

Francesca Bugiotti

Professore associato a CentraleSupélec e membro del team “Large-scale Heterogeneous Data and Knowledge” (LaHDAK) presso il laboratorio di ricerca LISN (Laboratoire Interdisciplinaire des Sciences du Numérique) dell’Università Paris Saclay.

CONTATTI

Attualmente
Enseignant-chercheur
CentraleSupélec - LISN

Sito internet
www.bugiotti.it

CentraleSupélec
3, rue Joliot-Curie
91192, Gif-sur-Yvette
France

LISN
1, rue Raimond Castaing
91190, Gif-sur-Yvette
France

E-mail
francesca.bugiotti@centralesupelec.fr

francesca.bugiotti@lisn.upsaclay.fr

ATTIVITA LAVORATIVA

CentraleSupélec
(*marzo 2015 - ad oggi*)

Enseignant-Chercheur: attività di insegnamento (Big Data, programmazione orientata agli oggetti, architetture software, algoritmi) e di ricerca (Big Data, basi di dati NoSQL ed intelligenza artificiale).

Inria - Institut National de Recherche en Informatique et en Automatique
(*novembre 2013 - febbraio 2015*)

Attività di studio e ricerca su meccanismi di memorizzazione di dati eterogenei sul cloud. Ricerca supportata dall’Inria in connessione con la tematica di ricerca correlata a KIC EIT ICT Labs Europa sulla memorizzazione di dati sul cloud.

Università Roma Tre
(*aprile 2012 - ottobre 2013*)

Attività di studio e ricerca su database non relazionali.

Inria - Institut National de Recherche en Informatique et en Automatique
(*aprile 2011 - luglio 2011*)

Attività di studio e ricerca sul salvataggio e l’indicizzazione di dati RDF e XML sulla piattaforma Amazon Cloud.

Università Roma Tre
(*gennaio 2008 - novembre 2008*)

Attività di studio e ricerca sulla gestione di modelli eterogenei nelle basi di dati.

ISA s.r.l. ¹
(*dicembre 2008 - gennaio 2010*)

Stage di ricerca su analisi di data mining su dati clinici.

Consip ²
(*dicembre 2008 - gennaio 2010*)

Collaborazione con Consip come consulente per la verifica della qualità del nuovo database relazionale sviluppato per la Ragioneria Generale dello Stato Italiano.

ISTRUZIONE

Università Roma Tre
(*novembre 2008 - aprile 2012*)

Dottorato di Ricerca in Ingegneria Informatica

- Titolo della tesi: “A model oriented approach to heterogeneity”.

¹ISA s.r.l. <http://www.isa.it/>, è una società italiana che sviluppa software per piccole-medio imprese. Si occupa prevalentemente di fornire servizi ERP e di business intelligence.

²Consip (Concessionaria Servizi Informativi Pubblici) s.p.a. è una impresa pubblica posseduta dal Ministero dell’economia e della finanza che agisce per conto dello Stato italiano.

- Tutor: prof. Paolo Atzeni³

Università Roma Tre - IBM - Formit ⁴

(*gennaio 2008 - marzo 2009*)

“Master” in IT governance.

Università Roma Tre

(*ottobre 2008*)

Abilitazione all’esercizio della professione di ingegnere.

Università Roma Tre

(*ottobre 2005 - dicembre 2007*)

Laurea magistrale in Ingegneria Informatica.

- Titolo della tesi: “Metodologie e strumenti per problemi di model management”.
- Relatore: prof. Paolo Atzeni
- Votazione: 110/110 e lode

Università Roma Tre (*ottobre 2002 - luglio 2005*)

Laurea in Ingegneria Informatica.

- Titolo della tesi: “Gestione di regole datalog per la traduzione di schemi e dati”.
- Relatore: prof. Paolo Atzeni
- Votazione: 110/110 e lode

AWARDS

Accenture - Università Roma Tre

(*marzo 2009*)

Premio di laurea “Accenture”.

IBM

(*luglio 2007*) Selezionata per l’evento “IBM EMEA Best Student Recognition Event”, Nizza.

COLLABORAZIONI

Transvalor

(*2021 - ad oggi*)

Partecipazione al contratto di ricerca per la prefigurazione di una chaire industriale.

Schlumberger

(*2017 - ad oggi*)

Supervisione di numerosi progetti durante il terzo anno e supervisione di una tesi di dottorato.

Vires - Msc Software

(*2018 - ad oggi*)

Supervisione di un progetto sulla guida automatizzata di autoveicoli.

RESPONSABILITÀ DIDATTICHE

BSc in Artificial Intelligence, Data & Management Sciences

(*2021 - ad oggi*)

Partecipazione all’ideazione del programma accademico del programma. Direttrice del corso di laurea a partire da settembre 2022.

³<http://www.dia.uniroma3.it/~atzeni/>

⁴FORMIT é una fondazione che esegue attività di ricerca, supporto tecnico, valutazione socio-economica per incentivare il processo di migrazione e di integrazione di sistemi software.

