

## Ruggero Gaetano Pensa

Università di Torino

Dipartimento di Informatica

Corso Svizzera 185

10149 Torino, Italy

Tel. : +39 011 670 6798

E-mail : [pensa@di.unito.it](mailto:pensa@di.unito.it)

Web : <http://www.di.unito.it/~pensa>

## Curriculum Vitæ

---

### Interessi di ricerca

---

- Data Mining e algoritmi per la gestione e l'analisi di dati
- Algoritmi per la privacy dei dati spaziali e spazio-temporali
- Algoritmi per l'analisi di reti sociali
- Bioinformatica e algoritmi per l'analisi di microarray
- Metodi e modelli per la gestione, interrogazione e analisi del dato multimediale

### Posizione attuale

---

Ricercatore presso il Dipartimento di Informatica dell'Università di Torino.

### Esperienze didattiche in ambito accademico

---

- 2011- **Professore aggregato** presso il Dipartimento di Informatica dell'Università degli Studi di Torino.
- 2011-2012 **Tutor** presso il Dipartimento di Informatica dell'Università del Piemonte Orientale nell'ambito del Master Erasmus Mundus DMKM (Data Mining and Knowledge Management).
- 2004-2006 **Esercitatore** presso il Primo Ciclo dell'INSA di Lione (Francia) **Professore a contratto** (annuale) presso il Dipartimento di Informatica dell'Università "Jean Monnet" di Saint-Etienne (Francia)
- 2004-2006 **Esercitatore** presso il Primo Ciclo dell'INSA di Lione (Francia)

### Cursus studiorum

---

- 20/11/2006 **Dottorato in informatica** - Laboratorio di Informatica Grafica e Sistemi di Informazione LIRIS (UMR CNRS 5205, INSA-Lyon, Francia). Equipollenza con il titolo di Dottore di Ricerca riconosciuta dal D.M. del 12 ottobre 2007.
- 18/11/2003 **Laurea** in Ingegneria Informatica, voto : 110/110, Politecnico di Torino.
- 08/2002 - 07/2003 **Periodo di scambio ERASMUS** nel Dipartimento di Informatica dell'INSA-Lyon.
- 1998 **Maturità** scientifica (60/60).

### Conoscenze linguistiche

---

- Francese **Bilingue**. Pratica quotidiana (letto, scritto, parlato) dal 2002.
- Inglese **Buona** (letto, scritto, parlato).
- Tedesco **Nozioni**. Certificato Start in Deutsch I nel 2005.

## Attività di ricerca

---

Le principali attività di ricerca di Ruggero G. Pensa coinvolgono tutti i diversi aspetti del processo KDD (Knowledge Discovery in Databases), dal preprocessing del dato grezzo, al postprocessing dei pattern/modelli, passando per il design e implementazione di nuovi solver e la salvaguardia della privacy del dato.

Durante la tesi di dottorato presso l'INSA di Lione (2003-2006), ha svolto la sua attività nell'ambito dei due progetti europei (IST-FET) cInQ e IQ, interessandosi di algoritmi di data mining per l'analisi dei dati di espressione genica. In particolare, ha sviluppato un metodo per la valutazione delle tecniche di discretizzazione dei microarray [18,37], si è occupato di estrazione di pattern in presenza di dati rumorosi [16,25] e ha sviluppato un metodo per l'interpretazione dei risultati del clustering attraverso l'utilizzo di pattern locali [12,24]. Ha inoltre contribuito allo sviluppo di un'applicazione web per l'analisi dei dati di espressione [10,39,54]. La tesi di dottorato è focalizzata sulla definizione di un framework teorico per la costruzione di bipartizioni con vincoli a partire da collezioni di pattern locali [1,15,22,23,44].

Nell'ambito del Post-Doc all'Università di Saint-Etienne prima (2006-2007), e dell'istituto ISTI del CNR di Pisa poi (2007-2009), Ruggero G. Pensa ha continuato a lavorare in due diversi settori : il clustering con vincoli e la salvaguardia della privacy nell'ambito di dati spazio-temporali. In particolare, ha sviluppato un algoritmo di co-clustering basato su un framework generico per l'ottimizzazione delle divergenze di Bregman in grado di processare diversi tipi di vincoli [9,31,42,43,52]. Nell'ambito del privacy-preserving data mining, come partecipante al progetto europeo GeoPKDD, ha proposto un algoritmo per l'anonimizzazione di dati sequenziali in grado di preservare i risultati del sequential pattern mining e del clustering e alcune statistiche di base [4,34,41]. Durante la permanenza all'ISTI, ha inoltre lavorato a tecniche di data mining per social network [30].

Come assegnista di ricerca all'Università di Torino (Febbraio 2009 - Ottobre 2010), Ruggero G. Pensa ha collaborato allo sviluppo di nuovi algoritmi per l'analisi di dati categorici e digital libraries. In particolare ha proposto un algoritmo di co-clustering gerarchico per collezioni di documenti sia in ambito offline [20,53] che incrementale [6,47]. Ha inoltre collaborato alla definizione di un nuovo approccio di co-clustering per dati rappresentati da features eterogenee [7]. Ruggero G. Pensa ha inoltre partecipato allo sviluppo di un framework per l'apprendimento di distanze tra variabili categoriche basato sulla mutual information [8,21,51]. Durante la permanenza all'Università di Torino, ha inoltre partecipato alla definizione di una metodologia per l'analisi di microarray con tecniche di bi-clustering e co-clustering vincolato, utilizzando come background knowledge l'informazione estratta dalla Gene Ontology [5,13,19,46,50].

Da Novembre 2010 a Ottobre 2011, come assegnista presso l'IRPI del CNR di Torino, ha contribuito allo sviluppo di un'applicazione web per il monitoraggio di fenomeni franosi [29]. Ha inoltre progettato e sviluppato il nuovo sito dell'International Association for Engineering Geology and the Environment [49].

Attualmente lavora alla definizione di modelli per la gestione, interrogazione ed analisi dei dati provenienti da social media, con un'attenzione particolare ai contenuti televisivi [27,28,45]. Lavora inoltre alla definizione di metodi di anomaly/fraud detection per dati categorici e in transazioni bancarie, e a tecniche di clustering per la sommarizzazione per dati geografici.

