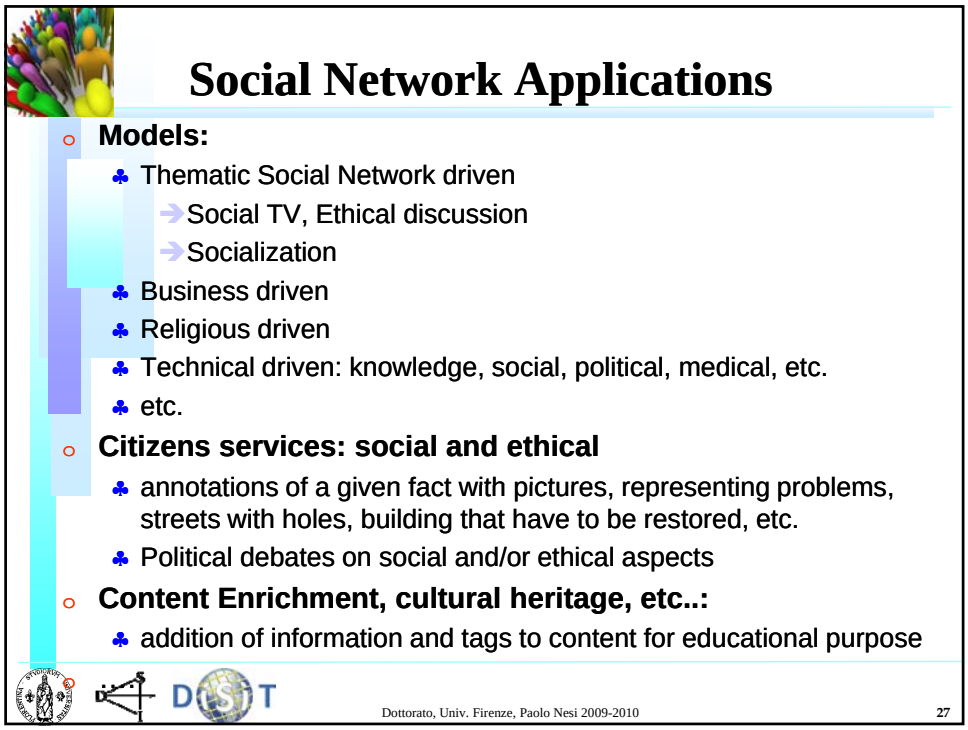
1. Era of Social Relationships:
   - People connect to others and share
   - P2P and present Social networks with UCG
2. Era of Social Functionality:
   - Social networks become like operating system
3. Era of Social Colonization:
   - Every experience can now be social
4. Era of Social Context:
   - Personalized and accurate content
5. Era of Social Commerce:
   - Communities define future products and services

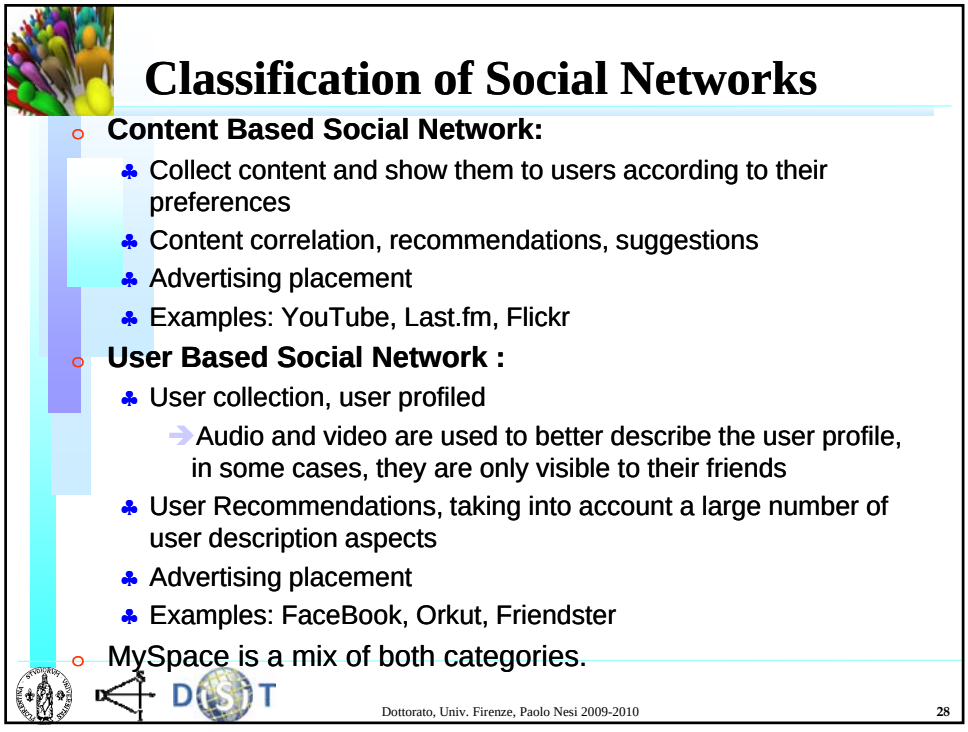

Social Network Motivations

- **Creating Social relationships and contacts**
  - Finding new friends
  - Sharing content with friends
  - Get knowledge about what other people do in their life
- **Increasing Knowledge of users**
  - on specific topics, the subject of the UGC and of the SN
  - on how content can be created and shared
- **Personal advantages for the users**
  - Increasing visibility in the community and in the job
  - Taking the leadership, be observed by a community
- **Save money for the users**
  - Storing user content permanently and making it accessible for its own usage (making it public as side effect)
  - making content public for friends


Social Network Applications

- **Creating a community to provide a service**
  - Objective: share experience, collect/provide knowledge
  - knowledge production (content, comments, annotations, etc.)
  - Collaborative work with users
  - Sharing Improving community knowledge
- **Creating a community to make business on advertising**
  - Objective: increment number of users, minimizing the costs
  - Get Content for placing advertising
  - Stimulating viral propagation
  - Sharing friendship
  - Attracting new users, replacing those that abandon the SN



Social Network Applications
Models:
Thematic Social Network driven
Social TV, Ethical discussion
Socialization
Business driven
Religious driven
Technical driven: knowledge, social, political, medical, etc.
etc.
Citizens services: social and ethical
annotations of a given fact with pictures, representing problems, streets with holes, building that have to be restored, etc.
Political debates on social and/or ethical aspects
Content Enrichment, cultural heritage, etc..:
addition of information and tags to content for educational purpose
DIST
Dottorato, Univ. Firenze, Paolo Nesi 2009-2010
27


Classification of Social Networks
Content Based Social Network:
Collect content and show them to users according to their preferences
Content correlation, recommendations, suggestions
Advertising placement
Examples: YouTube, Last.fm, Flickr
User Based Social Network :
User collection, user profiled
Audio and video are used to better describe the user profile, in some cases, they are only visible to their friends
User Recommendations, taking into account a large number of user description aspects
Advertising placement
Examples: FaceBook, Orkut, Friendster
MySpace is a mix of both categories.
DIST
Dottorato, Univ. Firenze, Paolo Nesi 2009-2010
28

Examples of Content Social Networks

- **Multimedia based Social Network:**
  - Flick for images
  - YouTube for video
  - Imeem for audio
- **Entertainment based Social Network:**
  - Second Life for 3D
  - Online gaming…
- **News/opinions based Social Network:**
  - Digg, Reddit for social news
  - Yelp for reviews
- **Fast Communications and social services**
  - Twitter a sort of microblog
  - Linkedin, Myspace, facebook, etc…

Content Searching in Social Network

- Traditional Classification based on Metadata
- Free Tags, such as Folksonomy
- Geotagging, GPS data
- votes

User Generated Content, UGC

- **Conditions that Facilitated the grown of UGC**
  - Reduced costs for equipments which allow the personal content production: cameras, smart phones, etc.
  - Reduced costs of connection, increment of broadband diffusion
  - More Web Interactive capabilities: Ajax, JSP
  - Creative Commons Licensing/formalisms, increment of confidence
- **Pros and Facilitations**
  - Growing of WEB sites that host your content and provide some tools to make them accessible on web for your friends
  - Natural selection/emergence of better UGC items, increment of visibility for some of UGC users…
  - Annotation and reuse of UGC of others users and friends


User Generated Content Cons 1/2

- **Cons and problems (1/2)**
  - Restricted social penetration since only User with are ICT skilled and have a certain economical capability may access to internet and spend time to ejoy SN
  - Lack of formal Privacy control
  - IPR problems
  - Lack of interoperability for users and content among different social networks
  - Content is not completely defined in terns of Metadata
  - Competitions of UGC against professional content, producers are against their support and diffusion
  - Growing costs for the SN providers


Sn: User Classification

- **Lurkers: passive users,**
  - take and do not contribute: no content, no other users, ….
  - can be even frequent users to read
  - they are tpyically invited and does not invite
- **Occasional users:**
  - sometimes they also contribute with UGC
  - marginal active in terms of invitations
- **Active users:**
  - frequently contribute
  - The first source of inviations of users and content
- **Pushers:**
  - Typically active users paid to stimulate activities with content, discussions, users, mailing, etc.

