Alexandra Olteanu

alexandra@aolteanu.com
http://www.aolteanu.com

Research Interests

I am interested in how data, methodological, and ethical constraints bound what we can learn from online social and behavioral data. I am particularly concerned about the ramifications of these boundaries when building tools and/or addressing questions about important societal issues such as discrimination, hate speech, and disaster relief.

Keywords: Computational Social Science, Social Computing, Crisis Computing, Social Systems, Social Media, Data Biases, Data Quality, Algorithmic Discrimination, Social Issues

Education

École Polytechnique Fédérale de Lausanne (EPFL), Switzerland　　Sept. 2011 -
Ph.D. in Computer and Communication Sciences　　**Jan. 2016**
Distributed Information Systems Laboratory, under the direction of Prof. Karl Aberer
Thesis: *Probing the Limits of Social Data: Biases, Methods and Domain Knowledge*

Vrije University of Amsterdam, Netherlands &　　Sept. 2009 -
"Politehnica" University of Bucharest, Romania　　**Aug. 2011**
Double-degree Research M.Sc. in Parallel and Distributed Computer Systems
Thesis: *P2P Social Networking*, under the direction of Prof. Guillaume Pierre

"Politehnica" University of Bucharest, Romania　　Sept. 2005 -
B.Sc. in Computer Science. Specialization in System Programming.　　**Aug. 2009**
Thesis: *Rescheduling and Error Recovery Algorithm for GRID Environments*,
under the direction of Prof. Valentin Cristea & Prof. Florin Pop.

Research & Work Experience

IBM Research, Thomas J. Watson Research Center, US　　Feb. 2017 - Present
Social Good Postdoctoral Fellow in the IBM Data Science Department　　*Yorktown Heights, New York, US*
and part of the IBM Science for Social Good Initiative

Data science projects that promote social good, with a focus on how offline events impact the prevalence and type of hate speech online and how various user aspects affect the perception of what constitutes hate speech. I also offered internal seminars on online social data biases, and co-mentored one of the summer interns.

École Polytechnique Fédérale de Lausanne (EPFL), Switzerland　　Sept. 2011 - Jan. 2016
Research Assistant in the Distributed Information Systems Laboratory,　　*Lausanne, Switzerland*
working with Prof. Karl Aberer

My work has focused on how to better gauge and assess web and social media data. Specifically, I worked on identifying and leveraging distinctive traits of data (e.g., source, type, credibility) within particular domains (e.g., climate change, crisis situations) and media (e.g., Twitter, news, websites) to build or guide the development of dedicated tools (e.g., data collection and filtering, credibility prediction), as well as to quantify various data pitfalls (e.g. collection, media or population biases).

Microsoft Research, US　　Jun. 2015 - Sept. 2015
Research Intern in the Context, Learning, and User Experience for　　*Redmond, Washington, US*
Search (CLUES) Group, working with Emre Kıcıman

I worked on the experimental design and evaluation methodology of the QuantifiedAll system which aims to extract information about the outcomes of people's actions and situations from social media in order to help others with e.g. decision-making or goal achievement.

Qatar Computing Research Institute (QCRI), Qatar　　Sept. 2013 - Feb. 2014
Research Associate Intern in the Social Computing Group, working　　*Doha, Qatar*
with Carlos Castillo

We showed that using a generic crisis-lexicon to collect migroblogged posts during mass emergencies leads to collections with higher recall than those obtained with crisis-specific keywords manually chosen by experts; it also helps to preserve the original distribution of message types. Since then, I also maintain crisislex.org, a repository of crisis-related social media data and tools.

Vrije University of Amsterdam (VU), Netherlands   Jan. - Aug. 2011
Research Master Project, working with Prof. Guillaume Pierre   *Amsterdam, Netherlands*

We showed that building robust and scalable online social networks on top of P2P systems is feasible from an architectural standpoint. In this regard, I implemented a dedicated simulation framework that scales to millions of nodes.

"Politehnica" University of Bucharest, Romania   Oct. 2009 - Aug. 2010
Research Assistant, working with Prof. Costin Boiangiu   *Bucharest, Romania*

I worked on designing image segmentation algorithms that improve the input of Optical Character Recognition (OCR) systems. Part of my work became part of a course on analysis and extraction of information in documents. My work was supported by CCS Content Conversion Specialists GmbH.

