STARS

Semistructured Temporal clinicAl GeogRaphical Systems

17 November, 2010

Outline

Outline

Who we are

Carlo Combi
Full Professor
(Head of Computer Science Department)

Alberto Belussi
Associate Professor

Barbara Oliboni
Researcher

Roberto Posenato
Researcher

Pietro Sala
Post-Doc

Mauro Gambini
Ph.D. Student

Sara Migliorini
Ph.D. Student

Gabriele Pozzani
Ph.D. Student

Vladan Mijatovic
Research Associate

What we do

The group focuses on the

- theoretical
- technological
- methodological

aspects of the information management in complex organizations.

The research activities comprise:

- information systems on the web
- temporal information systems
- information systems in medicine
- geographical information systems
- semistructured databases
- process-aware information systems

Where we work

The STARS lab is located at the floor -2 in Ca' Vignal 2.

It's a lab where students collaborating with the group can develop their projects and theses.

The lab has 10 workspaces and includes 5 PCs equipped with Ubuntu and any software useful for the research and thesis work, e.g., PostgreSQL, MySQL, Java SDK, Pentaho.

The PCs in the lab are connected with the STARS server that manages the users' authentication and disk space.

Outline

Provided courses

People in the research group (will) give the following courses:

- databases for bioinformatics
- information systems in medicine
- databases
- multimedia and geographical information systems
- biomedical and bioinformatics databases
- information systems
- advanced algorithms
- complexity

Databases for Bioinformatics

Degree in Bioinformatics (Laurea in bioinformatica)

Carlo Combi

The class focuses on:

1. the design of a database and its applications, including:
   - conceptual and logical design of a database
   - the query language SQL and the relational algebra
2. the design of data-intensive web applications, including:
   - ways to interact with a DBMS
   - the MVC model
3. the management of bioinformatics information in databases:
   - XML for bioinformatics

The course includes also lab practices about:

- the use of PostgreSQL/PostBio
- implementation of web applications based on Servlet

Distribution of positive results													
18	19	20	21	22	23	24	25	26	27	28	29	30	30 cum laude
0%	0%	3%	3%	9%	9%	9%	13%	6%	6%	13%	13%	6%	9%
Data from AA 2009-2010 based on 67 students.													

Information systems in medicine

Master Degree in Bioinformatics and Medical Biotechnologies (Laurea magistrale in bioinformatica e biotecnologie mediche)

Carlo Combi

The class focuses on:

- temporal modeling, reasoning, and databases
- temporal reasoning and maintenance in medicine:
  - temporal clinical databases
  - abstraction of time-oriented clinical data
- time in clinical tasks:
  - time in clinical diagnosis
  - automated support to clinical guidelines and care plans
- the display of time-oriented clinical data and knowledge

Databases

Degree in Computer Science (Laurea in informatica)

Alberto Belussi

The course focuses on two main arguments:

1. the design of a database and its applications, including:
   - conceptual and logical design of a database
   - the query language SQL and the relational algebra
2. the design of data-intensive web applications, including:
   - ways to interact with a DBMS
   - the approach “MVC-2 Servlet centric”
   - the treatment of multimedia datasets in a DBMS

The course includes also lab practices about:

- the use of PostgreSQL
- implementation of web applications based on Servlet and JSP

Distribution of positive results													
18	19	20	21	22	23	24	25	26	27	28	29	30	30 cum laude
0%	0%	0%	14%	29%	14%	29%	0%	0%	0%	0%	0%	0%	14%
Data from AA 2009-2010 based on 19 students.													

Multimedia and geographical information systems

Master Degree in Engineering and Computer Science
(Laurea magistrale in ingegneria e scienze informatiche)

Alberto Belussi

The course provides the concepts and methodologies for the design and implementation of spatial and multimedia databases:

- object-oriented databases: ODMG, ODL, and OQL standards
- conceptual modeling of a geographical database in GeoUML
- the mappings of a GeoUML conceptual schema onto geo-relational databases
- an extended version of the relational algebra for querying a geo-relational database

Biomedical and bioinformatics databases

Master Degree in Bioinformatics and Medical Biotechnologies (Laurea magistrale in bioinformatica e biotecnologie mediche)

Barbara Oliboni

The class focuses on:

- biomedical databases:
  - design
  - query languages for biomedical DBs
  - an example: medical records and their querying
- bioinformatics databases:
  - management of semi-structured data
  - techniques for information retrieval on semi-structured data
  - querying of bioinformatics databases
  - XML and bioinformatics

Information systems

Master Degree in Engineering and Computer Science (Laurea magistrale in ingegneria e scienze informatiche)

Barbara Oliboni

The class focuses on the design and use of information systems within complex organizations:

- structure and functions of an information system:
  - economic, organizational, and management issues
  - business process re-engineering
  - ERP, MRP, and CRM systems
- decision support systems:
  - data warehouse systems, OLAP and data mining techniques
- designing cooperative information systems:
  - workflow systems
- distributed information systems

Distribution of positive results													
18	19	20	21	22	23	24	25	26	27	28	29	30	30 cum laude
0%	13%	0%	0%	0%	0%	13%	0%	13%	0%	25%	0%	13%	25%
Data from AA 2009-2010 based on 13 students.													

