Federico Croce

Curriculum Vitæ ai fini della pubblicazione

May 26, 2023

Documento conforme a quanto prescritto dall'art. 4 del Codice in materia di protezione dei dati personali e dall'art. 26 del D. Lgs. 14 marzo 2013, n. 33.

Part I – General Information

Full Name	Federico Croce

Part II – Education

(IIA) – Academic Achievements

Type	Year	Institution	Notes
Ph.D.	2022	Sapienza University of Rome	Ph.D. in Engineering in Computer Science (Cycle XXXIII). Ph.D. Thesis: "Explaining Datasets in Ontology-based Data Management". Advisor: Prof. Maurizio Lenzerini.
Licensure	2020	Order of Engineers of the Province of Rome	License to practise as an engineer. Section A: Information Technology Engineering.
University graduation	2017	Sapienza University of Rome	Master of Science in Engineering in Computer Science. Final mark: 110/110 cum laude.
Excellence path	2017	Sapienza University of Rome	In-depth educational pathway accessed on merit-based selection for worthy students in the Master of Science in Engineering in Computer Science degree program.
University graduation	2015	Sapienza University of Rome	Bachelor of Science in Engineering in Computer Science and Automation. Final mark: 110/110 cum laude.
Excellence path	2015	Sapienza University of Rome	In-depth educational pathway accessed on merit-based selection for worthy students in the Bachelor of Science in Engineering in Computer Science and Automation degree program.

(IIB) – Partecipation to Training Schools and Workshops

Type	Year	Institution	Notes
Series of Lectures	2020	Goethe University Frankfurt	Series of Lectures regarding the Ethical Implications of AI organized by the Frankfurt Big Data Lab (link). Curatorship: Prof. Roberto V. Zicari.
Workshop	2019	Intel Technology Corporation (link)	Software Development Workshop for Technical Computing and HPC. Speakers: Vojtech Cima, Stephen Blair-Chappell.
Workshop	2019	California Institute of Technology	R Workshop for Caltech Postdocs. The usage of the tools R and RStudio to build reprodoucible experiments (link).
Ph.D. Winter School	2018	West University of Timisoara	4th International Winter School on Big Data (BigDat2018). Chairs: Prof. Bing Liu, Prof. Jeffrey Ullman.
Ph.D. Summer School	2017	University of Genoa	Extending Database Technologies Summer School (EDBT2017). Chairs: Prof. Georg Gottlob, Prof. Barbara Catania, Prof. Giovanna Guerrini, Prof. Leonid Libkin.
Ph.D. Summer School	2017	University of Birbeck	13th Reasoning Web Summer School (RW2017). Chairs: Prof. Giovambattista Ianni, Prof. Domenico Lembo.

Part III – Appointments

(IIIA) – Academic Appointments

Start	End	Institution	Position
2022-04-30	CURRENT	Sapienza University of Rome	**Research Fellow** at the Department of Computer, Control, and Management Engineering. Research project: "Principles for the explanation of open data". Funded by: Hope – PRIN 2017 (link).

Start	End	Institution	Position
2021-04-30	2022-04-29	Sapienza University of Rome	**Research Fellow** at the Department of Translational and Precision Medicine. Research project: "Data engineering services and researching tools and methodologies for data preparation in health care". Funded by: Department of Translational and Precision Medicine, Sapienza University of Rome.
2017-11-01	2022-09-26	Sapienza University of Rome	**Ph.D. Student** at the Department of Computer, Control, and Management Engineering. Research topics: Artificial Intelligence, Knowledge Representation, Semantic Technologies. Funded by: full ministerial scholarship.