Ruggero G. Pensa è presente nei comitati di programma di diverse conferenze internazionali (IJCAI, ICDM, SIAM SDM, ECML-PKDD, IDA) e serve abitualmente come reviewer per diverse riviste internazionali di settore (VLDBJ, TKDD, TKDE, KAIS, DAMI, DKE, TDSC, TBME, BMC Bioinformatics). Nel 2010 ha organizzato il Workshop internazionale DyNaK (International Workshop on Dynamic Networks and Knowledge Discovery), tenutosi a Barcellona il 24 settembre del 2010 [3]. È stato guest editor della rivista Intelligent Data Analysis [2] ed è membro del Guest Editorial Board della conferenza ECML-PKDD per i journal "Data Min. Knowl. Disc." e "Machine Learning".

## Incarichi e esperienze di ricerca

---

- 12/2011 Titolare di **Assegno di ricerca** presso il Dipartimento di Informatica dell'Università di Torino sul tema *Caratterizzazione, Specifica, Riconoscimento e Predizione di Eventi nei Social Media*. Responsabile : Prof.ssa **Maria Luisa Sapino**.
- 11/2010 - 10/2011 Titolare di **Assegno di ricerca** presso il Consiglio Nazionale delle Ricerche - Istituto di Ricerca per la Protezione Idrogeologica (IRPI-CNR) sul tema *Data mining ed analisi dati relativi a reti integrate di sensori per il monitoraggio e modellazione di fenomeni di dissesto e delle deformazioni del suolo*. Responsabile : Ing. **Giorgio Lollino**.
- 02/2009 - 10/2010 Titolare di **Assegno di ricerca** presso il Dipartimento di Informatica dell'Università di Torino sul tema *Analisi multidimensionale di repositories documentali e di librerie digitali tramite tecniche di data mining*. Responsabile : Prof.ssa **Rosa Meo**.
- 10/2007 - 01/2009 Titolare di **Assegno di ricerca** presso il Laboratorio KDD dell'ISTI-CNR di Pisa sul tema *Tecniche di Data Mining per l'estrazione di pattern locali e globali da dati con strutture complesse*. Responsabile : Dott.ssa **Fosca Giannotti**.
- 10/2006 - 08/2007 **Professore a contratto** presso l'Università di Saint-Etienne (Francia) e **assistente alla ricerca** presso il Laboratorio Hubert Curien (CNRS UMR5516) di Saint-Etienne (Francia).
- 12/2003 - 11/2006 **Dottorato** in Informatica.  
Tesi discussa il 20 novembre 2006, presso l'Istituto Nazionale di Scienze Applicate, sotto la direzione di **Jean-François Boulicaut** (Professore, INSA Lyon) e di **Céline Robardet** (Assistente, INSA de Lyon) sul tema *A generic framework for constraint-based co-clustering : application to transcriptomic analysis*, e realizzata nel Laboratorio di Informatica Grafica e Sistemi d'Informazione (LIRIS, UMR CNRS 5205).  
**Altri membri della commissione :**  
Presidente : Marc Sebban (Professore, Università di Saint-Etienne)  
Reviewers : Bruno Crémilleux (Professore, Università di Caen)  
Céline Rouveirol (Professore, Università Paris Nord)  
Esaminatore : Luc De Raedt (Professore, Università di Freiburg)
- 03/2003 - 10/2003 **Tesi di Laurea** sotto la direzione di **Jean-François Boulicaut** sul tema *Algoritmi di clustering e di caratterizzazione di clusters : applicazione a dei dati di espressione di geni*, presso il Laboratorio di Informatica Grafica e Sistemi d'Informazione (LIRIS, UMR CNRS 5205).  
Laurea discussa il 18 novembre 2003 al Politecnico di Torino (relatore : **Elena Baralis**).

## Partecipazione attiva a progetti di ricerca

---

- SMAT-F2 2013-2015 Il **progetto regionale SMAT-F2 Sistema di Monitoraggio Avanzato del Territorio** è sovvenzionato dalla Regione Piemonte e dal Fondo Europeo di Sviluppo Regionale (F.E.S.R.). Il consorzio è costituito da un gruppo di lavoro coordinato da Alenia Aermacchi, costituito in Associazione Temporanea di Scopo (ATS) e formato da grandi aziende (Alenia Aermacchi, Selex ES, Altec), Istituti di Ricerca (Politecnico di Torino, Università di Torino, Istituto Superiore Mario Boella, Cnr Irpi) e tredici Piccole e Medie Imprese piemontesi (Blue Engineering, Ceped, CFMAir, Digisky, Divitech, Eicas, Envisens, Eurix, Flyby, Mito, Nimbus, Pros3, Synarea). **Contributo** : Metodi di estrazione di summary dai dati spaziali.

