Michele De Bonis

[redacted]

WORK EXPERIENCE

01/01/2023 – CURRENT Pisa, Italy
RESEARCH FELLOW ISTITUTO DI SCIENZA E TECNOLOGIE DELL'INFORMAZIONE "A. FAEDO" (ISTI - CNR)

Subject: Design and implementation of data wrangling and artificial intelligence techniques for the unsupervised identification of anomalies in large knowledge graphs in the area of scientific communication
Project: InfraScience
ID: ISTI 028/2022 - PI
Protocol number:
- ISTI n. 4115 - 07/12/2022

Activities:
- study of artificial intelligence techniques for entity blocking and deduplication quality evaluation in graphs
- study and implementation of new functionalities of a framework for entity linking and deduplication
- use of Natural Language Processing (NLP) techniques to code graph entities

Tools and technologies:
- D-NET software toolkit
- Graph Neural Networks
- Apache Hadoop, Maven, Spark
- Deep Graph Library (DGL)
- Software management: svn, git, Jenkins, maven
- Programming languages: Java, Python, Scala
- Intellij IDE

14/06/2021 - 31/12/2022 Pisa, Italy
RESEARCH FELLOW ISTITUTO DI SCIENZA E TECNOLOGIE DELL'INFORMAZIONE "A. FAEDO" (ISTI - CNR)

Subject: Design and implementation of data wrangling and artificial intelligence techniques for unsupervised anomaly detection in big scholarly communication graphs
Project: Blue-Cloud - Piloting innovative services for Marine Research & the Blue Economy
ID: ISTI 020/2021 - PI
Protocol number:
- ISTI n. 2090 - 08/06/2021
- ISTI n. 1746 - 08/06/2022 (first renewal)

Activities:
- study of artificial intelligence techniques for anomaly detection and entity classification in graphs
- development of a web portal for the deduplication of organization entities
- study and implementation of new functionalities of a framework for entity linking and deduplication

Tools and technologies:
- D-NET software toolkit
- Graph Neural Networks
- mongoDB
- Apache Hadoop, Maven, Spark
- Software management: svn, git, Jenkins, maven
- Programming languages: Java, Python, Scala

[redacted]

- Intellij IDE

30/11/2018 – 30/05/2021 Pisa, Italy
RESEARCH FELLOW ISTITUTO DI SCIENZA E TECNOLOGIE DELL'INFORMAZIONE "A. FAEDO" (ISTI - CNR)

Subject: Development of clustering based solution for entity classification and de-duplications on authors and organizations entities in Big Scholarly Communication Graphs
Project: OpenAIRE-Connect & OpenAIRE-Advance
ID: ISTI 018/2018 - PI
Protocol number:
- ISTI n. 4375 - 27/11/2018
- ISTI n. 4021 - 06/11/2019 (first renewal)
- ISTI n. 3638 - 25/11/2020 (second renewal)

Activities:
- deep study of Apache Spark framework and MapReduce model for processing and generating big sets with a parallel and distributed algorithm on a cluster
- development of a mechanism to identify group of organizations in a large graph by identifying similar entities
- development of system for the collection of user's feedback to enhance the de-duplication precision
- study of deep learning architecture for the identification of anomalies of big knowledge graphs

Tools and technologies:
- D-NET software toolkit
- Graph Neural Networks
- mongoDB
- Apache Hadoop, Maven, Spark
- Software management: svn, git, Jenkins, maven
- Programming languages: Java, Python, Scala
- Intellij IDE

31/05/2017 – 29/11/2018 Pisa, Italy
RESEARCH FELLOW ISTITUTO DI SCIENZA E TECNOLOGIE DELL'INFORMAZIONE "A. FAEDO" (ISTI - CNR)

Subject: Development of clustering based solution for entity classification and de-duplications on big graphs
Project: Curiosity Driven - NeMIS
ID: ISTI 005/2017 - PI
Protocol number:
- ISTI n. 2665 - 26/05/2017
- ISTI n.1896 - 15/05/2018 (first renewal)

