



16535 Grant Bishop Ln.
Los Gatos, CA 95032

May 18, 2023

2023 مايو 18

Esteemed Patrons of the King Salman Global Academy Prize for Arabic Language,

On behalf of aiXplain Inc., it is my pleasure to nominate Dr. Kareem Darwish for the King Salman Global Academy Prize for Arabic Language. Dr. Darwish is one of the pioneers in the automated processing of Arabic language including Modern Standard Arabic (MSA), Classical Arabic, and Arabic dialects.

The contributions of Dr. Darwish to Arabic NLP include community service, productization of Arabic NLP, and scientific contributions. He is well recognized in the Arabic NLP community as one of its earliest practitioners of Arabic NLP and one the community leaders.

Community service:

Dr. Darwish has been active in building and supporting the Arabic NLP community over the past 20 years. He co-organized one of the earliest workshops involving Arabic NLP back in 2002 with Dr. Mona Diab and Dr. Nizar Habash, namely the *Workshop on Computational Approaches to Semitic Languages*. He successfully co-organized multiple iterations of the *Workshop on Arabic Natural Language Processing (WANLP)* starting with the first iteration of the Workshop in 2014 along with the iterations in 2015 and 2017. He also co-organized the 2016, 2018, 2020, and 2022 iterations of the *Workshop on Open-Source Arabic Corpora and Processing Tools (OSACT)* with Dr. Hend Al-Khalifa, Dr. Walid Magdy, Dr. Tamer Elsayed, and Dr. Abdulmohsen Al-Thubaity. Dr. Darwish continues to serve on the advisory board for SIGARAB, which is the special interest group that organizes both WANLP and OSACT.

He was responsible for submitting successful bids to host two major NLP and Social Computing conferences to the Middle East, namely EMNLP 2014 and SocInfo 2019, both of which were held in Doha.

In 2011, Dr. Darwish founded the Arabic Language Technologies group at the Qatar Computing Research Institute. This group soon became one of the premier Arabic NLP and speech groups in the Middle East and the World. The group was actively publishing in top tier conferences such as ACL, EMNLP, AACL, NAACL, InterSpeech, and ICASSP. Further, the group produced two state-of-the-art toolkits for Arabic NLP and text processing, namely *Farasa*, which included morphological segmentation,

السادة الأفاضل القائمين على جائزة مجمع الملك سلمان العالمية للغة العربية،

نيابة عن شركة aiXplain Inc، يسعدني ترشيح الدكتور كريم درويش لجائزة مجمع الملك سلمان العالمية للغة العربية. الدكتور كريم درويش هو أحد الرواد في مجال المعالجة الآلية للغة العربية بما في ذلك اللغة العربية الفصحى واللهجات المختلفة. تشمل مساهمات الدكتور درويش في المعالجة الآلية للغة الكم الوافر من المساهمات العلمية من أبحاث وأوراق علمية، وخدمة للمجتمع البحثي، وإنتاج البرامج المتطورة لمعالجة اللغة العربية.

الدكتور درويش يشار إليه بالبنان في مجال المعالجة الآلية للغة العربية باعتباره واحدًا من أوائل الممارسين في المجال وأحد أبرز الباحثين في المجال.

خدمة المجتمع:

كان وما زال للدكتور كريم درويش دور نشط في بناء ودعم المجتمع البحثي المتخصص في المعالجة الآلية للغة العربية على مدار العشرين عامًا الماضية. فقد شارك في تنظيم إحدى أولى ورش العمل التي تتضمن المعالجة الآلية للغة العربية في عام 2002 مع الدكتورة منى دياب والدكتور نزار حبش، وهي ورشة عمل المناهج الحسابية للغات السامية. شارك بنجاح في تنظيم نسخ متعددة لورشة عمل معالجة اللغة العربية – (Workshop on Arabic NLP – WANLP). بدءًا من النسخة الأولى لورشة العمل في عام 2014 بجانب نسخ 2015 و 2017. كما شارك في تنظيم نسخ 2016 و 2018 و 2020 و 2022 لورشة عمل المدونات العربية مفتوحة المصدر وأدوات المعالجة (Workshop on Open-Source Arabic Corpora and Processing Tools – OSAC) مع الدكتورة هند الخليفة، والدكتور وليد مجدي، والدكتور تامر السيد، والدكتور عبدالمحسن الثبتي. يواصل الدكتور درويش العمل كعضو في المجلس الاستشاري لـ SIGARAB، وهي مجموعة ذات اهتمامات خاصة تقوم بتنظيم كلا من WANLP و OSACT.

