

Calcolo Distribuito e Coordinazione

Università di Camerino

Corso di Laurea Specialistica in Informatica
12 CFU

I periodo didattico

Emanuela Merelli
emanuela.merelli@unicam.it

Struttura del Corso

In corso consiste in

24 ore di lezione
16 di esercitazione
16 di laboratorio
2 ore settimanali di ricevimento
studenti

Collaboratore didattico
Leonardo.mariani@unicam.it

Il corso viene valutato in

12 CFU (Crediti Formativi
Universitari)

Sono previsti seminari su
argomenti specifici

Orario Lezioni

dal 16 ottobre al 5 dicembre 2003

**Giovedì dalle 11:00 alle 13:00
dalle 15:00 alle 17:00 ****

**Venerdì dalle 8:30 alle 10:00 *
dalle 12:00 alle 13:00**

**Le ore del *venerdì* dalle 8:00 alle 10:00 si spostano al
mercoledì dalle 11:00 alle 13:00 nei seguenti giorni:**

Orario Ricevimento

**Mercoledì
dalle 14:30 alle 16:30**

**2° Piano stanza no.9
*Polo d'Informatica***

Materiale del corso

Pagina Web

<http://dmi.unicam.it/merelli/calcolo04>

Obiettivi Formativi

- Analizzare le principali caratteristiche dei sistemi distribuiti
- Progettare applicazioni software complesse in moderni ambienti distribuiti
- Utilizzare i concetti fondamentali che caratterizzano i vari modelli di calcolo distribuito e di coordinazione
- Classificare, valutare e applicare alcuni linguaggi per la coordinazione
- Affrontare in modo autonomo la progettazione di applicazioni distribuiti

Testi di Riferimento

TR1: Andrew Tanenbaum,
Distributed Systems,
Prentice Hall, 2002

TR3: C. Lee, A. Omicini ed.,
*Process Coordination and
Ubiquitous Computing*,
CRC Press, 2002

TR2: Wolfgang Emmerich,
*Engineering Distributed
Objects*, Wiley, 2000

Note: *Articoli e appunti
distribuiti durante le
lezioni*

Programma del Corso - Teoria

Introduzione ai sistemi distribuiti

Modello client-server
Paradigmi di comunicazione
Tipi di processi
Naming, synchronization, consistency and
replication, fault tolerance, security
P2P
Sistemi distribuiti basati sulla coordinazione (TIB
Rendezvous-bus Jini)
TR1

Calcolo distribuito

Algebre di processo e process behaviours
CCS
CSP
Petri Nets
UML Activity diagrams

Modelli di coordinazione

Linda e derivati

Linguaggi di coordinazione

Linda e derivati

Middleware per la coordinazione

Object-based
Agent-based

Agenti software e coordinazione

KLAIM
Lime
MARS
TUCSON

Architetture Software

Client-server
Peer-to-peer
Master-worker
...

Modalità d'esame

L'esame consiste nello sviluppo di un progetto e in un seminario per la presentazione dei risultati ottenuti

Per sostenere l'esame è necessario iscriversi all'esame tramite pagina web:
web.unicam.it/matinf

Appelli d'esame

I periodo

lunedì ... dicembre 2003
lunedì 12 gennaio 2004

II Periodo

lunedì 22 marzo 2004 ore 9:00

III Periodo

lunedì 14 giugno 2004
lunedì 12 luglio 2004

Sessione Recupero

lunedì 27 settembre 2004

Coordination

Coordination is a key concept for studying the activities of complex dynamic systems

Coordination is managing dependencies between activities

Such a definition implies that all instances of coordination include agents performing activities that are interdependent [Malone and Crowston 94]

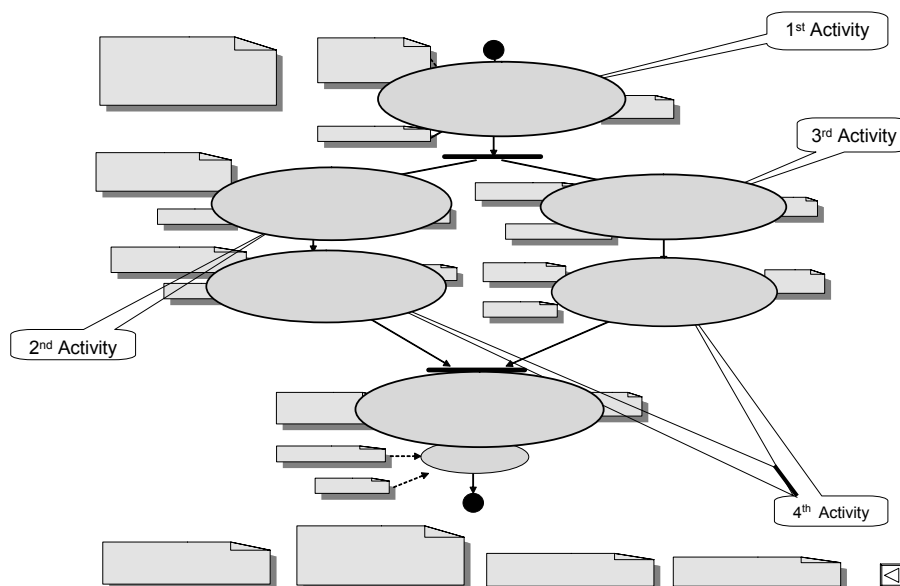
Due to its fundamentality, this notion covers a lot of facets, for instance in distributed artificial intelligence, robotics, biology, and organisational sciences