Advanced algorithms

Master Degree in Engineering and Computer Science (Laurea magistrale in ingegneria e scienze informatiche)

Roberto Posenato

“Advanced algorithms” is the first module of the “Algorithms” course.

The module introduces some advanced paradigms for algorithms development and analysis:

- greedy paradigm
- backtracking technique
- branch & bound technique
- dynamic programming paradigm
- approximations algorithms
- probabilistic algorithms

Complexity

Master Degree in Engineering and Computer Science (Laurea magistrale in ingegneria e scienze informatiche)

Roberto Posenato

"Complexity" is the second module of the "Algorithms" course.

The module introduces computational complexity theory, the NP-completeness theory, and computational analysis of problems w.r.t. their approximability:

- computational models
- time complexity
- space complexity
- reductions and completeness

Distribution of positive results													
18	19	20	21	22	23	24	25	26	27	28	29	30	30 cum laude
12%	9%	6%	9%	21%	12%	3%	6%	6%	6%	3%	0%	3%	3%
Data from AA 2009-2010 based on 113 students. N.B.: these results comprise all modules in the course													

Outline

Carlo Combi

Research interests include:

- temporal functional dependencies
- (temporal) OLAP and data mining on biomedical/bioinformatics data warehouses
- dealing with clinical guidelines through workflow management systems

Alberto Belussi

Research interests include:

- Conceptual design of geographical databases
  - models for the conceptual design of geographical databases
  - languages for the specification of spatial integrity constraints
  - comparison of conceptual designs
  - validation of spatial data wrt conceptual designs
  - methods for the updating of geographical databases maintaining topological properties
- Approximate spatial queries
  - similarity functions for spatial relations
  - approximate spatial queries
  - optimization methods of approximate spatial queries
  - spatial queries in recommendation systems

Barbara Oliboni (I)

Research topics include:

- XML and Data Warehouses:
  - integration of XML documents in DW: ETL technologies for integrating XML documents in classic DWs
  - XML Data Warehouses: “OLAP-style” querying of native XML DBs
  - XML Warehouses: “OLAP-style” querying of XML documents
- XML and bioinformatics:
  - representation and management of biological data in XML documents
- XML and data mining:
  - association rules for mining XML documents

Barbara Oliboni (II)

- Temporal aspects of semistructured data:
  - definition of temporal data models and query languages
  - definition of constraints for managing time dimensions
- Temporal aspects of XML data:
  - evolution and versioning of XML documents and schemata
  - minimization of sequences of XML document and schema update operations
- Temporal aspects of data warehouses

Roberto Posenato (I)
Temporal Workflow Model Analysis

- A workflow is an automation of a business process during which information or tasks are passed from one participant to another for actions according to a set of constraints in order to realize a business goal.
- A workflow management system (*WfMS*) is a software system that defines, creates and manages the execution of workflows.
- Currently, existing *WfMSs* offer only a limited support for modeling and managing time constraints associated to processes and their activities.

Since 2009, I have been working with Carlo Combi on:

- a proposal of a Time-aware Conceptual Workflow Model (*Temporal Workflow Model*);
- the computational analysis of the controllability of Temporal Workflow Model schemata.

Roberto Posenato (II)
Magnetic Resonance Imaging Tractografy Analysis

- Diffusion Magnetic Resonance Imaging (MRI) is a powerful non-invasively method producing images of biological tissues exploiting the water molecules diffusion into the living tissues under a magnetic field.
- Considering the brain white matter, it is interesting to reconstruct the fiber bundle trajectories of the brain from the MRI data (tractography algorithm).
- Currently, existing MRI mathematical models do not fully consider the density of fibers of a bundle trajectory preventing any evaluation of the real physical dimension of neuronal fiber bundles.

Since 2009, I have been working with A. Daducci, G. Orlandi and A. Marigonda on:

- a proposal of two different mathematical models for MRI in order to fully represent the fiber density information.
- the computational analysis of new resulting tractography algorithms.

Roberto Posenato (III)
Advanced Web Application Framework Analysis

- The term Web 2.0 is commonly associated with web applications that facilitate interactive information sharing, interoperability, user-centered design,[1] and collaboration on the World Wide Web.
- Currently, there is a lot of software frameworks allowing to develop web 2.0 applications in a rapid way following one or more design patterns.
- To the best our knowledge, none of such frameworks provide simple or well-defined patterns to develop data-centric web applications based on already defined databases.

Since 2008, I have been working with A. Belussi on:

- a systematical analysis of the most common frameworks w.r.t the data-centric development issues.
- proposal of simple design patterns for data-centric web application development based on frameworks as Liferay or Portlet technology.