(IIIB) – Other Appointments

Start	End	Institution	Position
2019-12-01	2021-11-30	GMATICS S.r.l. (link)	**Head of Artificial Intelligence**. The main activities involved the design, implementation, and management of Artificial Intelligence solutions to remote sensing with satellites and airborne imageries.
2018-10-01	2019-09-30	NASA Jet Propulsion Laboratory (link) and California Institute of Technology (link)	**Visiting Researcher** working on the application of Artificial Intelligence solutions to remote sensing with satellites and airborne imageries. Collaboration in the join ESA/NASA Joint Multi-Mission Algorithm and Analysis Platform (MAAP – link)
2017-01-01	2017-10-30	Sapienza University of Rome	**Library Assistant** at the Department of Computer, Control, and Management Engineering. The main activities involved the gathering, sorting, and management of technical study material and of scientific papers.
2015-09-01	2016-07-31	Nuove Visioni S.r.l. (link)	**Full Stack Developer**. The main activity involved the design, development and management of a Customer Relationship Management application.

Part IV – Teaching Experience

(IVA) – Teaching in Academia

Year	Institution	Course
2022/23	Sapienza University of Rome	**Lecturer** for the Databases (Basi di Dati) course in the Bachelor of Science in Management Engineering. 6 ECTS.
2021/22	Sapienza University of Rome	**Teaching Assistant** for the Databases (Basi di Dati) course in the Bachelor of Science in Engineering in Computer Science. Professorship: Prof. Maurizio Lenzerini.
2021/22	Sapienza University of Rome	**Teaching Assistant** for the Databases (Basi di Dati) course in the Bachelor of Science in Management Engineering. Professorship: Prof.ssa Tiziana Catarci, Prof. Francesco Leotta.
2020/21	Sapienza University of Rome	**Teaching Assistant** for the Databases (Basi di Dati) course in the Bachelor of Science in Management Engineering. Professorship: Prof.ssa Tiziana Catarci, Prof. Francesco Leotta.
2020/21	Sapienza University of Rome	**Teaching Assistant** for the Databases (Basi di Dati) course in the Bachelor of Science in Engineering in Computer Science. Professorship: Prof. Maurizio Lenzerini.
2019/20	Sapienza University of Rome	**Teaching Assistant** for the Databases (Basi di Dati) course in the Bachelor of Science in Management Engineering. Professorship: Prof.ssa Tiziana Catarci, Prof. Francesco Leotta.
2017/18	Sapienza University of Rome	**Teaching Assistant** for the Advanced Computer Science Fundamentals (Fondamenti di informatica II) course in the Bachelor of Science in Engineering in Computer Science. Professorship: Prof. Fabrizio D'Amore, Prof. Giuseppe De Giacomo.
2017/18	Sapienza University of Rome	**Teaching Assistant** for the Basic Computer Science Fundamentals (Fondamenti di informatica I) course in the Bachelor of Science in Telecommunication Engineering, and in the Bachelor of Science of Electronic Engineering. Professorship: Prof. Silvio Salza.

(IVB) – Lectures and Specialized Seminars

Year	Institution	Course
2022	Sapienza University of Rome	**Speaker** for a seminar titled "Data preparation e data analytics: the case of diabetes treatment.", held in collaboration with the Sapienza Information-Based Technology InnovaTion Center for Health (STITCH).
2021	Sapienza University of Rome	**Speaker** for a seminar titled "Use of big data: what potential? Which methodology? The diabetic patients data model as an example", held as part of the Doctoral Program in Data Science.

(IVC) – Teaching in Industry

Year	Institution	Course
2021	Teleconsys S.p.A. (link)	**Lecturer** for a course on NoSQL Databases and Microservices Architectures. 10 hours.

Part V – Society memberships, Awards and Honors

Year	Title
2020	Member of the executive committee for the Italian Society for the Ethics of Artificial Intelligence (SIpEIA – link).
2017	Award for outstanding graduate at Sapienza University of Rome. Selected among the top 400 (∼2%) of all graduates of the university.
2015	Honor scholarship titled "alla memoria dei Coniugi Ernesto e Iole DE MAGGI" awarded by the Sapienza Foundation for the Promotion of Study and Research (link).