Coordination is the process of building programs by gluing together active pieces [Carriero and Gelernter 92]

Active pieces here can mean processes, objects with threads, agents, or whole applications

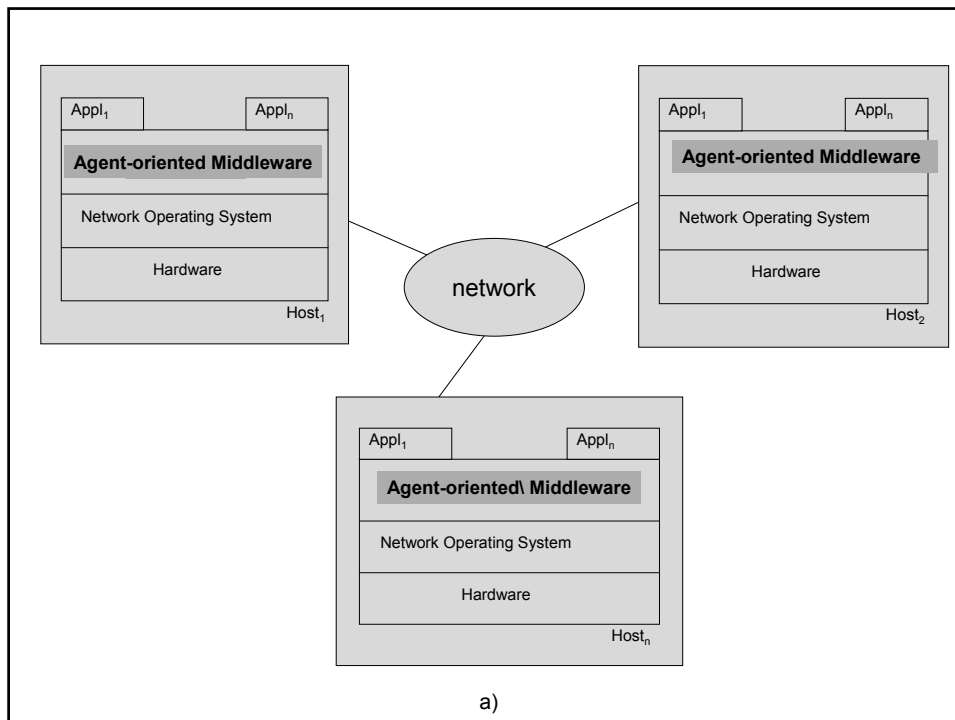
programming = coordination + computation

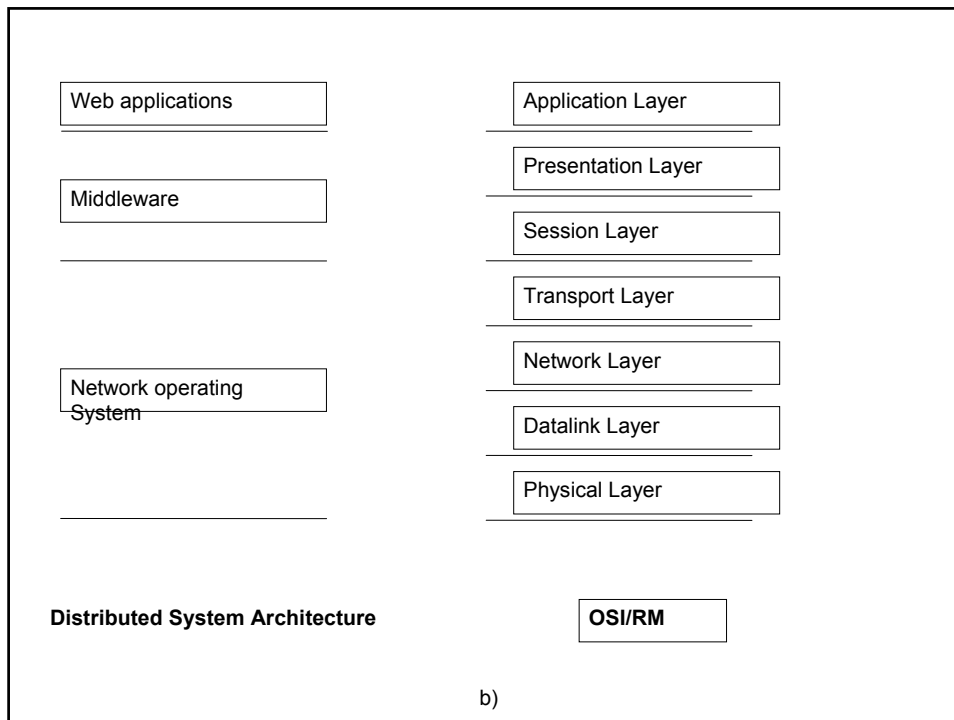
Workflow-based Activity Coordination



Internet and Coordination

The Internet can be enhanced by coordination middleware to a programmable platform offering support for large-scale groupware, agent-based applications, and high-performance computational services (eg. the GRID)





What is a Coordination Model?

A coordination model is the formal basis (semantics) for a coordination language;

It proved that it is possible to “think in a coordinated way” abstracting from low level mechanisms for concurrent, parallel, or distributed programming

A *coordination model* is an abstract (semantic) framework useful to study and understand problems in designing concurrent and distributed programs

“A coordination model is the glue that binds separate activities into an ensemble” [CarGel92]

Historically, Linda was introduced as a new model for parallel programming, more flexible and high level wrt its competitors (“Linda Is Not aDA”)

What is a Coordination Model? Cont.