Pietro Sala (I)

Research Interests:

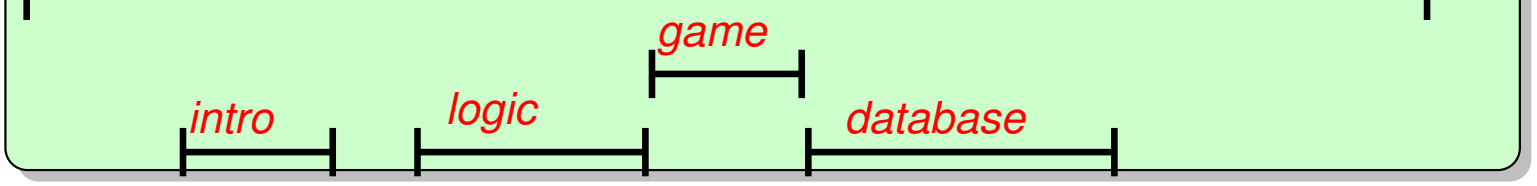
- Logic in Computer Science (decidability, verification, synthesis);
- Game Theory;
- Temporal Databases.

Pietro Sala (I)

Logic

Decidability of interval temporal logic:

$$\langle D \rangle (intro \wedge \langle L \rangle (logic \wedge (\langle A \rangle (game \wedge \langle A \rangle database))))$$

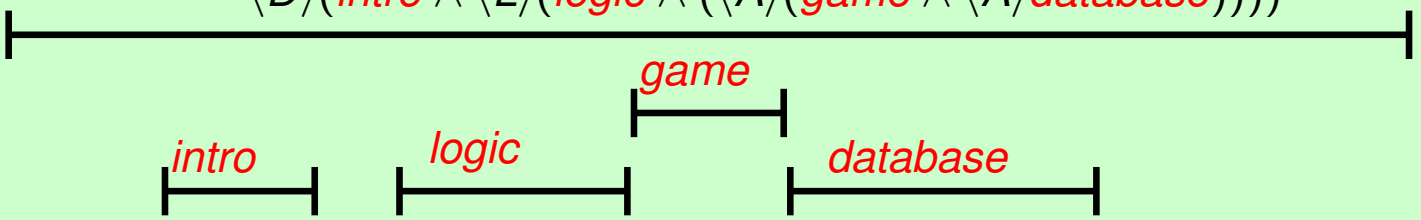


Pietro Sala (I)

Logic

Decidability of interval temporal logic:

$\langle D \rangle (intro \wedge \langle L \rangle (logic \wedge (\langle A \rangle (game \wedge \langle A \rangle database))))$



Publications (this year)

A. Montanari, I. Pratt-Hartmann, P. Sala. Decidability of the logic of the reflexive sub-interval relation over finite linear orders.

D. Bresolin, D. Della Monica, A. Montanari, P. Sala, G. Sciavicco. A decidable spatial generalization of metric interval temporal logic.

A. Montanari, G. Puppis, and P. Sala. Maximal decidable fragments of Halpern and Shoham's modal logic of intervals.

D. Bresolin, P. Sala, G. Sciavicco. Begin, after, and later: a maximal decidable interval temporal logic.

A. Montanari, G. Puppis, P. Sala, G. Sciavicco. Decidability of the interval temporal logic $AB\overline{B}$ over the natural numbers.

Pietro Sala (II)

Game Theory

In recent years the application of Game Theory has spread over many aspects of theoretical and applied computer science.

Pietro Sala (II)

Game Theory

In recent years the application of Game Theory has spread over many aspects of theoretical and applied computer science.

From formal verification to algorithms the benefits of applying game theoretic arguments are well recognized.

Pietro Sala (II)

Game Theory

In recent years the application of Game Theory has spread over many aspects of theoretical and applied computer science.

From formal verification to algorithms the benefits of applying game theoretic arguments are well recognized.

To the best of our knowledge there is no application of game theoretic results to Data Mining in particular regarding the extraction of temporal functional dependencies (TFDs) and we think that is a way which is worth to explore.

Pietro Sala (III)

Temporal Databases

Pietro Sala (III)

Temporal Databases

Together with prof. Carlo Combi we submitted a work regarding the verification of Interval Temporal Functional Dependencies (ITFDs) over a given database.

Pietro Sala (III)

Temporal Databases

Together with prof. Carlo Combi we submitted a work regarding the verification of Interval Temporal Functional Dependencies (ITFDs) over a given database.

ITFDs express constraints over a temporal database considering intervals, instead of points, as basic time entities:

Pietro Sala (III)

Temporal Databases

Together with prof. Carlo Combi we submitted a work regarding the verification of Interval Temporal Functional Dependencies (ITFDs) over a given database.

ITFDs express constraints over a temporal database considering intervals, instead of points, as basic time entities:

Example: "$Contract \rightarrow_{\mathbf{S}} Calling\_Rate$" on $R = (ClientId, Contract, Calling\_Rate)$

Pietro Sala (III)

Temporal Databases

Together with prof. Carlo Combi we submitted a work regarding the verification of Interval Temporal Functional Dependencies (ITFDs) over a given database.

