Introduction to JSONiq

The SQL of NoSQL

Ghislain Fourny <g@28.io>

Introduction to JSONiq: The SQL of NoSQL

by Ghislain Fourny

Abstract

This book introduces the JSONiq language, which is used to query NoSQL databases such as MongoDB.

Table of Contents

List of Figures

List of Examples

Chapter 1. Introduction

Why JSONiq?

JSONiq is a query and processing language specifically designed for the popular JSON data model. The main ideas behind JSONiq are based on lessons learned in more than 30 years of relational query systems and more than 15 years of experience with designing and implementing query languages for semi-structured data like XML and RDF.

The main source of inspiration behind JSONiq is XQuery, which has been proven so far a successful and productive query language for semi-structured data (in particular XML). JSONiq borrowed a large numbers of ideas from XQuery, like the structure and semantics of a FLWOR construct, the functional aspect of the language, the semantics of comparisons in the face of data heterogeneity, the declarative, snapshot-based updates. However, unlike XQuery, JSON is not concerned with the peculiarities of XML, like mixed content, ordered children, the confusion between attributes and elements, the complexities of namespaces and QNames, or the complexities of XML Schema, and so on.

The power of the XQuery's FLWOR construct and the functional aspect, combined with the simplicity of the JSON data model result in a clean, sleek and easy to understand data processing language. As a matter of fact, JSONiq is a language that can do more than queries: it can describe powerful data processing programs, from transformations, selections, joins of heterogeneous data sets, data enrichment, information extraction, information cleaning, and so on.

Technically, the main characteristics of JSONiq (and XQuery) are the following:

- It is a *set-oriented language*. While most programming languages are designed to manipulate one object at a time, JSONiq is designed to process sets (actually, sequences) of data objects.
- It is a *functional language*. A JSONiq program is an expression; the result of the program is the result of the evaluation of the expression. Expressions have fundamental role in the language: every language construct is an expression, and expressions are fully composable.
- It is a *declarative language*. A program specifies what is the result being calculated, and does not specify low level algorithms like the sort algorithm, the fact that an algorithm is executed in main memory or is external, on a single machine or parallelized on several machines, or what access patterns (aka indexes) are being used during the evaluation of the program. Such implementation decisions should be taken automatically, by an optimizer, based on the physical characteristics of the data, and of the hardware environment. Just like a traditional database would do. The language has been designed from day one with optimizability in mind.
- It is designed for *nested, heterogeneous, semi-structured data*. Data structures in JSON can be nested with arbitrary depth, do not have a specific type pattern (i.e. are heterogeneous), and may or may not have one or more schemas that describe the data. Even in the case of a schema, such a schema can be open, and/or simply partially describe the data. Unlike SQL, which is designed to query tabular, flat, homogeneous structures. JSONiq has been designed from scratch as a query for nested and heterogeneous data.

Structure of a JSONiq program.

Figure 1.1. Module

Figure 1.2. MainModule

A main JSONiq program is made of two parts: an optional prolog, and an expression, which is the main query.

The result of the main JSONiq program is the result of its main query.

In the prolog, it is possible to declare global variables and functions. Mostly, you will recognize a prolog declaration by the semi-colon it ends with. The main query does not contain semi-colons (at least in core JSONiq).

Figure 1.3. LibraryModule

Library modules do not contain any main query, just global variables and functions. They can be imported by other modules.

Part I. JSON and the JSONiq Data Model

Table of Contents

Chapter 2. The JSON Syntax

The JSONiq query language was specifically designed for querying and processing JSON.

As stated on its home http://www.json.org, JSON is a “ lightweight data-interchange format. It is easy for humans to read and write. It is easy for machines to parse and generate. ”

JSON itself is only about syntax: a string may or may not match the JSON grammar. If it does, then it is well-formed JSON. The JSON syntax is made of the following building blocks: objects, arrays, strings, numbers, booleans and nulls. Let us begin with a quick overview of all these building blocks.

JSON Strings

Strings are double-quoted. To put it simply, they are sequences of any Unicode characters with absolutely no restriction:

```
"foo"
"What NoSQL solutions are out there?"
```

However, syntactically, some of these characters must be escaped with backslashes (escape sequence). This includes double quotes, escaped as \" -- because otherwise they could be confused with the end of a string -- and backslahes themselves, escaped as \\ -- because otherwise you would not know if you mean a backslash character, or if you are escaping the following character.

```
"What \"NoSQL\" solutions are out there?"
```

Finally, all Unicode control characters (null, new line, line feed, delete...) are not allowed directly and must be built with an escape sequence. Any Unicode character, including control characters, can be built with \u followed by the four hexadecimal digits that identify it within Unicode. The most frequent control characters even have their own shortcuts: \n (new line), \t (tab), \r (carriage return), \b (backspace), \f (line feed). The slash also can be obtained with \/, although it is fine too if it appears alone. This is useful in JSON-hosting environments where slashes are special.

```
.
"What \"NoSQL\" solutions are out there:\n"
"MapReduce\u000AMongoDB\n\u0085"
```

JSON Numbers

Numbers cover the entire decimal space. There is no range restriction. Although there is no formal distinction in JSON, Numbers can be grouped in three subcategories. This subcategories play an important role in JSONiq.