In other words, a coordination model provides a minimum framework in which the interaction of individual agents can be expressed

Coordination framework covers the issues of dynamic agent creation and destruction, control of communication flows among agents, control of spatial distribution and mobility of agents, as well as synchronization and distribution of actions over time

What is coordination language?

A Coordination language is the linguistic embodiment of a coordination model, and consists of some coordination mechanisms that are added to an *host* (sequential) language

In practice, a coordination language includes clearly defined mechanisms for communication, synchronization, distribution, and concurrency control.

References

[Malone and Crowston 94]

T. Malone and K. Crowstone, *The interdisciplinary study of coordination*, ACM Computing Survey, vol.26, no.1, pp.87-119, 1994

[Carriero and Gelernter 92]

N. Carriero and D. Gelernter, *Coordination languages and their significance*, Communication of ACM, vol.35, no.2,

Digital Library

- ACM
 - www.acm.org
 - User-id: emerelli
 - Pwd: costanza

- IEEE
 - www.computer.org
 - User-id: miaweb
 - Pwd: memacu02

A case study

- Distributed Applications in Bioinformatics Domain
- Activity-based applications
- Tool integration distributed middleware
- Agent-based coordination model

The Tool Integration Problem in Activity-Based Applications

Problem: To integrate and coordinate multiple software tools for retrieving and integrating heterogeneous, distributed and frequently redundant data

Objective: To integrate and coordinate several software tools in order to provide a uniform way and an high level of abstraction for users

Aim: To define an *integrated environment* freeing the user from the need to know details on data repository and to coordinate the intermediate steps of an experiment (tasks)

Proposed Approach: To define an application as a workflow of tasks; to coordinate the execution of cooperative tasks by using software agent tools

Activity-Based Application in Bioinformatics Domain

Problem: To find the crystallographic structure of the 10 proteins more similar to that of a given genetic sequence, e.g., *atggag ... tga*,

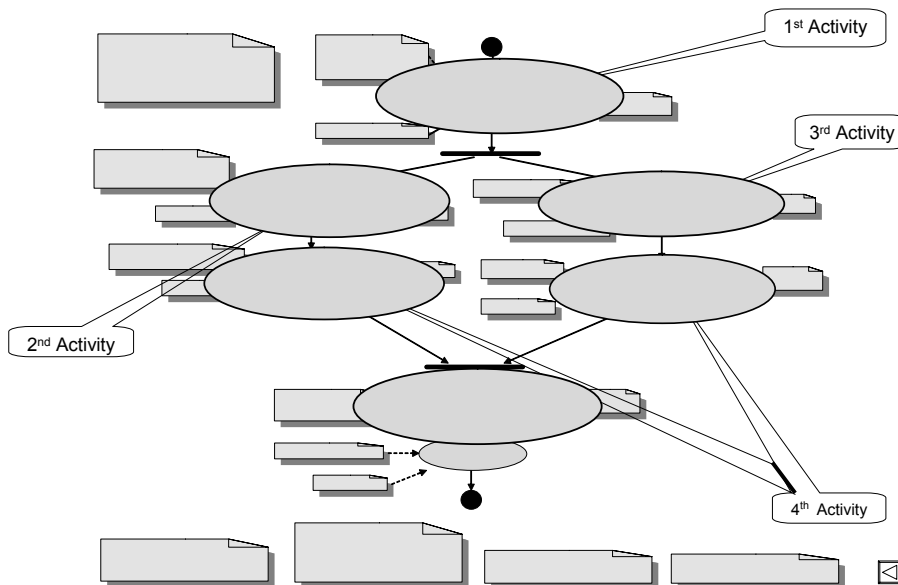
Objective: To use several Bioinformatics Software Tools available on Internet in order to find the wanted result

1. Select the 10 proteins more similar to the given sequence *atggag ... tga*
 - by using BLASTn in GenBank at [NCBI](#) 1st Activity
2. Search for the PDB ID (crystallographic structure identifier) of each selected proteins. 2nd Activity
 - by using BLASTp in SWISS-PROT at EMBL-EBI
 - by retrieving from PubMed via [Entrez Retrieval System](#) at NCBI, abstracts containing PDB-ID information
3. Search for the Crystallographic Structure of any selected PDB ID 3rd Activity
 - find 3-D biological macromolecular structure in [Protein DataBank](#) repository

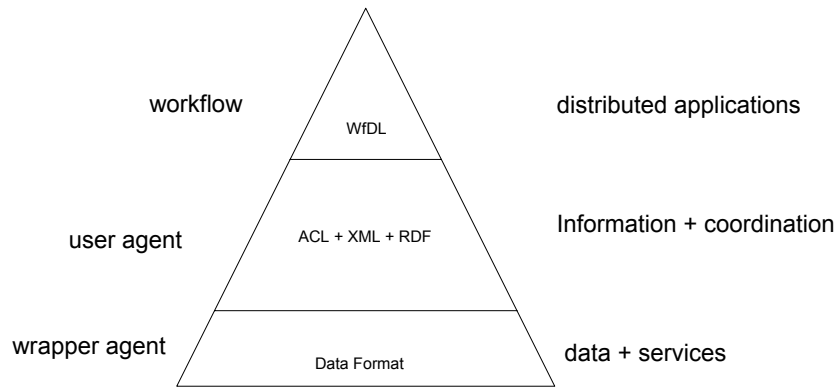
Aim: the automatic execution of the activities by coordinating several Bioinformatics tools, freeing the Bioscientist from the need to know details on data and to continuous interact with remote sites.

