Project details:

Applications are now invited from DCU PhD students for the following projects (two internships will be offered) from the following two locations:

**Location – Rio de Janeiro**

**Proponent: Paulo Cavalin**

**Project**: Social Media Analytics

**Period**: Summer 2014 (10 weeks) **Place**: IBM Research - Brazil, Rio de Janeiro
**Major**: Visualization or Infoviz

**Skills**: Data and information visualizations

**Description**: Intern will propose and develop visualization interfaces for Social Media Analytics (SMA). Activities will include understanding major projects of the lab related to SMA, and research, design and development of visualization interfaces that can leverage the way end-users can extract relevant information from SMA. Programming skills and interface design needed.

**Proponent: Bruno Flach:**

**Project 1**: Optimal Resource Allocation and Scheduling **Period**: Summer 2014 (10 weeks) **Place**: IBM Research - Brazil, Rio de Janeiro
**Major**: Mathematics, Computer Science or *Operations Research*-related areas  **Skills**: mathematical programming (optimization under uncertainty, integer programming) **Description**: Intern will develop, implement and test models and algorithms for optimization problems that arise in the context of construction and operations of complex projects. In particular, we are interested in investigating optimal strategies for resource allocation and scheduling in uncertain environments by using / combining techniques from robust optimization and stochastic programming.

**Project 2**: Optimization in Oil & Gas Exploration and Production **Period**: Summer 2014 (10 weeks) **Place**: IBM Research - Brazil, Rio de Janeiro **Major**: Mathematics, Computer Science or *Operations Research*-related areas  **Skills**: simulation-based optimization, optimization under uncertainty **Description**: Intern will develop, implement and test models and algorithms for optimization problems in the area of oil & gas exploration and production. An important aspect of this project is the incorporation of relevant sources of uncertainties which result from an incomplete knowledge of geological and petro-physical characteristics of reservoirs. Matlab programming skills are a plus.

**Proponent: Bianca Zadrozny**

**Project 1**: Deep Learning for Natural Language Processing **Period**: summer 2014 (12 weeks) **Place**: IBM Research - Brazil, Rio de Janeiro **Major**: Computer Science or Computer Engineering **Skills**: computer programming, machine learning, natural language processing (NLP) **Description**: Intern will develop, code, and test algorithms and software for: (1) representation learning for NLP; and (2) training deep neural networks to solve NLP tasks. The work developed by the intern will be part of a large study on the use of Deep Learning to minimize the use of labeled data and feature engineering to solve complex NLP tasks such as Question & Answering. Strong programming skills and knowledge of machine learning are needed; knowledge of NLP is a plus.

**Project 2**: Linking Open Urban Data
**Period**: summer 2014 (10 weeks)
**Place**: IBM Research - Brazil, Rio de Janeiro
**Major**: Computer Science
**Skills**: computer programming
**Description**: Open Urban data is a relatively recent phenomenon and, as a result, very little work has been done in integrating these datasets. More importantly, these datasets are usually disconnected and very heterogeneous which makes it very hard for the user to get insight and useful information from the data. For example, an important question that arises is what type of useful queries the user can do with a defined dataset. In this project, the intern will develop new algorithms and tools to organize and integrate Open Urban datasets. Strong programming skills are required.

**Project 3**: Effectively Managing, Assembling and Querying Big Trajectory Data
**Period**: summer 2014 (10 weeks)
**Place**: IBM Research - Brazil, Rio de Janeiro
**Major**: Computer Science
**Skills**: computer programming
**Description**: The immense growth of mobile and tracking devices has lead to a vast amount of data in the form of trajectories. This immense amount of ''raw'' generated data has lead to several open challenges, including techniques to efficiently store, manage and visualize trajectory datasets. One very interesting problem is to find the entire possible route of a trajectory given its partial results (e.g., users from the Foursquare social website can disclose their locations using check in/out, but the entire route is not known), and also the other trajectory's locations (e.g., where are the frequent routes for that time frame). In this project, the intern will develop new algorithms to manage and query large amounts of trajectory datasets. The proposed algorithms will then be compared with existing ones. Strong programming skills are required.