Part VI – Funding Information

Year	Role	Program	Grant value
2022	Principal Investigator	European Social Fund Program Plus (FSE+) 2021 – 2027. "Contributi premiali per i ricercatori e assegnisti di ricerca per rafforzarne la condizione professionale e potenziare il sistema della ricerca del Lazio" (link).	e 2000

Part VII – Research Activities

Federico Croce is affiliated to the Data and Service Integration Lab (DASILab) at the Department of Computer, Control, and Management Engineering of Sapienza University of Rome. He is a member of both the Data Management and Service-Oriented Computing, and the Artificial Intelligence and Knowledge Representation research groups. His main

research interest is Artificial Intelligence (AI), particularly the field of Knowledge Representation.

His research has been primarily focused on the Ontology-based Data Management (OBDM) paradigm, i.e. an approach to interact with data through a conceptual representation of a domain of interest called an ontology. The OBDM paradigm offers methods and tools to achieve the integration, preparation and governance of data stored in an information system in order to both form a single point of conceptual access to information assets, and to realize at the conceptual level all the data governance services of a complex system. According to the OBDM paradigm, an information system is based on an architecture that includes the following layers: (1) the data sources, characterized by metadata, schemas and the actual data; (2) the conceptual representation of the domain of interest, i.e. the ontology, formulated through axioms in logic; (3) the correspondences (mappings) between the data sources and the concepts and relations of the ontology, also expressed through assertions in logic; (4) the various services that are exposed to the users of the information system such as query answering, data quality checking, data provenance, etc. Note that information systems based on the OBDM paradigm provide to their users not a merely data structure accommodating the data at the sources, but a semantically rich description of the relevant concepts in the domain of interest, as well as the relationships between such concepts. In what follows, we will refer to an information system organized accordin to the OBDM paradigm, or simply an OBDM system, as a tuple $\Sigma = \langle J, D \rangle$ where D represents the acutal data residing at the data source layer, and $J = \langle \mathrm{O}, \mathrm{S}, \mathrm{M} \rangle$ represents a so-called OBDM specification, which is composed by the ontology O, the schema of the data sources S, and the mapping M between the two.

More Recently, Federico Croce has also investigated the problems related to the effective adoption of Artificial Intelligence solutions to the healthcare domain, especially the ones using Machine Learning (ML). This problem is becoming increasingly important because, on the one hand there are promising results that focus on automating tasks and analyzing big patient data sets to deliver better healthcare faster. On the other hand, there are many problems related to the typical behaviour of higher accuracy models based on ML techniques, i.e. their limited possibility of explaining their findings, as well as the complexity and huge overhead needed in preparing the datasets to be suitable for the application of ML techniques.

The following table summarizes the main areas of research Federico Croce has been focused on.

Keywords	Brief Description
Query Explaination	It is problem of explaining why a specific tuple is the answer of a query with respect to an OBDM system. The results are based on the fundamental assumption that explanations in this context are strictly related to deductions. Therefore, one can rephrase the task described as the problem of explaining the deductive process that from the query and the ontology of the OBDM system, led to the specific input tuple. In the case of plain relational databases, the explanation of an answer to a given query can be expressed as an homomorphism, i.e., as the assignment of the query variables to the values in the database that makes the query true. Conversely, in OBDM, the role of the ontology must be taken into account in order to provide meaningful explanations. For instance, if b is a student , and the ontology sanctions that every student is a person, then explaining the answer "b is a person" involves exhibiting both the fact that b is a student, and the ontology axiom Student is-a Person. This problem also deals with providing explanations to tuples being certain answers to the negation of queries. Intuitively, this can be seen as the task of explaining why a given tuple can never be in the certain answers of the corresponding positive query. For example, if b is a student, and the ontology sanctions that students and professors are disjoint sets, then explaining the answer "b cannot be a professor" can be done by exhibiting both the fact that b is a student, and the ontology axiom Student is-not-a Professor.
Query Characterization and Query Separation	This problem deals with building reverse-engineered queries from datasets in Ontology-based Data Management (OBDM). The most natural scenario is the case in which an OBDM system and a set of tuples are given as inputs, and one is interested in finding a query over the ontology so that the answers of the query with respect to the OBDM system are exactly the tuples in the input dataset. This problem is referred to as the Query Characterization problem, as the found reverse-engineered query characterizes the input dataset by using the ontology. A more general scenario is the case in which one is given an OBDM system, and two different sets of tuples. The two datasets represent positive and negative examples, and one is interested in searching for a query over the ontology of the OBDM system, so that the answers of the query, include all the tuples in the set of positive examples, and none of the tuples in the set of negative examples. This problem is referred to as the Query Separation problem, as the found reverse-engineered query separates the two sets of positive and negative examples according to the ontology. As one might expect, it is often the case that there exist no query with the aforementioned properties. To deal with these cases, for both problems, one is interested in finding approximated results in terms of precision and recall.