User Activities on Social Networks

- **Wikipedia (2006)**
  - 68000: active users
  - 32 millions of lurkers
  - While the 1000 more active users produced the 66% of changes.
- **Similar numbers in other portals:**
  - 90% lurkers
  - 9% occasional users
  - 1% active users
  - 90% is produced by the 1% of active users
  - 10% is generated by the 9% of users including the occasional

Centrality of User Profile

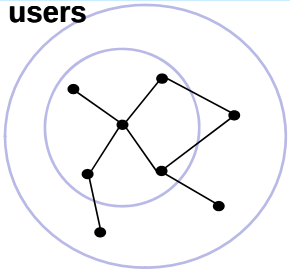
- **Static**
  - generically provided during registration.
  - frequently not so much detailed in generic Social Networks, since users prefer to avoid filling in 'useless' forms and/or to provide false data.
  - In small thematic and business oriented Social Networks the information is much more reliable.
- **Dynamic**
  - collected on the basis of the activities users perform on the portal elements,
  - such as those on content, on other users:
  - changed by users, Inferred by relationships

Profilo degli utenti

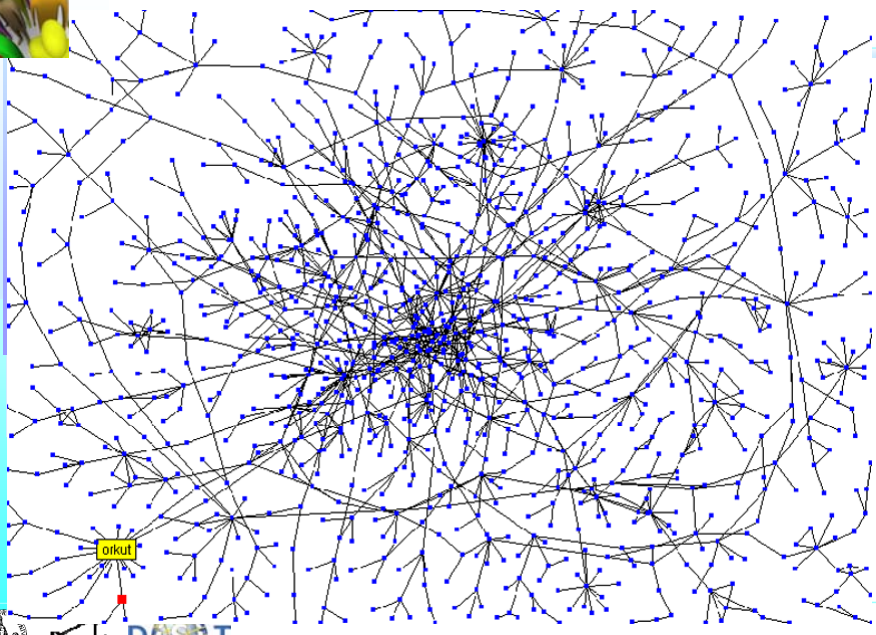
- *Informazioni statiche:*
  - *Informazioni generali:*
    - *nome, cognome, sesso,*
    - *foto, data di nascita,*
    - *descrizione personale,*
    - *località di provenienza (ISO 3166),*
      - *Nazione*
      - *Suddivisione*
      - *Provincia*
    - *lingue parlate (ISO 369)*
  - *Informazioni di contatto:*
    - *lista di contatti di instant messaging*
  - *Scuola e Lavoro:*
    - *scelta del livello scolastico,*
    - *nome della scuola,*
    - *tipo di lavoro,*
    - *nome del posto di lavoro*
  - *Interessi:*
    - *Vettore contenente la lista di valori del campo*
- *Informazioni dinamiche:*
  - *Lista di oggetti preferiti*
  - *Lista di amici*
  - *Lista gruppi*
  - *Voti positivi ad oggetti*
  - *Commenti ad oggetti*
  - *…*
  - *…*
  - *Informazioni sulle preferenze sulla base delle visualizzazioni degli oggetti*
    - *Format*
    - *Type*
    - *Taxonomy*

Relevance of Users

- **Number of Connections with other users**
  - Direct connections,
  - Second and third level connections,
  - Etc.
- **Number of accesses to their**
  - profile page (if any)
  - posted and/or preferred content
  - Comments
  - groups
- **Users' Activities**
  - Number of posted content in time
  - Number of posted comments, on content, on area…
  - Number of votes per content, per area, etc.
  - Number of accesses to the network

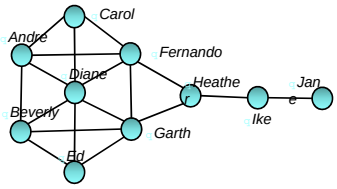


Stanford Social Web



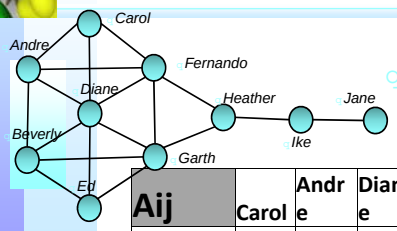
Social Network Analysis Metrics

- **Degree of Centrality** of a node
  - The number of connections to a certain node
  - Diane has 6 connections
  - Deg: can be non symmetric if the relationships are not symmetric, thus the graph is oriented.



  - Diane is connected to others which are in turn connected each other.
  - It is not true that to have many connections is the best model to identify the relevance of a certain node.
    - In this case Diana is connected to people that are in any case connected each other.
    - While Heather is central to keep Ike and Jane connected

Matrix of connections

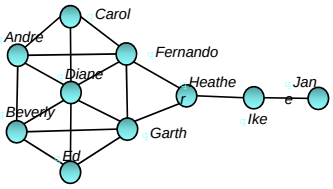


A[i][j]: matrix of connections

Aij	Carol	Andre	Diane	Fernando	Beverly	Ed	Garth	Heather	Ike	Jane	
Carol	0	1	1	1	0	0	0	0	0	0	3
Andre	1	0	1	1	1	0	0	0	0	0	4
Diane	1	1	0	1	1	1	1	0	0	0	6
Fernando	1	1	1	0	0	0	1	1	0	0	5
Beverly	0	1	1	0	0	1	1	0	0	0	4
Ed	0	0	1	0	1	0	1	0	0	0	3
Garth	0	0	1	1	1	1	0	1	0	0	5
Heather	0	0	0	1	0	0	1	0	1	0	3
Ike	0	0	0	0	0	0	0	1	0	1	2
Jane	0	0	0	0	0	0	0	0	1	0	1
	3	4	6	5	4	3	5	3	2	1	

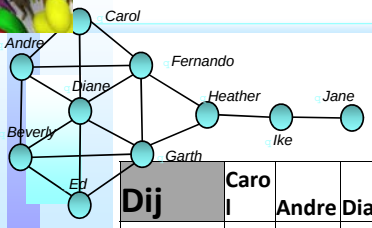
Averaged Number of connections

- Total number of connections divided for the number of Nodes
- According to the examples above:
  - Number of connections: 36
    - ➔ they are considered non bidirectional otherwise they should be 18
  - Number of nodes: 10
- Averaged number of connections:
  - 36/10, 3.6 connections per node, or
  - 18/10, 1.8 connections per node



- It is more similar to the user perception to say 3.6 connections rather then 1.8

Matrix of distances

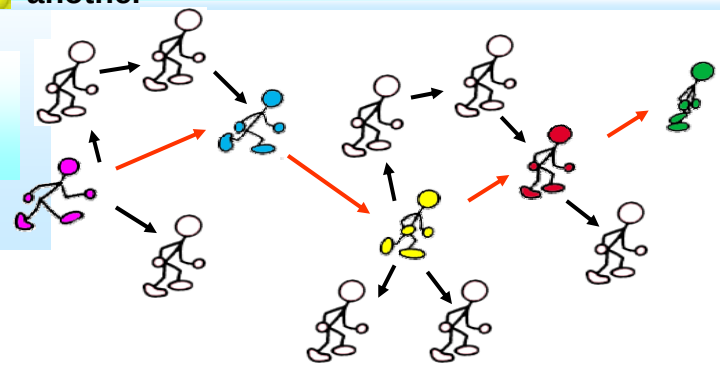


D[i][j]: matrix of distances

N(N-1)/2 elements*

Dij	Carol	Andre	Diane	Fernando	Beverly	Ed	Garth	Heather	Ike	Jane	
Carol	0	1	1	1	2	2	2	2	3	4	18
Andre		0	1	1	1	2	2	2	3	4	16
Diane			0	1	1	1	1	2	3	4	13
Fernando				0	2	2	1	1	2	3	11
Beverly					0	1	1	2	3	4	11
Ed						0	1	2	3	4	10
Garth							0	1	2	3	6
Heather								0	1	2	3
Ike									0	1	1
Jane										0	0
											89

- **Averaged shortest path from one person to another**

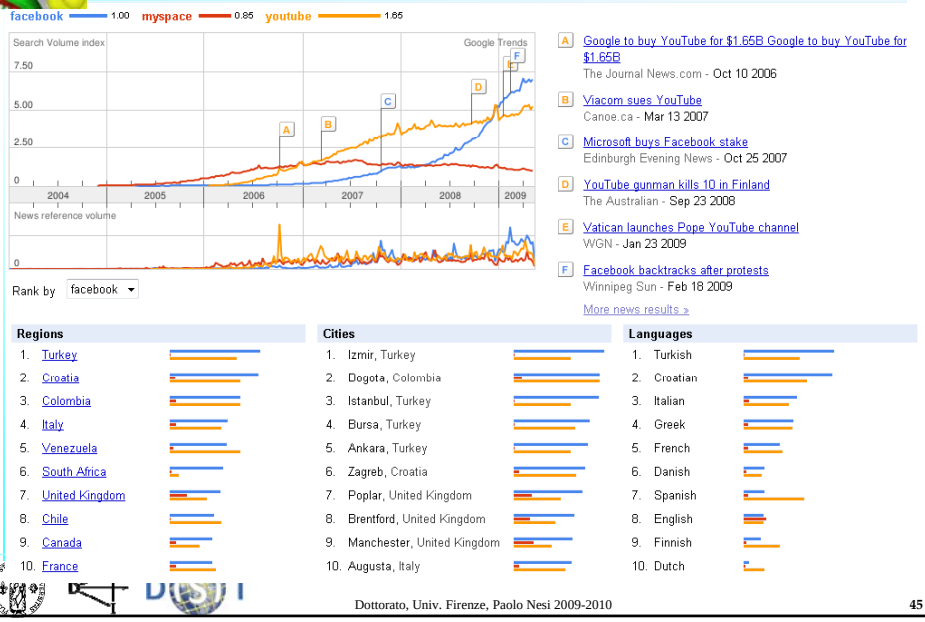


- *MIT: 6.4 hops*
- *Stanford: 9.2 hops*
- *Our example: 1.97 hops*
  - *Sum of shortest paths: 89*
  - *10 Nodes*
  - *45 possible connections*

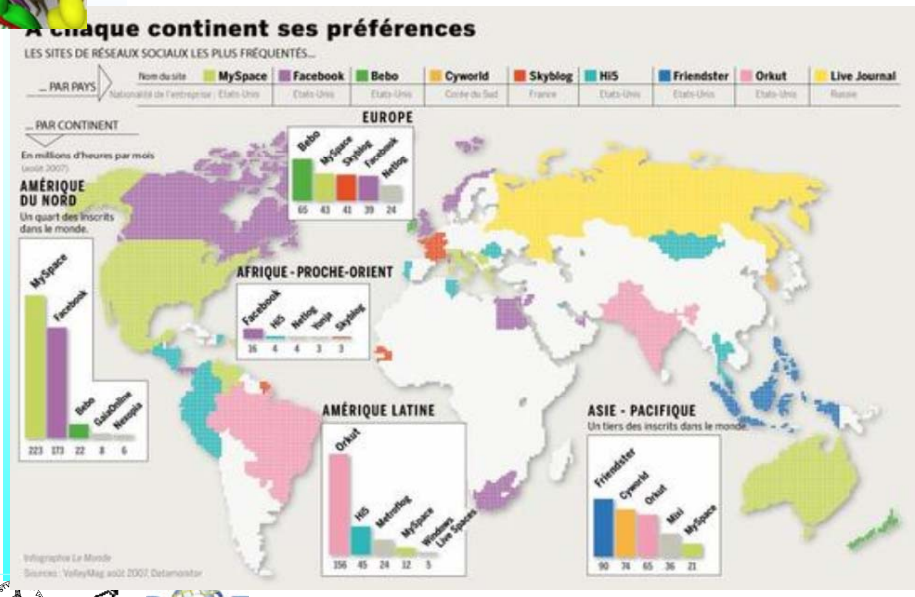
Business Models of UGC/SN

- **Advertising**
  - Publication of ads (banners) on the Social Network
  - A value proportional to the number of users
- **Targeted Advertising**
  - Placement of ads on the basis of Users, context, content, etc.
  - Cost per Click, cost per impression, etc.
- **Donations**
  - See Wikipedia
- **Pay per Item**
  - A small price for each item
  - A license for each item, DRM, CAS
- **Subscription**
  - A monthly subscription to have more power, see Second Life
- **Selling of Services**
  - Linkedin, Second Life, etc.

