Towards **CSpaces**: a new perspective for the Semantic Web

Francisco Martín-Recuerda

francisco.martin-recuerda@deri.org

IASW 2005 – Jyväskylä (Finland)

25th August 2005

Index

- The Semantic Web
- Open issues in the Semantic Web
- Conceptual Spaces (CSpaces)
- CSpaces in the "real World"
- Wrap-up

The Semantic Web

- What is the Semantic Web?
  - *"The Semantic Web is an extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation."* [W3C, 2001]
  - *"The Semantic Web is a project that intends to create a universal medium for information exchange by giving meaning, in a manner understandable by machines, to the content of documents on the Web."* [Wikipedia, 2004]

The Semantic Web

- What is the Semantic Web?
  - Annotating the Web with formal semantic descriptions together with domain theories (i.e., ontologies) will increase machine support for the interpretation and integration of information published in the Web
  - Semantic Annotations as a huge global Knowledge Base (or set of Knowledge Bases) that machines can interpret
  - It is expected that Inference Systems use these semantic annotations to provide sophisticated reasoning services

Open issues in the Semantic Web

- Dichotomy of the Semantic Web: Semantic Annotations + Web pages
  - Keep coherence between the information published in the Web and the associated semantic annotations is not an easy task
  - Still there is a separation between human information and machine information
- Scalability: "*Current proposals for Semantic Web languages such as the Web Ontology Language (OWL) are based on formal logic. Consequently they share the advantages and disadvantages of formal logic: a well-founded semantic can be used to derive implicit information, however, at the price of a high computational complexity*" [KW D2.1.1]

Open issues in the Semantic Web

- Inconsistency: The distributed and multi-authorship nature of the Semantic Web may introduce inconsistencies in the information published
  - Deal with inconsistency may affect performances and require extra computational resources
  - Different sources of information can drive to contradictory results
- Redundancy is one of the main problems of the Web and it is closely related with Information Overload and Scalability

francisco.martin-recuerda@deri.org

Open issues in the Semantic Web

- Heterogeneity: the Semantic Web is an open and distributed system in which heterogeneity cannot be avoided.
  - Heterogeneity is an old problem that has been studied during the last decades in data integration
  - RDF was introduced to ensure interoperability. However this interoperability is limited to RDF semantics (RDF has a very limited expressivity and cannot deal directly with semantics of more complex languages like OWL, SWRL and WRL)
  - Heterogeneity occurs in four different levels: syntactic, terminological, conceptual and semiotic [KW D2.2.1]
  - For a correct interoperation between participants in the Semantic Web, it is required a previous semantic agreement of the data used

Open issues in the Semantic Web

- Creation of Semantic Annotations: who will take the responsibility to annotate Web pages with semantic descriptions of the information published?
- Security and Trust:
  - How can we guarantee that only a restricted amount of users can visualize and edit concrete semantic annotations?
  - How can we guarantee that new semantic annotations do not introduces cycles, inconsistencies, wrong data and other not desire elements for a correct reasoning behave?

Open issues in the Semantic Web

- In summary:
  - How to guarantee usability and efficiency for reasoning services using a large amount of distributed and mapped semantic annotations that may be are inconsistent?
  - How to reconciliate heterogeneity in Semantic Web annotations?
  - How to limit the access to Semantic annotations?
  - How to guarantee the correctness of the semantic annotations stored?
  - How to generate and share semantic annotations?
  - Which architecture can be appropriate to support the requirements of the Semantic Web?

Perhaps a new perspective can help ....

- In s
  - H
    r
    d
    m
  - H
    a
  - H
  - H
    a
  - H
  - W
    r

that

Web

ns?

antic

ons?

