

Understanding JSON Schema

Release 7.0

Michael Droettboom, et al Space Telescope Science Institute

Contents

2	The 1 3.1 3.2 3.3	Hello, World!	4 4 7 11 11 12 13
_	The 1 3.1 3.2 3.3	basics Hello, World!	11 11 12
	3.1 3.2 3.3	Hello, World!	11 12
	3.4		13
	4.1 4.2 4.3 4.4	Type-specific keywords string 4.2.1 Length 4.2.2 Regular Expressions 4.2.3 Format Regular Expressions 4.3.1 Example Numeric types 4.4.1 integer 4.4.2 number 4.4.3 Multiples 4.4.4 Range object 4.5.1 Properties 4.5.2 Pattern Properties 4.5.3 Additional Properties 4.5.4 Required Properties	15 15 17 19 20 22 23 24 25 26 29 30 31 32
	4.6	4.5.6 Size 4.5.7 Dependencies array 4.6.1 Items	36 37 41 41

In	dev			75
6	Ackn	owledgn	nents	73
	5.3	Extendi	ng	70
	. .	5.2.1	Using \$id with \$ref	70
	5.2		property	69
		5.1.1	Recursion	67
	5.1	Reuse.		65
5	Struc	cturing a	complex schema	65
		4.13.1	Advanced	64
	4.13		hema keyword	64
	4.12	110	g subschemas conditionally	59
	4.10	4.11.5	Properties of Schema Composition	56
		4.11.4	not	56
		4.11.3	oneOf	55
		4.11.2	anyOf	54
		4.11.1	allOf	54
	4.11		Composition	53
	4 4 4		Examples	52
		4.10.2	contentEncoding	52
		4.10.1	contentMediaType	52
	4.10		string-encoding non-JSON data	52
		4.9.4	Constant values	51
		4.9.3	Enumerated values	50
		4.9.2	Comments	50
		4.9.1	Annotations	49
	4.9	Generic	keywords	49
	4.8	null		48
	4.7	boolean		47
		4.6.5	Uniqueness	47
		4.6.4	Length	46
		4.6.3	Contains	45

JSON Schema is a powerful tool for validating the structure of JSON data. However, learning to use it by reading its specification is like learning to drive a car by looking at its blueprints. You don't need to know how an electric motor fits together if all you want to do is pick up the groceries. This book, therefore, aims to be the friendly driving instructor for JSON Schema. It's for those that want to write it and understand it, but maybe aren't interested in building their own car—er, writing their own JSON Schema validator—just yet.

Note: This book describes JSON Schema draft 7. Earlier versions of JSON Schema are not completely compatible with the format described here, but for the most part, those differences are noted in the text.

Where to begin?

- This book uses some novel *conventions* (page 3) for showing schema examples and relating JSON Schema to your programming language of choice.
- If you're not sure what a schema is, check out What is a schema? (page 7).
- *The basics* (page 11) chapter should be enough to get you started with understanding the core *JSON Schema Reference* (page 15).
- When you start developing large schemas with many nested and repeated sections, check out *Structuring a complex schema* (page 65).
- json-schema.org has a number of resources, including the official specification and tools for working with JSON Schema from various programming languages.
- There are a number of online JSON Schema tools that allow you to run your own JSON schemas against example documents. These can be very handy if you want to try things out without installing any software.

Contents 1

2 Contents

CHAPTER 1

Conventions used in this book

- Language-specific notes (page 3)
- Draft-specific notes (page 4)
- Examples (page 4)

1.1 Language-specific notes

The names of the basic types in JavaScript and JSON can be confusing when coming from another dynamic language. I'm a Python programmer by day, so I've notated here when the names for things are different from what they are in Python, and any other Python-specific advice for using JSON and JSON Schema. I'm by no means trying to create a Python bias to this book, but it is what I know, so I've started there. In the long run, I hope this book will be useful to programmers of all stripes, so if you're interested in translating the Python references into Algol-68 or any other language you may know, pull requests are welcome!

The language-specific sections are shown with tabs for each language. Once you choose a language, that choice will be remembered as you read on from page to page.

For example, here's a language-specific section with advice on using JSON in a few different languages:

Python

In Python, JSON can be read using the json module in the standard library.

Ruby

In Ruby, JSON can be read using the json gem.



For C, you may want to consider using Jansson to read and write JSON.

1.2 Draft-specific notes

The JSON Schema standard has been through a number of revisions or "drafts". The the current version is Draft 7, but Draft 4 is still widely used as well.

The text is written to encourage the use of Draft 7 and gives priority to the latest conventions and features, but where it differs from earlier drafts, those differences are highlighted in special call-outs. If you only wish to target Draft 7, you can safely ignore those sections.

New in draft 7



This is where anything pertaining to an old draft would be mentioned.

1.3 Examples

There are many examples throughout this book, and they all follow the same format. At the beginning of each example is a short JSON schema, illustrating a particular principle, followed by short JSON snippets that are either valid or invalid against that schema. Valid examples are in green, with a checkmark. Invalid examples are in red, with a cross. Often there are comments in between to explain why something is or isn't valid.

Note: These examples are tested automatically whenever the book is built, so hopefully they are not just helpful, but also correct!

For example, here's a snippet illustrating how to use the number type:



Simple floating point number:



Exponential notation also works:



Numbers as strings are rejected:



1.3. Examples 5

CHAPTER 2

What is a schema?

If you've ever used XML Schema, RelaxNG or ASN.1 you probably already know what a schema is and you can happily skip along to the next section. If all that sounds like gobbledygook to you, you've come to the right place. To define what JSON Schema is, we should probably first define what JSON is.

JSON stands for "JavaScript Object Notation", a simple data interchange format. It began as a notation for the world wide web. Since JavaScript exists in most web browsers, and JSON is based on JavaScript, it's very easy to support there. However, it has proven useful enough and simple enough that it is now used in many other contexts that don't involve web surfing.

At its heart, JSON is built on the following data structures:

```
object:
{ "key1": "value1", "key2": "value2" }
array:
[ "first", "second", "third" ]
number:
42
3.1415926
string:
"This is a string"
boolean:
true
false
null:
```

null

These types have analogs in most programming languages, though they may go by different names.

Python

The following table maps from the names of JSON types to their analogous types in Python:

JSON	Python
string	string
number	int/float
object	dict
array	list
boolean	bool
null	None

45

Ruby

The following table maps from the names of JSON types to their analogous types in Ruby:

JSON	Ruby
string	String
number	Integer/Float
object	Hash
array	Array
boolean	TrueClass/FalseClass
null	NilClass

6

8

With these simple data types, all kinds of structured data can be represented. With that great flexibility comes great responsibility, however, as the same concept could be represented in myriad ways. For example, you could imagine representing information about a person in JSON in different ways:

```
"name": "George Washington",
"birthday": "February 22, 1732",
"address": "Mount Vernon, Virginia, United States"
}

{
    "first_name": "George",
```

(continues on next page)

 $^{^4}$ Since JSON strings always support unicode, they are analogous to unicode on Python 2.x and str on Python 3.x.

⁵ JSON does not have separate types for integer and floating-point.

⁶ JSON does not have separate types for integer and floating-point.

(continued from previous page)

```
"last_name": "Washington",
"birthday": "1732-02-22",
"address": {
    "street_address": "3200 Mount Vernon Memorial Highway",
    "city": "Mount Vernon",
    "state": "Virginia",
    "country": "United States"
}
```

Both representations are equally valid, though one is clearly more formal than the other. The design of a record will largely depend on its intended use within the application, so there's no right or wrong answer here. However, when an application says "give me a JSON record for a person", it's important to know exactly how that record should be organized. For example, we need to know what fields are expected, and how the values are represented. That's where JSON Schema comes in. The following JSON Schema fragment describes how the second example above is structured. Don't worry too much about the details for now. They are explained in subsequent chapters.

```
{ json schema }
{
  "type": "object",
  "properties": {
    "first_name": { "type": "string" },
    "last_name": { "type": "string" },
    "birthday": { "type": "string", "format": "date" },
    "address": {
      "type": "object",
      "properties": {
        "street_address": { "type": "string" },
        "city": { "type": "string" },
        "state": { "type": "string" },
        "country": { "type" : "string" }
      }
    }
  }
}
```

By "validating" the first example against this schema, you can see that it fails:

```
{
    "name": "George Washington",
    "birthday": "February 22, 1732",
    "address": "Mount Vernon, Virginia, United States"
}
```

However, the second example passes:

```
{
    "first_name": "George",
    "last_name": "Washington",
    "birthday": "1732-02-22",
    "address": {
        "street_address": "3200 Mount Vernon Memorial Highway",
        "city": "Mount Vernon",
        "state": "Virginia",
        "country": "United States"
    }
}
```

You may have noticed that the JSON Schema itself is written in JSON. It is data itself, not a computer program. It's just a declarative format for "describing the structure of other data". This is both its strength and its weakness (which it shares with other similar schema languages). It is easy to concisely describe the surface structure of data, and automate validating data against it. However, since a JSON Schema can't contain arbitrary code, there are certain constraints on the relationships between data elements that can't be expressed. Any "validation tool" for a sufficiently complex data format, therefore, will likely have two phases of validation: one at the schema (or structural) level, and one at the semantic level. The latter check will likely need to be implemented using a more general-purpose programming language.