ITFDs express constraints over a temporal database considering intervals, instead of points, as basic time entities:

Example: "$Contract \rightarrow_{\mathbf{S}} Calling\_Rate$" on $R = (ClientId, Contract, Calling\_Rate)$

We have proved that ITFDs are more expressive than other point-based TFDs, we are trying to apply our results to Data Mining in order to extract ITFDs from a given database instance.

Mauro Gambini

Research topic: Computational Models for Process-Aware Information Systems

- Process-Aware Information System (PAIS): software system driven by explicit process models, with the aim to coordinate and support agents in performing their activities
- Formalization of existing process modeling languages
- Process modeling language design and implementation
- Process models verification and correction

In collaboration with:

- Carlo Combi and Sara Migliorini

Sara Migliorini

Research topic: Supporting Distributed Geo-Processing Using Workflow Technology

- GeoProcess: Distributed long-running computation on spatial-related data that rely on loosely-coupled and interoperable services
- Applicability of Scientific and Business Workflow Management Systems in the geographical context
- Implementation of a modeling tool for geographical processes

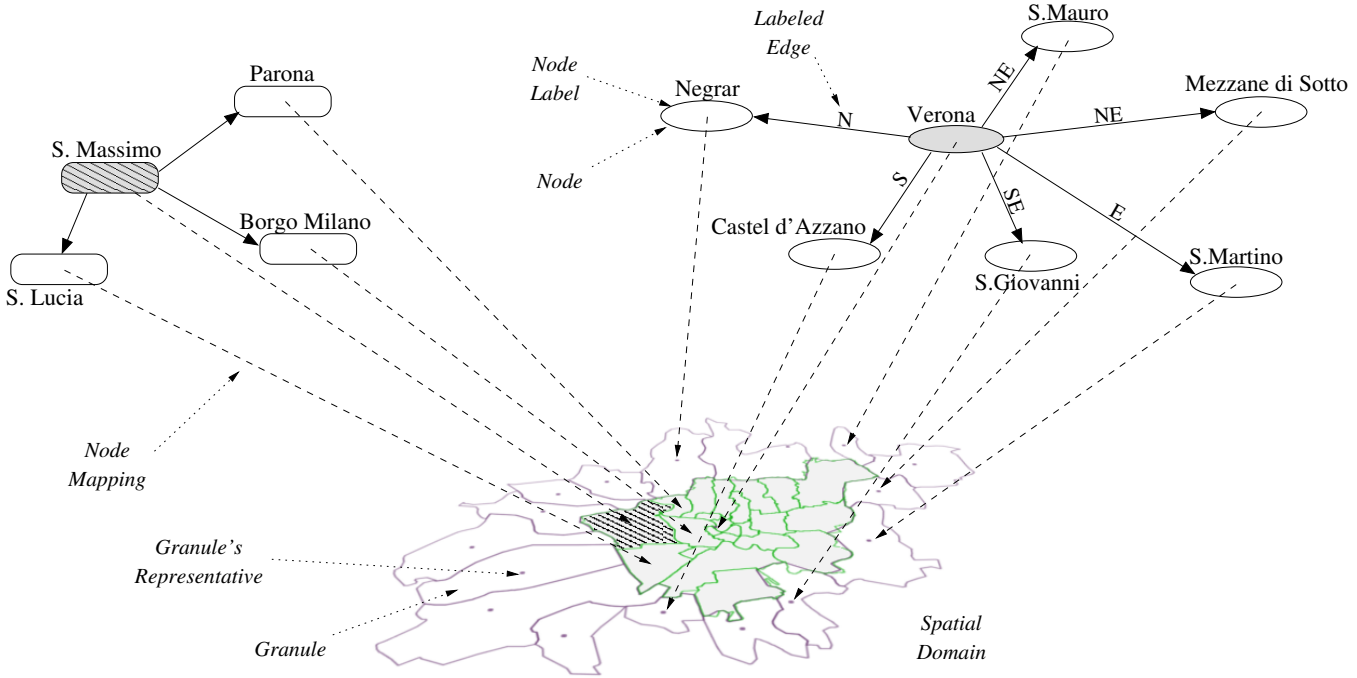
In collaboration with:

- Carlo Combi, Alberto Belussi, and Mauro Gambini

Gabriele Pozzani

Ph.D. thesis (in collaboration with C. Combi and A. Belussi):