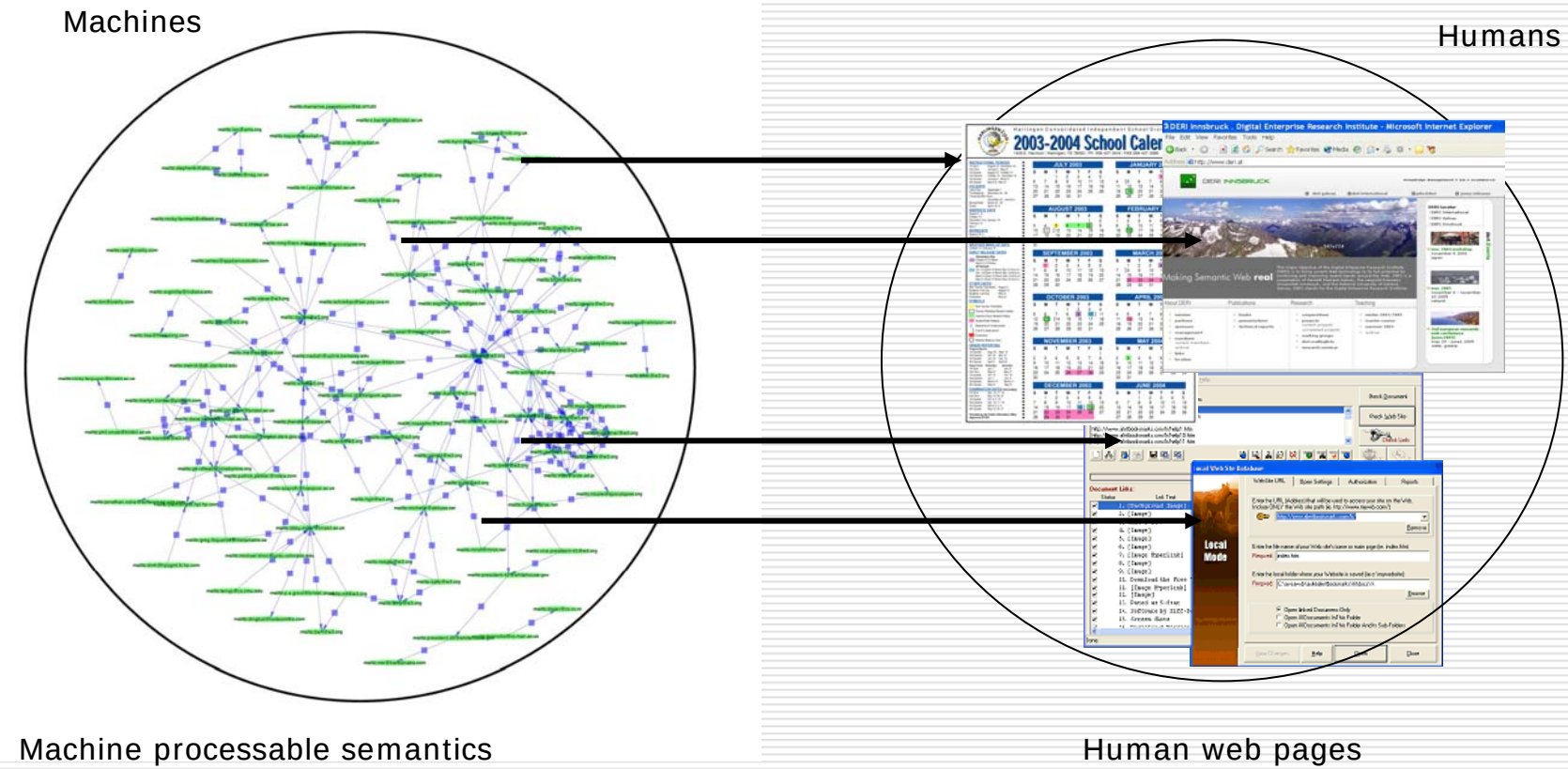
ort the

Conceptual Spaces (CSpaces)

- What are CSpaces?
  - Step behind current view of the Semantic Web
  - Increase machine support for information management by making machine processable semantics the prevalent representation formalism in this new infrastructure (reducing dichotomy between Web and Semantic annotations).
  - Interaction/Coordination model based on publication-subscription of machine processable semantics
  - Participants (humans and machines) publish their data semantically described in Individual CSpaces and shared this data in Shared CSpaces that can have associated access rights mechanisms
  - Natural language generation and graphical knowledge visualization techniques will make possible that humans deal with this "purest semantic" Web.

Conceptual Spaces (CSpaces)

- What are CSpaces?



francisco.martin-recuerda@deri.org

Conceptual Spaces (CSpaces)

- Building Blocks towards CSpaces:
  - Semantic data and organizational model
  - Coordination model
  - Semantic interoperability and consensus making model
  - Security and Trust model
  - Knowledge visualization model
  - Architecture model

Conceptual Spaces (CSpaces)

- Semantic data and organizational model:
  - A Conceptual Space (CSpace) is a finite set of ontologies, their instances, and mapping and transformation rules (alignment specification). All these elements are represented using a common formal language that allows ontologies to be enriched with rules, and exhibit some degree of semantic autonomy.
  - All logical statements have associated three identifiers:
    - id_context identifies univocally the context (CSpace) in which the logical statement was created
    - id_statement identifies univocally the logical statement
    - id_version identifies each version of a logical statement

francisco.martin-recuerda@deri.org

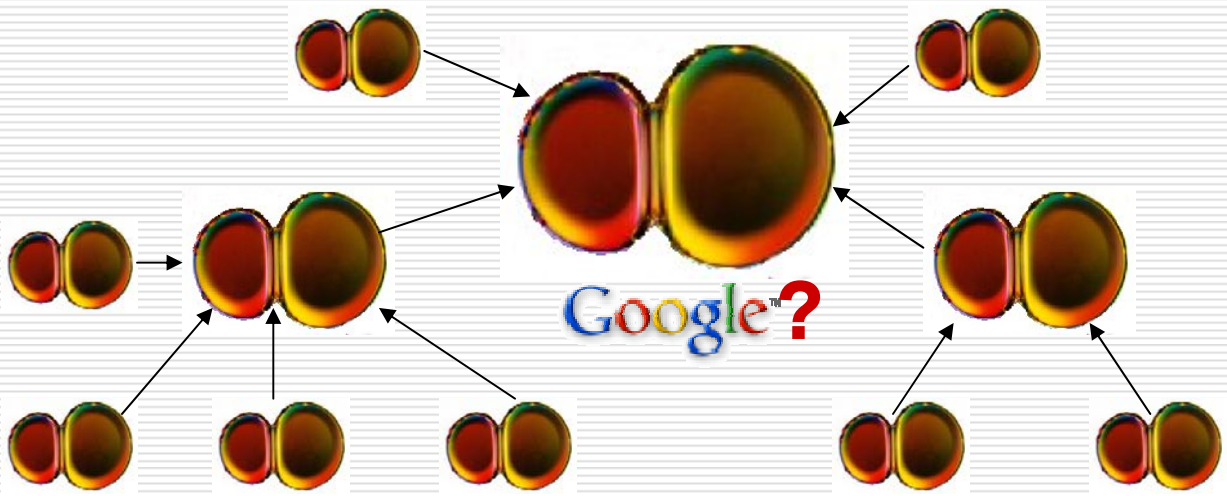
Conceptual Spaces (CSpaces)

- Semantic data and organizational model:
  - A Conceptual Space (CSpace) is composed of two sub-spaces:
    - A raw sub-space stores imported or local data, schemas, and alignment specifications (mapping and transformation rules) between these schemas.
    - A reasoning sub-space provides a compact representation of an associate raw sub-space. The main goal of this compact representation is to maximize reasoning performances.

  - Two types of CSpaces:
    - An Individual CSpaces is a formal representation of the perception that each individual (human or not) has about the Semantic Web (or a limited part of it)
    - Shared CSpaces are conceptual spaces shared by several participants that have reached an agreement on how to represent semantically common concepts. This requirement is fundamental to ensure interoperability between participants.