إضافة إلى ذلك، كان الدكتور درويش مسؤولاً عن تقديم عروض ناجحة لاستضافة مؤتمرات رئيسيين في المعالجة الآلية للغة والحوسبة الاجتماعية في الشرق الأوسط، وهما EMNLP 2014 و SocInfo 2019، حيث عقد كلاهما في الدوحة.

في عام 2011، أسس الدكتور درويش مجموعة تقنيات اللغة العربية في معهد قطر لبحوث الحوسبة. سرعان ما أصبحت هذه المجموعة واحدة من المجموعات الرائدة في المعالجة الآلية للغة العربية في الشرق الأوسط والعالم ككل. كانت المجموعة تنشر بنشاط في أفضل المؤتمرات العلمية مثل ACL و EMNLP و AACL و NAACL و InterSpeech و ICASSP. علاوة على ذلك، أنتجت المجموعة حزميتين من أدوات معالجة النصوص العربية ومعالجة النصوص، وهما Farasa، والتي تتضمن التجزئة الصرفية، والتحليل

aiXplain

16535 Grant Bishop Ln.
Los Gatos, CA 95032

syntactic analysis, named entity recognition, syntactic and dependency parsing, and diacritization, and *ASAD*, which included offensive language detection, emotion detection, sentiment analysis, dialect identification, and topic classification. The group produced the largest dataset of transcribed Arabic data, composed of more than 1,000 hours, which had a huge effect on Arabic speech recognition.

Dr. Darwish later became the director of the group. His efforts were recognized by Qatar Foundation, which awarded him the Distinguished Creator Award in 2020.

Productization:

Dr. Kareem Darwish took a leading role in developing Arabic NLP toolkits, which were made available to the research community, namely: **Farasa Arabic NLP Toolkit** (<https://farasa.qcri.org>): He was responsible for writing much of the code for Farasa, which performs segmentation, POS tagging, named entity recognition, diacritization, and syntactic and dependency parsing. Farasa achieves state-of-the-art results in all its components while being more than one order of magnitude faster than similar toolkits such as Camel or MADAMIRA.

Asad Arabic text processing toolkit (<https://asad.qcri.org>): He collaborated with other scientists and engineers to build a variety of models for Arabic emotion detection, sentiment analysis, offensive language detection, hate speech detection, and dialect identification.

Natiq Arabic text-to-speech (<https://tts.qcri.org>): He was instrumental in procuring data for TTS including data selection, preparation, and recording and in pre-processing text for TTS including custom diacritization and number conversion.

SADA Speech Recognition Dataset: Dr. Darwish was the team lead for a major annotation project with SDAIA involving the transcription and labeling of 6,000 hours of audio including: Designing guidelines and procedures; Developing pre-annotation models and speed up processing including quality estimation, gender identification, etc.; and Implementing a rigorous QA process to ensure annotation quality. Dr. Darwish is currently leading multiple initiatives in building large Arabic language models and state-of-the-art Arabic text to speech models.

Scientific contributions:

النحوي، والتعرف على أسماء الأعلام، والإعراب النحوي، والتشكيل، وحزمة ASAD، والتي تتضمن الكشف عن اللغة الهجومية وخطاب الكراهية، والكشف عن المشاعر، وتحليل المشاعر، وتحديد اللهجة، وتصنيف الموضوعات. كما أنتجت المجموعة أكبر مدونة مفتوحة للتفريغ الآلي والتي تضمنت أكثر من ألف ساعة صوتية، وكان لها كبير الأثر في تطوير التفريغ الآلي للغة العربية.