- AML 2012-2013 Il **progetto regionale Early Warning Anti-Money Laundering Tool** è sovvenzionato dalla Regione Piemonte nell'ambito del Polo ICT del programma operativo regionale "Competitività e occupazione" 2007-13. Le unità partecipanti al progetto sono : A&G S.r.l. (coordinatore), Step Ricerche S.r.l., Petra S.r.l., Università degli Studi di Torino.  
**Contributo** : Metodi di anomaly detection per l'individuazione di frodi bancarie riconducibili ad attività di riciclaggio.
- ANONIMO 2009-2010 Il **progetto italiano ANONIMO PRIN 2007** (Metodi informatici e legali per la protezione della privacy e l'anonimato, 2009-2010) è sovvenzionato dal Ministero della Ricerca nell'ambito dei Programmi di ricerca di Rilevante Interesse Nazionale (PRIN 2007). Le unità partecipanti al progetto sono : Università degli Studi di Pisa (coordinatore), Università degli Studi di Bologna, Università degli Studi di Milano, Università degli Studi Insubria Varese-Como, Consiglio Nazionale delle Ricerche. **Contributo** : Metodi di privacy e di knowledge discovery per mobility data analysis.
- GeoPKDD 2005-2008 Il **progetto europeo GeoPKDD IST-FET FP6-014915** (Geographic Privacy-aware Knowledge Discovery and Delivery, 2005-2008) è stato sovvenzionato dal ramo FET del programma IST. Il consorzio è costituito da : ISTI-CNR (coordinatore, Italia), Hasselt University (Belgio), École Polytechnique Fédérale de Lausanne (Svizzera), Fraunhofer Institute for Autonomous Intelligent Systems (Germania), Wageningen University (Olanda), Research Academic Computer Technology Institute (Grecia), Sabanci University (Turchia), WIND Telecomunicazioni SpA (Italia). **Contributo** : Metodi di privacy e di knowledge discovery per mobility data analysis.
- IQ 2005-2008 Il **progetto europeo IQ IST-FET FP6-516169** (Inductive Queries for Mining Patterns and Models, 2005-2008) è stato sovvenzionato dal ramo FET del programma IST. Il consorzio è costituito da : Institute Jozef Stefan (coordinatore, Slovenia), Albert-Ludwigs Universitaet Freiburg (Germania), INSA Lyon (Francia), University of Wales Aberysthwyth (Gran Bretagna), University of Helsinki HIIT (Finlandia), University of Antwerp (Belgio). **Contributo** : impiego di pattern locali per la costruzione/interpretazione di modelli globali e co-clustering con constraints, definizione di un nuovo tipo di pattern locale resistente al rumore.
- Bingo 2004-2007 Il **progetto francese ACI Masse de Données Bingo** (Basi dati Induttive e Genomica, 2004-2006) è stato sovvenzionato da ACI MD 46. Il progetto è costituito da : Università di Caen (coordinatore), Università di Saint-Etienne, Centro di Genetica Molecolare e Cellulare dell'Università Lyon 1, INSA di Lyon. **Contributo** : definizione di una metodologia d'analisi dei dati trascrittomici, impieghi molteplici dei pattern locali.
- clnQ 2001-2004 Il **progetto europeo clnQ IST-2000-26469** (consortium on knowledge discovery by Inductive Queries, 2001-2004) è stato sovvenzionato dal ramo FET del programma IST. Il consorzio era costituito da : Università degli Studi di Torino (Italia), Politecnico di Milano (Italia), Albert-Ludwigs Universitaet Freiburg (Germania), Nokia Research Center à Helsinki (Finlandia), Institute Jozef Stefan (Slovenia), INSA Lyon (coordinatore, Francia). **Contributo** : definizione di un metodo di valutazione della discretizzazione per dati trascrittomici transcriptomiques e di un metodo per la visualizzazione di collezioni di pattern locali.

## Preparazione richieste di finanziamento di progetti di ricerca

---

- 2012 preparazione richiesta di finanziamento come **Responsabile Scientifico dell'Unità di Ricerca** per il bando FIRB 2012 - Futuro in ricerca, per conto dell'Università di Torino. Titolo del progetto : RASHOMON (Recommendation And Search Of Multimedia data exploiting ONtologies). Coordinatore : Ilaria Bartolini (Università di Bologna). Altri partecipanti : Vincenzo Moscato (Università di Napoli).
- 2009 preparazione richiesta di finanziamento come **Responsabile Scientifico dell'Unità di Ricerca** per il bando FIRB 2008 - Futuro in ricerca, per conto dell'Università di Torino. Titolo del progetto : WeCare! : KnoWledge Extrac-tion and privaCy AwaREness in dynamic social networks. Coordinatore : Barbara Carminati (Università dell'Insubria). Altri partecipanti : Salvatore Rinzivillo (ISTI-CNR, Pisa).
- 2007 partecipazione alla preparazione della richiesta di finanziamento per il bando PRIN 2007, per conto dell'ISTI-CNR di Pisa. Titolo del progetto : ANONIMO : metodi informatici e legali per la protezione della privacy e l'anonimato.

## Partecipazione a congressi e convegni in qualità di relatore

---

- 2009 **Conferenza internazionale IDA 2009** Presentazione dell'articolo : *Context-based Distance Learning for Categorical Data Clustering*. Lyon (Francia), 31 agosto - 2 settembre 2009.
- 2008 **Conferenza internazionale ECML PKDD 2008 (workshop LeGo)** Presentazione dell'articolo : *A Constraint-Based Approach for Multispace Clustering*. Antwerp (Belgio), 15 settembre 2008.
- Conferenza internazionale SIAM SDM 2008** Presentazione dell'articolo : *Constrained Co-clustering of Gene Expression Data*. Atlanta, GA (USA), 24 - 26 aprile 2008.
- Congresso nazionale SEBD 2008** Presentazione dell'articolo : *Numerical Data Co-clustering via Sum-Squared Residue Minimization and User-defined Constraint Satisfaction*. Mondello (PA), 22 - 25 giugno 2008.
- Conferenza nazionale (francese) ECG 2008** Presentazione dell'articolo : *Co-classification sous contraintes par la somme des résidus quadratiques*. Sophia Antipolis (Francia), 29 gennaio - 1 febbraio 2008.
- 2006 **Conferenza internazionale ISMIS 2006** Presentazione dell'articolo : *Towards constrained co-clustering in ordered 0/1 data set*. Bari (Italia), 27 - 29 settembre 2006.
- Conferenza nazionale (francese) CAp 2006** Presentazione dell'articolo : *Co-classification sous contraintes*. Trégastel (Francia), 22 - 24 maggio 2006.
- 2005 **Conferenza internazionale AI\*IA 2005** Presentazione dell'articolo : *Towards fault-tolerant formal concept analysis*. Milano (Italia), 21 - 23 settembre 2005.
- Conferenza internazionale IDA 2005** Presentazione dell'articolo : *From local pattern discovery to relevant bi-cluster characterization*. Madrid (Spagna), 8 - 10 settembre 2005.
- 2004 **Conferenza internazionale DS 2004** Presentazione dell'articolo : *A methodology for biologically relevant pattern discovery from gene expression data*. Padova (Italia), 2 - 5 ottobre 2004.
- Conferenza internazionale ACM SIGKDD 2004 (workshop BIODDD)** Presentazione dell'articolo : *Assessment of discretization techniques for relevant pattern discovery from gene expression data*. Seattle, WA (USA), 22 - 25 agosto 2004.

