

Francesca Bugiotti

Enseignant-chercheur à CentraleSupélec et membre de l'équipe "Large-scale Heterogeneous DATA and Knowledge" (LaHDAK) du LISN (Laboratoire Interdisciplinaire des Sciences du Numérique) de l'Université Paris Saclay.

CONTACTS

Poste Actuel
Enseignant-chercheur
CentraleSupélec - LISN

Page personnelle
www.bugiotti.it

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

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

Adresse électronique
francesca.bugiotti@centralesupelec.fr

francesca.bugiotti@lisn.upsaclay.fr

EXPÉRIENCE PROFESSIONNELLE

CentraleSupélec
(*Depuis mars 2015*)

Enseignant-chercheur. Recherche, portant sur l'intelligence artificiel, le Big Data et le bases des données NoSQL.

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

Recherche postdoctorale, supervisée par Ioana Manolescu, sur le stockage efficace de grandes volumes de données hétérogènes dans le cloud.

Università Roma Tre
(*avril 2012 - octobre 2013*)

Recherche postdoctorale, supervisée par Paolo Atzeni, sur la gestion de modèles pour l'intégration de bases de données NOSQL.

Inria - Institut National de Recherche en Informatique et en Automatique
(*avril 2011 - juillet 2011*)

Séjour de recherche invité, supervisé par Ioana Manolescu, sur l'indexation et le stockage de données sémantiques à travers la plateforme Amazon Web Services (AWS).

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

Stage de recherche portant sur les problèmes de gestion de modèles dans les bases de données.

ISA s.r.l. ¹
(*décembre 2008 - janvier 2010*)

Stage de recherche portant sur la fouille de données appliquée aux données cliniques.

Consip ²
(*décembre 2008 - janvier 2010*)

J'ai travaillé pour Consip comme consultant sur la vérification de la qualité de la migration des données pour le nouveaux système d'information du Département Général de l'Etat Italien (*Ragioneria Generale dello Stato Italiano*).

DIPLÔMES

¹ISA s.r.l. <http://www.isa.it/>, est une société Italienne qui conçoit des outils pour les PME, spécialisée dans les services PGI (Progiciel de Gestion Intégré) et l'informatique décisionnelle.

²Consip (Concessionaria Servizi Informativi Pubblici) s.p.a. <http://www.consip.it/on-line/Home.html> est une société anonyme publique qui appartient au Ministère Italien de l'Economie et des Finances.

Università Roma Tre

(*novembre 2008 - avril 2012*)

Doctorat en Informatique - Dipartimento di Informatica ed Automazione

- Thèse : “A model oriented approach to heterogeneity”.
- Directeur de thèse : Pr. Paolo Atzeni³

Università Roma Tre - IBM - Formit ⁴

(*janvier 2008 - mars 2009*)

Master 2 en Gouvernance des technologies de l’information: développement, management et maintenance (*Governo dei sistemi informativi: gestione, sviluppo e monitoraggio*).

Università Roma Tre

(*octobre 2008*)

Diplôme habilitant à l’exercice de la profession d’ingénieur.

Università Roma Tre

(*octobre 2005 - décembre 2007*)

Master 2 en Génie informatique

- Mémoire : “Tools and methodology for model management problems”.
- Directeur : Pr. Paolo Atzeni
- Mention : très bien

Università Roma Tre (*octobre 2002 - juillet 2005*)

Licence en Génie informatique (“Laurea in Ingegneria informatica”).

- Mémoire : “Datalog rules management for data and schema translation”.
- Directeur : Prof. Paolo Atzeni
- Mention : très bien

PRIX

Accenture - Università Roma Tre

(*mars 2009*)

Mémoire récompensé par le prix “Accenture Outstanding Engineering Graduate Award”.

IBM

(*juillet 2007*)

Participation “IBM EMEA Best Student Recognition Event”, Nice.

COLLABORATIONS

Transvalor

(*Depuis 2021*)

Contract de recherche pour la prefiguration d’une chaire industrielle.

Schlumberger

(*Depuis 2017*)

Collaboration dans plusieurs projets de recherche avec la participations d’élèves du 3A.

