

Septina Dian Larasati

Natural Language Processing, Machine Learning, Data Engineering and Analytics

1200 N Irving St apt 4G • Arlington, VA 22201 • +1 571-719-1365
septina.larasati@gmail.com • septinalarasati.com • Green Card holder



EXPERIENCE

June 2017– August 2017

*Clarabridge
Reston VA. United States*

NLP Consultant

Assist the NLP team and implement Indonesian language pack to handle social media data into Clarabridge's platform using Java. The language pack finished and released 6 weeks ahead of schedule.

May 2016 – April 2017

*Oracle America Inc.
Washington DC. United States*

NLP Language Engineer

Working as a language engineer, responsible for several languages: Indonesian, Malay, Javanese, Tagalog, Cebuano, and Danish.

- Develop and maintain core language support capabilities for Oracle.
- Manage language data and create test sets for quality assurance.
- Communicate with linguist consultants.

May 2013 - August 2014

*Appen. (Remote)
Sydney, Australia*

NLP Contractor

Working as the computational linguist lead in the development team on Indonesian and Malay morphology project for Appen clients.

- Conducting product deliveries to the client.
- Delegating tasks to the development team.
- Developing language requirement and design documents.
- Creating and managing language data and test sets.

May 2013 - January 2014

*Lionbridge. (Remote)
Tampere, Finland*

Part-time NLP Contractor

Annotation of Indonesian language resources for Lionbridge clients.

- Adapting syntactic annotation guidelines to conform to Indonesian.
- Annotating syntax information for Indonesian.
- Cross review of multiple annotation results of all annotators.

August 2011 - April 2013

*Tilde, SIA.
Riga, Latvia*

CLARA Early Stage Researcher (ESR)

Working as an ESR in translation tools and resources for under-resourced languages. Exploring and investigating the possibilities to facilitate translation tools and resource development.

March 2012 - August 2012

Apertium + Google (Remote)

Google Summer of Code 2012 Mentor

Mentoring a National University of Singapore (NUS) student on his GSoC 2012 project: Apertium id-ms: Indonesian-Malaysian machine translation.

July 2009 - August 2009

Fondazione Bruno Kessler. Povo, Italy

Research Intern

January 2006 - August 2007

University of Indonesia. Depok, Indonesia

Part-time Database Migration Contractor

EDUCATION

●	2010 - present Ph.D. (ABD), Charles University in Prague, Prague, Czech Republic 🏆 2011 - 2013 Marie Curie CLARA Early Stage Researcher Grant Computational Linguistics Machine Translation Adviser: Prof. RNDr. Jan Hajič, Dr
●	2008 - 2010 M.Sc., Charles University in Prague, Prague, Czech Republic M.Sc., Free University of Bozen-Bolzano, Bolzano, Italy 🏆 2008 - 2010 European Masters Scholarship in Language & Communication Technologies European Masters in Language & Communication Technologies Machine Translation on Related Austronesian Languages Adviser: RNDr. Vladislav Kuboň, Ph.D
●	2003 - 2007 B.Sc., University of Indonesia, Depok, Indonesia Bachelor's in Computer Science Indonesian Semantic Analysis Adviser: Dr. Ruli Manurung

SKILLS

PERL, Java, Python, Android, PHP, R, Bash	Agile XFST, FOMA TrEd, brat
XML, Regular Expression, JSON	Moses, SRILM, IRSTLM, Apertium
GitHub, SVN MySQL, Oracle RESTful API	Stanford Parser GIMP, Inkscape

PROJECTS

Bohème	Mobile Vendor Tracker App <i>Co-founder</i>	Named Entity Extraction, Pattern Recognition, Tagging. PERL, Python, Android, Git, RESTful API, Map Geocode, Inkscape
MorphInd	Indonesian Morphological Analyzer	Natural Language Morphology. Perl, Linux Shell
IDENTIC	Indonesian-English Parallel Corpus	Data Collection
Brisq	Matching App through Q&A <i>Co-founder</i>	Machine Learning. Python, Android, Git, RESTful API, Inkscape
Votter	The Social Voting App <i>Co-founder</i>	Android, Git, RESTful API, Inkscape