**Proponent: Renato Cerqueira**

**Project:** Automatic Synthesis of Systems Connectors
**Period**: summer 2014 (10 weeks)
**Place**: IBM Research - Brazil, Rio de Janeiro
**Major**: Computer Science
**Skills**: computer programming, formal specifications, software engineering
**Description**: Intern will perform experiments to evaluate current approaches, such as the semantic-driven approach adopted in the Connect Project (https://www.connect-forever.eu), to synthetize on the fly connectors to enable heterogeneous systems to communicate. To perform this evaluation, a set of use cases will be defined and implemented. Based on this initial evaluation, further improvements to the selected approaches will be designed and implemented, such as better support for exception mapping and partial interface equivalence.

**Proponent: Alexandre Rademaker**

**Title:** WordNet Collaborative Editor

**Period**: summer 2014 (10 weeks)
**Place**: IBM Research - Brazil, Rio de Janeiro
**Description:** One of the basic requirements to process natural language is to analyze the semantic relationships between words. Computational systems do not have this information natively, they do not know, for instance, that Portuguese words such as "carro" and "automóvel" both refer to cars, or that "caminhão" (truck) is also related to these words as a more general hypernym. Lexical databases, such as WordNet provide these relationships, and therefore, are extremely valuable linguistic resources.OpenWordnet-PT (or OpenWN-PT is a lexical-semantic resource describing (Brazilian) Portuguese words and their relationships, modeled after WordNet 3.0. We would like to continue improving OpenWN-PT and other lexical resources for Portuguese. However, manually construction and maintenance of a lexical resource is quite expensive and time consuming. One approach is to tap into the ‘crowd.’ The main obstacle for using crowdsourcing to maintain OpenWN-PT is the lack of a good users interfaces, target for regular users, not a lexicographers. Existing interfaces for editing WordNet do not satisfy some of the most important features that we identify as necessary for the collaborative work: (1) version control; (2) game like tasks for motivations; (3) tracking of changes; and (4) target to mobile devices. Thus, we a looking for interns to help us in the developing of a novel web and mobile interface for guide the maintenance and improvement of openWN-PT and other Portuguese lexical resources.

**Skills**

- Experience with web developing
- Knowing JavaScript and libraries like Dojo, [http://backbonejs.org](http://backbonejs.org/), or [http://angularjs.org](http://angularjs.org/).
- Some experience with database, RDF (desired but not necessary).
- Interest in User Interface developing

**Location – São Paulo**

**Proponent: Rogério de Paula**

**Project Title 1**: Team Formation Project

**Period**: summer 2014 (10 weeks)
**Place**: IBM Research - Brazil, São Paulo
**Major**: Computer Science, Informatics, Design

**Desired skills:** information visualization tools (e.g. D3)

**Description**: Team formation project aims at investigating new algorithms and techniques based on social analytics and optimization to solve the complex problem of forming effective teams to tackle particular problems. For this internship, we aim at investigating and devising a new interactive interface that allows users to manage (visualize, select, and the like) different types of team-formation scenarios. The idea is to help users visualize and manipulate the various variables/attributes that affect the formation of different kinds of teams and in so doing being able identify the best teams to address a particular problem at hand.

**Project Title 2**: Sociometric Security

**Period**: summer 2014 (10 weeks)
**Place**: IBM Research - Brazil, São Paulo
**Major**: Human Computer Interaction, Computer Science, Informatics, Design

**Desired skills:** mobile application development for iOS and/or Android, information visualization tools (e.g. D3)

**Description**: Sensor technologies embedded in mobile devices have allowed the development of an increasingly larger number of context-aware solutions, such as, context-aware recommender-systems and context-aware security. This research will investigate how social-analytics (or data-mining people's everyday activities by means of mobile-sensor technologies) can reveal important information about people's safety, privacy, and security contexts. In this internship, we aims at investigating how to implement a mobile application that takes into account various sources of activity metrics to support better security, privacy, and safety issues. The intern will be thus asked to review and map out existing solutions and design and develop a prototypical solutions that attempts to address some facet of this project.

