

CURRICULUM VITAE

Alisa Zhila, PhD

July 2, 2017



1. Personal Data

Name: Alisa Zhila
Email: alisa.zhila@gmail.com, alisa_zh@mail.ru
URL: <http://nlp.cic.ipn.mx/~alisa/>
LinkedIn: [linkedin.com/in/alisazhila](https://www.linkedin.com/in/alisazhila)

2. Current Position

2016 – present, Software Engineer at IBM Watson, Visual Recognition Service Team

3. Research Interests

Broadly defined my area of research is **computational linguistics** and **natural language processing**. My current area of focus is **natural object naming** in images as well as image tag grouping based on ontology for multiple object tagging. Previously, my main research interest laid in **open information extraction** from text and its applications to **text quality evaluation**, in particular, text informativeness. Departing from these areas, I also worked on the problems of **human opinion collection** for evaluation of subjective aspects of text. Another direction of my research is the mapping of the output returned by open information extraction systems onto **RDF data representation model**. In parallel to my main focus, I have also worked on **semantic similarity measure** and **system log classification**. My other interests lie in **lexicography** and **semiotics**.

4. Education

4.1 Academic Education

PhD, Computer Science, Jan 2011 – Dec 2014, with honors

Natural Language Processing Laboratory (nlp.cic.ipn.mx),

Center for Computing Research (Centro de Investigación en Computación, CIC),

Instituto Politécnico Nacional (IPN), Mexico

PhD Thesis: Open Information Extraction based on Constraints over Part-of-Speech Sequences

Developed Open IE system ExtrHech for Spanish (https://bitbucket.org/alisa_ipn/extrhech)

Advisor: Dr. Prof. Alexander Gelbukh, www.cic.ipn.mx/gelbukh

M.Sc., Applied Physics and Mathematics, 2006 – 2008, with honors

Department of Conceptual Analysis and Design

Moscow Institute of Physics and Technology (MIPT, “PhysTech”), Russia.

Master Thesis: Basic Semiotic Concepts Explication in Species of Structures

Advisor: Dr. Yulia Garaeva

B.Sc., Applied Physics and Mathematics, 2002 – 2006, with honors

Department of Molecular and Biological Physics

Moscow Institute of Physics and Technology (MIPT, “PhysTech”), Russia.

Bachelor Thesis: Detection of Visual Evoked Potential Dipole Sources Trajectory over Human Brain Cortex

Thesis work fulfilled in the Institute of Higher Nervous Activity (IHNA) of Russian Academy of Science, www.ihna.ru/en/

Advisor: Dr. Elena Mikhailova

4.2 Professional Education

Associate Degree (Diploma), Translator in Professional Communications, 2005-2006

Moscow Institute of Physics and Technology (MIPT, “PhysTech”), Russia

5. Research Projects

5.1 IBM Projects

- A. Comparative Analysis of Text Search Systems Solr and Elasticsearch for Application for Documents with Complex Hierarchical Structure. *May-Sep 2016.*
 - Analysis of existing capabilities of the latest versions at the moment of analysis
 - Proposal of the improvements to Solr system for fuller flexibility
 - Resulted in a presentation at Lucene/Solr Revolution 2016 conference
- B. Enhancing Multiple Object Tagging in Images with Tag Clustering using Semantic Hierarchy, *Oct 2016 – present*
 - Design and development of the state-of-the-art image tagging enhancement through employment of a customized semantic hierarchy
 - In deploy to production
- C. Leveraging Semantic Correlation for Image Tag Augmentation. *June 2017 – present*
 - Comparison of effect of the semantic correlation computed on a large text corpus vs. semantic relations available from rich multirelational ontologies on image tagging augmentation

5.2 Institutional Projects

Research project 20113295: Detection of Textual Entailment and Lexical Relations in Natural Language Texts.

Director Dr. Alexander Gelbukh.

Text and Natural Language Processing Laboratory, Center for Computing Research, Instituto Politecnico Nacional

- Overview of related literature on the advanced methods for word sense disambiguation
- Development of methods for semantic representation of lexical relations between words in text
- Support in development of various tasks of the project; tests and documentation of the results
- Automatic detection of textual entailment in text and its use for the question answering task

Research project 20121823: Automatic disambiguation and clustering of word senses for applications in the computational processing of natural language

Director Dr. Alexander Gelbukh.

