

IBM projects for 2017

1. Project title: Enhanced Oil Recovery Science & Technology

Location: Rio de Janeiro

Major: Physics, Reservoir Engineering, Computer Science

Skills: Multi-scale modeling of flow in porous media; Image data analysis and processing

Description: Computational and experimental research at various physical scales enables cognitive simulation technologies that provide insights and strategies for improving oil recovery

2. Project title: Paper-based Microfluidics for Agriculture

Location: Rio de Janeiro

Major: Physics, Electrical Engineering, Computer Science

Skills: Computational Flow Simulation, Image data analysis and processing

Description: Low-cost and versatile paper-based microfluidics technology for agricultural, environmental and healthcare testing

3. Project Title: Automated knowledge base construction in specific domains

Place: São Paulo, Brazil

Major: Computer Science

Skills: Good knowledge of (maximize): Text and web mining, knowledge and data engineering, knowledge-based reasoning. Good programming skills (Python or Java).

Description: Question answering and other artificial intelligence tasks may benefit from previously acquired structured information that is encoded in a knowledge base. While for common-sense reasoning tasks most required information can be acquired from web sources only, domain-specific reasoning tasks require information to be extracted mostly from textbooks, technical standards and documents. In this project we aim at developing methods for extracting knowledge and enabling effective reasoning towards satisfying information needs in technical domains, with focus on natural resources ones, namely, oil and gas, and agriculture.

4. Project Title: Anomaly detection in crop fields using satellite images

Place: Rio de Janeiro, Brazil

Major: Engineering, Computer Science, Physics, Statistics, Mathematics, Geography or related fields.

Skills: Familiarity with one or more of these topics: image processing, remote sensing, machine learning, statistics.

Description:

Anomaly detection is an important problem that has been researched within diverse research areas and application domains. In agriculture, anomalies in the crop fields could be due to crop health variability, poor plant stand, bug infestation, etc. Often, anomalies are associated with productivity loss. Therefore, having an operational system that could emit early warnings of possible anomalous developments in farms could play a relevant role. In this context, remotely sensed data collected by orbital platforms, such as the Sentinel-2 satellite, provide repetitive image acquisitions covering both the visible and infrared portions of the electromagnetic spectrum. The objective of the internship is to investigate how the multivariate signals measured in each pixel can be used to address anomaly detection in crop fields. The work consists of developing and testing a method that can analyze observations from a single acquisition date, or eventually a set of multi-temporal acquisitions. The effectiveness of the technique will be tested using images collected over tropical farms and planted forests.

5. Project Title: Data repository for datasets with point-wise uncertainty for data machine learning experiments.

Place: Brazil Research Lab, Sao Paulo

Major: Computer Science

Skills: Machine learning, Python and/or R

Description:

There are several data sources producing datasets containing observations with point-wise uncertainty (microarray data, weather data, noisy data, astronomical data). Such observations usually are modeled with sets, probability distributions, intervals, etc., and some machine learning methods are specifically designed for dealing with that data (for example. measure machines, kernel methods with vicinal, fuzzy, or kernels on probability distributions). However, there is a lack of a common repository for that kind of data. This

project aims to identify, locate and gather all that data into a unique data repository. This would help researchers to validate their methods and to identify new problems to work on.

It is expected that in the process the student learn how to implement a machine learning model to deal with that kind of data, and to write a paper to present the repository to the scientific community.

6. Project Title: Real-time face tracking on the IBM Neuromorphic System

Place: Rio de Janeiro, Brazil

Major: Computer Science/Electrical Engineering

Skills:- Neural networks (crucial); Computer vision (very helpful);- Neuroinformatics (helpful);- FPGA programming (helpful);- Matlab (optional)

Description:

Have you ever considered that our neural networks process what we do or sense? That is the challenge here: implement a face-tracking mechanism in real-time using only hardware neurons and synapses. A few years ago, IBM presented the Neuromorphic System [1] consisting of a chip (TrueNorth) with 1,000,000 neurons and 256,000,000 synapses and a development toolkit in Matlab. The latest provides network templates and a training mechanism to brew neural networks for a given dataset [2, 3]. Although we could train a network to identify a face in a video, sometimes performance can be a major issue. That is the reason we are looking for a more elegant solution, such as Histogram of Oriented Gradients [5] implemented directly using TrueNorth neurons and synapses as mentioned in [4].

[1] Merolla, Paul A., et al. "A million spiking-neuron integrated circuit with a scalable communication network and interface." *Science* 345.6197 (2014): 668-673.

[2] Esser, Steven K., et al. "Convolutional networks for fast, energy-efficient neuromorphic computing." *Proceedings of the National Academy of Sciences* (2016): 201604850.

[3] Esser, Steve K., et al. "Backpropagation for energy-efficient neuromorphic computing." *Advances in Neural Information Processing Systems*. 2015.