## Organizzazione di conferenze/workshop

---

Workshop **Organizzatore** del workshop internazionale **DyNaK 2010** (International Workshop on Dynamic Networks and Knowledge Discovery) tenutosi a Barcellona (Spagna) il 24 settembre 2010, in concomitanza con la conferenza internazionale **ECML PKDD 2010**.

## Comitati di programma, comitati di organizzazione e referaggi

---

Comitati di programma **Membro del Guest Editorial Board** della conferenza internazionale **ECML PKDD 2014** (25th European Conference on Machine Learning and the 18th European conferences on Principles and practice of Knowledge Discovery in Databases) che si terrà a Nancy (Francia) dal 15 al 19 settembre 2014.

**Membro del comitato di programma** della conferenza internazionale **ACM CIKM 2012** (The 22nd ACM International Conference on Information and Knowledge Management) tenutasi a San Francisco, CA (USA), dal 27 ottobre al 1 novembre 2013.

**Membro del comitato di programma** della conferenza internazionale **IJCAI 2013** (The 23rd International Conference on Artificial Intelligence) tenutasi a Pechino (Cina) dal 3 al 9 agosto 2013.

**Membro del comitato di programma** della conferenza internazionale **IDA 2013** (The 12th International Symposium on Intelligent Data Analysis) tenutasi a Londra (UK) dal 17 al 19 ottobre 2013.

**Membro del comitato di programma** della conferenza internazionale **ECML PKDD 2013** (24th European Conference on Machine Learning and the 17th European conferences on Principles and practice of Knowledge Discovery in Databases) tenutasi a Praga (Repubblica Ceca) dal 23 al 27 settembre 2013.

**Membro del comitato di programma** della conferenza internazionale **ACM CIKM 2012** (The 21st ACM International Conference on Information and Knowledge Management) tenutasi a Maui, Hawaii (USA), dal 29 ottobre al 2 novembre 2012.

**Membro del comitato di programma** della conferenza internazionale **IDA 2012** (The 11th International Symposium on Intelligent Data Analysis) tenutasi a Helsinki (Finlandia) dal 25 al 27 ottobre 2012.

**Membro del comitato di programma** della conferenza internazionale **SIAM SDM 2012** (2012 SIAM International Conference on Data Mining) tenutasi a Anaheim, CA (USA), dal 26 al 28 aprile 2012.

**Membro del comitato di programma** della conferenza internazionale **IEEE ICDM 2011** (11th International Conference on Data Mining) tenutasi a Vancouver (Canada), dal 11 al 14 dicembre 2011.

**Membro del comitato di programma** della conferenza internazionale **ECML/PKDD 2011** (22nd European Conference on Machine Learning and the 15th European conferences on Principles and practice of Knowledge Discovery in Databases) tenutasi ad Atene (Grecia) dal 5 al 9 settembre 2011.

**Membro del comitato di programma** della conferenza internazionale **ECML/PKDD 2010** (21st European Conference on Machine Learning and the 14th European conferences on Principles and practice of Knowledge Discovery in Databases) tenutasi a Barcellona (Spagna) dal 20 al 24 settembre 2010.

**Membro del comitato di programma** della conferenza internazionale **IDA 2010** (The 9th International Symposium on Intelligent Data Analysis) tenutasi a Tucson, Arizona (USA) dal 19 al 21 maggio 2010.

**Membro del comitato di programma** della conferenza internazionale **ECML/PKDD 2009** (20th European Conference on Machine Learning and the 13th European conferences on Principles and practice of Knowledge Discovery in Databases) tenutasi a Bled (Slovenia) dal 7 all'11 settembre 2009.

**Membro del comitato di programma** della conferenza internazionale **IDA 2009** (The 8th International Symposium on Intelligent Data Analysis) tenutasi a Lyon (Francia) dal 31 agosto al 2 settembre 2009.

**Membro del comitato di programma** della conferenza internazionale **IEEE ICDM 2008** (8th International Conference on Data Mining) tenutasi a Pisa dal 15 al 19 dicembre 2008.

**Membro del comitato di programma** della conferenza internazionale **ECML/PKDD 2008** (19th European Conference on Machine Learning and the 12th European conferences on Principles and practice of Knowledge Discovery in Databases) tenutasi ad Anversa (Belgio), 15-19 settembre 2008.

Comitati di organizzazione **Membro del comitato di organizzazione** della conferenza internazionale **IEEE ICDM 2008** (8th International Conference on Data Mining) che tenutasi a Pisa dal 15 al 19 dicembre 2008.

**Membro del comitato di organizzazione** della conferenza internazionale **ECML/PKDD 2004** (15th European Conference on Machine Learning and the 8th European conferences on Principles and practice of Knowledge Discovery in Databases) tenutasi a Pisa, 20-24 settembre 2004.

Attività editoriali **Guest editor** della rivista internazionale ISI **Intelligent Data Analysis** (IOS Press), Special Issue on Dynamic Networks and Knowledge Discovery.

Referaggi Rivista **IEEE Transactions on Multimedia**  
Rivista **IEEE Transactions on Dependable and Secure Computing**  
Rivista **IEEE Transactions on Systems, Man, and Cybernetics**  
Rivista **BMC Bioinformatics**  
Rivista **Springer Data Min. Knowl. Discov. Journal**  
Rivista **Springer Mach. Learn. Journal**  
Rivista **Springer VLDB Journal**  
Rivista **Elsevier Data Knowl. Eng.**  
Rivista **IEEE Trans. Knowl. Data Eng.**  
Rivista **IEEE Trans. Biomed. Eng.**  
Rivista **Springer Knowl. Inf. Syst. Journal**  
Rivista **Elsevier Neurocomputing**