Conceptual Spaces (CSpaces)

- Semantic data and organizational model:
  - Individual CSpaces can be viewed as leaves and shared spaces can be graphically considered as the branches and the trunk of a fictitious tree following a very similar organization proposed in CO4 [Euzenat, 1995]



Conceptual Spaces (CSpaces)

- Building Blocks towards CSpaces:
  - Semantic data and organizational model
  - **Coordination model**
  - Semantic interoperability and consensus making model
  - Security and Trust model
  - Knowledge visualization model
  - Architecture model

Conceptual Spaces (CSpaces)

- Coordination Model:
  - The Web follows a "persistent publish and read" model
  - "persistent publish and read" has been applied in Semantic Web Services (Triple Space Computing [Fensel 2004]) and parallel computing (Tuple Space Computing [Gelernter, 1985])
  - Weblogs (blogs or bloggins) and RSS have popularized another paradigm in the Web "persistent publish, read and subscribe"
  - Publish-Subscription models was also applied to parallel computing and distributed software components [Hermes, 2005]
  - I proposed to follow "persistent publish, read and subscribe" model as a coordination/interaction model in CSpaces for humans and machines

Conceptual Spaces (CSpaces)

- Coordination Model:
  - Desirable orthogonal dimensions for "persistent publish, read and subscribe" model:
    - Space decoupling: processes involved in the interaction can run in completely different computational environments
    - Reference decoupling: processes involved in the interaction do not need to know each other (anonymous)
    - Time decoupling: processes do not need to be up at the same time during the interaction (asynchronous)
    - Flow decoupling: main flows of process are not affected for the generation of reception of data (no blocking read (receive) and write (send) operations)

Conceptual Spaces (CSpaces)

- Coordination Model (API):
  - void write (set tuples, IdCSpace id)
  - Tuple take (Template t, IdCSpace id)
  - Tuple read (Template t, IdCSpace id)
  - Tuple waitToRead (Template t, IdCSpace id)
  - Set scan (Template t, IdCSpace id)
  - long countN (Template t, IdCSpace id)
  - void subscribe (IdSubscriber s, Template t, IdCSpace id)
  - void unsubscribe (IdSubscriber s, Template t, IdCSpace id)
  - void advertise (IdSubscriber s, Template t, IdCSpace id)
  - void unadvertise (IdSubscriber s, Template t, IdCSpace id)

Conceptual Spaces (CSpaces)

- Building Blocks towards CSpaces:
  - Semantic data and organizational model
  - Coordination model
  - Semantic interoperability and consensus making model
  - Security and Trust model
  - **Knowledge visualization model**
  - Architecture model

francisco.martin-recuerda@deri.org

Conceptual Spaces (CSpaces)

- Knowledge visualization model:
  - Knowledge visualization comprises all the techniques and mechanisms that facilitate the exploration and visualization of semantic formal representation of information stored in knowledge bases.
  - Tree and graph visualization approaches have been intensively used for the representation of ontologies
  - "*NLG takes structured data in a knowledge base as an input and produces Natural Language text, tailored to the presentational and the target reader*" [Reiter 2000]

francisco.martin-recuerda@deri.org

Conceptual Spaces (CSpaces)

- Knowledge visualization model:


```
<?xml version="1.0" encoding="UTF-8" ?>
- <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:ns0="http://www.aktors.org/ontology/portal#">
  - <rdf:Description rdf:about="http://www.aktors.org/ontology/portal#tech-prof-1055846863-23538">
      <ns0:has-author rdf:resource="http://www.aktors.org/scripts/sheffield.dome#K.Bontcheva.dcs.shef.ac.uk" />
    </rdf:Description>
  - <rdf:Description rdf:about="http://www.aktors.org/ontology/portal#tech-prof-1055847699-23796">
      <ns0:has-author rdf:resource="http://www.aktors.org/scripts/sheffield.dome#K.Bontcheva.dcs.shef.ac.uk" />
    </rdf:Description>
  - <rdf:Description rdf:about="http://www.aktors.org/scripts/sheffield.dome#K.Bontcheva.dcs.shef.ac.uk">
      <ns0:family-name>Bontcheva</ns0:family-name>
      <ns0:full-name>Kalina Bontcheva</ns0:full-name>
      <ns0:given-name>Kalina</ns0:given-name>
      <ns0:has-appellation>Dr</ns0:has-appellation>
      <ns0:has-email-address>K.Bontcheva@dcs.shef.ac.uk</ns0:has-email-address>
      <ns0:has-telephone-number>+4401142221930</ns0:has-telephone-number>
      <ns0:has-web-address>http://www.dcs.shef.ac.uk/~kalina/</ns0:has-web-address>
      <ns0:works-in-unit rdf:nodeID="genid29.3f66a106" />
      <rdf:type rdf:resource="http://www.aktors.org/ontology/portal#Researcher-In-Academia" />
      <rdf:type rdf:resource="http://www.aktors.org/ontology/support#Thing" />
      <rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-schema#Resource" />
      <rdf:type rdf:resource="http://www.aktors.org/ontology/portal#Employee" />
      <rdf:type rdf:resource="http://www.aktors.org/ontology/portal#Generic-Agent" />
      <rdf:type rdf:resource="http://www.aktors.org/ontology/portal#Affiliated-Person" />
      <rdf:type rdf:resource="http://www.aktors.org/ontology/portal#Generic-Agent" />
      <rdf:type rdf:resource="http://www.aktors.org/ontology/portal#Person" />
      </rdf:Description>
  </rdf:RDF>
```

Conceptual Spaces (CSpaces)

- Knowledge visualization model:

<?xml version="1.0" encoding="UTF-8" ?>

Kalina Bontcheva (University of Sheffield) Profile

RDF representation from AKT Ontology (www.akt.org/ontology/)

