

Petra Galuščáková

335, Laboratoire d'Informatique de Grenoble
Bâtiment Imag
700 Avenue Centrale
Domaine Universitaire
38401 Saint Martin d'Hères, France
+33 (0)4 57 42 16 43
<http://galuscakova.github.io>

Résidence Le Château Grenoble
1/3 Rue Guy Allard
38000 Grenoble, France
galuscakova@gmail.com
+420 777 643 179

EDUCATION

- PhD Computational Linguistics, Institute of Formal and Applied Linguistics, Charles University, Prague, Czech Republic, 2018
- Mgr. Computational Linguistics (Equivalent to M.Sc.), Charles University, Prague, Czech Republic, 2011
- Bc. Computer Science (Equivalent to B.Sc.), Charles University, Prague, Czech Republic, 2008

PROFESSIONAL EXPERIENCE

- Since 2022 Junior Researcher, Laboratoire d'Informatique de Grenoble, Université Grenoble-Alpes
- 2021 Researcher, Summer Camp for Applied Language Exploration, Human Language Technology Center of Excellence, Johns Hopkins University
- 2018-2021 Postdoctoral Researcher, University of Maryland Institute for Advanced Computer Studies, University of Maryland
- Since 2017 Language expert, Lionbridge, online consulting
- 2010 - 2017 Research assistant, Institute of Formal and Applied Linguistics, Charles University
- 2010 Research programmer-linguist, Seznam a.s.

PUBLICATIONS

Book Chapters

- 2021 **Galuščáková Petra**, Oard Douglas W., Nair Suraj: Cross-language Information Retrieval. In: Alonso Omar, Baeza-Yates Ricardo: Advanced Topics for Information Retrieval, Springer-Verlag Berlin Heidelberg, 2021. To appear.

Conference Proceedings

- 2021 Jennifer Golbeck, Niloofarsadat Alavi, Hannah K. Bako, Calvin Bao, Matthew Barzola, Corie Brown, Kristen M. Byers, Melissa Carraway, Keaunna Cleveland, Brandon Y. Feng, Shayna Foote, **Petra Galuščáková**, Auguste H. Gezalyan, Rajesh Kumar Gnanasekaran, Saeed Hadadan, Seyyedpooya Hekmatiathar, Jeffrey Henrikson, Karan D Kaur, Nikita Khatri, Nitzan Koren, Anna Lenhart, Ruiqi Li, Yannan Liu, Jennifer MacDonald, Gloria Nzeka, Neviya Prakash, Joanna Purich, Robin Reiss, Aayushi Roy, David Rudd, Anup Sathya, Nicole R Schneider, Hilary Szu Yin Shiue, Astha Singhal, John Sutton, Gabrielle Trilling, Amulya Velamakanni, Jacob Walter, Tory White, Gareth Thomas Williams, Zachary Wilton, Peratham Wiriathamabhum, Mischa Wolfinger, Arthur Wu and Rachel Zehrung: Tweets and Social Network Data for Twitter Bot Analysis. In: 2021 International Conference on Social Computing, Behavioral-Cultural Modeling & Prediction and Behavior Representation in Modeling and Simulation, Online, 2021.
- 2021 Yanda Chen, Chris Kedzie, Suraj Nair, **Petra Galuščáková**, Rui Zhang, Douglas W. Oard and Kathleen McKeown: Cross-language Sentence Selection via Data Augmentation and Rationale Training. In: The Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2021), pp. 3881–3895, Online, 2021.
- 2021 Wan David, Kedzie Chris, Ladhak Faisal, Turcan Elsbeth, **Petra Galuščáková**, Zotkina Elena, Jiang Zhengping, Bell Peter and McKeown Kathleen: Segmenting Subtitles for Correcting ASR Segmentation Errors. In: The 16th Conference of the European Chapter of the Association for Computational Linguistics (EACL) , pp. 2842–2854, Online, 2021.
- 2020 Nair Suraj, **Galuščáková Petra**, Oard Douglas W.: Combining Contextualized and non-Contextualized Query Translations to Improve CLIR. In: 43rd International ACM SIGIR Conference of Research and Development in Information Retrieval, pp 1581–1584, Online, 2020.
- 2019 Nair Suraj, Ragni Anton, Klejch Ondrej, **Galuščáková Petra** and Oard W. Douglas: Experiments with Cross-Language Speech Retrieval for Lower-Resource Languages. In: Asia Information Retrieval Societies Conference, AIRS 2019, pp 145-157, Hong Kong, China, 2019.