WebWise.ro, Romania   Jul. - Aug. 2009
Software Engineer (project-based)   *Slatina, Romania*

I implemented, using a classic LAMP stack, an application-specific content management system consisting of two types of interfaces for different levels of user privileges.

"Politehnica" University of Bucharest, Romania   Dec. 2008 - Aug. 2010
Research student collaborator, working with Prof. Valentin Cristea &   *Bucharest, Romania*
Prof. Florin Pop

Research projects on resource management in large-scale distributed systems. I evaluated a set of scheduling heuristics, employed data mining techniques to uncover system behavior patterns and implemented various extensions to the MONARC II simulation framework.

Publications

[WebSci'17 EA]	**The Limits of Abstract Evaluation Metrics: The Case of Hate Speech Detection** Alexandra Olteanu, Kartik Talamadupula, and Kush R. Varshney. *Proc. of the 9th International ACM Web Science Conference (WebSci'17), Troy, NY, US, June 2017.*
[CSCW'17, NetSci'17 Poster**]**	**Distilling the Outcomes of Personal Experiences: A Propensity-scored Analysis of Social Media.** Alexandra Olteanu, Onur Varol, and Emre Kıcıman. *Proc. of the 20th ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW'17), Portland, OR, US, February 2017. A variant of this work has also been presented at International School and Conference on Network Science (NetSci'17), Indianapolis, Indiana, US, June 2017.*
[ICWSM'16 Short Paper, IC2S2'16 Talk, CI'16 Poster**]**	**Towards an Open-Domain Framework for Distilling the Outcomes of Personal Experiences from Social Media Timelines.** Alexandra Olteanu, Onur Varol, and Emre Kıcıman. *Proc. of 10th International AAAI Conference on Web and Social Media (ICWSM'16), Cologne, Germany, May 2016. This work has also been presented at Collective Intelligence (CI'16), New York, US, June 2016, and at the International Conference on Computational Social Science (IC2S2'16), Illinois, US, June 2016.*
[AAAI SSS'16 EA]	**Characterizing the Demographics Behind the #BlackLivesMatter Movement.** Alexandra Olteanu, Ingmar Weber, and Daniel Gatica-Perez. *In Proc. of AAAI Spring Symposia on Observational Studies through Social Media and Other Human-Generated Content (SSS'16 OSSM), Stanford, US, March 2016.*
[ICWSM'15]	**Comparing Events Coverage in Online News and Social Media: The Case of Climate Change.** Alexandra Olteanu, Carlos Castillo, Nicholas Diakopoulos, and Karl Aberer. *In Proc. of 9th International AAAI Conference on Web and Social Media (ICWSM'15), Oxford, UK, May 2015.*

