

## Diego Reforgiato Recupero

---

A.V. Williams Building  
UMIACS-UMD  
University of Maryland  
College Park, 20742

Phone: (301) 40-56730  
Fax: (301) 314-9658  
diegoref@umiacs.umd.edu  
<http://www.umiacs.umd.edu/~diegoref>

Dipartimento di Ingegneria Informatica  
e delle Telecomunicazioni  
DIIT  
University of Catania  
Viale A. Doria, 6 - 95125 Catania

Phone: +39 095 7382384  
Fax: +39 095 338280  
diegoref@diit.unict.it

Codice Fiscale

RFRDNG78B14C351V

### Education

Ph.D. Computer Science, University of Naples Federico II (Italy), 2004.  
Thesis title: Data Structures and Algorithms for Optimization and Searching Problems in Metric Spaces and Graphs.

B.S. Computer Science, University of Catania, 2001.  
Thesis title: Approximated Algorithms based on Random Tournaments and their applications to Optimization problems in Euclidean Spaces.  
*Summa Cum Laude*

### Research Experience

ICT Researcher June 2008–August 2008  
Alfio Lombardo University of Catania  
Research on MPEG-FGS encoder for P2P video streaming

Post Doc Researcher Apr 2005–May 2008  
VS Subrahmanian University of Maryland College Park - Maryland  
Development of a software for real-time mining of opinions expressed in written documents. Development of a software for querying annotated RDF data.

Visiting Researcher May 2008  
Sergio Greco Dipartimento di Elettronica Informatica e Sistemistica (DEIS) Cosenza - Italy  
Invited speaker by Professor Greco at DEIS on sentiment analysis and OASYS system.

Visiting Researcher September 2007  
Prendinger Helmut National Institute of Informatics (NII) Tokyo - Japan  
Invited speaker by Professor Prendinger at NII on sentiment analysis and Second Life.

Visiting Researcher August 2004–December 2004  
Mary Kathryn Harcombe Nashville - Tennessee  
PhD thesis supervised by a law clerk working for a Federal judge.

Visiting Researcher Jan 2004–July 2004  
Dennis Shasha New York University (NYU) - New York  
Development of a graph matching software and graph clustering tool.

## Employment

NeoDataGroup Consulting Researcher Text mining for documents classification and categorization	Catania, Italy March 2008 – to date
SentiMetrix Member of the Board of Directors Member of the Board of Directors at SentiMetrix Inc.	Bethesda, Maryland March 2008 – to date
SentiMetrix Co-founder and lead developer Co-founder of SentiMetrix and developer of OASYS engine at SentiMetrix Inc.	Bethesda, Maryland January 2007 – to date
NeoDataGroup Consulting Programmer Implementation of a system for automatic classification of Mediaset SMS messages and clustering on H3G user data.	Catania, Italy Nov 2004 – March 2005
Political Science Department, University of Catania Consulting Programmer Implementation of APQ (Accordo Programma Quadro).	Catania, Italy Nov 2004 – January 2005
Nokia Consulting Programmer Implementation of a billing system.	Catania, Italy Oct 2003 – Nov 2003
Istituto Maria Ausiliatrice High School Teacher Teacher of computer science and ECDL.	Catania, Italy Jan 2003
Tecno Work Consulting Programmer Implementation of a car antitheft system.	Catania, Italy Oct 2002 – Dec 2002
Cities On Line (TNET) Consulting Programmer Implementation of a web sites and a software to manage an internet point.	Catania, Italy Mar 2002 – Aug 2002
Department of Economic and Estimative Sciences, University of Catania Consulting Programmer and Researcher Implementation of a data mining system.	Catania, Italy Oct 2001 – Dec 2001

## Conference Papers

- “CIG: Cultural Island and Games”, John Dickerson, Maria Vanina Martinez, Diego Reforgiato, V.S. Subrahmanian, accepted to ICCCD 2008.
- “Sentiment Analysis: Adjectives and Adverbs are better than Adjectives Alone”, F. Benamara, C. Cesarano, A. Picariello, D. Reforgiato, VS Subrahmanian, in Proceedings of ICWSM 07 (2007), pp. 203-206.
- “The OASYS 2.0 Opinion Analysis System: A Demo”, C. Cesarano, A. Picariello, D. Reforgiato, VS Subrahmanian, in Proceedings of ICWSM 07 (2007) DEMO session, pp. 313-314.

