Improving JSON Schema Inference

by Incorporating User Inputs

Stijn Broekhuis & Vadim Zaytsev

BENEVOL, Namen, 21 November 2024

BENEVOL24

UNIVERSITY OF TWENTE. | Fm Formal Methods & Tools

Introduction

- Vadim Zaytsev aka @grammarware ( )
  - research (, , CWI)
  - teaching ( )
  - industry (RAIN CODE mainframe to .NET , raincode LABS)
- Relevant details:
  - grammars [2004..]
  - inference [2014..]
  - by example [GPCE 2017]

http://grammarware.net && http://grammarware.github.io

UNIVERSITY OF TWENTE. | Fm Formal Methods & Tools |
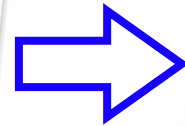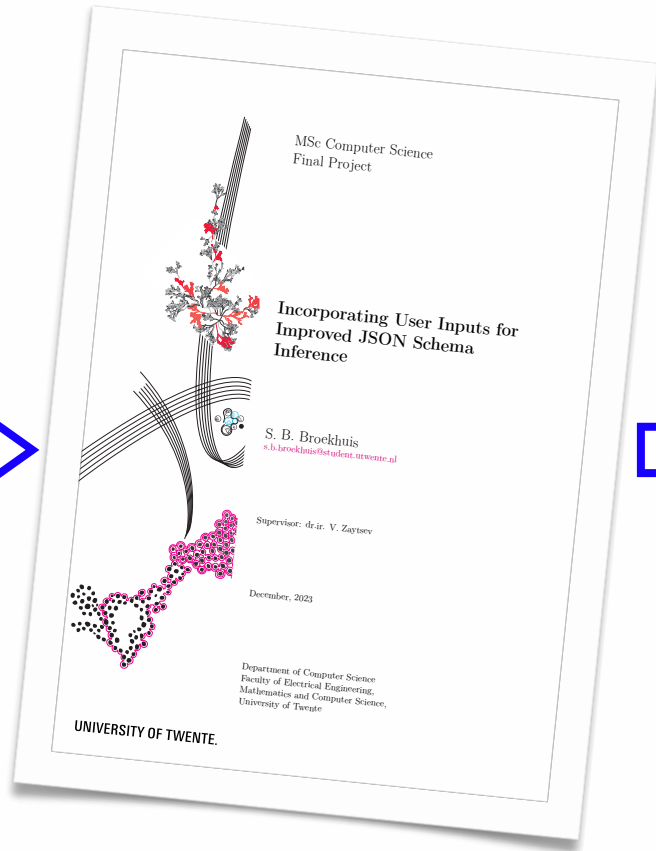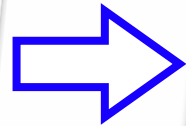
JSON vs XML
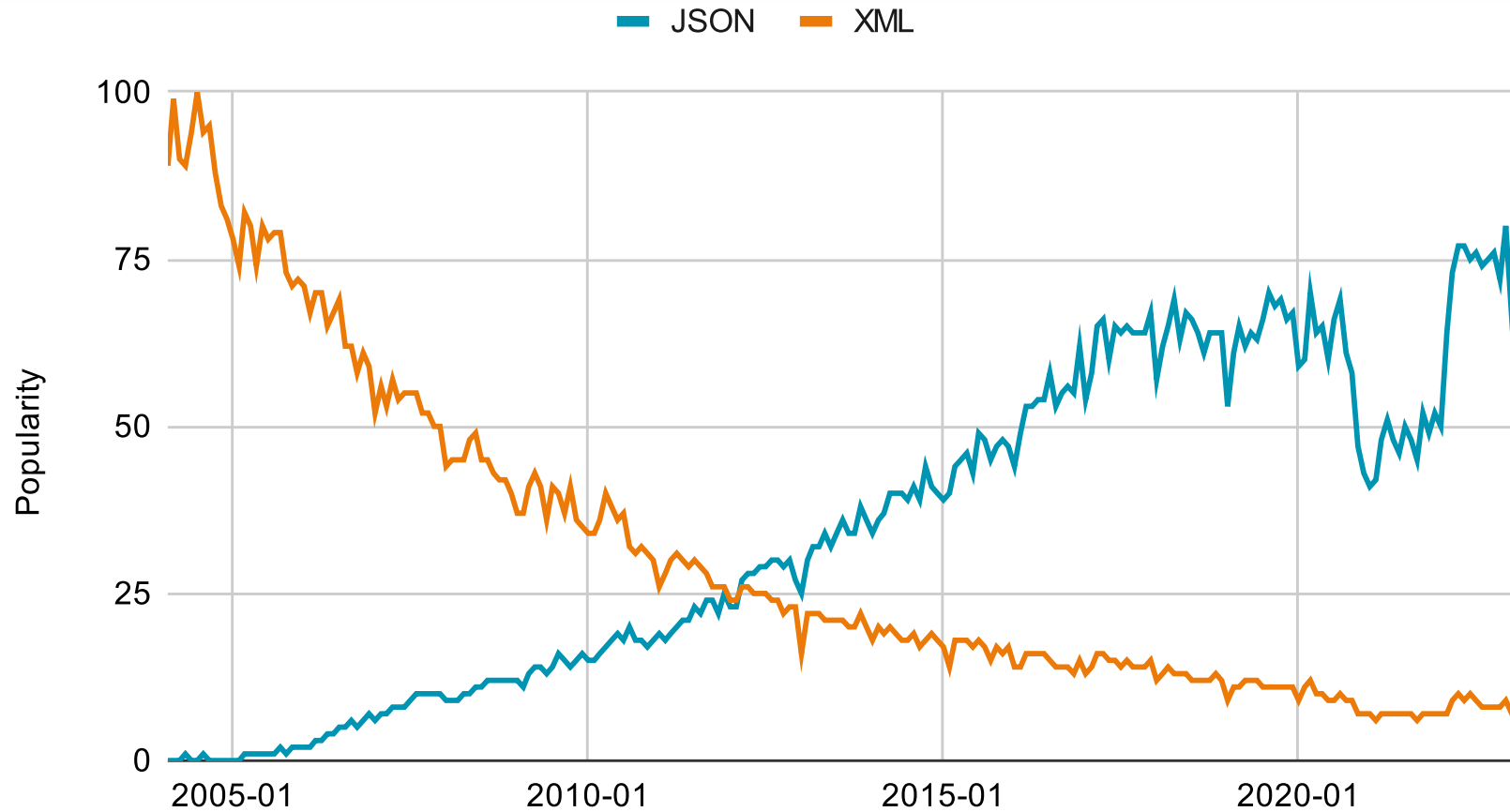
JSON in 🥜

```json
{
  "children" : [
    {
      "name": "Emma",
      "age": 11,
      "hobbies": ["football", "drawing"]
    },
    ...
  ]
}
```