Co-encadrement d’une thèse de doctorat.

Vires - Msc Software

(*Depuis 2018*)

Co-encadrement de plusieurs projets avec élèves de la 3A.

RESPONSABILITÉS PEDAGOGIQUES

BSc in Artificial Intelligence, Data & Management Sciences

(*Depuis 2021*)

- Directrice académique du programme depuis septembre 2022.
- Responsable scientifique pendant la phase de déploiement et conception programme.

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

⁴FORMIT est une fondation qui s’occupe d’activités de recherche scientifique, support technique, et de soutien de la migration et de l’intégration des systèmes technologiques.

- Participation au processus d'accréditation et à l'entretien face à la commission CTI (Commission des Titres d'Ingénieur) et CEFDG (Commission d'évaluation des formations et diplômes de gestion).

Matière	Nature	Niveau	Formation	Heures	Effectif
CentraleSupélec					
Bases de données Massives	C/TD/TP	CS (M2)	INI	25	20
Algorithmes pour systèmes distribués	C/TP/TD	CS (M2)	INI	24	20
Infrastructures modernes et cloud	C/TP/TD	CS (M2)	INI	24	40
Cloud computing et informatique distribuée	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	15	100
Génie Logiciel	TP/TD	CS (M1)	INI	15	25
Algorithmes et Structures de Données	TP/TD	CS (L3)	INI	25	60
Systèmes d'information et programmation	TP/TD	CS (L3)	INI	15	35
Modèles et systèmes pour la gestion de données Massives	TP/TD	CS (M1)	INI	15	30
Hardware Architecture	TP/TD	CS (M1)	INI	15	25
Systèmes d'information	TP/TD	CS (M1)	INI	15	25
EXED/CentraleSupélec					
Bases de données relationnelles et NoSQL	C/TP/TD	MS	APP	25	30
CentraleSupélec/Institut Villebon Charpak					
Analyse de donnée	C/TP/TD	MS	APP	25	15
Essec/CentraleSupélec					
Big Data Algorithmes, Techniques & Plateformes	C/TP/TD	M2	INI	25	120
CentraleSupélec/Erasmus Mundus					
Projet de recherche 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
Università Roma Tre					
Databases	TP/TD	L3	INI	25	60
Object Oriented Analysis and Design	TP/TD	M2	INI	25	15
IT Governance	C/TP/TD	M2	INI	25	40
Java programming and Algorithms	TP/TD	L3	INI	25	80
Remedial Mathematics	TP/TD	L2	INI	25	30
Università della Tuscia					
Big Data	C/TP/TD	M2	INI	25	15

ENSEIGNEMENT

Mes activités d'enseignement ont commencé en 2007, avec un cours portant sur la programmation Java et Algorithmique, et se sont poursuivies jusqu'à maintenant. A partir de mars 2015 le volume horaire est de 192H équiv. TD. Mes activités sont réparties entre cours magistraux, travaux dirigés et travaux pratiques. Les cours dans lesquels je suis intervenue ont visé un auditoire très diversifié, incluant des élèves de niveau L1, L2, L3, et M2.

CentraleSupélec

Bases de données Massives (2020-2022) - cours magistral, travaux dirigés. L'objectif approfondi les bases de données et les notions associées à l'exécution des opérations : i) comment les bases des données relationnelles et les systèmes noSQL stockent les données physiquement (ii) comment les requêtes sont traduites dans opérations de plus bas niveau qui agissent sur les données stockés physiquement dans le pattern spécifique des bases des données sous analyse (iii) comment écrire et re-écrire les requêtes en prenant in consideration cette organisation physique peut améliorer les performances.

Algorithmes pour systèmes distribués (2020-2022) - cours magistral, travaux dirigés. Le cours comment traiter des données hétérogènes, complexes et massives à l'aide de méthodes, d'algorithmes et de plateformes distribués. Chaque défi d'optimisation et aspect théorique sera analysé en référence à une technologie Big Data standard et dans le cadre de référence tel que Docker, Kubernetes et Argo Workflow.