“GraphBlast: A Fast and Universal Method for Querying Graphs”, R. Giugno, D. Skripin, A. Pulvirenti, G. Pigola, D. Reforgiato, M. Ragusa, C. Di Pietro, D. Shasha, A. Ferro, poster presentation at World Pharmaceutical Congress. Philadelphia May 23-25 (2006).

“Annotated RDF”, O. Udrea, D. Reforgiato Recupero, V.S. Subrahmanian, in Proceedings of ESWC European Semantic Web Conference (2006), Vol. 4011, pp. 487-501.

“OASYS: An Opinion Analysis System”, C. Cesarano, B. Dorr, A. Picariello, D. Reforgiato, A. Sagoff, V.S. Subrahmanian, in Proceedings of AAAI-2006 Spring Symposium on Computational Approaches to Analyzing Weblogs (2006), pp. 21-26.

“Clustered Trie Structures for Approximate Search in Hierarchical Objects Collections”, R. Giugno, A. Pulvirenti, D. Reforgiato in Proceedings of 3rd International Conference on Advances in Pattern Recognition ICAPR’05, LNCS (2005).

“Graphgrep: A Fast Universal Method for Querying Graphs”, R. Giugno, D. Shasha, D. Reforgiato, A. Pulvirenti, C. Di Pietro. M. Purrello, A. Ferro in Proceedings of 7th National Biotechnology Congress (2005).

“Fast Colorization of Gray Images”, G. Di Blasi, D. Reforgiato in Proceedings of Eurographics Italian Chapter (2003).

“Antipole Clustering For Fast Texture Synthesis”, S. Battiato, A. Pulvirenti, D. Reforgiato in Proceedings of Winter School of Computer Graphics WSCG (2003).

## Journal Papers

“Annotated RDF”, Octavian Udrea, Diego Reforgiato, VS Subrahmanian, accepted to ACM ToCL, 2008.

“AVA: Adjective-Verb-Adverb Combinations for Sentiment Analysis”, VS Subrahmanian, Diego Reforgiato, accepted to IEEE Intelligent Systems, 2008.

“A new unsupervised method for Document Clustering by using WordNet Lexical and Conceptual Relations”, Diego Reforgiato, Information Retrieval, Volume 10, Issue 6, Pages 563-579, 2007.

“Advanced Indexing Scheme for Imaging Applications: Three-Case Studies”, S. Battiato, G. Di Blasi, D. Reforgiato, IET Image Processing - September 2007 - Volume 1, Issue 3, p. 249-268.

“CARA: A Cultural Adversarial Reasoning Architecture”, VS Subrahmanian, M. Albanese, V. Martinez, D. Nau, D. Reforgiato, G. Simari, A. Sliva, O. Udrea, J. Wilkenfeld, IEEE Intelligent Systems, (2007) Vol. 22, N. 2, pp. 12-16.

“Antipole Tree Indexing to Support Range Search and K-Nearest-Neighbor Search in Metric Spaces”, D. Cantone, A. Ferro, A. Pulvirenti, D. Reforgiato, D. Shasha, IEEE Transactions on Knowledge and Data Engineering (TKDE), (2005) vol. 17, n.4, pp. 535-550.

## Book Chapters

“Efficient Graph Matching”, Diego Reforgiato, In Encyclopedia of Data Warehousing and Mining, Wang, J. (editor). To appear, 2008.

“Multimedia Presentation Databases”, V.S Subrahmanian, V. Martinez, D. Reforgiato. In Encyclopedia of Temporal Databases, (eds. T. Ozsu and L. Liu), Springer Verlag. 2008.

## Participation at Conferences

“International Conference on Scalable Uncertainty Management (SUM 2007)”, University of Maryland - College Park - Maryland 2007.

“Analyst Space for Exploration (A-SpaceX)”, Baltimore - Maryland 2007.

“First International Conference on Computational Cultural Dynamics”, College Park - Maryland 2007.

“International Conference on Weblogs and Social Media”, Boulder - Colorado 2007.