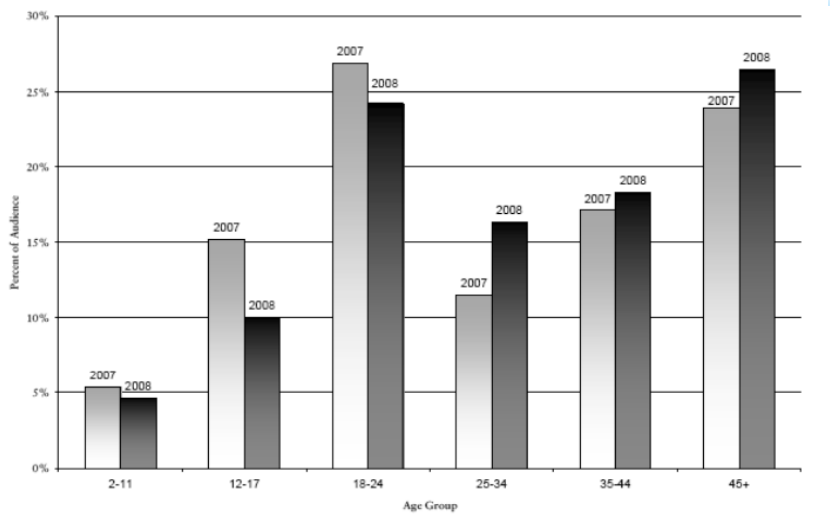
Trend of Facebook, MySpace and YouTube



General Distribution of SN



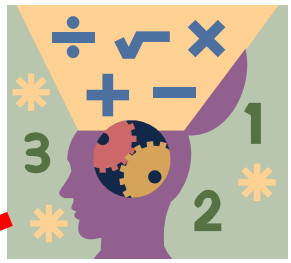
FACEBOOK AUDIENCE COMPOSITION JUNE 2008 VERSUS JUNE 2007

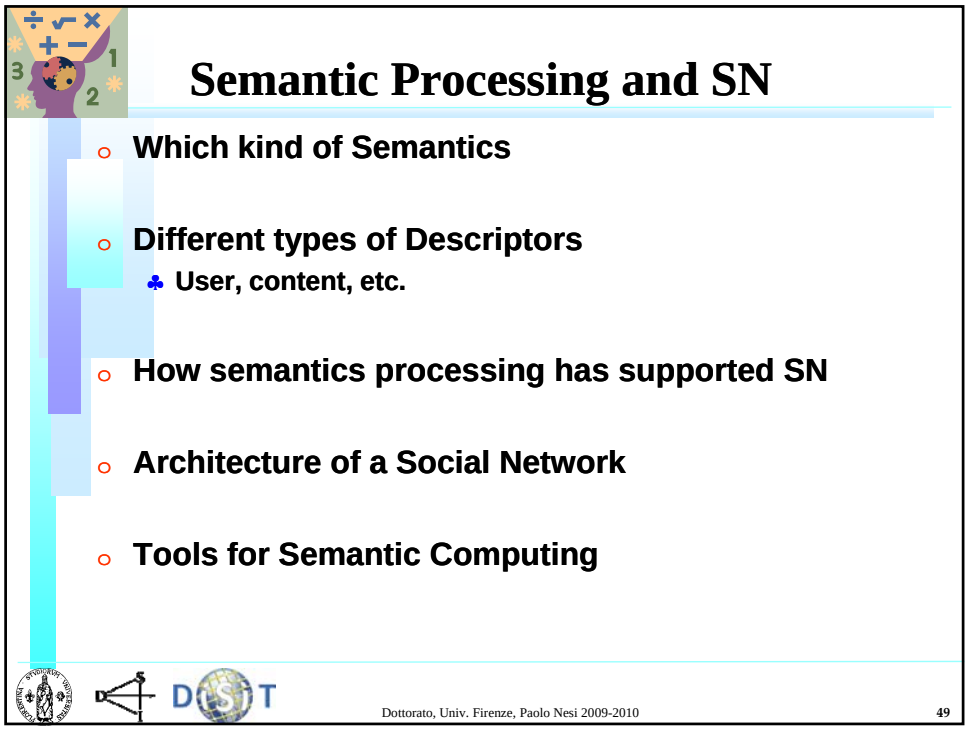


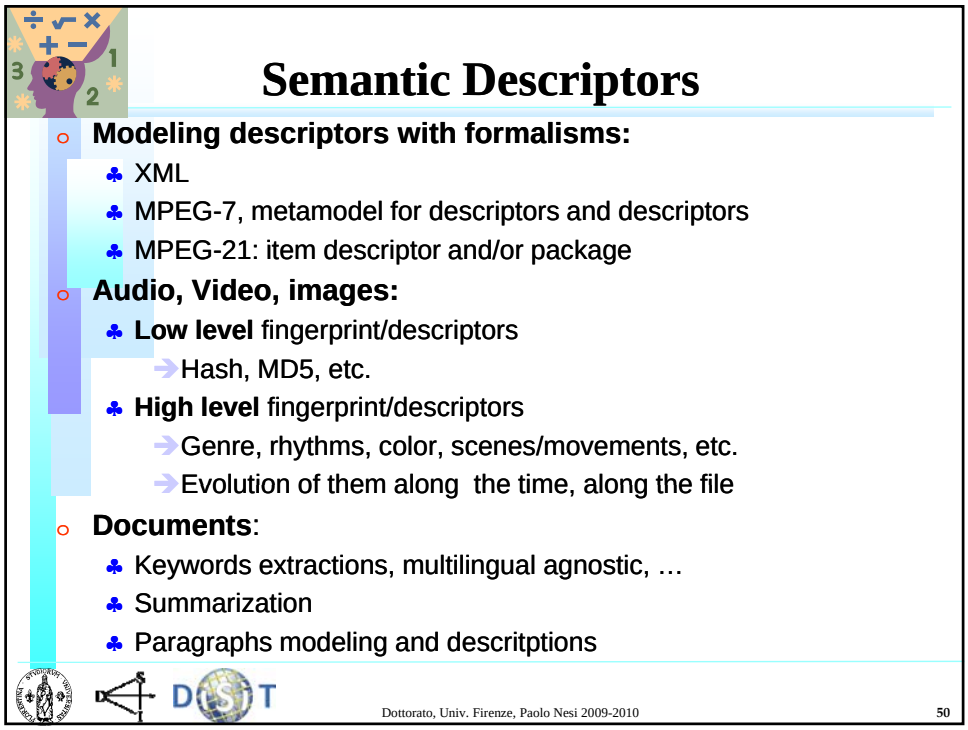
Source: comScore, Inc.

Struttura del Seminario

- Sistemi Distribuiti
- Sistemi Cooperativi, CSCW
- Sistemi collaborativi
- Social Networks in general
- Semantics and Social Networks
- Semantic processing
- Suggestions
- Architecture of a Social Network




Semantic Processing and SN
Which kind of Semantics
Different types of Descriptors
User, content, etc.
How semantics processing has supported SN
Architecture of a Social Network
Tools for Semantic Computing
DIST
Dottorato, Univ. Firenze, Paolo Nesi 2009-2010
49


Semantic Descriptors
Modeling descriptors with formalisms:
XML
MPEG-7, metamodel for descriptors and descriptors
MPEG-21: item descriptor and/or package
Audio, Video, images:
Low level fingerprint/descriptors
Hash, MD5, etc.
High level fingerprint/descriptors
Genre, rhythms, color, scenes/movements, etc.
Evolution of them along the time, along the file
Documents:
Keywords extractions, multilingual agnostic, …
Summarization
Paragraphs modeling and descritptions
DIST
Dottorato, Univ. Firenze, Paolo Nesi 2009-2010
50

Semantic Descriptors and info 1/2

- **user profile descriptions** collected via user registration and dynamically on the basis of user actions, migrated also on the mobile;
- **content descriptors** for simple and complex content, web pages, forums, etc.;
- **user groups descriptors** and their related discussion forums and web pages (with taxonomic descriptors and text);
- **relationships among users/colleagues** (similarly to friendships, group joining) that impact on the user profile and are created via registration, by inviting colleagues, performing registration to groups, etc.;
- **votes and comments on contents, forums, web pages**, etc., which are dynamic information related to users;


Semantic Descriptors and info 2/2

- **downloads and play/executions** of simple and/or complex content on PC and mobiles, to keep trace of user actions as references to played content, which are dynamic information related to users preferences;
- **lists of elements marked as preferred by users,** which are dynamic information related to users;
- **uploads and publishing** of user provided content on the portal (only for registered users, and supervised by the administrator of the group). Each Content element has its own static metadata, descriptors and taxonomy; while the related action of upload is a dynamic information associated with the User who performed it. In addition, Content elements can be associated with Groups.