- Integers, possibly with a negative sign and not beginning with a leading 0:

```
0
9
42
-96
123456789012345678901234567890123456789012345678901234
```

- "Plain" decimals, with a dot, both followed and preceded by at least by one digit (no leading dot):

  ```
  0.3
  9.6
  42.2346902834
  -96.01345023400
  ```

- Decimals in scientific notation, i.e., a plain decimal followed by an E (case does not matter) and by a power of ten (an integer with an optional sign):

  ```
  0.3e0
  9.6E+24
  42.2346902834e-2
  -96.01345023400E-02345
  ```

JSON Booleans

Booleans cover the two logical truth values true and false - unquoted. There is not much more to say about them...

```
true
false
```

JSON Null

Null is a special value that can be used to denote the absence of value.

```
null
```

JSON Objects

Objects are unordered sets of string/value pairs. A string is any JSON string as described above. A value is any JSON building block.

According to the JSON RFC, keys (the strings) should be unique -- and JSONiq does consider them unique.

You can see in the following examples that values can be also nested objects or arrays.

```
{
"_id" : "511C7C5C9A277C22D138802D",
"question_id" : 4419499,
"last_edit_date" : "2012-12-17T00:02:31",
"creation_date" : "2010-12-11T23:15:19",
"last_activity_date" : "2012-12-17T00:02:31",
"score" : 15,
"accepted_answer_id" : 4421601,
"title" : "MySQL and NoSQL: Help me to choose the right one",
```

```
"tags" : [ "php", "mysql", "nosql", "cassandra" ],
"view_count" : 3972,
"owner" : {
  "user_id" : 279538,
  "display_name" : "cedivad",
  "reputation" : 430,
  "user_type" : "registered",
  "profile_image" : "http://www.gravatar.com/avatar/b77f...",
  "link" : "http://stackoverflow.com/users/279538/cedivad",
  "accept_rate" : 74
},
"link" : "http://stackoverflow.com/questions/4419499/mys...",
"is_answered" : true
}
```

In the NoSQL world, top-level JSON objects are often referred to as JSON documents.

Chapter 3. The JSONiq Data Model

Having a JSON document as pure syntax is not very useful in itself, except to send it over a network or to store it in a document store of course. To make use of it in a database or in other processing environments, you need to bring it to a higher level of abstraction and give semantics to the building blocks. This is what a Data Model is for.

We now introduce the JSONiq data model.

Let us begin with some good news first: the JSON syntax that we have just introduced is a subset of JSONiq. Concretely, this means that any of these syntactic JSON building blocks can be copy-and-pasted, and executed as a JSONiq query. The output will be the counterpart of this JSON building block in the Data Model. So, if you are familiar with JSON, then you already know some JSONiq.

JSONiq Values: Items and Sequences

In JSONiq, the JSON building blocks described in the former section, on a more abstract level, are referred to as items. JSONiq manipulates sequences of these items. Hence, a JSONiq value is a sequence of items. So, in particular, a JSONiq query returns sequences of items. Actually, even inside a JSONiq query, sequences of items are passed around between the JSONiq building blocks internal to a query (called expressions).

Let us copy-and-paste a JSON Object and execute it as JSONiq:

Example 3.1. A sequence of just one item

```
{ "foo" : "bar" }
```

The above query generates a sequence of one item, an object item in this case. The result displayed above is the output of this query when run with the Zorba query processor, which is one of the JSONiq implementations.

Commas are all you need to begin building your own sequences. You can mix and match!

Example 3.2. A sequence of various items

```
"foo", 2, true, { "foo", "bar" }, null, [ 1, 2, 3 ]
```

There are three golden rules about sequences that are useful to keep in mind.

Rule #1: Sequences are flat and cannot be nested. This makes streaming possible, which is very powerful.

Example 3.3. Sequences are flat

```
( ("foo", 2), ( (true, 4, null), 6 ) )
```

Rule #2: A sequence can be empty. The empty sequence can be constructed with empty parentheses.

Example 3.4. The empty sequence

```
()
```

Rule #3: A sequence of just one item is considered the same as just this item. Whenever we say that an expression returns or takes one item, we really mean that it takes a singleton sequence of one item.

Example 3.5. A sequence of one item

```
("foo")
```

JSONiq classifies the items mentioned above in three categories:

- Objects: the counterparts of the syntactic JSON objects.
- Arrays: the counterparts of the syntactic JSON arrays.
- Atomics: the counterparts of JSON strings, JSON numbers, JSON booleans and JSON nulls - but with a very rich type system which includes dates, for example.

Objects

An object represents a JSON object: an unordered collection of string/item pairs.

Each pair consists of an atomic of type *string* and of an item which can be in any category.

No two pairs have the same name. Because of this, the word *field* is also used to refer to pairs.

Arrays

An array represents a JSON array: an ordered list of items -- items in any category.

An array can be seen as a sequence that is wrapped in one single item. And since an array is an item, arrays can nest -- like in JSON.

Atomics

An atomic is a non-structured value that is annotated with a type.

JSONiq defines many useful builtin atomic types. For now, let us introduce those that have a JSON counterpart. Note that JSON numbers correspond to three different types in JSONiq.