## Seminari e presentazioni

---

- Seminari su invito **Università di Torino** (Dipartimento di Informatica) nel 2008 : *A generic framework for constrained co-clustering : application to gene expression data analysis*. (15 partecipanti)  
**LIG (Laboratory of Informatics of Grenoble)**, Grenoble, France, nel 2007 : *Constrained co-clustering and cluster characterization*. (20 partecipanti)
- Presentazione **Seminario** del gruppo di lavoro **Biologia dei Sistemi e Modellizzazione Cellulare (BSMC)** all'**INSA Lyon** nel 2006 : *Co-clustering e caratterizzazione di clusters*. (20 partecipanti)
- Presentazioni ACI Bingo **INSA di Lyon** nel 2004 : *A methodology for biologically relevant pattern discovery from gene expression data*. (20 partecipanti)  
**INSA di Lyon** nel 2005 : *Post-processing di pattern : visualizzazione di formal concepts*. (20 partecipanti)  
**Università di Caen** nel 2005 : *Caratterizzazione di bipartizioni per mezzo di pattern locali*. (20 partecipanti)
- Presentazione **Black Forest Workshop (Titisee, Germania)** nel 2006 : *Co-clustering for gene expression data analysis : some promising discrete approaches*. (60 partecipanti)

## Premi e riconoscimenti

---

- 11/2011 Vincitore di **Assegno di ricerca** della durata di **24 mesi** co-finanziato dal MIUR, per attività presso il Dipartimento di Informatica dell'Università di Torino.
- 10/2010 Vincitore di **Assegno di ricerca** della durata di **24 mesi** co-finanziato dalla Regione Piemonte, per attività presso l'Istituto di Ricerca per la Protezione Idrogeologica (IRPI-CNR).
- 06/2010 **Idoneità** alla posizione di **Ricercatore CNR III Livello**.
- 01/2009 Vincitore di **Assegno di ricerca** della durata di **24 mesi** co-finanziato dalla Regione Piemonte, per attività presso il Dipartimento di Informatica dell'Università di Torino.
- 09/2007 Vincitore di **Assegno di ricerca** della durata di **11 mesi** per attività presso l'Istituto di Scienze e Tecnologia dell'Informazione (ISTI-CNR).

## Attività Didattica

---

- Legenda **C** : Corso.  
**E** : Esercitazione in aula.  
**L** : Esercitazione di laboratorio.

### Università di Torino come professore aggregato

- 2013-2014 **Informatica Generale** (Corso di Studi in Scienze della Comunicazione, 36h C) Lezioni in aula.
- 2013-2014 **Pattern Recognition and Data Mining** (Master in in ICT e Progettazione avanzata, 4h C) Lezioni in aula.
- 2012-2014 **Laboratorio di Basi di Dati** (Corso di Studi in Informatica, 30h E/L) Esercitazioni di laboratorio e correzione degli esercizi in aula. Realizzazione delle slide di supporto e degli esercizi.
- 2012-2013 **Mining II - Relazioni statistiche** (Master in Analisi dei Dati per la Business Intelligence, 18h C/L) Lezioni in aula ed esercitazioni di laboratorio.
- 2011-2013 **Laboratorio di Architettura degli Elaboratori** (Corso di Studi in Informatica, 30h E/L) Esercitazioni di laboratorio e correzione degli esercizi in aula. Realizzazione di fogli di esercitazione.



#### Università del Piemonte Orientale come tutor

- 2011-2012 **Information Retrieval** (10h L) Tutor locale del corso, nell'ambito del programma del Master Erasmus Mundus DMKM (Data Mining and Knowledge Management). Esercitazioni di laboratorio. Valutazione delle esercitazioni scritte e dei progetti di laboratorio.
- 2011-2012 **Numerical Learning** (10h L) Tutor locale del corso, nell'ambito del programma del Master Erasmus Mundus DMKM (Data Mining and Knowledge Management). Esercitazioni di laboratorio. Valutazione delle esercitazioni scritte e dei progetti di laboratorio.

#### Università di Saint-Etienne come professore a contratto

- 2007 **Tecniche internet di base** (14h C + 20h E + 16h L) Responsabile del corso. Iniziazione alla programmazione web in HTML, JavaScript, PHP e MySQL. Realizzazione di tutte le diapositive e dei supporti cartacei del corso, e di tutti i soggetti delle esercitazioni in aula e in laboratorio. Creazione di un progetto di concezione di un sito web. Preparazione dei temi d'esame. Valutazione dell'esame finale e dei progetti.
- 2007 **Iniziazione agli algoritmi** (48h E + 20h L)  
Programmazione in linguaggio C.
- 2006 **Programmazione in C** (20h C + 20h E + 20h L) Responsabile del corso. Iniziazione agli algoritmi in C. Realizzazione di tutte le slide e le dispense del corso e di tutte le esercitazioni. Preparazione di un progetto di programmazione e realizzazione di tutti i documenti necessari. Preparazione delle tracce d'esame. Correzione e valutazione degli esoneri, dell'esame finale e dei progetti.
- 2006 **Teoria dei grafi** (18h E) Iniziazione alla teoria dei grafi. Esercitazioni incentrate sulla risoluzione di piccole dimostrazioni e sulla modellizzazione di problemi per mezzo di grafi. Correzione e valutazione dell'esame finale

#### INSA de Lyon come esercitatore

- 2005-2006 **Supervisione di un progetto** (28h L), realizzato da gruppi di 4 studenti, incentrato sullo sviluppo di utility di query e di estrazione di conoscenze nei dati di espressione dei geni. La programmazione è realizzata in Java e JDBC + MySQL. Correzione delle relazioni e valutazione dei progetti e della presentazione finale.
- 2004-2006 **Informatica** (40h E + 80h L) Fondamenti di informatica. Esercitazioni pratiche di architettura, word processor e spreadsheet. Esercitazioni in aula e in laboratorio di basi di dati (algebra relazionale, SQL e progetto), e di algoritmi in Pascal. Correzione e valutazione di esoneri ed esami finali.

#### Tesi supervisionate

---

- 2008 **Co-relatore** della tesi di Michele Coscia (**Laurea Specialistica** in Informatica Umanistica, Università di Pisa). Titolo della Tesi : *DiBiMiner : Sistema di estrazione di conoscenza da biblioteche digitali basato su tecniche innovative di Social Network Analysis e Data Mining*. Voto conseguito dal candidato : 110 e Lode.
- 2009 **Supervisore** della tesi di Laura Caponi (**Laurea Specialistica** in Informatica, Università di Pisa). Titolo della Tesi : *Pubblicazione anonima di dati sequenziali*. Voto conseguito dal candidato : 110 e Lode.

