

# Eli Cortez

Redmond/WA, USA  
☎ (425) 516-1601  
✉ [elicortez@gmail.com](mailto:elicortez@gmail.com)  
📄 [cortez.me/](http://cortez.me/)

## Work and Research Interests

**Data Sciences, Big Data, Cloud Computing, data mining, information retrieval and machine learning**, with an emphasis on Data Analysis, information extraction, data integration, classification and web crawling.

## Education

2009-2012 **Ph.D. Computer Science**, Federal University of Amazonas - UFAM, December 2012.

Thesis Title: Unsupervised Information Extraction by Text Segmentation

Advisor: Dr. Altigran Soares da Silva

Comitee: Dr. Divesh Srivastava, Dr. Alberto Laender, Dr. Caetano Traina Jr., Dr. João Cavalcanti

2008-2009 **M.Sc. Computer Science**, Federal University of Amazonas - UFAM, April 2009.

Dissertation Title: A flexible approach for extracting metadata from bibliographic citations

Advisor: Dr. Altigran Soares da Silva

2004-2007 **B.S. Computer Science**, Federal University of Amazonas - UFAM, December 2007.

## Employment

2016-Present Senior Software Engineer (Machine Learning Specialist) - Azure Compute - **Microsoft**, Redmond, WA

**Project(s):**

- Applying Machine Learning to empower and manage Azure Data-centers and compute resources.

2014-2016 Senior Data & Applied Scientist - OneDrive and XBOX - **Microsoft**, Redmond, WA

**Project Lead:**

- Text analytics - Consumer Intelligence
- XBOX Studios and Live - Data Analysis and Machine Learning
- User Segmentation

- 2013-2014 Visiting Researcher - Microsoft Research - **Microsoft**, Redmond, WA
- Project(s):**
- Database Column Annotation
  - Tuning procedures from Schema Matching systems
- Contributions:**
- Proposal, Implementation and Development of a New Data Mining System in C# to perform Database column annotation using corporate data.
  - Implementation and Development in C# of a Optimization Tuning algorithm that is being tested in a Microsoft's Schema Matching systems.
- 2009-2013 Co-Founder and Chief Scientist at **Neemu Technologies**, Manaus, Brazil
- Project(s):**
- E-commerce product classification systems.
  - Crawling and Extraction of product specifications.
  - Enrichment and Data cleaning of E-commerce catalogs.
- Contributions:**
- Proposal, Implementation and Development of a state-of-the-art Machine Learning product classification system in JAVA used in e-commerce marketplaces.
  - Design of the architecture of a web-scale system in JAVA that automatically crawls and extract e-commerce information from the web.
  - Implementation of a new system in JAVA to remove noise information from e-commerce catalogues, mostly related to product specifications.
- Summer 2012 Research Intern within the Database group, **Microsoft Research**, Redmond, WA
- Project(s):**
- Memory management in streaming environments.
- Contributions:**
- Proposal, Implementation and Development of a memory management component in C# that enable streaming systems to increase their performance in single machines.
- February 2012 Visiting Researcher within the **Polytechnic Institute of New York University - NYU**, New York, NY
- Project(s):**
- Entity Extraction and Tracking in social networks
- Contributions:**
- Proposal and Implementation of a prototype in Perl that uses machine learning to allow the automatically extraction and tracking of entities (celebrities, enterprises, places) that are mentioned in Twitter.
- Summer 2009 Visiting Researcher within the Database Group, **University of Alberta - UofA**, Edmonton, Canada
- Project(s):**
- Flexible Data Exchange
- Contributions:**
- Proposal and Implementation of a system in JAVA that enables non-expert users to exchange their XML data in a Peer-to-Peer context relying on data mining methods.

2005-2012 Research Student within the Database and Information Retrieval Group, **Federal University of Amazonas - UFAM**, Manaus, Brazil

**Project(s):**

- Metadata Extraction from Bibliographic Citations
- Blocking Algorithms for replica detection
- Information Extraction from unstructured data records

**Contributions:**

- Proposal and Implementation of a state-of-the-art system in JAVA that performs metadata extraction from Bibliographic Citations.
- Proposal and Implementation of a Blocking system in JAVA that uses Genetic Programming to scale out replica detection in database records.
- Proposal and Implementation of 3 state-of-the-art systems in JAVA that, in a fully automatically and unsupervised fashion, allows for the extraction of relevant data from unstructured data records.