- 2019 Shing Han-Chin, Barrow Joe, **Galušćáková Petra**, Oard W. Douglas, Resnik Philip: Unsupervised System Combination for Set-based Retrieval with Expectation Maximization. In: International Conference of the Cross-Language Evaluation Forum for European Languages, CLEF 2019, pp. 191-197, Lugano, Switzerland, 2019.
- 2019 Suzuki Tokinori, Ikeda Daisuke, **Galušćáková Petra**, Oard W. Douglas: Towards Automatic Cataloging of Image and Textual Collections with Wikipedia. In: International Conference on Asian Digital Libraries, ICADL 2019, pp. 167-180, Kuala Lumpur, Malaysia, 2019.
- 2018 **Galušćáková Petra**, Neužilová Lucie: Low Resource Methods for Medieval Document Sections Analysis. In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation, LREC 2018, Miyazaki, Japan, 2018.
- 2017 **Galušćáková Petra**, Batko Michal, Čech Jan, Matas Jiří, Novák David, Pecina Pavel: Visual Descriptors in Methods for Video Hyperlinking. In: Proceedings of the International Conference on Multimedia Retrieval, ICMR 2017, pp. 294-300, Bucharest, Romania, 2017.
- 2016 **Galušćáková Petra**, Saleh Shadi, Pecina Pavel: SHAMUS: UFAL Search and Hyperlinking Multimedia System. In: Proceedings of the 38th European Conference on IR Research, ECIR 2016, pp. 853-856, Padua, Italy, 2016.
- 2014 **Galušćáková Petra**, Pecina Pavel: Experiments with Segmentation Strategies for Passage Retrieval in Audio-Visual Documents. In: Proceedings of International Conference on Multimedia Retrieval, ICMR 2014, pp. 217-224, Glasgow, UK, 2014.
- 2013 Eskevich Maria, Jones Gareth J.F., Aly Robin, Ordelman Roeland, Chen Shu, Nadeem Danish, Guinaudeau Camille, Gravier Guillaume, Sébillot Pascale, De Nies Tom, Debevere Pedro, Van de Walle Rik, **Galušćáková Petra**, Pecina Pavel, Larson Martha A.: Multimedia Information Seeking through Search and Hyperlinking. In: Conference ICMR'13 International Conference on Multimedia Retrieval, pp. 287-294, Dallas, Texas, 2013.
- 2012 **Galušćáková Petra**, Pecina Pavel, Hajič Jan: Penalty Functions for Evaluation Measures of Unsegmented Speech Retrieval. In: Information Access Evaluation. Multilinguality, Multimodality, and Visual Analytics - Third International Conference of the CLEF Initiative, Lecture Notes in Computer Science, Vol. 7488, pp. 100-111, Rome, Italy, 2012
- 2012 Bojar Ondřej, Žabokrtský Zdeněk, Dušek Ondřej, **Galušćáková Petra**, Majliš Martin, Mareček David, Maršík Jiří, Novák Michal, Popel Martin, Tamchyna Aleš: The Joy of Parallelism with CzEng 1.0. In: Proceedings of the 8th International Conference on Language Resources and Evaluation, LREC 2012, pp. 3921-3928, Istanbul, Turkey, 2012.

- 2012 **Galušćáková Petra**, Bojar Ondřej: Improving SMT by Using Parallel Data of a Closely Related Language. In: Proceedings of Human Language Technologies – The Baltic Perspective, HLT 2012 pp. 58-65, Tartu, Estonia, 2012.