http://www.hyphen.info/browse/

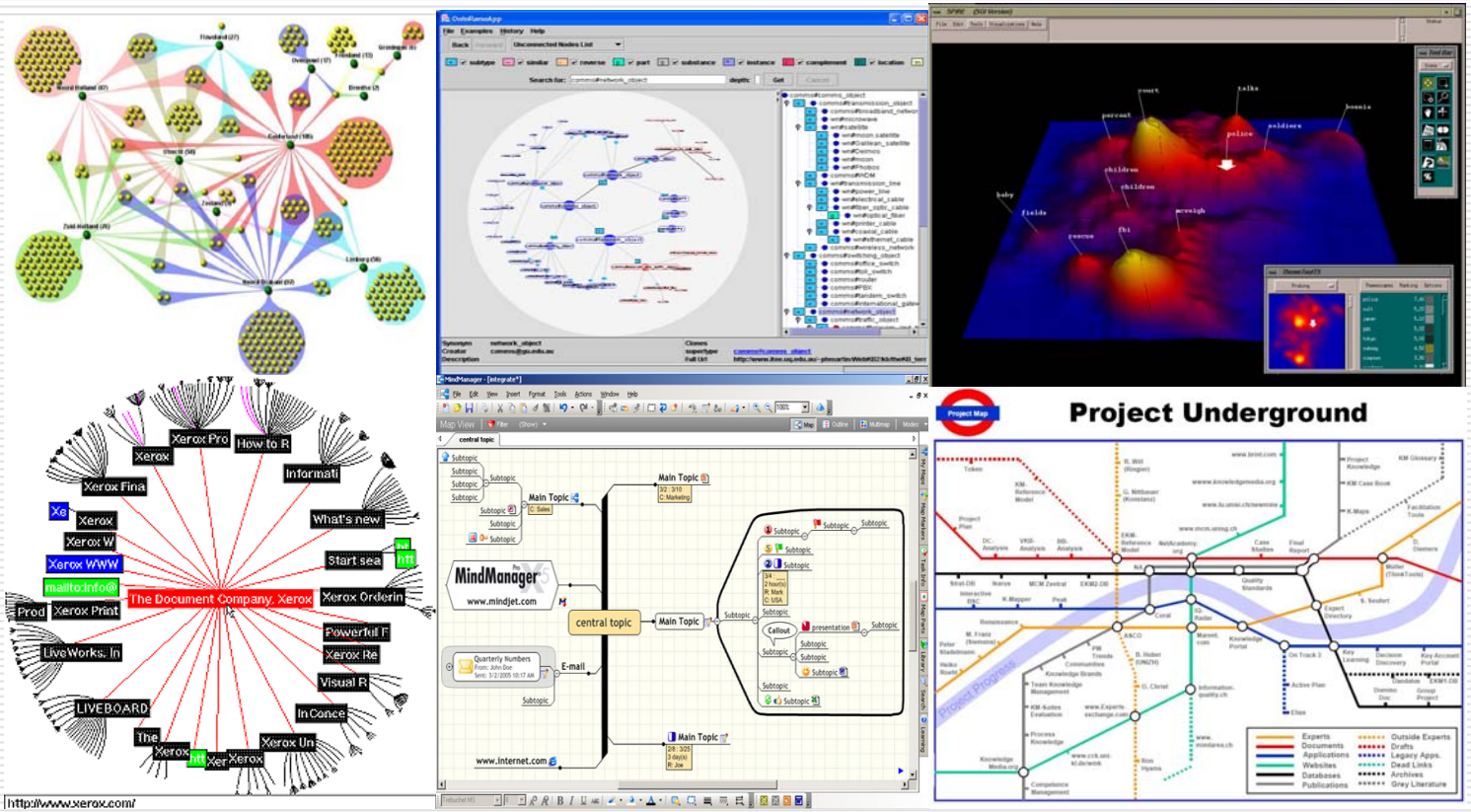
"*Kalina Bontcheva has a Dr appellation, a web page http://www.dcs.shef.ac.uk/kalina/, a phone number + 44 011 4222 1930, and email address K.Bontcheva@dcs.shef.ac.uk. Kalina Bontcheva works on SEKT and Knowledge Web projects, and GATE – General Architecture for Text Engineering and ANNIE – Open Source Information Extraction technologies.*"

Example generated with ONTOSUM and presented in the the paper "Generating Tailored Textual Summaries from Ontologies". Second European Semantic Web Conference (ESWC'2005). Crete. 2005.

</rdf:RDF>

Conceptual Spaces (CSpaces)

- Knowledge visualization model:



francisco.martin-recuerda@deri.org

Conceptual Spaces (CSpaces)

- Building Blocks towards CSpaces:
  - Semantic data and organizational model
  - Coordination model
  - Semantic interoperability and consensus making model
  - Security and Trust model
  - Knowledge visualization model
  - **Architecture model**

francisco.martin-recuerda@deri.org

Conceptual Spaces (CSpaces)

- Architecture Model:
  - Which suitable architecture can be used for the re-elaboration of the Semantic Web that CSpaces proposed?
  - We have to take into account:
    - Huge amount of semantic data described using a formal language
    - Reasoning/Inference support
    - Compatible with a coordination model that is focussed in asynchronous communication based on "*persistent publish, read and subscribe*" metaphor
    - Provide a infrastructure that facilitate semantic interoperability and cosensus making (not discussed)
    - Security and Trust issues (not discussed)

Conceptual Spaces (CSpaces)

- Architecture Model:
  - The Web has been described using an abstract model called REST (*Representational State Transfer*) [Fielding 2000]
    - The fundamental principle of REST is that resources are stateless and identified by URIs
    - HTTP is the protocol used to access to the resources and provides a minimal set of operations enough to model any application domain
    - HTTP operations can be easily mapped into Tuple-Space operations (READ, TAKE and WRITE in TSpaces [Wyckoff 1998])
    - REST cannot model properly asynchronous communication between clients and servers [Khare 2004].

Conceptual Spaces (CSpaces)

- Architecture Model:
  - Asynchronous communication is a key requirement in CSpaces.
  - I am exploring the feasibility of use P2P as a core architecture model for CSpaces.
  - "*A peer-to-peer system is a decentralized, distributed system that consists of symmetric nodes caller peers. It is self-organized and capable of adapting to changes such as failure*" [Pietzuch, 2004]
  - This preliminary proposal is strongly influenced in the work done in OceanStore, Edutella and SWAP projects
  - Like OceanStore, instead to follow a pure P2P model, I follow a hybrid model (P2P and Client/Server) that brings a backbone of reliable powerful and well-connected nodes FM1