Text and Natural Language Processing Laboratory, Center for Computing Research, Instituto Politecnico Nacional

- Development and programming of the methods for clustering of word senses for automatic translation
- Support in development of other modules
- Integration of the developed methods with the task of automatic detection of text entailment
- Application of the developed methods to classification of Wikipedia articles

Research project 20131702: Analysis of compound expressions; affectivity and personality in text with machine learning methods.

Director Dr. Alexander Gelbukh.

Text and Natural Language Processing Laboratory, Center for Computing Research, Instituto Politecnico Nacional

- Related literature overview
- Development of lexical resources
- Machine learning: refinement of the methods used in the previous research
- Analysis of opinions and personality: development of methods for automatic processing of personality data, affectivity, and opinion in text

Research project 20144534: Fact extraction and disambiguation in opinion and polarity detection in text

Director Dr. Alexander Gelbukh.

Text and Natural Language Processing Laboratory, Center for Computing Research, Instituto Politecnico Nacional

- Automatic extraction of facts from text in Spanish language
- Concept analysis and extraction

- Development of patterns based on named entities for relevant information detection in texts
- Implementation of the methods based on these patterns for automatic extraction of relevant information from news articles

5.3 Internship Projects

IBM, Littleton, MA, USA, Feb – Mar 2015

Division: Watson

Project: Open Relation Extraction based on Predicate-Argument Structures (Java)

- Designed and implemented a Relation Extraction system using predicate-argument structures
- Worked in a shared project environment
- Connected to the input from corpus storage
- Implemented output to JSON

Yahoo! Inc., Sunnyvale, CA, USA, July – Sep 2014

Project: Analysis of Entropy Calculation in Alert Monitoring Systems and Its Improvement (Python)

- The task was to detect what system logs are important
- Comparison of entropy to TF-IDF
- Feature engineering through regex pattern matching
- Compared simple threshold-based classifier (heuristics, no training) to LogReg classifier (in R)
- Data analysis in R

Oracle, Oracle MDC office, Guadalajara, Jalisco, Mexico, March – June 2014

Product Group: Oracle Spatial and Graph

Project: Relation Extraction System for Semantic Indexing Functionality Demo (Java)

- Relation Extraction from text based on syntactic constraints (basic rules + heuristics)
- Converting extractions into RDF/XML format
- Connection to DB
- Oracle Semantic Technologies, SPARQL queries

Microsoft Research, Redmond, WA, USA, June - September 2012

Project: Measuring Degrees of Relational Similarity between Word Pairs (Python)

- Research on the subject
- Feature engineering: relation-specific information, lexical patterns, vector space
- Development of a System for feature extraction and processing (in Python)
- Machine learning (LogReg) application using an existing tool
- Publication in NAACL'2013

5.4 International Projects

WIQ-EI project “Web information quality evaluation initiative”

Funded by European Commission within the FP7 People Programme (project no. 269180)
KNOW-IT Center, Graz, Austria, April 2013.

- Automatic methods for evaluation of text quality
- Development and implementation of the method for automatic informativeness evaluation of arbitrary internet text based on factual density using open information extraction tool
- Design and development of methods for collecting of human annotation of informativeness
- Conducting the experiment

6. Employment Record

IBM, Littleton, MA, 2016 – present

Software Engineer, Jan 2016 – Present

Watson Question Answering Project, Jan 2016 – May 2016

Watson Conversation Project, May 2016 – Oct 2016

Watson Visual Recognition Project, Oct 2016 – present

ABBYY Software Ltd, Moscow, Russia, 2007 – 2010

Acting Head of IT Terminology Group, 2009 – 2010

Compreno Machine Translation project: senior developer of syntactic-semantic rules for translation of words in context and collocations; managing IT terminology group working process

Linguist, IT Terminology Group, 2007 – 2009

Compreno Machine Translation project: extracting word senses from corpus and elaborating their formal descriptions including collocations, syntax, syntactic-semantic rules; other lexicographic tasks

7. Professional Service

7.1 National and International Conference Organizing Committees

- NAACL SRW 2015: Student Research Workshop at 2015 edition of the North American Chapter of the Association for Computational Linguistics conference
- CORE 2012: International Congress on Computer Science (www.core.cic.ipn.mx)

7.2 Reviewing for Conferences

- IJCNLP 2017 (invited, review period: July 14 - August 4, 2017): 8th International Joint Conference on Natural Language Processing
- ACL 2017: 55th annual meeting of the Association for Computational Linguistics
- COLING 2016: 26th International Conference on Computational Linguistics
- MICAI 2015, 2016: 14th and 15th Mexican International Conference on Artificial Intelligence
- EMNLP 2013: Conference on Empirical Methods in Natural Language Processing
- CICLing 2013: 14th International Conference on Computational Linguistics and Natural Language Processing