CentraleSupélec

Corso	Tipo	Livello	Formazione	Ore	Studenti
BDDM	C/TD/TP	CS (M2)	INI	25	20
Algoritmi per i sistemi distribuiti	C/TP/TD	CS (M2)	INI	24	20
Infrastrutture moderne e cloud	C/TP/TD	CS (M2)	INI	24	40
Cloud Computing	C/TP/TD	CS (M1)	INI	25	100
Big Data	C/TP/TD	CS (M1)	INI		80
Software Engineering	C/TP/TD	CS (M1)	INI		100
Programmazione orientata agli oggetti	TP/TD	CS (M1)	INI		25
Algoritmi	TP/TD	CS (L3)	INI	25	60
Sistemi informativi a programmazione	TP/TD	CS (L3)	INI		35
Models and Systems for Big Data Management	TP/TD	CS (M1)	INI		30
Hardware Architecture	TP/TD	CS (M1)	INI		25
Computer Engineering	TP/TD	CS (M1)	INI		25

EXED/CentraleSupélec

Big Data: algoritmi, piattaforme e metodologie	C/TP/TD	MS	APP	25	30
--	---------	----	-----	----	----

CentraleSupélec/Instytut Villebon Charnak

Analisi dei dati	C/TP/TD	MS	APP	25	15
------------------	---------	----	-----	----	----

Essec/CentraleSupélec

Big Data: algoritmi, piattaforme e metodologie	C/TP/TD	M2	INI	25	120
--	---------	----	-----	----	-----

CentraleSupélec/Erasmus Mundus

Progetto di ricerca Big Data	C/TP/TD	M2	INI	25	20
------------------------------	---------	----	-----	----	----

TU-Berlin

Data Analytics in Energy Sector Applications	C/TP/TD	M2	INI	25	15
--	---------	----	-----	----	----

Advanced Database Design, Data Management & Integration	C/TP/TD	M2	INI	25	15
---	---------	----	-----	----	----

Computer science and programming methods for Energy engineering	C/TP/TD	M1	INI	25	15
---	---------	----	-----	----	----

--	--	--	--	--	--

--	--	--	--	--	--

--	--	--	--	--	--

--	--	--	--	--	--

Basi di Dati II	TP/TD	L3	INI	25	60
-----------------	-------	----	-----	----	----

Programmazione Orientata agli oggetti	TP/TD	M2	INI	25	15
---------------------------------------	-------	----	-----	----	----

IT Governance	C/TP/TD	M2	INI	25	40
---------------	---------	----	-----	----	----

Programmazione Java	TP/TD	L3	INI	25	80
---------------------	-------	----	-----	----	----

Tutorato in Matematica	TP/TD	L2	INI	25	30
------------------------	-------	----	-----	----	----

--	--	--	--	--	--

Big Data	C/TP/TD	M2	INI	25	15
----------	---------	----	-----	----	----

INSEGNAMENTO

Ho iniziato ad insegnare nel 2007 con delle esercitazioni di programmazione Java ed algoritmi ed attualmente il mio volume orario è di 192H/TD.

CentraleSupélec

BDDM (2020-2022) - corso, esercitazioni, valutazione, ricevimento studenti. Il corso si incentra sui sistemi di gestione di dati e sui diversi processi di esecuzione delle query: (i) come i DBMS relazionali e i sistemi NoSQL archiviano i dati (ii) come le query vengono tradotte in operazioni di basso livello che vengono eseguite fisicamente sui dati memorizzato secondo il modello fisico del sistema in analisi (iii) Come scrivere e riscrivere query tenendo conto di questa organizzazione fisica per migliorare le prestazioni.

Algoritmi per i sistemi distribuiti (2020-2022) - corso, esercitazioni, valutazione, ricevimento studenti. Il corso illustra come trattare una grande quantità di dati eterogenei complessi, utilizzando metodi, algoritmi e piattaforme distribuiti. Ogni aspetto teorico è analizzato con riferimento a una tecnologia Big Data standard ed a un framework di riferimento come Docker, Kubernetes e Argo Workflow.

Infrastrutture moderne e cloud (2020-2022) - corso, esercitazioni, valutazione, ricevimento studenti. Questo corso fornisce una conoscenza precisa delle componenti fondamentali utilizzate nella progettazione di infrastrutture e soluzioni informatiche, dei

concetti, metodi e strumenti applicabili alle infrastrutture e ai servizi Cloud, e dei diversi approcci ai database, NoSQL, grafi, e dei loro usi principali

Cloud Computing (2019-2022) - corso, esercitazioni, valutazione, ricevimento studenti. Il corso introduce i concetti base relativi ai Big Data, Hadoop, Spark, Kubernetes e l'analisi dei dati attraverso algoritmi distribuiti (k-means, page rank, etc.).

Programmazione orientata agli oggetti (2015-2022) - esercizi, valutazione. Il corso copre la programmazione Java, la programmazione orientata agli oggetti, introduce UML e la programmazione concorrente.

Algoritmi (2015-2021) - esercitazione, valutazione, ricevimento studenti. Il corso copre gli algoritmi ed i concetti fondamentali della definizione degli algoritmi (indicizzazione, ricerca, ordinamento, ricorsione) ed anche degli algoritmi avanzati (costruzione di diagrammi di Voronoi, analisi dei grafi e colorazione, etc.).

Sistemi informativi e programmazione (2018-2021) - esercitazioni, valutazione. Il corso ha lo scopo di introdurre l'informatica e di insegnare come padroneggiare gli strumenti IT e i sistemi informativi. Il corso introduce anche i concetti di come progettare e scrivere correttamente il codice. Il corso si compone di due parti principali: Sistemi Informativi e Programmazione, la seconda incentrata sul linguaggio Python.

Big Data (2016, 2017, 2018) - corso, esercitazioni, valutazione, ricevimento studenti. Il corso introduce i concetti base relativi ai Big Data, i Databases, Hadoop, Spark e l'analisi dei dati attraverso algoritmi distribuiti (k-means, page rank, etc.).

Models and Systems for Big Data management (2019) - esercitazioni, valutazione. Il corso introduce i concetti base relativi ai DatabasesNoSQL.

Software Engineering (2015) - corso, esercitazioni, valutazione e ricevimento studenti. Il corso comprende la programmazione Java, i concetti della programmazione orientata agli oggetti, UML e la programmazione concorrente.

Hardware Architecture (2016, 2017, 2018) - : esercitazioni. Introduzione alla logica di funzionamento dei computer, all'architettura fisica del processore e ai linguaggi di programmazione di basso livello.