## **Pubblicazioni e Prodotti Scientifici**

---

L'informazione sul numero di citazioni è ripresa da Google Scholar. L'impact factor è ripreso dal Journal Citation Reports di ISI Web of Knowledge. Si forniscono solo le informazioni disponibili al 9/2/2014.

H-Index : 12.

### **Tesi di dottorato (1)**

- [1] **R.G. Pensa**. Un cadre générique pour la co-classification sous contraintes : application à l'analyse du transcriptome. Tesi di Dottorato dell'INSA di Lyon (Francia), presentata il 20 novembre 2006.  
**[Citazioni : 3]**

### **Curatele (2)**

- [2] **R.G. Pensa**, F. Cordero, C. Rouveirol. R. Kanawati Eds. *Intelligent Data Analysis - Special Issue on Dynamic Networks and Knowledge Discovery*, Vol. 17(1) 2013. IOS Press.
- [3] **R.G. Pensa**, F. Cordero, C. Rouveirol. R. Kanawati, J.A. Troyano, P. Rosso. Dynamic Networks and Knowledge Discovery. *Proceedings of the 1st Workshop on Dynamic Networks and Knowledge Discovery (DyNaK 2010), Barcelona, Spain, 24 september 2010*. CEUR-WS.org, Vol-655. <http://CEUR-WS.org/Vol-655/>.

### **Riviste internazionali con referee (9)**

- [4] A. Monreale, **R.G. Pensa**, F. Pinelli, D. Pedreschi. Anonymity Preserving Sequential Pattern Mining. *Artificial Intelligence and Law*. Vol. 22(2) 2014. Springer. To appear.
- [5] A. Visconti, F. Cordero, **R.G. Pensa**. Leveraging additional knowledge to support coherent bicluster discovery in gene expression data. *Intelligent Data Analysis*, Vol. 18(5) 2014. IOS Press. To appear.  
**[Impact Factor : 0.472]**
- [6] **R.G. Pensa**, D. Ienco, R. Meo. Hierarchical Co-Clustering : Off-line and Incremental Approaches. *Data Min. Knowl. Discov.* Vol. 28(1) 2014. pp 31-64. Springer  
**[Impact Factor : 2.877, Citazioni : 1]**
- [7] D. Ienco, C. Robardet, **R.G. Pensa**, R. Meo. Parameter-Less Co-Clustering for Star-Structured Heterogeneous. *Data. Data Min. Knowl. Discov.* Vol. 26(2) 2013. pp 217-254. Springer.  
**[Impact Factor : 2.877, Citazioni : 4]**
- [8] D. Ienco, **R.G. Pensa**, R. Meo. From Context to Distance : Learning Dissimilarity for Categorical Data Clustering. *Trans. in Knowledge Discovery in Data*. Vol. 6(1) 2012. pp 1 :1-1 :25. ACM Press.  
**[Impact Factor : 1.676, Citazioni : 7]**
- [9] **R.G. Pensa**, J-F. Boulicaut, F. Cordero, M. Atzori. Co-clustering Numerical Data under User-defined Constraints. *Statistical Analysis and Data Mining*, Vol. 3(1) 2010. pp. 38-55. Wiley-Blackwell.  
**[Citazioni : 12]**
- [10] J. Leyritz, S. Schicklin, S. Blachon, C. Keime, C. Robardet, J-F. Boulicaut, J. Besson, **R.G. Pensa**, O. Gandrillon. SQUAT : a web tool to mine human, murine and avian SAGE data. *BMC Bioinformatics* 2008, 9 :378.  
**[Impact Factor : 3.781, Citazioni : 15]**

- [11] S. Blachon, **R.G. Pensa**, J. Besson, C. Robardet, J-F. Boulicaut and O. Gandrillon. Clustering formal concepts to discover biologically relevant knowledge from gene expression data. *In Silico Biology* 7, 0033 (2007).  
**[Citazioni : 32]**
- [12] **R.G. Pensa**, C. Robardet, J-F. Boulicaut. Supporting co-cluster interpretation in 0/1 data by means of local patterns. *Intelligent Data Analysis journal*, Vol. 10(5) 2006. pp. 457-472. IOS Press.  
**[Impact Factor : 0.446, Citazioni : 9]**

#### Capitoli di libro a diffusione internazionale (6)

- [13] A. Visconti, F. Cordero, D. Ienco, **R.G. Pensa**. *Coclustering under Gene Ontology derived Constraints for Pathway Identification*. Biological Knowledge Discovery Handbook : Preprocessing, Mining and Postprocessing of Biological Data, M. Elloumi and A.Y. Zomaya (Eds), pp. 625-642. John Wiley & Sons, USA. 2013.
- [14] A. Monreale, D. Pedreschi, **R.G. Pensa**. *Anonymity Technologies for Privacy-Preserving Data Publishing and Mining*. Privacy-Aware Knowledge Discovery : Novel Applications and New Techniques, F. Bonchi, and E. Ferrari (Eds), pp. 3-33. Chapman & Hall/CRC Data Mining and Knowledge Discovery Series. December 2, 2010.
- [15] **R.G. Pensa**, C. Robardet, J-F. Boulicaut. *Constraint-driven co-clustering of 0/1 data*. Constrained Clustering : Advances in Algorithms, Theory, and Applications, S. Basu, I. Davidson, and K. Wagstaff (Eds), pp. 123-148. Chapman & Hall/CRC Data Mining and Knowledge Discovery Series. August 18, 2008.  
**[Citazioni : 9]**
- [16] J. Besson, **R.G. Pensa**, C. Robardet, and J-F. Boulicaut. *Constraint-based mining of fault-tolerant patterns from boolean data*. Knowledge Discovery in Inductive Databases (4th International Workshop KDID'05 Revised Selected and Invited Papers). Springer-Verlag LNCS Volume 3933, pp. 55-71, 2006. Springer-Verlag.  
**[Citazioni : 20]**
- [17] **R.G. Pensa**, J. Besson, C. Robardet, J-F. Boulicaut. *Contribution to gene expression data analysis by means of set pattern mining*. Constraint-based mining and Inductive Databases. Springer-Verlag LNCS Volume 3848, pp. 328-347, 2005. Springer-Verlag.
- [18] **R.G. Pensa**, J-F. Boulicaut. *Boolean property encoding for local set pattern discovery : an application to gene expression data analysis*. Local Pattern Detection : International Seminar, Dagstuhl Castle, Germany, April 12-16, 2004, Revised Selected Papers, Springer-Verlag LNAI Volume 3539, pp. 115-134, 2005. Springer-Verlag.  
**[Citazioni : 8]**