[4] Chevitarese, Daniel Salles, and Marcelo Nery Dos Santos. "Real-time face tracking and recognition on IBM neuromorphic chip." *Multimedia (ISM)*, 2016 IEEE International Symposium on. IEEE, 2016.

[5] Kazemi, Vahid, and Josephine Sullivan. "One millisecond face alignment with an ensemble of regression trees." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014.

7. Project Title: Sparse Multi-Order Relaxation Solver for Polynomial Optimization

Period: Summer or Fall 2017 (10 weeks)

Place: IBM Research - Brazil, São Paulo

Major: Computer Engineering, Mathematics, Computer Science

Skills: Mathematical optimization, Python programming, and C++ programming

Description:

Many mathematical optimization problems in science and engineering are modeled in terms of polynomials, so it is becoming more important to develop new computational tools to solve those problems. Exactly solving them is sometimes computationally intractable. In this scenario, approximation methods play an important role. Most polynomial optimization solvers are not flexible enough in terms of applying different relaxation orders to specific problem variables, thus making it cumbersome to implement computationally tractable solutions whilst maintaining global optimality certificates. In this summer internship you will have a hands-on experience in developing different sparse multi-order relaxation strategies for polynomial optimization using Python, C++, and existing semidefinite programming solvers.

8. Project title: Pseudo-well log generation based on stochastic rock physics models

Location: Campinas, Sao Paulo

Major: Mathematics, physics, mechanical engineering

Skills: Python programming, stochastic simulation, optimization, statistics

Description: Well log data is widely employed in both geophysical and petrophysical analysis workflows, including quantitative seismic interpretation and seismic inversion, just to name a few. Correlating seismic data and well log data is a very challenging task in highly heterogeneous porous media due to the large resolution mismatches between these datasets. In this project, you will have the opportunity to work on stochastic pseudo well log generation based on rock physics models that describe media pore types by means of inversion algorithms. The goal is to design, implement, and test algorithms that take seismic

interpretation, lithological and well log data as input to provide probabilistic simulations of high resolution reservoir data for improved characterization.

9. Project Title: Knowledge Workbench

Place: Rio de Janeiro

Major: Computer Science

Skills: Programming, Cloud, Linked-data, Security

Description: The knowledge workbench is a key component for building knowledge-intensive applications. The KW captures rich domain relationships, integrates diverse data sources and also provides rich search services.

This project proposal seeks for students with a strong interest in software engineering who would like to work with Cognitive Computing technologies from an architectural point of view. The intern could work on key aspects such as domain-independent query engines, scalability, performance, security or DevOps practices.

10. Project Title: Impact Analysis of Ontology Schema Changes to Source Code

Place: Rio de Janeiro

Major: Computer Science

Skills: Experience in programming languages (like Java, JavaScript, Python), SQL, RDF, OWL, SPARQL, Web Services RESTful

Description: Gruber (2008) defines ontology as “a set of representational primitives with which to model a domain of knowledge or discourse”. Ontologies can be used in several areas, such as (Damjanovic et al., 2004; McGuinness, 2005): database design; data integration; information retrieval and extraction; knowledge management; etc. The OWL (a W3C standard) is by far the most adopted language for ontology representation. Building an ontology is only one step in the lifecycle of an ontology within a corporate environment, since there is a need to envision ontology construction and evolution as part of an initiative towards information governance (Azevedo et al., 2009). One use of ontologies is for semantics interoperability between different information sources and applications. Therefore, one change in an ontology item can produce inconsistencies in other parts of it. The bigger is the ontology, the more complex is to understand completely the extension and meaning of changes (Stojanovic et al., 2002). The implementation of changes must be

propagated to other dependent ontologies and applications. In addition to this, ontologies' instances must be changed to preserve consistency.

This project aims to develop techniques to support the impact analysis of ontology evolution on source code. The intern will investigate approaches for handling ontology schema evolution, develop a technique to deal with existing challenges, and assess their use on existing ontologies broadly used in the literature. The project aims at dealing with issues like: identification and reduction of the impacts of changes in ontologies; evolution management of semantic applications; understanding of coupling and cohesion of components based on the data elements they manipulate.

Azevedo, L. G., Siqueira, S. W. M. , Baiao, F., Souza, J. F., Lopes, M., Cappelli, C., Santoro, F. (2009). Enterprise Ontology Management: An Approach Based on Information Architecture. In: 11th International Conference on Enterprise Information Systems, 2009.

Damjanovic, V., Devedžic, V., Djuric, D., et al., 2004. Framework for Analyzing Ontology Development Tools. AIS SIGSEMIS Bulletin, 1 (3), pp. 43-47.

Gruber, T.R., 2008. Ontology. "Encyclopedia of Database Systems", In: Liu, L. and Özsu, M. (Eds.) Springer-Verlag.