---

## Research Impact

The conducted research and its results were published in top tier venues. These results were published in 4 of the top 6 venues of the Databases & Information Systems area according to the rank provided by Google Scholar<sup>1</sup>. In July 2016 Google Scholar reported over 150 citations to 6 of the 24 articles listed below.

In the academic realm, the conducted research received Best Paper and Best Demo awards. Also, the PhD Thesis was awarded as the Best Computer Science PhD Thesis of 2012 by the Brazilian Computer Society. Besides, it has received an Award Honorable Mention by the Ministry of Education in Brazil. For this reason, the results were present in tutorials of Brazilian and International venues. The research has produced not only academic results, but also, industrial results. The developed techniques are in use in a startup company called Neemu Technologies<sup>2</sup>, which currently holds the largest market share in search systems for e-commerce in Brazil.

---

## Awards

- 2014 **HackDay 2014**, Winner of the OneDrive & SharePoint HackDay - Microsoft.
- 2013 **SBC Ph.D. Thesis Award**, Awarded by the Brazilian Computer Society - SBC as the best CS PhD thesis of 2012.
- 2013 **CAPES Ph.D. Thesis Award Honorable Mention**, Award given by the Coordination for the Improvement of Higher Education Personnel (CAPES) - Ministry of Education in Brazil.
- 2011 **Best Tool Award**, in Brazilian Symposium on Databases (SBBD).
- 2010 **Master Dissertation Award**, Dissertation ranked among the 11 best at the Brazilian Computer Society Thesis and Dissertation Contest (CTD).
- 2009 **Best Paper Award**, in Brazilian Symposium on Databases (SBBD).
- 2009-2012 **Ph.D. Scholarship Award**, given by UOL Inc.
- 2008 **M.Sc. Scholarship Award**, given by UOL Inc.
- 2008 **ACM SIGMOD Undergraduate Award**.

---

<sup>1</sup>[http://scholar.google.com/citations?view\\_op=top\\_venues&hl=en&vq=eng\\_databasesinformationsystems](http://scholar.google.com/citations?view_op=top_venues&hl=en&vq=eng_databasesinformationsystems)

<sup>2</sup><http://neemu.com>

2006-2007 **Youth Potential Research Award**, given by the Computer Science Department to the top five undergrad students with the highest GPA.

---

## Technical Skills

**Strong Knowledge on Information Retrieval, Big Data Applications and Machine Learning.**

**Programming Languages:** Java, Perl, C#, PHP and Shell Script

**Web Development:** HTML and JSON

**Databases:** MySQL, MongoDB and SQL Server

**Programming Tools:** Git, Eclipse and Visual Studio

---

## Teaching and Mentoring Experience

2011-2012 **Mentoring**, CS Department - Federal University of Amazonas , Manaus-Brazil.  
André Porto, undergraduate student

2011-2012 **Mentoring**, CS Department - Federal University of Amazonas , Manaus-Brazil.  
Elton Serra, graduate student

2010 **Teacher**, CS Department - Federal University of Amazonas , Manaus-Brazil.  
Course: Foundations and Theory of Computer Science

2009 **Teacher Assistant**, CS Department - Federal University of Amazonas , Manaus-Brazil.  
Course: Database Systems

2007 **Teacher Assistant**, CS Department - Federal University of Amazonas , Manaus-Brazil.  
Course: Algorithms

---

## Publications

### Books (1)

- [1] **Eli Cortez and Altigran Silva**, 2013, Unsupervised Information Extraction by Text Segmentation, In SpringerBriefs in Computer Science - Springer International Publishing.

### Refereed Conference Papers (12)

- [2] **Badrish Chandramouli, Justin Levandoski, Eli Cortez**, 2016, ICE: Managing Cold State for Big Data Applications, In IEEE International Conference on Data Engineering- ICDE..
- [3] **Eli Cortez, Altigran Silva, Daniel Oliveira, Edleno Moura and Alberto Laender**, 2011, Joint Unsupervised Structure Discovery and Information Extraction, In ACM International Conference on Management of Data - SIGMOD. (Acceptance rate: 25%).
- [4] **André Porto, Eli Cortez, Altigran Silva and Edleno Moura**, 2011, Unsupervised Information Extraction with the ONDUX Tool, In Brazilian Symposium on Databases - SBBD.