#### Conferenze internazionali con atti pubblicati su Springer LNCS/LNAI (8)

- [19] F. Cordero, **R.G. Pensa**, A. Visconti, D. Ienco, M. Botta. Ontology-driven Co-clustering of Gene Expression Data. In *Proceedings of the 11th Conference of the Italian Association for Artificial Intelligence AI\*IA 2009*. December 9-12, 2009, Reggio Emilia, Italy. LNAI 5883, pp 426-435. Springer.  
**[Citazioni : 5]**
- [20] D. Ienco, **R.G. Pensa**, R. Meo. Parameter-free Hierarchical Co-Clustering by n-Ary Splits. In *Proceedings of the 20th European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases ECML PKDD 2009*. September 7-11, 2009, Bled, Slovenia. LNCS 5781, pp 580-595. Springer.  
**[Citazioni : 12]**

- [21] D. Ienco, **R.G. Pensa**, R. Meo. Context-based Distance Learning for Categorical Data Clustering. In *Proceedings of the 8th International Symposium on Intelligent Data Analysis IDA 2009*. August 31 - September 2, 2009, Lyon, France. LNCS 5772, pp 83-94. Springer  
**[Citazioni : 24]**
- [22] **R.G. Pensa**, C. Robardet, J-F. Boulicaut. Towards constrained co-clustering in ordered 0/1 data sets. *Proceedings of the 16th International Symposium on Methodologies for Intelligent Systems ISMIS'06*, Bari, Italy, September 27-29, 2006. Springer-Verlag LNCS 4203, pp. 425-434. Springer-Verlag.  
**[Citazioni : 10]**
- [23] **R.G. Pensa**, C. Robardet, J-F. Boulicaut. A bi-clustering framework for categorical data. *Proceedings of the 9th European conferences on Principles and practice of Knowledge Discovery in Databases PKDD 2005*, Porto, Portugal. Springer-Verlag LNAI Volume 3721, pp. 643-650, 2005. Springer-Verlag.  
**[Citazioni : 30]**
- [24] **R.G. Pensa**, J-F. Boulicaut. From local pattern mining to relevant bi-cluster characterization. *Proceedings of the 6th International Symposium on Intelligent Data Analysis IDA 2005*, Madrid, Spain. Springer-Verlag LNCS Volume 3646, pp. 293-304, 2005. Springer-Verlag.  
**[Citazioni : 10]**
- [25] **R.G. Pensa**, J-F. Boulicaut. Towards fault-tolerant formal concept analysis. *Proceedings of the 9th Congress of the Italian Association for Artificial Intelligence AI\*IA'05*, Milan, Italy. Springer-Verlag LNAI Volume 3673, pp. 212-223, 2005. Springer-Verlag.  
**[Citazioni : 25]**
- [26] **R.G. Pensa**, J. Besson, and J-F. Boulicaut. A methodology for biologically relevant pattern discovery from gene expression data. *Proceedings of the 7th Int. Conf. on Discovery Science DS'04*, Padova, Italy. Springer-Verlag LNAI Volume 3245, pp. 230-241, 2004. Springer-Verlag.  
**[Citazioni : 18]**

#### Conferenze internazionali con atti IEEE/SIAM/ACM/Altro (6)

- [27] A. Antonini, **R.G. Pensa**, M.L. Sapino, C. Schifanella, R. Teraoni Prioletti, L. Vignaroli. Tracking and Analyzing TV Content on the Web through Social and Ontological Knowledge. In *Proceedings of the 11th European Interactive TV Conference EuroITV 2013*. June 24-26, 2013, Como, Italy. pp. 13-22. ACM Press.
- [28] A. Antonini, L. Vignaroli, C. Schifanella, **R.G. Pensa**, M.L. Sapino. MeSoOnTV : A Media and Social-driven Ontology-based TV Knowledge Management System. In *Proceedings of the 24th ACM Conference on Hypertext and Social Media HT'13*. May 1-3, 2013, Paris, France. pp. 208-213. ACM Press.  
**[Citazioni : 2]**
- [29] P. Allasia, **R.G. Pensa**, A. Manconi, M. Baldo, D. Giordan, G. Lollino. GeomonWeb : a Web-Based Monitoring System for Landslide Phenomena. In *Proceedings of the International Conference on Environmental Geosciences and Engineering Survey for Territory Protection and Population Safety EngeoPro-2011*. September 6-8, 2011, Moscow, Russia.
- [30] M. Coscia, F. Giannotti, **R.G. Pensa**. Social Network Analysis as Knowledge Discovery process : a case study on Digital Bibliography. *Proceedings of the 2009 International Conference on Advances in Social Networks Analysis and Mining ASONAM 2009*. July 20-22, 2009, Athens, Greece. pp 279-283, IEEE  
**[Citazioni : 3]**

- [31] **R.G. Pensa**, J-F. Boulicaut. Constrained Co-clustering of Gene Expression Data. *SIAM International Conference on Data Mining SDM 2008*, Atlanta, GA, USA, April 2008.  
[Citazioni : 38]
- [32] I. Slavkov, **R.G. Pensa**, S. Dzeroski. Using bi-sets that characterize bi-partitions as features for classification : an application to microarray data analysis. *Proceedings of the 9th International Multiconference Information Society IS'06*, Ljubljana, Slovenia, October 2006, pages 58–61.