- *string*: all JSON strings.
- *integer*: all JSON numbers that are integers (no dot, no exponent), infinite range.
- *decimal*: all JSON numbers that are decimals (no exponent)
- *double*: IEEE double-precision 64-bit floating point numbers (corresponds to JSON numbers with an exponent).
- *boolean*: the JSON booleans true and false.
- *null*: the JSON null.

JSONiq also offers many other types of atomics. Here is a little appetizer that showcases constructing a date, and then extracting the month with a function. All these features will be shown in details later in this book.

Example 3.6. Atomics with the type date

```
month-from-date(date("2013-06-18"))
```

Chapter 4. The JSONiq Type System

JSONiq manipulates semi-structured data: in general, JSONiq allows you, but does not require you to specify types. So you have as much or as little type verification as you wish.

Like in Java or C++, it is possible to create a variable with a given static type:

Example 4.1. Specifying a type

```
let $x as integer := 16
return $x * $x
```

Like in JavaScript, it is possible to create a variable without explicitly giving any static type. JSONiq is still strongly typed, so that you will be told if there is a type inconsistency or mismatch in your programs.

Example 4.2. Not specifying a type

```
let $x := 16
return $x * $x
```

Variables will be explained later more in details.

JSONiq supports types at the sequence level. They are called sequence types, and the syntax for designing types is called the sequence type syntax. The type "integer" that was shown above in a query matches singleton sequences of one atomic item of type integer.

Whenever you do not specify the type of a variable or the type signature of a function, the most general type for any sequence of items, *item**, is assumed. But it is not forbidden for the processor to be smart and warn you if it can detect that a type issue can arise at runtime.

There are many JSONiq expressions (cast, instance of, ...) which perform type operations and that make use of the sequence type syntax. In the remainder of this section, we will introduce sequence types using an "instance of" expression that returns true or false depending on whether or not the type on the right side is matched by the value on the left side -- like in Java.

Example 4.3. The instance of operator

```
16 instance of integer
```

Item Types

Atomic Types

Atomic types are organized in a tree hierarchy.

JSONiq defines the following build-in types that have a direct relation with JSON:

- *string*: the value space is all strings made of Unicode characters.

  All string literals build an atomic which matches string.

- *integer*: the value space is that of all mathematical integral numbers (N), with an infinite range. This is a subtype of *decimal*, so that all integers also match the item type *decimal*.

  All integer literals build an atomic which matches integer.

- *decimal*: the value space is that of all mathematical decimal numbers (D), with an infinite range.

  All decimal literals build an atomic which matches decimal.

- *double*: the value space is that of all IEEE double-precision 64-bit floating point numbers.

  All double literals build an atomic which matches double.

- *boolean*: the value space contains the booleans true and false.

  All boolean literals build an atomic which matches boolean.

- *null*: the value space is a singleton and only contains null.

  All null literals build an atomic which matches null.

- *atomic*: all atomic types.

  All literals build an atomic which matches atomic.

Example 4.4. Atomic types

```
16 instance of integer,
16 instance of decimal,
16.6 instance of decimal,
16.6e10 instance of double,
"foo" instance of string,
true instance of boolean,
null instance of null,
"foo" instance of atomic
```

JSONiq also supports further atomic types, which were borrowed from XML Schema 1.1.

These datatypes are already used as a set of atomic datatypes by the other two semi-structured data formats of the Web: XML and RDF, as well as by the corresponding query languages: XQuery and SPARQL, so it is natural for a complete JSON data model to reuse them.

- Further number types: long, int, short, byte, float.
- Date or time types: date, dateTime, dateTimeStamp, gDay, gMonth, gMonthDay, gYear, gYearMonth, time.
- Duration types: duration, dayTimeDuration, yearMonthDuration.
- Binary types: base64Binary, hexBinary.
- An URI type: anyURI.

Atomic items that have these builtin atomic types can only be built with a constructor -- again similar to JavaScript.

Example 4.5. Further builtin atomic types

```
date("2013-06-18") instance of date,
long("1234567890123") instance of long
```

JSON Item Types : Object Types and Array Types

All objects match the item type *object* as well as *json-item*.

All arrays match the item type *array* as well as *json-item*.

Example 4.6. Further builtin atomic types

```
{ "foo" : "bar" } instance of object,
{ "foo" : "bar" } instance of json-item,
{} instance of object,
[ 1, 2, 3, 4 ] instance of array,
[ 1, 2, 3, 4 ] instance of json-item
```

The Most General Item Type.

All items match the item type *item*.

Example 4.7. The most general item type: item

```
{ "foo" : "bar" } instance of item,
[ 1, 2, 3, 4 ] instance of item,
"foo" instance of item,
42 instance of item,
false instance of item,
null instance of item
```

Sequence Types

All sequences match the sequence type *item**.

Example 4.8. The most general sequence type: item*

```
{ "foo" : "bar" } instance of item*,
() instance of item*,
([ 1, 2, 3 ], 2, { "foo" : "bar" }, 4) instance of item*
```

But sequence types can be much more precise than that. In general, a sequence type is made of an item type, as presented above, followed by an occurrence indicator among the following:

- * stands for a sequence of any length (zero or more)
- + stands for a non-empty sequence (one or more)
- ? stands for an empty or a singleton sequence (zero or one)
- The absence of indicator stands for a singleton sequence (one).