[CSCW'15]	**What to Expect When the Unexpected Happens: Social Media Communications Across Crises.** Alexandra Olteanu, Sarah Vieweg, and Carlos Castillo. *Proc. of 18th ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW'15), Vancouver, BC, Canada, March 2015.*
[WISE'14]	**Comparing the Predictive Capability of Social and Interest Affinity for Recommendations.** Alexandra Olteanu, Anne-Marie Kermarrec, and Karl Aberer. *Proc. of 15th International Conference on Web Information Systems Engineering (WISE'14), Thessaloniki, Greece, October 2014 (**Best paper award**).*
[ICWSM'14]	**CrisisLex: A Lexicon for Collecting and Filtering Microblogged Communications in Crises.** Alexandra Olteanu, Carlos Castillo, Fernando Diaz, and Sarah Vieweg. *Proc. of 8th International AAAI Conference on Weblogs and Social Media (ICWSM'14), Ann Arbor, US, June 2014.*
[CHI'13 EA]	**CredibleWeb: A Platform for Web Credibility Evaluation**. Zhicong Huang, Alexandra Olteanu, and Karl Aberer. *Proc. of CHI'13 Extended Abstracts on Human Factors in Computing Systems, Paris, France, May 2013.*
[ECIR'13]	**Web Credibility: Features Exploration and Credibility Prediction.** Alexandra Olteanu, Stanislav Peshterliev, Xin Liu, and Karl Aberer. *Proc. of the 34th European Conference on Information Retrieval (ECIR'13), Moscow, Russia, March 2013.*
[ICDM'12 Demo]	**Topick: Accurate Topic Distillation from Users Streams.** Anton Dimitrov, Alexandra Olteanu, Luke McDowell, and Karl Aberer. *Proc. of 13th IEEE International Conference on Data Mining (ICDM'12), Demo Session, Brussels, Belgium, December 2012.*
[CIKM'12]	**A Decentralized Recommender System for Effective Web Credibility Assessment.** Thanasis Papaioannou, Jean-Eudes Ranvier, Alexandra Olteanu, and Karl Aberer. *In Proc. of the 21th ACM International Conference on Information and Knowledge Management (CIKM'12), Hawaii, USA, November 2012.*
[SNS'12]	**Towards Robust and Scalable Peer-to-Peer Social Networks.** Alexandra Olteanu, and Guillaume Pierre. *Proc. of the 5th EuroSys' Workshop on Social Network Systems (SNS'12), Bern, Switzerland, May 2012 (**Best paper award**).*
[Elsevier CMA'12]	**A Dynamic Rescheduling Algorithm for Resource Management in Large Scale Dependable Distributed Systems.** Alexandra Olteanu, Florin Pop, Ciprian Dobre, and Valentin Cristea. *Computers & Mathematics with Applications 63(9), Elsevier; Volume 63(9), May 2012.*

Preprints

Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries. Alexandra Olteanu, Carlos Castillo, Fernando Diaz, and Emre Kıcıman. *Pre-print available at SSRN:* https://ssrn.com/abstract=2886526

Peer-reviewed Posters and Talks (not published)

[DSSG'17 Talk]	**Demystifying Social Entrepreneurship: An NLP Based Approach to Finding a Social Good Fellow.** Aditya Garg, Alexandra Olteanu, Richard B. Segal, Dmitriy A. Katz-Rogozhnikov, Keerthana Kumar, Joana Maria, Liza Mueller, Ben Beers, and Kush R. Varshney. *Data Science for Social Good Conference, Chicago, IL, September 2017.*
[GHC'10 Poster]	**Adaptive Scheduling Approach Used for Rescheduling in Distributed Systems.** Alexandra Olteanu, and Florin Pop. *Grace Hopper Celebration of Women in Computing Conference, Atlanta, USA, November 2010.*
[GHC'10 Poster]	**Adaptive Binarization Method with Variable Window Size.** Alexandra Olteanu, Alexandru Stefanescu, and Costin-Anton Boiangiu. *Grace Hopper Celebration of Women in Computing Conference , Atlanta, USA, November 2010.*

Tutorials

A Critical Review of Online Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries. <u>Alexandra Olteanu</u>, Emre Kıcıman, Carlos Castillo, and Fernando Diaz. *At the 23rd SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'17), Halifax, Canada, August 2017.* This is a revised version of the ICWSM'16 Tutorial.

A Critical Review of Online Social Data: Limitations, Ethical Challenges, and Current Solutions. Carlos Castillo, Fernando Diaz, Emre Kıcıman, and <u>Alexandra Olteanu</u> (Alphabetical Order). *At the 10th International AAAI Conference on Web and Social Media (ICWSM'16), Cologne, Germany, May 2016.*

Contributed Blog Posts

Demystifying Social Entrepreneurship: A Data-Driven Approach. IBM Research Blog, Science for Social Good, *September 2017.* https://www.ibm.com/blogs/research/2017/09/demystifying-social-entrepreneurship/

CrisisLex: A Lexicon for Collecting and Filtering Microblogged Communications in Crises. Follow the Crowd: A Blog for Researchers Studying Crowdsourcing, Human Computation, and Social Computing, *June 2014.* https://humancomputation.com/blog/?p=8818

Talks (this list does not include research paper presentations)