FM1 Super-peer or hybrid model? Check in the literature

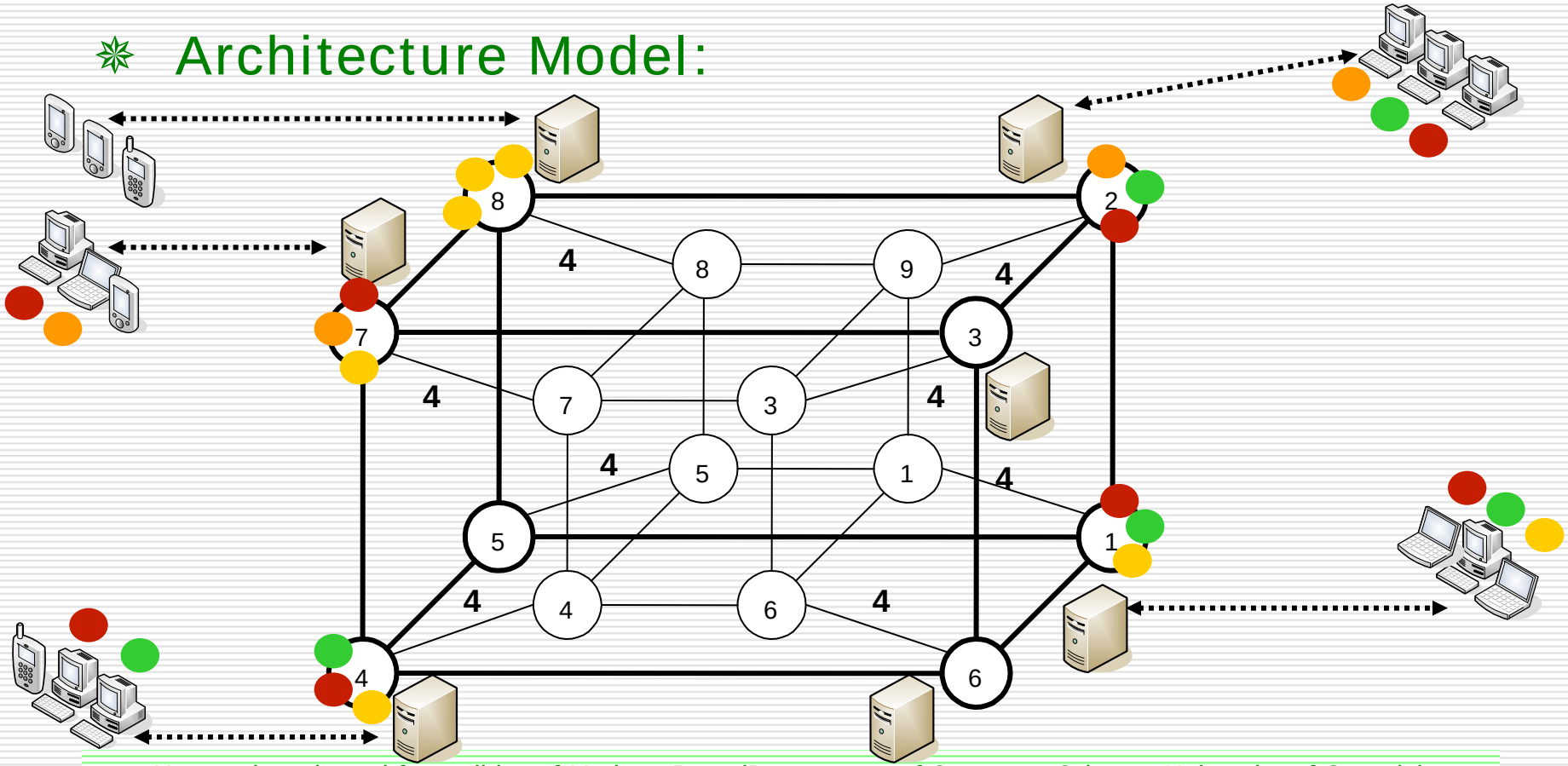
Francisco J. Martin-Recuerda Moyano; 24.08.2005

Conceptual Spaces (CSpaces)

- Architecture Model:
  - Three different kind of nodes are currently proposed:
    - CSpaces-servers store mainly primary (secondary for heavy-clients) replicas of the data published in CSpaces; support versioning services; provide an access point for CSpace clients to the peer network; include reasoning services for evaluating complex queries; implement subscription mechanisms related with the contents stored; and balance workload
    - CSpaces-heavy-clients are focused to provide CSpaces-server infrastructure for users that have to work off-line. These clients also include a service for presentation and edition of semantic data, and a replication mechanism
    - CSpaces-light-clients only include the presentation infrastructure to query, edit and visualize knowledge contents stored on CSpace-servers.

Conceptual Spaces (CSpaces)

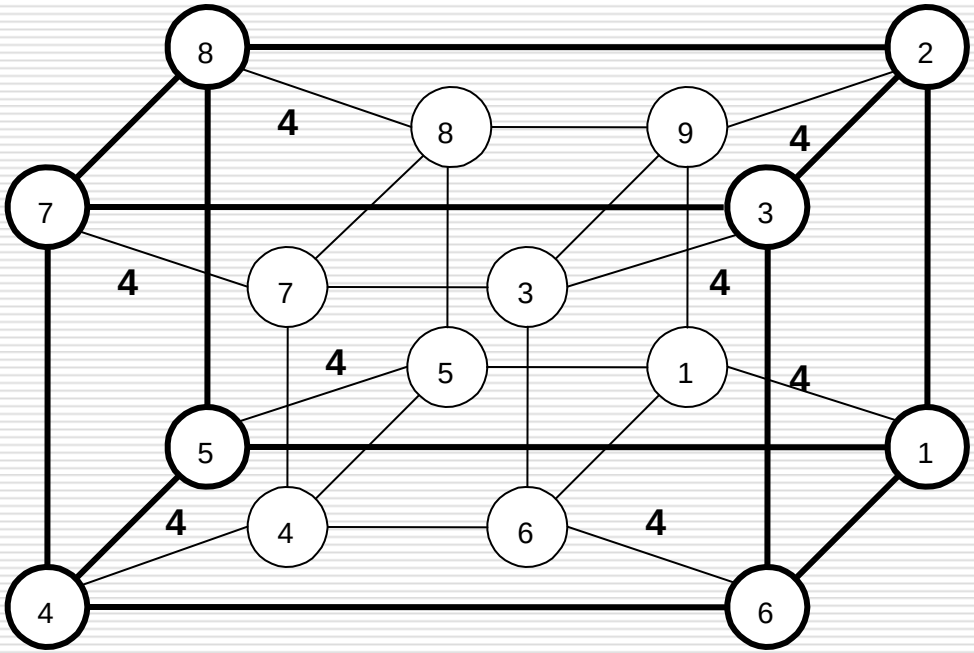
- Architecture Model:

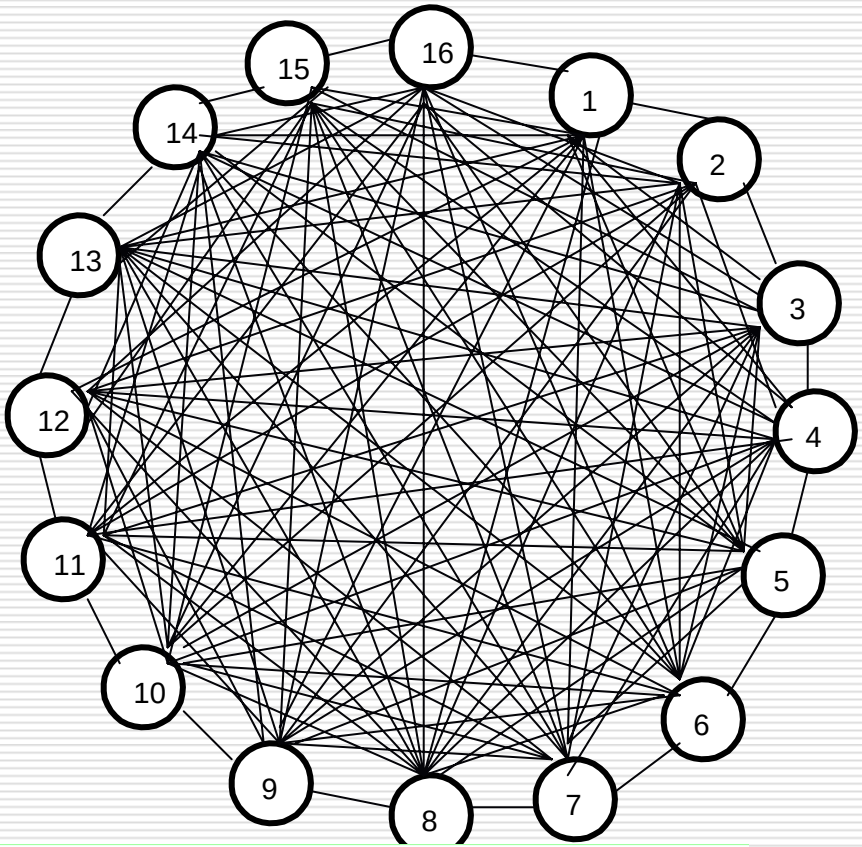


Hypercube adapted from slides of Mathew Ross (Department of Computer Science, University of Georgia)

Conceptual Spaces (CSpaces)

- Which is more efficient?





Example from slides of Mathew Ross (Department of Computer Science, University of Georgia)

Index

- The Semantic Web
- Open issues in the Semantic Web
- Conceptual Spaces (CSpaces)
- **CSpaces in the "real World"**
- Wrap-up

CSpaces in the "*real world*"

- Personal and Distributed Knowledge Management (PDKM)
- Enterprise Application Integration (EAI)
- Distributed Sofware Components
- Semantic Web Services
- Ubiquitous Computing

CSpaces in the "*real world*"

- Personal and Distributed Knowledge Management
  - "*Information overload occurs when the volume of the information supply exceeds the limited human information processing capacity. Dysfunctional effects such as stress or confusion are the result*". [Eppler and Mengis 2003].
  - Print, film, magnetic, and optical storage media produced about 5 exabytes (2^{60} bytes) of new information in 2002 [SIMS 2003]:
    - 92% of the new information was stored on magnetic media, mostly hardisks,
    - the estimated size of Internet is 532.897 terabytes accessed by around 600 million people.
    - About 31 million emails are sent daily
    - Unsolicited emails (also called spam) represent 33% of today's email traffic

CSpaces in the "*real world*"

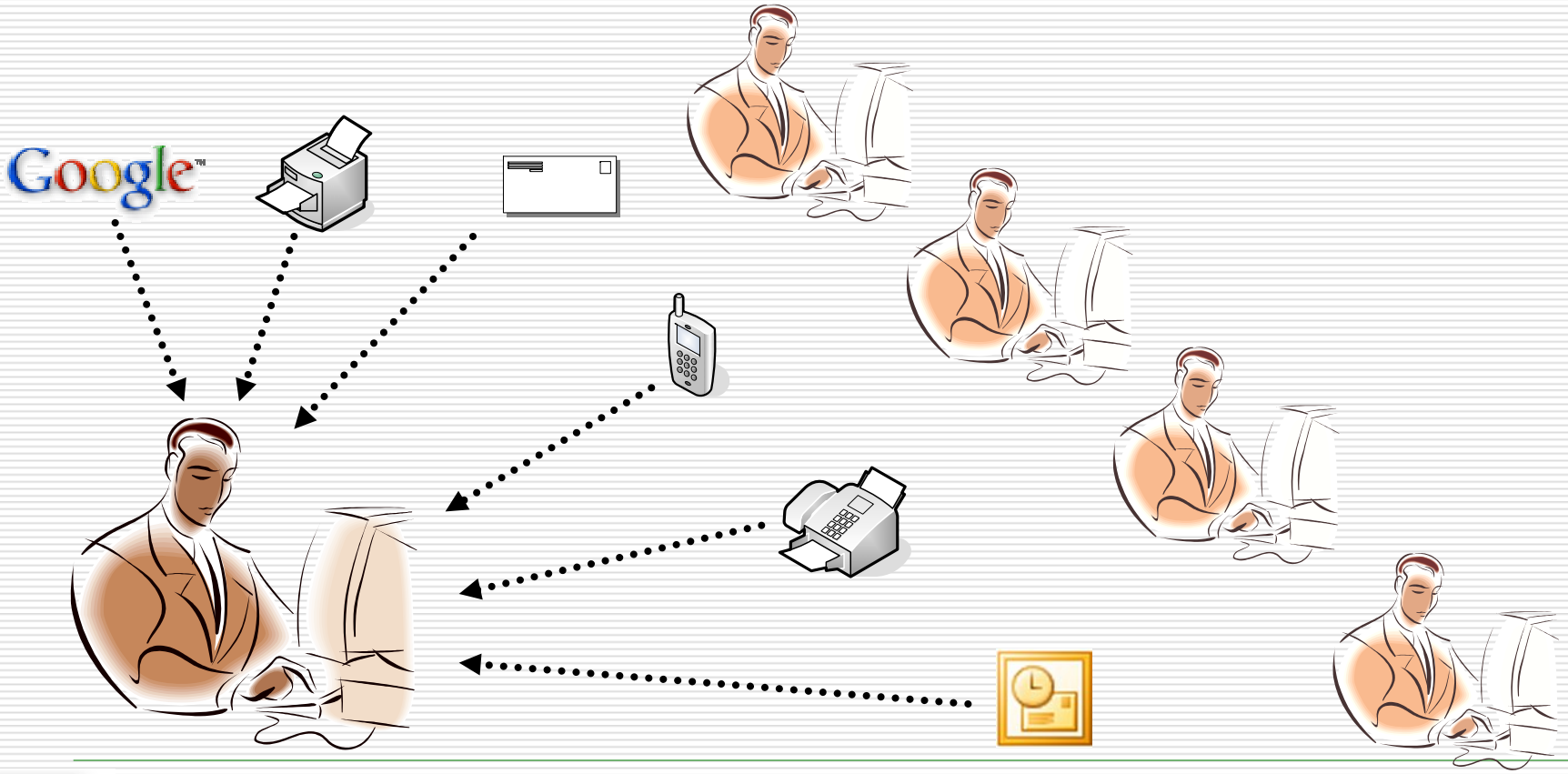
- Personal and Distributed Knowledge Management
  - "Personal Knowledge Management (PKM) is a collection of processes that an individual needs to carryout in order to gather, classify, store, search and retrieve knowledge in his/her daily activities. Activities are not confined to business/work-related tasks but also include personal interests, hobbies, home, family and leisure activities" [Tsui 2002]
  - Knowledge workers often have to deal with data, information and knowledge specified in various formats (e.g. hard copy, video, picture, texts, voice message etc.), distributed using different information channels (e.g. emails, fax, instant messages, file systems, etc) and stored using multiple electronic devices for communications, planning and recording purposes.