Example 4.9. Further sequence types

```
( { "foo" : "bar" } , {} ) instance of object*,
() instance of object*,
( [ 1, 2, 3 ] , {} ) instance of json-item+,
[ 1, 2, 3 ] instance of array?,
() instance of array?,
"foo" instance of string
```

There is also a special type that matches only empty sequences, denoted () as well:

Example 4.10. Empty sequence type: ()

```
() instance of ()
```

Part II. Construction of Items and JSON Navigation

Table of Contents

Chapter 5. Construction of Items

As we just saw, the items (objects, arrays, strings, ...) mentioned in the former section are constructed exactly as they are constructed in JSON. In a way, any JSON building block is also a well-formed JSONiq query which just "returns itself" (more precisely: its counterpart in the JSONiq Data Model)!

Atomic Literals

String Literals

The syntax for creating strings is identical to that of JSON. No surprise here. JSON's backslash escaping is supported, and like in JSON, double quotes are required and single quotes are forbidden.

Example 5.1. String literals

```
  "foo",
  "This is a line\nand this is a new line",
  "\u0001",
  "This is a nested \"quote\""
```

Number Literals

The syntax for creating numbers is identical to that of JSON.

Example 5.2. Integer literals

```
42,
3.14,
-6.022E23
```

Well, not quite. Actually, JSONiq allows a more flexible superset. In particular:

- leading 0s are allowed
- a decimal literal can begin with a dot
- a number may begin with a + sign

Example 5.3. A more general literal syntax

```
042,
.1415926535,
+6.022E23
```

Remember that JSONiq distinguishes between integers (no dot, no scientific notation), decimals (dot but no scientific notation), and doubles (scientific notation). As expected, an integer literal creates an atomic of type integer, and so on. No surprises either.

Boolean and Null Literals

There is not much to say actually -- boolean literals build boolean atomics, the null literal builds a null atomic, so no worries here, the world is in order. You might as well want to move to the next section.

Example 5.4. Boolean and Null literals

```
true,
false,
null
```

Object Constructors

The syntax for creating objects is also identical to that of JSON. You can use for an object key any string literal, and for an object value any literal, object constructor or array constructor.

Example 5.5. Object Constructors

```
{},
{ "foo" : "bar" },
{ "foo" : [ 1, 2, 3, 4, 5, 6 ] },
{ "foo" : true, "bar" : false },
{ "this is a key" : { "value" : "a value" } }
```

Again, JSONiq is more flexible here. Like in JavaScript, if your key is simple enough (like alphanumerics, underscores, dashes, this kind of things), you are welcome to omit the quotes. The strings for which quotes are not mandatory are called *unquoted names*. This class of strings can be used for unquoted keys, but also in later sections for variable and function names, and for module aliases.

Example 5.6. Object constructors with unquoted keys

```
{ foo : "bar" },
{ foo : [ 1, 2, 3, 4, 5, 6 ] },
{ foo : "bar", bar : "foo" },
{ "but you need the quotes here" : null }
```

Array Constructors

The syntax for creating arrays is identical to that of JSON (do you sense a growing feeling that we are repeating ourselves? But it feels so good to say it): square brackets, comma separated values.

Example 5.7. Empty array constructors

```
[],
[ 1, 2, 3, 4, 5, 6 ],
[ "foo", 3.14, [ "Go", "Boldly" ], { "foo" : "bar" }, true ]
```

Square brackets are mandatory. Things can only be pushed so far.

Composing Constructors

Of course, JSONiq would not be very interesting if all you could do is copy and paste JSON documents. So now is time to get to the meat.

Because JSONiq expressions are fully composable, in objects and arrays constructors, you can put way more than just atomic literals, object constructors and array constructors: you can put any JSONiq *expression*. An expression is the JSONiq building block. You already know some (literals, constructors, comma, cast, instance of) and plenty more will be introduced in the next part (arithmetics, logic, comparison, if-then-else, try-catch, FLWORS that allow you to join, select, group, filter, project, stream in windows, ...)

In order to illustrate composability, the following examples use a few of many operators you can use:

- "to" for creating sequences of consecutive integers,
- "||" for concatenating strings,
- "+" for adding numbers,
- "," for appending sequences (yes, you already know this one).

So here we go.

In an array, the operand expression inside the square bracket will evaluated to a sequence of items, and these items will be copied and become members of the newly created array.

Example 5.8. Composable array constructors

```
[ 1 to 10 ],
[ "foo" || "bar", 1 to 3, 2 + 2 ]
```

In an object, the expression you use for the key must evaluate to an atomic - if it is not a string, it will just get cast to it.

An error is raised if the key expressions is not an atomic.

Example 5.9. Composable object keys

```
  { "foo" || "bar" : true },
  { 1 + 1 : "foo" }
```

And do not worry about the value expression: if it is is empty, null will be used as a value, and if it contains two items or more, they will be wrapped into an array.

Example 5.10. Composable object values

```
  { "foo" : 1 + 1 },
  { "foo" : (), "bar" : (1, 2) }
```

The {| |} syntax can be used to merge several objects.

Example 5.11. Merging object constructor

```
  {| { "foo" : "bar" }, { "bar" : "foo" } |}
```

An error is raised if the operand expression does not evaluate to a sequence of objects.

Chapter 6. Collections

Even though you can build your own JSON values with JSONiq by copying-and-pasting JSON documents, most of the time, your JSON data will be in a collection.