By defining an *integrated environment* supporting activity coordination data and services integration large scale distributed management system in order to automatically execute an experiment

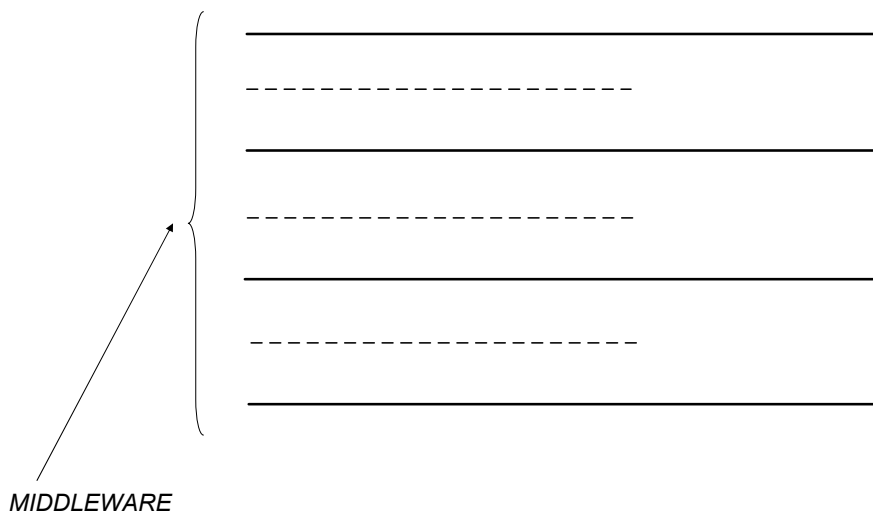
Workflow-based Activity Coordination



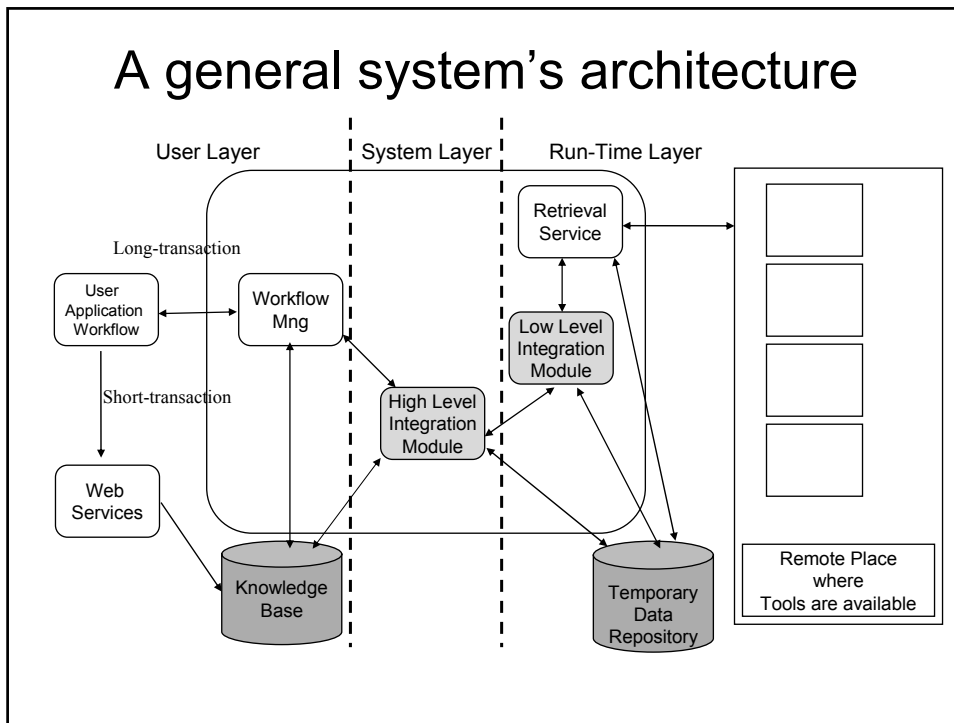