Infrastructures modernes et cloud (2020-2022) - cours magistral, travaux dirigés. A l'issue de ce cours les étudiants auront une connaissance précise: des équipements essentiels utilisées dans la conception d'infrastructures informatiques des différentes approches relatives aux bases de données, et de leurs principaux usages des concepts, méthodes et outils particulièrement applicables aux infrastructures et services Cloud.

Cloud computing et informatique distribuée (2019-2022) - cours magistral, travaux dirigés. Le cours introduit le concept de Big Data, Hadoop, Spark, Kubernetes et l'analyse des données à travers des algorithmes standard (k-means, page rank, etc.).

Génie Logiciel (2015-2022) - cours magistral, travaux dirigés et pratiques. Fondamentaux de génie logiciel, Programmation Java, Conception d'applications orientées objet, programmation Agile, approche de développement itératif et incrémental.

Algorithmes et Structures de Données (2015-2021) - travaux dirigés et pratiques. Méthodes informatiques de résolution de problèmes d'ingénierie: structures de données usuelles (arbres, graphes, etc.). Techniques de résolution de problèmes (glouton, backtracking, diviser pour régner, approches par flots, heuristiques, etc.) programmation dynamique.

Systèmes d'information et programmation (2018-2021) - travaux dirigés et pratiques. Le cours donne les bases des systèmes d'information qu'ils utiliseront dans leur carrière et à leur permettre de concevoir et d'écrire "proprement" un programme informatique.

Big Data (2016, 2017, 2018) - cours magistral, travaux dirigés, CentraleSupélec. Le cours introduit le concept de Big Data, les bases de données NoSQL, Hadoop, Spark, et l'analyse des données à travers des algorithmes standard (k-means, page rank, etc.).

Modèles et systèmes pour la gestion de données Massives (2019) - travaux dirigés. Le cours introduit le concept de Big Data et des bases de données NoSQL.

Software Engineering (2015) - cours magistraux, exercices, notation des tests, heures de bureau. Le cours couvre la programmation Java, l'analyse et la conception orientées objet, introduit à la programmation UML et concurrent.

Hardware Architecture (2015, 2016, 2017, 2018) - travaux dirigés et pratiques. Introduction, logique et arithmétique, structure d'un ordinateur, traduction des langages de programmation, amélioration des performances.

Systèmes d'information (2014, 2015, 2016) - travaux dirigés et pratiques. Introduction à la distribution des traitements et des données, architecture des systèmes de communication, Bases de données.

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

Mes activités d'enseignement ont aussi prévu des interventions dans un master spécialisé: Mastère Spécialisé en Architecte des Systèmes d'Information.

Bases de données relationnelles et NoSQL (2019, 2021, 2022) - le cours présente les systèmes de gestion de bases de données et compare en détail les bases de données relationnelles classiques et les bases de données NoSQL.

Mes activités d'enseignement ont aussi prévu des interventions dans l'institut Villebon Charpak⁵. Dans ce contexte je suis en train de développer et mettre en place des techniques d'enseignement innovantes.

Analyse de donnée (2021-2022) - 30H équiv. TD. L'objectif de ce cours est de rendre les étudiants autonomes dans l'étude et l'exploration de données à l'aide d'outils numériques avec une ouverture sur le machine learning à partir d'un ensemble d'exemples.

CentraleSupélec/ESSEC - Master in Data sciences & Business analytics

Mes activités d'enseignement ont aussi prévu des interventions dans un master international avec un effectif de 120 élèves.

Big Data : Algorithmes, Techniques & Plateformes (2020-2022) - 30H équiv. TD, cours magistral, travaux dirigés et pratiques. Le cours présente les principales caractéristiques du Big Data. La première partie décrit les systèmes NoSQL montrant comment leurs caractéristiques permettent de stocker une quantité importante et hétérogène de données. La deuxième partie analyse des méthodes de programmation permettant aux utilisateurs de traiter ces grandes quantités de données. Enfin, le cours présente le paradigme de programmation map-reduce et montre comment il peut être utilisé dans des frameworks informatiques basés sur le cloud (comme Hadoop) et nativement dans certains magasins de données NoSQL (comme MongoDB).