We now introduce collections, because collections are perfect to illustrate the JSON navigation syntax which will be introduced in the next section.

Collections are sequences of objects, identified by a name which is a string.

Adding or deleting collections from the set of known collections to a query processor, and loading the data in a collection are implementation-dependent and outside of the scope of this book.

We will just assume that there is a function named collection() that returns all objects associated with the provided collection name.

Example 6.1. Getting all objects from a collection

```
collection("one-object")
```

Collections Used Throughout This Book

For illustrative purposes, we will assume that we have the following collections:

- *collection("one-object")*

```
{ "question" : "What NoSQL technology should I use?" }
```

- *collection("faq")* - this is a collection of StackOverflow FAQs.

```
{
  "_id" : "511C7C5C9A277C22D138802D",
  "question_id" : 4419499,
  "last_edit_date" : "2012-12-17T00:02:31",
  "creation_date" : "2010-12-11T23:15:19",
  "last_activity_date" : "2012-12-17T00:02:31",
  "score" : 15,
  "accepted_answer_id" : 4421601,
  "title" : "MySQL and NoSQL: Help me to choose the right one",
  "tags" : [ "php", "mysql", "nosql", "cassandra" ],
  "view_count" : 3972,
  "owner" : {
    "user_id" : 279538,
    "display_name" : "cedivad",
    "reputation" : 430,
    "user_type" : "registered",
    "profile_image" : "http://www.gravatar.com/avatar/b77fadd2ba791...",
    "link" : "http://stackoverflow.com/users/279538/cedivad",
    "accept_rate" : 74
  },
  "link" : "http://stackoverflow.com/questions/4419499/mysql-and-n...",
  "is_answered" : true
```

```
}
{
  "_id" : "511C7C5C9A277C22D138802F",
  "question_id" : 282783,
  "last_edit_date" : "2012-04-30T22:43:02",
  "creation_date" : "2008-11-12T02:02:42",
  "last_activity_date" : "2012-04-30T22:43:02",
  "score" : 42,
  "accepted_answer_id" : 282813,
  "title" : "The Next-gen Databases",
  "tags" : [ "sql", "database", "nosql", "non-relational-database" ],
  "view_count" : 5266,
  "owner" : {
    "user_id" : 3932,
    "display_name" : "Randin",
    "reputation" : 585,
    "user_type" : "registered",
    "profile_image" : "http://www.gravatar.com/avatar/d9d7ba9c17d671d...",
    "link" : "http://stackoverflow.com/users/3932/randin",
    "accept_rate" : 100
  },
  "link" : "http://stackoverflow.com/questions/282783/the-next-gen-data...",
  "is_answered" : true
}
...
```

- *collection("answers")* - this is a collection of StackOverflow answers (to the previous FAQ).

```
{
  "_id" : "511C7C5D9A277C22D13880C3",
  "question_id" : 37823,
  "answer_id" : 37841,
  "creation_date" : "2008-09-01T12:14:38",
  "last_activity_date" : "2008-09-01T12:14:38",
  "score" : 7,
  "is_accepted" : false,
  "owner" : {
    "user_id" : 2562,
    "display_name" : "Ubiguchi",
    "reputation" : 1871,
    "user_type" : "registered",
    "profile_image" : "http://www.gravatar.com/avatar/00b87a917ec763c0c...",
    "link" : "http://stackoverflow.com/users/2562/ubiguchi"
  }
}
{
  "_id" : "511C7C5D9A277C22D13880C4",
  "question_id" : 37823,
  "answer_id" : 37844,
  "creation_date" : "2008-09-01T12:16:40",
  "last_activity_date" : "2008-09-01T12:16:40",
  "score" : 4,
  "is_accepted" : false,
  "owner" : {
    "user_id" : 2974,
    "display_name" : "Rob Wells",
```

```
      "reputation" : 17543,
      "user_type" : "registered",
      "profile_image" : "http://www.gravatar.com/avatar/8769281d99f8fe9c2...",
      "link" : "http://stackoverflow.com/users/2974/rob-wells",
      "accept_rate" : 94
    }
  }
  {
    "_id" : "511C7C5F9A277C22D1388211",
    "question_id" : 4419499,
    "answer_id" : 4419542,
    "creation_date" : "2010-12-11T23:24:21",
    "last_edit_date" : 1292112046,
    "last_activity_date" : "2010-12-12T00:00:46",
    "score" : 17,
    "is_accepted" : false,
    "owner" : {
      "user_id" : 236047,
      "display_name" : "Victor Nicollet",
      "reputation" : 14632,
      "user_type" : "registered",
      "profile_image" : "http://www.gravatar.com/avatar/e083220ac33b47364d3...",
      "link" : "http://stackoverflow.com/users/236047/victor-nicollet",
      "accept_rate" : 95
    }
  }
  {
    "_id" : "511C7C5F9A277C22D1388212",
    "question_id" : 4419499,
    "answer_id" : 4419578,
    "creation_date" : "2010-12-11T23:30:42",
    "last_activity_date" : "2010-12-11T23:30:42",
    "score" : 1,
    "is_accepted" : false,
    "owner" : {
      "user_id" : 510782,
      "display_name" : "descent89",
      "reputation" : 33,
      "user_type" : "registered",
      "profile_image" : "http://www.gravatar.com/avatar/d15c0949f7e051e9d86...",
      "link" : "http://stackoverflow.com/users/510782/descent89"
    }
  }
  ...
```

Many queries on this book can be directly input into 28.io's try-it-now sandbox, as these collections are preloaded (this is real-world data).

