
Article

[Michelle Stolwyk](#) · May 25, 2017 2m read

The Interns are Coming!

The Data Platforms department here at InterSystems is gearing up for this year's crop of interns, and I for one am very excited to meet them all next week!

We've got folks from top technical colleges with diverse specialties from hard core engineers to pure computer scientists to mathematicians to business professionals. They come from countries around the world like Vietnam, China, and Finland and they all come with impressive backgrounds. We're sure they will do very well this summer.

Rather than providing a set list of projects this year, Data Platforms interns will be developing their own projects which satisfy a short list of requirements. I can't wait to see what they come up with!

For an idea of what to expect, check out the white paper attached from my group last year who ported Deep Feature Synthesis research done at MIT for SQL platforms to Caché. Sarat Vysyaraju and Ryan St. Pierre did me and all the other mentors who provided guidance proud.

Here's a taste from their abstract.

Data science involves analyzing and deriving insights from large sets of data. Such a process requires a data science team to invest a lot of time and resources in order to draw meaningful conclusions from the data. We have designed an end to end product for data residing on Caché by adapting the Deep Feature Synthesis (DFS) algorithm [1]. This product automates the process, by aiding data science teams in the discovery of hidden links, generation of meaningful features from the data, and creation of accurate predictive models in a more time efficient manner.

The rest of the paper is organized as follows. We begin by describing the DFS algorithm. Then we state the problems and solutions we encountered while integrating DFS to Caché. Next, we explain the graph optimizations that can be applied to a relational structure to decrease the complexity of the DFS algorithm. Following this, we explain the steps taken in the machine learning process to produce predictive models for the data. Lastly, we outline the possibility for future work regarding the improvement of the tool.

[#Analytics](#) [#Caché](#) [#Interoperability](#) [#SQL](#) [#Unstructured Data](#)

Source URL: <https://community.intersystems.com/post/interns-are-coming>