أصبح الدكتور درويش فيما بعد مديراً للمجموعة. وقد نالت جهوده تقدير مؤسسة قطر التي منحتة جائزة المبدع المتميز في عام 2020.

تحويل الأبحاث إلى منتجات:

لعب الدكتور كريم درويش دورًا رائدًا في تطوير حزم أدوات لمعالجة اللغة العربية آلياً، والتي تم توفيرها لمجتمع البحث مجاناً، وهي: حزمة فراسا للمعالجة الآلية للغة العربية (<https://farasa.qcri.org>): كان مسؤولاً عن برمجة الجزء الأكبر من Farasa، والتي تقوم بالتجزئة الصرفية، والتوسيم بأجزاء الكلم، والتعرف على أسماء الأعلام، والتشكيل، والتحليل النحوي. تحقق Farasa أفضل النتائج على الإطلاق في جميع مكوناتها مع كونها أسرع بأكثر من عشر مرات مقارنة بالحزم المماثلة مثل Camel و MADAMIRA.

حزمة أسد لمعالجة النصوص العربية (<https://asad.qcri.org>): تعاون مع علماء ومهندسين آخرين لبناء نماذج متنوعة لاكتشاف المشاعر في النصوص العربية، وتحليل المشاعر، واكتشاف اللغة العدوانية، واكتشاف الكلام الذي يحض على الكراهية، وتحديد اللهجات.

ناطق آلي للغة العربية (<https://tts.qcri.org>): كان له دور فعال في تجهيز البيانات المطلوبة لتدريب الناطق الآلي بما في ذلك اختيار البيانات وإعدادها وتسجيلها والمعالجة المسبقة للنص، بما في ذلك التشكيل المناسب للنطق الآلي وتحويل الأرقام.

مدونة صدى للتفريغ الآلي: كان الدكتور درويش قائد الفريق لمشروع مع SDAIA يتضمن تفريغ وتوسيم 6,000 ساعة صوتية بما في ذلك: تصميم الإرشادات التوجيهية والإجراءات التفريغ والتوسيم؛ وتطوير نماذج ما قبل التوسيم لتسريع المعالجة بما في ذلك تقدير الجودة وتحديد جنس المتحدث وما إلى ذلك؛ وتنفيذ خطوات محكمة لضمان الجودة لضمان أفضل تفريغ وتوسيم.

يقود الدكتور درويش حالياً مبادرات متعددة في بناء نماذج كبيرة للغة العربية وتدريب نماذج للنطق الآلي.

المساهمات العلمية:



16535 Grant Bishop Ln.
Los Gatos, CA 95032

Dr. Darwish authored and co-authored more than 100 scientific books, book chapters, and refereed papers in top journals and conferences that cover different aspects of Arabic processing, including Arabic Information Retrieval (IR) and natural language processing.

For Arabic IR

1. He wrote the reference book on Arabic IR entitled: *Arabic Information Retrieval*, which covers retrieval methods for different modalities such as webpages, news, printed documents, audio documents, and images.
2. He did seminal work on searching printed Arabic documents, cross-lingual retrieval between Arabic and other languages, and image search. Select publications include:
 - a. Ph.D. Thesis: *Probabilistic Methods for Searching OCR-Degraded Arabic Text*
 - b. Darwish, K. and D. Oard. *Term Selection for Searching Printed Arabic*. In ACM SIGIR: 261-268, 2002.
 - c. Magdy, W., K. Darwish, and M. El-Sabban. *Efficient Language-Independent Retrieval of Printed Documents without OCR*. Symposium on String Processing and Information Retrieval (SPIRE'09), 334-343, 2009.
 - d. Hefny, A., K. Darwish, Ali El-kahki. *Crowdsourcing for Cross Language and Multilingual Web Search*. ECIR-2011, 2011.
 - e. Darwish, K., W. Magdy, O. Emam, A. Abdelsapor, N. Adly, M. Nagi. *Building a Heterogeneous Information Retrieval Collection of Arabic Document Images*. LREC 2006.
 - f. Darwish, K. and W. Magdy. 2007. *Error correction vs. query garbling for Arabic OCR document retrieval*. ACM Transactions on Information Systems (TOIS) Vol. 26, Issue 1, Article 5, 2007.
 - g. Magdy, W. and Darwish, K. Arabic. 2006. *OCR error correction using character segment correction, language modeling, and shallow morphology*. Empirical Methods in Natural Language Processing (EMNLP'06). Sydney, Australia, 408-414, 2006.
 - h. Clough, P., A. Al-Maskari, and K. Darwish. 2006. *Providing Multilingual Access to Flickr for Arabic Users*. Evaluation of Multilingual and Multi-modal Information Retrieval, 205-216, 2006.
3. He authored many papers on Arabic preprocessing for IR in the context of news and social media search. Select publications include:
 - a. Darwish, K. and D. Oard. *Adapting Morphology for Arabic Information Retrieval*. Arabic Computational Morphology:

ألف الدكتور درويش وشارك في تأليف أكثر من 100 كتاب علمي وفصل في كتب وأوراق علمية محكمة في أهم المجالات والمؤتمرات التي تغطي جوانب مختلفة من معالجة اللغة العربية، بما في ذلك استرجاع المعلومات العربية ومعالجة اللغة الطبيعية.
بالنسبة لاسترجاع البيانات:

1. قام بتأليف المرجع الرئيس في استرجاع البيانات العربية بعنوان: Arabic Information Retrieval، والذي يغطي وسائل البحث في مختلف الوثائق مثل صفحات الويب والمقالات الأخبارية والوثائق المطبوعة والمستندات الصوتية والصور.
2. قام بعمل الأبحاث المؤسسة في مجال البحث في المستندات العربية المطبوعة، واسترجاع البيانات بين اللغة العربية واللغات الأخرى، والبحث في الصور. هذه قائمة ببعض الأبحاث المختارة:
 - a. Ph.D. Thesis: *Probabilistic Methods for Searching OCR-Degraded Arabic Text*
 - b. Darwish, K. and D. Oard. *Term Selection for Searching Printed Arabic*. In ACM SIGIR: 261-268, 2002.
 - c. Magdy, W., K. Darwish, and M. El-Sabban. *Efficient Language-Independent Retrieval of Printed Documents without OCR*. Symposium on String Processing and Information Retrieval (SPIRE'09), 334-343, 2009.
 - d. Hefny, A., K. Darwish, Ali El-kahki. *Crowdsourcing for Cross Language and Multilingual Web Search*. ECIR, 2011.
 - e. Darwish, K., W. Magdy, O. Emam, A. Abdelsapor, N. Adly, M. Nagi. *Building a Heterogeneous Information Retrieval Collection of Arabic Document Images*. LREC 2006.
 - f. Darwish, K. and W. Magdy. 2007. *Error correction vs. query garbling for Arabic OCR document retrieval*. ACM Transactions on Information Systems (TOIS) Vol. 26, Issue 1, Article 5, 2007.
 - g. Magdy, W. and Darwish, K. Arabic. 2006. *OCR error correction using character segment correction, language modeling, and shallow morphology*. Empirical Methods in Natural Language Processing (EMNLP'06). Sydney, Australia, 408-414, 2006.
 - h. Clough, P., A. Al-Maskari, and K. Darwish. 2006. *Providing Multilingual Access to Flickr for Arabic Users*. Evaluation of Multilingual and Multi-modal Information Retrieval, 205-216, 2006.
3. قام بتأليف العديد من الأبحاث المعنية بطرق التحليل الصرفي للنصوص للحصول على أفضل نتائج في البحث. هذه بعض الأبحاث المختارة:
 - a. Darwish, K. and D. Oard. *Adapting Morphology for Arabic Information Retrieval*. Arabic Computational Morphology:



16535 Grant Bishop Ln.
Los Gatos, CA 95032

- Knowledge-based and Empirical Methods. In Kluwer/Springer's series on Text, Speech, and Language Technology, 2006.
- b. Darwish, K., H. Hassan, and O. Emam. Examining the Effect of Improved Context Sensitive Morphology on Arabic Information Retrieval. In ACL Workshop on Computational Approaches to Semitic Languages, Ann Arbor, 2005.
- c. Kareem Darwish, Ahmed A. Ali, Ahmed Abdelali, 2014, Query Term Expansion by Automatic Learning of Morphological Equivalence Patterns from Wikipedia. In Proceedings of the Workshop on Semantic Matching in Information Retrieval (SMIR 2014), July 2014.
- d. Kareem Darwish, Ahmed Ali. *Arabic Retrieval Revisited: Morphological Hole Filling*. ACL-2012, 2012.
- e. Kareem Darwish, Walid Magdy, Ahmed Mourad. 2012. *Language Processing for Arabic Microblog Retrieval*. Proceedings of the 21st ACM international conference on Information and knowledge management, CIKM-2012, pp. 2427-2430.
- Knowledge-based and Empirical Methods. In Kluwer/Springer's series on Text, Speech, and Language Technology, 2006.
- b. Darwish, K., H. Hassan, and O. Emam. Examining the Effect of Improved Context Sensitive Morphology on Arabic Information Retrieval. In ACL Workshop on Computational Approaches to Semitic Languages, 2005.
- c. Kareem Darwish, Ahmed A. Ali, Ahmed Abdelali, 2014, Query Term Expansion by Automatic Learning of Morphological Equivalence Patterns from Wikipedia. In Proceedings of the Workshop on Semantic Matching in Information Retrieval (SMIR 2014), July 2014.
- d. Kareem Darwish, Ahmed Ali. Arabic Retrieval Revisited: Morphological Hole Filling. ACL-2012, 2012.
- e. Kareem Darwish, Walid Magdy, Ahmed Mourad. 2012. *Language Processing for Arabic Microblog Retrieval*. Proceedings of the 21st ACM international conference on Information and knowledge management, CIKM-2012, pp. 2427-2430.

For Arabic NLP:

1. He authored the first paper on statistical Arabic morphological analysis and followed with multiple papers on Arabic morphology:
 - a. Darwish, K. Building a Shallow Morphological Analyzer in One Day. In ACL Workshop on Computational Approaches to Semitic Languages: 47-54, 2002.
 - b. Kareem Darwish and Hamdy Mubarak. Farasa: Fast and Accurate Arabic Word Segmenter. LREC'16. 2016
2. He worked on Arabic syntactic analysis with the state-of-the-art part of speech tagger for MSA and dialects:
 - a. Kareem Darwish, Hamdy Mubarak, Ahmed Abdelali, Mohmed Eldesouki. 2017. Arabic POS Tagging: Don't Abandon Feature Engineering Just Yet. Proceedings of the Third Arabic Natural Language Processing Workshop, pp. 130--137, 2017.
 - b. Darwish, Kareem, Hamdy Mubarak, Ahmed Abdelali, Mohamed Eldesouki, Younes Samih, Randah Alharbi, Mohammed Attia, Walid Magdy, and Laura Kallmeyer. 2018. Multi-Dialect Arabic POS Tagging: A CRF Approach. In Proceedings of the Eleventh International Conference on

بالنسبة للمعالجة الآلية للغة العربية:

1. ألف الدكتور كريم درويش أول بحث في استخدام الوسائل الإحصائية في التحليل الصرفي للغة العربية وتبع ذلك عدة أبحاث أخرى في نفس الموضوع:
 - a. Darwish, K. Building a Shallow Morphological Analyzer in One Day. In ACL Workshop on Computational Approaches to Semitic Languages: 47-54, 2002.
 - b. Kareem Darwish and Hamdy Mubarak. Farasa: Fast and Accurate Arabic Word Segmenter. LREC'16. 2016
2. عمل على تدريب أفضل محللات سياقية، وإنتاج موسمات بأجزاء الكلم للغة العربية الفصحى واللهجات العامية:
 - a. Kareem Darwish, Hamdy Mubarak, Ahmed Abdelali, Mohmed Eldesouki. 2017. Arabic POS Tagging: Don't Abandon Feature Engineering Just Yet. Proceedings of the Third Arabic Natural Language Processing Workshop, pp. 130--137, 2017.
 - b. Darwish, Kareem, Hamdy Mubarak, Ahmed Abdelali, Mohamed Eldesouki, Younes Samih, Randah Alharbi, Mohammed Attia, Walid Magdy, and Laura Kallmeyer. 2018. Multi-Dialect Arabic POS Tagging: A CRF Approach. In Proceedings of the Eleventh International Conference on