Usage/Prod of Semantic Information

- **content ingestion.** semantic tagging while technical descriptors about digital resources are added during the automated adaptation and icon production;
- **repurposing and publication for several kinds of end-user devices**
- **extraction of semantic technical descriptors** from simple and complex essences,
- **content indexing** to prepare and accelerate the process of search.
- **packaging content and semantics into MPEG-21/AXMEDIS** binary format: integrating digital essences with metadata and descriptors
- **exporting content** to other databases, or posting them on other social networks or portals, publishing on P2P networks
- **estimating similarities among users, objects/content,** to pose the basis of generating suggestions and reasoning;
- **producing suggestions** about potential colleagues, interesting content, and groups;

Content Descriptors

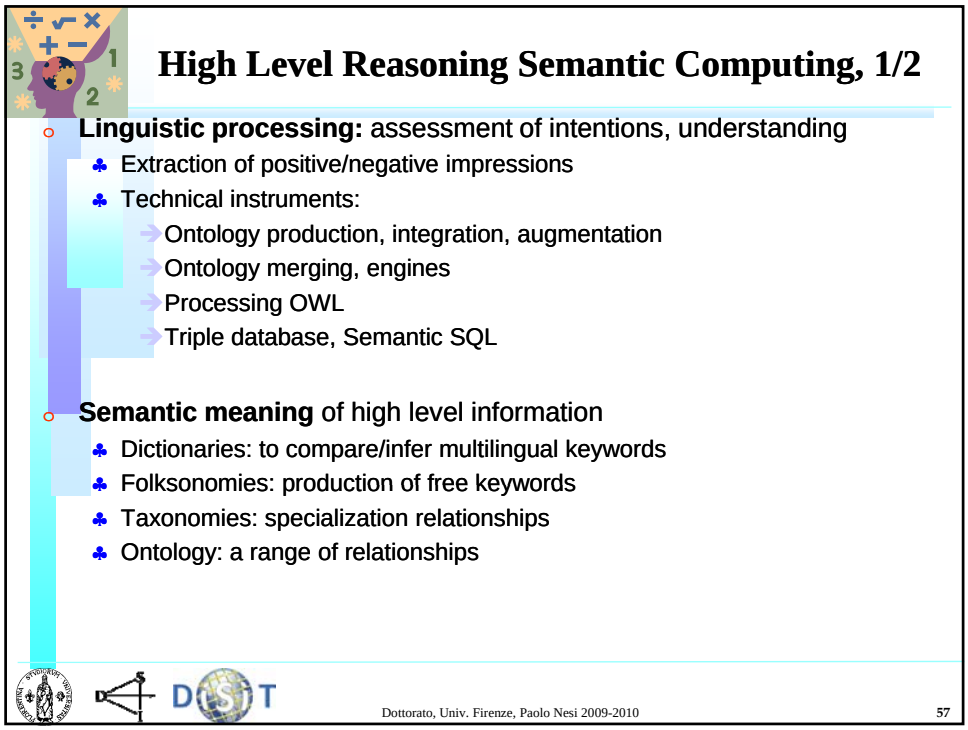
- **Static aspects** : more relevant since the content description is typically not changing over time. They are:
  - metadata, keywords extracted from description, comments, etc.;
  - technical description (as the Format in the following): audio, video, document, cross media, image,..;
  - content semantic descriptors such as: rhythm, color, etc.; genre, called Type in the following;
  - groups to which the content has been associated with;
  - taxonomies classification to which the content has been associated, taking into account also the general taxonomy;
- **dynamic aspects** are marginally changed and may be related to:
  - user's votes, user's comments;
  - number of votes, comments, download, direct recommendations, etc.;

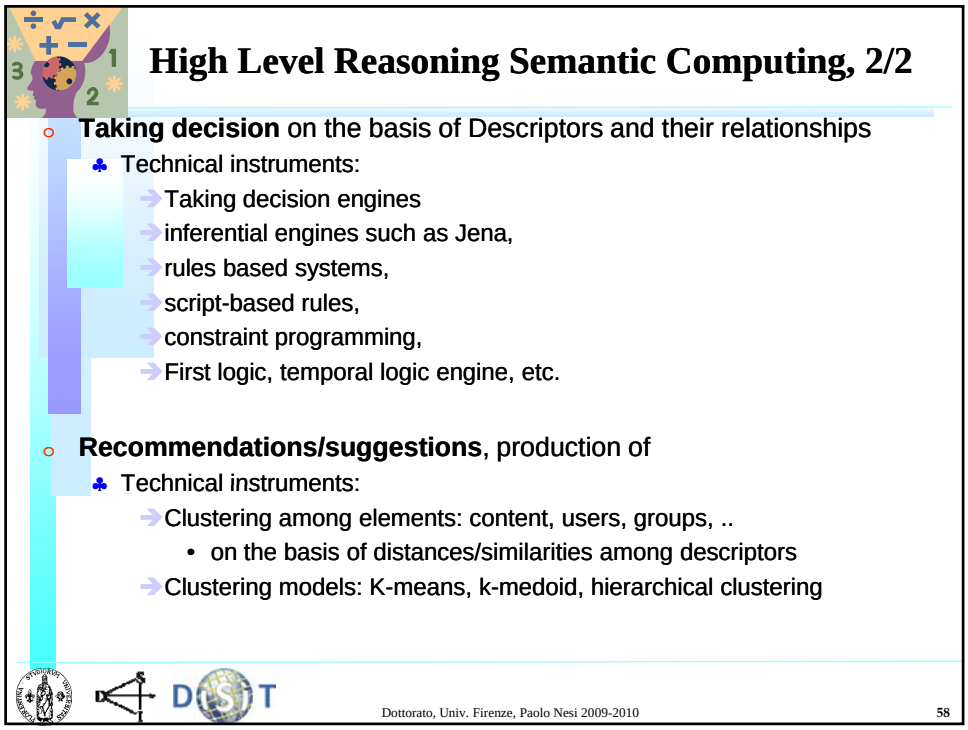
Group Descriptors

- **Groups** of users they may have specific descriptors and those inherited by the users:
- **static** aspects of the groups such as:
  - objectives, topics, web pages, keywords, taxonomy, etc.;
- **dynamic** aspects related to:
  - users belonging to the group; users may: join and leave the group, be more or less active over time;
  - content associated with the group: files, comments, etc., with their taxonomical classification, metadata and descriptors.

Struttura del Seminario

- Sistemi Distribuiti
- Sistemi Cooperativi, CSCW
- Sistemi collaborativi
- Social Networks in general
- Semantics and Social Networks
- Semantic processing
- Suggestions
- Architecture of a Social Network


High Level Reasoning Semantic Computing, 1/2
Linguistic processing: assessment of intentions, understanding
Extraction of positive/negative impressions
Technical instruments:
Ontology production, integration, augmentation
Ontology merging, engines
Processing OWL
Triple database, Semantic SQL
Semantic meaning of high level information
Dictionaries: to compare/infer multilingual keywords
Folksonomies: production of free keywords
Taxonomies: specialization relationships
Ontology: a range of relationships
Dottorato, Univ. Firenze, Paolo Nesi 2009-2010
57


High Level Reasoning Semantic Computing, 2/2
Taking decision on the basis of Descriptors and their relationships
Technical instruments:
Taking decision engines
inferential engines such as Jena,
rules based systems,
script-based rules,
constraint programming,
First logic, temporal logic engine, etc.
Recommendations/suggestions, production of
Technical instruments:
Clustering among elements: content, users, groups, ..
on the basis of distances/similarities among descriptors
Clustering models: K-means, k-medoid, hierarchical clustering
Dottorato, Univ. Firenze, Paolo Nesi 2009-2010
58

Recommendations

- **They are a means for the**
  - Usage of content/object info to find/propose users
  - Usage of users info to find/propose content
  - Usage of users info to find/propose other users
  - Etc..
- **Different Recommendations/Suggestions**
  - U → U: a user to another user on the basis of his profile
  - O → U: an object at a user on the basis of his profile
  - O → O: an object on the basis of a played object of a user
  - G → U: a group to a user
  - Etc…
- **Objects can be Advertising, Ads, Content, Events, Groups, etc.….**

Different Recommendations

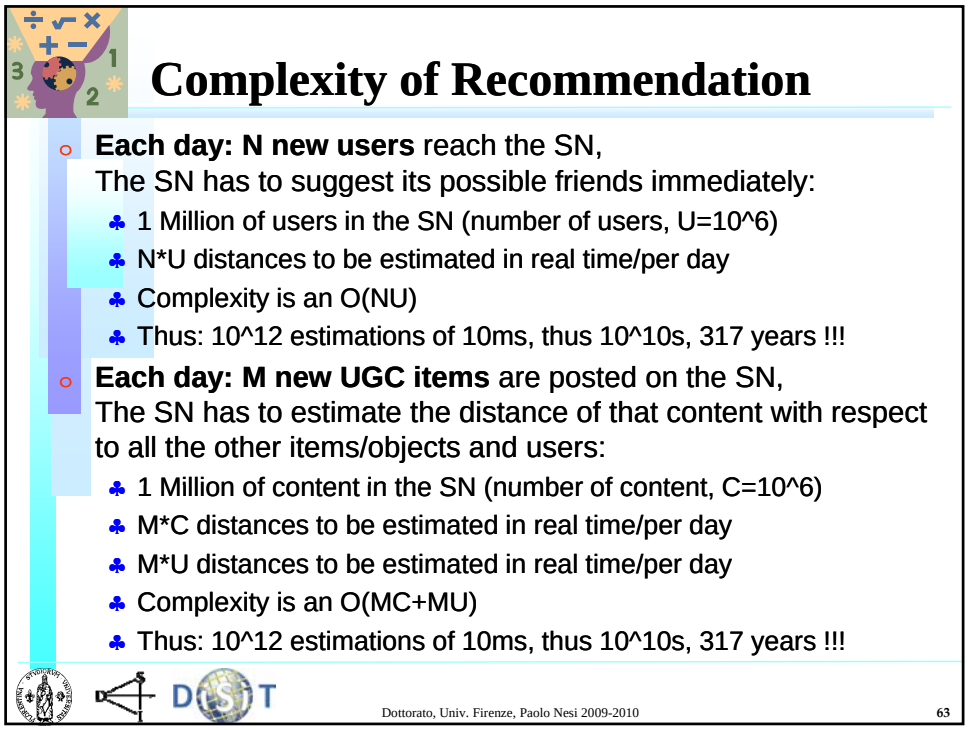
- **FOR YOU: Suggested objects/contents/events/groups since they**
  - are the less, most viewed, most played, most played in your group, ..
  - are similar to your highest voted/ranked objects
  - are similar to what you usually play, pay, print, upload, etc.
    - The most played/../voted in absolute
    - The most played/../voted in the last Month/Day, week, etc…
    - The most played/../voted in your area, country, group, etc..
  - are new for the SN
  - belongs to the preferred of your friends, …
  - have been posted/commented by your friends, in your group, …
  - have been recommended by a your friend
- **FOR BUSINESS: Suggested objects/…./groups since they**
  - are new for the SN, and thus are new for the market/business of the SN
  - are commercially proposed and have to be commercially promoted for the business of the SN
  - belong on the log tail of the content distribution/usage