LANGUAGE

PROFICIENCY*

		Understanding		Speaking		Writing	
Indonesian	C2	Native		C2	Native	C2	Native
English	C2	Proficient User		C2	Proficient User	C2	Proficient User
Malay	B1	Independent User		B1	Independent User	B1	Independent User
German	A2	Basic User		A2	Basic User	A1	Basic User
Italian	A1	Basic User		A1	Basic User	A1	Basic User
Czech	A1	Basic User		A1	Basic User	A1	Basic User

(* Self-Assessment using Common European Framework of Reference for Languages)

PUBLICATIONS

- S.D. Larasati**, T. Järvinen, E. Bertol, M.M. Rizea, M.R. Santabalbina, and M. Souček, “*Towards Cross-language Application of Dependency Grammar*”, Proceedings of the 3rd International Conference on Dependency Linguistics (Depling 2015), Uppsala, Sweden, 2015
- S.D. Larasati** and N. Green, “*Votter Corpus: A Corpus of Social Polling Language*”, Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014), Reykjavik, Iceland, 2014
- S.D. Larasati**, R.H. Susanto, and F.M. Tyers, “*Rule-based Machine Translation between Indonesian and Malaysian*”, Proceedings of the 24th International Conference on Computational Linguistics (COLING 2012) WSSANLP Workshop, Mumbai, India, 2012
- S.D. Larasati**, “*Handling Indonesian Clitics: A Dataset Comparison for an Indonesian-English Statistical Machine Translation System*”, Proceedings of the 26th Pacific Asia Conference on Language, Information, and Computation (PACLIC 26), Bali, Indonesia, 2012  Presented 11/2012
- S.D. Larasati**, N. Green, and Z. Žabokrtský, “*Indonesian Dependency Treebank: Annotation and Parsing*”, Proceedings of the 26th Pacific Asia Conference on Language, Information, and Computation (PACLIC 26), Bali, Indonesia, 2012  Presented 11/2012
- S.D. Larasati**, “*Improving Word Alignment by Exploiting Adapted Word Similarity*”, Proceedings of the 10th Association for Machine Translation in the Americas Conference (AMTA 2012) MONOMT Workshop, San Diego, USA, 2012  Presented 11/2012
- S.D. Larasati**, “*Towards an Indonesian-English SMT System: A Case Study of an Under-Studied and Under-Resourced Language, Indonesian*”, Proceedings of the 21st Annual Student Conference, Week of Doctoral Students (WDS 2012), Prague, Czech Republic, 2012  Presented 05/2012
- S.D. Larasati**, “*IDENTIC Corpus: Morphologically Enriched Indonesian - English Parallel Corpus*”, Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012), Istanbul, Turkey, 2012  Presented 05/2012
- S.D. Larasati**, V. Kuboň, and D. Zeman, “*Indonesian Morphology Tool (MorphInd): Towards an Indonesian Corpus*”, Proceedings of Workshop on Systems and Frameworks for Computational Morphology (SFCM 2011), Zurich, Switzerland, 2011  Presented 08/2011
- S.D. Larasati** and V. Kuboň, “*A study of Indonesian-to-Malaysian MT System*”, Proceedings of Workshop on Malay and Indonesian Language Engineering (MALINDO), Jakarta, Indonesia, 2010  Presented 08/2010
- S.D. Larasati**, R. Manurung, and R. Mahendra, “*Extending an Indonesian Semantic Analysis-based Question Answering System with Linguistic and World Knowledge Axioms*”, Proceedings of the 22nd Pacific Asia Conference on Language, Information, and Computation (PACLIC 22), Cebu City, Philippines, 2008
- S.D. Larasati** and R. Manurung, “*Towards a Semantic Analysis of Bahasa Indonesia for Question Answering*”, Proceedings of the 10th Conference of the Pacific Association for Computational Linguistics (PACLING 10), Melbourne, Australia, 2007