Activities:
- deep study of Apache Spark framework and MapReduce model for processing and generating big sets with a parallel and distributed algorithm on a cluster
- extraction of authors from a dataset of publications by identifying a set of attribute to describe them
- building a large graph with authors as vertexes
- development of a clustering mechanism to identify group of authors in a large graph by drawing edges between similar authors
- identification of semantic relations between entities in a graph
- development of a mechanism to identify group of organizations in a large graph by identifying similar entities

Tools and technologies:
- D-NET software toolkit
- mongoDB
- Apache Hadoop, Maven, Spark
- Software management: svn, git, Jenkins, maven
- Programming languages: Java, Python, Scala
- Intellij IDE

EDUCATION AND TRAINING

30/11/2020 – CURRENT
PHD STUDENT IN COMPUTER ENGINEERING University of Pisa

Title: Artificial Intelligence for Open Science
Description: The research program of the PhD is focused on the study of new solution based on the Artificial Intelligence for the identification of anomalies in Big Scholarly Communication Graphs.
Supervisors: Paolo Manghi (ISTI - CNR), Fabrizio Falchi (ISTI - CNR), Marco Avvenuti (UniPi)

08/2013 – 04/2017 Pisa, Italy
LAUREA MAGISTRALE IN COMPUTER ENGINEERING (110/110 CUM LAUDE) Università di Pisa

Thesis title: Development of a mobile application for Food Recognition using Convolutional Neural Networks
Principal subjects/occupational skills covered: Intelligent Systems (29/30), Robotics and Automated Systems (24/30), Advanced Topics in Networked Computer Systems (30/30), Information Systems (30/30), Mobile and Pervasive Systems (30/30), Performance Evaluation (28/30), Software System Engineering (30/30), Electronics and Communication Systems (26/30), Advanced Network Architectures and Wireless Systems (30/30), Concurrent and Distributed Systems (27/30), Security in Networked Computer Systems (29/30), Multimedia Information Management (30/30)

09/2010 – 12/2013 Pisa, Italy
LAUREA TRIENNALE IN INGEGNERIA INFORMATICA (100/110) Università di Pisa

08/2005 – 05/2010 San Giovanni Rotondo, Italy
PERITO INDUSTRIALE CON SPECIALIZZAZIONE IN INFORMATICA (78/100) I.S.I.S. Luigi Di Maggio

LANGUAGE SKILLS

Mother tongue(s): **ITALIAN**

Other language(s):

	UNDERSTANDING		SPEAKING		WRITING
	Listening	Reading	Spoken production	Spoken interaction	
ENGLISH	C1	C1	C1	C1	C1

Levels: A1 and A2: Basic user; B1 and B2: Independent user; C1 and C2: Proficient user

ADDITIONAL INFORMATION

MASTER DEGREE THESIS

09/2016 – 05/2017
Development of a mobile application for Food Recognition using Convolutional Neural Networks

Supervisors: Giuseppe Amato (ISTI-CNR), Fabrizio Falchi (ISTI-CNR), Claudio Gennaro (ISTI-CNR)
Description: The work done in this thesis provides an efficient solution to implement a mobile application for food recognition using Convolutional Neural Networks (CNNs). Different CNNs architectures have been trained and tested on two datasets available in literature and the best one in terms of accuracy has been chosen. Since the CNN runs on a mobile phone, efficiency measurements have also taken into account both in terms of memory and computational requirements. The mobile application has been implemented relying on RenderScript and the weights of every layer have been serialized in different files stored in the mobile phone memory. Extensive experiments have been carried out to choose the optimal configuration and tuning parameters.
Article link: https://dx.doi.org/10.1007/978-3-319-98678-4_31
Principal subjects/occupational skills covered: Knowledge of different Convolutional Neural Networks architectures (VGG, AlexNet, GoogLeNet, SqueezeNet, Residual Networks).
Final mark: 30/30

[redacted]

PUBLICATIONS

Graph-based methods for Author Name Disambiguation: a survey – 2023
Authors: De Bonis M., Manghi P., Falchi F.
Source: PeerJ Computer Science
Type: Article
Link: https://peerj.com/articles/cs-1536/
DOI: 10.7717/peerj-cs.1536