From Data to Knowledge and vice versa



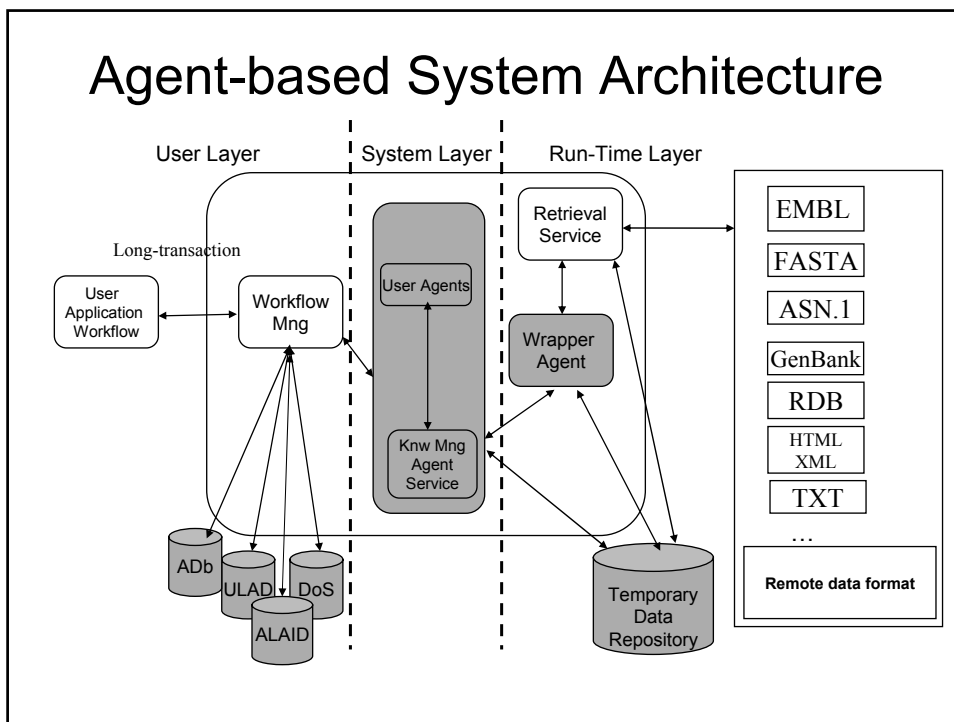
Software Architecture



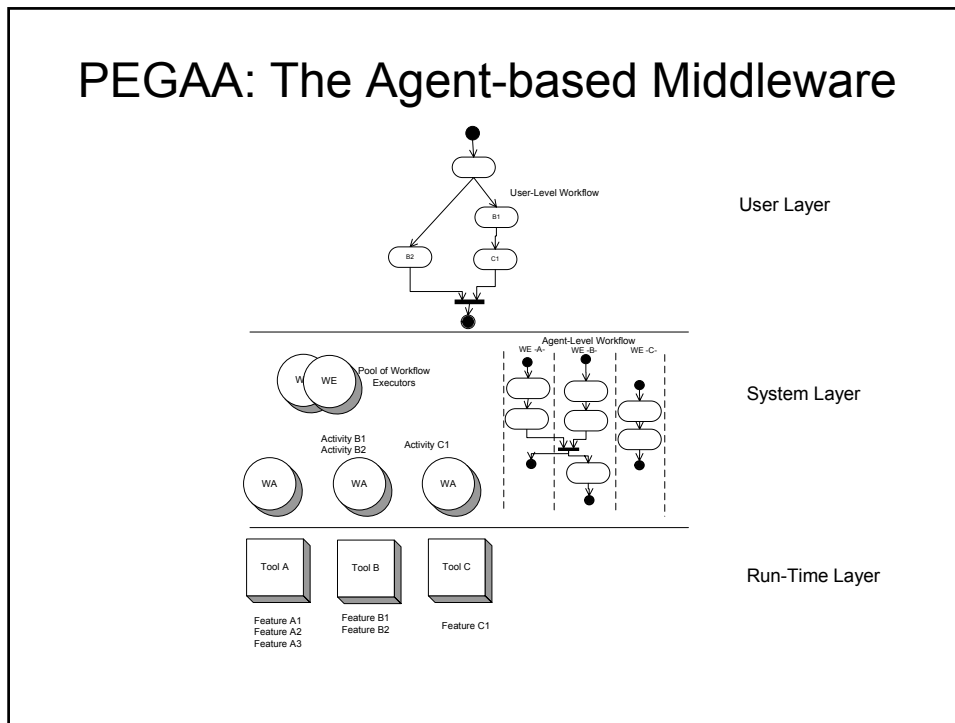
A general system's architecture



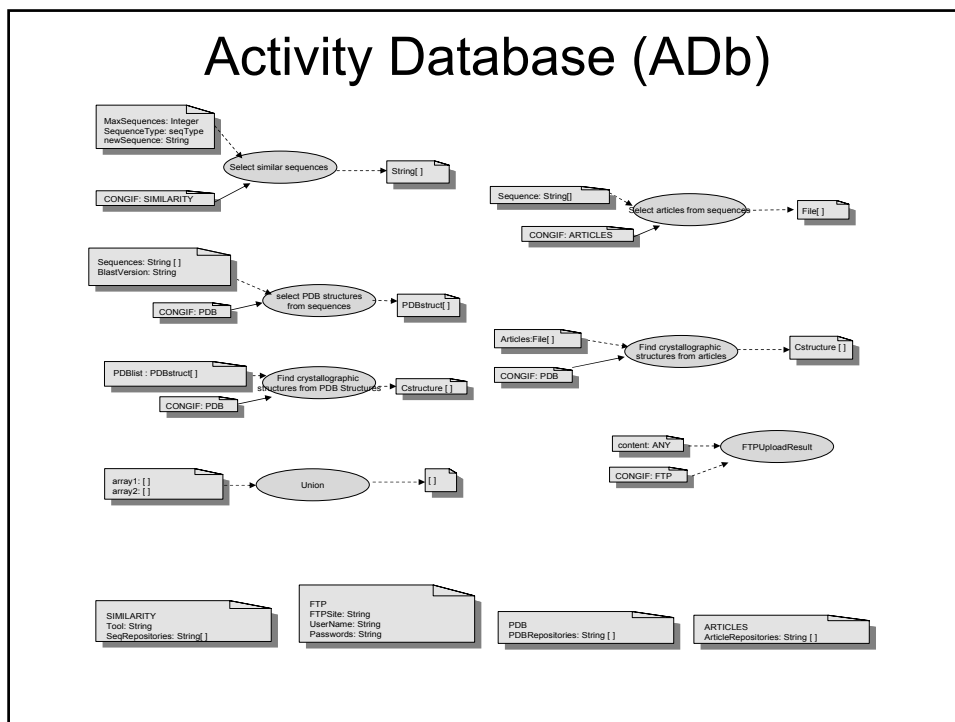
Agent-based System Architecture



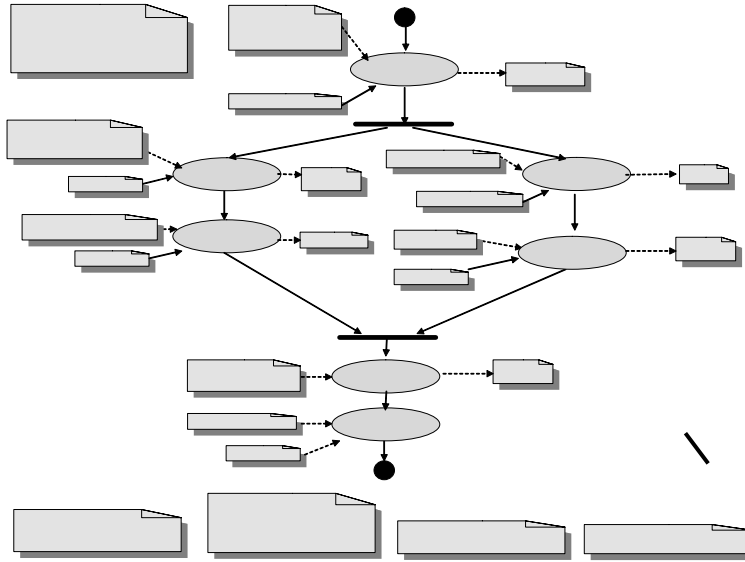
PEGAA: The Agent-based Middleware