CHAPTER 3

The basics

- Hello, World! (page 11)
- *The type keyword* (page 12)
- Declaring a JSON Schema (page 13)
- Declaring a unique identifier (page 13)

In *What is a schema?* (page 7), we described what a schema is, and hopefully justified the need for schema languages. Here, we proceed to write a simple JSON Schema.

3.1 Hello, World!

When learning any new language, it's often helpful to start with the simplest thing possible. In JSON Schema, an empty object is a completely valid schema that will accept any valid JSON.

```
{ json schema }
```

This accepts anything, as long as it's valid JSON



```
"I'm a string"

{ "an": [ "arbitrarily", "nested" ], "data": "structure" }
```

New in draft 6

You can also use true in place of the empty object to represent a schema that matches anything, or false for a schema that matches nothing.

```
{ json schema }
true
```

This accepts anything, as long as it's valid JSON

```
42

"I'm a string"

{ "an": [ "arbitrarily", "nested" ], "data": "structure" }
```

```
{ json schema }
false
```

```
"Resistance is futile... This will always fail!!!"
```

3.2 The type keyword

Of course, we wouldn't be using JSON Schema if we wanted to just accept any JSON document. The most common thing to do in a JSON Schema is to restrict to a specific type. The type keyword is used for that.

12 Chapter 3. The basics

Note: When this book refers to JSON Schema "keywords", it means the "key" part of the key/value pair in an object. Most of the work of writing a JSON Schema involves mapping a special "keyword" to a value within an object.

For example, in the following, only strings are accepted:

```
{ "type": "string" }

"I'm a string"

42
```

The type keyword is described in more detail in *Type-specific keywords* (page 15).

3.3 Declaring a JSON Schema

It's not always easy to tell which draft a JSON Schema is using. You can use the \$schema keyword to declare which version of the JSON Schema specification the schema is written to. See *The \$schema keyword* (page 64) for more information. It's generally good practice to include it, though it is not required.

Note: For brevity, the \$schema keyword isn't included in most of the examples in this book, but it should always be used in the real world.

```
{ json schema }

{ "$schema": "http://json-schema.org/draft-07/schema#" }
```

Draft 4

In Draft 4, a \$schema value of http://json-schema.org/schema# referred to the latest version of JSON Schema. This usage has since been deprecated and the use of specific version URIs is required.

3.4 Declaring a unique identifier

It is also best practice to include an \$id property as a unique identifier for each schema. For now, just set it to a URL at a domain you control, for example:

```
{ "$id": "http://yourdomain.com/schemas/myschema.json" }
```

The details of *The \$id property* (page 69) become more apparent when you start *Structuring a complex schema* (page 65). New in draft 6

Draft 4

In Draft 4, \$id is just id (without the dollar-sign).

14 Chapter 3. The basics

CHAPTER 4

JSON Schema Reference

4.1 Type-specific keywords

The type keyword is fundamental to JSON Schema. It specifies the data type for a schema.

At its core, JSON Schema defines the following basic types:

- string (page 17)
- number (page 25)
- *integer* (page 24)
- *object* (page 29)
- array (page 41)
- boolean (page 47)
- *null* (page 48)

These types have analogs in most programming languages, though they may go by different names.

Python

The following table maps from the names of JSON types to their analogous types in Python:

Python
string
int/float
dict
list
bool
None

45

Ruby

The following table maps from the names of JSON types to their analogous types in Ruby:

JSON	Ruby
string	String
number	Integer/Float
object	Hash
array	Array
boolean	TrueClass/FalseClass
null	NilClass

6

The type keyword may either be a string or an array:

- If it's a string, it is the name of one of the basic types above.
- If it is an array, it must be an array of strings, where each string is the name of one of the basic types, and each element is unique. In this case, the JSON snippet is valid if it matches *any* of the given types.

Here is a simple example of using the type keyword:

```
{ "type": "number" }
```



42

⁴ Since JSON strings always support unicode, they are analogous to unicode on Python 2.x and str on Python 3.x.

⁵ JSON does not have separate types for integer and floating-point.

⁶ JSON does not have separate types for integer and floating-point.



This is not a number, it is a string containing a number.



In the following example, we accept strings and numbers, but not structured data types:

```
{ "type": ["number", "string"] }

42

"Life, the universe, and everything"

["Life", "the universe", "and everything"]
```

For each of these types, there are keywords that only apply to those types. For example, numeric types have a way of specifying a numeric range, that would not be applicable to other types. In this reference, these validation keywords are described along with each of their corresponding types in the following chapters.

4.2 string

```
Length (page 19)
Regular Expressions (page 19)
Format (page 20)

Built-in formats (page 20)
* Dates and times (page 20)
* Email addresses (page 21)
* Hostnames (page 21)
```

4.2. string 17

```
* IP Addresses (page 21)
* Resource identifiers (page 21)
* URI template (page 21)
* JSON Pointer (page 22)
* Regular Expressions (page 22)
```

The string type is used for strings of text. It may contain Unicode characters.

Python

In Python, "string" is analogous to the unicode type on Python 2.x, and the str type on Python 3.x.

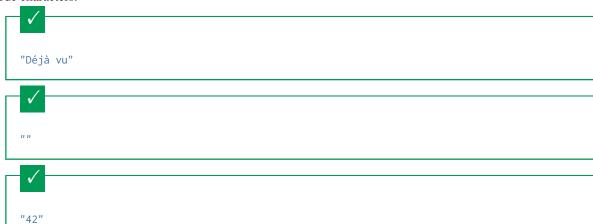
Ruby

In Ruby, "string" is analogous to the String type.

```
{ "type": "string" }
```

"This is a string"

Unicode characters:





4.2.1 Length

The length of a string can be constrained using the minLength and maxLength keywords. For both keywords, the value must be a non-negative number.

4.2.2 Regular Expressions

The pattern keyword is used to restrict a string to a particular regular expression. The regular expression syntax is the one defined in JavaScript (ECMA 262 specifically). See *Regular Expressions* (page 22) for more information.

Note: When defining the regular expressions, it's important to note that the string is considered valid if the expression matches anywhere within the string. For example, the regular expression "p" will match any string with a p in it, such as "apple" not just a string that is simply "p". Therefore, it is usually less confusing, as a matter of course, to surround the regular expression in ^...\$, for example, "^p\$", unless there is a good reason not to do so.

The following example matches a simple North American telephone number with an optional area code:

4.2. string 19

```
{
    "type": "string",
    "pattern": "^(\\([0-9]{3}\\))?[0-9]{3}-[0-9]{4}$"
}

"555-1212"

"(888)555-1212"

*

"(888)555-1212 ext. 532"

*

"(800)FLOWERS"
```

4.2.3 Format

The format keyword allows for basic semantic validation on certain kinds of string values that are commonly used. This allows values to be constrained beyond what the other tools in JSON Schema, including *Regular Expressions* (page 22) can do.

Note: JSON Schema implementations are not required to implement this part of the specification, and many of them do not.

There is a bias toward networking-related formats in the JSON Schema specification, most likely due to its heritage in web technologies. However, custom formats may also be used, as long as the parties exchanging the JSON documents also exchange information about the custom format types. A JSON Schema validator will ignore any format type that it does not understand.

Built-in formats

The following is the list of formats specified in the JSON Schema specification.

Dates and times

Dates and times are represented in RFC 3339, section 5.6. This is a subset of the date format also commonly known as ISO8601 format.

- "date-time": Date and time together, for example, 2018-11-13T20: 20: 39+00:00.
- "time": New in draft 7 Time, for example, 20:20:39+00:00
- "date": New in draft 7 Date, for example, 2018-11-13.

Email addresses

- "email": Internet email address, see RFC 5322, section 3.4.1.
- "idn-email": New in draft 7 The internationalized form of an Internet email address, see RFC 6531.

Hostnames

- "hostname": Internet host name, see RFC 1034, section 3.1.
- "idn-hostname": New in draft 7 An internationalized Internet host name, see RFC5890, section 2.3.2.3.