7.3 Reviewing for Journals

- Polibits, ISSN: 1870-9044 (Print) 2395-8618 (Online)
- Computación y Systemas (Eng: Computation and Systems), ISSN 1405-5546 (Print) 2007-9737 (Online)
- Research in Computing Science, ISSN 1870-4069
- Procesamiento del Lenguaje Natural (SEPLN, Eng: Natural Language Processing, Journal), ISSN: 1135-5948 (Print) 1989-7553 (Online)
- Cognitive Computation, ISSN: 1866-9956 (Print) 1866-9964 (Online)
- International Journal of Computational Linguistics and Applications (IJCLA), ISSN 0976-0962
- Information Processing Letters, ISSN: 0020-0190

7.4 Reviewing for Publishers

Editorial Division of the Mexican Society of Artificial Intelligence in 2014, 2015

8. Professional Memberships

Association for Computational Linguistics (ACL), 2013 – present

Mexican Association for Natural Language Processing (AMPLN), 2011 – present

Mexican Society for Artificial Intelligence (SMIA), 2011 – 2014

Semiotic Society of America, 2008 – 2009

9. Distinguished Memberships

Distinguished Membership in Mexican Society for Artificial Intelligence (SMIA), 2014 - present

10. Awards

Microsoft Research 2012 Latin America Fellow (only 2 fellows are chosen from 23 countries per year)

11. Invited Talks

1. *(upcoming, invitation accepted)* Talk at the University Conference in the Universidad de Guanajuato (University of Guanajuato). Sep 28, 2017.
2. Talk at Center for Computing Research, Instituto Politécnico Nacional. Sep 5, 2016.
3. **MICAI 2014**, 13th Mexican International Conference on Artificial Intelligence. Tuxtla Gutierrez, Chiapas, Mexico. Nov 17 – 21, 2014.
Mexican Students at Microsoft Research: Experience and Lessons Learnt. (In collaboration with 2 other presenters).
4. **9° Taller de Tecnologías del Lenguaje Humano** (9th Workshop on Human Language Technologies). Tonantzintla, Puebla, Mexico. Oct 18 – 19, 2012.
My Research Internship in Microsoft Research (Presented in Spanish: *Mi estancia de investigación en Microsoft Research*)

5. **COMIA 2012** (www.comia.org.mx), 4° Congreso Mexicano de Inteligencia Artificial (4th Mexican Conference on Artificial Intelligence). Xicotepec de Juarez, Puebla, Mexico. Jun 12-15, 2012. *How to Win an Important Award.* (Presented in Spanish: *Como ganar un premio importante*)
6. **Foro PIFI 2012**, 7° Foro del Programa Institucional de Formación de Investigadores (7th Forum of the Institutional Program of Researcher Education). Mexico City, DF, Mexico. May 17, 2012. *How to Win an Important Scientific Award?* (Presented in Spanish: *¿Cómo ganar un premio científico importante?*)

