

ADAPT Undergraduate Internship Programme 2018

PROJECT DESCRIPTION

Institution/Team:	ADAPT Centre, Theme E - TCD	
Project Title:	A Data-Driven Dashboard for Exploring Data Quality Problems.	
Suitable for students who are studying in the following areas:	Web Programming , Visualisation, Data Analysis, Linked Data and the Semantic Web	
Skills needed:	Strong javascript / NodeJs programming skills, usage of third-party libraries such as D3.js. Basic knowledge of Java and Github. Knowledge of statistical methods and/or Linked Data is a benefit.	
Project Description:	<p>Modern machine Learning, analytics and data driven processes all depend on data quality – having the right data for the job. But how do you assess data quality? The Luzzu data quality assessment framework is a flexible, open source Java-based toolset for assessing the quality of Linked Data that is being developed by the ADAPT Centre at TCD. Luzzu supports semantic (self-describing) reporting of data quality assessments. Luzzu uses the dataset quality vocabulary, the quality problem ontology and the Luzzu metric implementation ontology for these reports. However Luzzu is still a command-line tool and the reports it generates are optimised for machine readability. In this project we will build a data quality dashboard that visualises the Luzzu data quality reports and makes it easy for quality managers or data stewards to understand a data quality assessment reports visually explore the links between quality metric values and data quality issues. The project is part of our ongoing collaboration with Ordnance Survey Ireland who manage over 1.2 petabytes of national spatial reference data.</p>	
The Role of the student & benefits gained from participation in this project:¹	<p>Skills:</p> <ul style="list-style-type: none"> • Learning about Linked Data and the RDF data model, data quality management • Working with real world customers/users as part of a team • Get more familiar with different visualisation tools <p>Experience:</p> <ul style="list-style-type: none"> • Working within a research group which might involve scientific writing to introduce and foster the student into the academic world • Working with a small team to improve upon a state-of-the-art tool which is widely used in academia but yet has to reach out further industrial partners. The output of this tool will be used in an industry-led project. • Working within an agile environment 	
Who will be working with you?	The student will be working closely with research fellow Dr. Jeremy Debattista as part of the research team of Dr. Rob Brennan who is leads research on the ADAPT-Ordnance Survey Ireland collaboration and will be available for support. The student will participate in weekly project meetings and will also be able to attend the regular ADAPT Theme E research meetings.	
Short description of the group:	We have 13 people in our ADAPT Theme E group in TCD – Prof. Declan O’Sullivan and Prof. Dave Lewis theme E leaders, 5 post-docs, and 6 PhD students.	
Recommended Reading Material:	The Luzzu framework, http://eis-bonn.github.io/Luzzu/ Data Quality, https://en.wikipedia.org/wiki/Data_quality Linked Data - The Story So Far, http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf	
Other information:		
For further details on this project please contact:	Name: Phone: E-Mail: Website:	Jeremy Debattista or Rob Brennan 01 896 4244 (Rob Brennan) jeremy.debattista@adaptcentre.ie http://kdeg.scss.tcd.ie/

¹ This is an initial description of the role of the student and it is liable to change following discussions between the investigators and the student.