IP Addresses

- "ipv4": IPv4 address, according to dotted-quad ABNF syntax as defined in RFC 2673, section 3.2.
- "ipv6": IPv6 address, as defined in RFC 2373, section 2.2.

Resource identifiers

- "uri": A universal resource identifier (URI), according to RFC3986.
- "uri-reference": New in draft 6 A URI Reference (either a URI or a relative-reference), according to RFC3986, section 4.1.
- "iri": New in draft 7 The internationalized equivalent of a "uri", according to RFC3987.
- "iri-reference": New in draft 7 The internationalized equivalent of a "uri-reference", according to RFC3987

If the values in the schema have the ability to be relative to a particular source path (such as a link from a webpage), it is generally better practice to use "uri-reference" (or "iri-reference") rather than "uri" (or "iri"). "uri" should only be used when the path must be absolute.

Draft 4

Draft 4 only includes "uri", not "uri-reference". Therefore, there is some ambiguity around whether "uri" should accept relative paths.

URI template

• "uri-template": New in draft 6 A URI Template (of any level) according to RFC6570. If you don't already know what a URI Template is, you probably don't need this value.

4.2. string 21

JSON Pointer

- "json-pointer": New in draft 6 A JSON Pointer, according to RFC6901. There is more discussion on the use of JSON Pointer within JSON Schema in *Structuring a complex schema* (page 65). Note that this should be used only when the entire string contains only JSON Pointer content, e.g. /foo/bar. JSON Pointer URI fragments, e.g. #/foo/bar/ should use "uri-reference".
- "relative-json-pointer": New in draft 7 A relative JSON pointer.

Regular Expressions

• "regex": New in draft 7 A regular expression, which should be valid according to the ECMA 262 dialect.

Be careful, in practice, JSON schema validators are only required to accept the safe subset of *Regular Expressions* (page 22) described elsewhere in this document.

4.3 Regular Expressions

• Example (page 23)

The *pattern* (page 19) and *Pattern Properties* (page 31) keywords use regular expressions to express constraints. The regular expression syntax used is from JavaScript (ECMA 262, specifically). However, that complete syntax is not widely supported, therefore it is recommended that you stick to the subset of that syntax described below.

- A single unicode character (other than the special characters below) matches itself.
- .: Matches any character except line break characters. (Be aware that what constitutes a line break character is somewhat dependent on your platform and language environment, but in practice this rarely matters).
- ^: Matches only at the beginning of the string.
- \$: Matches only at the end of the string.
- (...): Group a series of regular expressions into a single regular expression.
- |: Matches either the regular expression preceding or following the | symbol.
- [abc]: Matches any of the characters inside the square brackets.
- [a-z]: Matches the range of characters.
- [^abc]: Matches any character *not* listed.
- [^a-z]: Matches any character outside of the range.
- +: Matches one or more repetitions of the preceding regular expression.
- *: Matches zero or more repetitions of the preceding regular expression.
- ?: Matches zero or one repetitions of the preceding regular expression.
- +?, *?, ??: The *, +, and ? qualifiers are all greedy; they match as much text as possible. Sometimes this behavior isn't desired and you want to match as few characters as possible.
- (?!x), (?=x): Negative and positive lookahead.
- {x}: Match exactly x occurrences of the preceding regular expression.
- {x,y}: Match at least x and at most y occurrences of the preceding regular expression.

- {x,}: Match x occurrences or more of the preceding regular expression.
- $\{x\}$?, $\{x,y\}$?, $\{x,\}$?: Lazy versions of the above expressions.

4.3.1 Example

The following example matches a simple North American telephone number with an optional area code:

```
{
    "type": "string",
    "pattern": "^(\\([0-9]{3}\\))?[0-9]{3}-[0-9]{4}$"
}

    "(888)555-1212"

    "(888)555-1212 ext. 532"

    "(800)FLOWERS"
```

4.4 Numeric types

```
integer (page 24)
number (page 25)
Multiples (page 25)
Range (page 26)
```

There are two numeric types in JSON Schema: *integer* (page 24) and *number* (page 25). They share the same validation keywords.

Note: JSON has no standard way to represent complex numbers, so there is no way to test for them in JSON Schema.

4.4. Numeric types 23

4.4.1 integer

The integer type is used for integral numbers. JSON does not have distinct types for integers and floating-point values. Therefore, the presence or absence of a decimal point is not enough to distinguish between integers and non-integers. For example, 1 and 1.0 are two ways to represent the same value in JSON. JSON Schema considers that value an integer no matter which representation was used.

Python

In Python, "integer" is analogous to the int type.



In Ruby, "integer" is analogous to the Integer type.

```
{ "type": "integer" }
```



42



_ -

Numbers with a zero fractional part are considered integers



1.0

Floating point numbers are rejected:



Numbers as strings are rejected:



4.4.2 number

The number type is used for any numeric type, either integers or floating point numbers.

Python

In Python, "number" is analogous to the float type.

Ruby

In Ruby, "number" is analogous to the Float type.

{ "type": "number" }



42



.

Simple floating point number:



5.0

Exponential notation also works:



2.99792458e8

Numbers as strings are rejected:



"42"

4.4.3 Multiples

Numbers can be restricted to a multiple of a given number, using the multipleOf keyword. It may be set to any positive number.

4.4. Numeric types 25

```
{
    "type": "number",
    "multipleOf": 10
}

10

20
```

Not a multiple of 10:



4.4.4 Range

Ranges of numbers are specified using a combination of the minimum and maximum keywords, (or exclusiveMinimum and exclusiveMaximum for expressing exclusive range).

If *x* is the value being validated, the following must hold true:

- $x \ge \min \min$
- *x* > exclusiveMinimum
- $x \leq \max$
- x < exclusiveMaximum

While you can specify both of minimum and exclusiveMinimum or both of maximum and exclusiveMaximum, it doesn't really make sense to do so.

```
{
    "type": "number",
    "minimum": 0,
    "exclusiveMaximum": 100
}
```

Less than minimum:

-1

minimum is inclusive, so 0 is valid:

0

10

99

exclusiveMaximum is exclusive, so 100 is not valid:

Greater than maximum:



4.4. Numeric types 27

Draft 4

In JSON Schema Draft 4, exclusiveMinimum and exclusiveMaximum work differently. There they are boolean values, that indicate whether minimum and maximum are exclusive of the value. For example:

- if exclusiveMinimum is false, $x \ge minimum$.
- if exclusiveMinimum is true, x > minimum.

This was changed to have better keyword independence.

Here is an example using the older Draft 4 convention:

```
{ json schema }

{
  "type": "number",
  "minimum": 0,
  "maximum": 100,
  "exclusiveMaximum": true
}
```

Less than minimum:



exclusiveMinimum was not specified, so 0 is included:

```
0 10 99
```

exclusiveMaximum is true, so 100 is not included:



Greater than maximum:

```
101
```

4.5 object

Properties (page 30)
Pattern Properties (page 31)
Additional Properties (page 32)
Required Properties (page 34)
Property names (page 36)
Size (page 36)
Dependencies (page 37)

Property dependencies (page 37)
Schema dependencies (page 39)

Objects are the mapping type in JSON. They map "keys" to "values". In JSON, the "keys" must always be strings. Each of these pairs is conventionally referred to as a "property".

Python

In Python, "objects" are analogous to the dict type. An important difference, however, is that while Python dictionaries may use anything hashable as a key, in JSON all the keys must be strings.

Try not to be confused by the two uses of the word "object" here: Python uses the word object to mean the generic base class for everything, whereas in JSON it is used only to mean a mapping from string keys to values.

Ruby

In Ruby, "objects" are analogous to the Hash type. An important difference, however, is that all keys in JSON must be strings, and therefore any non-string keys are converted over to their string representation.

Try not to be confused by the two uses of the word "object" here: Ruby uses the word Object to mean the generic base class for everything, whereas in JSON it is used only to mean a mapping from string keys to values.

```
{ "type": "object" }
```

```
{
    "key": "value",
    "another_key": "another_value"
}
```

4.5. object 29

```
{
    "Sun": 1.9891e30,
    "Jupiter": 1.8986e27,
    "Saturn": 5.6846e26,
    "Neptune": 10.243e25,
    "Uranus": 8.6810e25,
    "Earth": 5.9736e24,
    "Venus": 4.8685e24,
    "Mars": 6.4185e23,
    "Mercury": 3.3022e23,
    "Moon": 7.349e22,
    "Pluto": 1.25e22
}
```

Using non-strings as keys is invalid JSON:

```
{
    0.01: "cm",
    1: "m",
    1000: "km"
}

"Not an object"

["An", "array", "not", "an", "object"]
```

4.5.1 Properties

The properties (key-value pairs) on an object are defined using the properties keyword. The value of properties is an object, where each key is the name of a property and each value is a schema used to validate that property. Any property that doesn't match any of the property names in the properties keyword is ignored by this keyword.

