

Debora Donato

CONTACT INFORMATION C/ Maestranza 37-39, 2,1
08003 Barcelona
Catalunya, Spain

Voice: +34 653715424
E-mail: debora.donato@yahoo.it
WWW: www.donade.net

CITIZENSHIP Italy

CURRENT POSITION Research scientist in Web Information Retrieval, working at Yahoo! Research Barcelona in the "Usage and Link Mining" group.

RESEARCH INTERESTS My main research interests include (but are not limited to):

- Characterization and modeling of Complex Networks (especially focused on the Web and Link Analysis Algorithms)
- Search and Relevance (especially focused on Web Spam Detection)
- Usage Mining
- Social Networks and Social Media Analysis

INTERNAL PROJECTS I am involved in a number of internal projects concerning search enhancement applications based on the analysis of query-logs.

Science team leader for Yahoo! Search Pad:

"Search Pad helps you track sites and make notes by intelligently detecting user research intent and automatically collecting sites the user visits. Search Pad turns on automatically when youre doing research, tracking sites to make document authoring a snap."
<http://www.ysearchblog.com/2009/07/07/unveiling-yahoo-search-pad/>

Science team for Search Assist Tools improvements:

This project is devoted to create analysis tools that focus on the behavior of users interacting with automatic recommendation system. The objective is to develop features and a ranking model to score suggestions in order to reduce the cognitive effort of users engaged in search activities.

Science team for developing a Reference Tool for Query-log Analysis:

This project aims to develop a reference tool for the analysis of query-logs. The tool is meant to minimize the effort of finding logical sessions and extracting features for a number of different applications ranging from query suggestion to ad-keywords generation.

EDUCATION **Università di "La Sapienza", Dipartimento di Informatica e Sistemistica**

Ph.D. in Computer Science Engineering, November 2002 -November 2005

- Dissertation: *Web Mining and Exploration: Algorithms and Experiments.*, 23 February 2006
- Advisor: Professor Stefano Leonardi
- External Committee: Professor Ricardo Baeza-Yates, Professor Massimo Santini

Degree in Electronic Engineering, December 2001. Honor thesis on PPP protocols

- Advisor: Professor Stefano Leonardi
- Rating: 110/110

I am listing some of my publications. For a complete list please refer to my website (www.donade.net).

Web Characterization and modeling of Complex Networks

Web characterization and modeling of Complex Networks was the original topic of my PhD thesis. Complex Networks is a new science devoted to understand the properties and model the behavior of scale-free networks. Research in this field requires handling huge data sets and hence studying and implementing secondary memory or data stream algorithms. Link analysis algorithms, spectral methods and machine learning (and more general data mining) skills are of fundamental importance for such kind of analysis.

Book Chapters

- [1] D. Donato: An introduction to the World Wide Web. In G. Caldarelli (eds) "Communication Networks: WWW and social networks", The Encyclopedia of Life Support Systems (EOLSS) (Invited, under preparation)
- [2] D. Donato, A. Gionis: Graph Mining for the Web. In H. Wang and C. Aggarwal (eds) "Managing and mining graph data", Springer. 2009 (invited, to appear)
- [3] D. Donato, L. Laura, S. Leonardi and S. Millozzi: Modeling the Webgraph: how far we are. In G. Caldarelli and A. Vespignani (eds) "Large structure and Dynamics of Complex Networks", World Scientific pp. 133-161. 2007

Journals

- [4] D. Donato, L. Laura, S. Leonardi, U. Meyer, S. Millozzi, J.F. Sibeyn. Algorithms and Experiments for the Webgraph. *Journal of Graph Algorithms and Applications* 10 (2), pp. 219 - 236, 2006.
- [5] A. Capocci, V.D.P. Servedio, F. Colaiori, L.S. Buriol, D. Donato, S. Leonardi, G. Caldarelli. Preferential attachment in the growth of social networks: The internet encyclopedia Wikipedia, *Physical Review E* 74, 036116, 6 pp., 2006.
- [6] D. Donato, L. Laura, S. Leonardi, S. Millozzi. Large Scale properties of the Webgraph. *European Journal of Physics B* 38, pp. 239-243, 2004.
- [7] D. Donato, L. Laura, S. Leonardi, S. Millozzi. Simulating the Webgraph: A Comparative Analysis of Models. *Computing in Science & Engineering* 6 (6), pp. 84-89, 2004.
- [8] N. Perra, V. Zlatic, A. Chessa, C. Conti, D. Donato, G. Caldarelli: PageRank Schroedinger-like equation: ranking top web pages through a local potential (Submitted for publication)
- [9] D. Donato, S. Leonardi, S. Millozzi, P. Tsaparas. Mining The Inner Structure of the Web Graph. *J. Phys. A: Math. Theor.* 41 224017, 12 pp., 2008.
- [10] D. Donato, S. Leonardi, P. Tsaparas. Stability and Similarity of Link Analysis Ranking Algorithms. Special Issue of *Internet Mathematics* devoted to the ANAW workshop 3 (4), pp. 479-507, 2008.
- [11] D. Donato, L. Laura, S. Leonardi and S. Millozzi. The Web as a graph: how far we are. *ACM Trans. Internet Technol.* 7 (1), 23 pp., 2007.

