

Tarfah Alrashed

tarfah@mit.edu
www.tarfahalrashed.com

CSAIL, Massachusetts Institute of Technology
32 Vassar St. 32G-598 Cambridge, MA, USA 02139

RESEARCH INTEREST

Human Computer Interaction, Information Retrieval, Data Science, and Semantic Web.

EDUCATION

- 2016 - CURRENT PhD. Computer Science
Massachusetts Institute of Technology
Computer Science and Electrical Engineering Department, Cambridge, MA
Advisor: David Karger
- 2016 - 2019 M.S. Computer Science
Massachusetts Institute of Technology
Computer Science and Electrical Engineering Department, Cambridge, MA
Advisor: David Karger
- 2012 - 2014 M.S. Computer Science
University of California San Diego
Computer Science and Engineering Department, San Diego, CA
- 2004 - 2008 B.S. Computer and Information Sciences
King Saud University
College of Computer and Information Sciences, Riyadh, Saudi Arabia

WORK EXPERIENCE

- Sep 2016 – Current Graduate Research Assistant
Massachusetts Institute of Technology, CSAIL, Cambridge, MA
Advisor: David Karger
- May 2020 – Aug 2020 Research Intern
Google Research, Mountain View, CA, USA
Advisor: Natasha Noy
- June 2018 – Aug 2018 Research Intern
Microsoft, Bellevue, WA, USA
Advisor: Peter Bailey
- June 2017 – Aug 2017 Research Intern
Microsoft Research, Redmond, WA, USA
Advisor: Susan Dumais

- Oct 2014 – Aug 2016 Research Affiliate
Massachusetts Institute of Technology, Cambridge, MA, USA
- Oct 2014 – Aug 2016 Research Associate
Center for Complex Engineering Systems, KACST, Riyadh, Saudi Arabia
- June 2012 – May 2014 Software Engineer/Intern
University of California San Diego, CALIT, San Diego, CA, USA
- July 2009 – Nov 2011 Research Specialist
National Center for Computation Technology and Applied Mathematics, KACST
Riyadh, Saudi Arabia

RESEARCH PROJECTS

- May 2020–Aug 2020 **Dataset or Not? Automatic Detection of Dataset Web Pages**
Google There are millions of datasets published on the Web. Search engines, such as Google’s Dataset Search, index these pages and provide a search interface over them. These search engines often rely on schema.org metadata to determine which pages have datasets and to find the dataset metadata. However, relying solely on dataset metadata poses two key challenges: First, many pages that have dataset markup are not actually about datasets, in which some websites would use the wrong metadata type to describe the content of their pages. Second, many of the pages that are describing datasets do not have the markup. Thus, in order to guarantee both quality and coverage, we need to develop a classifier that would identify dataset pages. Because datasets on the Web can be part of any page and can cover any topic imaginable, a classifier that relies on specific features of a vertical (real estate listings, events, etc.) or a function (course page, faculty page, etc.) may not be very accurate in this domain. In this paper, we develop a classifier that outperforms state-of-the-art Web classifiers by at least 20%.
- Jan 2019–Aug 2019 **ScrAPIr: Making Web Data APIs Accessible to End Users**
CSAIL, MIT Users have long struggled to extract and repurpose data from websites by laboriously copying or scraping content from web pages. An alternative is to write scripts that pull data through APIs. This provides a cleaner way to access data than scraping; however, APIs are effortful for programmers and nigh impossible for non-programmers to use. In my research, we empower users to access APIs without programming. We evolve a schema for declaratively specifying how to interact with a data API. We then develop ScrAPIr, a standard query GUI that enables users to fetch data through any API for which a specification exists, and a second GUI that enables users to author the specification for a given API. From a lab evaluation, we find that even non-programmers can access APIs using ScrAPIr, while programmers can access APIs 3.8 times faster on average using ScrAPIr than using programming.
- June 2018 – Aug 2018 **Evaluating User Actions as a Proxy for Email Significance**
Microsoft The number of emails people receive every day can be overwhelming. Having a good estimate of the significance of emails forms the foundation for many downstream tasks (e.g.

email prioritization); but determining significance at scale is expensive and challenging. In this work, we hypothesize that the cumulative set of actions on any individual email can be considered as a proxy for the perceived significance of that email. We propose two approaches to summarize observed actions on emails, which we then evaluate against the perceived significance. First approach is a fixed-form utility function parameterized on a set of weights. Second, we build machine-learning models to capture users' significance directly based on the observed actions. Our analysis suggests that there is a positive correlation between actions and significance of emails and that actions performed on personal and work emails are different. We also find that the degree of correlation varies across people, which may reflect the individualized nature of email activity patterns or significance.