Note: See *Additional Properties* (page 32) for how to disallow properties that don't match any of the property names in properties.

For example, let's say we want to define a simple schema for an address made up of a number, street name and street type:

```
{
    "type": "object",
    "properties": {
        "number": { "type": "number" },
        "street_name": { "type": "string" },
        "street_type": { "enum": ["Street", "Avenue", "Boulevard"] }
    }
}
```

```
{ "number": 1600, "street_name": "Pennsylvania", "street_type": "Avenue" }
```

If we provide the number in the wrong type, it is invalid:

```
{ "number": "1600", "street_name": "Pennsylvania", "street_type": "Avenue" }
```

By default, leaving out properties is valid. See Required Properties (page 34).

```
{ "number": 1600, "street_name": "Pennsylvania" }
```

By extension, even an empty object is valid:

```
{ }
```

By default, providing additional properties is valid:

```
{ "number": 1600, "street_name": "Pennsylvania", "street_type": "Avenue", "direction": "NW" }
```

4.5.2 Pattern Properties

Sometimes you want to say that, given a particular kind of property name, the value should match a particular schema. That's where patternProperties comes in: it maps regular expressions to schemas. If a property name matches the given regular expression, the property value must validate against the corresponding schema.

Note: Regular expressions are not anchored. This means that when defining the regular expressions for patternProperties, it's important to note that the expression may match anywhere within the property name. For example, the regular expression "p" will match any property name with a p in it, such as "apple", not just a property whose name is simply "p". It's therefore usually less confusing to surround the regular expression in ^...\$,

4.5. object 31

for example, "^p\$".

In this example, any properties whose names start with the prefix S_ must be strings, and any with the prefix I_ must be integers. Any properties that do not match either regular expression are ignored.

```
{
    "type": "object",
        "patternProperties": {
            "^S_": { "type": "string" },
            "^I_": { "type": "integer" }
      }
}
```

```
{ "S_25": "This is a string" }
```

```
{ "I_0": 42 }
```

If the name starts with S_, it must be a string

```
{ "S_0": 42 }
```

If the name starts with I_, it must be an integer

```
{ "I_42": "This is a string" }
```

This is a key that doesn't match any of the regular expressions:

```
{ "keyword": "value" }
```

4.5.3 Additional Properties

The additionalProperties keyword is used to control the handling of extra stuff, that is, properties whose names are not listed in the properties keyword or match any of the regular expressions in the patternProperties keyword. By default any additional properties are allowed.

The value of the additionalProperties keyword is a schema that will be used to validate any properties in the instance that are not matched by properties or patternProperties. Setting the additionalProperties schema to

false means no additional properties will be allowed.

Reusing the example from *Properties* (page 30), but this time setting additional Properties to false.

```
{
    "type": "object",
    "properties": {
        "number": { "type": "number" },
        "street_name": { "type": "string" },
        "street_type": { "enum": ["Street", "Avenue", "Boulevard"] }
    },
    "additionalProperties": false
}
```

```
{ "number": 1600, "street_name": "Pennsylvania", "street_type": "Avenue" }
```

Since additionalProperties is false, this extra property "direction" makes the object invalid:

```
{ "number": 1600, "street_name": "Pennsylvania", "street_type": "Avenue", "direction": "NW" }
```

You can use non-boolean schemas to put more complex constraints on the additional properties of an instance. For example, one can allow additional properties, but only if they are each a string:

```
{
    "type": "object",
    "properties": {
        "number": { "type": "number" },
        "street_name": { "type": "string" },
        "street_type": { "enum": ["Street", "Avenue", "Boulevard"] }
    },
    "additionalProperties": { "type": "string" }
}
```

```
{ "number": 1600, "street_name": "Pennsylvania", "street_type": "Avenue" }
```

This is valid, since the additional property's value is a string:

```
{ "number": 1600, "street_name": "Pennsylvania", "street_type": "Avenue", "direction": "NW" }
```

4.5. object 33

This is invalid, since the additional property's value is not a string:

You can use additionalProperties with a combination of properties and patternProperties. In the following example, based on the example from *Pattern Properties* (page 31), we add a "builtin" property, which must be a number, and declare that all additional properties (that are neither defined by properties nor matched by patternProperties) must be strings:

```
{
    "type": "object",
    "properties": {
        "builtin": { "type": "number" }
    },
    "patternProperties": {
        "^S_": { "type": "string" },
        "^I_": { "type": "integer" }
    },
    "additionalProperties": { "type": "string" }
}
```

```
{ "builtin": 42 }
```

This is a key that doesn't match any of the regular expressions:

```
{ "keyword": "value" }
```

It must be a string:

```
{ "keyword": 42 }
```

4.5.4 Required Properties

By default, the properties defined by the properties keyword are not required. However, one can provide a list of required properties using the required keyword.

The required keyword takes an array of zero or more strings. Each of these strings must be unique.

Draft 4

In Draft 4, required must contain at least one string.

In the following example schema defining a user record, we require that each user has a name and e-mail address, but we don't mind if they don't provide their address or telephone number:

```
{
    "type": "object",
    "properties": {
        "name": { "type": "string" },
        "email": { "type": "string" },
        "address": { "type": "string" },
        "telephone": { "type": "string" }
},
    "required": ["name", "email"]
}
```

```
{
    "name": "William Shakespeare",
    "email": "bill@stratford-upon-avon.co.uk"
}
```

Providing extra properties is fine, even properties not defined in the schema:

```
{
    "name": "William Shakespeare",
    "email": "bill@stratford-upon-avon.co.uk",
    "address": "Henley Street, Stratford-upon-Avon, Warwickshire, England",
    "authorship": "in question"
}
```

Missing the required "email" property makes the JSON document invalid:

```
{
    "name": "William Shakespeare",
    "address": "Henley Street, Stratford-upon-Avon, Warwickshire, England",
}
```

In JSON a property with value null is not equivalent to the property not being present. This fails because null is not of type "string", it's of type "null"

4.5. object 35

```
{
    "name": "William Shakespeare",
    "address": "Henley Street, Stratford-upon-Avon, Warwickshire, England",
    "email": null
}
```

4.5.5 Property names

New in draft 6

The names of properties can be validated against a schema, irrespective of their values. This can be useful if you don't want to enforce specific properties, but you want to make sure that the names of those properties follow a specific convention. You might, for example, want to enforce that all names are valid ASCII tokens so they can be used as attributes in a particular programming language.

```
{
    "type": "object",
    "propertyNames": {
        "pattern": "^[A-Za-z_][A-Za-z0-9_]*$"
    }
}
```

```
{
    "_a_proper_token_001": "value"
}
```

```
{
  "001 invalid": "value"
}
```

Since object keys must always be strings anyway, it is implied that the schema given to propertyNames is always at least:

```
{ "type": "string" }
```

4.5.6 Size

The number of properties on an object can be restricted using the minProperties and maxProperties keywords. Each of these must be a non-negative integer.

```
{ "type": "object",
    "minProperties": 2,
    "maxProperties": 3
}

()

("a": 0 )

("a": 0, "b": 1, "c": 2 )

("a": 0, "b": 1, "c": 2, "d": 3 }
```

4.5.7 Dependencies

The dependencies keyword allows the schema of the object to change based on the presence of certain special properties.

There are two forms of dependencies in JSON Schema:

- Property dependencies declare that certain other properties must be present if a given property is present.
- Schema dependencies declare that the schema changes when a given property is present.

Property dependencies

Let's start with the simpler case of property dependencies. For example, suppose we have a schema representing a customer. If you have their credit card number, you also want to ensure you have a billing address. If you don't have their credit card number, a billing address would not be required. We represent this dependency of one property on another using the dependencies keyword. The value of the dependencies keyword is an object. Each entry in the object maps from the name of a property, p, to an array of strings listing properties that are required if p is present.

4.5. object 37

In the following example, whenever a credit_card property is provided, a billing_address property must also be present:

```
{
    "type": "object",

    "properties": {
        "name": { "type": "string" },
        "credit_card": { "type": "number" },
        "billing_address": { "type": "string" }
},

    "required": ["name"],

    "dependencies": {
        "credit_card": ["billing_address"]
}
```

```
{
    "name": "John Doe",
    "credit_card": 555555555555555,
    "billing_address": "555 Debtor's Lane"
}
```

This instance has a credit_card, but it's missing a billing_address.