16535 Grant Bishop Ln.
Los Gatos, CA 95032

- Language Resources and Evaluation (LREC-2018). Miyazaki, Japan, May 2018.
- c. Kareem Darwish, Ahmed Abdelali, Hamdy Mubarak, 2014, Using Stem-Templates to Improve Arabic POS and Gender/Number Tagging. In Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2014), Reykjavik, Iceland, May 2014.
- d. Kareem Darwish, Hamdy Mubarak, Ahmed Abdelali, Mohamed Eldesouki, Younes Samih, Rehab Alharbi, Walid Magdy. 2018. Multi-Dialect Arabic POS Tagging: A CRF Approach. LREC 2018.
3. He wrote papers on Arabic Named Entity Recognition (NER) including the development of core datasets and producing state-of-the-art results:
- a. Kareem Darwish. 2013. Named Entity Recognition using Cross-lingual Resources: Arabic as an Example. Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, pp. 1558–1567.
- b. Kareem Darwish and Wei Gao. Simple Effective Microblog Named Entity Recognition: Arabic as an Example. In Proceedings of the 9th Language Resources and Evaluation Conference (LREC 2014), Reykjavik, Iceland, May 2014.
- c. Abdul-Hamid, A., K. Darwish. 2010. *Simplified Feature Set for Arabic Named Entity Recognition*, ACL 2010 NEWS workshop, 2010.
4. For Arabic diacritization, he developed multiple models including the current state-of-the-art in Arabic diacritization for MSA, classical Arabic, and dialectal Arabic:
- a. Kareem Darwish, Ahmed Abdelali, Hamdy Mubarak, Mohamed Eldesouki. 2021. Arabic diacritic recovery using a feature-rich bilstm model. Transactions on Asian and Low-Resource Language Information Processing, (20) 2: 1-18.
- b. Hamdy Mubarak, Ahmed Abdelali, Kareem Darwish, Mohamed Eldesouki, Younes Samih, Hassan Sajjad. 2019. A System for Diacritizing Four Varieties of Arabic. Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP'2019): 2019.
- c. Hamdy Mubarak, Ahmed Abdelali, Hassan Sajjad, Younes Samih, Kareem Darwish. 2019. Highly effective Arabic
- Language Resources and Evaluation (LREC-2018). Miyazaki, Japan, May 2018.
- c. Kareem Darwish, Ahmed Abdelali, Hamdy Mubarak, 2014, Using Stem-Templates to Improve Arabic POS and Gender/Number Tagging. In Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2014), Reykjavik, Iceland, May 2014.
- d. Kareem Darwish, Hamdy Mubarak, Ahmed Abdelali, Mohamed Eldesouki, Younes Samih, Rehab Alharbi, Walid Magdy. 2018. Multi-Dialect Arabic POS Tagging: A CRF Approach. LREC 2018.
3. كتب عدة أبحاث عن التعرف على أسماء الأعلام، والتي حققت أفضل نتائج مقارنة بالنماذج الأخرى، وقام بإعداد عدة مدونات للتعرف على أسماء الأعلام في النصوص الإخبارية ووسائل التواصل الاجتماعي: Kareem Darwish. 2013. Named Entity Recognition using Cross-lingual Resources: Arabic as an Example. Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, pp. 1558–1567.
- b. Kareem Darwish and Wei Gao. Simple Effective Microblog Named Entity Recognition: Arabic as an Example. In Proceedings of the 9th Language Resources and Evaluation Conference (LREC 2014), Reykjavik, Iceland, May 2014.
- c. Abdul-Hamid, A., K. Darwish. 2010. *Simplified Feature Set for Arabic Named Entity Recognition*, ACL 2010 NEWS workshop, 2010.
4. درب أفضل نماذج على الإطلاق لتشكيل اللغة العربية سواء الفصحى أو اللهجات العامية: Kareem Darwish, Ahmed Abdelali, Hamdy Mubarak, Mohamed Eldesouki. 2021. Arabic diacritic recovery using a feature-rich bilstm model. Transactions on Asian and Low-Resource Language Information Processing, (20) 2: 1-18.
- b. Hamdy Mubarak, Ahmed Abdelali, Kareem Darwish, Mohamed Eldesouki, Younes Samih, Hassan Sajjad. 2019. A System for Diacritizing Four Varieties of Arabic. Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP'2019): System Demonstrations, Hong Kong, November 2019.
- c. Hamdy Mubarak, Ahmed Abdelali, Hassan Sajjad, Younes Samih, Kareem Darwish. 2019. Highly effective Arabic