CentraleSupélec/Erasmus Mundus - Master Big Data Management & Analytics

Mes activités d'enseignement ont aussi prévu des interventions dans un master Erasmus Mundus: BDMA (Big Data Management & Analytics).

Projet de recherche Big Data : (2015-2022) - le cours introduit à la méthodologie de la recherche. L'objectif principal du cours est dans le développement d'un projet de recherche portant sur l'analyse de données hétérogènes et massives.

TU - Berlin

Mes activités d'enseignement ont aussi prévu des interventions dans le cadre international avec une collaboration avec le TU-Berlin (2018-2021) dans le cadre du master "IT for Energy."

Data Analytics in Energy Sector Applications (2018, 2019, 2020) - cours magistral, travaux dirigés et pratiques. Ce cours détaille les meilleures pratiques en matière de conception de bases de données NoSQL. Cette classe fournit également une introduction aux concepts d'intégrité des données et donne une démonstration pratique de la façon dont ces concepts sont gérés dans les bases de données relationnelles et NoSQL.

Advanced Database Design, Data Management & Integration (2019) - cours magistral, travaux dirigés et pratiques. Ce cours se concentre sur la conception de bases de données relationnelles et montre comment elles peuvent être déployées, remplies et interrogées à l'aide du langage de requête SQL et du système MySQL.

⁵La mission fondamentale de l'institut Villebon - Georges Charpak consiste à offrir à de jeunes bacheliers intéressés par les sciences mais peu à l'aise avec l'enseignement classique, l'occasion d'intégrer après leur licence une formation scientifique de qualité en master ou en école d'ingénieur.

Computer science and programming methods for energy engineering - campus El-Goonah (2019) - cours magistral, travaux dirigés et pratiques. Le cours présente les principes fondamentaux de l'informatique, de la programmation orientée objet et du langage Java.

Università della Tuscia

Mes activités d'enseignement à Università della Tuscia dans un cadre d'internationalisation de leur programme.

Big Data (2018) - cours magistral, travaux dirigés et pratiques. Le cours explique les fondamentaux de la science des données : Big Data, partitionnement et distribution de données, clustering de données.

Alternanza scuola-lavoro (13-24 Avril 2018). Le projet se focalisait sur la supervision de 23 élèves dans le projet de définition d'une application Java dans un lycée avec le développement d'un algorithme NP et l'utilisation d'une base de données.

Università Roma Tre

Mes activités d'enseignement at Università Roma Tre (2008-2012).

Bases de données (2008, 2009, 2011) - 20h équiv. TD, travaux dirigés et pratiques, niveau L2. Fondamentaux des bases de données, SGBDs (IBM DB2, PostgreSQL, Oracle, SimpleDB), entrepôts de données (Pentaho).

Méthode d'analyse et de conception d'applications orientées objet (2009, 2011, 2012) - 30h équiv. TD, travaux dirigés et pratiques, niveau L3. Fondamentaux de génie logiciel, Conception d'applications orientées objet, programmation Agile, approche de développement itératif et incrémental.

Gouvernance des technologies de l'information (2011, 2012) - 15h équiv. TD, travaux dirigés et pratiques, niveau M2. Principes à la base du cycle de vie d'un système informatique.

Programmation Java et Algorithmique (2007) - 30h équiv. TD, travaux dirigés et pratiques, niveau L1. Fondamentaux de la programmation Java.

Fondamentaux de Mathématiques (2009) - heures équiv. TD, cours magistral et travaux pratiques, niveau L1. Algèbre linéaire, géométrie, trigonométrie et analyse de fonctions.

Contribution au livre "Databases - Models and query languages", McGraw Hill 2009. J'ai conçu 100 exercices qui apparaissent à la fin de chaque chapitre <http://www.ateneonline.it/atzeni3e/areastudenti.asp>.