#### Workshop internazionali (6)

- [33] I. Bartolini, V. Moscato, **R.G. Pensa**, A. Penta, A. Picariello, C. Sansone, M.L. Sapino. Recommending Multimedia Objects in Cultural Heritage Applications. In *Proceedings of the 2nd International Workshop on Multimedia for Cultural Heritage MM4CH 2013*. September 9, 2013, Naples, Italy. LNCS 8158, pp 257-267. Springer.  
[Citazioni : 2]
- [34] **R.G. Pensa**, A. Monreale, F. Pinelli, D. Pedreschi. Pattern-Preserving k-Anonymization of Sequences and its Application to Mobility Data Mining. *Proceedings of the International Workshop on Privacy in Location-Based Applications PiLBA'08 in conjunction with ESORICS 2008*. October 9, 2008 - Malaga, Spain. CEUR-WS.org/Vol-397.  
[Citazioni : 30]
- [35] **R.G. Pensa**, M. Nanni. A Constraint-Based Approach for Multispace Clustering. *Proceedings of LeGo-08 Workshop (From Local Patterns to Global Models) co-located with ECML/PKDD-08*, September 2008, Antwerp, Belgium.  
[Citazioni : 3]
- [36] J. Besson, **R.G. Pensa**, C. Robardet, and J-F. Boulicaut. Constraint-based mining of fault-tolerant patterns from boolean data. *Proceedings 4rd Int. Workshop on Knowledge Discovery in Inductive Databases KDID'05 co-located with ECML/PKDD*, pages 13–26, Porto, Portugal, october 2005.  
[Citazioni : 1]
- [37] **R.G. Pensa**, C. Leschi, J. Besson, and J-F. Boulicaut. Assessment of discretization techniques for relevant pattern discovery from gene expression data. *Proceedings 4th ACM Workshop on Data Mining in Bioinformatics BIODDD'04 co-located with SIGKDD'04*, pages 24–30, Seattle - WA, USA, august 2004. ACM.  
[Citazioni : 47]
- [38] C. Robardet, **R.G. Pensa**, J. Besson, and J-F. Boulicaut. Using classification and visualization on pattern databases for gene expression data analysis. *Proceedings of the Int. Workshop on Pattern Representation and Management PaRMA'04 co-located with EDBT 2004*, Heraclion - Crete, Greece, pages 107–118. CEUR Proceedings volume 16, march 2004.  
[Citazioni : 9]

#### Posters selezionati per conferenze in biologia e bioinformatica (2)

- [39] J. Leyritz, C. Keime, S. Blachon, C. Robardet, J-F. Boulicaut, J. Besson, **R.G. Pensa**, and O. Gandrillon. SQUAT : a web tool to browse chicken SAGE data. *International Chick Meeting*, Barcelona, 2007.
- [40] S. Blachon, **R.G. Pensa**, J. Besson, C. Robardet, J-F. Boulicaut. Using formal concepts from biological knowledge discovery from human SAGE data. In *Journées ouvertes Biologie Informatique Mathématiques - JOBIM*, Juillet 2005.

#### Conferenze nazionali (4)

- [41] **R.G. Pensa**, A. Monreale, F. Pinelli, D. Pedreschi. Anonymous Sequences from Trajectory Data. *Proceedings of the 17th Italian Symposium on Advanced Database Systems, SEBD 2009*, Camogli (GE), Italy, June 21-24, 2009.
- [42] **R.G. Pensa**, J-F. Boulicaut. Numerical Data Co-clustering via Sum-Squared Residue Minimization and User-defined Constraint Satisfaction. *Proceedings of the Sixteenth Italian Symposium on Advanced Database Systems, SEBD 2008*, Mondello (PA), Italy, June 22-25, 2008.
- [43] **R.G. Pensa**, J-F. Boulicaut. Co-classification sous contraintes par la somme des résidus quadratiques. *Actes des 8ème Journées Francophones Extraction et Gestion de Connaissances EGC'08*, Sophia-Antipolis, 29 janvier - 1 février 2008, pages 655–666. Cépaduès éditions.  
[Citazioni : 3]
- [44] **R.G. Pensa**, C. Robardet, J-F. Boulicaut. Co-classification sous contraintes. *Actes de la conférence francophone d'apprentissage automatique CAp 2006*, Trégastel, 22-24 mai 2006, pages 155–170, Presses Universitaires de Grenoble.  
[Citazioni : 5]

#### Software, Applicazioni Web, Siti Web (6)

- [45] C. Schifanella, L. Vignaroli, R. Teraoni Prioletti, A. Antonini, **R.G. Pensa**, M.L. Sapino. MeSoOnTV : Tracking and Analyzing TV Content on the Web through Social and Ontological Knowledge. *Released under GPL V3.0 license, 2013*.
- [46] A. Visconti, F. Cordero, **R.G. Pensa**. AID-ISA : Additional Information-Driven Bi-clustering of Gene Expression Data. *Released under GPL V3.0 license, 2013*.
- [47] D. Ienco, **R.G. Pensa**, R. Meo. iHiCC : Incremental Flat and Hierarchical Co-Clustering. *Released under GPL V3.0 license, 2012*.
- [48] D. Ienco, C. Robardet, **R.G. Pensa**, R. Meo. CoStar : Parameter-less Co-Clustering for Star-structured Heterogeneous Data. *Released under GPL V3.0 license, 2012*.
- [49] **R.G. Pensa**. Website of the International Association for Engineering Geology and the Environment (IAEG). *Web Editorial Board : G. Lollino, M. Arattano, R.G. Pensa. Released on September 2011*. Available at <http://www.iaeg.info>.
- [50] A. Visconti, D. Ienco, F. Cordero, **R.G. Pensa**. GOClust : Gene Ontology driven Co-clustering of Gene Expression Data. *Released under GPL V3.0 license, 2011*.
- [51] D. Ienco, **R.G. Pensa**, R. Meo. DILCA : Distance Learning for Categorical Attributes. *Released under GPL V3.0 license, 2011*.
- [52] **R.G. Pensa**, J-F. Boulicaut, F. Cordero, M. Atzori, D. Ienco. CoCoClust : Constrained Co-clustering via Sum-Squared Residue Minimization. *Released under GPL V3.0 license, 2010*.
- [53] D. Ienco, **R.G. Pensa**, R. Meo. HiCC : Hierarchical Co-Clustering. *Released under GPL V3.0 license, 2009*.
- [54] J. Leyritz, S. Schicklin, S. Blachon, C. Keime, C. Robardet, J-F. Boulicaut, J. Besson, **R.G. Pensa**, O. Gandrillon. SQUAT : SAGE Querying and Analysis Tool. *Released in 2008*. Available at <http://bsmc.insa-lyon.fr/squat/>.