- June 2017 – Aug 2017
Microsoft Research
- The Lifetime of Email Messages: A Large-Scale Analysis of Email Revisitation**
Communication, leading to a number of challenges related to information overload and email management. To better understand email management practices in detail, we examined the distribution of visits to emails over time. During their lifetime, emails may be visited once or several times, and with each visit different actions may be taken. Emails that are revisited over time are especially interesting because they represent an opportunity to improve email management and search. We examine a large-scale log analysis of email revisitation, the activities that people perform on revisited email messages, and the strategies they use to go back to these emails. Most emails have a short lifetime, with more than 33% having a lifetime of less than 5 minutes. Our findings have implications for designing email clients and intelligent agents that support both short- and long-term revisitation patterns.
- Oct 2014 – June 2016
CCES – MIT
- City Schema – Collaborative Tangible Interface**
Developed simulations for the complex systems of city infrastructures as system of systems to simulate the behaviors of these systems and their interdependencies as well as assist stakeholders in predicting future scenarios. Designed and built a collaborative tangible user interface (TUI), which provides multi-touch interactive capabilities with analytical and visualization components, as a decision support system to support collaborative city planning. Studied the usability patterns in TUI systems and their associations with system factors and group dynamics, by evaluating the different components of TUI that could correlate with user choices and affect how they use these systems. Conducted observational studies to analyze users interactions with TUI in the context of urban planning, and its impact on communication and decision-making.
- 2013 – 2014
Calit2/UCSD
- A Crowd sourced Approach for Wait Time Estimation**
Developed an iOS app “Best Time to Cross the Border”, which provides commuters with wait time estimations at the US land borders to help them plan their trips ahead. Adopted Crowdsourcing approach in our iReport feature, which leverages the users as sensors by empowering them to report the border delays they experience. Validated the data submitted by the users by restricting the posting of the wait time reports to three miles radius from the border. Calculated the average wait time estimations for the last three months, using the Customs and Border Protection (CBP) wait time estimations for the US land borders along with the wait times reported by our users. Provided historical wait time trends in graphical form.