Peer-reviewed Workshop Papers

- 2020 **Galušćáková Petra**, Nair Suraj, Oard W. Douglas, Barrow Joseph, Shing Han-Chin, Eskander Ramy Nagah, Zotkina Elena, Zhang Rui: MATERIALizing Cross-Language Information Retrieval: A Snapshot. In: LREC Workshop on Cross-Language Search and Summarization of Text and Speech, pp 14-21, Marseilles, France, 2020.
- 2019 Oard W. Douglas, Carpuat Marine, **Galušćáková Petra**, Barrow Joseph, Nair Suraj, Niu Xing, Shing Han-Chin, Xu Weijia, Zotkina Elena, McKeown Kathleen, Muresan Smaranda, Kayi Efsun Selin, Eskander Ramy, Kedzie Chris, Virin Yan, Radev Dragomir, Zhang Rui, Gales Mark, Ragni Anton, Heafield Kenneth: Surprise Languages: Rapid-Response Cross-Language IR. In: The 9th International Workshop on Evaluating Information Access, EVIA 2019, pp. 23-28, Tokyo, Japan, 2019.
- 2015 **Galušćáková Petra**, Pecina Pavel: Audio Information for Hyperlinking of TV Content. In: Proceedings of SLAM Workshop, pp. 27-30 Brisbane, Australia, 2015.
- 2013 **Galušćáková Petra**, Popel Martin, Bojar Ondřej: PhraseFix: Statistical Post-Editing of TectoMT. In: Proceedings of the Eight Workshop on Statistical Machine Translation, pp. 141-147, Sofia, Bulgaria, 2013.
- 2012 Tamchyna Aleš, **Galušćáková Petra**, Kamran Amir, Stanojević Miloš, Bojar Ondřej: Selecting Data for English-to-Czech Machine Translation. In: Proceedings of the Seventh Workshop on Statistical Machine Translation, pp. 374-381, Montréal, Canada, 2012.
- 2011 Mareček David, Rosa Rudolf, **Galušćáková Petra**, Bojar Ondřej: Two-step Translation with Grammatical Post-processing. In: Proceedings of the Sixth Workshop on Statistical Machine Translation, WMT 2011, pp. 426-432, Edinburgh, Scotland, 2011.

Other Workshop Papers

- 2020 **Galušćáková Petra**, Nair Suraj, Oard Douglas W.: Combine and Re-Rank: The University of Maryland at the TREC 2020 Podcasts Track, Gaithersburg, MD, USA, 2020.
- 2015 **Galušćáková Petra**, Batko Michal, Kruliš Martin, Lokoč Jakub, Novák David, Pecina Pavel: CUNI at TRECVID 2015 Video Hyperlinking Task. In: TRECVID 2015 Workshop Notebook, Gaithersburg, MD, USA, 2015.
- 2015 **Galušćáková Petra** and Pecina Pavel: CUNI at MediaEval 2015 Search and Anchoring

in Video Archives: Anchoring via Information Retrieval. In: MediaEval 2015 Workshop, Wurzen, Germany, 2015.

- 2014 **Galušáková Petra**, Kruliš Martin, Lokoč Jakub and Pecina Pavel: CUNI at MediaEval 2014 Search and Hyperlinking Task: Visual and Prosodic Features in Hyperlinking. In: Proceedings of the MediaEval 2014 Multimedia Benchmark Workshop, Barcelona, Spain, 2014
- 2014 **Galušáková Petra** and Pecina Pavel: CUNI at MediaEval 2014 Search and Hyperlinking Task: Search Task Experiments. In: Proceedings of the MediaEval 2014 Multimedia Benchmark Workshop, Barcelona, Spain, 2014.
- 2013 **Galušáková Petra** and Pecina Pavel: CUNI at MediaEval 2013 Similar Segments in Social Speech Task. In: MediaEval 2013 Workshop, Barcelona, Spain, 2013.
- 2012 **Galušáková Petra** and Pecina Pavel: CUNI at MediaEval 2012 Search and Hyperlinking Task. In: Working Notes Proceedings of the MediaEval Workshop, Pisa, Italy, 2012

Data and Software (7643 cumulative downloads)