Using Social Media to Understand Collective and Personal Events: Challenges and Applications	Sept. 2017
• *Invited/Keynote Talk, ECML/PKDD Data Science for Social Good Workshop, Skopje, Macedonia*	
Hate Speech: Thoughts on the Role of User Aspects and External Events	Sept. 2017
• *Talk, Dagstuhl Seminar on Cybersafety in Modern Online Social Networks, Dagstuhl, Germany*	
Challenges, Problem Selection, and Impact: A Biased Perspective on PhD Studies	
• *Invited Talk, PhD Forum, International Conference on Data Mining, Barcelona, Spain*	Dec. 2016
Social Data Biases and the "Garbage In-Garbage Out" Principle	
• *Guest Lecture (remote), Applied Data Analysis Course, EPFL, Switzerland*	Dec. 2016
The True Colors of Depicting Real-World Events with Social Media	
• *Seminar, UN Global Pulse, NY, US*	Apr. 2017
• *Seminar, IBM T.J. Watson Research Centre, US*	Nov. 2016
• *Seminar, IBM Research-Almaden, US*	Sept. 2016
• *Seminar (remote), Northeastern University, US*	Sept. 2016
• *Seminar, Columbia University, New York, US*	Aug. 2016
• *Seminar, NEC Labs, Princeton, US*	Aug. 2016
• *Seminar, Microsoft Research, Redmond, US*	May 2016

Awards & Grants

ACM-W Scholarship	2015
for 2015 ACM Conference on Computer-Supported Cooperative Work and Social Computing	
Student Travel Grant	2015
for 8th and 9th International AAAI Conference on Web and Social Media	-2014
Best Paper Award	2014
at 15th International Conference on Web Information Systems Engineering (WISE'14)	
Best Paper Award	2012
at 5th EuroSys' Workshop on Social Network Systems (SNS'12)	
Google Global Community Scholarship	2010
for Grace Hopper Celebration of Women in Computing	

"Dinu Patriciu" Private Scholarship	2010

"Dinu Patriciu" Private Scholarship 2010
for excellent academic results

Research scholarship 2009
in cooperation with CCS Content Conversion Specialists GmbH

3rd place Image Processing Session 2010
3rd place Managerial Communication Session 2006
Student Scientific Communications at "Politehnica" University of Bucharest

Selected Media Coverage & Impact

CrisisLex Taxonomies Available in GDELT Global Knowledge Graph
GDELT Project, July 2015
http://blog.gdeltproject.org/crisislex-taxonomies-now-available-in-gkg/

Why You Shouldn't Only Get Your Climate Change News from the Mainstream Media
The Washington Post, April 2015
http://www.washingtonpost.com/news/energy-environment/wp/2015/04/21/how-to-get-twitter-riled-up-about-global-warming/

Earthquakes, Hurricanes and Other Disasters on Twitter
La Stampa (in Italian), March 2015
http://www.lastampa.it/2015/03/26/multimedia/tecnologia/terremoti-uragani-e-altre-calamitcos-li-racconta-twitter-jQGdKvF6Dq61Q3axYkfmUI/pagina.html

Crisis Tweets Study Identifies What to Expect During Emergencies
Emergency Management, February 2015
http://www.emergencymgmt.com/training/Crisis-Tweets-Study-Expect-During-Emergencies.html

The Role of Twitter During Emergencies and Disasters
Sky TG25 (in Italian), February 2015
http://tg24.sky.it/tg24/mondo/2015/02/18/twitter_ruolo_disastri_studio_differenti_tipi_di_informazioni.html

Could This Be The Most Comprehensive Study of Crisis Tweets Yet?
iRevolutions Blog, February 2015
http://irevolution.net/2015/02/16/comprehensive-crisis-tweets-study/

World Humanitarian Data and Trends 2014
United Nations Office for the Coordination of Humanitarian Affairs, December 2014
http://www.unocha.org/data-and-trends-2014/

Professional Service

Steering Committee:

Conference on Fairness, Transparency, and Accountability (FAT*): 2018

Program Committee (Conferences):