JURIES THESES
DOCTORAT

20 Octobre 2020

- Candidate: Amine GHRAB
- Titre: Graph data warehousing
- Directeur: Dr. Oscar Romero Moral (Universitat Politècnica de Catalunya) Co-directeur: Dr. Esteban Zimanyi (Université libre de Bruxelles)
- Jurie: Prof. Stijn Vansummeren (Université libre de Bruxelles), Dr. Hannes Voigt (Empreses d'Alemanya), Prof. Francesca Bugiotti (CentraleSupélec)

RECHERCHE
ENCADREMENT

Direction des Thèses

J'ai dirigé plusieurs étudiants (Master 2 et Licence) pour leur projet de thèse en génie informatique avec mes responsables de recherche prof. Paolo Atzeni (50% de la supervision),

prof. Tetiana Morozyuk (TU-Berlin - 50 % de la supervision) et prof. Nacéra Seghouani (50% de la supervision). Publications avec Luca Rossi [4, 7, 10], Marco De Leonardis [17], Fabrizio Celli [2], et Moditha Hewasinghage.

Co-Encadrement doctoral

Molood Arman (*mars 2019 - décembre 2022*), direction et co-encadrement avec Nacéra Seghouani (Professeur à CentraleSupélec et membre du LISN) et Sylvain Wlodarczyk (Schlumberger). Sujet de la thèse : Approches faiblement non supervisées pour la construction de bases de connaissances à partir de données géologiques et pétrophysique sources de données hétérogènes. Financement : Thèse CIFRE.

COMPÉTENCES TECHNIQUES

- *Langages*: Italien (maternel), Français, Anglais.
- *Langages de programmation*: Python, Java, C/C++, Prolog, OCaml, Assembly.
- *Base de données*: relationnelles, No-SQL (MongoDB, Redis, Oracle NoSQL, HBase, DynamoDB).
- *Systèmes d'exploitation*: Linux, Mac Os, Windows.

DISSEMINATION SCIENTIFIQUE

- 22 avril 2016** Séminaire - Roma Tre - équipe recherche data, Titre: “Flexible Stores and Data”.
- 4 décembre 2017** Tutoriel invité TU-Berlin - équipe de recherche DIMA/DFKI, Titre: “Database Design for NoSQL Systems”.
- 5 décembre 2017** Séminaire - TU-Berlin - équipe de recherche DIMA / DFKI, Titre: “Modeling Methodology for a uniform access to NoSQL systems”.
- 9 juillet 2018** Séminaire - TU-Berlin - équipe de recherche DIMA / DFKI, Titre: “Interpreting Reputation through Frequent Named Entities in Twitter”.
- 4 juillet 2021** Exposé - CentraleSupélec, Titre: “Weakly supervised Named Entity Recognition using Deep Neural Networks”.
- 12 décembre 2022** Séminaire - CEA - In the art - DataIa, Titre: “La cartographie de la recherche en AI à CentraleSupélec”.

RESPONSABILITES SCIENTIFIQUE

Correspondante HUB AI CentraleSupélec - LISN (*Depuis 2019*) Je fais partie du Copil de l’HUB AI de CentraleSupélec et je suis correspondante entre le HUB et le laboratoire de recherche LISN.

Le Hub IA a été lancé en 2020, avec l’appui de la direction générale et de la Fondation CentraleSupélec. Au carrefour de l’enseignement, de la recherche et l’innovation, le Hub a pour vocation de faire rayonner l’IA “made in CS” à l’externe mais aussi à l’interne, et ainsi créer un écosystème d’élèves-ingénieur, de doctorants et post-doctorants et d’enseignants-chercheurs et industries au travers de partenariats et des actions autour de l’entreprenariat.

Co-responsable de l’organisation des séminaires de l’équipe LaHDAK et du Département Science de Données - (*Depuis 2020*)

Activités stratégiques et de collaboration pour l’organisation hebdomadaire (équipe LaHDAK) et mensuelle (Département Science de Données) des séminaires.

Member of the PhD copil “Engineering for Energy and Environment” - Università della Tuscia - (*Depuis 2018*)

Les principaux objectifs du conseil doctoral sont de planifier l’ensemble des activités stratégiques du doctorat de recherche et de vérifier l’état d’avancement des activités envisagées.

