

Mario Cataldi
74 Rue du Faubourg Poissoniere, 75010 Paris
Tel. : +33 650849336
E-mail : mario.cataldi@ecp.fr
Web : <http://perso.ecp.fr/~cataldim/>

Date of Birth: 29/09/1981
Place of Birth: Moncalieri (TO), Italy
Nationality: Italian
Gender: Male

Curriculum Vitæ

Current Position

2012 - **Post-Doctoral Researcher** at the École Centrale Paris
(Advisor: Prof. Marie-Aude Aufaure).

Education

- 2012 **Qualification aux fonctions de Maître de Conférences** (Section 27 - *Informatique*) got on February 2012 (qualification number: 12227226397)
- 2007-2010 **Ph.D. in Computer Science** got on 19th November 2010 at the University of Torino, Department of Computer Science with the Ph.D. thesis entitled "Adaptation of Hierarchical Meta-Data for Efficient Large Data Set Exploration" under the supervision of Prof. Maria Luisa Sapino.
- 2003-2006 **Master's Degree** in Virtual Reality and Multimedia, final Mark: 110 / 110, got on 27 March 2006 at the University of Torino, Department of Computer Science.
- 2000-2003 **Bachelor's Degree** in Computer Science, final Mark: 100 / 110, got on 25 September 2003 at the University of Torino, Department of Computer Science.

Languages

Italian **Mother tongue.**
English **Proficient level, C1** (in reading, writing, and speaking).
Spanish **Proficient level, C2** (in reading, writing, and speaking).
French **Intermediate level, B2** (in reading, writing, and speaking).

Research Experiences

- 2012 - **Post-Doctoral Fellowship** at the École Centrale Paris, on *Graph Management, Social Graphs, Social Network Analysis*.
Supervisor: Prof. Marie-Aude Aufaure.
- 2011 - 2012 **Post-Doctoral Fellowship** at the University of Torino, Interdepartmental Research Center on Multimedia and Audiovisuals (CIRMA) on *Semantic Data, Ontologies and AI*.
Supervisor: Prof. Vincenzo Lombardo.
- 2008 **Visiting Ph.D. Student** at the Arizona State University (ASU) from August 30 2008 until December 15 2008 for research on *Data Structures Management, Database Summarization/Anonymization*.
Supervisor: Prof. K.S. Candan.
- 2007- 2010 **Ph.D** at the University of Torino, Dept. of Computer Science on *Information Retrieval, Data Mining, Knowledge Management, Large Data Sets Exploration*.
Supervisor: Prof. M.L. Sapino.

Date.....

1

Sign.....

Research Interests

- 2009 - **Large Data Sets Management:** during the last years, the amount of available information (on- and off-line) has enormously increased. For this, it is of practical importance to be able to automatically analyze large data set in order to provide novel mechanisms for efficient analysis, indexing, exploration and retrieval purposes. Within my research activities, in collaboration with the research groups I worked with, I investigated these problems developing innovative statistical-based data mining methodologies for retrieval, clustering and exploration purposes; in particular, we took into account user preferences, specific content domain properties and aggregations techniques to help the user analyze and explore very large data sets (especially databases and large text collections). To cope with this, we studied novel approaches for efficiently represent the data and help the user leverage hidden semantic relationships that exist among them. We then exploits these information to provide novel search and exploration methods that help the user navigate into the data space.
- 2010 - **Social Networks Analysis:** In order to facilitate the analysis of user-generated contents (always enormous in number), we studied novel topic detection and tracking (TDT) techniques to analyze and follow the evolution of the information expressed by a social network. Moreover, we also proposed novel measures for estimating authority and influence among the users in social environments. For this, we studied novel metrics to identify the relationships that exist among users, pages and contents, and therefore we map these information in a social graph where it is possible to analyze the information spread and contagion (by considering temporal conditions) within a social network community.
- 2009 - **Database Management, Summarization/Anonymization:** Considering the enormous amount of contents usually stored in databases, there is an emerging need for novel data summarization techniques that can reduce the amount of data by preserving the relevant information and reduce, where it is possible, the overall redundancy. Moreover, we also studied novel algorithms for anonymizing the information (where it is necessary for security reasons) by presenting to the user an abridged version of the data where the original contents are not uniquely associable to the presented ones. In order to do that, we developed novel efficient data graph management algorithms that help handle very large databases and simplify their access and the exploration through summarization and/or anonymization pre-processes.
- 2009 - **Ontology/Taxonomy Management and Semantic Data:** Ontologies and/or taxonomies embody formalized knowledge and define aggregations between concepts/categories in a given domain that can facilitate the organization of large data sets and can make their content easily accessible to the users. Thus, we developed novel taxonomy and ontology management algorithms for automatically use and adapt existing structures to the content expressed by dynamics data sets (information news, user-generated contents, etc). Finally, we leverage and exploit this ontological knowledge for efficiently indexing (and share) data contents.