Conference on Fairness, Transparency, and Accountability (FAT*): 2018
The Web Conference (former WWW Conference): 2018 (Web and Society Track)
AAAI International Conference on Web and Social Media (ICWSM): 2016, 2017
ACM Conference on Research and Development in Information Retrieval (SIGIR): 2016, 2017 (short papers)
ACM Conference on Information and Knowledge Management (CIKM): 2017 (short papers, case studies)
ACM Web Science Conference (WebSci): 2017

Program Committee (Symposia, Workshops, Contests):

Workshop on Social Web for Disaster Management (SWDM, co-located with WSDM): 2018
Workshop on Data Science for Social Good (SoGood, co-located with ECML PKDD): 2017
ACM SIGAI Essay Contest: 2017
Internation Conference on Computational Social Science (IC2S2): 2017
Fairness, Accountability, and Transparency on World Wide Web (FAT/WEB, co-located with WWW): 2017
Observational Studies on Social Media Workshop (OSSM, co-located with ICWSM): 2017

External Reviewer:

ACM Conference on Computer Human Interaction (CHI): 2016, 2017, 2018
ACM Conference on Computer Supported Collaborative Work & Social Computing (CSCW): 2016
International Conference on Web Information Systems Engineering (WISE): 2012, 2013, 2014
ACM Conference on Very Large Data Bases (VLDB): 2013
IEEE International Conference on Data Engineering (ICDE): 2013
International Workshop on Self-Organizing Systems (IWSOS): 2013
International Semantic Web Conference (ISWC): 2012
International Conference on Distributed Computing Systems (ICDCS): 2012

Journal Reviewer:

Communications of the ACM: 2017
International Journal On Semantic Web and Information Systems (Special Issue on the Ubiquitous Semantic Web): 2013

Teaching and Advising Activities

IBM Research, T.J. Watson Research Center
Intern co-mentoring

- *Demystifying Social Entrepreneurship: A Data Driven Approach* (Aditya Garg, Columbia University) — Summer 2017

Seminars series

Biases, Methodological Pitfalls, and Ethical Boundaries When Analyzing Social Data — Jun. 2017
- Session 1: *General Challenges When Analyzing Social Data*
- Session 2: *Issues at the Data Source and Along the Data Processing Pipeline*
- Session 3: *Evaluation Challenges and Ethical Boundaries*

École Polytechnique Fédérale de Lausanne (EPFL)
Teaching Assistant

- *Probabilities and Statistics* (bachelor level) — Spring 2015
- *Analysis I* (bachelor level) — Fall 2014
- *Algorithms* (bachelor level) — Fall 2012

Advising Activities

- *Twitter as an Indicator of Food Security* (M.Sc. Thesis, Busser Alexander) — Spring 2015
- *The Impact of Twitter on Stock Markets* (Semester Project, Victor Kristof) — Fall 2014
- *An Empirical Analysis of the Evolution of Climate Change Discussions in Twitter* (Semester Project, Antoine Bastien) — Fall 2014
- *Debating Climate Change: How to Identify Credible Information and Sources from Twitter* (M.Sc. Thesis, Ashish Bindal) — Spring 2014
- *Robust Web Content Evaluation Systems* (M.Sc. Thesis, Bogdan Stoica) — Fall 2013
- *A Platform for Web Content Evaluation* (Semester Project, Zhicong Huang) — Fall 2012
- *Assess Web Credibility* (Semester Project, Stanislav Peshterliev) — Spring 2012
- *Inferring User Expertise from Online Social Network Content* (Semester Project, Anton Dimitrov) — Spring 2012

"Politehnica" University of Bucharest
Advising Activities

- *Data Transmissions for Multimedia* (class projects, M.Sc. level) — Fall 2009
- *Analysis and Extraction of Information in Documents* (class projects, M.Sc. level) — Spring 2010
- *Computational Geometry* (class projects, M.Sc. level) — Spring 2010
- *Summer school for developing image processing applications* (instructor) — Summer 2010

Skills

Coding: Python (current projects); C++, Java, C (past projects); PHP, CSS, HTML (occasionally)

Miscellaneous

Languages: English (fluent), Romanian (native), French, Spanish, Italian (basic)

Hobbies: Snowboarding, cooking, swimming. In a past life I also used to paint clocks and mirror frames. Once I also cooked for 120 people.