COMITÉS DE PROGRAMME ET REVISEUR

European Conference on Advances in Databases and Information Systems (ADBIS) 2023, workshop-track chair.

Data-driven Smart Cities (DASC) 2023, ICDE Workshops, membre du comité de programme.

Multi-Armed Bandits for Knowledge Discovery (MAB-KG) 2023, ICDM Workshops, membre du comité de programme.

Handiversite 2023, membre du comité de programme.

Gestion de Données - Principes, Technologies et Applications DBA (2022), membre du comité de programme pour les démos BDA.

Membre du comité de programme pour le 26th International Conference on Scientific and Statistical Database Management, 2014.

Reviseur pour le Conference on Extending Database Technology (EDBT) en 2012, le Data & Knowledge Engineering (DKE) Journal en 2013, le Proceedings of Very Large Data Bases en 2014, le ACM SIGMOD Conference in 2014, et le Journal of Information Systems en 2022, MENACIS 2022.

ANIMATION SCIENTIFIQUE

Mentorat Solinum (2020/21) - Mentorat pour l'analyse de données et l'application d'algorithmes AI en collaboration avec la cellule Entrepreneuriat de l'école. L'objectif de cette collaboration était de fournir un environnement d'expertise scientifique permettant de consolider de projets nées pendant la crise COVID. Dans mon action j'ai participé à l'amélioration de Soliguide, util de l'entreprise Solinum, qui permet d'orienter l'action sociale grâce à l'IA.

“A volte ritornano” (2018-2023) - Cycle de séminaires de vulgarisation scientifique organisé par d'anciens élèves du Liceo Paolo Ruffini de Viterbo. L'objectif des séminaires est de rendre la recherche accessible à tous et de transmettre la passion des sciences. Pour les lycéens, l'initiative vise à fournir des outils utiles dans le choix de leur future carrière, et à informer sur les axes actuels de la recherche scientifique.

Dalkia: Women's Energy In Transition (2019 - 2020 - 2022 - 2023) - Membre du jury - *Dalkia challenge* vise à promouvoir la place des femmes dans le domaine de l'énergie, de mettre en lumière des parcours riches et exemplaires et d'inciter les jeunes femmes à rejoindre ces métiers, en participant à sa promotion auprès des élèves. J'ai contribué à ce projet en participant à l'évaluation des dossiers des candidates et au jury.

Projet “Ecole été filles” (2021) - L'Université Paris-Saclay vise à initier à des disciplines traditionnellement délaissées par les filles, dans un vaste bassin de recrutement qui inclut des zones prioritaires. J'ai participé à deux semaines d'initiatives pour des élèves L2 et du collège à travers un séminaire : “Big Data : les incroyables opportunités du rangement et de l'interprétation” et la participation à plusieurs groupes de discussion.

EUROPEAN PROJECT

VRAILEXIA European Erasmus+ Project - (2020-2023)

Le projet VRAILEXIA est un projet Européen Erasmus+, reconnu par Unesco, qui vise à changer la perception et à développer un modèle d'outil pour surmonter les principales difficultés des dyslexiques en renforçant leur motivation et leur estime de soi. Le projet a comme objectif de développer une plateforme numérique de soutien aux étudiants dyslexiques basée sur l'IA. Mon rôle est dans l'intégration de données qui vient de plusieurs tests, en plusieurs langages, pour l'évaluation du profil de la dyslexie et des effets de l'utilisation de la plateforme sur les aspects psychologiques.

RESEARCH PROJECTS

Mes activités de recherche portent principalement sur l'analyse et l'intégration efficace de bases de données hétérogènes, ce qui a aussi été le noyau de ma thèse et poursuivant sur : (i) transformation indépendante des modèles de schémas et de données, (ii) accès uniforme à des entrepôts NoSQL, (iii) gestion de données issues du Web Sémantique dans une architecture cloud, (iv) intégration efficace de “big data” indépendante du modèle, et dans les dernières années (v) sur l'analyse efficace de ces données à travers l'Intelligence Artificielle.