Computer Engineering (2014, 2015, 2016) - esercitazioni. Il corso illustra le regole e i metodi che descrivono la funzionalità, l'organizzazione e l'implementazione dei sistemi informatici.

EXED CentraleSupélec - Mastère Spécialisé en Architecte des Systèmes d'Information

Le mie attività di insegnamento hanno previsto anche interventi in un master specialistico: Master in Information Systems Architect.

Basi di dati relazionali e NoSQL (2019, 2021, 2022) - il corso presenta una introduzione alle basi di dati e confronta l'approccio di gestione ed uso classico dei database relazionali con i modelli dei database NoSQL.

CentraleSupélec - Institut Villebon Charpak

La mia attività didattica comprende anche corsi all'Istituto Villebon Charpak⁵. In questo contesto ho sviluppato e implementando tecniche di insegnamento innovative.

⁵La missione fondamentale dell'istituto Villebon - Georges Charpak é quella di offrire ai giovani diplomati interessati alle scienze ma non a proprio agio con i sistemi di insegnamento tradizionale, l'opportunità di integrare dopo la loro licenza una formazione scientifica di qualità in ingegneria.

Analisi dei dati (2021-2022) - 30H/TD. L'obiettivo di questo corso è rendere gli studenti autonomi nello studio e nell'esplorazione dei dati utilizzando strumenti digitali con un'introduzione all'apprendimento automatico attraverso una serie di esempi.

CentraleSupélec/ESSEC - Master in Data sciences & Business analytics

La mia attività didattica ha previsto anche interventi in un master internazionale in un corso di 120 studenti.

Big Data: algoritmi, piattaforme e metodologie (2020-2022) - 30H équiv. TD, corso, esercitazioni, valutazione, ricevimento studenti. Il corso introduce le principali caratteristiche dei Big Data. La prima parte è incentrata sui database NoSQL e illustra come le loro caratteristiche consentano di archiviare una grande ed eterogenea quantità di dati. La seconda parte analizza i metodi di programmazione di base che consentono agli utenti di gestire grandi quantità di dati. Infine, il corso presenta il paradigma di programmazione map-reduce e mostra come può essere utilizzato un framework di cloud computing (come Hadoop) e nativamente in alcuni datastore NoSQL (come MongoDB).

CentraleSupélec/Erasmus Mundus - Master Big Data Management & Analytics

La mia attività di insegnamento prevede anche la partecipazione ad un master Erasmus Mundus: BDMA (Big Data Management & Analytics).

Progetto di ricerca Big Data (2015-2022) - il corso introduce gli studenti alla ricerca. L'obiettivo principale del corso è lo sviluppo di un progetto di ricerca sull'analisi di dati eterogenei.

TU - Berlin

La mia attività didattica ha previsto anche interventi in ambito internazionale con una collaborazione con TU-Berlino (2018-2021) nell'ambito del master "IT for Energy."

Data Analytics in Energy Sector Applications (2018, 2019, 2020) - corso, esercitazioni, valutazione, ricevimento studenti. Questo corso introduce la progettazione di database NoSQL. In questo corso si presenta anche un'introduzione ai concetti di integrità dei dati e una dimostrazione pratica di come tali concetti vengono gestiti nei database relazionali e NoSQL.

Advanced Database Design, Data Management & Integration (2019) - corso, esercitazioni, valutazione, ricevimento studenti. Questo corso si concentra sulla progettazione delle basi di dati relazionali e mostra come i dati possono essere create ed interrogate utilizzando il linguaggio di query SQL e il DBMS MySQL.

Computer science and programming methods for energy engineering corso presso il campus El-Goona (2019) - corso, esercitazioni, valutazione, ricevimento studenti. Il corso si focalizza sui principi fondamentali dell'informatica, della programmazione orientata agli oggetti e del linguaggio Java.

Università della Tuscia

La mia attività di insegnamento ha previsto degli interventi anche presso l'Università della Toscana nel contesto di internazionalizzazione del loro programma.

Big Data (2018) - corso, esercitazioni, valutazione, ricevimento studenti. Il corso illustra i concetti fondamentali della scienza dei dati: Big Data, partizionamento e distribuzione dei dati, clustering dei dati.

Alternanza scuola-lavoro (13-24 Aprile 2018). Progetto in collaborazione con il Liceo scientifico "Paolo Ruffini" nell'ambito della supervisione di 23 studenti nella definizione di un software per acquisire le conoscenze relative alla programmazione Java ed alla programmazione di un algoritmo NP e all'utilizzo di una piccola base di dati.

Università Roma Tre

La mia attività di insegnamento presso l'Università degli studi Roma Tre (2008-2012).

Basi di dati II (2008, 2009, 2011) - corso per **studenti del secondo anno** con il prof. Paolo Atzeni, esercitazioni, votazione test, orari di ricevimento. Il corso introduce la progettazione di database e i sistemi di gestione di database (DBMS). Ho curato la parte pratica del corso che richiede di eseguire esperimenti in alcuni DBMss (IBM DB2, PostgreSQL, Oracle, SimpleDB) e in alcuni sistemi per di warehousing (Pentaho).

Programmazione orientata agli oggetti (2009, 2011, 2012) - corso per **studenti del terzo anno** con il prof. Luca Cabibbo, esercitazioni, votazione test, ricevimento. Il corso copre l'analisi e la progettazione orientata agli oggetti, introduce alla programmazione agile ed orientata agli oggetti e alla metodologia di sviluppo software iterativo e incrementale.