“AAAI Computational Approaches to Analyzing Weblogs”, Stanford - California 2006.

“Workshop on the Establishment of an International Center for Global Infectious Disease, Climate, Water and Health”, Washington DC 2005.

“7th National Biotechnology”, Catania - Italy 2004.

“Eurographic Italian Chapter”, Milan - Italy 2003.

“WSCG International Conferences in Central Europe on Computer Graphics, Visualization and Computer Vision”, Pilzen - Czech Republic 2003.

## Scientific Software

V. Martinez, D. Reforgiato, G. Simari, VS Subrahmanian (2007): **CAGE, a Cultural Adversarial Game Engine.**

CAGE is a game engine which allows the creation of customized games with the following characteristics. The games may consist of a set of game states, where each game state offers the player an array of information and a set of possible moves he can make. In this application, the game state consists of a multimedia (audio, video, image) presentation. The objects to be shown in a given game state (e.g. audio clip, video clip, text bar, etc.) can be obtained by querying a database and selecting a presentation method for the answer returned to the query. Each game state might also pose a question to the player with a set of answers to the question. Independently of which answer is correct, the game proceeds to a new state (similarly defined) that shows the player another body of information—perhaps real, perhaps hypothetical—and invites him to do something else. This continues until the end of the game when the player can be told what he did right and what he did wrong.

D. Reforgiato and VS Subrahmanian (2005): **GIDSTAR, Global Infectious Disease Surveillance Tracking and Analysis Repository.**

This is a software platform to gather information about diseases occurring around the world and track them in real-time so as to provide alerts to relevant public health officials well before a widespread outbreak occurs. This requires: (i) Extracting information about outbreaks from a wide variety of text-rich sources such as DCO and PDF files; (ii) Culling news reports from various countries around the world for outbreak information; (iii) Geo-referencing outbreak or symptom data and correlating this with land cover data, population density data, drainage data, temperature and weather data, road map data, and other demographic data; (iv) Mining this data for possible correlations between outbreak occurrences and such phenomena - this can often serve as an effective predictor of future outbreaks; (v) Providing appropriate notifications and suggested public health actions to take when an outbreak occurs.

C. Cesarano, B. Dorr, A. Picariello, D. Reforgiato, A. Sagoff and VS Subrahmanian (2005): **OASYS, an Opinion Analysis System.**

OASYS is a software for automatic opinion analysis of text document. Our OASYS system can

look at a collection of documents (in multiple languages) and assign an "intensity" of opinion of a given document on a given topic. The intensity of opinion of document  $d$  with regard to topic  $t$  depends not only on the terms used in the document, but also on the perceptions of the reader. As a consequence, we have developed statistical algorithms conditioned by human responses/input to create an intensity scoring model. Our prototype shows how the system behaves on English, Spanish, French, Arabic, Russian, Chinese, Korean, and Italian documents.

O. Udrea, D. Reforgiato and VS Subrahmanian (2005): **Annotated RDF**.

This research software is a general framework for RDF extensions and query answering. The Annotated RDF is an extension of RDF that builds upon annotated logic. In aRDF, you can start with any partial ordered set that you like as long as it has a bottom element. In essence an aRDF triple consists of an ordinary RDF triple together with an annotation.

S. Battiato, G. Di Blasi and D. Reforgiato (2005): **Antipole Tree for Fast Super Resolution**.

This software is built on top of the Antipole Tree Indexing. Given an input image  $s$ , the software learns how to sharpen  $s$  from a training set of sharp images containing low, mid and high frequency data and, at the end, it outputs the input image  $s$  at higher resolution. The reconstruction process is fast thank to the employment of the Antipole Tree data structure. The software is based on the method described by W. T. Freeman, T. R. Jones and E. C. Pasztor in *Example-based superresolution*, published in MERL TR-2001-340, 2001.

G. Di Blasi and D. Reforgiato (2005): **Antipole Tree for Fast Colorization of Gray Images**.

This software is built on top of the Antipole Tree Indexing; given a black and white image  $b$  and a set of colored images  $c_1, c_2, \dots, c_n$ , the software colors  $b$  according to the colors present in the images  $c_i, i = 1, \dots, c_n$ . The Antipole Tree data structure is used to speed up the reconstruction process. This software is based on the method described by E. Reinhard, M. Ashikhmin, B. Gooch and P. Shirley in *Color Transfer between Images*, published in Proc. of IEEE Computer Graphics and Applications, pp. 34-40, 2001.