A Graph Neural Network Approach for Evaluating Correctness of Groups of Duplicates – 2023
Authors: De Bonis M., Minutella F., Falchi F., Manghi P.
Source: TPDL 2023 - The 27th International Conference on Theory and Practice of Digital Libraries - 26-29 September - Zadar, Croatia
Type: Article
Link: http://dx.doi.org/10.1007/978-3-031-43849-3_18
DOI: 10.1007/978-3-031-43849-3_18

(Semi)automated disambiguation of scholarly repositories – 2023
Authors: Baglioni M., Mannocci A., Pavone G., De Bonis M., Manghi P.
Type: Article
Link: http://dx.doi.org/10.48550/arXiv.2307.02647
DOI: 10.48550/arXiv.2307.02647

FDup: a framework for general-purpose and efficient entity deduplication of record collections – 2022
Authors: De Bonis M., Manghi P., Atzori C.
Source: PeerJ Computer Science
Type: Article
Link: https://peerj.com/articles/cs-1058/
DOI: 10.7717/peerjcs.1058

A preliminary assessment of the article deduplication algorithm used for the OpenAIRE Research Graph
– 2022
Authors: Vichos K., De Bonis M., Kanellos I., Chatzopoulos S., Atzori C.
Source: IRCDL2022 - 18th Italian Research Conference on Digital Libraries
Type: Article
Link: http://ircdl2022.dei.unipd.it/downloads/papers/IRCDL_2022_paper_26.pdf

Towards unsupervised machine learning approaches for knowledge graphs – 2022
Authors: Minutella F., Falchi F., Manghi P., De Bonis M., Messina N.
Source: IRCDL2022 - 18th Italian Research Conference on Digital Libraries
Type: Article
Link: http://ircdl2022.dei.unipd.it/downloads/papers/IRCDL_2022_paper_10.pdf

OpenAIRE OpenOrgs Database – 2021
Authors: Artini, Michele; De Bonis, Michele; Manghi, Paolo; Atzori, Claudio; Bardi, Alessia; Baglioni, Miriam
Type: Software
Source: OpenOrgs project
DOI: 10.5281/zenodo.4751149

OpenAIRE Covid-19 publications, datasets, software and projects metadata – 2021
Authors: Manghi P., Atzori C., Bardi A., Baglioni M., Schirrwagen J., Dimitropoulos H., La Bruzzo S., Foufoulas I., Lohden A., Backer A., Mannocci A., Horst M., Czerniak A., Kiatropoulou K., Kokogiannaki A., De Bonis M., Artini M., Ottonello E., Lempesis A., Ioannidis A., Summan F.
Type: Dataset
DOI: 10.5281/zenodo.4736827

OpenAIRE research graph: dumps for research communities and initiatives – 2021
Authors: Manghi P., Atzori C., Bardi A., Baglioni M., Schirrwagen J., Dimitropoulos H., La Bruzzo S., Foufoulas I., Lohden A., Backer A., Mannocci A., Horst M., Czerniak A., Kiatropoulou K., Kokogiannaki A., De Bonis M., Artini M., Ottonello E., Lempesis A., Ioannidis A., Summan F.
Type: Dataset
DOI: 10.5281/zenodo.3974604

Entity deduplication in Big Data Graph for Scholarly Communication – 2020
Authors: Paolo Manghi, Claudio Atzori, Michele De Bonis, Alessia Bardi
Type: Article
DOI: 10.1108/DTA-09-2019-0163

The OpenAIRE research graph: third- party publishing APIs – 2019
Authors: Atzori C., Baglioni M., Bardi A., Manghi P., La Bruzzo S., De Bonis M., Dell'Amico A., Artini M., Mannocci A., Ottonello E.
Type: Technical report
Link: https://openportal.isti.cnr.it/doc?id=people_____::c745b5f9d10e021be60d81403d8c5125