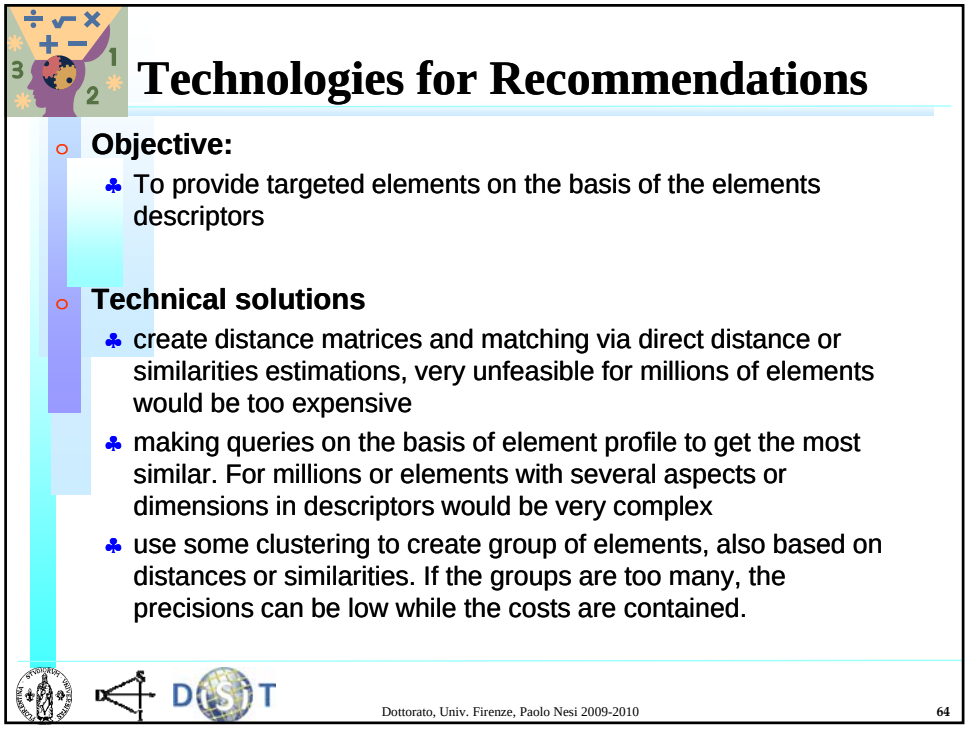
Recommendations

		Recipient of the suggestions		
		User	**Content (played by a user)**	**Group (leader or members)**
Suggested elements	**Users**	Proposing to a user possible colleagues / friends	--no sense--	Proposing at a group responsible possible interested colleagues to be invited
	Contents	Proposing to a user possible interesting contents	Proposing at a play of a content similar content items	Proposing at a group members possible interesting content (not much different with respect to C-C combination)
	Groups	Proposing to a user possible interesting groups	Proposing at a play of a content possible interesting groups in which similar contents are discussed	--no sense--
	Ads	Proposing to a user possible interesting ads	Proposing at a play of a content the possible interesting ads	Proposing at a/all group member/s possible interesting ads

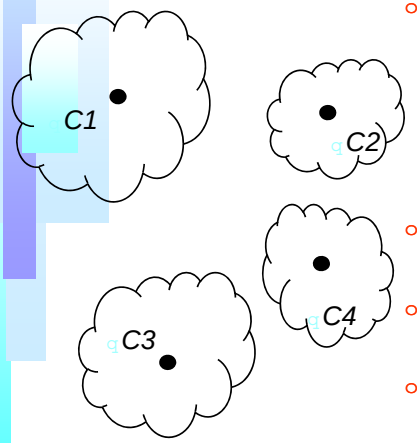
Why to recommend

- **The SN owners aim to:**
  - reduce the number of queries to reduce costs
  - push for the long tail content to increment of revenues
  - stimulate the socialization to have more connections
  - get more users,
  - get more value for advertising
- **To create more connected SNs**
  - More cohesion leads to have more resistance to close
  - More connections means more solidity and activity
  - Etc.


Complexity of Recommendation
Each day: N new users reach the SN,
The SN has to suggest its possible friends immediately:
1 Million of users in the SN (number of users, U=10^6)
N*U distances to be estimated in real time/per day
Complexity is an O(NU)
Thus: 10^12 estimations of 10ms, thus 10^10s, 317 years !!!
Each day: M new UGC items are posted on the SN,
The SN has to estimate the distance of that content with respect to all the other items/objects and users:
1 Million of content in the SN (number of content, C=10^6)
M*C distances to be estimated in real time/per day
M*U distances to be estimated in real time/per day
Complexity is an O(MC+MU)
Thus: 10^12 estimations of 10ms, thus 10^10s, 317 years !!!
Dottorato, Univ. Firenze, Paolo Nesi 2009-2010
63

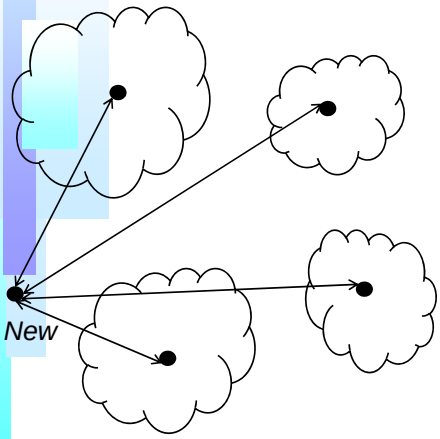

Technologies for Recommendations
Objective:
To provide targeted elements on the basis of the elements descriptors
Technical solutions
create distance matrices and matching via direct distance or similarities estimations, very unfeasible for millions of elements would be too expensive
making queries on the basis of element profile to get the most similar. For millions or elements with several aspects or dimensions in descriptors would be very complex
use some clustering to create group of elements, also based on distances or similarities. If the groups are too many, the precisions can be low while the costs are contained.
Dottorato, Univ. Firenze, Paolo Nesi 2009-2010
64

Clustering among descriptors



- *K-Means* clustering
  - Based on a multidimensional distance model among each other
  - Define the number of clusters
  - Estimation process to maximize the cohesion among clusters
- Some items can be spare
  - They are classified in any case
- Millions of content items, thousands of clusters, …
- Periodic re-clustering taking into account all the content/objects/users

Clustering among descriptors

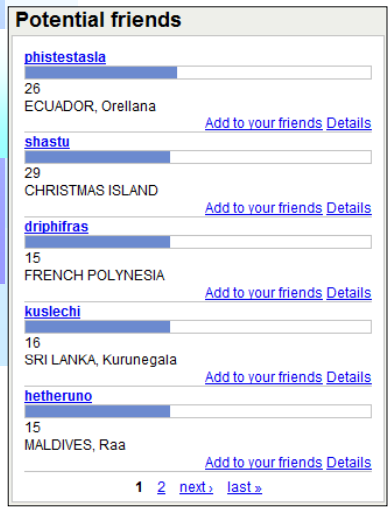


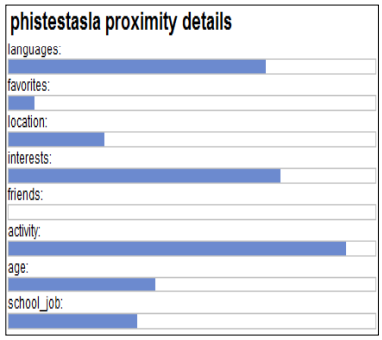
- Millions of content items,
- ONLY thousands of clusters
- At each New Object
  - Distance of the new object with respect to cluster Centers
  - Reduction of complexity
- Usable on recommendations:
  - UU, UO, OO, etc.