S. Battiato, A. Pulvirenti and D. Reforgiato (2005): **Antipole Tree for Fast Texture Synthesis**.

This software is built on top of the Antipole Tree Indexing. Given a texture image  $t$ , the software replicates the texture patterns on a given noise image (usually bigger than  $t$ ). The Antipole Tree allows an exact and fast texture synthesis. This software is based on the method described by L. Y. Wey and M. Levoy in *Fast texture synthesis using tree-structured vector quantization*, published in Proc. of ACM-SIGGRAPH, pp. 479-488, 2000.

R. Giugno, A. Pulvirenti and D. Reforgiato (2005): **Clustered Trie Structures for Approximate Search in Hierarchical Objects Collections**.

Given a dataset of hierarchical objects and an object query, this software first represents the dataset in a trie structure and then returns all the objects in the dataset with distance within a certain threshold from the input query. The Antipole Tree provides a clustering of trie elements which allows a more effective filtering of the dataset improving the query time.

D. Cantone, A. Ferro, A. Pulvirenti, D. Reforgiato and D. Shasha (2005): **Antipole Tree Indexing to Support Range Search and K-Nearest-Neighbor Search in Metric Spaces**.

The Antipole Tree is a simple and efficient indexing scheme to support range queries and  $k$ -nearest neighbor queries in generic metric spaces. Its construction begins by first allocating a root  $r$  and then selecting two splitting points  $c_1, c_2$  in the input set, which become the children of  $r$ . Subsequently, the points in the input set are partitioned according to their proximity to the points  $c_1, c_2$ . Then one recursively constructs the tree rooted in  $c_i$  associated with the partition set of the elements closer to  $c_i$ , for  $i = 1, 2$ .

D. Reforgiato and D. Shasha (2004): **GraphClust**.

GraphClust is a tool that, given a dataset of labeled (directed and undirected) graphs, clusters the

graphs based on their topology. GraphClust consists of 16 different algorithms broken down along four binary dimensions: Number of clusters, Definition of substructures, Graph Type, Distance metric. For all algorithms, the procedure starts in the same way. First, all substructures are found for each graph. Then a matrix A is formed whose columns consist of the union of all substructures and for which there is one row for each graph. Each entry  $A[i,j]$  represents the number of substructures j in graph i. Once the matrix A is created both algorithms take all rows and cluster them using distances - either inner product or Euclidean distance - chosen by the user.

### Scientific Projects

Laboratory of Computational Cultural Dynamics : <http://www.umiacs.umd.edu/research/LCCD/>

GraphClust : <http://ftp.cs.nyu.edu/shasha/papers/GraphClust.html>

Bioinformatics : <http://alpha.dmi.unict.it/bioinformatics>

University of Catania and New York University : <http://alpha.dmi.unict.it/~ctnyu>

Texture Synthesis : <http://alpha.dmi.unict.it/~texture>

### Honors and Awards

Winner of Marie Curie International Reintegration Grant (IRG). Call: FP7-PEOPLE-IRG-2008 - Cut-off date 3 April 2008. Project Title: *P2P middleware for the deployment of an innovative business model for the provision of a QoS-aware video multicast transport service over the Internet*. Project Acronym: *PROVIDEO*.

Intellectual Property for UM Invention Disclosure IS-2006-076 OASYS version 1.0 (Opinion Analysis System), 2006.

Intellectual Property for UM Invention Disclosure IS-2006-085 Multilingual Opinion Analysis Architecture, 2006.

OASYS, an Opinion Analysis System, *Computer World Horizon Award Winner*, 2006.

OASYS Version 1.0 (Opinion Analysis System), finalist for *Invention of the Year 2006 at Office of Technology Commercialization* at University of Maryland, College Park.

*Premio di incentivazione* award (500 Euros) from University of Catania, 2000.

*Premio di incentivazione* award (500 Euros) from University of Catania, 1999.