Projects and Collaborations

DynamicTV 2007-2009 The main aim of the project was the development of a television contents recommender based on a novel automatic categorization approach; in fact, the considered categories can evolve in number and meanings depending on the real user context (for example, where the user lives and what happened in her real world). The project considered a very large set of un-structured and semi-structured data (blogs, news web sites, newspapers) as user context, and enriched the existing categories depending on the meanings assumed by them in the most recent contents. Therefore, the categories were organized in structured taxonomies depending on the semantics expressed by the considered domains. These taxonomies have been finally used for indexing the television contents by using novel classification algorithms; using these associations, the system was able to highlight hidden semantic relationships among indexed TV contents, defining the basis for an intelligent television contents recommender. The results of this project have been patented (Pub. No.: WO/2010/075889).
Position held : Research collaborator
Institution : Telecom Italia Lab, Torino, Italy

Cadmos 2011- The CADMOS project aims at designing a novel semantic annotation model for vast multimedia data sets (video, text, audio) that, following the principles of linked data, permits to annotate semantic concepts by using Artificial Intelligence techniques. The project consists in devising a BDI annotation model (which permits to model characters actions, goals, beliefs etc) that also leverages and exploits vast commonsense ontological knowledge (i.e., large scale ontologies) to help the user annotate resources within a semantic framework. Within this project, we also developed a web-based software platform to annotate and access the archived documents.
Position held : Research collaborator
Institution : CIRMA, Italy

CoOperare 2009-2010 The main aim of the project is to study how, in a Web 2.0 reality, the most important Italian museum institutions communicate with the potential visitors and how these visitors understand the message vehiculated by them. In order to do that, we analyzed a very large number of contents, related to the considered museums, extracted from the blogosphere (i.e., representing the users point of views) and the official web pages of the examined institutions (i.e. defining the message vehiculated by the museums to their public); thus, we compare these contents by using novel lexical and semantic mechanisms in order to highlight, from a sociological point of view, where and how they significantly differ.
Position held : Research collaborator
Institution : Dept. of Social Sciences, University of Torino, Italy

Publications

International Journals

- 2013 M. Cataldi, L. Di Caro, C. Schifanella.
Personalized Emerging Topic Detection based on a Term Aging Model.
ACM Transactions on Intelligent Systems and Technology (ACM TIST), 2013
(to appear)
- 2012 L. Di Caro, M. Cataldi, C. Schifanella.
The d-index: Discovering Dependences among Scientific Collaborators from their
Bibliographic Data Records.
International Journal of Scientometrics, 2012
- 2012 M. Cataldi, K.S. Candan, M.L. Sapino.
Narrative-based Taxonomy Distillation for Effective Indexing of Text Collections
International Journal of Data & Knowledge Engineering (DKE), 2012
- 2010 M. Cataldi, C. Schifanella, K.S. Candan, M.L. Sapino, L. Di Caro.
Context-informed Knowledge Extraction from Document Collections to Support
User Navigation.
Journal of Multimedia Processing and Technologies (JMPT), 2010

Chapters

- 2012 M. Cataldi, R. Damiano, V. Lombardo, A. Pizzo.
Lexical Mediation for Ontology-based Annotation of Multimedia
New Trends of Ontologies and Lexical Resources (Springer), 2012