JSON in 🥜

```json
{
  "orderId" : "2022343-34AZEEF",
  "userId"  : 433,
  "reason" : 1
}
```

UNIVERSITY
OF TWENTE.

Fm Formal
Methods
& Tools

JSON Schema in 🥜

```
{                                                        {
  "$schema": "http://json-schema.org/draft-04/schema#",      "orderId" : "2022343-34AZEEF",
  "type": "object",                                          "userId"  : 433,
    "properties": {                                          "reason"  : 1
      "orderId": {                                      }
    "description": "Unique identifier of the order", "type": "string"
  },
  "userId": {
    "description": "Unique identifier of the user", "type": "integer"
  },
  "reason": {
    "description": "Reason for the return", "type": "string"}
  },
  "required": ["orderId","userId","reason"]
}
```

incorrect

format

minimum

$ref

UNIVERSITY OF TWENTE.

Fm Formal Methods & Tools

Grammar/Schema Inference

- a lot of prior work
  - no ultimate solution
- based on
  - finite automata
  - structure identification graphs
  - equivalence relations
  - class diagrams
- there are tools for JSON

Highlights

- **const**
  - always has the same value
- **enum**
  - restricted to a set of possible values
- **default**
  - semantic equivalent of a missing value
- **uniqueItems** / **multipleOf** / **anyOf**+**contains**
  - restricts structure of a value

UNIVERSITY
OF TWENTE.

Formal
Methods
& Tools

Evaluation

Informational Keys: Main Challenge

```
"people" : [
  {
    "name": "Alis", "age": 34, "email" : "alis@example.com"
  }
]
```

Informational Keys: It Gets Worse!

```json
{
  "variants": {
    "powered=false": {
      "model": "minecraft:block/oak_pressure_plate"
    },
    "powered=true": {
      "model": "minecraft:block/oak_pressure_plate_down"
    }
  }
}
```

Conclusion

- JSON Schema inference from data
- when in doubt, ask the user
  - ~~assume~~ propose **default** values
  - more work
  - better results
- challenges left
  - custom Booleans (**none**/**frozen**)
  - instrumental keys