- [5] **Alberto Laender, Eli Cortez et al**, 2011, Building a Research Social Network from an Individual Perspective, In ACM/IEEE Joint Conference on Digital Libraries - JCDL. (Acceptance rate: 23%).
- [6] **Alberto Laender, Eli Cortez et al**, 2011, CiênciaBrasil - The Brazilian Portal of Science and Technology, In Integrated Seminar of Software and Hardware - SEMISH.
- [7] **Eli Cortez, Altigran Silva, Marcos Gonçalves and Edleno Moura**, 2010, ON-DUX: on-demand unsupervised learning for information extraction, In ACM International Conference on Management of Data - SIGMOD. (Acceptance rate: 21%).
- [8] **Guilherme Toda, Eli Cortez, Filipe Mesquita, Altigran Silva, Edleno Moura and Marden Neubert**, 2009, Automatically filling form-based web interfaces with free text inputs, In International Conference on World Wide Web - WWW. (Acceptance rate: 24%).
- [9] **Luiz Evangelista, Eli Cortez, Altigran Silva and Wagner Meira**, 2009, Adaptable Record Blocking using Genetic Programming, In Brazilian Symposium on Databases - SBBD.
- [10] **Eli Cortez and Altigran Silva**, 2008, A Flexible Approach for Extracting Metadata from Bibliographic Citations, In ACM International Conference on Management of Data - SIGMOD - Undergraduate Posters.
- [11] **Eli Cortez, Altigran Silva, Marcos Gonçalves, Filipe Mesquita and Edleno Moura**, 2008, FLUX-CiM Tool: flexible unsupervised extraction of citation metadata, In Brazilian Symposium on Databases - SBBD.
- [12] **Eli Cortez, Altigran Silva, Marcos Gonçalves, Filipe Mesquita and Edleno Moura**, 2007, FLUX-CiM: flexible unsupervised extraction of citation metadata, In ACM/IEEE Joint Conference on Digital Libraries - JCDL. (Acceptance rate: 36%).
- [13] **Roberto Santos, Filipe Mesquisa, Altigran Silva and Eli Cortez**, 2006, Data and metadata extraction in semi-structured text using HMMs, In Brazilian Symposium on Databases - SBBD.

#### Journal Articles (6)

- [14] **Eli Cortez, Philip A. Bernstein, Yeye He, Lev Novik:**, 2015, Annotating Database Schemas to Help Enterprise Search, In Proceedings of the VLDB Endowment - PVLDB.
- [15] **Eli Cortez, Mauro Herrera, Altigran Silva, Edleno Moura and Marden Neubert**, 2011, Lightweight methods for Large-Scale Product Categorization, In Journal of the American Society for Information Science and Technology - JASIST.
- [16] **Elton Serra, Eli Cortez, Altigran Silva and Edleno Moura**, 2011, On Using Wikipedia to Build Knowledge Bases for Information Extraction by Text Segmentation, In Journal of the Information and Data Management - JIDM.
- [17] **Guilherme Toda, Eli Cortez, Altigran Silva and Edleno Moura**, 2011, A Probabilistic Approach for Automatically Filling Form-Based Web Interfaces, In Proceedings of the VLDB Endowment - PVLDB.
- [18] **Luiz Evangelista, Eli Cortez, Altigran Silva and Wagner Meira**, 2010, Adaptive and Flexible Blocking for Record Linkage Tasks, In Journal of the Information and Data Management - JIDM.

- [19] **Eli Cortez, Altigran Silva, Marcos Gonçalves, Filipe Mesquita and Edleno Moura**, 2009, A flexible approach for extracting metadata from bibliographic citations, In Journal of the American Society for Information Science and Technology - JASIST.