1. LISN

PROCLAIM - (2020-2023)

L'objectif de ce projet c'est d'explorer et de définir des approches d'extraction d'information et de construire des modèles d'apprentissage pour obtenir une base de connaissance à partir de documents issus des boues de forages (cuttings), des rapports de laboratoire d'analyse de données carottés et des études géologiques afin de fournir et ce de manière automatique les informations a priori nécessaires à l'interprétation des diagraphies en utilisant des connaissances et règles métier.

B-GRAP - (2018-2020)

La définition de stratégies efficaces pour le partitionnement de graphes est un défi majeur dans les environnements distribués car un partitionnement efficace permet d'améliorer considérablement les performances des calculs d'analyse de données sur des graphes volumineux. Dans ce projet, nous avons étudié et défini un algorithme de partitionnement de GRAPh (B-GRAP) pour produire des partitions de graphes équilibrés. B-GRAP est basé sur l'approche Label Propagation (LP) et définit différentes fonctions objectives pour traiter les contraintes de distributions des sommets ou des arêtes tout en tenant compte de la direction des arêtes dans les graphes. Les expériences sont réalisées sur différents graphes en faisant varier le nombre de partitions.

SATT DataForYou - (2018-2020)

Participation au projet SATT DataForYou visant à accompagner la création de la start-up DataForYou qui a pour objectif de construire des outils pour soutenir les collectivités territoriales (par exemple, les mairies, les administrations départementales) en France. Le projet avait comme objectif l'intégration des données pour l'optimisation des services rendus au citoyens en s'appuyant sur des outils d'analyse comportementale. Dans ce projet, j'ai été impliquée dans l'encadrement d'un ingénieur sur le lot d'intégration de données.

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

Ce projet avait pour objectif de définir une méthodologie pour retourner des réponses complètes aux requêtes sur les données accessibles via les API Web. Le projet s'est concentré sur les données du graphique Twitter pour le début. Le projet a défini un moteur de requête qui intègre des requêtes en temps réel (ou en ligne) via l'API Twitter avec une source de données locale (ou hors ligne). Il était possible de construire et de maintenir les données à l'aide d'un magasin de données de graphes NoSQL. De plus, nous nous sommes concentrés sur différentes façons d'organiser les données sur le magasin de données hors ligne, afin d'améliorer les performances des requêtes et l'exhaustivité des résultats.

2. Università Roma Tre

SOS - (2011-2012)

Save Our Systems (SOS) est une interface de programmation vers des systèmes NOSQL hétérogènes. SOS vise à cacher les détails des systèmes afin de simplifier le développement des applications.

J'ai contribué à la définition de l'architecture de la plateforme, ses opérations et ses stratégies d'optimisation des requêtes. J'ai participé à la définition de stratégies portant à l'intégration d'entrepôts NOSQL dans le système. J'ai participé aussi à la définition des techniques de mémorisation utilisées pour chaque entrepôt [4, 7, 11].

NOAM and ONDM - (2013-2017)

NoSQL Abstract Model (NOAM) est une approche logique aux problèmes de conception des bases de données NoSQL qui vise à identifier et exploiter les points communs entre les différents systèmes NoSQL. Elle est fondée sur un modèle de données intermédiaire et abstrait, où des objets complexes sont introduits pour supporter le passage à l'échelle et la cohérence. ONDM (Object-NoSQL Datastore Mapper) est un outil qui implémente le modèle NOAM. ONDM fournit aux développeurs d'applications une interface de programmation uniforme et peut être utilisé pour faire des expériences afin de valider l'efficacité d'une représentation de données choisie par rapport à un ensemble de requêtes [9, 18].

MIDST - (2005-2012)

Model-Independent Schema and Data Translation (MIDST) est une plateforme pour la traduction, indépendante du modèle, de schémas et de données entre plusieurs modèles. La traduction est basée sur une approche meta-modèle qui prend en considération plusieurs modèles tels que le modèle relationnel, objet-relationnel, orienté objet, entité-relation ou XML. J'ai contribué à l'extension du *super-modèle* MIDST (un modèle de données abstrait décrivant différents modèles de données en termes d'un petit ensemble de *concepts*) en ajoutant de nouveaux concepts. J'ai aussi réalisé des modules du logiciel MIDST, ce qui a permis une évolution de la plateforme [15].

