OMAR F. ZAIDAN

Mailing Address
P.O. Box: 930175
Zip Code 11193
Amman
JORDAN

v-omfaye@microsoft.com
http://cs.jhu.edu/~ozaidan
++962-79-707-9453

Street Address
14 Al-Hareeri Street
Apt. #8
Amman
JORDAN

EXECUTIVE SUMMARY

A track record of research in computational linguistics, machine learning, and annotator modeling. Specific expertise in machine translation and crowdsourcing, coupled with command of software development practices, a keen eye for detail, and strong written and spoken communication skills.

EDUCATION

Johns Hopkins University, Baltimore, MD, USA *August 2004 – December 2011*

Ph.D., Computer Science (M.S.Eng. conferred May 2007; **GPA:** 4.00/4.00)

Thesis: *Crowdsourcing Annotation for Machine Learning in Natural Language Processing Tasks*
Advisor: Chris Callison-Burch

St. Lawrence University, Canton, NY, USA *August 2000 – May 2004*

B.Sc., double major in Computer Science (with Honors) and Mathematics (with Honors) and a minor in Chemistry.

GPA: 3.96/4.00 (*summa cum laude*, junior-year election to Phi Beta Kappa, 4.00 majors GPA)

Computer Science Thesis: *Creating Computer Othello Players Using a Genetic Algorithm*
Mathematics Thesis: *Coloring Random Graphs: A Statistical Analysis*

PROFESSIONAL EXPERIENCE

Microsoft Corporation, Redmond, WA, USA *February 2012 – present*
Software Development Engineer II, Machine Translation Group

Design and build automated metrics to evaluate the performance of *Bing Translator*, the world's second largest Machine Translation service, a large-scale product with releases every 6–8 weeks.

- Designed a filtering pipeline to distinguish spmmy queries from genuine ones, achieving 95% accuracy for previously unknown words.
- Implemented data mining of web resources for high-value named entities (country names, heads of states, etc), and automated their testing across 40 languages.

Johns Hopkins University, Baltimore, MD, USA *December 2006 – December 2011*
Research Assistant
Department of Computer Science & Center for Language and Speech Processing, funded by:

BBN Technologies – *Translation of Informal Texts via Mechanical Turk*

- Created and managed a large annotation effort for dialectal Arabic identification.
- Created parallel datasets for *dialectal* Arabic-to-English to aid training of MT systems.

The European Commission – *EuroMatrix* and *EuroMatrixPlus*

- Developed the RYPT metric and a method for human-in-the-loop tuning of MT systems.
- Ran the evaluation campaigns of the 2010 and 2011 Workshops on Machine Translation, involving 150+ systems over 8 different language pairs per year.

IBM – *DARPA's Global Autonomous Language Exploitation (GALE)*

- Created the *Arabic Online Commentary* dataset, a 52M-word corpus of informal Arabic.
- Designed methods to filter crowdsourced translations, yielding near-professional quality.
- Member of the development team for Joshua (open-source MT toolkit in Java).

JHU WSE-APL Partnership Fund – *Learning with Less*

- Designed a new paradigm for statistical learning, using *annotator rationales*, and applied it to sentiment analysis and dialect identification, achieving significant improvements.

TEACHING EXPERIENCE

Johns Hopkins University, Baltimore, MD, USA *Summer 2007*
Instructor
Department of Computer Science

Taught department's summer offering of *Introduction to Java*, using self-prepared course materials.

Johns Hopkins University, Baltimore, MD, USA *August 2004 – December 2006*
Teaching Assistant
Department of Computer Science

Head TA for various courses at the graduate level, including *Natural Language Processing*, *Artificial Intelligence*, *Database Systems*, and *Modern Complexity Theory*.

HONORS AND AWARDS

Finalist, Best Teaching Assistant Award, JHU Whiting School of Engineering *Spring 2007*
Pi Mu Epsilon Award for an Outstanding Senior *May 2004*
Phi Beta Kappa (National academic honorary) *Inducted Fall 2003*
Pi Mu Epsilon (National mathematics honorary) *Inducted Fall 2002*
Dean's List, St. Lawrence University *Fall 2000 – Spring 2004* (All semesters)

SELECTED REFEREED PUBLICATIONS

+S: released software

+D: released dataset

O. Zaidan and C. Callison-Burch. 2013. Arabic Dialect Identification. ***Computational Linguistics***.

R. Zbib, E. Malchiodi, J. Devlin, D. Stallard, S. Matsoukas, R. Schwartz, J. Makhoul, **O. Zaidan** and C. Callison-Burch. 2012. Machine Translation of Arabic Dialects. ***NAACL***, pp. 49–59.

O. Zaidan. 2011. MAISE: A Flexible, Configurable, Extensible Open Source Package for Mass AI System Evaluation. ***EMNLP*** *Workshop on Statistical Machine Translation*, pp. 130–134. +S

C. Callison-Burch, P. Koehn, C. Monz, and **O. Zaidan**. 2011. Findings of the 2011 Workshop on Statistical Machine Translation. ***EMNLP*** *Workshop on Statistical Machine Translation*, pp. 22–64.