- formalization of granularities

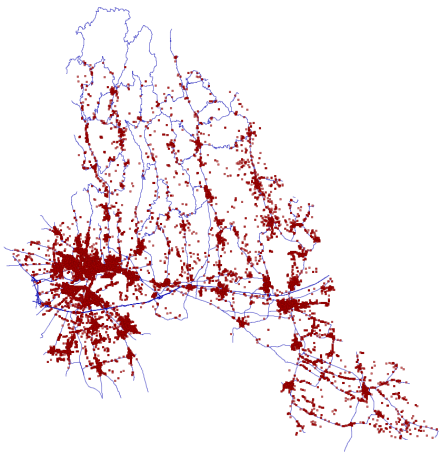


- querying of spatio-temporal databases

Gabriele Pozzani

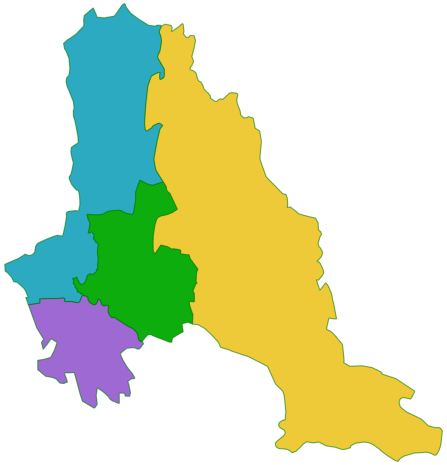
Ph.D. thesis (in collaboration with C. Combi and A. Belussi):

- formalization of granularities
- querying of spatio-temporal databases



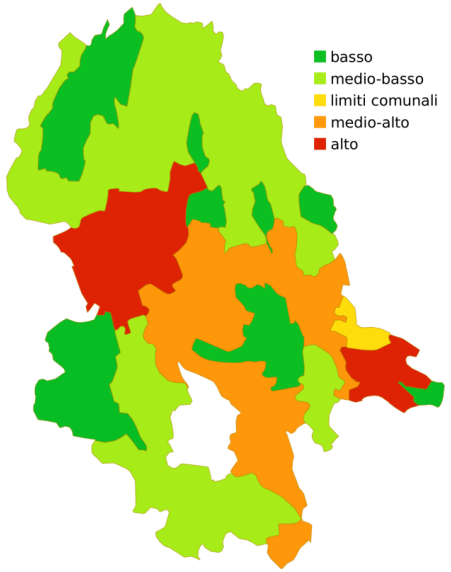
patients' domicile

+



ULSS 20 areas

+



pollution areas

=

area domicilio	livello acustico diurno	num contatti	num totale contatti area	percentuale
1	alto	3830	3830	100.00
2	alto	3370	3370	100.00
3	alto	5672	5938	95.52
3	basso	248	5938	4.18
3	limiti_comunali	6	5938	0.10
3	medio-alto	12	5938	0.20

Vladan Mijatovic

Implementation of BI solutions in clinical and business applications:

- Extraction, Transformation, and Loading (ETL) of data coming from heterogeneous sources
- data mart and data warehouse design
- report design
- OLAP analysis
- dashboard design
- data mining

In collaboration with:

- section of immunology
- section of pharmacology
- NovaRatio s.r.l.
- section of genetics

External projects and collaborations (I)

- XML and time. Versioning of XML Schemata: minimization of operation sequences for the schemata modification
  - B. Oliboni in collaboration with Giovanna Guerrini (University of Genova) and Marco Mesiti (University of Milan)
- XML and time: versioning and evolution of XML schemata
  - B. Oliboni in collaboration with Fabio Grandi (University of Bologna) and Zouhaier Brahmia (University of Sfax, Tunisia)
- Petri Nets Simulated Annealing (PNSA): a genetic algorithm to rectify Petri nets based process models
  - M. Gambini and S. Migliorini in collaboration with Queensland University of Technology (QUT), Brisbane (Australia)
- Comparative analysis of Scientific and Business Workflow Management Systems
  - M. Gambini and S. Migliorini in collaboration with Queensland University of Technology (QUT), Brisbane (Australia)

External projects and collaborations (II)

- Spatial (approximate) query processing and optimization
  - A. Belussi and S. Migliorini in collaboration with Prof.ssa B. Catania (University Of Genova)
- Conceptual modeling of spatial databases
  - A. Belussi and S. Migliorini in collaboration with Prof. G. Pelagatti (Politecnico di Milano)
- REVIVAL - REstauro dell'archivio VIcentini di Verona e sua accessibilità come Audio e-Library
  - B. Oliboni and F. Fontana in collaboration with Archivio Vicentini and Eye-Tech S.r.l.

Publications statistics

Our research work has led to:

- 113 publications in international conferences. 41 since 2005
- 40 journal publications. 10 since 2005
- 17 book chapters. 11 since 2005
- 5 books
- 2 books/proceedings editing

Participation in conference committees

We participate in several international conference committees:

- Conference on Artificial Intelligence in Medicine (AIME)
- ACM International Conference on Information and Knowledge Management (CIKM)
- International Symposium on Temporal Representation and Reasoning (TIME)
- East-European Conference on Advances in Databases and Information Systems (ADBIS)
- International Workshop on Evolution and Change in Data Management (ECDM)
- International Conference and Workshop on Database and Expert Systems Applications (DEXA)
- Flexible Database and Information Systems Technology (FlexDBIST)
- European Symposium on Algorithm

