

RYAN GEORGI

PERSONAL INFORMATION

email rgeorgi@gmail.com
website <http://ryan.georgi.cc>
phone +1 (510) 508 2131

PROFILE

Researcher in computational linguistics with focus on projection-based transfer and noisy, resource-poor domains. Experience with supervised and semi-supervised machine learning algorithms and implementation in parallel environment on university cluster.

EDUCATION

- Ph.D* The University of Washington, Seattle 2009–2016
Ph.D, Computational Linguistics · Aug. 2016
Thesis: *From Aari to Zulu: Massively Multilingual Creation of Language Tools using Interlinear Glossed Text*
Advisors: FEI XIA & WILLIAM D. LEWIS
- MA* The University of Washington, Seattle 2006–2009
MA, Computational Linguistics · Jun. 2009
Thesis: *Grammar Induction with Prototypes Derived from Interlinear Text*
Advisors: FEI XIA & WILLIAM D. LEWIS
- BA* The University of California, Berkeley 2001–2005
BA, Major in Linguistics · Minor in Computer Science · May 2005

WORK EXPERIENCE

- University of Washington* Research Assistant 2007–2016
Member of the *information engineering and synthesis for Resource-Poor Languages (RiPLes)* research group. Developed tools to automatically enrich Interlinear Glossed Text (IGT) for the Online Database of INterlinear text (ODIN) and bootstrap NLP tools for resource-poor languages. Created python libraries and web interfaces for machine learning and corpus enrichment tasks.
Supervisor: FEI XIA
- Microsoft Research* Linguistics Test Engineer Feb–Sep 2008
Member of the Machine Translation Incubation Team, developed metrics and experimented with different methods of corpora selection with the aim of improving MT system performance on different target domains. Adapted new tools to existing C# codebase.
Supervisor: WILLIAM D. LEWIS
- Voxify, Inc* QA Engineer 2004–2006
Developed QA procedures for testing Voice user interfaces (VUIs) in customer-facing telephony systems, as well as backend database creation and testing.
Supervisor: BOB BIRSS

SOFTWARE PACKAGES

INTENT

INTENT

INTENT, or the INterlinear Text ENrichment Toolkit, implements machine learning techniques for automatically enriching IGT data and extracting NLP tools including POS taggers, dependency parsers, and translation lexicons. This package is the implementation of my doctoral dissertation and was recently used to enrich 158K IGT instances for 1.5K languages for the Online Database of Interlinear Text (ODIN) v2.1 — <http://xigt.org/odin/>.

Link: <https://github.com/rgeorgi/intent>

IGT-DETECT

IGT-DETECT for ODIN

Version 2.1 of ODIN contains approximately 158,000 IGT instances, covering over 1,500 languages, and culled from approximately 3,000 PDF-to-text extracted documents. In order to expand coverage for ODIN 3.0, this package was created to perform feature extraction and automatic detection of IGT instances embedded among documents obtained by ODIN's web scraper, with over 1.3 million documents currently targeted. Current tests show 96% accuracy detecting IGT instances in linguistic documents.

Link: <https://github.com/xigt/igtdetect>

IGT Editor

IGT EDITOR

AJAX-enabled browser-based IGT editing toolkit that enables annotators to clean IGT instances, rate cleanliness, and visualize the INTENT-enriched output. Example application where underlying INTENT software is used to facilitate annotation by a non-expert user.

Link: <http://editor.xigt.org/user/demo>

PUBLICATIONS

Journal Papers

Fei Xia, William D. Lewis, Michael W. Goodman, Glenn Slayden, Ryan Georgi, Joshua Crowgey, & Emily Bender. Enriching a Massively Multilingual Database of Interlinear Glossed Text. *Language Resources and Evaluation*, Jan. 2016.

<http://link.springer.com/article/10.1007%2Fs10579-015-9325-4>

Ryan Georgi, Fei Xia, William D. Lewis. Capturing Divergence in Dependency Trees to Improve Syntactic Projection. *Language Resources and Evaluation*, Oct. 2014.

<http://link.springer.com/article/10.1007%2Fs10579-014-9273-4>

Conference Papers

Ryan Georgi, Michael W. Goodman, and Fei Xia. A Web-framework for ODIN Annotation. *ACL 2016*, Aug. 2016.

<https://www.aclweb.org/anthology/P/P16/P16-4006.pdf>

Ryan Georgi, Fei Xia & William D. Lewis. Enriching Interlinear Text using Automatically Constructed Annotators. *LaTeCH-2015*, Jul. 2015.

<http://www.aclweb.org/anthology/W15-3709>

Fei Xia, Michael W. Goodman, Ryan Georgi, Glenn Slayden, and William D. Lewis. Enriching, Editing, and Representing Interlinear Glossed Text. *Computational Linguistics and Intelligent Text Processing*, Apr. 2015.

http://link.springer.com/chapter/10.1007%2F978-3-319-18111-0_3

Ryan Georgi, Fei Xia, William D. Lewis. Enhanced and Portable Dependency Projection Algorithms Using Interlinear Glossed Text. *ACL 2013*, Aug. 2013.

<http://www.aclweb.org/anthology/P13-2055>

Ryan Georgi, Fei Xia, William D. Lewis. Improving Dependency Parsing with Interlinear Glossed Text and Syntactic Projection. *COLING 2012*, Dec. 2012.

<http://www.aclweb.org/anthology/C12-2037>

Ryan Georgi, Fei Xia, William D. Lewis. Measuring the Divergence of Dependency Structures Cross-Linguistically to Improve Syntactic Projection Algorithms. *LREC 2012*, May 2012.

<http://research.microsoft.com/pubs/168001/GeorgietalLREC2012.pdf>

Ryan Georgi, Fei Xia, William D. Lewis. Comparing Language Similarity across Genetic and Typologically-Based Groupings. *COLING 2010*, Aug. 2010.

<http://www.aclweb.org/anthology/C10-1044>

TALKS

Ryan Georgi. Using IGT with INTENT: Automatically Enriching Interlinear Glossed Text (IGT). *MS/UW Symposium*, Nov. 6, 2015.

<http://research.microsoft.com/apps/video/dl.aspx?id=258553>

Ryan Georgi, Fei Xia, and William D. Lewis. Measuring the Divergence of Dependency Structures Cross-Linguistically to Improve Syntactic Projection Algorithms. *NW-NLP*, May 11, 2012.

<http://research.microsoft.com/apps/pubs/default.aspx?id=168001>

COMPUTER SKILLS

NLP Tasks

Classification, Clustering, Parsing, Semi-supervised Learning, Syntactic Projection

General

Web development (LAMP servers), Parallelization using HTCondor

ADVANCED · Python, Java, L^AT_EX
INTERMEDIATE · PHP, Javascript/JQuery, C#, SQL
BASIC · Matlab, Perl
OTHER TOOLS · Bash, Git+Subversion

TEACHING

Teaching Assistant

Advanced Statistical Methods for NLP

*Winter 2014,
2010 & 2009*

Course for introducing Computational Linguistics Masters (CLMS) students to machine learning tasks including classifiers, sequence labeling, and Expectation Maximization.

Instructor: FEI XIA · fxia@uw.edu

Teaching Assistant

NLP Systems & Applications

Spring 2013

Capstone course for CLMS program. Students work in teams to develop an NLP system. For this quarter, students created a Question Answering system to run on TREC QA-Track data.

Instructor: GINA LEVOW · levow@uw.edu

LANGUAGES

ENGLISH · Native
GERMAN · Basic