**Proponent: Alécio Binotto**

**Project Title**: Dynamic Load-balancing over Heterogeneous Platforms

**Period**: summer 2014 (10 weeks)
**Place**: IBM Research - Brazil, São Paulo
**Major**: Computer Science, Computational Engineering, or Mechanical Engineering with applied informatics knowledge

**Desired skills:** C/C++; optional other programming environments (MatLab, Fortran); CUDA (GPU); OpenMP/MPI

**Description**: Distributing the workload upon all available Processing Units (PUs) of a high-performance heterogeneous platform (e.g., PCs composed by CPU, GPU, Xeon Phi) is a challenging task, since the execution cost of a task on distinct PUs is non-deterministic and affected by parameters not known a priori. This problem was previously addressed by a context-aware runtime and tuning system based on a compromise between reducing the execution time of specific tasks and the cost of tasks' scheduling on a CPU-GPU platform. Based on this introduction, this summer project benefits from such previous development and intends to extend the approach in order to insert the Xeon Phi PU on the heterogeneous platform. For test purposes, a simple task can be used as a benchmark, like matrix-vector multiplication (Ax=b). The intern will have all the support needed for performing the experiment.

**Proponent: Leonardo Tizzei**

**Project:** Model-driven Software Product Line engineering
**Period**: summer 2014 (10 weeks)
**Place**: IBM Research - Brazil, São Paulo
**Major:** Computer Science or Computer Engineering
**Skills:** software engineering
**Description:** Clone and own approach is often used in practice to build similar software products and cloning practices are usually performed in an ad-hoc manner. Manual reverse engineering is cumbersome, expensive, and error-prone. Existing tools support to build structural models based on source code and few of them also support behavioral modeling. However, to build architectural product lines is necessary to handle variability. This proposal aims at developing a model-driven software tool to support the creation of software product line architectures by identifying variability on legacy structural and behavioural models. Intern will develop, implement, and test that software tool. The software developed by the intern will be used to support the development of a large project on enterprise cloud environment. This cloud environment should be customized according to client's needs like a software product line. Strong programming and software engineering (e.g. requirements, software architecture design) skills are required; knowledge of model-driven development or software product lines is a plus.

**Proponent: Marco Netto**

**Project**: Elasticity in Cloud Computing
**Period**: summer 2014 (10 weeks)
**Place**: IBM Research - Brazil, São Paulo
**Major**: Computer Science or Computer Engineering
**Skills**: computer programming and computer networks
**Description**: Intern will design, develop, optimize, and test algorithms for managing Cloud resources focusing on elasticity aspects. The algorithms developed/investigated by the intern will be part of a large project on Cloud Computing Platforms in IBM. Strong programming and computer networks skills needed, virtualization and operating systems are a plus.

**Proponent: Eduardo Rodrigues**

**Project**: **High Performance Fracture Simulation**
**Period**: Summer 2014 (10 weeks)
**Place**: IBM Research - Brazil, **Sao Paulo**
**Major**: Computer Sciences, Math, Engineering
**Skills**: Parallel programming; Some Knowledge on Finite Element Methods
**Description**: Accelerators and co-processors have become important tools for High Performance Computing. Programming these devices efficiently is not simple though. In this project, we are going to design and evaluate alternatives to implement eXtended Finite Element Method (XFEM) software in accelerators and co-processors. This method has applications in many areas, for example, simulation of fractures and material interfaces. In addition, it has some very attractive features and is been used in many state-of-art simulators. In this project we expect to develop a competitive implementation of XFEM for High Performance Computing machines.