Chapter 7. JSON Navigation

Like in JavaScript or SQL or Java, it is possible to navigate through data.

JSONiq supports:

- Looking up the value of a field (given its string key) in an object.
- Looking up the item at a given position (integer) in an array.
- Filtering items from a sequence, retaining only the items that match a given criterium.

Object Navigation

The simplest way to navigate in an object is similar to JavaScript. This will work as soon as you do not push it too much: alphanumerical characters, dashes, underscores - just like unquoted keys in object constructors, any NCName is allowed.

Example 7.1. Object lookup

```
{ "question" : "What NoSQL technology should I use?" }.question
```

Since JSONiq expressions are composable, you can also use any expression for the left-hand side. You might need parentheses depending on the precedence.

Example 7.2. Lookup on a single-object collection.

```
collection("one-object").question
```

The dot operator does an implicit mapping on the left-hand-side, i.e., it applies the lookup in turn on each item. Lookup on any item which is not an object (arrays and atomics) results in the empty sequence.

Example 7.3. Object lookup with an iteration on several objects

```
({ "foo" : "bar" }, { "foo" : "bar2" } ).foo,
{ "ids" : collection("faq").question_id }
```

Example 7.4. Object lookup on non-objects

```
"foo".foo,
({ "question" : "What NoSQL technology should I use?" },
 [ "question", "answer" ],
 { "question" : "answer" },
 "question").question
```

Of course, unquoted keys will not work for strings that are not NCNames, e.g., if the field contains a dot or begins with a digit. Then you will need quotes. If you use a more general expression on the right side of the dot, it must always have parentheses.

Example 7.5. Quotes and parentheses for object lookup

```
{ "my question" : "What NoSQL technology should I use?" }
  ."my question",
{ "my question" : "What NoSQL technology should I use?" }
  .("my " || "question")
```

If it does not evaluate to a single string, an error is raised.

Example 7.6. Object lookup with a nested expression

```
{ "question" : "What NoSQL technology should I use?" }
  .("question", "answer"),
{ "question" : "What NoSQL technology should I use?" }.(1)
```

Variables, or a context item reference, do not need parentheses. Variables are introduced later, but here is a sneak peek:

Example 7.7. Object lookup with a variable

```
let $field := "my " || "question"
return {
  "my question" : "What NoSQL technology should I use?"
}.$field
```

Array Navigation

Array lookup uses normal parentheses - square brackets are for sequence filtering shown below. If the operand does not evaluate to a single positive integer, an error is raised. If the integer exceeds the size of the array, the empty sequence is returned.

Example 7.8. Array lookup

```
[ "question", "answer" ](2),
{
  questions: [
      "What NoSQL technology should I use?",
      { "faq" : "What is the bottleneck in the MapReduce framework?" }
  ]
}.questions(2).faq
```

The behavior on mixed sequences is similar to object lookup, i.e., an implicit iteration is done and an array lookup on non-array items (objects, atomics) returns the empty sequence.

Example 7.9. Array lookup on a mixed sequence (pending implementation in Zorba, should return "foo" "Jim")

```
("question",
 [ "question", "answer" ],
```

```
 { "question" : "answer" },
 [ "NoSQL, "JSONiq" ])(1)
```

You can also extract all items from an array (i.e., as a sequence) with an empty selector.

Example 7.10. Extracting all items from an array

```
[ "What NoSQL technology should I use?",
  "What is the bottleneck in the MapReduce framework?" ]()
```

Sequence Filtering

A predicate allows filtering a sequence, keeping only items that fulfill it.

The predicate is evaluated once for each item in the left-hand-side sequence. The predicate expression can use $$ to refer to the item being processed, called the context item.

If the predicate evaluates to an integer, it is matched against the item position in the left-hand side sequence automatically.

Example 7.11. Predicate expression

```
(1 to 10)[5],
("What NoSQL technology should I use?",
 "What is the bottleneck in the MapReduce framework?")[2]
```

Otherwise, the result of the predicate is converted to a boolean.

All items for which the converted predicate result evaluates to true are then output.

Example 7.12. Predicate expression

```
("What NoSQL technology should I use?",
 "What is the bottleneck in the MapReduce framework?")[contains($$, "NoSQL")],
(1 to 10)[$$ mod 2 eq 0]
```

Part III. JSONiq Expressions

Table of Contents

Chapter 8. Basic Operations

Now that we have showed how expressions can be composed, we can begin the tour of all JSONiq expressions. First, we introduce the most basic operations.

Construction of Sequences

Comma Operator

The comma allows you to concatenate two sequences, or even single items. This operator has the lowest precedence of all, so do not forget the parentheses if you would like to change this.

Also, the comma operator is associative -- in particular, sequences do not nest. You need to use arrays in order to nest.

Example 8.1. Comma

```
1, 2, 3, 4, 5,
{ "foo" : "bar" }, [ 1  ],
1 + 1, 2 + 2,
(1, 2, (3, 4), 5)
```

Range Operator

With the binary operator "to", you can generate larger sequences with just two integer operands.