Keywords	Brief Description
Query Expressivity	This problem deals with changing the expressivity of the class of queries that a user can make over an OBDM system, and in studying the related changes (if any) to the computational complexity of the reasoning tasks offered by the system. In particular, Federico Croce studied the problem of adding forms of inequalities to unions of conjunctive queries, i.e. the most popular class of queries both for traditional databases and for OBDM systems.
Data Preparation in Healthcare	In the context of healthcare, an AI solution is generally developed for a specific data analytics task, based on a specific dataset that is relevant for the analysis at hand. In this task-oriented scenario, data preparation is often carried out through ad-hoc methods, with little attention to reusability and generalizability. Conversely, this problem is framed in a context-oriented scenario, where a set of clinical data sources that are relevant for a specific context (e.g., a particular disease) is available and can be used for a variety of data analytics tasks, often carried out by different research groups. Federico Croce worked on a method, centred around the Ontology-based Data Management paradigm, for the two main steps of data preparation, namely data modeling and data cleaning, in such a context-oriented scenario. Furthermore, he studied the application of this method in a project dealing with data regarding the treatment of diabetes and its complications.

Part VIII – Summary of Scientific Achievements

Product type	Number	Database	Start	End
Papers [international]	7	Google Scholar	2018	2023
Papers [national]	2	Google Scholar	2018	2023
Papers [international]	4	Scopus	2018	2023

Metric	Google Scholar	Scopus
Total Citations	35	8
Average Citations per Product	3.8	2
Hirsch (H) index	4	2
Normalized H index [1]	0.8	0.4

Part IX – Publications

(IXA) – All publications

Marianna Maranghi, Aris Anagnostopoulos, Irene Cannistraci, Ioannis Chatzigiannakis, Federico Croce, Giulia Di Teodoro, Michele Gentili, Giorgio Grani, Maurizio Lenzerini, Stefano Leonardi, Andrea Mastropietro, Laura Palagi, Massimiliano Pappa,

[1] H index divided by the academic seniority (2018-2023)

Riccardo Rosati, Riccardo Valentini, Paola Velardi: **AI-based Data Preparation and Data Analytics in Healthcare: The Case of Diabetes.** In *Computing Research Repository (CoRR)*, arXiv abs/2206.06182, 2022. (pdf – Open Access)

Gianluca Cima, Federico Croce, Maurizio Lenzerini: **Query Definability and Its Approximations in Ontology-based Data Management.** In *Proceedings of the 30th ACM International Conference on Information and Knowledge Management (CIKM)*, pp. 271-280, 2021. (pdf)

Federico Croce, Maurizio Lenzerini: **Explaining Classifiers in Ontology-Based Data Access.** *The 17th International Conference on Principles of Knowledge Representation and Reasoning (XLoKR2020).* Held in conjunction with the 17th International Conference on Principles of Knowledge Representation and Reasoning, 2020. (pdf – Open Access)

Federico Croce, Gianluca Cima, Maurizio Lenzerini, Tiziana Catarci: **Ontology-based explanation of classifiers.** In *Proceedings of the 2nd International Workshop on Processing Information Ethically (PiE).* Volume 2578 of *CEUR Workshop Proceedings*, 2020. (pdf – Open Access)