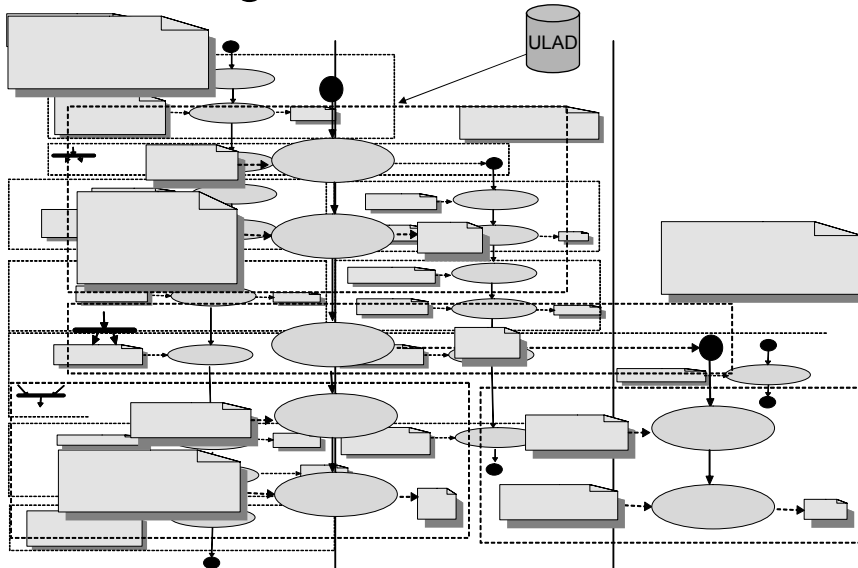
Activity Database (ADb)



User Level Workflow



Agent level workflow



GLOBAL VARIABLES
 sequence: String []
 PDBs: PDBstruct []
 Cstructures: Cstructure
 articles: File []

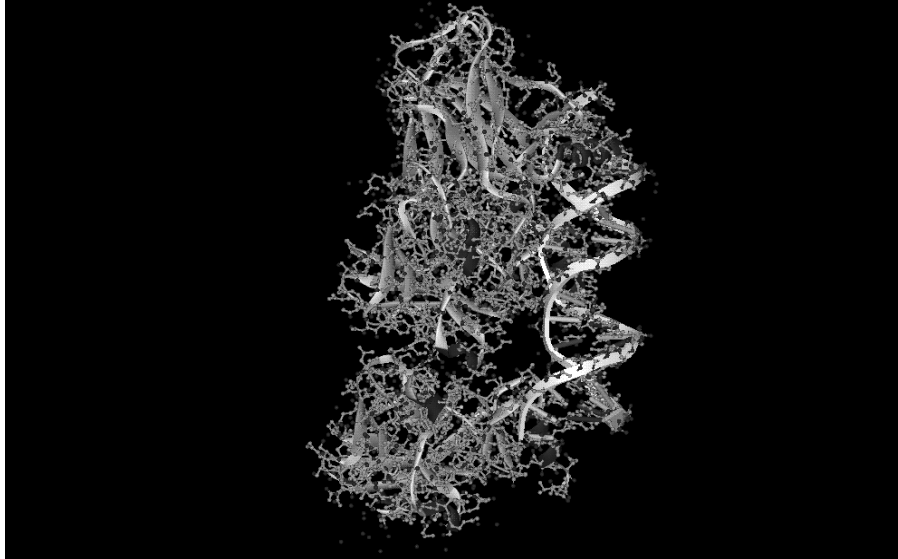
Sequences=sequence
 BlastVersion= "Blastp"