16535 Grant Bishop Ln.
Los Gatos, CA 95032

- Diacritization using Sequence to Sequence Modeling.
Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT'2019), 2019.
5. He worked on improving Arabic ↔ English transliteration and translation for MSA and dialects and Arabizi to Arabic conversion:
- Ali El Kahki, Kareem Darwish, Ahmed Saad El Din, Mohamed Abd El-Wahab. *Transliteration Mining Using Large Training and Test Sets*. NAACL-2012, 2012.
 - Ali El-Kahky, Kareem Darwish, Ahmed Saad Aldein, Mohamed Abd El-Wahab, Ahmed Hefny, Waleed Ammar. *Improved Transliteration Mining Using Graph Reinforcement*. EMNLP-2011, 2011.
 - Hany Hassan, Kareem Darwish. 2014. *Statistical Machine Translation*. Natural Language Processing of Semitic Languages, pp. 199-219, Springer Berlin Heidelberg.
 - Hassan Sajjad, Kareem Darwish and Yonatan Belinkov. *Translating Dialectal Arabic to English*. In Proceedings of the Association for Computational Linguistics (ACL 2013), Sofia, Bulgaria, August 2013
 - Darwish K. *Arabizi Detection and Conversion to Arabic*. In Proceedings of the EMNLP 2014 Workshop on Arabic Natural Language Processing (ANLP) 2014 Oct (pp. 217-224).
6. He co-authored some of the earliest works on Arabic text classification in offensive language, hate speech detection, dialect identification, topic classification, and emotion detection:
- Hamdy Mubarak, Kareem Darwish, Walid Magdy, Tamer Elsayed, Hend Al-Khalifa. 2020. *Overview of OSACT4 Arabic Offensive Language Detection Shared Task*. In: OSACT 4: The 4th Workshop on Open-Source Arabic Corpora and Processing Tools (LREC-2020). Marseille, France, May 2020.
 - Hamdy Mubarak, Kareem Darwish. 2019. *Arabic Offensive Language Classification on Twitter*. International Conference on Social Informatics (SocInfo'2019). 2019.
 - Sabit Hassan, Shaden Shaar, Kareem Darwish. 2022. *Cross-Lingual Emotion Detection*. Proceedings of the Thirteenth Language Resources and Evaluation Conference.
- Diacritization using Sequence to Sequence Modeling.
Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT'2019), Minneapolis, MN, USA, June 2019.
5. شارك في عدة أبحاث معنية بالترجمة الآلية، والترجمة الصوتية سواء لأسماء الأعلام أو ما يسمى بالعريزي (اللغة العربية المكتوبة بالحروف اللاتينية) لتحويلها لنصوص عربية عادية:
- Ali El Kahki, Kareem Darwish, Ahmed Saad El Din, Mohamed Abd El-Wahab. *Transliteration Mining Using Large Training and Test Sets*. NAACL-2012, 2012.
 - Ali El-Kahky, Kareem Darwish, Ahmed Saad Aldein, Mohamed Abd El-Wahab, Ahmed Hefny, Waleed Ammar. *Improved Transliteration Mining Using Graph Reinforcement*. EMNLP-2011, 2011.
 - Hany Hassan, Kareem Darwish. 2014. *Statistical Machine Translation*. Natural Language Processing of Semitic Languages, pp. 199-219, Springer Berlin Heidelberg.
 - Hassan Sajjad, Kareem Darwish and Yonatan Belinkov. *Translating Dialectal Arabic to English*. In Proceedings of the Association for Computational Linguistics (ACL 2013), Sofia, Bulgaria, August 2013
 - Darwish K. *Arabizi Detection and Conversion to Arabic*. In Proceedings of the EMNLP 2014 Workshop on Arabic Natural Language Processing (ANLP) 2014 Oct (pp. 217-224).
6. شارك في أولى الأبحاث المعنية بالتصنيف الآلي للنصوص العربية وذلك فيما يخص التعرف على اللغة الهجومية، وخطاب الكراهية، المشاعر، والمواضيع، والتعرف على اللهجات:
- Hamdy Mubarak, Kareem Darwish, Walid Magdy, Tamer Elsayed, Hend Al-Khalifa. 2020. *Overview of OSACT4 Arabic Offensive Language Detection Shared Task*. In: OSACT 4: The 4th Workshop on Open-Source Arabic Corpora and Processing Tools (LREC-2020). Marseille, France, May 2020.
 - Hamdy Mubarak, Kareem Darwish. 2019. *Arabic Offensive Language Classification on Twitter*. International Conference on Social Informatics (SocInfo'2019). Doha, Qatar, November 2019.
 - Sabit Hassan, Shaden Shaar, Kareem Darwish. 2022. *Cross-Lingual Emotion Detection*. Proceedings of the Thirteenth Language Resources and Evaluation Conference.