Gianluca Cima, Federico Croce, Maurizio Lenzerini, Antonella Poggi, Elian Toccacieli: **On Queries with Inequalities in *DL-Lite*$_R^{\neq}$.** In *Proceedings of the 32nd International Workshop on Description Logics (DL)*. Volume 2373 of *CEUR Workshop Proceedings*, 2019. (pdf – Open Access)

Federico Croce, Maurizio Lenzerini: **A Framework for Explaining Query Answers in DL-Lite.** In *Proceedings of the 21st International Conference on Knowledge Engineering and Knowledge Management (EKAW)*, Volume 11313 of *Springer Lecture Notes in Computer Science*, pp. 83-97, 2018. (pdf)

Raffaella Maria Aracri, Adele Maria Bianco, Roberta Radini, Monica Scannapieco, Laura Tosco, Federico Croce, Domenico Fabio Savo, Maurizio Lenzerini: **On the Experimental Usage of Ontology-based Data Management for the Italian Integrated System of Statistical Registers: Quality Issues.** *The 9th European Conference on Quality in Official Statistics (Q2018)*, 2018. (pdf – Open Access)

(IXB) – Selected publications

Gianluca Cima, Federico Croce, Maurizio Lenzerini: **Query Definability and Its Approximations in Ontology-based Data Management.** In *Proceedings of the 30th ACM International Conference on Information and Knowledge Management (CIKM)*, pp. 271-280, 2021. (pdf). Citations: 2.

Federico Croce, Gianluca Cima, Maurizio Lenzerini, Tiziana Catarci: **Ontology-based explanation of classifiers.** In *Proceedings of the 2nd International Workshop on Processing Information Ethically (PiE).* Volume 2578 of *CEUR Workshop Proceedings*, 2020. (pdf – Open Access). Citations: 2.

Gianluca Cima, Federico Croce, Maurizio Lenzerini, Antonella Poggi, Elian Toccacieli: **On Queries with Inequalities in *DL-Lite*$_R^{\neq}$.** In *Proceedings of the 32nd International Workshop on Description Logics (DL)*. Volume 2373 of *CEUR Workshop Proceedings*, 2019. (pdf – Open Access). Citations: 1.

Federico Croce, Maurizio Lenzerini: **A Framework for Explaining Query Answers in DL-Lite.** In *Proceedings of the 21st International Conference on Knowledge Engineering and Knowledge Management (EKAW)*, Volume 11313 of *Springer Lecture Notes in Computer Science*, pp. 83-97, 2018. (pdf). Citations: 3.

Part X – Paper Presentation

Year	Event	Paper
2022	The 2nd CINI National Conference on Artificial Intelligence (Ital-IA).	AI-based Data Preparation and Data Analytics in Healthcare: The Case of Diabetes.
2021	The 30th ACM International Conference on Information and Knowledge Management (CIKM).	Query Definability and Its Approximations in Ontology-based Data Management.
2020	The 17th International Conference on Principles of Knowledge Representation and Reasoning (XLoKR2020).	Explaining Classifiers in Ontology-Based Data Access.
2020	The 2nd International Workshop on Processing Information Ethically (PiE).	Ontology-based explanation of classifiers.
2018	The 9th European Conference on Quality in Official Statistics (Q2018).	On the Experimental Usage of Ontology-based Data Management for the Italian Integrated System of Statistical Registers: Quality Issues.

Part XI – Academic Community Service

REVIEWER:

- Journal of Data and Information Quality (JDIQ): 2020, 2021, 2022, 2023.
- Journal Springer Nature for Computer Science: 2021, 2022, 2023.
- Journal of Multidisciplinary Digital Publishing Institute (MDPI): 2023.
- International Conference on Principles of Knowledge Representation and Reasoning (KR): 2020.

SUB-REVIEWER:

- International Joint Conference on Artificial Intelligence (IJCAI): 2022.

Firma: .....