If the left operand is greater than the right operand, an empty sequence is returned.

If an operand evaluates to something else than a single integer, an error is raised. There is one exception with the empty sequence, which behaves in a particular way for most operations (see below).

Example 8.2. Range operator

```
1 to 10,
10 to 1
```

Parenthesized Expressions

Expressions take precedence on one another. For example, addition has a higher precedence than the comma. Parentheses allow you to change precedence.

If the parentheses are empty, the empty sequence is produced.

Example 8.3. Empty sequence

```
( 2 + 3 ) * 5,
()
```

Arithmetics

JSONiq supports the basic four operations, as well integer division and modulo. You should keep in mind that, as is the case in most programming languages, multiplicative operations have precedence over additive operations. Parentheses can override it, as explained above.

Example 8.4. Basic arithmetic operations with precedence override

```
1 * ( 2 + 3 ) + 7 idiv 2 - (-8) mod 2
```

Dates, times and durations are also supported in a natural way.

Example 8.5. Using basic operations with dates.

```
date("2013-05-01") - date("2013-04-02")
```

If any of the operands is a sequence of more than one item, an error is raised.

If any of the operands is not a number, a date, a time or a duration, or if the operands are not compatible (say a number and a time), an error is raised.

Do not worry if the two operands do not have the same number type, JSONiq will do the adequate conversions.

Example 8.6. Basic arithmetic operations with different, but compatible number types

```
2.3e4 + 5
```

String Concatenation

Two strings or more can be concatenated using the concatenation operator. An empty sequence is treated like an empty string.

Example 8.7. String concatenation

```
"Captain" || " " || "Kirk",
"Captain" || () || "Kirk"
```

Comparison

Atomics can be compared with the usual six comparison operators (equality, non-equality, lower-than, greater-than, lower-or-equal, greater-or-equal), and with the same two-letter symbols as in MongoDB.

Comparison is only possible between two compatible types, otherwise, an error is raised.

Example 8.8. Equality comparison

```
1 + 1 eq 2,
1 lt 2
```

null can be compared for equality or inequality to anything - it is only equal to itself so that false is returned when comparing if for equality with any non-null atomic. True is returned when comparing it with non-equality with any non-null atomic.

Example 8.9. Equality and non-equality comparison with null

```
1 eq null,
"foo" ne null,
null eq null
```

For ordering operators (lt, le, gt, ge), null is considered the smallest possible value (like in JavaScript).

Example 8.10. Ordering comparison with null

```
null lt 1
```

Comparisons and logic operators are fundamental for a query language and for the implementation of a query processor as they impact query optimization greatly. The current comparison semantics for them is carefully chosen to have the right characteristics as to enable optimization.

The Empty Sequence in Range, Arithmetics and Comparison Operations

In range operations, arithmetics and comparisons, if an operand is the empty sequence, then the result is the empty sequence as well.

Example 8.11. The empty sequence used in basic operations

```
() to 10,
1 to (),
1 + (),
() eq 1,
() ge 10
```

Logic

JSONiq logics support is based on two-valued logics: there is just true and false and nothing else.

Non-boolean operands get automatically converted to either true or false, or an error is raised. The boolean() function performs a manual conversion. The rules for conversion were designed in such a way that it feels "natural". Here they are:

- An empty sequence is converted to false.
- A singleton sequence of one null is converted to false.
- A singleton sequence of one string is converted to true except the empty string which is converted to false.

- A singleton sequence of one number is converted to true except zero or NaN which are converted to false.
- Operand singleton sequences of any other item cannot be converted and an error is raised.
- Operand sequences of more than one item cannot be converted and an error is raised.

Example 8.12. Conversion to booleans

```
{
  "empty-sequence" : boolean(()),
  "null" : boolean(null),
  "non-empty-string" : boolean("foo"),
  "empty-string" : boolean(""),
  "zero" : boolean(0),
  "not-zero" : boolean(1e42)
},
null and "foo"
```

Propositional Logic

JSONiq supports the most famous three boolean operations: conjunction, disjunction, and negation. Negation has the highest precedence, then conjunction, then disjunction. Comparisons have a higher precedence than all logical operations. Parentheses can override.

Example 8.13. Logics with booleans

```
  true and ( true or not true ),
  1 + 1 eq 2 or not 1 + 1 eq 3
```

A sequence with more than one item, or singleton objects and arrays cannot be converted to a boolean. An error is raised if it is attempted.

Unlike in C++ or Java, you cannot rely on the order of evaluation of the operands of a boolean operation. The following query may return true or may return an error.

Example 8.14. Non-determinism in presence of errors.

```
  true or (1 div 0)
```

First-Order Logic (Quantified Variables)

Given a sequence, it is possible to perform universal or existential quantification on a predicate.

Example 8.15. Universal quantifier

```
  every $i in 1 to 10 satisfies $i gt 0,
  some $i in -5 to 5, $j in 1 to 10 satisfies $i eq $j
```

Variables can be annotated with a type. If no type is specified, item* is assumed. If the type does not match, an error is raised.

Example 8.16. Existential quantifier with type checking

```
some $i as integer in -5 to 5, $j as integer in 1 to 10 satisfies $i eq $j
```

Builtin Functions

The syntax for function calls is similar to many other languages.