This is okay, since we have neither a credit_card, or a billing_address.

```
{
   "name": "John Doe"
}
```

Note that dependencies are not bidirectional. It's okay to have a billing address without a credit card number.

```
{
    "name": "John Doe",
    "billing_address": "555 Debtor's Lane"
}
```

To fix the last issue above (that dependencies are not bidirectional), you can, of course, define the bidirectional dependencies explicitly:

```
{
    "type": "object",

    "properties": {
        "name": { "type": "string" },
        "credit_card": { "type": "number" },
        "billing_address": { "type": "string" }
},

    "required": ["name"],

    "dependencies": {
        "credit_card": ["billing_address"],
        "billing_address": ["credit_card"]
}
}
```

This instance has a credit_card, but it's missing a billing_address.

This has a billing_address, but is missing a credit_card.

```
{
    "name": "John Doe",
    "billing_address": "555 Debtor's Lane"
}
```

Schema dependencies

Schema dependencies work like property dependencies, but instead of just specifying required properties, they specify a schema that must be valid against the instance.

4.5. object 39

For example, here is another way to write the above:

This instance has a credit_card, but it's missing a billing_address:

This has a billing_address, but is missing a credit_card. This passes, because here billing_address just looks like an additional property:

```
{
    "name": "John Doe",
    "billing_address": "555 Debtor's Lane"
}
```

4.6 array

```
Items (page 41)

List validation (page 42)
Tuple validation (page 42)

Additional Items (page 44)
Contains (page 45)
Length (page 46)
Uniqueness (page 47)
```

Arrays are used for ordered elements. In JSON, each element in an array may be of a different type.

Python

In Python, "array" is analogous to a list or tuple type, depending on usage. However, the json module in the Python standard library will always use Python lists to represent JSON arrays.

Ruby

In Ruby, "array" is analogous to a Array type.

```
{ "type": "array" }

[1, 2, 3, 4, 5]

[3, "different", { "types" : "of values" }]

**

{"Not": "an array"}
```

4.6.1 Items

There are two ways in which arrays are generally used in JSON:

• List validation: a sequence of arbitrary length where each item matches the same schema.

4.6. array 41

• **Tuple validation:** a sequence of fixed length where each item may have a different schema. In this usage, the index (or location) of each item is meaningful as to how the value is interpreted. (This usage is often given a whole separate type in some programming languages, such as Python's tuple).

List validation

List validation is useful for arrays of arbitrary length where each item matches the same schema. For this kind of array, set the items keyword to a single schema that will be used to validate all of the items in the array.

Note: When items is a single schema, the additionalItems keyword is meaningless, and it should not be used.

In the following example, we define that each item in an array is a number:

```
{
    "type": "array",
    "items": {
        "type": "number"
    }
}
```

```
[1, 2, 3, 4, 5]
```

A single "non-number" causes the whole array to be invalid:

```
[1, 2, "3", 4, 5]
```

The empty array is always valid:

Tuple validation

Tuple validation is useful when the array is a collection of items where each has a different schema and the ordinal index of each item is meaningful.

For example, you may represent a street address such as:

```
1600 Pennsylvania Avenue NW
```

as a 4-tuple of the form:

[number, street name, street type, direction]

Each of these fields will have a different schema:

- number: The address number. Must be a number.
- street_name: The name of the street. Must be a string.
- street_type: The type of street. Should be a string from a fixed set of values.
- direction: The city quadrant of the address. Should be a string from a different set of values.

To do this, we set the items keyword to an array, where each item is a schema that corresponds to each index of the document's array. That is, an array where the first element validates the first element of the input array, the second element validates the second element of the input array, etc.

Here's the example schema:

```
[1600, "Pennsylvania", "Avenue", "NW"]
```

"Drive" is not one of the acceptable street types:

```
[24, "Sussex", "Drive"]
```

This address is missing a street number

```
["Palais de l'Élysée"]
```

It's okay to not provide all of the items:

```
[10, "Downing", "Street"]
```

And, by default, it's also okay to add additional items to end:

4.6. array 43

```
[1600, "Pennsylvania", "Avenue", "NW", "Washington"]
```

4.6.2 Addtional Items

The additionalItems keyword controls whether it's valid to have additional items in a tuple beyond what is defined in items. The value of the additionalItems keyword is a schema that all additional items must pass in order for the keyword to validate. This keyword is ignored if there is not a "tuple validation" items keyword present in the same schema.

Draft 4

In Draft 4, additionalItems does not require a "tuple validation" items keyword to be present. There are no constraints on any of the items, so all items are considered to be additional items.

Here, we'll reuse the example schema above, but set additionalItems to false, which has the effect of disallowing extra items in the array.

```
[1600, "Pennsylvania", "Avenue", "NW"]
```

It's ok to not provide all of the items:

```
[1600, "Pennsylvania", "Avenue"]
```

But, since additionalItems is false, we can't provide extra items:

```
[1600, "Pennsylvania", "Avenue", "NW", "Washington"]
```

You can express more complex constraints by using a non-boolean schema to constrain what value additional items can have. In that case, we could say that additional items are allowed, as long as they are all strings:

```
{
    "type": "array",
    "items": [
        { "type": "number" },
        { "type": "string" },
        { "enum": ["Street", "Avenue", "Boulevard"] },
        { "enum": ["NW", "NE", "SW", "SE"] }
],
    "additionalItems": { "type": "string" }
}
```

Extra string items are ok ...

```
[1600, "Pennsylvania", "Avenue", "NW", "Washington"]
```

... but not anything else

```
[1600, "Pennsylvania", "Avenue", "NW", 20500]
```

Note: Because "list validation" (items is an object) applies to all items in the list three are no additional items and therefore additionalItems has nothing to apply its schema to and will have no effect.

4.6.3 Contains

New in draft 6

While the items schema must be valid for every item in the array, the contains schema only needs to validate against one or more items in the array.

```
{
    "type": "array",
    "contains": {
        "type": "number"
    }
}
```

A single "number" is enough to make this pass:

```
["life", "universe", "everything", 42]
```

4.6. array 45

But if we have no number, it fails:

```
["life", "universe", "everything", "forty-two"]
```

All numbers is, of course, also okay:

```
[1, 2, 3, 4, 5]
```

4.6.4 Length

The length of the array can be specified using the minItems and maxItems keywords. The value of each keyword must be a non-negative number. These keywords work whether doing *List validation* (page 42) or *Tuple validation* (page 42).

4.6.5 Uniqueness

A schema can ensure that each of the items in an array is unique. Simply set the uniqueItems keyword to true.

```
{
  "type": "array",
  "uniqueItems": true
}

[1, 2, 3, 4, 5]
```

The empty array always passes:



4.7 boolean

The boolean type matches only two special values: true and false. Note that values that *evaluate* to true or false, such as 1 and 0, are not accepted by the schema.

Python

In Python, "boolean" is analogous to bool. Note that in JSON, true and false are lower case, whereas in Python they are capitalized (True and False).

Ruby

In Ruby, "boolean" is analogous to TrueClass and FalseClass. Note that in Ruby there is no Boolean class.

```
{ "type": "boolean" }
```

4.7. boolean 47



Values that evaluate to true or false are still not accepted by the schema:



4.8 null

When a schema specifies a type of null, it has only one acceptable value: null.

Note: It's important to remember that in JSON, null isn't equivalent to something being absent. See *Required Properties* (page 34) for an example.

Python

In Python, null is analogous to None.

Ruby

In Ruby, null is analogous to nil.

```
{ "type": "null" }
```



4.9 Generic keywords

- Annotations (page 49)
- Comments (page 50)
- Enumerated values (page 50)
- Constant values (page 51)

This chapter lists some miscellaneous properties that are available for all JSON types.

4.9.1 Annotations

JSON Schema includes a few keywords, that aren't strictly used for validation, but are used to describe parts of a schema. None of these "annotation" keywords are required, but they are encouraged for good practice, and can make your schema "self-documenting".

The title and description keywords must be strings. A "title" will preferably be short, whereas a "description" will provide a more lengthy explanation about the purpose of the data described by the schema.

The default keyword specifies a default value. This value is not used to fill in missing values during the validation process. Non-validation tools such as documentation generators or form generators may use this value to give hints to users about how to use a value. However, default is typically used to express that if a value is missing, then the value is semantically the same as if the value was present with the default value. The value of default should validate against the schema in which it resides, but that isn't required.

New in draft 6 The examples keyword is a place to provide an array of examples that validate against the schema. This isn't used for validation, but may help with explaining the effect and purpose of the schema to a reader. Each entry should validate against the schema in which it resides, but that isn't strictly required. There is no need to duplicate the default value in the examples array, since default will be treated as another example.