McGuinness, D.L., 2005. Ontologies Come of Age. In: Fensel, D., et al. (eds), Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential, MIT Press.

Stojanovic, L., Maedche, A., Motik, B., et al., 2002. User-driven ontology evolution management. In Proceedings of the 13th European Conference on Knowledge Engineering and Knowledge Management, Ontologies and the Semantic Web, pp. 285-300.

11. Project Title: Text Entailment

Place: Rio de Janeiro

Major: CS

Skills: Programming, basic NLP understanding

Description:

We propose a project to evaluate different techniques for text entailment test (https://en.wikipedia.org/wiki/Textual_entailment) using deep parsing. We were particularly interested in 'deep' linguistic processing of sentences. The goal is the combination of linguistic and statistical processing methods for getting at the meaning of texts and utterances. For the experiments, we propose the use of the

<http://clic.cimec.unitn.it/composes/sick.html> corpus and it was the corpus used in the <http://alt.qcri.org/semEval2014/task1/>. Some tools/ideas under consideration are:

1. <https://arxiv.org/abs/1702.03196>
2. <https://github.com/percyliang/sempr>
3. <http://nlp.stanford.edu/pubs/schuster2016enhanced.pdf>
4. <https://github.com/uwnlp/EasySRL>
5. <http://amr.isi.edu/>

12. Project Title: Dependency Parser for Portuguese in FreeLing

Place: Rio de Janeiro

Major: CS

Skills: Programming, basic NLP understanding

Description:

FreeLing (<http://nlp.lsi.upc.edu/freeling/>) is a developer-oriented library providing language analysis services. FreeLing has already a good support for Portuguese in all its base modules (tokenizer, sentence splitter, POS tagger, WSD, etc.). We want to extend that support with a dependency parser for Portuguese. This project is about to understand how to train a parser in FreeLing and make it for Portuguese, evaluating the result. For the training we can use the recently released UD_Portuguese data under the Universal Dependencies <http://universaldependencies.org/project>.

13. Project Title: an SUO-KIF translator to TPTP

Place: Rio de Janeiro

Major: CS

Skills: Programming in Haskell, Common Lisp or Prolog

Description:

We have rewritten the translation from SUO-KIF logic language (<http://www.adampease.org/OP/>) to TPTP (<http://www.cs.miami.edu/~tptp/>). The code is <https://github.com/own-pt/cl-krr>. In this project we want to expand the translation of high-order construction to TPTP/THF. In the sequence, we want to make the output readable to

<http://www.ai.sri.com/~stickel/snark.html> prover to explore its support to Procedural Attachments.

Ideally, the translator should be written in logic or functional programming style using: Prolog, Haskell or Common Lisp, etc.

14. Project Title: CoNLL-U and Universal Dependencies toolset

Place: Rio de Janeiro

Major: CS

Skills: Programming in Common Lisp

Description:

The creation of an annotated corpus with dependencies is a hard task and very time-consuming. We are collaborating with the UD Project, <http://universaldependencies.org/>, with a Portuguese Corpus (UD_Portuguese). After release 2.0, we are now preparing for the next version expanding and solving errors in the current 2.0 corpus. In this project, we are interested in improving the necessary tools

that we use:

1. <https://github.com/own-pt/cl-conllu> : A Common Lisp library for work with CoNLL-U files
2. <https://github.com/own-pt/conll-workbench> : a set of opensource tools that we use for searching and editing the corpus.

In particular, the CL library needs better support for rules and functions for comparing different trees and help in the identification of common patterns of errors.

15. Project Title: Improving the openWordnet-PT interface

Place: Rio de Janeiro

Major: CS

Skills: Web Programming

Description:

The OpenWordnet-PT (<http://wnpt.brcloud.com/wn/>, abbreviated as OpenWN-PT or simply OWN-PT) is a open access wordnet for Portuguese, originally developed by Valeria de Paiva, Alexandre Rademaker and Gerard de Melo as a syntactic projection of Universal WordNet

(UWN) of de Melo and Weikum. Like many other open wordnets we believe that lexical resources need to be open to be useful. The OpenWN-PT is available in RDF/OWL, following and expanding, when necessary, the mappings from the original Princeton WordNet. Both the data and the RDF template settings (classes and properties) of the OpenWN-PT are freely available for download here. Besides being downloadable, the data can be retrieved via SPARQL in the endpoint and one can consult and compare it with other wordnets at the generic interface provided by the Open MultiLingual WordNet project. This project is about helping our team in the improving of the web interface for our openWordnet-PT. Code code and issues are in github: <https://github.com/own-pt/cl-wnbrowser> In particular, we need to: (1) simplify the architecture; (2) improve the interface for votes and suggestions; (3) improve navigation and data visualization.