12. Conference Presentations

1. **Lucene/Solr Revolution 2016**, Boston, MA. October 11-14, 2016. *Working with Deeply Nested Documents in Apache Solr* (jointly with Anshum Gupta).
2. **ACL SRW 2014**, Student Research Workshop at the 52nd Annual Meeting of the Association for Computational Linguistics. Baltimore, MD, USA. Jun 22-27, 2014. *Open Information Extraction for Spanish Language based on Syntactic Constraints.*
3. **6° CoLiCo**, VI Coloquio de Lingüística Computacional en la UNAM (VI Colloquium on Computational Linguistics in UNAM). Mexico City, DF, Mexico. Aug 19-20, 2013. *Open Information Extraction for Spanish.* (Presented in Spanish: *La extracción abierta de información para el español*).
4. **NAACL 2013**, the 2013 Conference of the North American Chapter of the Association for Computational Linguistics. Atlanta, GA, USA. Jun 10 – 12, 2013. *Combining Heterogeneous Models for Measuring Relational Similarity.*
5. **Dialogue 2013** (www.dialog-21.ru/en/), A Major Conference on Computational Linguistics in Russia. Bekasovo, Moscow Region, Russia. May 29 – Jun 3, 2013. *Comparison of Open Information Extraction for Spanish and English.*
6. **9° Taller de Tecnologías del Lenguaje Humano** (9th Workshop on Human Language Technologies). Tonantzintla, Puebla, Mexico. Oct 18-19, 2012. *Measuring Degrees of Relational Similarity.*
7. **Dialogue 2012**, A Major Conference on Computational Linguistics in Russia. Bekasovo, Moscow Region, Russia. May 30 – Jun 3, 2012. *Exploring Context Clustering for Term Translation.*
8. **COMIA 2012** (www.comia.org.mx), 4° Congreso Mexicano de Inteligencia Artificial (4th Mexican Conference on Artificial Intelligence). Xicotepec de Juarez, Puebla, Mexico. Jun 12 – 15, 2012. *Exploration of a Multilingual Application of Text Clustering* (Presented in Spanish: *Exploración de una aplicación multilingüe del agrupamiento de textos.*)
9. **MICAI 2011**, Doctoral Consortium at 10th Mexican International Conference on Artificial Intelligence. Puebla, Puebla, Mexico. Nov 26 – Dec 4, 2011. *Improving Machine Translation with Automatic Cross-Lingual Word Sense Discrimination.*
10. **COMIA 2011**, 3° Congreso Mexicano de Inteligencia Artificial (3th Mexican Conference on Artificial Intelligence). Atizapan de Zaragoza, Estado de Mexico, Mexico. Oct 18 – 21, 2011. *Methods for Evaluation of Word Sense Disambiguation.*

11. **SSA 2008**, 33rd Annual Meeting of the Semiotic Society of America (SSA). Houston, TX, USA. Oct 16-19, 2008.

Basic Semiotic Concepts Explication in Species of Structures for Their Further Formal Systematization with Advantages of Extensional Approach.

12. **50th Scientific Conference of the Moscow Institute of Physics and Technology** (mipt.ru/dasr/news/n_38hrb6). Moscow, Russia. Nov 23 – 26, 2007. (Student conference)
Analysis and Synthesis of Models used by Melnikov in System Classification of Languages
(Presented in Russian).

13. Tutorial

COMIA 2012 (www.comia.org.mx), 4^o Congreso Mexicano de Inteligencia Artificial (4th Mexican Conference on Artificial Intelligence). Xicotepec de Juarez, Puebla, Mexico. Jun 12-15, 2012.
Sentiment Analysis (Presented in Spanish: *Análisis de Sentimientos*), 3 hour workshop for undergraduate students

14. Posters

1. **KESW 2015**, International Conference on Knowledge Engineering and Semantic Web. Moscow, Russia, Sep 30 – Oct 2, 2015.
Alisa Zhila, Elena Yagunova, and Olga Makarova. Bringing The Output of Open Information Extraction to The RDF/XML Format: A Case Study. (Presented by Olga Makarova).
2. **Tapia 2014**, Richard Tapia Celebration of Diversity in Computing, 2014. Seattle, WA, USA. Feb 5 – 8, 2014.
Alisa Zhila, Alexander Gelbukh. Informativeness and Objectivity of Texts on the Web.
3. **MICAI 2013**, 12th Mexican International Conference on Artificial Intelligence. Mexico City, DF, Mexico. Nov 24 – 30, 2013.
Alisa Zhila, Christofer Horn, Alexander Gelbukh. Automatic Assessment of Text Quality on the Web via Fact Extraction.
4. **GHC 2013**, Grace Hopper Celebration of Women in Computing. Minneapolis, MN, USA. Oct 2 – 5, 2013.
Alisa Zhila, Christofer Horn, Alexander Gelbukh. Automatic Assessment of Web Text Quality via Fact Extraction.
5. **Tapia 2013**, Richard Tapia Celebration of Diversity in Computing, 2013. Washington, DC, USA. Feb 7 – 10, 2013.
Alisa Zhila, Christofer Horn, Alexander Gelbukh. Open Information Extraction for Spanish and Its Application to Measuring Informativeness of Web Documents.
6. **8^o Taller de Tecnologías del Lenguaje Humano** (8th Workshop on Human Language Technologies). Puebla, Puebla, Mexico. Nov 28 – 29, 2011.
Alisa Zhila, Alexander Gelbukh. Unsupervised Cross-Lingual Sense Tagging for Statistical Machine Translation.