IT Governance (2012) - corso per **studenti del secondo anno di laurea specialistica**, con il prof. Paolo Atzeni, lezioni, esercitazioni, ricevimento, valutazione. Il corso introduce i principi che regolano il ciclo di vita di un sistema informativo, descrivendone le problematiche metodologiche, le regole economiche e gli aspetti giuridici.

Programmazione Java (2007) - corso per studenti del **terzo anno**, con il prof. Luca Cabibbo, lezioni di laboratorio e votazione della prova d'esame. Il corso introduce le metodologie di programmazione Java.

Tutorato Matematica (2009) - corso per gli studenti del primo anno. Lezioni di matematica con esercizi. Il corso introduce e rivede i concetti principali algebra lineare, geometria, trigonometria ed analisi di funzioni.

Contribuzione al libro, "Basi di dati - Modelli e linguaggi di interrogazione", McGraw Hill 2009, con un contratto con l'editore. Redazione di 100 esercizi con soluzione per i vari capitoli del libro (risorse accessibili on-line all'indirizzo: <http://www.ateneonline.it/atzeni3e/areastudenti.asp>).

COMMISSIONI
TESI
DOTTORATO

20 Ottobre 2020

- Candidato: Amine GHRAB
- Titolo: Graph data warehousing
- Direttore: Dr. Oscar Romero Moral (Universitat Politècnica de Catalunya) Co-direttore: Dr. Esteban Zimanyi (Université libre de Bruxelles)
- Commissione: Prof. Stijn Vansummeren (Université libre de Bruxelles), Dr. Hannes Voigt (Empreses d'Alemania), Prof. Francesca Bugiotti (CentraleSupélec)

RICERCA
SUPERVISIONE

Supervisione di Tesi

Ho supervisionato diversi studenti (laurea triennale e specialistica) per il loro progetto di tesi in ingegneria informatica con il prof. Paolo Atzeni (50% della supervisione), la prof. Tetiana Morozjuk (TU-Berlin - 50 % della supervisione) e la prof. Nacèra Seghouani (50% della supervisione).

PhD tutor

Molood Arman (*marzo 2019 - dicembre 2022*), supervisione con Nacéra Seghouani (Professore presso CentraleSupélec e membro della LISN) e Sylvain Wlodarczyk (Schlumberger). Argomento della tesi: Approcci debolmente non supervisionati per la costruzione di basi di conoscenza da dati geologici e petrofisica fonti di dati eterogenee. Finanziamento: Tesi CIFRE.

COMPETENZE TECNICHE

- *Linguaggi*: Python, Java, C/C++, Prolog, OCaml, Assembly.
- *Basi di Dati*: Basi di dati relazionali, SQL, sistemi NoSQL (MongoDB, Redis, Oracle NoSQL, HBase, DynamoDB).
- *Sistemi operativi*: Linux, Mac Os, Windows.
- *Progettazione del software*: Analisi e progettazione orientata agli oggetti, processo UP.

INTERVENTI SCIENTIFICI

22 aprile 2016 - Seminario - Roma Tre - “Flexible Stores and Data”.

4 dicembre 2017 - Tutorial - TU-Berlin - gruppo di ricerca DIMA / DFKI, Titolo: “Database Design for NoSQL Systems”.

5 dicembre 2017 - Seminario - TU-Berlin - gruppo di ricerca DIMA / DFKI, “Modeling Methodology for a uniform access to NoSQL systems”

9 luglio 2018 - Seminario - TU-Berlin - gruppo di ricerca DIMA / DFKI, “Interpreting Reputation through Frequent Named Entities in Twitter”

4 luglio 2021 - Presentazione - CentraleSupélec, Titolo: “Analisi della ricerca in Intelligenza Artificiale a CentraleSupélec”

12 Dicembre 2022 - Seminario - CEA - In the art - DataIA, Titolo: “Weakly supervised Named Entity Recognition using Deep Neural Networks”

RESPONSABILITÀ SCIENTIFICHE

Corrispondente HUB AI CentraleSupélec - LISN (*2019 - ad oggi*) Faccio parte del Copil dell’HUB AI di CentraleSupélec e sono corrispondente tra l’HUB e il laboratorio di ricerca LISN.

L’AI Hub è stato lanciato nel 2020, con il supporto della direzione generale e della Fondazione CentraleSupélec. Al crocevia di insegnamento, ricerca e innovazione, l’Hub ha lo scopo di diffondere l’IA “ made in CS ” all’interno della scuola ma anche fuori, creando così un ecosistema di studenti, dottorandi, ricercatori, professori e imprese.

Co-responsabile dell’organizzazione di seminari per il team LaHDAK e il Dipartimento di Data Science - (*2020 - ad oggi*)

Attività strategiche e di collaborazione per l’organizzazione di seminari settimanali (team LaHDAK) e mensili (Dipartimento di Data Science).

Membro del collegio di dottorato “Engineering for Energy and Environment” - Università della Toscana - (*2018 - ad oggi*)

Il collegio di dottorato ha come obiettivi principali di programmare il complesso delle attività del dottorato di ricerca e di verificarne l’attuazione.

COMITATI DI PROGRAMMA E REVISIONI

European Conference on Advances in Databases and Information Systems (ADBIS) 2023, co-responsabile dell’organizzazione dei workshops.

Data-driven Smart Cities (DASC) 2023, ICDE Workshops, membro del comitato di programma.

Multi-Armed Bandits for Knowledge Discovery (MAB-KG) 2023, ICDM Workshops, membro del comitato di programma.

Handiversite 2023, membro del comitato di programma.

Gestion de Données - Principes, Technologies et Applications DBA (2022), membro del comitato di programma per le Demo.

Membro del comitato di programma della 26ma International Conference on Scientific and Statistical Database Management, 2014.