International Conferences and Workshops

- 2013 M. Cataldi, N. Mittal, M-A. Afaure
Estimating Domain-based User Influence in Social Networks.
SAC 2013, *the 28th Symposium On Applied Computing*, 2013.
- 2012 L. Di Caro, C. Schifanella. , M. Cataldi, M-A. Afaure
D-INDEX: a web environment for analyzing dependences among scientific col-
laborators.
KDD 2012, *18th ACM SIGKDD Conference on Knowledge Discovery and Data
Mining*, 2012.
- 2012 M. Cataldi, R. Damiano, V. Lombardo, A. Pizzo.
An Agent-based Annotation Model for Narrative Media
AAMAS 2012, *11th International Conference on Autonomous Agents and Mul-
tiagent Systems*, 2012
- 2011 M. Cataldi, L. Di Caro, C. Schifanella.
ImmEx: IMMersive Text Documents EXploration System.
CBMi 2011, *9th International Workshop on Content-based Multimedia Indexing*,
2011

- 2011 M. Cataldi, R. Damiano, V. Lombardo, A. Pizzo, D. Sergi.
Integrating Commonsense Knowledge into the Semantic Annotation of Narrative Media Objects.
AIXIA 2011, *12th International Conference on Advances in Artificial Intelligence*, 2011, Springer
- 2011 M. Cataldi, R. Damiano, V. Lombardo, A. Pizzo.
Representing Dramatic Features of Stories through an Ontological Model.
ICIDS 2011, *4th International Conference on Interactive Digital Storytelling*, 2011, Springer
- 2010 M. Cataldi, K.S. Candan, M.L. Sapino.
ANITA: A Narrative Interpretation of Taxonomies for their Adaptation to Text Collections.
CIKM 2010, *19th International Conference on Information and Knowledge Management*, 2010, ACM.
- 2010 K.S. Candan, M. Cataldi, M.L. Sapino.
Reducing metadata complexity for faster table summarization.
EDBT 2010, *13th International Conference on Extending Database Technology* 2010, pp. 240-251, ACM.
- 2010 M. Cataldi, L. Di Caro, C. Schifanella.
Emerging Topic Detection on Twitter based on Temporal and Social Terms Evaluation.
MDM-KDD 2010 - 10th Multimedia Data Mining Workshop of KDD, 2010, ACM.
- 2010 S. Monaci, C. Schifanella, M. Cataldi.
Museums discourses: a comparative analysis of web communication and the blogosphere
EUROMED 2010 International Conference, Museum Futures: Emerging Social and Technological Paradigms, 2010, Springer.
- 2009 M. Cataldi, C. Schifanella, K.S. Candan, M.L. Sapino, L. Di Caro.
CoSeNa: a context-based search and navigation system.
MEDES 2009: International ACM Conference on Management of Emergent Digital EcoSystems, 2009. pp. 218-225, ACM.
- 2008 K.S. Candan, M. Cataldi, M.L. Sapino, and C. Schifanella.
Structure- and Extension-Informed Taxonomy Alignment.
ODBIS 2008, *4th Workshop of VLDB on Ontologies-based Techniques for DataBases in Information Systems and Knowledge Systems*, 2008. pp. 1-8, ACM.
- 2008 A. Favenza, M. Cataldi, M.L. Sapino, A. Messina.
Development Based Refinement of Audio-Segmented Television News.
NLDB 2008, *13th International Conference on Applications of Natural Language to Information Systems*, 2008. pp. 226-232, ACM.

Patents

- 2010 F. Antonelli, D. Mana, R. Simeoni, M. Geymonat, M. Cataldi, L. Di Caro, M.L. Sapino, K.S. Candan.
Method and System for Content Classification.
Pub. No.: WO/2010/075889 International Application No.:
PCT/EP2008/068356

