

Improving Schema Generation for Multi-Entity JSON Arrays

Davis Goulet¹

The Javascript Object Notation (JSON) is one of the primary document formats used for transmitting and storing information. Because of its prominence, an entire ecosystem of tools has risen up around it. One such tool is another document format that can be used to create a schema document that describes the structure of a JSON file.

As these schema documents are tedious and error-prone for humans to create manually, the task has largely been allocated to computers. However, existing programs do not adequately handle the scenario where a JSON array contains elements with differing sub-structures. In essence, current approaches involve mashing these different structures together into one schema which can lead to unintended side effects.

In this talk, I will first go over the JSON and JSON schema formats. Following this, I will discuss the issues with existing approaches and finally define the problem and algorithmic solution my work entails.

¹Department of Computer Science, University of Northern British Columbia, Prince George, B.C., V2N4Z9, Canada (gouletd@unbc.ca).