

## Newsletter

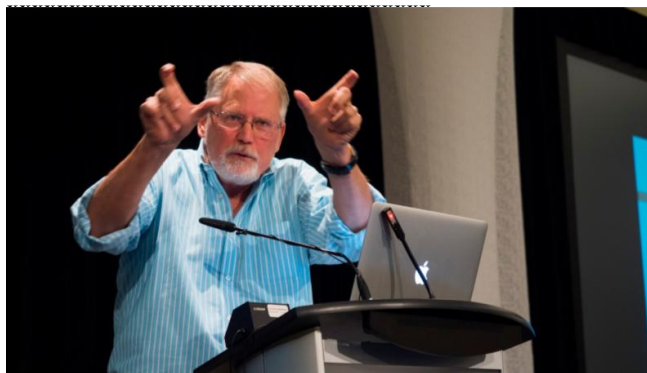
Autumn 2015

### Big Data for Productivity Congress

In October Halifax hosted a large international meeting with nearly 500 delegates from 12 countries focusing on the role that Big Data technologies can play in the constant search for productivity improvements in the government and private sectors. It was a unique event, combining the forces and the interests of two different meetings and taking advantage of Halifax's growing prominence in the field of Big Data Analytics.

For the past couple of years a Big Data Congress had been held annually in Saint John, bringing together both industry and academics and Halifax had been chosen as the site for 2015. However, when Mike Shepherd (Dalhousie) was approached by the World Confederation of Productivity Science (WCPS) to also host the World Productivity Congress in Halifax in 2015 it only made sense to bring the two meetings together. "Innovation drives productivity", says Mike and together with his co-chair Peter Watkins from WCPS they put together an innovative and productive program that has had lots of positive feedback from the participants.

As well as a diverse range of talks, presentations and networking opportunities the program included an NSERC funded Springboard connector event, organized by Evangelos Milios (Dalhousie) bringing industry people together with academic researchers to explore the possibilities of research collaboration. And there was a day of activities for high school students, mentored by volunteers from industry, including the participation of the delegates from both Google and Boeing. The Nova Scotia Dept of Education bussed in 640 students



Dan Russell from Google

from around the province to take part. The final day of the Congress included a technical track that was organized by Stan Matwin (Dalhousie Institute for Big Data Analytics) and Danny Silver (Acadia Institute for Data Analytics).

Still tidying up the loose ends and paperwork Mike Shepherd describes the Congress as an important step along the path for the Atlantic Region in establishing itself as a national centre of excellence in this field. With the Institute for Big Data Analytics at Dalhousie, and the recent decisions of big industry players such as IBM, CGI and Ernst & Young to set up operations here there is a momentum that is growing. Plans are now underway for next year's Big Data Congress 4.0 to be held here in Halifax.

### At the Digital Humanities Conference



Dr. Rob Warren represented the Institute for Big Data Analytics at this summer's Digital Humanities Conference in Australia, presenting his paper on his work with data from the First World War. His paper explores the notion of place, feature and geometry in the context of the Great War using Linked Open Data. In previous works, the translation of obsolete military coordinates through API's (Application Program Interface)

was covered. He reviewed their use as an efficient and effective means of indexing archival documents about the war. Most war diaries, operations orders and dispatches in British and Dominions records refer to locations using both named features and coordinates. This permits the geo-referencing of each statements within a document to find the current location in question while segmenting the document according to different spatial component.

# News

## An Interview with Dr. Ken Barker - researcher with IBM's Watson Group



Ken Barker is a researcher at IBM at the Thomas J. Watson Research Center, Yorktown Heights, New York. His current research attempts to leverage structured knowledge and reasoning both to guide

Watson's natural language understanding and bridge conceptual gaps in its resulting interpretations. Prior to joining IBM in 2011, he was a Research Scientist at the University of Texas at Austin, where he was an Investigator on DARPA's Rapid Knowledge Formation and Machine Reading Projects, as well as on Vulcan's Project Halo to build intelligent scientific textbooks. Previously he was an Assistant Professor of Computer Science at the University of Ottawa. His research there focused on Natural Language Semantics and Semi-Automatic Interpretation of English Text. Ken Barker delivered a seminar at Dalhousie in October on the subject of "Knowledge Acquisition from Knowledge -- Digging Deeper to Find Signal in Structured Sources". We caught up with Ken afterwards to ask him a few questions:

**Q.** How did you get into this field of research?

**A.** I had always been interested in languages, having taken Latin in high school. Seeing the commonality and regularity in languages sparked an interest in linguistics. And then at university I studied computer science but also took courses in linguistics on the side. My PhD studies were in the area of computer linguistics and then in Texas I researched knowledge representation and reasoning. Now at IBM I am trying to bring the two together. The fields of Natural Language Processing (NLP) and Knowledge used to be quite close together at one time, and then a rift occurred which sent each field off in its own direction. NLP researchers abandoned knowledge as it was seen to be too brittle. But NLP has hit a ceiling, presenting an interesting opportunity to bring these fields back together again.