Revisore esterno per: International Conference on Extending Database Technology (EDBT) nel 2012, per la Data & Knowledge Engineering (DKE) Journal nel 2013, per Proceedings of Very Large Data Bases nel 2014, and per ACM SIGMOD Conference nel 2014, Journal of Information Systems nel 2022, MENACIS 2022.

PROMOZIONE SCIENTIFICA

Tutorato Solinum (2020/21) - Tutorato per l'analisi dei dati e l'applicazione di algoritmi di IA in collaborazione con la cellula per Imprenditorialità di CentraleSupélec. L'obiettivo di questa collaborazione è stato di fornire le competenze scientifiche per consolidare i progetti nati durante la crisi del COVID. L'attività svolta è consistita inoltre nel miglioramento di Soliguide, uno strumento dell'azienda Solinum, che permette di guidare l'azione sociale grazie all'IA.

“A volte ritornano” (2018-2023) - Serie di di divulgazione scientifica tenuti da ex-studenti e ex-studentesse del Liceo Paolo Ruffini di Viterbo. L'obiettivo dei seminari è quello di rendere la ricerca accessibile a tutti e di tramandare la passione verso la scienza. Per gli studenti del liceo l'iniziativa vuole fornire strumenti utili di fronte alla scelta della propria carriera futura, ad informare sullo stato dell'arte nella ricerca scientifica.

Dalkia: Women's Energy In Transition (2019 - 2020 - 2022 - 2023) - Membro della giuria - *Dalkia challenge* mira a promuovere il posto delle donne nel campo dell'energia, evidenziare curriculum ed iniziative esemplari e incoraggiare le giovani donne ad entrare in queste professioni, partecipando alla promozione della scienza legata all'energia tra gli studenti. Ho contribuito a questo progetto partecipando alla valutazione delle schede dei candidati e alla giuria di valutazione.

Progetto “Summer school for girls” (2021) - L'Università di Paris-Saclay mira a introdurre gli studenti a discipline tradizionalmente trascurate dalle ragazze. Ho partecipato a due settimane di iniziative attraverso un seminario: “Big Data: le incredibili opportunità dell'archiviazione e dell'interpretazione” ed a molteplici gruppi di discussione.

EUROPEAN PROJECT

VRAILEXIA European Erasmus+ Project - (2020-2023)

Il progetto VRAILEXIA è un progetto europeo Erasmus+, riconosciuto e premiato dall'Unesco, che mira a cambiare la percezione della Dislessia ed a sviluppare un modello per superare le principali difficoltà dei dislessici rafforzando la loro motivazione e autostima. Il progetto ha come obiettivo chiave lo sviluppo di una piattaforma digitale per supportare gli studenti dislessici basata sull'IA. Il mio ruolo attiene l'integrazione dei dati che provengono da diversi test, in più lingue, per la valutazione del profilo della dislessia e degli effetti dell'uso della piattaforma sugli aspetti psicologici.

RESEARCH PROJECTS

Le mie attività di ricerca si concentrano principalmente sull'efficienza nell'analisi e nell'integrazione di database eterogenei, che è stato anche il tema di ricerca principale della mia tesi di dottorato. I miei interessi si sono quindi ampliati e sto continuando a studiare: (i) la trasformazione di basi di dati secondo un approccio indipendente dagli schemi e dai modelli di dati, (ii) le tecniche di accesso uniforme ai database NoSQL, (iii) la gestione dei dati efficiente nelle architetture cloud, (iv) l'integrazione efficiente di basi di dati “big data”, e negli ultimi anni (v) l'applicazione di algoritmi di Intelligenza Artificiale su dati eterogenei ed integrati attraverso le tecniche precedenti.

1. LISN

PROCLAIM - (2020-2023)

L'obiettivo di questo progetto è esplorare e definire approcci di estrazione delle informazioni e costruire modelli di apprendimento per ottenere una base di conoscenza da documenti di

perforazione (cuttings), rapporti di analisi dei dati fondamentali e studi geologici, al fine di fornire automaticamente le informazioni a priori necessarie all'interpretazione dei dati di terreno utilizzando la conoscenza pregressa le informazioni di dominio.

B-GRAP - (2018-2020)

La definizione di strategie efficaci per il partizionamento dei grafi è una sfida importante negli ambienti distribuiti poiché un efficace partizionamento dei grafi consente di migliorare considerevolmente le prestazioni dei calcoli di analisi dei dati dei grafi di grandi dimensioni. In questo progetto abbiamo studiato e definito un algoritmo Balanced GRAPh Partitioning (B-GRAP) scalabile per produrre partizioni di grafi bilanciate. B-GRAP si basa sull'approccio Label Propagation (LP) e definisce diverse funzioni-obiettivo per gestire i vincoli di bilanciamento dei vertici o degli archi considerando la direzione degli archi. Gli esperimenti sono stati effettuati su vari grafi e variando il numero di partizioni [IJ7].

SATT DataForYou - (2018-2020)

Partecipazione al progetto SATT DataForYou finalizzato a supportare la creazione della start-up DataForYou, che mira a costruire strumenti a supporto degli enti locali (ad esempio municipi, amministrazioni dipartimentali) in Francia. L'obiettivo del progetto era quello di integrare i dati per l'ottimizzazione dei servizi forniti ai cittadini facendo affidamento su strumenti di analisi comportamentale. In questo progetto, sono stata coinvolta nel coaching di un ingegnere nell'integrazione dei dati.

APIQA - (2016-2017) - [budget : 10KE]

Questo progetto ha avuto come obiettivo la definizione di una metodologia per fornire risposte complete alle query sui dati accessibili tramite Web API. Il progetto si è concentrato inizialmente sui dati del grafico di Twitter. Il progetto ha definito un motore di query che integra le query in tempo reale (o online) sull'API di Twitter con un'origine dati locale (o offline). Grazie a questo progetto è stato possibile creare e mantenere i dati utilizzando un datastore grafico NoSQL. Inoltre, ci siamo concentrati su diversi modi di organizzare i dati sul datastore offline, al fine di migliorare le prestazioni delle query e la completezza dei risultati.