The OpenAIRE Research Community Dashboard: on Blending Scientific Workflows and Scientific Publishing
– 2019
Authors: Miriam Baglioni, Alessia Bardi, Argiro Kokogiannaki, Paolo Manghi, Katerina Iatropoulou, Pedro Príncipe, André Vieira, Lars Nielsen, Harry Dimitropoulos, Ioannis Foufoulas, Natalia Manola, Claudio Atzori, Sandro La Bruzzo, Emma Lazzeri, Michele Artini, Michele De Bonis and Andrea Dell'Amico
Source: 23rd International Conference on Theory and Practice of Digital Libraries, Oslo Metropolitan University, 9-12 September 2019
Type: Article
DOI: 10.1007/978-3-030-30760-8_5
DOI: 10.5281/zenodo.3467104
DOI: 10.5281/zenodo.3467103

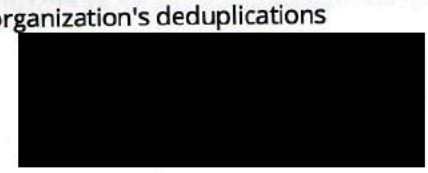
OpenAIRE Research Graph Dump – 2019
Authors: Manghi P., Atzori C., Bardi A., Schirrwagen J., Dimitropoulos H., La Bruzzo S., Foufoulas I., Loehden A., Baecker A., Mannocci A., Horst M., Baglioni M., Czerniak A., Kiatropoulou K., Kokogiannaki A., De Bonis M., Artini M., Ottonello E., Lempesis A., Nielsen L. H., Ioannidis A., Bigarella C., Summan F.
Type: Article
DOI: 10.5281/zenodo.3516918

Deep Learning Techniques for Visual Food Recognition on a Mobile App – 2018
Authors: De Bonis M., Manghi P., Amato G., Falchi F., Gennaro C.
Source: The eleventh edition of International Conference on Multimedia & Network Information Systems, University of Wroclaw, 12-14 September 2018
Type: Article
DOI: 10.1007/978-3-319-98678-4_31
Paper nominated to the Best Paper Award of MISSI 2018

OpenAIRE: The OpenScience European Infrastructure – 2017
Authors: Artini M., Atzori C., Baglioni M., Bardi A., Biagioni S., De Bonis M., Dell'Amico A., La Bruzzo S., Manghi P.
Source: Nineteenth International Conference on Grey Literature - Public Awareness and Access to Grey Literature, pp. 13–13, Roma, CNR, 23-24 October 2017
Type: Poster
Link: https://openportal.isti.cnr.it/doc?id=people_____::b24ac6f9527e59d0c4f5852a8a26c82a

PROJECTS

12/2019 – CURRENT
Openorgs Database
Description: In the OpenAIRE context, research organizations are aggregated from several datasources. This often leads to a duplication problem because an organization can be provided by multiple datasources.
The software available in this release provides the OpenOrgs web application: a web interface for the collection of user's feedbacks in the context of organizations deduplication.
An user can edit organization's metadata and approve or reject similarity relations suggested by the deduplication algorithm.
The deduplication algorithm takes advantage of user's feedback to increase the precision and the recall of the results.
The organizations resulting from the deduplication enhanced by the user feedback are indexed and subsequently exposed by the OpenAIRE portal.
Activities:
- project leader, programmer and designer of organization de-duplication tools
- Design and implementation of a web portal for the management of organization's deduplications



Link to the software: https://doi.org/10.5281/zenodo.4751149
Coordinators:
- Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo", ISTI-CNR
Project Managers: Michele Artini, Michele De Bonis

12/2017 – CURRENT
Recolecta Aggregator for FECYT
Typology: Technology transfer
Description: This project provides a data infrastructure for the aggregation of publications collected by various repositories.
Activities:
- project leader, programmer and designer of new functionalities for the aggregator
- Design and implementation of a web portal for the search (old portal: https://www.recolecta.fecyt.es/buscador)
- extension of D-NET functionalities to support specific needs for the aggregator
Coordinators:
- Fundación Española para la Ciencia y Tecnología, FECYT
- Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo", ISTI-CNR
Project Managers: Miriam Baglioni, Michele De Bonis