- 2018 **Galušáková Petra**, Neužilová Lucie: Medieval Charter Sections Corpus, <http://hdl.handle.net/11234/1-1952>, 2018
- 2017 **Galušáková Petra**, Pecina Pavel, Hoffmannová Petra, Hajič, Jan, Ircing Pavel, Švec Jan: Czech Malach Cross-lingual Speech Retrieval Test Collection, <http://hdl.handle.net/11234/1-1912>, 2017
- 2017 **Galušáková Petra**, Saleh Shadi: SHAMUS (UFAL Search and Hyperlinking Multimedia System), <https://ufal.mff.cuni.cz/shamus>, 2012
- 2012 **Galušáková Petra**, Bojar Ondřej: WMT 2011 Testing Set, <http://hdl.handle.net/11858/00-097C-0000-0006-AADA-9>, 2012
- 2012 **Galušáková Petra**, Garabík Radovan, Bojar Ondřej: Czech-Slovak Parallel Corpus, <http://hdl.handle.net/11858/00-097C-0000-0006-AADF-0>, 2012
- 2012 **Galušáková Petra**, Garabík Radovan, Bojar Ondřej: English-Slovak Parallel Corpus, <http://hdl.handle.net/11858/00-097C-0000-0006-AAE0-A>, 2012
- 2012 **Galušáková Petra**, Bojar Ondřej: Manually Classified Errors in Cs->Sk Translation,

- <http://hdl.handle.net/11858/00-097C-0000-0006-AADB-7>, 2012
- 2012 **Galuščáková Petra**, Bojar Ondřej; Manually Classified Errors in En->Sk Translation, <http://hdl.handle.net/11858/00-097C-0000-0006-AADC-5>, 2012
- 2012 Bojar Ondřej, **Galuščáková Petra**: Manually Ranked Translation Output, <http://hdl.handle.net/11858/00-097C-0000-0006-AADD-3>, 2012
- 2011 Bojar Ondřej, Žabokrtský Zdeněk, Dušek Ondřej, **Galuščáková, Petra**, Majliš Martin, Mareček David, Maršík Jiří, Novák Michal, Popel Martin Tamchyna Aleš: Czech-English Parallel Corpus 1.0 (CzEng 1.0), <http://hdl.handle.net/11234/1-1458>, 2011
- 2011 **Galuščáková Petra**: Antenna Noise Temperature Calculator, <http://om6aa.eu/antc.php>, 2011

AWARDS

- 2022 Innovative solution award for data extraction in the NASA AI Risk Prediction Challenge (\$5,000)
- 2020 Ranked first in the Podcasts Challenge at TREC
- 2016 Part of Nuages team ranked first at the Archives Unleashed Hackathon
- 2015 Ranked first in the Search and Hyperlinking Tasks at the MediaEval Benchmark
- 2012 Ranked first in the Search Task at the MediaEval Benchmark

AWARDED GRANTS

- 2016 ECIR Student Travel Grant (500€)
- 2016 Archives Unleashed 2.0: Web Archive Datathon travel grant (\$1,000 CAD)
- 2016 Archives Unleashed: Web Archive Hackathon travel grant (\$1,000 CAD)
- 2015 Google PhD NLP Summit travel grant
- 2014-2016 Principal Investigator
Interactive information retrieval in audiovisual dialogue corpora project (732k CZK, equivalent to ~\$33k)
- 2013 SIGIR Student Travel Grant (\$1642.66)
- 2013 PROMISE Winter School Information Retrieval meets Information Visualization Scholarship (500€)

INVITED TALKS

- 2021 Multi-stage Information Retrieval
Massachusetts Institute of Technology, DeCoDE lab, Online, July 7
- 2020 Combining Evidence for Cross-language Information Retrieval
Fordham University, Online, March 31
Combining Modalities for Hyperlinking Navigation in Multimedia Collections
University of Wisconsin, Madison, WI, USA, February 17
- 2019 Cross-Language Information Retrieval in the MATERIAL Program
Qatar Computing Research Institute, Doha, Qatar, March 28
- 2015 Multimodal Features for Search and Hyperlinking of TV Content
Center for Machine Perception, Czech Technical University, Prague, Czech Republic, February 12
- 2014 Multimodal Features for Search and Hyperlinking of Video Content
IDIAP Research Institute, Martigny, Switzerland, October 29

CAMPUS TALKS

- 2019 MATERIALizing Cross-Language Information Retrieval
University of Maryland, MD, USA, CLIP Talk
Information Retrieval in the CLIP Lab
University of Maryland, MD, USA, Computer & Computational Sciences Research Day
- 2018 Information Retrieval and Navigation in Audio-visual Archives
University of Maryland, MD, USA, CLIP Talk
- 2016 Czech Malach Cross-lingual Speech Retrieval Test Collection
Malach Centre for Visual History, Prague, Czech Republic
- 2014 Content-based Segmentation for Audio-Visual Information Retrieval
Seminar of the Institute of Formal and Applied Linguistics, Prague, Czech Republic