CSpaces in the "*real world*"

- Personal and Distributed Knowledge Management
  - Unify information channels and sources in a suitable one that can be shared by humans and machines.
  - Increase machine support for information management by making machine processable semantics the prevalent representation formalism in this new infrastructure.
  - Reduce heterogeneity by introducing mechanisms that allow humans to find agreements in the representation and definition of common terminology and semantic data specifications.
  - Identify redundant information
  - Promote the use of ontologies as articulation means for organizing data and identifying redundant data.

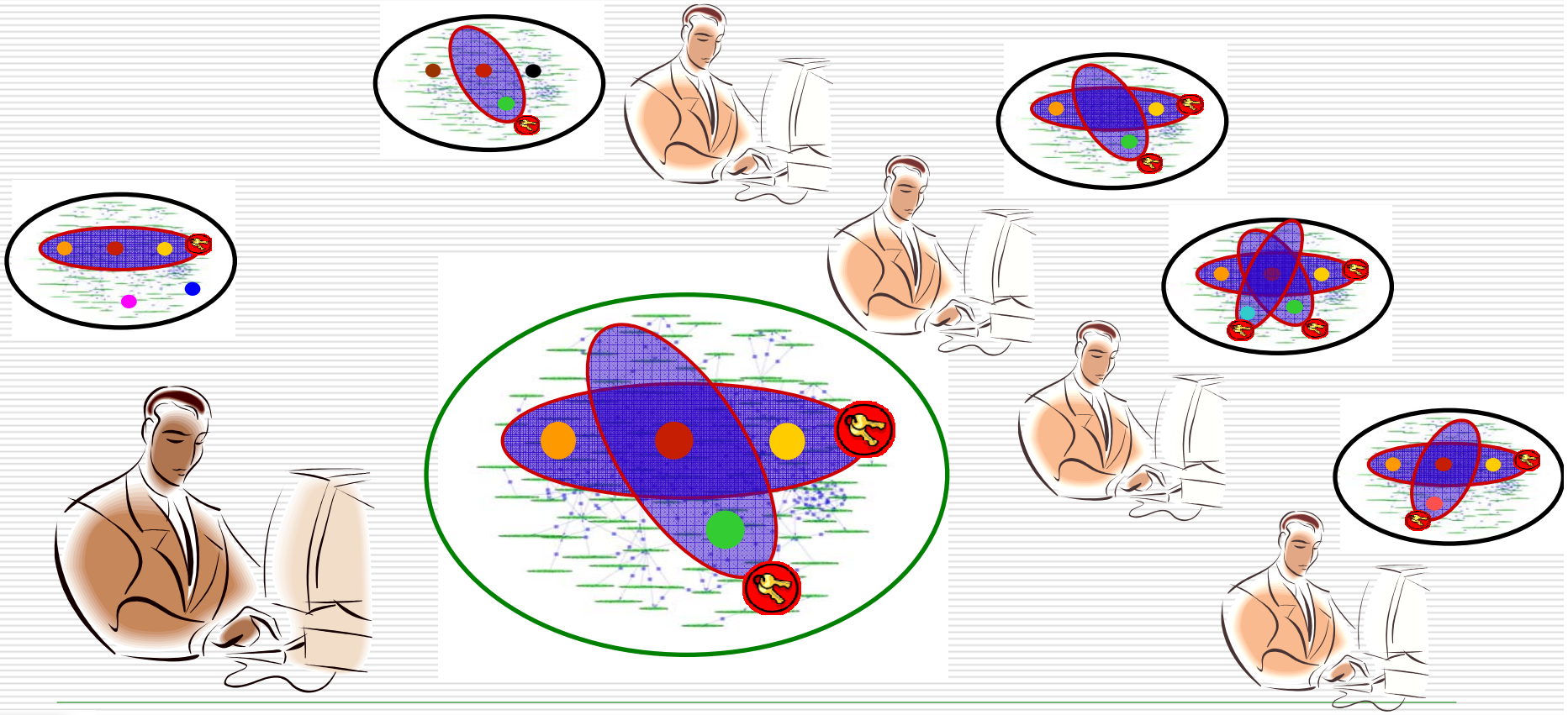
CSpaces in the "*real world*"

