

# SAKHAR ALKHEREYF

www.sakhar.co ◊ sakhar@cs.columbia.edu

## EDUCATION

---

### Columbia University

PhD in Computer Science, Natural Language Processing

**Advisor:** Dr. Owen Rambow

**Thesis:** *Text Classification: Exploiting the Social Network.*

**Research interests:** Document Classification, Social Network Analysis, and Arabic Dialects NLP.

- MPhil conferred in February 2018.
- MS conferred in February 2016.

*Sep 2014 – Present*

*New York, NY*

### King Saud University

BS in Computer Science (with Honors).

*Sep 2006 – Jun 2010*

*Riyadh, Saudi Arabia*

## WORK EXPERIENCE

---

### Verisk Analytics

*Machine Learning Engineer*

*Jun 2017 - Dec 2017*

*Jersey City, NJ*

- Joint Development Environment (JDE) team.
- Developed machine learning models for automatic classification of insurance documents.

### KACST

*Research Assistant*

*Nov 2010 - Sep 2013*

*Riyadh, Saudi Arabia*

- Worked on projects in Natural Language Processing and High Performance Computing.
- Project manager of the SAFI project, a commercial real-time automatic text filtration system.
- Project manager of the SANAM supercomputer project (ranked second in the Green500 list, 2012).

### Advanced Electronics Company

*Associate Software Engineer*

*Sep 2010 - Nov 2010*

*Riyadh, Saudi Arabia*

- Hardware Development Team, Research & Development.
- Developed embedded systems for the smart electric meter ADDAD.

## TEACHING EXPERIENCE

---

### Columbia University

*Teaching Assistant*

*Sep 2015 - Jun 2016*

*New York, NY*

- Spring 2016, Natural Language Processing (COMS 4705). Instructor: Prof. Dragomir Radev.
- Fall 2015, Natural Language Processing (COMS 4705). Instructor: Prof. Dragomir Radev.

## PUBLICATIONS

---

- **Sakhar Alkhereyf** and Owen Rambow. Email classification incorporating social networks and thread structure. In *LREC 2020 Proceedings: The International Conference on Language Resources and Evaluation*. European Language Resources Association, 2020
- Faisal Alshargi, Shahd Dibas, **Sakhar Alkhereyf**, Reem Faraj, Basmah Abdulkareem, Sane Yagi, Ouafaa Kacha, Nizar Habash, and Owen Rambow. Morphologically annotated corpora for seven Arabic dialects: Taizi, Sanaani, Najdi, Jordanian, Syrian, Iraqi and Moroccan. In *Proceedings of the Fourth Arabic Natural Language Processing Workshop*, pages 137–147, 2019

- Nizar Habash, Fadhil Eryani, Salam Khalifa, Owen Rambow, Dana Abdulrahim, Alexander Erdmann, Reem Faraj, Wajdi Zaghrouani, Houada Bouamor, Nasser Zalmout, Sara Hassan, Faisal Al-Shargi, **Sakhar Alkhereyf**, et al. Unified guidelines and resources for Arabic dialect orthography. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, 2018
- **Sakhar Alkhereyf** and Owen Rambow. Work hard, play hard: Email classification on the Avocado and Enron corpora. In *Proceedings of TextGraphs-11: the Workshop on Graph-based Methods for Natural Language Processing*, pages 57–65, 2017
- Abdulqadir A Alaqeeli, Ibrahim Almosallam, **Sakhar Alkhereyf**, and Suliman Alsowelim. Automated text-evaluation of user generated text, August 1 2017. US Patent 9,720,901
- David Rohr, Sebastian Kalcher, Matthias Bach, Abdulqadir A Alaqeeliy, Hani M Alzaid, Dominic Eschweiler, Volker Lindenstruth, **Sakhar Alkhereyf**, Ahmad Alharthiy, Abdulelah Almubarak, et al. An energy-efficient multi-gpu supercomputer. In *2014 IEEE Intl Conf on High Performance Computing and Communications, 2014 IEEE 6th Intl Symp on Cyberspace Safety and Security, 2014 IEEE 11th Intl Conf on Embedded Software and Syst (HPCC, CSS, ICESS)*, pages 42–45. IEEE, 2014

---

## PROFESSIONAL ACTIVITIES

### Paper Reviews

- 12<sup>th</sup> edition of the Language Resources and Evaluation Conference (LREC 2020).

### Conference Organizing

*Registration Committee*

- KACST International Information Technology Conferece (KACSTIT 2012).

### Research Advising

- Advised summer interns at KACST: Suliman Alsowelim (2011), and Azzam Alsudais (2011).

---

## TECHNICAL SKILLS

<b>Programming Languages</b>	Python, Java, C/C++
<b>Libraries</b>	PyTorch, Keras, scikit-learn, Stanford CoreNLP, pandas
<b>Software</b>	L <sup>A</sup> T <sub>E</sub> X, Matlab, Gephi