New in draft 7 The boolean keywords readOnly and writeOnly are typically used in an API context. readOnly indicates that a value should not be modified. It could be used to indicate that a PUT request that changes a value would result in a 400 Bad Request response. writeOnly indicates that a value may be set, but will remain hidden. In could be used to indicate you can set a value with a PUT request, but it would not be included when retrieving that record with a GET request.

4.9.2 Comments

New in draft 7 \$comment

The \$comment keyword is strictly intended for adding comments to a schema. Its value must always be a string. Unlike the annotations title, description, and examples, JSON schema implementations aren't allowed to attach any meaning or behavior to it whatsoever, and may even strip them at any time. Therefore, they are useful for leaving notes to future editors of a JSON schema, but should not be used to communicate to users of the schema.

4.9.3 Enumerated values

The enum keyword is used to restrict a value to a fixed set of values. It must be an array with at least one element, where each element is unique.

The following is an example for validating street light colors:

```
{
  "enum": ["red", "amber", "green"]
}

//

"red"

//

"blue"
```

You can use enum even without a type, to accept values of different types. Let's extend the example to use null to indicate "off", and also add 42, just for fun.

4.9.4 Constant values

New in draft 6

The const keyword is used to restrict a value to a single value.

For example, if you only support shipping to the United States for export reasons:

```
{ json schema }

{    "properties": {
        "country": {
            "const": "United States of America"
        }
    }
}
```

```
{ "country": "Canada" }
```

4.10 Media: string-encoding non-JSON data

- contentMediaType (page 52)
- contentEncoding (page 52)
- Examples (page 52)

New in draft 7

JSON schema has a set of keywords to describe and optionally validate non-JSON data stored inside JSON strings. Since it would be difficult to write validators for many media types, JSON schema validators are not required to validate the contents of JSON strings based on these keywords. However, these keywords are still useful for an application that consumes validated JSON.

4.10.1 contentMediaType

The contentMediaType keyword specifies the MIME type of the contents of a string, as described in RFC 2046. There is a list of MIME types officially registered by the IANA, but the set of types supported will be application and operating system dependent. Mozilla Developer Network also maintains a shorter list of MIME types that are important for the web

4.10.2 contentEncoding

The contentEncoding keyword specifies the encoding used to store the contents, as specified in RFC 2054, part 6.1.

The acceptable values are 7bit, 8bit, binary, quoted-printable and base64. If not specified, the encoding is the same as the containing JSON document.

Without getting into the low-level details of each of these encodings, there are really only two options useful for modern usage:

- If the content is encoded in the same encoding as the enclosing JSON document (which for practical purposes, is almost always UTF-8), leave contentEncoding unspecified, and include the content in a string as-is. This includes text-based content types, such as text/html or application/xml.
- If the content is binary data, set contentEncoding to base64 and encode the contents using Base64. This would include many image types, such as image/png or audio types, such as audio/mpeg.

4.10.3 Examples

The following schema indicates the string contains an HTML document, encoded using the same encoding as the surrounding document:

```
{ json schema }

{
   "type": "string",
   "contentMediaType": "text/html"
}
```

```
✓
```

"<!DOCTYPE html><html xmlns=\"http://www.w3.org/1999/xhtml\"><head></head></html>"

The following schema indicates that a string contains a PNG image, encoded using Base64:

```
{
    "type": "string",
        "contentEncoding": "base64",
        "contentMediaType": "image/png"
}
```



"iVBORw0KGgoAAAANSUhEUgAAABgAAAAYCAYAAADgdz34AAAABmJLR0QA/wD/AP+gvaeTAAAA..."

4.11 Schema Composition

- *allOf* (page 54)
- *anyOf* (page 54)
- *oneOf* (page 55)
- *not* (page 56)
- Properties of Schema Composition (page 56)
 - Subschema Independence (page 56)
 - *Illogical Schemas* (page 58)
 - Factoring Schemas (page 59)

JSON Schema includes a few keywords for combining schemas together. Note that this doesn't necessarily mean combining schemas from multiple files or JSON trees, though these facilities help to enable that and are described in *Structuring a complex schema* (page 65). Combining schemas may be as simple as allowing a value to be validated against multiple criteria at the same time.

These keywords correspond to well known boolean algebra concepts like AND, OR, XOR, and NOT. You can often use these keywords to express complex constraints that can't otherwise be expressed with standard JSON Schema keywords.

The keywords used to combine schemas are:

- allOf (page 54): (AND) Must be valid against all of the subschemas
- anyOf (page 54): (OR) Must be valid against any of the subschemas
- oneOf (page 55): (XOR) Must be valid against exactly one of the subschemas

All of these keywords must be set to an array, where each item is a schema.

In addition, there is:

• not (page 56): (NOT) Must not be valid against the given schema

4.11.1 allOf

To validate against al10f, the given data must be valid against all of the given subschemas.

Note: *allOf* (page 54) can not be used to "extend" a schema to add more details to it in the sense of object-oriented inheritance. Instances must independently be valid against "all of" the schemas in the allof. See the section on *Subschema Independence* (page 56) for more information.

4.11.2 anyOf

To validate against any 0f, the given data must be valid against any (one or more) of the given subschemas.

4.11.3 oneOf

To validate against oneOf, the given data must be valid against exactly one of the given subschemas.

Not a multiple of either 5 or 3.

```
2
```

Multiple of both 5 and 3 is rejected.

```
15
```

4.11.4 not

The not keyword declares that an instance validates if it doesn't validate against the given subschema.

For example, the following schema validates against anything that is not a string:

```
{ "not": { "type": "string" } }

42

{ "key": "value" }

"I am a string"
```

4.11.5 Properties of Schema Composition

Subschema Independence

It is important to note that the schemas listed in an *allOf* (page 54), *anyOf* (page 54) or *oneOf* (page 55) array know nothing of one another. For example, say you had a schema for an address in a definitions section, and want to "extend" it to include an address type:

```
{ json schema }
{
  "definitions": {
    "address": {
      "type": "object",
      "properties": {
        "street_address": { "type": "string" },
       "city": { "type": "string" },
       "state": { "type": "string" }
      },
      "required": ["street_address", "city", "state"]
    }
  },
  "allOf": [
    { "$ref": "#/definitions/address" },
      "properties": {
        "type": { "enum": [ "residential", "business" ] }
    }
  ]
}
```

```
{
    "street_address": "1600 Pennsylvania Avenue NW",
    "city": "Washington",
    "state": "DC",
    "type": "business"
}
```

This works, but what if we wanted to restrict the schema so no additional properties are allowed? One might try adding the highlighted line below:

```
{ json schema }
  "definitions": {
    "address": {
      "type": "object",
      "properties": {
        "street_address": { "type": "string" },
        "city": { "type": "string" },
        "state": { "type": "string" }
      },
      "required": ["street_address", "city", "state"]
    }
  },
  "allOf": [
   { "$ref": "#/definitions/address" },
    {
      "properties": {
        "type": { "enum": [ "residential", "business" ] }
   }
  ],
 "additionalProperties": false
}
```

```
{
    "street_address": "1600 Pennsylvania Avenue NW",
    "city": "Washington",
    "state": "DC",
    "type": "business"
}
```

Unfortunately, now the schema will reject *everything*. This is because additionalProperties knows nothing about the properties declared in the subschemas inside of the *allOf* (page 54) array.

To many, this is one of the biggest surprises of the combining operations in JSON schema: it does not behave like inheritance in an object-oriented language. There are some proposals to address this in the next version of the JSON schema specification.

Illogical Schemas

Note that it's quite easy to create schemas that are logical impossibilities with these keywords. The following example creates a schema that won't validate against anything (since something may not be both a string and a number at the same time):

Factoring Schemas

Note that it's possible to "factor" out the common parts of the subschemas. The following two schemas are equivalent.

4.12 Applying subschemas conditionally

New in draft 7 if, then and else keywords

The if, then and else keywords allow the application of a subschema based on the outcome of another schema, much like the if/then/else constructs you've probably seen in traditional programming languages.

If if is valid, then must also be valid (and else is ignored.) If if is invalid, else must also be valid (and then is ignored).

If then or else is not defined, if behaves as if they have a value of true.

If then and/or else appear in a schema without if, then and else are ignored.