MIDST-RT - (2008-2012)

Model-Independent Schema and Data Translation-RunTime (MIDST-RT) est une plateforme basée sur MIDST, qui fournit des transformations de données à temps d'exécution. J'ai contribué à la définition, la conception et la réalisation de l'algorithme MIDST-RT qui, à partir du schéma d'une base de données source et du modèle d'une base de données cible, génère des vues qui représentent les données de la source selon le modèle cible [2, 6, 16].

MISM - (2009-2012)

Model Independent Schema Management (MISM) est une plateforme pour le *model management* réalisant un ensemble d'opérateurs pour la transformation de schémas et la manipulation de modèles. La gestion de données à travers des niveaux superposés d'abstraction conduit au problème de *round-trip engineering*⁶. J'ai mis en oeuvre une solution utilisant des opérateurs appropriés (au niveau du schéma et du modèle) qui permettent la propagation correcte des changements de schéma vers les niveaux supérieurs du modèle [3, 17].

MATRIX and EXL - (2012)

J'ai collaboré avec la Banque d'Italie pour la réalisation de EXLEngine, un logiciel pour la manipulation de données statistiques sous la forme de séries temporelles. Nous avons proposé (i) le langage EXL pour la spécification de programmes statistiques, (ii) une approche pour la traduction du code EXL en fichiers exécutables sur plusieurs systèmes, et (iii) une implémentation concrète (la plateforme EXLEngine), qui a été intégrée dans le système informatique de la Banque d'Italie INFOSTAT. L'approche utilise des mappages de schéma comme une étape intermédiaire, afin de faciliter la traduction de EXL vers plusieurs systèmes cibles [5, 19].

GENDATA - (2013)

Le travail portant sur les modèles de données se poursuit dans le contexte du projet GENDATA <http://gendata.weebly.com/>. GENDATA vise à concevoir les modèles et les protocoles pour faciliter le partage et l'analyse de gros volumes de données génomiques qui résident sur des serveurs repartis dans plusieurs laboratoires de biologie dans le monde entier.

3. INRIA

AMADA - (2011)

Mon séjour invité à l'INRIA m'a amené à contribuer au projet AMADA, consistant dans la définition d'une plateforme pour stocker des données issues du Web Sémantique (documents XML et graphes RDF) en utilisant l'infrastructure cloud Amazon Web Services (AWS). J'ai travaillé sur le problème d'indexation d'entrepôts RDF en utilisant SimpleDB, une base de données key-value fournie par AWS. J'ai contribué à la définition et au développement de quatre stratégies d'indexation, qui ont été intégrées et évaluées dans la plateforme AWS [1, 10, 13, 14].

ESTOCADA - (2013-2016)

La production rapide de données digitales est devenue un élément central de la société moderne. Dans ce contexte, Data-Intensive Applications (DIA) ont les caractéristiques suivantes : (i) utilisation de données hétérogènes de grande taille (big data), ayant une structure potentiellement complexe, et (ii) manipulation de données de façon efficace afin

⁶Le *round-trip engineering*, est défini comme suit. Si l'exploitation des données entraîne des changements au niveau d'implémentation inférieur (par ex. le schéma relationnel est altéré), la spécification de haut niveau doit être altérée aussi afin de garder la cohérence entre les deux niveaux.

de garantir des performances optimales. Ma recherche porte sur la création d'entrepôts de données pour une manipulation efficace et automatique de données de grande taille et hétérogènes, afin de supporter au mieux les applications DIA [8, 12].

PUBLICATIONS

Toutes mes publications sont à l'adresse suivant:

<http://www.bugiotti.it/publications.html>

Chapitres de livre

- [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.

Journaux internationaux

- [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.

Conférences internationales

- [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.

Ateliers internationaux

- [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.

Conférences nationales

- [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.

Articles soumis

- [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.