TEACHING

University of Maryland, College Park, MD, Sole Instructor

Decision-Making for Information Science (Fall 2021)
Information Retrieval Reading Group (Spring, Summer, Fall 2020, Spring 2021)
Information Retrieval Systems (Fall 2019)

**Advanced Solutions and Technologies Summer School, Prague Czech Republic,
Sole Instructor**

Data Mining (Summer 2017)

PROFESSIONAL SERVICE

Peer Review

Program Committee member for SIGIR, 2021 - 2022

Reviewer for Concurrency and Computation: Practice and Experience, 2022

Reviewer for ACM Computing Surveys, 2021

Reviewer for The Scientific Grant Agency of the Slovak Republic and of the Slovak Academy of Sciences, 2021

Reviewer for the The Prague Bulletin of Mathematical Linguistics, 2020

Reviewer for the Journal of Linguistics of the Slovak Academy of Science, 2020

Program Committee member for Cross-Language Search and Summarization of Text and Speech Workshop, 2020

Program Committee member for LREC, 2020

Reviewer for RANLP (Recent Advances in Natural Language Processing), 2017

Program Committee member for ICMI (International Conference on Multimodal Interaction), 2015 - 2016

To Profession

Poster Session Chair at ICMR (International Conference on Multimedia Retrieval), 2017

To Community

Volunteer, Umsemumtam.cz, 2015-2017

Volunteer, LanguageTool, 2011

Website administrator, Dražkovce municipality, 2007-2010

PROFESSIONAL ASSOCIATIONS

Association for Computing Machinery, since 2014

ACM Special Interest Group on Information Retrieval, since 2019

LANGUAGES

Slovak: native speaker

English: fluent, The First Certificate in English (grade A)

Czech: fluent

German, French: basic

CERTIFICATES

An Introduction to Evidence-Based STEM Undergraduate Teaching (provided by The Center for the Integration of Research Teaching and Learning)

Machine Learning (online, provided by Stanford University on coursera.org)

Neural Networks and Deep Learning (online, provided by deeplearning.ai on coursera.org)

Data Mining with Weka (online, provided by the University of Waikato)

More Data Mining with Weka (online, provided by the University of Waikato)

Creative Writing (provided by Czech News Agency)

TRAININGS AND SEMINARS

NASA Second AI and Data Science Workshop for Earth and Space Sciences, online, 2021

Machine Translation Marathon, College Park, MD, USA, 2019

NASA Goddard Workshop on Artificial Intelligence, Greenbelt, MD, USA, 2018

Digital Editing: Encoding, tagging, and online publishing, Siena, Italy, 2017

McKinsey & Company Centered Leadership Workshop, Prague, Czech Republic, 2016

Archives Unleashed 2.0: Web Archive Datathon, Washington D.C., USA, 2016

Archives Unleashed: Web Archive Hackathon, Toronto, Canada, 2016

Google PhD NLP Summit, Zurich, Switzerland, 2015

Designing for Multimedia Archives - Training Event. Lausanne, Switzerland, 2014

International PhD School in Language and Speech Technologies, Tarragona, Spain 2012

PROMISE Winter School 2012 Information Retrieval meets Information Visualization, Zinal, Switzerland 2012

REFERENCES

Douglas W. Oard (including teaching reference)

Professor and Associate Dean for Faculty
University of Maryland, College of Information Studies
Patuxent 1109C, 20740 College Park, MD, United States
oard@umd.edu

Kathleen McKeown

Henry and Gertrude Rothschild Professor of Computer Science
Columbia University, Department of Computer Science
Room 722, Schapiro CESPR, New York NY 10027, United States
kathy@cs.columbia.edu

Dragomir Radev

A. Bartlett Giamatti Professor of Computer Science
Yale University, Department of Computer Science,
Room 319, 17 Hillhouse Avenue, New Haven, CT 06511, United States
dragomir.radev@yale.edu

Pavel Pecina

Assistant Professor
Charles University in Prague
Malostranské náměstí 25, 11800, Prague 1, Czech Republic
pecina@ufal.mff.cuni.cz