

発表概要

JSONのデータ格納効率を高めるバイナリフォーマット TBONの設計と実装

李彦志^{1,a)} 菅谷 みどり^{1,b)}

2021年7月20日発表

JSONはECMAScriptから独立したデータ記述言語である。XMLフォーマットに比べ、解析負荷が低い利点があることから、データの保存やアプリケーション間のデータ交換などの場合に広く使われている。しかし、JSONはテキストフォーマットであるため、データの保存と交換を高速で利用するためには、データ格納効率が低い問題がある。ZIPによるデータ圧縮はこの問題を解決する一方、解析負荷は高くなることから、十分な解決ではない。そこで、本研究では、JSONのデータ格納効率が低いという問題に対し、データをバイナリ化保存するデータフォーマットを提案する。提案したデータフォーマットはJSONのバイナリ化としてTBONとする。TBONは、データをバイナリフォーマットで保存する。保存においては、JSONが記述できるデータはすべて格納でき、かつ、値の小さい変数でも、より効率的に保存するために、8bitデータ型なども扱えるようにした。さらに、提案したデータフォーマットを人が編集可能とするために、JSONを拡張したテキストフォーマット(TON)を提案した。TONは、JSONと同じようなテキスト形式のフォーマットであるため、人が編集しやすい利点がある。これらのJSONの拡張部分を評価するため、C言語とRubyを用いて提案フォーマットを用いてプログラムを実装した。結果は、本提案のバイナリフォーマット拡張(TBON)は、JSONと比べ、データの格納効率が向上したことを確認した。

Presentation Abstract

Design and Implementation of Binary Format TBON to Improve JSON Data Storage Efficiency

YANZHI LI^{1,a)} MIDORI SUGAYA^{1,b)}

Presented: July 20, 2021

JSON is a data notation language based on a subset of ECMAScript. Because of the advantage of lower parsing overhead than the XML, it is widely used for data storage and data interchange. However, since JSON is a text format, it has a low data storage efficiency, which will be a problem in data storage and interchange usage. While data compression such as ZIP may solve this problem, the increases of the parse overhead make it not a sufficient solution. Therefore, in this research, we propose a binary data format for storing data which can be stored by JSON data storage in an efficient way. The proposed data format TBON is a binary extension of JSON. TBON saves the data in binary format. All the data that can be contained in JSON can be stored by TBON, for the variables holds small values can be stored in a more efficient way such as 8-bit data type. Furthermore, in order to make the proposed data format editable by humans, we also proposed a text format TON which extends JSON. Since TON is a text format similar to JSON, it can be easy for humans to read or write. In order to evaluate these JSON extensions, we implemented the program using the proposed format using C language and Ruby. As a result, it was confirmed that the binary format extension TBON of the present proposal improved the data storage efficiency compared to JSON.

This is the abstract of an unrefereed presentation, and it should not preclude subsequent publication.

¹ 芝浦工業大学
Shibaura Institute of Technology, Koto, Tokyo 135-8548,
Japan

^{a)} ma20505@shibaura-it.ac.jp

^{b)} doly@shibaura-it.ac.jp