Proceedings

- [12] L. Becchetti, C. Castillo, D. Donato, A. Fazzone. A Comparison of Sampling Techniques for Web Characterization. *Procs. of LinkKDD*, Philadelphia, Pennsylvania, 8 pp. 2006

Search and Relevance

I am interested in studying methods to improve the relevance of the search engine results. A large fraction of this work was devoted to detect and demote Web spam. Web spam can significantly deteriorate the quality of results. Thus there is a large incentive for commercial search engines to detect spam pages efficiently and accurately.

Book Chapters

[13] D. Donato, A. Gionis: Next Generation Search. In G. Cormode (eds). Algorithms for Next Generation Networks . Springer. 2009

Journals [14] L. Becchetti, C. Castillo, D. Donato, S. Leonardi, R. Baeza-Yates. Link Analysis for Web Spam Detection. ACM Trans. Web 2 (1), pp.1-42, 2008.

[15] J. Xavier-Parreira, C. Castillo, D. Donato, S. Michel, G. Weikum. The JXP Method for Robust PageRank Approximation in Peer-to-Peer Web Search Network. VLDB Journal 17 (2), pp. 291-313, 2008.

Proceedings

[16] A. Ukkonen, C. Castillo, D. Donato, A. Gionis. Searching the Wikipedia with contextual information. (Poster) Procs of CIKM, Napa Valley, California, 2008.

[17] C. Castillo, D. Donato, A. Gionis, V. Murdock, F. Silvestri. Know your Neighbors. Web Spam Detection Using the Web Topology. Procs. of the 30th SIGIR, Amsterdam, The Netherlands, pp. 423-430, 2007

[18] J. Xavier-Parreira, D. Donato, S. Michel, G. Weikum. Efficient and Decentralized PageRank Approximation in a Peer-to-Peer Web Search Network. Procs. of the 32th VLDB06, Seoul, Korea, pp. 415-426, 2006

Usage Mining

My research is focused on the analysis of graphs that can be inferred by the semantic relations implicitly contained in the query log of a search engine. These graphs are based on different sources of information, like words in the query or clicked URLs in their answers, as well as their links or terms. Some potential uses might be the recognition of polysemous words, similarities measure among queries according to some closeness functions and the clustering of queries for various purposes, like ranking, query suggestions, logical session finding.

Proceedings

[19] F. Bonchi, C. Castillo, D. Donato, A. Gionis. Taxonomy-driven lumping for sequence mining.

[20] P. Boldi, F. Bonchi, C. Castillo, D. Donato, S. Vigna. Query Suggestions Using Query-Flow Graphs. Procs. of the ACM Workshop on Web Search Click Data (WSDM09). Barcelona, Spain, 10 pp 2009.

[21] P. Boldi, F. Bonchi, C. Castillo, D. Donato, A. Gionis, S. Vigna. The query-flow graph: model and applications. Procs of CIKM, Napa Valley, California, 10 pp., 2008.

[22] F. Bonchi, C. Castillo, D. Donato, A. Gionis. Topical query decomposition. Procs. of 14th ACM KDD'08, Las Vegas, Nevada, pp. 52-60, 2008

[23] C. Castillo, C. Corsi, D. Donato, P. Ferragina, A. Gionis. Query-log mining for detecting polysemy and spam. Procs. of WebKDD, Las Vegas, Nevada, 2008

Social Network and Social Media Analysis

The tasks of filtering and ranking in social systems are more complex compared to the classical domain since the distribution of quality has much higher variance: from very high-quality items to very low-quality. I have worked on the study of content quality evaluation and prediction. This research work is exploiting the rich amount and types of content and social interaction available in social media. This characteristic is an inherent advantages with respect to traditional collections of documents: in addition to document content and link structure social media exhibit a wide variety of user-to-document relation types and user-to-user interactions.