*Premio di incentivazione* award (500 Euros) from University of Catania, 1998.

*Premio di incentivazione* award (500 Euros) from University of Catania, 1997.

### Media Citations

Work on OASYS has been written about in the journal Science.

<http://www.sciencemag.org/cgi/content/full/316/5824/534>

Work on OASYS was featured at a seminar on Capitol Hill organized by the American Association for the Advancement of Science. <http://www.aaas.org/news/releases/2007/0625insurgents.shtml>

Work on OASYS covered by *Sci.Cam.* (<http://www.sci-cam.it/>) online magazine, May 31, 2007.

RAI-TV (Italy) coverage of OASYS on November 6, 2006.

Work on OASYS covered by *Panorama* magazine, September 21, 2006.

Work on OASYS covered by *ComputerWorld* on August 21, 2006.

Work on OASYS covered by *Datamining.com* on July 11, 2007.

Work on OASYS covered by *A. James Clark School of Engineering* on April 20, 2008.

Work on OASYS covered by *Prnewswire.com* on April 24, 2008.

### Activities

Member of the Board of Directors for SentiMetrix Inc., (<http://www.sentimetrix.com>) Bethesda, USA.

Member of the Editorial Review Board of Scientific Journals International - area of "Computer Science, Computer Engineering, BCIS, MIS". ([http://www.scientificjournals.org/editorial\\_board.htm](http://www.scientificjournals.org/editorial_board.htm))

Reviewer, Book "Domain Driven Data Mining: Domain Problems and Applications" published by Springer.

Reviewer, Encyclopedia of Data Warehousing and Mining - 2nd Edition.

Reviewer, TKDE.

### Languages

Italian (native).

English (fluent level writing and speaking).

### Courses

*Foundations of Wide Area Network Programming.*

Lipari, July 2001. 13th International School for Computer Science Researchers.

*Interuniversity Mathematical School.*

Perugia, August 2003.

*Algorithmics for Data Mining and Pattern Discovery.*

Lipari, July 2003. 15th International School for Computer Science Researchers.

*Mobile Networks: Algorithms and Systems.*

Lipari, July 2004. 16th International School for Computer Science Researchers.

### Programming Skills

*Programming Languages:* Java, C, C++, Javascript, CSS, Visual Basic, SQL, HTML, ASP, Pascal, assembler, PHP, C#, JSP, Python, Perl, K, Scheme, Lingo, ActionScript.

*IDE:* IntelliJ IDEA, NetBeans 5.5, Sun Java Studio Creator, Visual Studio.

*Multimedia:* Macromedia Flash and Director MX 2004.

*Database management:* Oracle 8i, Postgresql, SQLServer, MySQL, MiniSQL.

*Operating Systems:* Linux, Unix, Windows (9x, 2000, XP, VISTA), MacOSx, Minix.

*Programming Paradigms:* object oriented (UML), structured, concurrent, declarative, functional.

*Security protocols, certificates and digital signatures.*

*P2P networks.*

*Semantic Web:* metadata, ontologies and Semantic Web Services: XML, RDF, OWL.

*Information Retrieval and searching:* crawling, indexing, ranking search engines, entity extraction.

*Cryptography:* DES, MD5.

*Data Fusion:* Dempster Shafer Theory, Kalman Filter.

*Neural networks and genetic algorithms.*

*3D-Modeling:* Google SketchUp.

*Natural Language Processing:* Link Grammar and Parser, WordNet.

*3D-Worlds:* SecondLife - Linden Script language and libsecondlife.

*Cell Phone:* Android Architecture and API.

## References

VS Subrahmanian, Director and Full Professor of UMIACS, University of Maryland College Park.  
Email: vs@cs.umd.edu Tel. (+1) 301.405.2711.

Dennis Shasha, Full Professor at New York University.  
Email: shasha@cs.nyu.edu Tel. (+1) 212.998.3078.

Antonio Picariello, Associate Professor at University of Naples Federico II.  
Email: picus@unina.it Tel. (+39) 081.768.38.26.

Vadim Kagan, President of SentiMetrix. <http://www.sentimetrix.com>  
Email: kagan@sentimetrix.com Tel. (+1) 240.498.5285.