31/12/2020 – 30/06/2023
OpenAIRE-Nexus
Supervisor: Dr. Paolo Manghi, CNR - ISTI
Grant agreement ID: 101017452
Description: Development of clustering based solution for entity classification and de-duplications on authors and organizations entities in Big Scholarly Communication Graphs

30/11/2018 – 30/05/2021
OpenAIRE-Advance
Supervisor: Dr. Paolo Manghi, CNR - ISTI
Grant agreement ID: 777541
Description: Design and implementation of solutions for clustering and element classifications for entity resolutions in Big Scholarly Communication Graphs

30/11/2018 – 29/06/2019
OpenAIRE-Connect
Supervisor: Dr. Paolo Manghi, CNR - ISTI
Grant agreement ID: 731011
Description: Design and implementation of solutions for clustering and element classifications for entity resolutions in Big Scholarly Communication Graphs

31/05/2017 – 29/11/2018
Curiosity Driven - NeMIS
Supervisor: Dr. Paolo Manghi, CNR - ISTI
Description: Design and implementation of neural network based solution for clustering support and entity de-duplication and classification in a large graph

06/2016 – 08/2016
Development of a mobile application for geolocalization (AppHere)
Description: Project for "Mobile and Pervasive Systems" university course
Principal subjects/occupational skills covered: Android programming

04/2015 – 06/2015
Use of an Artificial Neural Network for predicting the trend in stock exchange market
Description: Project for "Intelligent Systems" university course
Principal subjects/occupational skills covered: MATLAB

ORGANISATIONAL SKILLS

Problem solving attitude Experience gained during several University Projects:
- Use of an Artificial Neural Network for predicting the trend in stock exchange market
- Development of a web site for reservations in multiplex cinemas
- Development of a mobile application for geolocalization
- Configuration and Setup of a Software Defined Network which allows the client to request and release a certain number of servers

And projects managed at ISTI-CNR as Research Fellow:

- Recolecta Aggregator for FECYT
- OpenOrgs Database

COMMUNICATION AND INTERPERSONAL SKILLS

Good communication skills gained through my experience as sales manager in tailor shop
Description: Summer job during high school period. I was responsible for the maintenance and the configuration of automatic embroidery machines.
Empoloyer: "Abiti da Lavoro" Fania - Via Antonio Rosmini, 32 - 71013 - San Giovanni Rotondo (FG)

JOB-RELATED SKILLS

Programming Skills

Programming Languages: Java, PHP, Python, Scala, SQL, JavaScript, C++
Databases: Postgres, MongoDB
Tools and Frameworks: Apache Maven, Apache Spark, Hadoop, HDFS
IDE: Eclipse, Intellij, PyCharm
All the skills have been obtained during the activity as Research Fellow at ISTI-CNR in the following projects:
- Curiosity Driven - NeMIS
- OpenAIRE-Connect
- OpenAIRE-Advance
- InfraScience
- FAIRCORE4EOSC

Deduplication skills

Experience gained during work at ISTI-CNR as Research Fellow in the following projects:
- OpenAIRE-Connect
- OpenAIRE-Advance
- OpenAIRE-Nexus
- FAIRCORE4EOSC

Clustering skills

Experience gained during work at ISTI-CNR as Research Fellow in the following projects:
- Curiosity Driven
- OpenAIRE-Connect
- OpenAIRE-Advance
- FAIRCORE4EOSC

Neural Networks and Deep Learning on Images and Knowledge Graphs

Deep Learning on Images skills gained at the University and for the work in the Master Degree Thesis.
Experience on graphs gained during work at ISTI-CNR as a Research Fellow and during PhD program at the University of Pisa.
Principal subject/occupational skills covered:
- knowledge of the Caffe and Torch frameworks for the implementation of Convolutional Neural Networks
- knowledge of Deep Graph Library for the implementation of AI architectures

Understanding of metadata for representation of data (XML)

Skill gained during work as Research Fellow at ISTI-CNR in the project "Recolecta Aggregator for FECYT".

Unix-like operating systems

Skill gained in whole career starting from high school, going through University and then at ISTI-CNR.