**Q.** How is it different to work for IBM, compared to a university?

**A.** I have spent a lot of time, 15 years, working in the university environment and there is a lot to be liked about it. However, a big attraction of IBM was the team of people that I would get to work with; the scale and the focus of the research is on a whole other level. We have 20 PhDs working on the same problem at the same time.

**Q.** What accomplishments are you most proud of?

**A.** After experiencing the divide between the Statistical Natural Language Processing community and the Knowledge community for many years, I get satisfaction from having been able to convince my NLP colleagues that there is great value in being able to leverage Knowledge resources for language tasks; and I get equal satisfaction from having been able to convince

and I get equal satisfaction from having been able to convince my Knowledge colleagues that there is power in using statistics and learning to exploit Knowledge more flexibly.

**Q.** What concerns you most in your field?

**A.** One of my greatest concerns is that people tend to see particular solutions as cure-alls; this goes in trends. It seems more likely that progress will be dependent not on a particular technique, but on being creative in applying different techniques. Without that creativity, whole communities can get stuck in a rut.

**Q.** How did things change after the big Jeopardy success?

**A.** Although I wasn't working directly on the Jeopardy project, I was a consultant for IBM leading up to that event, and was also working with the Jeopardy team on a different project. It certainly made a big splash in the public consciousness at the time and the most positive result was that researchers at IBM were then given a very long leash. However, the public is already beginning to forget about it.

**Q.** Do you have any advice for students looking for a career in your field?

**A.** I would advise that students should remember to stay open to applying techniques and ideas from a variety of sources. Don't get stuck in an arms race on a particular metric. There's a lot more to be found in the space of creative solutions.

## Big Data and Flooding in Nova Scotia

To live in Nova Scotia is to accept that water is a significant element in your daily environment. Our beautiful coastline and scenic lakes are often a big attraction to visitors and although we grumble we even get used to the rain. However, when the water levels rise to the point of flooding it's much more serious, potentially causing large amounts of damage to private property and infrastructure. Good data about flooding is thus hugely important and when Nova Scotia Environment approached Dalhousie about the problem it was quickly taken up by Rob Warren from the Institute for Big Data Analytics and Mike Smit from the School of Information Management. Rob and Mike teamed up with mathematics graduate Amelia Yzaguirre, and with access to newspaper data provided by the Chronicle Herald were able to create algorithms to draw out more than 14,000 flood occurrences since 1992 and plot them on a map. This information can be supplied to actuaries, municipal engineers and planners to enable better decisions going forward. More data added from past years, from other sources, or from future reports of floods will only make the tool more valuable.



# News

## Social Media in the Middle East

### Eman Alyami: Student Research Profile



The increasing interpenetration of computer science into the widest range of research topics across the university has meant that an interdisciplinary approach has become increasingly common amongst our students. Many research questions these days cannot be answered from only one perspective. One such student is Eman Alyami, working towards an interdisciplinary PhD researching how women in the Middle East are using social media and how it is changing their lives. This broad topic draws on sociology, management studies and computer science, where her supervisors are Dr. Stan Matwin, Dr. Carolyn Watters, Dr. Chris Helland and Dr. Dawn Jutla.

In the beginning the interdisciplinary approach created a confusing situation, with different libraries for each different discipline; requiring her to create her own library drawing from all three sources. Her research began with a look at women's issues, such as power, politics and social life, but also necessitated a study of the background of this diverse geographical area. The computer science aspect of her research concerns the huge amounts of data, Twitter traffic in this case, which must be tracked and processed. The volume of data is so large that keeping up with it is a real problem. Algorithms to deal with large volumes of text are some of the big data tools that are required. There are also language problems with differences between modern standard Arabic and more informal spoken variations. Help from colleagues is also necessary to confirm Arabic/English translations; and, Twitter traffic also does not always explicitly identify the gender of the writer and so methods to identify or infer gender have to be adopted or created.

Eman came from Saudi Arabia in 2010 to do her Masters in Electronic Commerce. When she completed her degree and was looking to doctoral studies Dr. Vlado Keselj pointed her in the direction of the interdisciplinary program. She started her PhD in 2014. Her great ambition after she graduates is to teach - she would like to go back to Saudi Arabia and become a professor. In Canada she has enjoyed the open and multicultural society, as well as the abundant natural environment. Even the winters have not been too bad.

## MEOPAR / exactEarth AIS Workshop

In October 2014, the Marine Environmental Observation Prediction and Response Network (MEOPAR), based at Dalhousie, and exactEarth Ltd, the private provider of international AIS data, entered into a mutually beneficial partnership in order to facilitate workshops on best practices with AIS data management, to promote the use of AIS data for MEOPAR researchers, and to build an internship program between the two organizations. The first workshop co-hosted through this partnership along with the Institute for Big Data Analytics, "AIS: Applications, Analysis and Data Management Techniques" was held in Halifax at the Hollis Hotel in September.

