

Michelin Project Discussion with Professor Dan Wooster

By Troy Butler

Prof. Wooster is a Professor of Practice with Clemson University's School of Computing. During his career in Computer Science, he has also instructed at USC Upstate and Bob Jones University, and he has acquired decades of experience in software development. I met with Prof. Wooster on Tuesday 12 November 2024 to discuss the Capstone Michelin project and gain his insights.

After explaining the nature of the project, Prof. Wooster was very interested. As an advocate for the use of generative AI as a tool, he was very enthusiastic about the benefits of an interface that is capable of both extracting business rules and answering natural language queries about them. He agreed that this kind of solution could help tens of thousands of companies as well as Government agencies. Prof. Wooster has worked with COBOL-based systems in the past and is familiar with the complexities of embedded rules.

I described our current LLM architecture to Prof. Wooster, to which he was interested but also expressed concern with regard to potential LLM memory and the privacy of proprietary data. "What are the guarantees that the LLM won't remember (or transmit) information," he asked. While it is my understanding that the LLM model we are using is entirely local to the system it operates on, this demands further research to ensure privacy and security of data is maintained.

When asked about business rules, Prof. Wooster was curious to know if Michelin's COBOL system is a single system or multisystem. He then pondered the scope of the enterprise system and whether it is used for all Michelin departments. He stated that it may include rules for payroll, invoices, HR, international operations, interfaces with Federal Government, interfaces with customers and vendors, testing and development, and more.

He suggested that we identify a set of large public-domain COBOL source code that we could experiment with, to which I informed him of the NASA files that were provided to us. He stated that the larger the codebase is, the more accurately we can determine if the generative AI interface works.

He then noted that there is a downside to using the NASA code – we know nothing about the business rules of NASA. "At some point, we need an expert in that field to say 'yes, that's a viable business rule.' "

He also noted that in order to fully understand the context of extracted business rules, access to database schemas would become necessary. "The COBOL code is probably interfacing with multiple databases. At some point, for the rules we extract out of the system to make any sense, it's got to be referencing specific fields in the database, so we would need the COBOL code and the database schemas." He then speculated that the Data Division section within the COBOL code likely contains schema information, which I later confirmed to be the case.

Prof. Wooster then expressed that, given the size of an enterprise system, this is a massive project. He began to speculate the number of embedded business rules in Michelin's codebase and estimated that it is potentially in the order of tens of thousands. He then expressed interest in how Michelin intends to keep track of which rules have been implemented in the new system at any given time and which ones have not, given the number of rules that will be extracted. He was also curious about how rules will be tested in the new system during the migration, and whether regression testing will be used as more rules are implemented. Finally, he wondered if all of Michelin's business rules are contained within the COBOL code but admitted that is likely outside the scope of this project.

Overall, Prof. Wooster was very interested and enthusiastic about the project and raised valuable points of interest. He is eager to hear more about the project in future and is more than willing to provide advice as needed. He also gave his consent to be contacted by future Capstone teams with respect to this project.