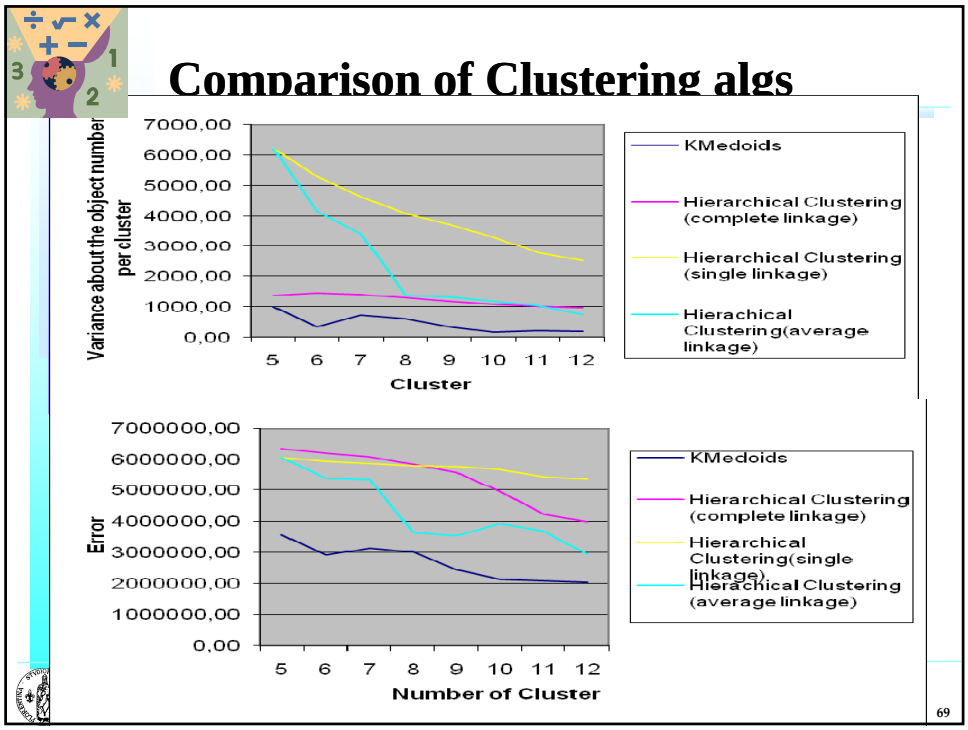
Similarity Distances

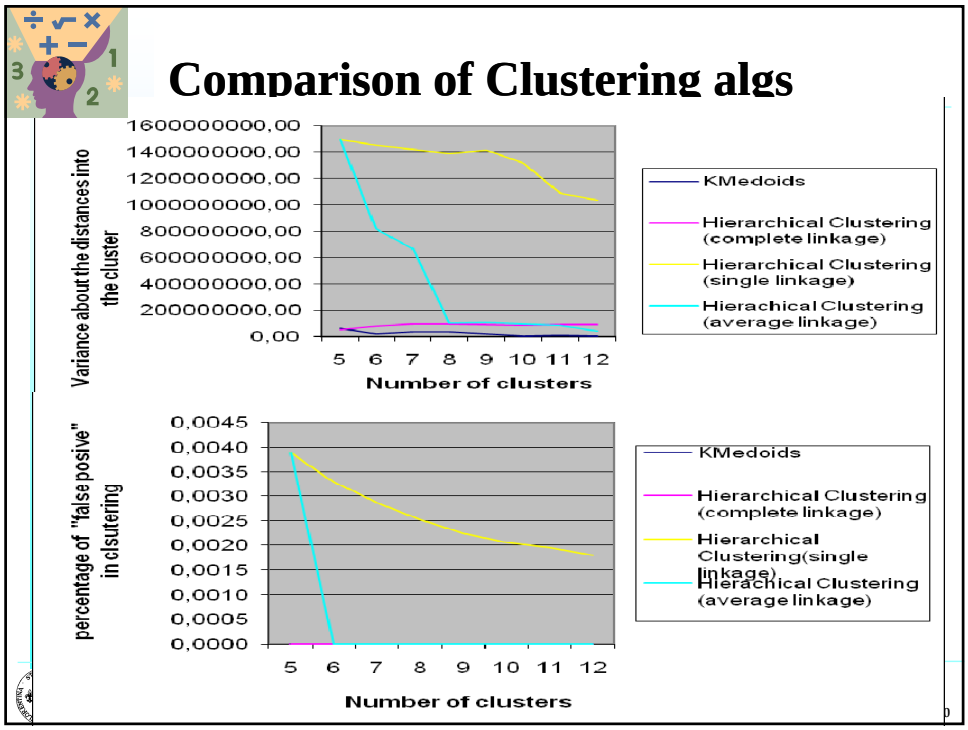
Suggested	Recipient of the suggestions: User	Content (played by a user)	Group (leader or members)
Users	D(U(s,d);U(s,d))	**--no sense--**	D(U(s,d);G(s,d))
Contents	D(C(s);U(s,d))	D(C(s);C(s))	D(C(s);G(s,d))
Groups	D(G(s,d);U(s,d))	D(G(s,d);C(s))	**--no sense--**

Visualizzazione di Suggerimenti e dist





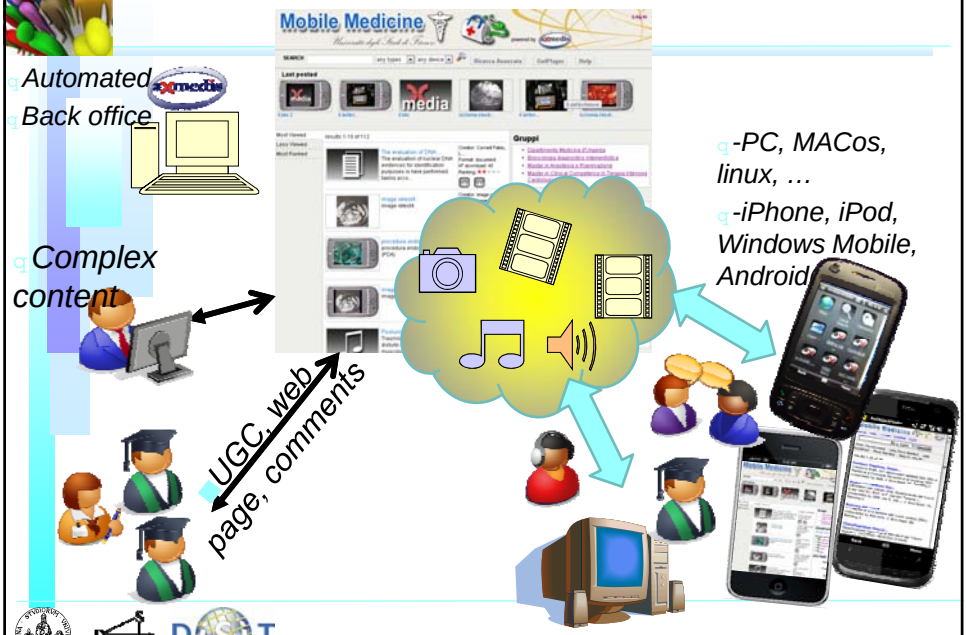

Comparison of Clustering algs
Variance about the object number per cluster
7000,00
6000,00
5000,00
4000,00
3000,00
2000,00
1000,00
0,00
5 6 7 8 9 10 11 12
Cluster
KMedoids
Hierarchical Clustering (complete linkage)
Hierarchical Clustering (single linkage)
Hierachical Clustering(average linkage)
Error
7000000,00
6000000,00
5000000,00
4000000,00
3000000,00
2000000,00
1000000,00
0,00
5 6 7 8 9 10 11 12
Number of Cluster
KMedoids
Hierarchical Clustering (complete linkage)
Hierarchical Clustering(single linkage)
Hierachical Clustering (average linkage)
69


Comparison of Clustering algs
Variance about the distances into the cluster
1600000000,00
1400000000,00
1200000000,00
1000000000,00
800000000,00
600000000,00
400000000,00
200000000,00
0,00
5 6 7 8 9 10 11 12
Number of clusters
KMedoids
Hierarchical Clustering (complete linkage)
Hierarchical Clustering (single linkage)
Hierachical Clustering (average linkage)
percentage of "false posive" in clsutering
0,0045
0,0040
0,0035
0,0030
0,0025
0,0020
0,0015
0,0010
0,0005
0,0000
5 6 7 8 9 10 11 12
Number of clusters
KMedoids
Hierarchical Clustering (complete linkage)
Hierarchical Clustering(single linkage)
Hierachical Clustering (average linkage)

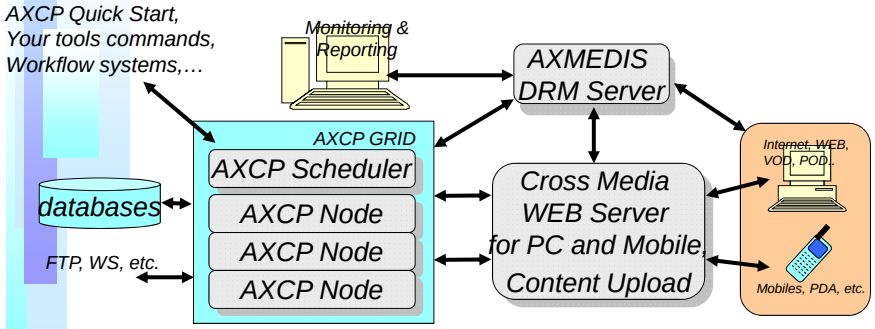
Struttura del Seminario

- Sistemi Distribuiti
- Sistemi Cooperativi, CSCW
- Sistemi collaborativi
- Social Networks in general
- Semantics and Social Networks
- Semantic processing
- Suggestions
- Architecture of a Social Network

Mobile Medicine

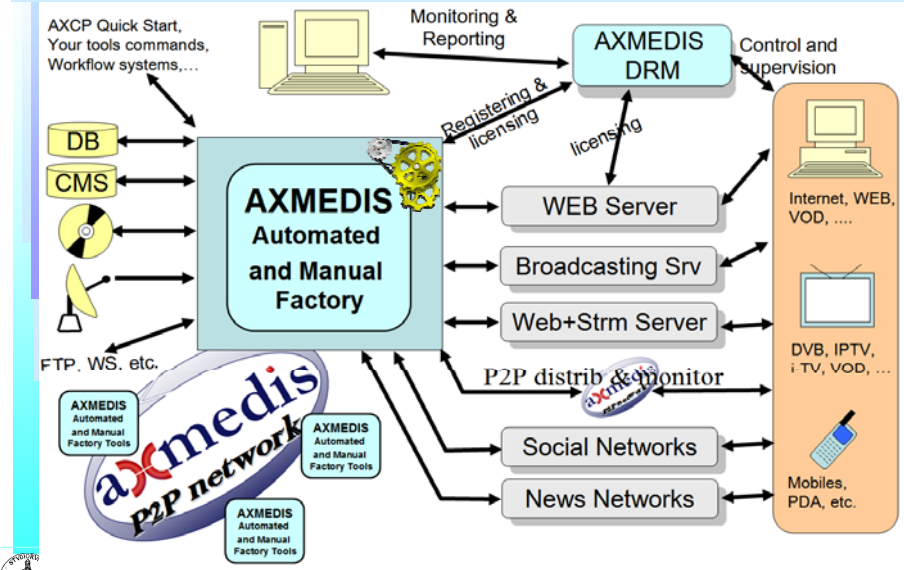


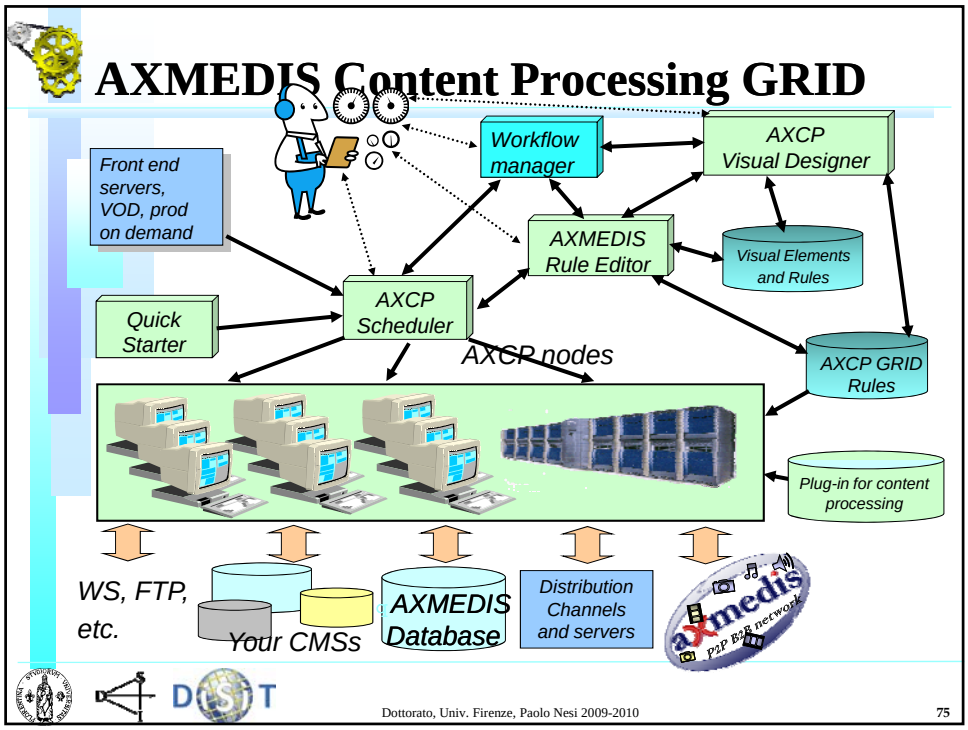
Mobile Medicine



- Cross Media distribution portal for multichannel:
  - PC, PDA, iPhone, iPod, mobile, etc.
- Production tools and players, for PC and PDA
- AXMEDIS AXCP GRID backoffice server: **semantic computing**
- AXMEDIS DRM: for rights control and security