We can put this in the form of a truth table, showing the combinations of when if, then, and else are valid and the resulting validity of the entire schema:

if	then	else	whole schema
T	T	n/a	T
T	F	n/a	F
F	n/a	T	T
F	n/a	F	F
n/a	n/a	n/a	T

For example, let's say you wanted to write a schema to handle addresses in the United States and Canada. These countries have different postal code formats, and we want to select which format to validate against based on the country. If the address is in the United States, the postal_code field is a "zipcode": five numeric digits followed by an optional four digit suffix. If the address is in Canada, the postal_code field is a six digit alphanumeric string where letters and numbers alternate.

```
{ json schema }
  "type": "object",
  "properties": {
    "street_address": {
      "type": "string"
    },
    "country": {
      "default": "United States of America",
      "enum": ["United States of America", "Canada"]
    }
  },
  "if": {
    "properties": { "country": { "const": "United States of America" } }
  "then": {
    "properties": { "postal_code": { "pattern": "[0-9]{5}(-[0-9]{4})?" } }
  },
  "else": {
    "properties": { "postal_code": { "pattern": "[A-Z][0-9][A-Z] [0-9][A-Z][0-9]" } }
  }
}
```

```
{
    "street_address": "1600 Pennsylvania Avenue NW",
    "country": "United States of America",
    "postal_code": "20500"
}
```

```
{
    "street_address": "1600 Pennsylvania Avenue NW",
    "postal_code": "20500"
}

{
    "street_address": "24 Sussex Drive",
    "country": "Canada",
    "postal_code": "KIM 1M4"
}

{
    "street_address": "24 Sussex Drive",
    "country": "Canada",
    "postal_code": "10000"
}

*

{
    "street_address": "1000 Pennsylvania Avenue NW",
    "postal_code": "KIM 1M4"
}
```

Note: In this example, "country" is not a required property. Because the "if" schema also doesn't require the "country" property, it will pass and the "then" schema will apply. Therefore, if the "country" property is not defined, the default behavior is to validate "postal_code" as a USA postal code. The "default" keyword doesn't have an effect, but is nice to include for readers of the schema to more easily recognize the default behavior.

Unfortunately, this approach above doesn't scale to more than two countries. You can, however, wrap pairs of if and then inside an allof to create something that would scale. In this example, we'll use United States and Canadian postal codes, but also add Netherlands postal codes, which are 4 digits followed by two letters. It's left as an exercise to the reader to expand this to the remaining postal codes of the world.

```
{ json schema }
  "type": "object",
  "properties": {
   "street_address": {
     "type": "string"
    "country": {
     "default": "United States of America",
      "enum": ["United States of America", "Canada", "Netherlands"]
   }
  },
  "allOf": [
   {
     "if": {
       "properties": { "country": { "const": "United States of America" } }
     },
      "then": {
        "properties": { "postal_code": { "pattern": "[0-9]{5}(-[0-9]{4}))?" } }
      }
   },
    {
      "if": {
       "properties": { "country": { "const": "Canada" } },
       "required": ["country"]
      "then": {
        "properties": { "postal_code": { "pattern": "[A-Z][0-9][A-Z] [0-9][A-Z][0-9]" } }
      }
   },
    {
      "if": {
       "properties": { "country": { "const": "Netherlands" } },
       "required": ["country"]
     },
        "properties": { "postal_code": { "pattern": "[0-9]{4} [A-Z]{2}" } }
      }
   }
 ]
}
```

```
{
    "street_address": "1600 Pennsylvania Avenue NW",
    "country": "United States of America",
    "postal_code": "20500"
}
```

```
"street_address": "1600 Pennsylvania Avenue NW",
  "postal_code": "20500"
  "street_address": "24 Sussex Drive",
  "country": "Canada",
  "postal_code": "K1M 1M4"
}
  "street_address": "Adriaan Goekooplaan",
  "country": "Netherlands",
  "postal_code": "2517 JX"
}
  "street_address": "24 Sussex Drive",
  "country": "Canada",
  "postal_code": "10000"
  "street_address": "1600 Pennsylvania Avenue NW",
  "postal_code": "K1M 1M4"
}
```

Note: The "required" keyword is necessary in the "if" schemas or they would all apply if the "country" is not defined. Leaving "required" off of the "United States of America" "if" schema makes it effectively the default if no "country" is defined.

Note: Even if "country" was a required field, it's still recommended to have the "required" keyword in each "if" schema. The validation result will be the same because "required" will fail, but not including it will add noise to error results because it will validate the "postal_code" against all three of the "then" schemas leading to irrelevant errors.

4.13 The \$schema keyword

• Advanced (page 64)

The \$schema keyword is used to declare that a JSON fragment is actually a piece of JSON Schema. It also declares which version of the JSON Schema standard that the schema was written against.

It is recommended that all JSON Schemas have a \$schema entry, which must be at the root. Therefore most of the time, you'll want this at the root of your schema:

"\$schema": "http://json-schema.org/draft/2019-09/schema#"

4.13.1 Advanced

You should declare that your schema was written against a specific version of the JSON Schema standard and include the draft name in the path, for example:

- http://json-schema.org/draft/2019-09/schema#
- http://json-schema.org/draft-07/schema#
- http://json-schema.org/draft-06/schema#
- http://json-schema.org/draft-04/schema#

The possibility to declare \$schema without specific version (http://json-schema.org/schema#) was deprecated after Draft 4 and should no longer be used.

Additionally, if you have extended the JSON Schema language to include your own custom keywords for validating values, you can use a custom URI for \$schema. It must not be one of the predefined values above, and should probably include a domain name you own.

Structuring a complex schema

```
Reuse (page 65)
Recursion (page 67)
The $id property (page 69)
Using $id with $ref (page 70)
Extending (page 70)
```

When writing computer programs of even moderate complexity, it's commonly accepted that "structuring" the program into reusable functions is better than copying-and-pasting duplicate bits of code everywhere they are used. Likewise in JSON Schema, for anything but the most trivial schema, it's really useful to structure the schema into parts that can be reused in a number of places. This chapter will present some practical examples that use the tools available for reusing and structuring schemas.

5.1 Reuse

For this example, let's say we want to define a customer record, where each customer may have both a shipping and a billing address. Addresses are always the same—they have a street address, city and state—so we don't want to duplicate that part of the schema everywhere we want to store an address. Not only would that make the schema more verbose, but it makes updating it in the future more difficult. If our imaginary company were to start doing international business in the future and we wanted to add a country field to all the addresses, it would be better to do this in a single place rather than everywhere that addresses are used.

So let's start with the schema that defines an address:

```
{
  "type": "object",
  "properties": {
    "street_address": { "type": "string" },
```

(continues on next page)

(continued from previous page)

```
"city": { "type": "string" },
    "state": { "type": "string" }
},
    "required": ["street_address", "city", "state"]
}
```

Since we are going to reuse this schema, it is customary (but not required) to put it in the parent schema under a key called definitions:

```
{
  "definitions": {
    "address": {
        "type": "object",
        "properties": {
            "street_address": { "type": "string" },
            "city": { "type": "string" },
            "state": { "type": "string" }
        },
        "required": ["street_address", "city", "state"]
    }
}
```

We can then refer to this schema snippet from elsewhere using the \$ref keyword. The easiest way to describe \$ref is that it gets logically replaced with the thing that it points to. So, to refer to the above, we would include:

```
{ "$ref": "#/definitions/address" }
```

This can be used anywhere a schema is expected. You will always use \$ref as the only key in an object: any other keys you put there will be ignored by the validator.

The value of \$ref is a URI-reference, and the part after # sign (the "fragment" or "named anchor") is in a format called JSON Pointer.

Note: JSON Pointer aims to serve the same purpose as XPath from the XML world, but it is much simpler.

If you're using a definition from the same document, the \$ref value begins with the pound symbol (#). Following that, the slash-separated items traverse the keys in the objects in the document. Therefore, in our example "#/definitions/address" means:

- 1) go to the root of the document
- 2) find the value of the key "definitions"
- 3) within that object, find the value of the key "address"

\$ref can resolve to a URI that references another file, so if you prefer to include your definitions in separate files, you can also do that. For example:

```
{ "$ref": "definitions.json#/address" }
```

would load the address schema from another file residing alongside this one.

Now let's put this together and use our address schema to create a schema for a customer:

```
{ json schema }
{
  "$schema": "http://json-schema.org/draft-07/schema#",
  "definitions": {
    "address": {
      "type": "object",
      "properties": {
        "street_address": { "type": "string" },
       "city": { "type": "string" },
                         { "type": "string" }
        "state":
      },
      "required": ["street_address", "city", "state"]
  },
  "type": "object",
  "properties": {
    "billing_address": { "$ref": "#/definitions/address" },
    "shipping_address": { "$ref": "#/definitions/address" }
 }
}
```

```
{
    "shipping_address": {
        "street_address": "1600 Pennsylvania Avenue NW",
        "city": "Washington",
        "state": "DC"
},
    "billing_address": {
        "street_address": "1st Street SE",
        "city": "Washington",
        "state": "DC"
}
```

Note: Even though the value of a \$ref is a URI-reference, it is not a network locator, only an identifier. This means that the schema doesn't need to be accessible at the resolved URI, but it may be. It is basically up to the validator implementation how external schema URIs will be handled, but one should not assume the validator will fetch network resources indicated in \$ref values.