PUBLICATIONS

- T. Alrashed**, J. Almahmoud, A. X. Zhang, D. R. Karger. “*ScrAPIr: Making Web Data APIs Accessible to End Users*”. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20). Association for Computing Machinery, New York, NY, USA, 1-12.
- T. Alrashed**, C.J. Lee, P. Bailey, C Lin, M. Shokouhi & S. Dumais. “*Evaluating User Actions as a Proxy for Email Significance*”. In The World Wide Web Conference (WWW'19), pp. 26-36. San Francisco, USA, 2019.
- L. Verou, **T. Alrashed**, & D. Karger. “*Extending a Reactive Expression Language with Data Update Actions for End-User Application Authoring*.” The 31st Annual ACM Symposium on User Interface Software and Technology (UIST'18). Berlin, Germany, 2018.
- L. H. Gilpin, D. M. Olson, and **T. Alrashed**. “*Perception of Speaker Personality Traits Using Speech Signals*.” Extended Abstracts of the Conference on Human Factors in Computing Systems (CHI'18). Montreal Canada, 2018.
- T. Alrashed**, A. H. Awadallah, & S. Dumais. “*The Lifetime of Email Messages: A Large-Scale Analysis of Email Revisitation*”. In Proceedings of the 2018 Conference on Human Information Interaction & Retrieval (CHIIR '18). New York, NY, USA, 120-129, 2018.
- T. Alrashed**, J. Almahmoud, M. Alrashed, S. Alsubaiee, M. Alsaleh, & C. S. Olascoaga. “*Social Communities in Urban Mobility Systems*.” In International Conference on Social Computing and Social Media, pp. 177-187. Springer International Publishing, 2016.
- J. Almahmoud, A. Almalki, **T. Alrashed** & A. Alwabil. “*Prototyping Complex Systems: A Diary Study Approach to Understand the Design Process*.” In International Conference of Design, User Experience, and Usability, pp. 187-196. Springer International Publishing, 2016.
- T. Alrashed**, A. Almalki, S. Aldawood, T. Alhindi, I. Winder, A. Noyman, A. Alfariis & A. Alwabil. “*An Observational Study of Usability in Collaborative Tangible Interfaces for Complex Planning Systems*.” AHFE 2015 Proceedings, Volume 3, pp. 1-6614, 2015.
- T. Alrashed**, A. Almalki, S. Aldawood, A. Alfariis & A. Al-Wabil. “*Coding Schemes for Observational Studies of Usability in Collaborative Tangible User Interfaces*.” In HCI International 2015-Posters' Extended Abstracts, pp. 3-6. Springer International Publishing, 2015.
- S. Aldawood, F. Aleissa, A. Almalki, **T. Alrashed**, T. Alhindi, R. Alnasser, M. Hadhrawi, A. Alfariis & A. Al-Wabil. “*Collaborative Tangible Interface (CoTI) for Complex Decision Support Systems*.” In Design, User Experience, and Usability: Users and Interactions, pp. 415-424. Springer International Publishing, 2015.

SELECTED PRESS

- May 2020 **Meet ScrAPIr, MIT's Swiss army-knife for non-coders to shake data out of APIs (It's useful for pro devs, too),**
The Register, https://www.theregister.com/2020/05/21/mit_scrapir_api/
- May 2020 **New From MIT/CSAIL: "Searching Websites the Way You Want"; Introducing scrAPIr,**
InfoDocket News, <https://www.infodocket.com/2020/05/18/new-from-mit-csail-searching-websites-the-way-you-want-introdcing-scrapir/>
- May 2020 **Searching websites the way you want,**
MIT CSAIL News, <https://www.csail.mit.edu/news/searching-websites-way-you-want>
- March 2013 **Smartphone app speeds up border crossings,**
Los Angeles Times, <http://articles.latimes.com/2013/mar/26/local/la-me-abcarian-border-20130327>
- March 2013 **Calit2 Border App Wins Third Place at Mobile World Congress,**
UC San Diego News Center,
http://ucsdnews.ucsd.edu/pressrelease/calit2_border_app_wins_third_place
- November 2012 **New Mobile App Tells Users When to Cross the Border,**
The Guardian UC San Diego, <http://ucsdguardian.org/2012/11/19/new-mobile-app-tells-users-when-to-cross-the-border/>
- November 2012 **iPhone app lets border crossers determine best time to cross U.S. border,**
Homeland Security News Wire,
<http://www.homelandsecuritynewswire.com/dr20121109-iphone-app-lets-border-crossers-determine-best-time-to-cross-u-s-border>
- November 2012 **Students Build App To Determine The Best Time To Cross Border,**
KPBS, <http://www.kpbs.org/news/2012/nov/12/students-build-best-border-crossing-times-app/>
- November 2012 **Crowdsourcing Feature Lets iPhone Users Determine Best Time to Cross U.S. Border,**
UC San Diego News Center,
http://ucsdnews.ucsd.edu/pressrelease/best_time_to_cross_us_b

AWARDS

- 2016 Graduate PhD Scholarship, Center for Complex Engineering Systems, King Abdulaziz City for Science and Technology (KACST).
- 2014 Letter of Recognition, Von Liebig Entrepreneurism Center Graduate Curriculum, UCSD.

2013 Second place, Mobile World Congress (MWC), in Barcelona, Spain.

2011 Graduate Masters Scholarship, King Abdulaziz City for Science and Technology (KACST).

COMPUTER SKILLS

Programming Languages: C, C++, C#, Objective C, Processing, and Python.

Web Programming: HTML, CSS, JavaScript, D3, Leaflet, and SQL.