CONGIF:
 CONGIF

PDBlist = PDBs
 PDBlist = PDBs

CONGIF
 CONGIF

P53 Crystallographic Structure

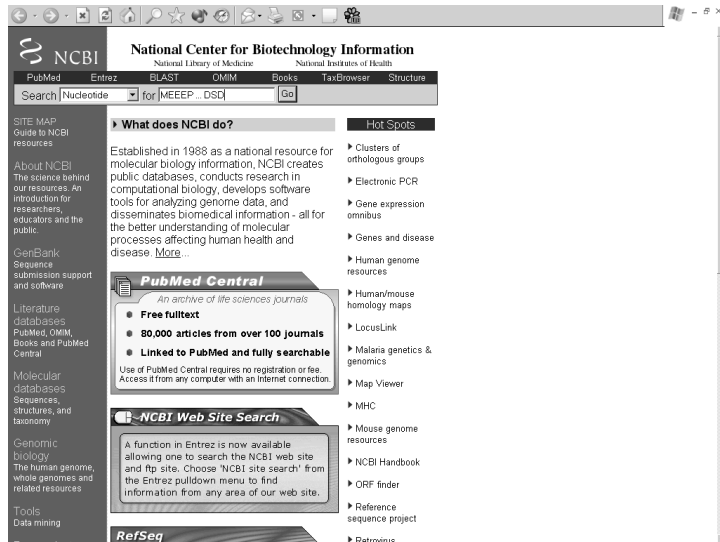


A case study

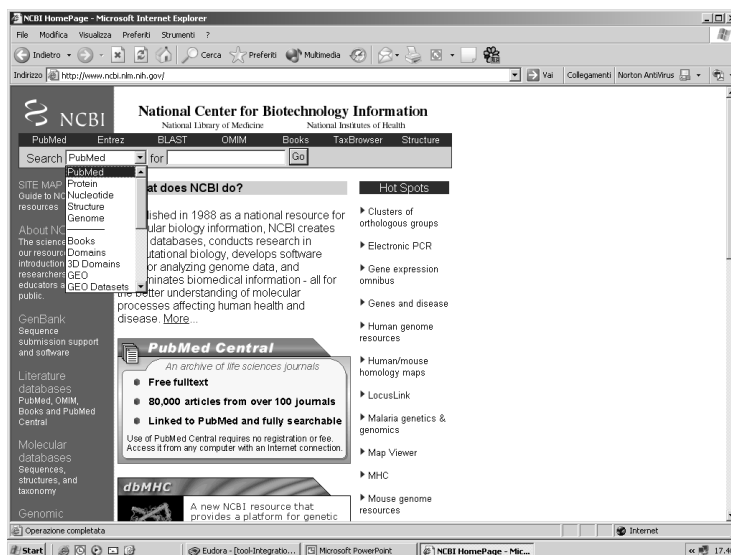
- Distributed Application in Bioinformatics Domain
- Activity-based application
- Tool integration distributed middleware
- Agent-based coordination model

- What for “coordination language?”

NCBI – Home page



NCBI – main databases



NCBI – site map

This site map is also a **guide to NCB resources**. Each link leads to a **brief description of the resource** on this page, then to the resource itself. A **Quick Links** table is also available. It provides only an **alphabetical list of the major resources with direct links** to those resources, bypassing the descriptions.

RESOURCE CATEGORIES

About NCB
 programs and services, NCB handbook, what's new, NCB news, postdoctoral fellowships, organizational structure, contact information, e-mail lists, site search

GenBank
 quickly submit sequences, submit genomes, sample record, GenBank divisions, statistics, release notes, international collaboration, FTP, GenBank

Molecular Databases
 nucleotides, proteins, structures, taxonomy

Literature Databases
 PubMed, PubMed Central, OMM, Books, Citation Matcher

Genomes and Maps
 organism collections (including Entrez Genomes, Map Viewer, and UniGene), human, mouse, rat, cow, zebrafish, *Drosophila*, nematode, plant genomes, yeast, malaria, microbial genomes, viruses, viroids, plasmids, eukaryotic organelles

Tools
 Entrez, LinkOut, Cubby, BLAST, nucleotide sequence analysis, protein sequence analysis, 3-D structure display and similarity searching

Research at NCB
 Computational Biology Branch (CBB), senior

ALPHABETICAL INDEX WITH LINKS TO RESOURCE DESCRIPTIONS
 (To bypass descriptions, use the Quick Links table.)

About NCB	Genomes and Maps	PubMed
ASN.1	GEO (Expression)	PubMed Central
BankIt	Glossaries	RefSeq
BLAST	HTGs	Research at NCB
Books	HomoloGene	Retroviruses
CDART	Human Genome Resources	SAGEmap
CDD	Human/Mouse Homology Maps	Science Primer
CGAP	LinkOut	Seminars
Clones	LocustLink	Seqin
Cn3D	Malaria	Site Search
Coffee Break	Map Viewer	SKYDGH
COGs	MISC	Software Engineering
Computational Biology Branch	Microbial Genomes	Spidey
dbEST	MIMDB	Structures
dbGSS	Model Maker	Submit Data
dbMIM	Multiple Database	Transcript

NCBI - Entrez

Entrez is a retrieval system for searching several linked databases. It provides access to:

- PubMed: Biomedical literature
- Nucleotide: sequence database (GenBank)
- Protein: sequence database
- Structure: three-dimensional macromolecular structures
- Genome: complete genome assemblies
- Bookshelf: online books
- Conserved Domains (CD):
- 3D Domains: domains from Entrez Structure
- GEO: Gene Expression Omnibus
- GEO Datasets: curated GEO data sets
- Journals: journals in Entrez
- MEDLINE: medical subject headings
- NCBI Web Site: NCBI Web site search
- OMIM: Online Mendelian Inheritance in Man
- PMC: full-text digital archive of life sciences journal literature
- PopSet: population study datasets
- SNP: single nucleotide polymorphisms
- Taxonomy: organisms in GenBank
- UniGene: gene-oriented clusters of transcript sequences
- UniSTS: markers and mapping data

Pre-computed similarity searches are available for most database records, which produce a list of related sequences, structure models, as well as related articles.