16535 Grant Bishop Ln.
Los Gatos, CA 95032

- d. Fouzi Harrag, Maria Dabbah, Kareem Darwish, Ahmed Abdelali. 2020. *Bert Transformer model for Detecting Arabic GPT2 Auto-Generated Tweets*. Proceedings of the Fifth Arabic Natural Language Processing Workshop.
- e. Ahmed Abdelali, Hamdy Mubarak, Younes Samih, Sabit Hassan, Kareem Darwish. 2021. *QADI: Arabic dialect identification in the wild*. Proceedings of the Sixth Arabic Natural Language Processing Workshop, pp. 1-10.
- f. Shammur Absar Chowdhury, Ahmed Abdelali, Kareem Darwish, Jung Soon-Gyo, Joni Salminen, Bernard J Jansen. 2020. *Improving Arabic text categorization using transformer training diversification*. Proceedings of the Fifth Arabic Natural Language Processing Workshop.

- d. Fouzi Harrag, Maria Dabbah, Kareem Darwish, Ahmed Abdelali. 2020. *Bert Transformer model for Detecting Arabic GPT2 Auto-Generated Tweets*. Proceedings of the Fifth Arabic Natural Language Processing Workshop.
- e. Ahmed Abdelali, Hamdy Mubarak, Younes Samih, Sabit Hassan, Kareem Darwish. 2021. *QADI: Arabic dialect identification in the wild*. Proceedings of the Sixth Arabic Natural Language Processing Workshop, pp. 1-10.
- f. Shammur Absar Chowdhury, Ahmed Abdelali, Kareem Darwish, Jung Soon-Gyo, Joni Salminen, Bernard J Jansen. 2020. *Improving Arabic text categorization using transformer training diversification*. Proceedings of the Fifth Arabic Natural Language Processing Workshop.

We hope you accept our nomination for Dr. Kareem Darwish. His scientific and practical contributions were instrumental in advancing the state-of-the-art for Arabic NLP and contributed to the building of a vibrant research community in the Arab World.

نتمنى قبول ترشيحنا للدكتور كريم درويش. حيث إن مساهماته العلمية والعملية كان لها عظيم الأثر في تطوير تقنيات اللغة العربية، وساهمت في بناء مجتمع بحثي نابض في العالم العربي.

Sincerely,

Hassan Sawaf, CEO of aiXplain Inc.
Los Gatos, CA, USA



ولكم منا جزيل الشكر،

د. حسان صواف، المدير العام لشركة إكسبلين (aiXplain Inc.)
لوس جاتوس، كاليفورنيا، الولايات المتحدة الأمريكية