Factory and integration





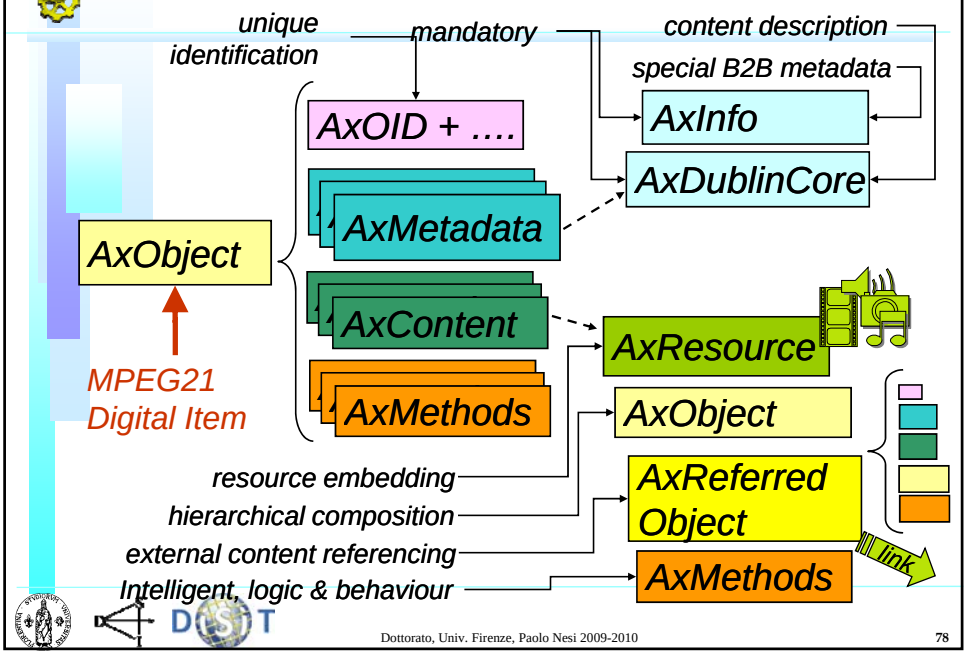
Multimedia GRIDs for the future applications Comparison (IEEE Multimedia March 2009)

	Content management	Content analysis	Media streaming	Interactive controls	Parallel processing
Access Grid	Y	Y	Y		
GridCast	Y		Y	Y	
mmGrid		Y	Y	Y	
gMOD	Y		Y	Y	
MediaGrid	Y		Y	Y	Y
AE@SG	Y				
Parallel-Horus	Y	Y			Y
Context Aware MM Middleware	Y	Y	Y		Y
AXMEDIS	Y	Y	Y	Y	Y

Intelligent Cross Media Content

- **Evolved Business Models**:
  - **Educational**:
    - Sliding Shows, video, document, audio, images…
  - **Procedures/protocols**: (mini applications)
    - Assessing conditions: emergency..
    - Guidelines, routines/procedures, flows, …
  - **Calculators** for several aspects: (mini applications)
    - Dosages and formulas for intensive therapy
    - Estimation of rule for assessing conditions
    - Risk analysis, …e.g.: pulmonary emboli….
    - Classification of conditions/damages, …
  - **Wizards: active and proactive content**
    - Self-unpacking, guiding the user

AXMEDIS Object/Package Model

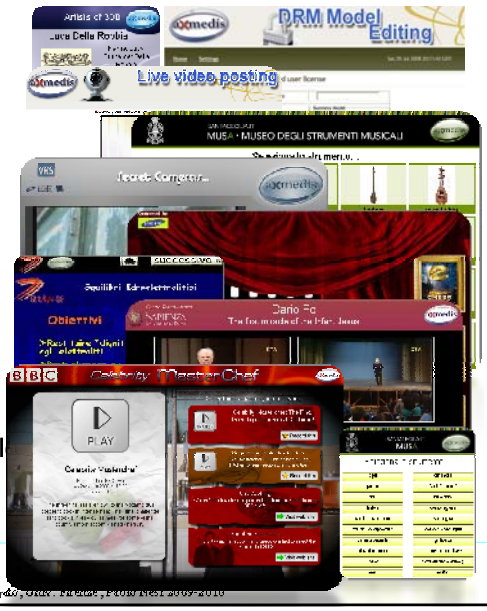


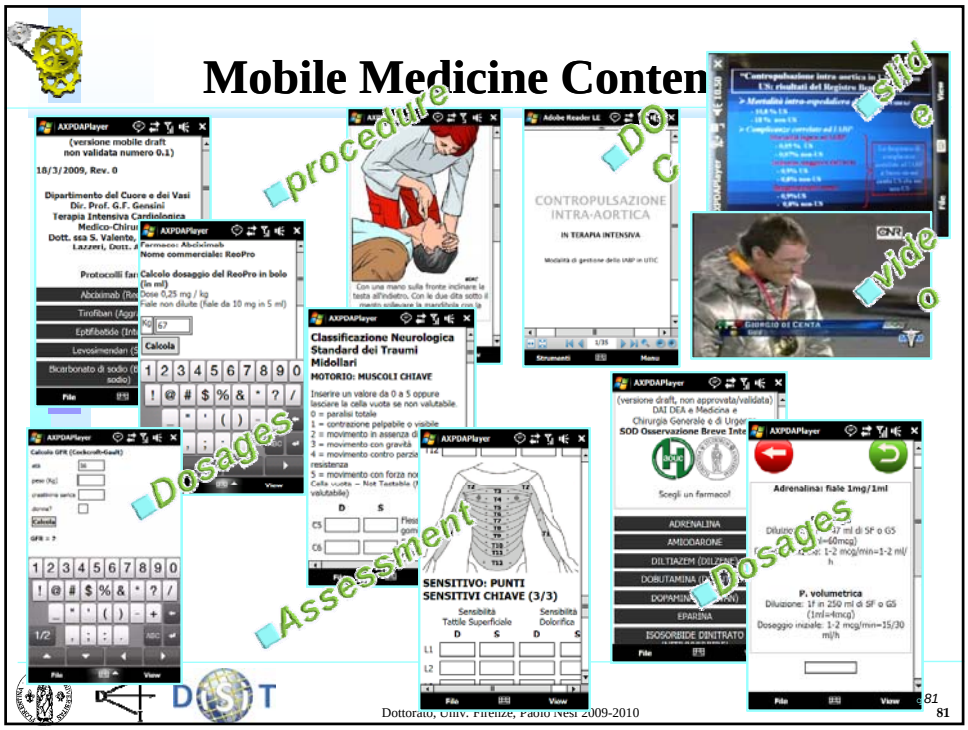
Intelligent Cross Media Content

- **AXMEDIS** Cross media content
  - **Structure**: ISOMEDIA + mpeg-21 + hierarchical + links, etc.
  - **Packaging**: Metadata + descriptors + essences
  - **Classification**: DC + taxonomy + descriptors, any MD,
  - **File format**: direct play video without unpack, ….
  - **Distribution**: download, streaming, progressive, P2P,..
  - **Intelligence**: profiling, decision, scripting, proaction, forms,
  - **Presentational/interaction**: SMIL, MPEG-4, HTML, …
  - **Protection**: sign, CAS vs DRM
  - **Tools**:
    - **Players and Authoring** tools
    - **Automated production**, repurposing, delivering
    - **Annotations**: …….

Possibili Contenuti

- **File singoli**:
  - audio, video, documenti, immagini, etc..
- **Contenuti interattivi**:
  - HTML o SMIL come tecnologi di interazione
  - Guide, giochi, etc.
  - Valoriz. Beni Culturali
  - Contenuti educazionali
- **Wizard proattivi**:
  - Registra video messaggio
  - Upload assistito di liste di fil
  - Emissione licenze
  - ….



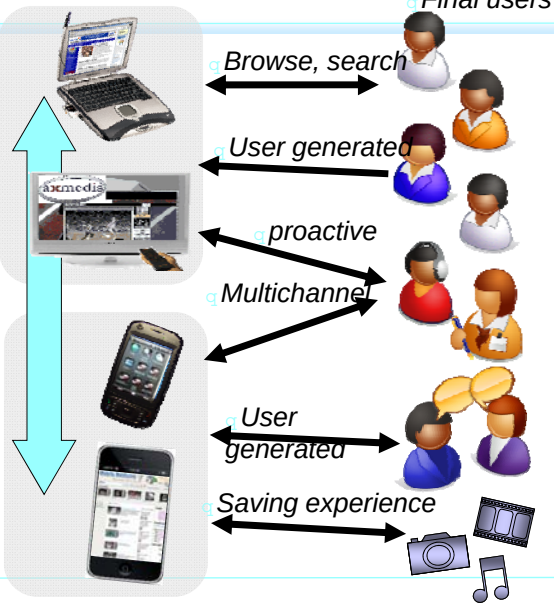


Intelligence Mobile Content

- Collect content on mobile device, PDA,….
- Access to personal collection in any conditions
- Navigate into the collection via several views:
  - medical, taxonomy, classif., description, etc.
  - Use data based: less used, most, recent, etc.
- Querying into the collection
- Keep updated the content collection automatically
- Keep the same content accessible on PC/PDA
- Licensing and rights controls to access and use (patience info and/or record)