#### Workshop Papers and Technical Reports (5)

- [20] **Eli Cortez and Altigran Silva**, 2013, Unsupervised Information Extraction by Text Segmentation, In Brazilian Computer Society Thesis and Dissertation Contest - SBC CTD.
- [21] **Eli Cortez and Altigran Silva**, 2010, Unsupervised strategies for information extraction by text segmentation, In ACM SIGMOD PhD Workshop on Innovative Database Research - IDAR.
- [22] **Eli Cortez and Altigran Silva**, 2010, A flexible approach for extracting metadata from bibliographic citations, In Brazilian Computer Society Thesis and Dissertation Contest - SBC CTD.
- [23] **Filipe Mesquita, Denilson Barbosa, Eli Cortez and Altigran Silva**, 2010, FleDEX: flexible data exchange, In International workshop on Web Information and Data Management - WIDM.
- [24] **Filipe Mesquita, Denilson Barbosa, Eli Cortez and Altigran Silva**, 2010, A Lightweight Framework for Exchanging Web Data, In Technical Report - University of Calgary.

#### Tutorials (2)

- [25] **Altigran Silva and Eli Cortez**, 2012, Methods and Techniques for Information Extraction by Text Segmentation, In International Workshop on Foundations of Data Management - AMW.
- [26] **Altigran Silva and Eli Cortez**, 2011, Methods and Techniques for Information Extraction by Text Segmentation, In Brazilian Symposium on Databases - SBBD.

---

#### Talks

- 2014 **Exploring Structured Data to Perform Information Extraction: Methods and Real Applications**, University of Alberta.
- 2014 **Empowering Text Analytics with Structured Data**, Mitre Media.
- Summer 2012 **Managing Cold State for Streaming Big Data Applications**, Microsoft Research.
- Summer 2012 **Unsupervised Information Extraction by Text Segmentation**, Microsoft Research.
- 2012 **Large Scale Information Extraction**, Polytechnic Institute of New York University - NYU.
- 2011 **Information Extraction over Textual Sources**, UOL TechDay.
- 2009 **Flexible Information Extraction from Bibliographic References**, University of Alberta.
- 2007 **FLUX-CiM: flexible unsupervised extraction of citation metadata**, University of Calgary.

---

## Research Community Services

**Conference Program Committee and Reviewer:** VLDB 2017, ICDE Demos 2016, MobiWis 2015, SBBD 2015, SIGMOD Demos 2015, SBBD 2014, WWW Demo 2014, SIGMOD Reproducibility 2013, SBBD 2013 Short Papers, WWW Demos 2013, ACM SIGIR Short Papers 2013, SIGMOD Reproducibility 2012, ACM SIGIR Posters 2012

**External Reviewer:** ACM DocEng 2013, ACM SIGIR 2012, ACM/IEEE JCDL Conference 2012, EDBT Conference 2011, ACM/IEEE JCDL Conference 2010

---

## Idioms

**English** - fluent

**Portuguese** - mother language

---

## Participation in Events

- 2015 **International Conference on Very Large Data Bases - VLDB**, Kohala Coast-USA.
- 2013 **ACM SIGMOD Conference**, New York City-USA.
- 2013 **Congress of the Brazilian Computer Society - CSBC**, Maceió-Brazil.
- 2013 **EMC<sup>2</sup> Summer School on Big Data**, Rio de Janeiro-Brazil.
- 2012 **ACM SIGMOD Conference**, Scottsdale-USA.
- 2011 **ACM SIGMOD Conference**, Athens-Greece.
- 2011 **International Conference on Very Large Data Bases - VLDB**, Seattle-USA.
- 2011 **Brazilian Symposium on Databases - SBBD**, Florianópolis-Brazil.
- 2010 **ACM SIGMOD Conference**, Indianapolis-USA.
- 2010 **Congress of the Brazilian Computer Society - CSBC**, Belo Horizonte-Brazil.
- 2009 **Brazilian Symposium on Databases - SBBD**, Fortaleza-Brazil.
- 2008 **ACM SIGMOD Conference**, Vancouver-Canada.
- 2008 **Brazilian Symposium on Databases - SBBD**, Campinas-Brazil.
- 2007 **ACM/IEEE Joint Conference on Digital Libraries - JCDL**, Vancouver-Canada.