MEOPAR research projects span a wide range of topics, several of which depend on accurate and timely shipping data. Vessel impacts on the environment can include noise, air pollution, accidental spills, invasive species transfer through ballast water, collisions with mammals, interactions with other vessels, and other operational ramifications. Investigations to help improve policies and practices to mitigate such risks depend fundamentally on information about ship patterns, types, densities, trends and anomalies, enabled through the Automatic Identification System. Thus, access to shipping data will produce significant advances for the understanding and management of such issues. Furthermore, research on how best to access, visualize, clean, and query the data, and produce derived products such as shipping tracks, were also key areas of focus in this workshop. This event addressed cutting-edge research in all of these areas by bringing together industry, government and academia. According to Dr. Stan Matwin,

"AIS data is certainly Big Data in every sense of this term, and as such is a truly exciting challenge for the Institute for Big Data Analytics. We want our Institute to become a leader in AIS data analytics, and that can only happen in partnership with many other experts from this community."

Workshop speakers included a range of experts from academia, industry and the government and security sectors.

## TRIBE: In its Second Year

Thanks to a generous six-year CREATE grant from NSERC and our partners at Simon Fraser and l'Université de Montréal, the graduate training program in big text Data (TRIBE) is entering a second year with a full complement of engaged and industrious students and a busy year of activities ahead. As well as continuing their studies in their subject area the students this year will also be participating in industrial internships in a diverse range of organizations across the country, attending online seminars with their counterparts at our partner universities and taking training in the softer skills such as ethics and entrepreneurship. We also hope, this year, to get our dedicated TRIBE repository set up to hold the datasets that we are working with.



## Award for the "Most Innovative Use of Big Data"

A team of 3 students from the Institute for Big Data Analytics have walked away with the prize of the "Most Innovative Use of Big Data", and a cheque for \$800 from the Sports Hack 2015 competition, held Nov 27-29th in Vancouver, Toronto and Halifax and sponsored by a range of big name tech companies, universities and the CFL. Forty-one teams competed in the challenge to produce an app to encourage fan engagement using datasets from the CFL. Our team created an app that tracked the Tweets of CFL fans providing them with rankings of their activity levels and a points system to reward their activity and their participation in mini-games about the football matches, making use of both machine learning techniques and sentiment analysis. The app could also provide CFL organizations with information about their fans activities and locations and provide the opportunity to advertize to fans, to convert fan points into merchandise of the organizations or their sponsors. It would be fan-centric rather than team-centric. The contest was a big challenge, given the hard work, the lack of sleep, and the fact that they lost one team member at the very end - the member who was central to putting together their final presentation.



Gurcan Gercek, Parsa Sarshar and Behrouz Haji Soleimani Mamaghani from the Institute for Big Data Analytics

## First Prize in the Smart Energy Apps Challenge



Stephen Duff (CEO Innovacorp) presents prize to Parsa Sarshar, Pedram Adibi and Mehran Zamani

Congratulation to Masters student Parsa Sarshar and Research Assistant Pedram Adibi from the Institute for Big Data Analytics, and their colleague, Dalhousie alumnus Mehran Zamani, who, as Team Sol-Ops, won first prize in the Smart Energy Apps Challenge sponsored by Innovacorp, HRM, NSCC, Shiftkey Labs and Dalhousie University. The challenge was to create an app which made use of the data collected by Halifax Solar City on a number of parameters relating to installed solar hot water systems. Team Sol-Ops created an app which made use of techniques of computer science, data science and engineering to provide customers with an analysis of their usage of hot water, and recommendations for changing that behaviour to maximize the environmental and economic benefits they could receive from their systems. It took 6 weeks to create the winning

app, beating 10 other teams and claiming the \$6000 first prize. Success in the competition may also be the stepping stone to other opportunities as the team finds themselves in conversations with investors and other organizations who are taking in interest in the commercial potential of their creation.

### Director

**Dr. Stan Matwin PhD, CRC**

### Executive Committee

**Dr. Michael Bliemel**

*Faculty of Management (representing the Dean of Management)*

**Dr. Tom Marrie**

*Dean of Medicine*

**Dr. Evangelos Milios**

*Associate Dean, Research, Faculty of Computer Science*

**Dr. Andrew Rau-Chaplin**

*Faculty of Computer Science*

**Dr. Nur Zincir-Heywood**

*Faculty of Computer Science*

**Dr. Stan Matwin**

*Director, (ex officio), Faculty of Computer Science*

### Advisory Board

**Sue Abu-Hakima Amika**

*President and CEO, AmikaMobile, Ottawa, ON*

**Martin Davis**

*Vice-President of IT, JD Irving, Moncton, NB*

**Peter Hickey**

*Co-founder, Oris4, Halifax, NS*

**David Kasik**

*Senior Technical Fellow, The Boeing Company, Seattle, Washington*

**Stephen Perelgut**

*IBM Canada University Relations Manager, Markham, ON*

*Newsletter Editor:*

*David Langstroth  
dll@cs.dal.ca*

*"Data matures like wine, applications like fish.."*

*- James Governer, Founder of RedMonk*