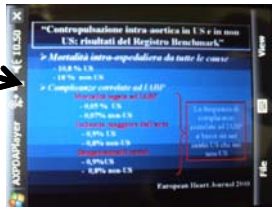
Interactive and intelligent content

- integrated media info
- proactive with the users
- attractive experience
- personalized
- multichannel interoperable
- device interoperable

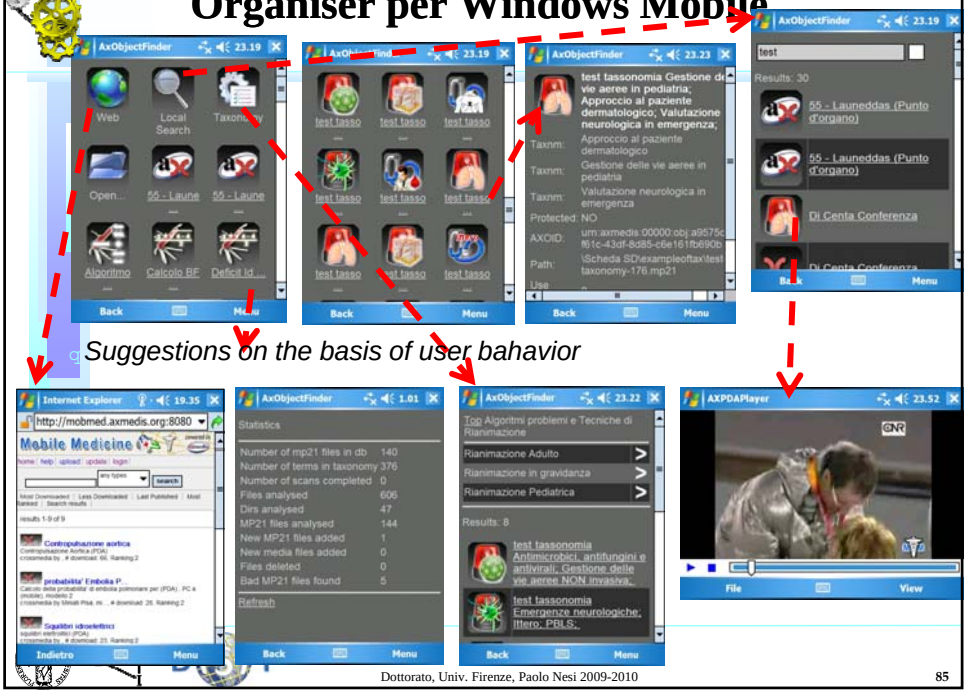


Organizer for PDA

- Mini applications
- PDA
- IE browse
- Mobile Medicine Organiser
- Access OFFLINE



Organiser per Windows Mobile



iPhone/iPod and other devices

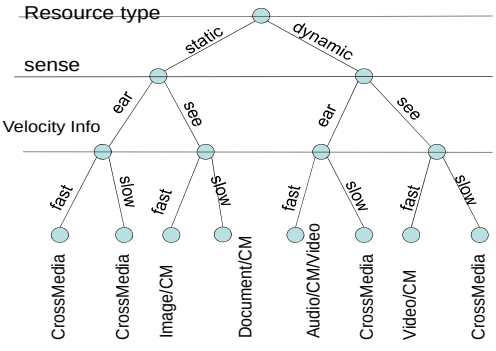
- Direct access to WEB portal as PC
- Play of resources:
  - Video, audio, img, doc. Etc.
  - Calcolators, procedures, …
- Preferred, groups, etc.
- Orgnization of cotnent on the web
- Stream of vido, audio
- Not accessible offline

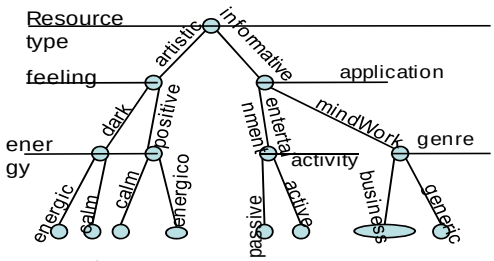
Recommendations and suggestions

- **Users**:
  - Static: User Profile as provided
  - Dynamic: play, UGC post, friends, votes, comments, ..
- **Cross Media Content (Objects):**
  - Technical description + semantics
  - MD + semantic description as Taxonomy
- **Distances among dynamic symbolic entities**:
  - **U→U, C→U, C→C, G→C, ...**
- **Problems:**
  - Computational Complexity:
    - AXCP scalable
    - Clustering, K-means, K-Medoids, ….
  - To find distances among symbolic descriptors

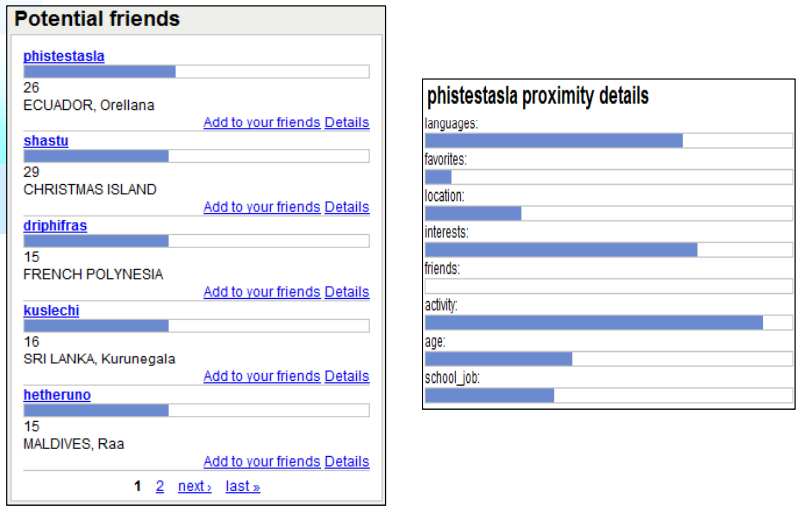
Distances for symbolic descriptors

- **Definition of metrics** among semantics symbolic descriptors for user, content, actions, etc.
- **UU:** D (U(s,d), U(s,d))
  - U(.,d): play, UGC, friends, comments, votes, etc.
- **UO:** D (U(s,d), O(s))
  - …..

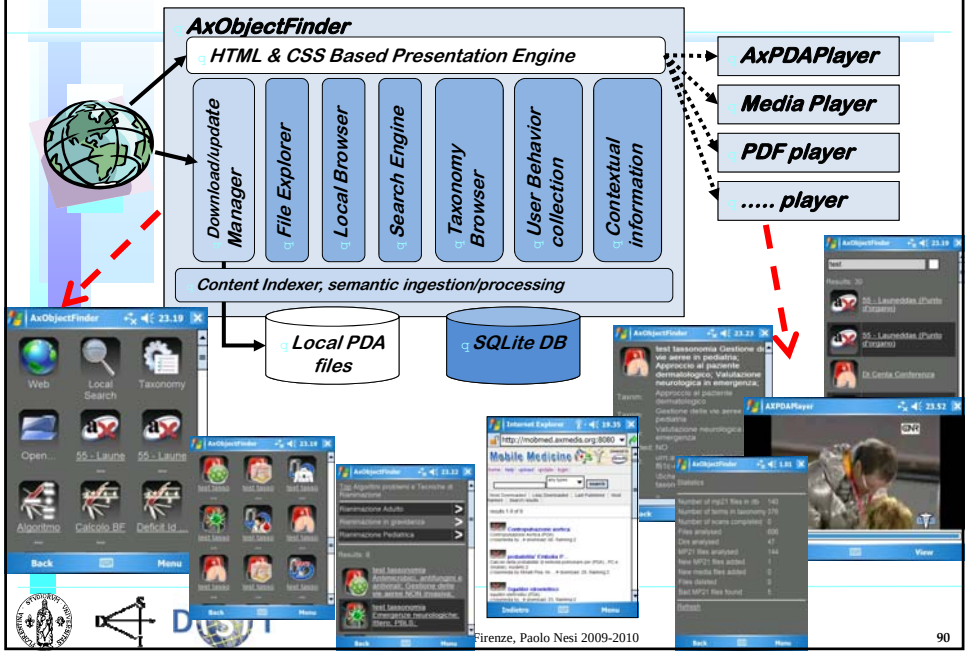




Recommendations: UU, UC, CC, ..



Personal Mobile Social Intelligence



Semantic flows

- User Local Side

User Profile
User behavior
Use data
Content
- DC+IDs
- AXInfo: ver, prod., rights,..
- Descriptors
- Taxonomy
- Groups

Local User Profile
Local User behavior
Local Use data
Content
- DC+IDs
- AXInfo: ver, prod, rights, ...
- Descriptors
- Taxonomy
- Groups

- + Content action data

Recommendation
Suggestions on the basis of user behavior

Local Recommendation
Local Suggestions on the basis of user behavior and local content

Recent improvement

- **platforms**: iPhone, iPod
- **Groups on iPhone and PDA**
- **Intelligent Frontal query**
  - Fuzzy model for frontal query on metadata, taxonomy for object, web pages, comments, forum messages, etc.
  - Very robust with respect to the user intention and wrong writing
- **Tracking:**
  - IP, GPS, download, OS, date, time, users, etc.
- **C→C, U→C, U→U suggestions**:
  - clustering of Content

Some DISIT Projects

- Multimedia Content Modeling and distribution:
  - MOODS, cooperative work on Music notation
  - WEDELMUSIC platform (chair), IST Fp5
    - WEDELMUSIC conference series
    - WEDELAUTHORING (chairs)
  - MUSICNETWORK Environment (chair), IST Fp5
    - Workshops, emerging European associations
  - IMUTUS, music tuition, distance learning, IST Fp5
  - MPEG-SMR integration (co-chair)
  - MPEG M3W, Multimedia Middleware
  - AXMEDIS, Automating cont. prod. and protection
  - IMAESTRO, music education, cooperative, gesture, etc.
  - Other minor projects: archives, mobile distribution, etc.

References

- DISIT http://www.disit.dsi.unifi.it/
  - Per slide complete si veda materiale dei corsi di Sistemi Distribuiti e di Sistemi Collaborativi e di Protezione, SCP, specialmente su SN, intelligent content, protezione
- Mobile Medicine Social network:
  - http://mobmed.axmedis.org
  - Manuale e strumenti di produzione, player
- AXMEDIS: http://www.axmedis.org
  - AXCP tool e players, intelligent content, mpeg-21, mobile
- IMAESTRO: http://www.imaestro.org
  - Collaborative tools for music education, mpeg smr
- MOODS: http://www.dsi.unifi.it/~moods/