2. Università Roma Tre

SOS - (2011-2012)

Save Our Systems (SOS) è un'interfaccia di programmazione comune [D10] per sistemi NoSQL. Il suo obiettivo è supportare lo sviluppo di applicazioni nascondendo i dettagli specifici dei vari sistemi.

Ho contribuito alla definizione dell'architettura della piattaforma, delle operazioni che espone e le strategie di query che implementa. Sono stata coinvolta nella definizione delle strategie per l'integrazione del datastore NoSQL nel sistema. Ho anche partecipato alla definizione delle tecniche di archiviazione dei dati che vengono utilizzate in ogni datastore per eseguire le operazioni che l'interfaccia espone [IC7,IJ4].

NOAM and ONDM - (2013-2017)

NoSQL Abstract Model (NOAM) è un approccio logico al problema di progettazione del database NoSQL [N15] e mira a sfruttare le caratteristiche comuni di vari sistemi NoSQL. Si basa su un modello di dati intermedio e astratto in cui si definiscono come unità di accesso gli aggregati: unità di distribuzione (per supportare la scalabilità) e coerenza (nella misura necessaria). Alcune rappresentazioni intermedie possono essere implementate nei datastore NoSQL di destinazione, tenendo conto delle loro caratteristiche specifiche e fornendo un supporto efficace per scalabilità, coerenza e prestazioni [M16]. ONDM (Object-NoSQL Datastore Mapper), è il framework [M17] che supporta l'approccio NOAM. Fornisce agli sviluppatori di applicazioni un'interfaccia di programmazione uniforme, nonché la possibilità di mappare i dati delle applicazioni a diverse rappresentazioni di dati e può essere utilizzato, in modo efficace, per eseguire gli esperimenti durante la progettazione di un database NoSQL.

MIDST - (2005-2012)

Model-Independent Schema and Data Translation (MIDST) è una piattaforma per la traduzione di schemi in modo indipendente dal modello e la traduzione dei dati. L'approccio si

basa su un meta-livello che agisce su costrutti di un'ampio numero di modelli con differenti caratteristiche (Relazionale, OR, OO, ER, XML).

Ho contribuito all'estensione di MIDST *supermodel* (un modello generale gestito dalla piattaforma che descrive i vari modelli di dati in termini di un piccolo insieme di *costrutti di base*) e ho anche implementato alcuni componenti software come il motore di traduzione Datalog-SQL fornendo alcune idee sull'evoluzione della piattaforma [N12].

MIDST-RT - (2008-2012)

Model-Independent Schema and Data Translation-RunTime (MIDST-RT) è una piattaforma basata su MIDST ma implementa un approccio di runtime.

Ho contribuito alla definizione, progettazione e implementazione dell'algoritmo MIDST-RT che, dato lo schema del database di origine e il modello di quello di destinazione, genera viste sul sistema operativo che espongono i dati sottostanti secondo lo schema corrispondente nel modello di destinazione. L'approccio implementato genera viste in modo automatico, sulla base delle regole del Datalog per la traduzione dello schema [IJ2,IC6,N14].

MISM - (2009-2012)

Model Independent Schema Management (MISM) è una piattaforma per la gestione dei modelli che offre una serie di operatori per manipolare gli schemi.

Ho progettato e implementato l'algoritmo che fornisce una soluzione al problema ingegneristico di round-trip engineering considerando i principali operatori di gestione dei modelli (merge, diff e modelgen). L'algoritmo agisce secondo un approccio model-independent e model-aware sulla base del supermodello MIDST [IJ3,IN13].

MATRIX and EXL - (2012)

Ho collaborato con Banca d'Italia, supportando l'implementazione di EXLEngine, uno strumento che manipola dati statistici ad alto livello, in termini di entità di modelli statistici come le serie temporali. Abbiamo proposto: (i) un linguaggio EXL, per la specifica dichiarativa di programmi statistici, (ii) un approccio per la traduzione del codice EXL in script eseguibili in vari sistemi target, e (iii) un'implementazione concreta, EXLEngine. L'approccio sfrutta le mappature degli schemi come passaggio di specifica intermedio, al fine di facilitare la traduzione da EXL verso diversi sistemi target [IC5].

GENDATA - (2013)

Nell'ambito dello studio sulla modellazione dei dati ho collaborato al progetto europeo GEN-
DATA <http://gendata.weebly.com/> che mira a definire delle astrazioni, dei modelli e dei protocolli per supportare una rete di dati genomici, rendendoli disponibili per i server nei principali laboratori di biologia. Ho iniziato a collaborare al progetto all'interno dei work packages che coinvolgono Università Roma Tre indagando sul design del modello, il linguaggio di query e la standardizzazione del modello.

3. INRIA

AMADA - (2011)

Durante il periodo di ricerca all'estero del mio dottorato di ricerca ho contribuito al progetto AMADA: una piattaforma [D9,N11] per la memorizzazione di dati Web (Documenti XML e grafici RDF) che si basa sull'infrastruttura cloud di Amazon Web Services (AWS).

Ho fornito una soluzione al problema dell'indicizzazione dei set di dati RDF utilizzando SimpleDB, un archivio chiave-valore fornito da AWS. Ho contribuito alla definizione e allo sviluppo di quattro strategie di indicizzazione [B1,IC8].