15. Publications

1. **Alisa Zhila**, Alexander Gelbukh. Open Information Extraction from Real Internet Texts in Spanish Using Constraints over Part-Of-Speech Sequences: Problems of the Method, Their Causes, and Ways for Improvement. *Revista Signos. Estudios de Lingüística* (Journal Signos. Linguistics Studies.), 90, vol. 49, pp. 119-142, **2016**.
2. **Alisa Zhila**, Alexander Gelbukh, Helena Gomez-Adorno. Fast Named Entity Driven Open Information Extraction with Shallow Semantic Interpretation. *Information Sciences*. **Submitted Sept. 22, 2015**.
3. **Alisa Zhila**, Alexander Gelbukh. Open Information Extraction for Spanish Language based on Syntactic Constraints. In *Proceedings ACL SRW*, pp. 78-85, **2014**.
4. **Alisa Zhila**, Scott Yih, Chris Meek, Geoffrey Zweig and Tomas Mikolov. Combining Heterogeneous Models for Measuring Relational Similarity. In *Proceedings HLT-NAACL'2013*, pp. 1000-1009, **2013**.
5. Christopher Horn, **Alisa Zhila**, Alexander Gelbukh, Roman Kern, Elisabeth Lex. Using Factual Density to Measure Informativeness of Web Documents. In *Proceedings NoDaLiDa'13*, pp. 227-238, **2013**.
6. **Alisa Zhila**, Alexander Gelbukh. Comparison of Open Information Extraction for Spanish and English. *Computational Linguistics and Intellectual Technologies. Proceedings Dialogue'2013*, 12, vol. 1, pp. 794-802, **2013**.
7. **Alisa Zhila**, Alexander Gelbukh. Exploring context clustering for term translation. In *Computational Linguistics and Intellectual Technologies, Proceedings Dialogue'2012*, 11, Vol. 1, pp. 716-725, **2012**.
8. **Alisa Zhila**, Alexander Gelbukh. Analysis of a cross-lingual application of context clustering (in Spanish: *Análisis de una aplicación multilingüe del agrupamiento de textos*). In *Research in Computing Science*, vol. 55, Special Issue: *Avances en Inteligencia Artificial* (Advances in Artificial Intelligence), pp. 45-57, Mexican Society for Artificial Intelligence (SMIA), **2012**.
9. **Alisa Zhila, Alexander Gelbukh**. Classification of methods for improvement of WSD and the corresponding evaluation methods (in Spanish: *Clasificación de los métodos par ala mejora de WSD y de los métodos de evaluación correspondientes*). Book chapter in: M. Gonzalez Mendoza and O. Herrera Alcántara (Eds.), *Avances recientes en sistemas inteligentes (Recent Advances in Intelligent Systems)*, pp. 232-241, Mexican Society for Artificial Intelligence (SMIA), **2011**.
10. **Alisa Zhila**. Formalization of basic semiotic notions in set theoretic terms. *Polibits* 42, pp. 83-97, **2010**.
11. **Alisa Zhila**. Basic semiotic concepts explication in species of structures for their further formal systematization with advantages of extensional approach. In: *The Proc. of the 33rd Annual Meeting of the Semiotic Society of America (SSA)*, pp. 751-771, **2009**.
12. **Alisa Zhila**, Victor Kapoustyan. Review of semantic models and investigation of the possibilities of their applying to C.S. Pierce's sign categories interpretation (in Russian). In: *Proc. of 10th conference "Grigorievskie chteniya"* (Readings in honor of Prof. Grigoriev): *Symbols, codes, signs*. Moscow, Russia, pp.141-147, **2008**.

13. **Alisa Zhila**, Yulia Garaeva. Analysis and synthesis of models used by Melnikov in System Classification of Languages (in Russian), In *Abstract Collection for 50th Scientific Conference of the Moscow Institute of Physics and Technology, vol. 9: Innovations and High Technologies*, pp. 4–6, **2007**.
14. **Elena Mikhailova**, Alisa Zhila, Anna Slavutskaya, Mikhail Kulikov, Igor Shevelev. Trajectories of Visual Evoked Potentials Dipole Sources Shifting over Human Brain Cortex (in Russian), In *Journal of Higher Nervous Activity* 56(6), pp. 555–564, **2007**.

16. Patent

Open patent application:

US 20140249799 A1. Wen-tau Yih, Geoffrey Zweig, Christopher Meek, **Alisa Zhila**, Tomas Mikolov. Relational similarity measurement. Publication Date: Sept. 4, **2014**

17. Languages

English (fluent), Spanish (fluent), German (basic reading), Russian (native)