PDB –Home page

DEPOSIT data
DOWNLOAD files
Browse LINKS
DATA TEST: new features
DATA TEST: new features
DATA TEST: new features

Current Holdings
21838 Structures
Last Update: 22-Jul-2003
PDB Statistics

Molecule of the Month:
Src Tyrosine Kinase

The Protein Data Bank (PDB) is operated by Rutgers, The State University of New Jersey; the San Diego Supercomputer Center at the University of California, San Diego; and the Center for Advanced Research in Biotechnology of the National Institutes of Standards and Technology – three members of the Research Collaboratory for Structural Bioinformatics (RCSB). The PDB is supported in part by the National Science Foundation, the Department of Energy, and two units of the National Institutes of Health: the National Institute of General Medical Sciences and the National Library of Medicine.

Welcome to the PDB, the single worldwide repository for the processing and distribution of 3-D biological macromolecular structure data.

Search the Archive
Enter a **PDB ID** or **keyword**
Query Tutorial
[TSR] Find a structure
 query by PDB ID only match exact word
 remove sequence homologs

PDB Mirrors
Please bookmark a mirror site
San Diego Supercomputer Center
Rutgers University
National Institute of Standards and Technology
Cambridge Crystallographic Data Centre, UK
National University of Singapore
Osaka University, Japan
Universidade Federal de Minas Gerais, Brazil
Max Delbrück Center for Molecular Medicine, Germany

News
Complete News Newsletter
pub | Archive Subscribe

22-Jul-2003
Demonstrations, Posters, and More: PDB at the ACA Annual Meeting and the IUB Symposium of the Protein Society: The PDB would like to thank those ISMB 2003 attendees who provided valuable feedback at our demonstration session and exhibit during this wonderful event. PDB staff members will also participate in several other meetings in the near future, including the annual meetings of the Protein Society and the American Crystallographer Association. [MORE...]

OTHER SITES
*RCSB partner

In citing the PDB please refer to:
H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Weissig, I.N. Shindyalov, P.E. Bourne. The Protein Data Bank. *Nucleic Acids Research*, 28 pp. 235-242 (2000)

NCBI - BLAST

Basic Overview

- [What is BLAST?](#)
- [What is the best BLAST for me?](#)
- [How can I access the BLAST services at NCBI?](#)
- [What is the Standalone WWW BLAST Server?](#)

Q: What is BLAST?

BLAST® (Basic Local Alignment Search Tool) is a set of similarity search programs designed to explore all of the available sequence databases regardless of whether the query is protein or DNA. The BLAST programs have been designed for speed, with a minimal sacrifice of sensitivity to distant sequence relationships. The scores assigned in a BLAST search have a well-defined statistical interpretation, making real matches easier to distinguish from random background hits. BLAST uses a heuristic algorithm that seeks local as opposed to global alignments and is therefore able to detect relationships among sequences that share only isolated regions of similarity (Altschul et al., 1990). For a better understanding of BLAST, you can refer to the [BLAST Course](#), which explains the basics of the BLAST algorithm. There is also a description of BLAST services located [here](#). Also for details on BLAST and the theory of similarity search, see the [References](#) section.

Q: What is the Best BLAST for me?

There are many different types of BLAST available from the main [BLAST web page](#). Choosing the right one depends on the type of sequence you are searching with (long, short, nucleotide protein), and the desired database. In order to help guide you to the right service for your search please consult the "[BLAST Program Selection Guide](#)".

This guide provides information on what database and BLAST service (megaBLAST, translation BLAST etc) is ideal for your search. Plus there is also additional information describing how the tools work and why they are best for certain types of searches.

Q: How can I access the NCBI's BLAST services?

- Sistemi distribuiti e coordinazione
- Modelli di coordinazione
- Architetture software
- Linguaggi di coordinazione
- Middleware per la coordinazione
- Agenti software e coordinazione

Sistemi distribuiti e coordinazione

- Coordination-based systems
 - TIBCO's Rendezvous Bus
 - Sun Microsystem's Jini

Articoli

La parte di coordinazione di un sistema distribuito gestisce la comunicazione e la cooperazione tra processi. Rappresenta “glue” che lega le attività dei processi come se fosse unica.

- D. Gelemter and N. Carriero, *Coordination languages and their significance*, *Communication of ACM*, vol.35, no.2, pp.96-107, Feb. 1992

Una tassonomia di modelli di coordinazione per agenti mobili, generalizzabile ad altri tipi di sistema distribuito

- G. Cabri, L. Leonardi and F. Zambonelli, *Mobile-Agent coordination models for Internet Applications*, *IEEE Computer*, vol.33, no.2, pp.82-89, feb. 2000

Modelli di coordinazione

- Direct coordination
- Meeting-oriented coordination

Architetture software e coordinazione

Linguaggi di coordinazione

- LINDA

Middleware per la coordinazione

Agenti sw e coordinazione

- CCS: un calcolo di base per la concorrenza, semantiche operazionali ed osservazionali, assiomatizzazioni
- Pi-calcolo: un calcolo con mobilità di interconnessioni,
- semantiche operazionali ed osservazionali, sistemi di tipo.
-