ESTOCADA - (2013-2016)

Un nuovo sistema in grado di utilizzare una varietà di DMS praticamente illimitata, garantendo nel contempo la solidità e la completezza del sistema e cercando di ottenere le migliori prestazioni dai vari DMS. Il nostro sistema sfrutta i recenti progressi nell'area della riscrittura delle query, e gestisce i vari modelli descrivendo opportunamente i frammenti archiviati da ciascun DMS. [N15]

Capitoli di libri

- [1] Francesca Bugiotti, Jesús Camacho-Rodríguez, François Goasdoué, Zoi Kaoudi, Ioana Manolescu, and Stamatis Zampetakis. SPARQL query processing in the cloud. In *Linked Data Management.*, pages 165–192. CRC Press, 2014.

Articoli in giornali internazionali

- [IJ2] Paolo Atzeni, Luigi Bellomarini, Francesca Bugiotti, Fabrizio Celli, and Giorgio Gianforme. A runtime approach to model-generic translation of schema and data. *Information Systems*, 37(3):269–287, 2012.
- [IJ3] Paolo Atzeni, Luigi Bellomarini, Francesca Bugiotti, and Giorgio Gianforme. MISM: A Platform for Model-Independent Solutions to Model Management Problems. *Journal of Data Semantics*, 14:133–161, 2009.
- [IJ4] Paolo Atzeni, Luigi Bellomarini, Francesca Bugiotti, and Marco De Leonardis. Executable schema mappings for statistical data processing. *Distributed Parallel Databases*, 36(2):265–300, 2018.
- [IJ5] Paolo Atzeni, Francesca Bugiotti, Luca Cabibbo, and Riccardo Torlone. Data modeling in the nosql world. *Comput. Stand. Interfaces*, 67, 2020.
- [IJ6] Paolo Atzeni, Francesca Bugiotti, and Luca Rossi. Uniform access to nosql systems. *Inf. Syst.*, 43:117–133, 2014.
- [IJ7] Adnan El Moussawi, Nacéra Bennacer Seghouani, and Francesca Bugiotti. BGRAP: balanced graph partitioning algorithm for large graphs. *J. Data Intell.*, 2(2):116–135, 2021.
- [IJ8] Nacéra Bennacer Seghouani, Francesca Bugiotti, Moditha Hewasinghage, Suela Isaj, and Gianluca Quercini. A frequent named entities-based approach for interpreting reputation in twitter. *Data Sci. Eng.*, 3(2):86–100, 2018.

Articoli in conferenze internazionali e workshops

- [IC9] Molood Arman, Sylvain Wlodarczyk, Nacéra Bennacer Seghouani, and Francesca Bugiotti. PROCLAIM: an unsupervised approach to discover domain-specific attribute matchings from heterogeneous sources. In Nicolas Herbaut and Marcello La Rosa, editors, *International Conference on Advanced Information Systems Engineering (CAiSE)*, volume 386, pages 14–28. Springer, 2020.
- [IC10] Paolo Atzeni, Luigi Bellomarini, and Francesca Bugiotti. EXLEngine: executable schema mappings for statistical data processing. In *International Conference on Extending Database Technology (EDBT)*, pages 672–682, 2013.
- [IC11] Paolo Atzeni, Luigi Bellomarini, Francesca Bugiotti, and Giorgio Gianforme. A runtime approach to model-independent schema and data translation. In *International Conference on Extending Database Technology (EDBT)*, pages 275–286, 2009.
- [IC12] Paolo Atzeni, Francesca Bugiotti, and Luca Rossi. Uniform Access to Non-relational Database Systems: The SOS Platform. In *International Conference on Advanced Information Systems Engineering (CAiSE)*, pages 160–174, 2012.
- [IC13] Nacéra Bennacer, Francesca Bugiotti, Jorge Galicia, Mariana Patricio, and Gianluca Quercini. Eliminating incorrect cross-language links in wikipedia. In *Web Information Systems Engineering (WISE)*, pages 109–116, 2017.
- [IC14] Nacéra Bennacer, Francesca Bugiotti, Moditha Hewasinghage, Suela Isaj, and Gianluca Quercini. Interpreting reputation through frequent named entities in twitter. In *Web Information Systems Engineering (WISE)*, pages 49–56, 2017.
- [IC15] Francesca Bugiotti, Damian Bursztyn, Alin Deutsch, Ioana Ileana, and Ioana Manolescu. Invisible glue: Scalable self-tuning multi-stores. In *Conference on Innovative Data Systems Research (CIDR)*, 2015.

- [IC16] Francesca Bugiotti, Luca Cabibbo, Paolo Atzeni, and Riccardo Torlone. Database Design for NoSQL Systems. In *International Conference on Conceptual Modeling (ER)*, pages 1–7, 2014.
- [IC17] Francesca Bugiotti, François Goasdoué, Zoi Kaoudi, and Ioana Manolescu. RDF Data Management in the Amazon Cloud. In *Workshop on Data analytics in the Cloud (DanaC 2012)*, 2012.
- [IC18] Moditha Hewasinghage, Nacéra Bennacer Seghouani, and Francesca Bugiotti. Modeling strategies for storing data in distributed heterogeneous nosql databases. In Juan Trujillo, Karen C. Davis, Xiaoyong Du, Zhanhuai Li, Tok Wang Ling, Guoliang Li, and Mong-Li Lee, editors, *International Conference on Conceptual Modeling (ER)*, volume 11157 of *Lecture Notes in Computer Science*, pages 488–496. Springer, 2018.
- [IC19] Ahmad Khazaie, Nacéra Bennacer Seghouani, and Francesca Bugiotti. Smart crawling: A new approach toward focus crawling from twitter. volume abs/2110.06022, 2021.
- [IC20] Siying Li, José Alves, Francesca Bugiotti, and Frédéric Magoulès. A comparison study of graph neural network and support vector machine. In *Distributed Computing and Applications for Business Engineering and Science (DCABES)*, 2022.
- [IC21] René Gómez Londono, Sylvain Włodarczyk, Molood Arman, Francesca Bugiotti, and Nacéra Bennacer Seghouani. Weakly supervised named entity recognition for carbon storage using deep neural networks. In *International Conference on Discovery Science (DS)*, 2022.
- [IC22] Venkata Shivaditya Meduri, José Alves, Francesca Bugiotti, and Frédéric Magoulès. Point-cloud-based deep learning models for finite element analysis. In *Distributed Computing and Applications for Business Engineering and Science (DCABES)*, 2022.
- [IC23] Venkata Shivaditya Meduri, Francesca Bugiotti, and Frédéric Magoulès. Point-cloud-based deep learning models for finite element analysis. In *Distributed Computing and Applications for Business Engineering and Science (DCABES)*, 2022.
- [IC24] Adnan El Moussawi, Nacéra Bennacer Seghouani, and Francesca Bugiotti. A graph partitioning algorithm for edge or vertex balance. In Sven Hartmann, Josef Küng, Gabriele Kotsis, A Min Tjoa, and Ismail Khalil, editors, *Database and Expert Systems Applications - 31st International Conference, DEXA 2020, Bratislava, Slovakia, September 14-17, 2020, Proceedings, Part I*, volume 12391 of *Lecture Notes in Computer Science*, pages 23–37. Springer, 2020.