Proceedings

- [24] D. Donato, S. Leonardi, M. Panicia. Combining Transitive Trust and Negative Opinions for better Reputation Management in Social Networks. Procs. of SNA-KDD, Las Vegas, Nevada, 10 pp., 2008
- [25] E. Agichtein, C. Castillo, D. Donato, A. Gionis, G. Mishne. Finding high quality content in social media with an application to community-based question answering. Procs. of WSDM, Stanford, California, pp. 183-194, 2008
- [26] J. Xavier-Parreira, D. Donato, C. Castillo, G. Weikum. Computing Trusted Authority Scores in Peer-to-Peer Network. Procs. of AIRWeb, Banff, Canada, 10 pp., 2007
- [27] C. Castillo, D. Donato, A. Gionis. Estimating the Number of Citations using Author Reputation. Procs. of SPIRE, Santiago, Chile, pp.107-117, 2007.

PATENTS

I filed 3 patents in the field of spam detection (2007) and 8 patents concerning search enhancements using query-logs (2008-2009).

ACADEMIC EXPERIENCE

Università di “La Sapienza”, Dipartimento di Informatica e Sistemistica

Invited Lecturer May 2009

- Course of Seminars of Computer Networks, second level laurea degree.

Lecturer May 2008

- Course of Seminars of Computer Networks, second level laurea degree.

Teaching Assistant January 2002 - May 2006

- Courses of “Algorithms and Data Structures” and “Web Information Retrieval”.

Undergraduate Researcher January 2002 - August 2006

- Topics: Link and Temporal Analysis of the Web, Models and Algorithms for a dynamic study of the Web, Web graph measurements.

Max Planck Insitut fur Informatik, Saarbrucken, Germany, D5: Databases and Information Systems

Undergraduate Researcher May 2005 - August 2005

- Topic “Decentralized Ranking Algorithms”.

Basic Research Unit of University of Helsinki, Finland

Undergraduate Researcher

September 2004 - January 2005

- Topic “The inner structure of the Webgraph”.

RESEARCH PROJECT

Dynamically Evolving Large Scale Information Systems (DELIS), within the 6th framework of the EU - IST - FET initiative, 2004 - 2008

- In charge of deliverables and technical reports

COevolution and Self-organization In dynamical Networks (COSIN) - EU programme Information Society Technologies, IST-2001-33555, 2002 - 2005

- In charge of deliverables and technical reports

Algorithms for the Next Generation Internet and Web (ALGO-NEXT) Italian Ministry of University and Scientific and Technological Research, 2004.

Algorithms for Internet and the Web (ALINWEB). Italian Ministry of University and Scientific and Technological Research, 2002.

SERVICE

Organization of Events

Sponsorship chair for ECML-PKDD 2010.

Program Committee Co-Chair of WAW2009

Treasurer for WSDM 2009

Organizing Committee for MI09

PC Member

WWW 2010, ICALP 2010, HT 2009, ESA 2009, M3ESN-ICDE'09, SIGIR 2008, WWW08, HYPERTEXT 2008, INFOSCALE08, ECML PKDD 2008, SIGIR07, INFOSCALE07, WAW06

Journal Reviewer

Journal on Communications and Networking (JNC)

ACM Transactions on the Web (TWEB)

IEEE Transactions on Knowledge and Data Engineering (TKDE)

Theoretical Computer Science (TCS)

Knowledge and Information Systems (KAIS)

Journal of Web Semantics (JWS)

Information Processing & Management (IPM)

Invited Editor

Special Issue of Internet Mathematics devoted to the Workshop of Algorithm for the Web

National Science Foundation(NFS) Proposal Reviewer

OTHER
PROFESSIONAL
EXPERIENCE

Translation from English to Italian of the book "Great Ideas in Computer Science with JAVA", Alan W. Biermann and Dietolf Ramm, for Apogeo Publisher. 2004

Marketing Teacher in the course "95 RM tecnico di Web Marketing" - I.S.I.S 2003 - 2004

Computer Science Teacher at the "Collegio Universitario Lamaro Pozzani" Rome. 2002

Java programmer for Copernico, New Media Factory Company, Rome. 1997 - 1998

TECHNICAL SKILLS Programming: Java, Perl C, C++, SQL.

Machine Learning Tools: Expert with Weka.

Operating Systems: Microsoft Windows XP/2000, Apple OS X, Linux.

FOREIGN
LANGUAGES

Proficient in English and Italian(mother tongue). Intermediate Spanish.