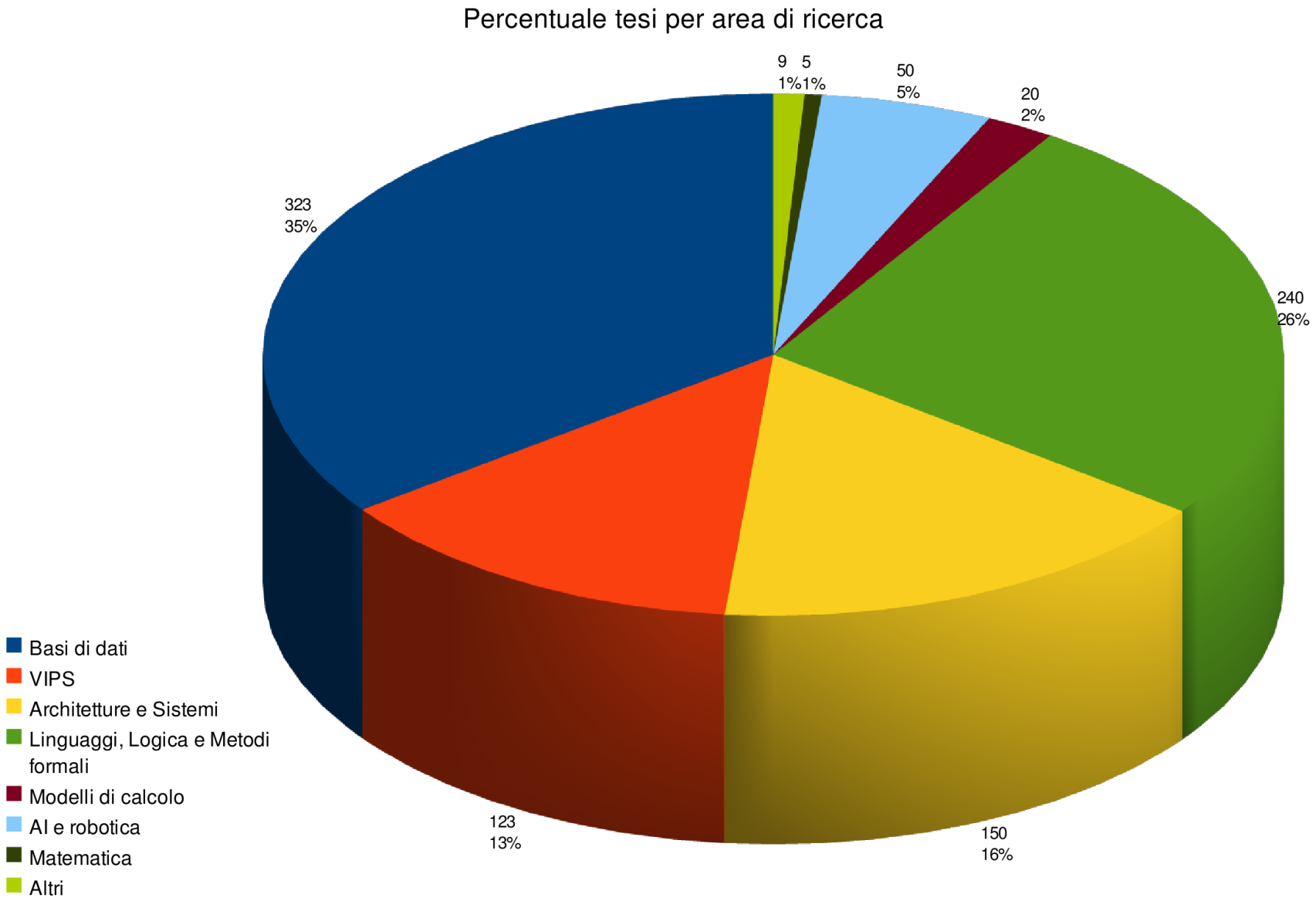
Participation in journal activities

We participate also in several journal activities:

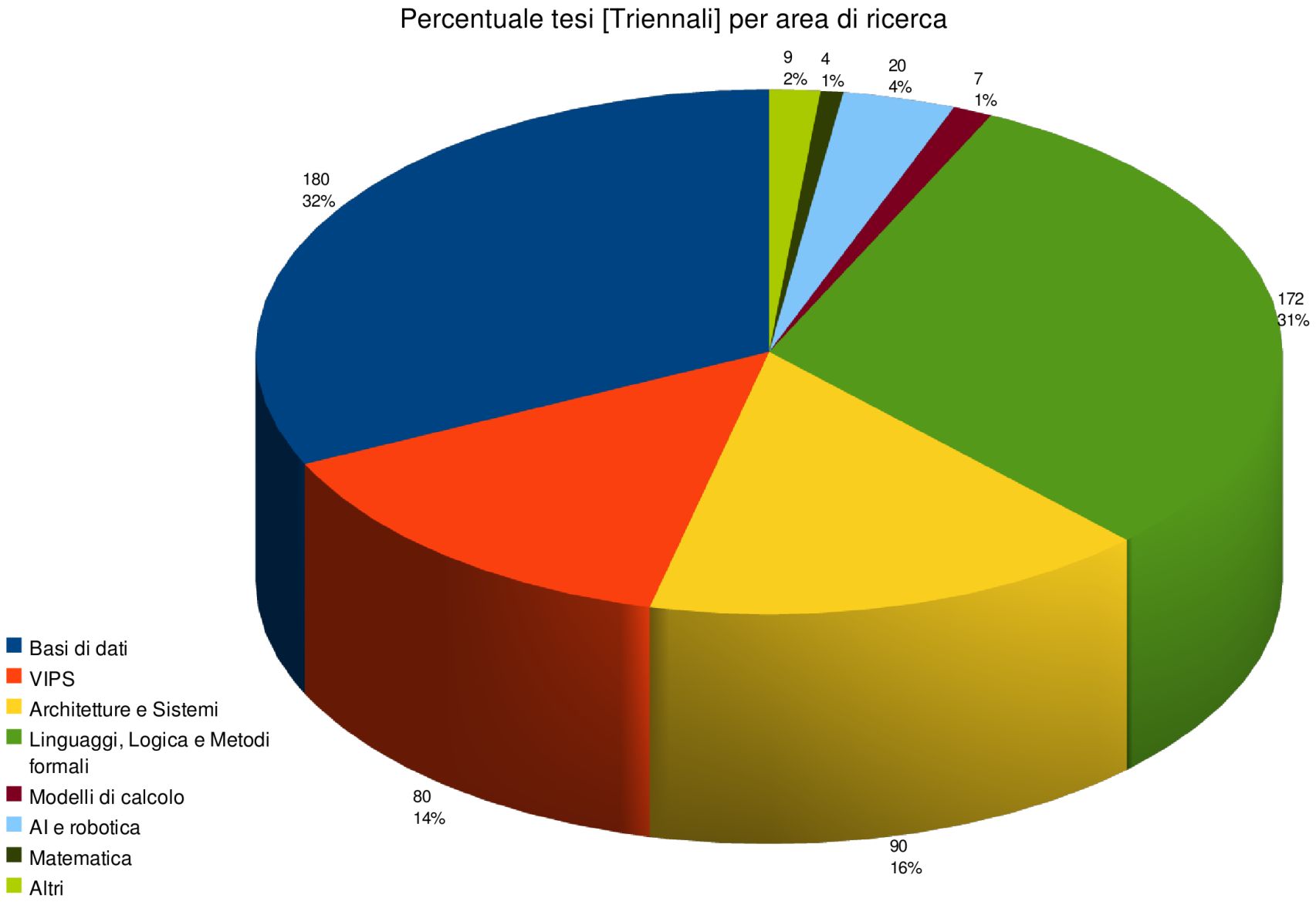
- Information Systems
- Journal of Web Engineering and Technology
- Artificial Intelligence in Medicine
- Data and Knowledge Engineering
- International Journal of Knowledge and Learning
- Methods of Information in Medicine
- Annals of Mathematics and Artificial Intelligence
- ACM Transactions on Autonomous and Adaptive Systems
- IEEE Transactions on Neural Networks