5.1.1 Recursion

\$ref elements may be used to create recursive schemas that refer to themselves. For example, you might have a person schema that has an array of children, each of which are also person instances.

5.1. Reuse 67

```
{ json schema }
  "$schema": "http://json-schema.org/draft-07/schema#",
  "definitions": {
    "person": {
      "type": "object",
      "properties": {
        "name": { "type": "string" },
        "children": {
          "type": "array",
          "items": { "$ref": "#/definitions/person" },
          "default": []
      }
   }
  },
  "type": "object",
  "properties": {
    "person": { "$ref": "#/definitions/person" }
  }
}
```

A snippet of the British royal family tree

```
{
  "person": {
    "name": "Elizabeth",
    "children": [
        "name": "Charles",
        "children": [
          {
            "name": "William",
            "children": [
              { "name": "George" },
                "name": "Charlotte" }
          },
          {
            "name": "Harry"
        ]
      }
    ]
  }
}
```

Above, we created a schema that refers to another part of itself, effectively creating a "loop" in the validator, which is both allowed and useful. Note, however, that a loop of \$ref schemas referring to one another could cause an infinite loop in the resolver, and is explicitly disallowed.

5.2 The \$id property

The \$id property is a URI-reference that serves two purposes:

- It declares a unique identifier for the schema.
- It declares a base URI against which \$ref URI-references are resolved.

It is best practice that every top-level schema should set \$id to an absolute-URI (not a relative reference), with a domain that you control. For example, if you own the foo.bar domain, and you had a schema for addresses, you may set its \$id as follows:

```
{ json schema }

{ "$id": "http://foo.bar/schemas/address.json" }
```

This provides a unique identifier for the schema, as well as, in most cases, indicating where it may be downloaded.

But be aware of the second purpose of the \$id property: that it declares a base URI for \$ref URI-references elsewhere in the file. For example, if you had:

```
{ json schema }

{ "$ref": "person.json" }
```

in the same file, a JSON schema validation library that supported network fetching may fetch person.json from http://foo.bar/schemas/person.json, even if address.json was loaded from somewhere else, such as the local filesystem. The drafts do not define this area of behaviour very clearly, and validator implementations may vary in exactly how they try to locate the referenced schema.

New in draft 6

```
Draft 4
```

In Draft 4, \$id is just id (without the dollar sign).

5.2. The \$id property 69

The \$id property should never be the empty string or an empty fragment (#), since that doesn't really make sense.

5.2.1 Using \$id with \$ref

\$id also provides a way to refer to subschema without using JSON Pointer. This means you can refer to them by a unique name, rather than by where they appear in the JSON tree.

Reusing the address example above, we can add an \$id property to the address schema, and refer to it by that instead.

```
{ json schema }
  "$schema": "http://json-schema.org/draft-07/schema#",
  "definitions": {
    "address": {
      "$id": "#address",
      "type": "object",
      "properties": {
        "street_address": { "type": "string" },
        "city":
                          { "type": "string" },
        "state":
                          { "type": "string" }
      "required": ["street_address", "city", "state"]
    }
  },
  "type": "object",
  "properties": {
    "billing_address": { "$ref": "#address" },
    "shipping_address": { "$ref": "#address" }
  }
}
```

Note: This functionality isn't currently supported by the Python jsonschema library.

5.3 Extending

The power of \$ref really shines when it is used with the combining keywords allof, anyof and oneOf (see *Schema Composition* (page 53)).

Let's say that for a shipping address, we want to know whether the address is a residential or business address, because the shipping method used may depend on that. For a billing address, we don't want to store that information, because it's not applicable.

To handle this, we'll update our definition of shipping address:

```
"shipping_address": { "$ref": "#/definitions/address" }
```

to instead use an allOf keyword entry combining both the core address schema definition and an extra schema snippet for the address type:

Tying this all together,

```
{ json schema }
  "$schema": "http://json-schema.org/draft-06/schema#",
  "definitions": {
    "address": {
     "type": "object",
      "properties": {
        "street_address": { "type": "string" },
       "city": { "type": "string" },
                         { "type": "string" }
        "state":
      "required": ["street_address", "city", "state"]
   }
  },
  "type": "object",
  "properties": {
    "billing_address": { "$ref": "#/definitions/address" },
    "shipping_address": {
      "allOf": [
       { "$ref": "#/definitions/address" },
        { "properties":
         { "type": { "enum": [ "residential", "business" ] } },
          "required": ["type"]
        }
      ]
   }
  }
}
```

This fails, because it's missing an address type:

5.3. Extending 71

```
{
    "shipping_address": {
        "street_address": "1600 Pennsylvania Avenue NW",
        "city": "Washington",
        "state": "DC"
    }
}
```

```
{
    "shipping_address": {
        "street_address": "1600 Pennsylvania Avenue NW",
        "city": "Washington",
        "state": "DC",
        "type": "business"
    }
}
```

From these basic pieces, it's possible to build very powerful constructions without a lot of duplication.

CHAPTER 6

Acknowledgments

Michael Droettboom wishes to thank the following contributors:

- Alexander Kjeldaas
- Alexander Lang
- Anders D. Johnson
- Armand Abric
- Ben Hutton
- Brandon Wright
- Brent Tubbs
- Chris Carpenter
- Christopher Mark Gore
- David Branner
- · David Michael Karr
- David Worth
- E. M. Bray
- Fenhl
- forevermatt
- goldaxe
- Henry Andrews
- Hervé
- Hongwei
- Jesse Claven
- Koen Rouwhorst

- Mike Kobit
- Oliver Kurmis
- Sam Blackman
- Vincent Jacques

Symbols	default, 49
\$comment, 50	dependencies, 37
\$id, 69	description, 49
\$ref,66	Е
\$schema, 63	
A additionalItems, 41, 44 additionalProperties, 32 allOf, 54 annotation, 49 anyOf, 54	else, 59 email, 21 enum, 50 enumerated values, 50 examples, 49 exclusiveMaximum, 26 exclusiveMinimum, 26
array, 40	F
contains, 45 items, 41 length, 46 list validation, 42 tuple validation, 42 tuple validation; additionalItems, 44 uniqueness, 46	format, 20 date, 20 date-time, 20 email, 21 hostname, 21 idn-email, 21 idn-hostname, 21
В	ipv4, 21
boolean, 47	ipv6, 21
C comment, 50 conditionals, 59 else, 59 if, 59 then, 59 const, 51 constant values, 51 contains, 41, 45 contentEncoding, 52	iri, 21 iri-reference, 21 json-pointer, 21 regex, 22 relative-json-pointer, 21 time, 20 uri, 21 uri-reference, 21 uri-template, 21 H hostname, 21
contentMediaType, 52	
D	l
date, 20 date-time, 20	<pre>idn-email, 21 idn-hostname, 21 if, 59</pre>

integer, 23	S
ipv4, 21	schema
ipv6, 21	keyword, 63
iri,21	schema composition, 53
iri-reference,21	all0f, 54
items, 41	anyOf, 54
	not, 56
	oneOf, 55
json-pointer,21	subschema independence, 56
	string, 17
M	format, 20
maximum, 26	length, 18
maxItems, 46	regular expression, 19
maxLength, 18	structure, 64
maxProperties, 36	Structure, or
media, 52	T
contentEncoding, 52	then, 59
contentMediaType, 52	time, 20
minimum, 26	title, 49
minItems, 46	type, 15
minLength, 18	types
minProperties, 36	basic, 15
multipleOf, 25	
	numeric, 23
N	U
non-JSON data, 52	uniqueItems, 46
not, 56	uri, 21
nul1, 48	uri-reference, 21
number, 23	uri-template, 21
multiple of, 25	ui i tempiate, 21
range, 26	W
	writeOnly,49
0	wificeonly, 49
object, 29	
dependencies, 37	
properties, 30, 32	
properties; regular expression, 31	
property names, 36	
required properties, 34	
size, 36	
one0f, 55	
Р	
pattern, 19	
patternProperties,31	
properties, 30	
propertyNames, 36	
R	
readOnly, 49	
regex, 22	
regular expressions, 22	
relative-json-pointer,21	
required 34	

76 Index