O. Bojar, M. Ercegovčević, M. Popel, and **O. Zaidan**. 2011. A Grain of Salt for the WMT Manual Evaluation. ***EMNLP*** *Workshop on Statistical Machine Translation*, pp. 1–11.

O. Zaidan and C. Callison-Burch. 2011. Crowdsourcing Translation: Professional Quality from Non-Professionals. ***ACL***, pp. 1220–1229. +D

O. Zaidan and C. Callison-Burch. 2011. The Arabic Online Commentary Dataset: an Annotated Dataset of Informal Arabic with High Dialectal Content. ***ACL*** *Short Paper Track*, pp. 37–41. +D

O. Zaidan and C. Callison-Burch. 2010. Predicting Human-Targeted Translation Edit Rate via Untrained Human Annotators. ***NAACL-HLT*** *Short Paper Track*, pp. 369–372.

O. Zaidan and J. Ganitkevitch. 2010. An Enriched MT Grammar for Under $100. ***NAACL-HLT*** *Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*, pp. 93–98.

Z. Li, C. Callison-Burch, C. Dyer, J. Ganitkevitch, A. Irvine, S. Khudanpur, L. Schwartz, W.N.G. Thornton, Z. Wang, J. Weese, and **O. Zaidan**. 2010. Joshua 2.0: a Toolkit for Parsing-based Machine Translation with Syntax, Semirings, Discriminative Training and Other Goodies. ***ACL*** *Workshop on Statistical Machine Translation and MetricsMATR*, pp. 133–137. +S

O. Zaidan and C. Callison-Burch. 2009. Feasibility of Human-in-the-loop Minimum Error Rate Training. ***EMNLP***, pp. 52–61.

O. Zaidan. 2009. Z-MERT: A Fully Configurable Open Source Tool for Minimum Error Rate Training of Machine Translation Systems. *The* ***Prague Bulletin*** *of Mathematical Linguistics*, No. 91, pp. 79–88. +S

O. Zaidan and J. Eisner. 2008. Modeling Annotators: A Generative Approach to Learning from Annotator Rationales. ***EMNLP***, pp. 31–40.

O. Zaidan, J. Eisner, and C. Piatko. 2007. Using "Annotator Rationales" to Improve Machine Learning for Text Categorization. ***NAACL-HLT***, pp. 260–267. +D

RELEASED SOFTWARE

Z-MERT (licenced under LGPL; first release Jan. 2009)

A demonstrably time- and space-efficient tool for tuning MT systems, used by researchers at many institutions, including Carnegie Mellon, RWTH Aachen, and University of Edinburgh.

MAISE (licenced under LGPL; first release Nov. 2010)

An extensible package for "mass" evaluation of AI systems, greatly streamlining the process of crowdsourced system evaluation, using the workforce on Amazon's Mechanical Turk.

RELEASED DATASETS

The Dialectal Arabic Dataset

A set of 108K Arabic sentences each annotated for which dialect it contains by multiple annotators.

The Arabic Online Commentary Dataset

A 52M-word corpus of informal Arabic, harvested from reader commentary on online articles.

The Movie Review Sentiment Polarity Dataset, Enriched with Annotator Rationales

A version of Pang & Lee's dataset of 2,000 movie reviews, each enriched with annotator rationales.

SKILLS

Programming Languages: Java, C#, C++, SQL, MATLAB©, HTML, JavaScript, Prolog, and extensive experience with the Java API for Mechanical Turk

Software: LaTeX, SVN, R, MySQL, SRILM, Photoshop, Dreamweaver, Fireworks

Languages: English (fluent) and Arabic (native)

SERVICE

Organizing Committee:

- 6th Workshop on Statistical Machine Translation (at EMNLP 2011)
- 5th Workshop on Statistical Machine Translation and MetricsMATR (at ACL 2010)
- 4th North-East Student Colloquium on Artificial Intelligence (NESCAI 2010)

Conference Reviewing: AAAI, COLING, Journal of MT, NESCAI, WMT

Departmental Service: Student-Faculty Liaison, Department of Computer Science (2008–2010)

GRADUATE COURSES

Natural Language Processing • Database Systems • Machine Learning • Artificial Intelligence • Computer Vision • Information Extraction • Information Theory • Modern Complexity Theory •

REFERENCES

Vishal Chowdhary	Chris Callison-Burch	Jason Eisner
Senior SDET Lead	Associate Research Professor	Associate Professor
Machine Translation Group Microsoft Corporation	Dept. of Computer Science Johns Hopkins University	Dept. of Computer Science Johns Hopkins University
++1–425–722–3917	++1–410–516–3427	++1–410–516–8438
`vishalc@microsoft.com`	`ccb@cs.jhu.edu`	`jason@cs.jhu.edu`