- Personal and Distributed Knowledge Management



CSpaces in the "*real world*"

- Personal and Distributed Knowledge Management



francisco.martin-recuerda@deri.org

CSpaces in the "*real world*"

- Personal and Distributed Knowledge Management
  - Perhaps it is a bit extreme what I proposed in the previous slides but …
    - Semantic Web Desktop promotes the generation and share of personal metadata.
    - Semantic Email Processes (SEP) encourage the use of machine processable semantics in the automation of concrete processes through email
    - Semantic Blogs introduce machine processable semantics in the data published, and facilitate the execution of RSS processes --> "persistent publish, read and subscribe"
    - P2P oriented to store and share machine processable semantics (SWAP, Edutella, Bibster, …)
    - Email communication has become one of the main means of communication in a very short period of time

CSpaces in the "*real world*"

- Personal and Distributed Knowledge Management
  - CSpaces can contribute to organize and share knowledge using a bottom-up approach:
    - Instead of centralized systems that forces users to agree in a set of rules, schemas and data, CSpaces offer a distributed infrastructure where users can publish personal knowledge that can be shared with other users with common information/interests.
    - This approach is inspired in an earlier proposal called Distributed Knowledge Management [Bonifacio 2002, 2003]
    - Several tests in real scenarios have showed users were more favorable to this kind of approach because it takes into account the different perspectives and understandings that users have about the world and more concretely about the information, processes and interactions of their organizations or working groups [Bonifacio 2002, 2003]

CSpaces in the "*real world*"

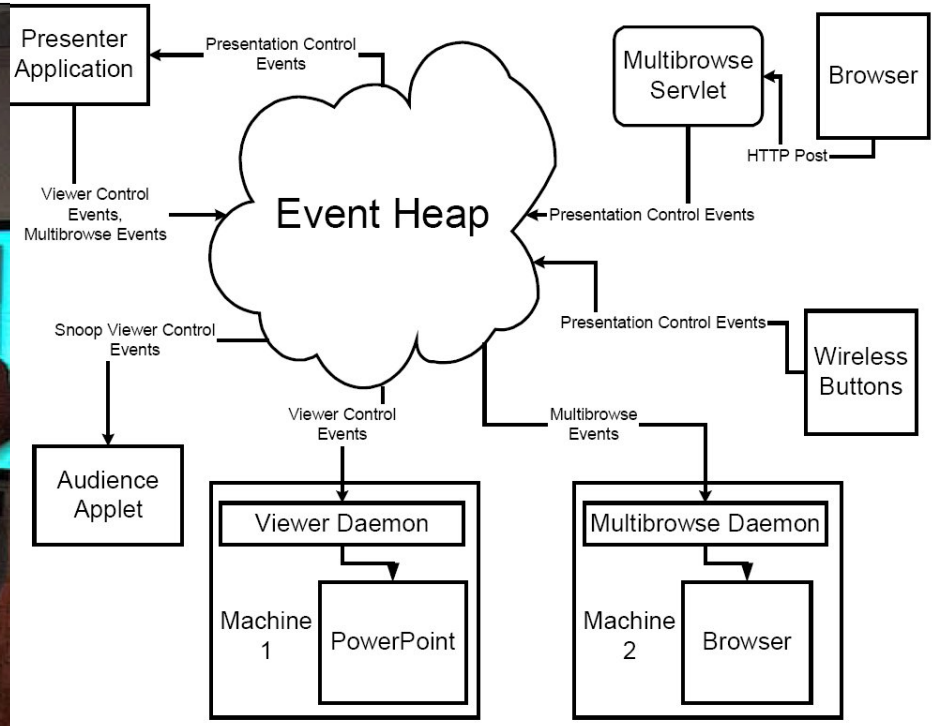
- Ubiquitous Computing
  - Ubiquitous computing was the vision of Mark Weiser for a World saturated with computing and wireless communication gracefully integrated with human users.
  - Active environments are sensor-rich environments with computational and communication facilities that analyze users behave to anticipate potential new requirements or facilitate the developing of certain tasks in a natural way.
  - Interactive Workspaces [Fox 2000] project is a concrete example of application of active environments in laboratories and collaborative e-learning spaces
  - One of the components of this infrastructure is Event Heap [Johanson 2002], a coordination mechanism derived from a tuple space model

CSpaces in the "*real world*"

- Ubiquitous Computing
  - Several initiatives promote the integration of machine processable semantics in the tuple space model: Triple Space Computing (UIBK and NUIG), RDF/Semantic Web Space (Free U. Berlin), XSVM+ (TU Vienna), sTuples (NOKIA-Lassila).
  - The coordination model of CSpaces extends tuple space model with a more sophisticate notification service based on publish-subscribe paradigm
  - Flow decoupling from the client side and anticipation of future client requests and/or data format published are two of the advantage of this extension

CSpaces in the "*real world*"

- Ubiquitous Computing



Wrap-up

- The Semantic Web increases machine support for the interpretation and integration of information published in the Web
- Open issues in the Semantic Web:
  - Keep coherence between Web and semantic annotations
  - Scalability
  - Inconsistency and Redundancy
  - Heterogeneity
  - Security and Trust
  - Generation and sharing of semantic annotations

Wrap-up

- CSpaces proposes a new step behind the current view of the Semantic Web to facilitate the resolution of the open issues discussed before
- Main contributions of the proposal:
  - Make machine processable semantics the prevalent representation formalism
  - Include Knowledge visualization mechanism to allow humans to deal with this kind of formalisms
  - Define an organizational model that facilitate the production, distribution and sharing of machine processable semantics
  - Introduce a simple but powerful coordination model that allow interaction between machines and humans
  - Propose of Architecture based on a hybrid model (P2P and Client/Server)

Wrap-up

- Like the Semantic Web, many possible application scenarios:
  - Personal and Distributed Knowledge Management (PDKM)
  - Enterprise Application Integration (EAI)
  - Distributed Sofware Components
  - Semantic Web Services
  - Ubiquitous Computing
  - e-Commerce
  - GRID Computing
  - …

Conceptual Spaces (CSpaces)

Thanks very much

Questions?

francisco.martin-recuerda@deri.org