Teaching

- 2012 **Professor of "Information Retrieval"**: Professor of the 24 hours course "Information Retrieval" within the Bachelor's Degree in Computer Science at the University of École Centrale Paris, France.
Program in short: Information filtering and retrieval, Standard Boolean model, Vector space model, Probabilistic Model, Relevance feedback.
- 2012 **Professor of "Algorithms and Programming"**: Professor of the 25 hours course "Algorithms and Programming" within the Bachelor's Degree in Computer Science at the University of École Centrale Paris, France.
Program in short: Programming in Python with Arrays, Lists, Tree Structures and Graphs.
- 2012 **Assistant Professor of "Database"**: Assistant Professor of the 25 hours course "Database" within the Bachelor's Degree in Computer Science at the University of École Centrale Paris, France.
Program in short: SQL, Relational algebra, Advanced Query Formulation.
- 2011 **Professor of "Digital Documentation for the Arts"**: Professor of the 54 hours course "Digital Documentation for the Arts" within the Bachelor's Degree in Drama, Art and Music Studies at the University of Torino, Italy.
Program in short: Semantic Web, Resource Description Framework (RDF), Web Ontology Language (OWL), eXtensible Markup Language (XML) Data Bases and Content Management Systems (CMS).

Seminars & Invited Talks

- Invited Talk **INRIA (Paris)** on 2012: *Influences and Dependences among Scientific Collaborators.*
- Seminar **University of Torino (Dept. of CIRMA)** on 2011: *Communication analysis in social media and social network environments.*
- Seminar **University of Torino (Dept. of Computer Science)** on 2010: *Intelligent approaches for exploration of very large text collections.*

Referees

- Conference **DEXA 2008**: Int. Conference on Database and Expert Systems Applications.
ACM Multimedia 2008: ACM Int. Conference on Multimedia.
SUM 2008: Int. Conference on Scalable Uncertainty Management.
SAC 2009: ACM Symposium on Applied Computing.
SIGMOD 2009: Int. Conference on Management of Data.
CIVR 2010: Int. Conference on Image and Video Retrieval.
KES 2010: Int. Conference on Knowledge-Based and Intelligent Information & Engineering Systems
SAC 2011: ACM Symposium on Applied Computing.
FOSINT-SI 2012: International Symposium on Foundation of Open Source Intelligence and Security Informatics.
DOLAP 2012: International Workshop On Data Warehousing and OLAP.
CIKM 2012: ACM Conference on Information and Knowledge Management.
ACM TOIT 2012: ACM Journal Transactions on Internet Technology.

Conference Talks

- 2012 **KDD 2012** Presentation of the paper: *D-INDEX: a web environment for analyzing dependences among scientific collaborators..* Beijing (China), August 2012.
- 2011 **AI*IA 2011** Presentation of the paper: *Integrating Commonsense Knowledge into the Semantic Annotation of Narrative Media Objects..* Palermo (Italy), September 2011.
- 2010 **MDM-KDD 2010** Presentation of the paper: *Emerging Topic Detection on Twitter based on Temporal and Social Terms Evaluation.* Washington (USA), July 2010.
- 2010 **EDBT 2010** Presentation of the paper: *Reducing Metadata complexity for faster table summarization.* Lausanne (Switzerland), March 2010.
- 2008 **ODBIS 2008** Presentation of the paper: *Structure- and Extension-Informed Taxonomy Alignment.* Auckland (New Zealand), August 2008.
- 2008 **NLDB 2008** Presentation of the paper: *Topic Development Based Refinement of Audio-Segmented Television News.* London (U.K.), June 2008.

Technical Skills and Competencies

- Languages Java, Python, C/C++/C#, SQL, XML, OWL, RDF.
O.S. MicroSoft Windows 95/98/Me/2000/XP/Vista/7, Unix and Unix-like systems, Mac OS.
- Tools MySql, Oracle, Matlab, Protege, Pellet, BigOwlim, Eclipse, NetBeans IDE, R, IntelliJ IDE, Microsoft Visual Studio .NET, Apache.

References

- 1st Prof. **M.L. Sapino**
Professor at the University of Torino, Italy
Dept. of Computer Science
email: mlsapino@di.unito.it
phone: +39 (011) 6706745
- 2nd Prof. **K. Selçuk Candan**
Professor at the Arizona State University, USA
Ira Fulton School of Engineering
email: candan@asu.edu
phone: +1 (480) 965-2770
- 3th Prof. **V. Lombardo**
Professor at the University of Torino, Italy
Dept. of Computer Science
email: lombardo@di.unito.it
phone: +39 (011) 6706722
- 4th Prof. **Marie-Aude Afaure**
Professor at the École Centrale Paris, France
MAS Laboratory
email: Marie-Aude.Afaure@ecp.fr
phone: +33 141 131 885