Outline

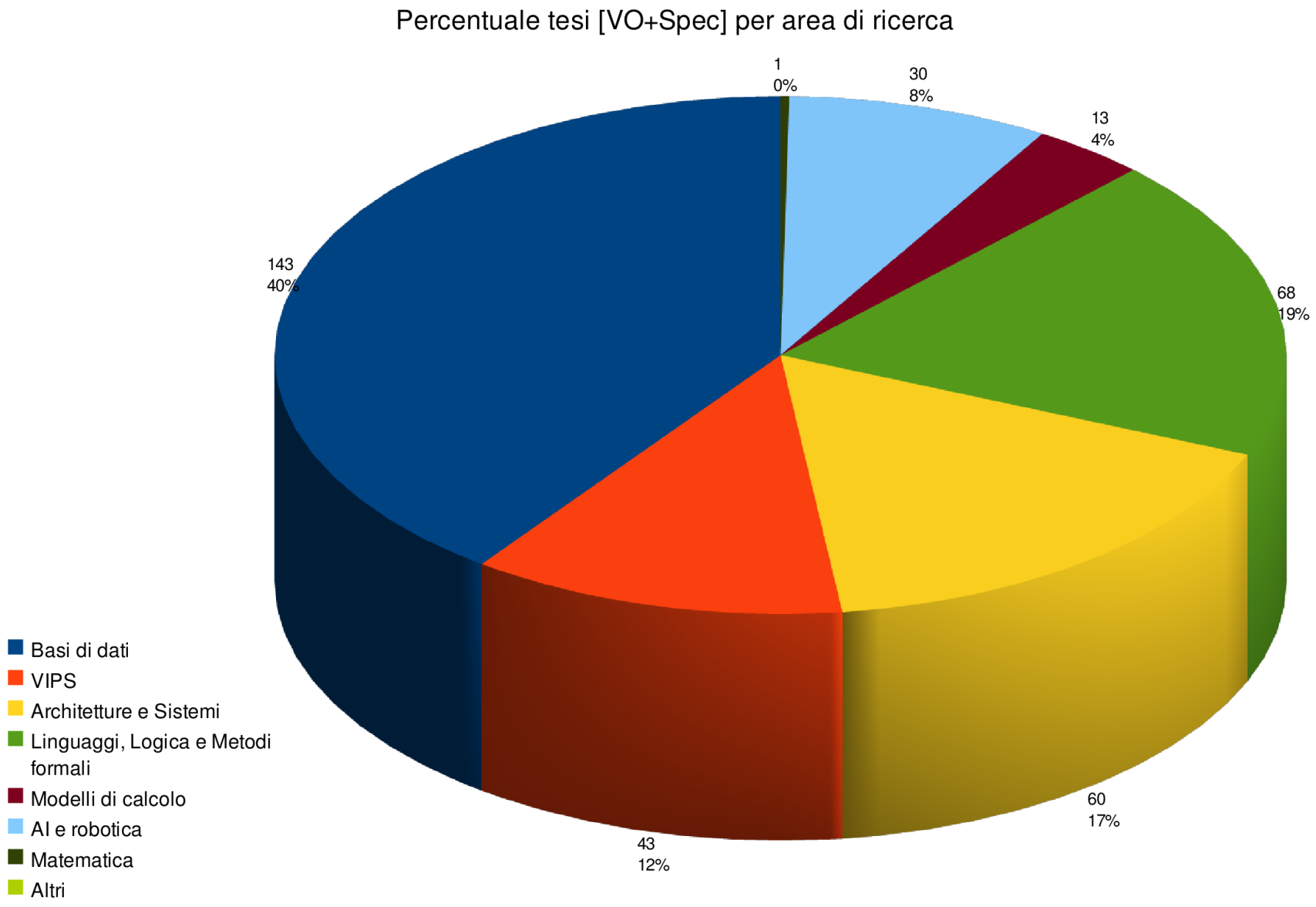
Theses percentage: 2004 – Oct 2010



Bachelor theses percentage: 2004 – Oct 2010



Master theses percentage: 2004 – Oct 2010



Thesis proposals (I)
. . . and some internships

Alcune proposte di tesi e/o stage riguardano:

- (Triennali, Posenato) Tesi applicative sulla sperimentazione e personalizzazione del framework Liferay per la realizzazione di applicazioni web
- (Posenato) Tesi di analisi complessità computazionale di problemi di controllabilità nell'ambito di modelli di workflow temporali
- (Posenato) Tesi di studio e implementazione di algoritmi di controllabilità nell'ambito di modelli di workflow temporali
- (Combi) Analisi OLAP e data mining su dati (temporali) biomedici, nei seguenti ambiti
  - immunologia
  - psichiatria
  - farmacologia
  - emodialisi
  - genetica

Thesis proposals (II)
. . . and some internships

- (Combi) Modellazione e gestione di workflow temporali in medicina
  - gestione di linee guida
  - gestione di processi sanitari
- (Combi) Modelli computazionali per process-aware information systems
- (Combi) Petri Nets Simulated Annealing (PNSA): un algoritmo genetico per la correzione di modelli di processo basati su reti di Petri
- (Oliboni) Integrazione di XML e data warehouse
- (Oliboni) XML in bioinformatica
- (Oliboni) Aspetti temporali nella gestione di dati XML
- (Oliboni, Combi) Aspetti temporali nella gestione di dati semistrutturati
- (Oliboni, Combi) XML e data mining

Thesis proposals (III)
. . . and some internships

- (Triennali, Belussi) Strumenti innovativi per la progettazione e lo sviluppo di applicazioni web (Portlet, . . .)
- (Triennali, Belussi) Progettazione concettuale di basi di dati geografiche: il modello GeoUML
  - analisi e confronto di specifiche
  - modello dei dati
  - modello geometrico: analisi e confronto con altri approcci
  - vincoli di integrità spaziale: analisi e confronto con altri approcci
  - analisi degli strumenti per la gestione di uno schema concettuale GeoUML
- (Triennali, Belussi) Analisi di strumenti per la visualizzazione e gestione del dato geografico
- (Triennali, Belussi) Altri argomenti avanzati correlati alle basi di dati

Thesis proposals (IV)
. . . and some internships

- (Belussi) Confronto tra modelli concettuali per basi di dati geografiche
- (Belussi) Tecniche per il confronto di specifiche concettuali di basi di dati geografiche
- (Belussi) Studio e progettazione di strumenti per la visualizzazione di dati geografici guidati dallo schema concettuale
- (Belussi) Studio di tecniche guidate dallo schema concettuale per la migrazione di dati geografici da uno schema fisico all'altro
- (Belussi) Tecniche per la gestione di interrogazioni spaziali approssimate: top-k, skyline basate sulle relazioni di similarità tra relazioni topologiche
- (Belussi, Combi) Supporto al geo-processing distribuito tramite tecnologie di workflow