Demo

- [D25] Andrés Aranda-Andújar, Francesca Bugiotti, Jesús Camacho-Rodríguez, Dario Colazzo, François Goasdoué, Zoi Kaoudi, and Ioana Manolescu. AMADA: Web Data Repositories in the Amazon Cloud. In *International Conference on Information and Knowledge Management (ACM CIKM)*, 2012.
- [D26] Paolo Atzeni, Francesca Bugiotti, and Luca Rossi. SOS (Save Our Systems): a uniform programming interface. for non-relational systems. In *International Conference on Extending Database Technology (EDBT)*, pages 582–585, 2012.
- [D27] Francesca Bugiotti, Damian Bursztyn, Alin Deutsch, Ioana Manolescu, and Stamatis Zampetakis. Flexible hybrid stores: Constraint-based rewriting to the rescue. In *International Conference on Data Engineering (ICDE)*, pages 1394–1397, 2016.

Articoli e demo in conferenze nazionali

- [N28] Rana B. AL-Otaibi, Francesca Bugiotti, Damian Bursztyn, Alin Deutsch, Ioana Manolescu, and Stamatis Zampetakis. Estocada : Stockage hybride et ré-écriture sous contraintes d'intégrité. In *Journées des Bases de Données Avancées (BDA)*, 2016.
- [N29] Andrés Aranda-Andújar, Francesca Bugiotti, Jesús Camacho-Rodríguez, and Zoi Kaoudi. AMADA: Web Data Repositories in the Amazon Cloud. In *Journées des Bases de Données Avancées (BDA)*, 2012.
- [N30] Paolo Atzeni, Luigi Bellomarini, Francesca Bugiotti, and Giorgio Gianforme. From Schema and Model Translation to a Model Management System. In *British National Conference on Databases (BNCOD)*, pages 227–240, 2008.
- [N31] Paolo Atzeni, Luigi Bellomarini, Francesca Bugiotti, and Giorgio Gianforme. A platform for model-independent solutions to model management problems. In *Italian Symposium on Advanced Database Systems (SEBD)*, pages 310–317, 2008.
- [N32] Paolo Atzeni, Luigi Bellomarini, Francesca Bugiotti, and Giorgio Gianforme. A runtime approach to model-independent schema and data translation. In *Italian Symposium on Advanced Database Systems (SEBD)*, pages 245–252, 2009.
- [N33] Quinio Bernard, Antoine Harfouche, and Francesca Bugiotti. Human-centric ai to mitigate ai biases: The advent of augmented intelligence. In AIM, editor, *Conférence de l'Association Information et Management (AIM)*, 2022.
- [N34] Francesca Bugiotti, Damian Bursztyn, Alin Deutsch, Ioana Ileana, and Ioana Manolescu. Toward Scalable Hybrid Stores. In *Italian Symposium on Advanced Database Systems (SEBD)*, pages 312–319, 2015.
- [N35] Francesca Bugiotti and Luca Cabibbo. A Comparison of Data Models and APIs of NoSQL Datastores. In *Italian Symposium on Advanced Database Systems (SEBD)*, pages 63–74, 2013.
- [N36] Francesca Bugiotti, Luca Cabibbo, Paolo Atzeni, and Riccardo Torlone. How I Learned to Stop Worrying and Love NoSQL Databases. In *Italian Symposium on Advanced Database Systems (SEBD)*, pages 216–223, 2015.
- [N37] Nacéra Bennacer Seghouani, Francesca Bugiotti, Jorge Galicia, Mariana Patricio, and Gianluca Quercini. Élimination des liens inter-langues erronés dans wikipédia. In Mustapha Lebbah, Christine Largeron, and Hanane Azzag, editors, *Conférence francophone sur l'Extraction et la Gestion des Connaissances (EGC)*, volume E-34 of *RNTI*, pages 397–402. Éditions RNTI, 2018.

Articoli sottomessi

- [M38] Molood Arman, Sylvain Wlodarczyk, Nacéra Bennacer Seghouani, and Francesca Bugiotti. Ocrana: Optical character recognition analytics. 2022.
- [M39] René Gómez Londono, Sylvain Wlodarczyk, Molood Arman, Francesca Bugiotti, and Nacéra Bennacer Seghouani. Weakly supervised named entity recognition for carbon storage using deep neural network. 2022.

Autorizzo al trattamento dei dati personali ai sensi del D.Lgs. 196/03.