Like in C++ (namespaces) or Java (packages, classes), functions live in namespaces that are URIs.

Although it is possible to fully write the name of a function, namespace included, it can be cumbersome. Hence, for convenience, a namespace can be associated with a prefix that acts as a shortcut.

JSONiq supports three sorts of functions:

- Builtin functions: these have no prefix and can be called without any import.
- Local functions: they are defined in the prolog, to be used in the main query. They have the prefix *local:*. Chapter Chapter 12, *Prologs* describes how to define your own local functions.
- Imported functions: they are defined in a library module. They have the prefix corresponding to the alias to which the imported module has been bound to. Chapter Chapter 13, *Modules* describes how to define your own modules.

For now, we only introduce how to call builtin functions -- these are the simplest, since they do not need any prefix or explicit namespace.

Example 8.17. A builtin function call.

```
keys({ "foo" : "bar", "bar" : "foo" }),
concat("foo", "bar")
```

Some builtin functions perform aggregation and are particularly convenient:

Example 8.18. A builtin function call.

```
sum(1 to 100),
avg(1 to 100),
count( (1 to 100)[ $$ mod 5 eq 0 ] )
```

Remember that JSONiq is a strongly typed language. Functions have signatures, for example sum() expects a sequence of numbers. An error is raised if the actual types do not match the expected types.

Chapter 9. Control Flow Expressions

JSONiq supports control flow expressions such as conditional expressions (if then else), switch, and typeswitch. At least the first two should be familiar to any programmer.

Conditional Expressions

A conditional expression allows you to pick the one or the other value depending on a boolean value.

Example 9.1. A conditional expression

```
if (1 + 1 eq 2) then { "foo" : "yes" } else { "foo" : "false" }
```

The behavior of the expression inside the if is similar to that of logical operations (two-valued logics), meaning that non-boolean values get converted to a boolean. The exists() builtin function can be useful to know if a sequence is empty or not.

Example 9.2. A conditional expression

```
if (null) then { "foo" : "yes" } else { "foo" : "no" },
if (1) then { "foo" : "yes" } else { "foo" : "no" },
if (0) then { "foo" : "yes" } else { "foo" : "no" },
if ("foo") then { "foo" : "yes" } else { "foo" : "no" },
if ("") then { "foo" : "yes" } else { "foo" : "no" },
if (()) then { "foo" : "yes" } else { "foo" : "no" },
if (exists(collection("faq"))) then { "foo" : "yes" } else { "foo" : "no" }
```

Note that the else clause is mandatory (but can be the empty sequence)

Example 9.3. A conditional expression

```
if (1+1 eq 2) then { "foo" : "yes" } else ()
```

Switch expressions

Switch expressions are very similar to C++. A switch expression evaluates the expression inside the switch. If it is an atomic, it compares it in turn to the provided atomic values (with the semantics of the eq operator) and returns the value associated with the first matching case clause.

Example 9.4. A switch expression

```
switch ("foo")
case "bar" return "foo"
case "foo" return "bar"
default return "none"
```

If the provided value is not an atomic, an error is raised (this is also similar to C++).

If the value does not match any of the expected values, the default is used.

Example 9.5. A switch expression

```
switch ("no-match")
case "bar" return "foo"
case "foo" return "bar"
default return "none"
```

The case clauses support composability of expressions as well - an opportunity to remind you about the precedence of the comma.

Example 9.6. A switch expression

```
switch (2)
case 1 + 1 return "foo"
case 2 + 2 return "bar"
default return "none",
switch (true)
case 1 + 1 eq 2 return "1 + 1 is 2"
case 2 + 2 eq 5 return "2 + 2 is 5"
default return "none of the above is true"
```

Try-Catch expressions

A try catch expression evaluates the expression inside the try block and returns its resulting value.

However, if an error is raised during execution, the catch clause is evaluated and its result value returned.

Example 9.7. A try catch expression

```
try { 1 div 0 } catch * { "Caught!" }
```

Only errors raised within the lexical scope of the try block are caught.

Example 9.8. An error outside of a try-catch expression

```
let $x := 1 div 0
return try { $x }
catch * { "Caught!" }
```

Errors that are detected statically within the try block, for example syntax errors, are still reported statically.

Note that this applies also if the engine is capable of detecting a type error statically, while another engine might only discover it at runtime and catch it. You should keep this in mind, and only use try-catch expressions as a safety net.

Example 9.9. A try catch expression with a syntax error

```
try { x } catch * { "Caught!" }
```

Example 9.10. A try catch expression with a type error

```
try { "foo" + "bar" } catch * { "Caught!" }
```

Chapter 10. FLWOR Expressions

FLWOR expressions are probably the most powerful JSONiq construct and correspond to SQL's SELECT-FROM-WHERE statements, but they are more general and more flexible. In particular, clauses can almost appear in any order (apart that it must begin with a for or let clause, and end with a return clause).

Here is a bit of theory on how it works.

A clause binds values to some variables according to its own semantics, possibly several times. Each time, a tuple of variable bindings (mapping variable names to sequences) is passed on to the next clause.

This goes all the way down, until the return clause. The return clause is eventually evaluated for each tuple of variable bindings, resulting in a sequence of items for each tuple. It is not to be confused with Java or C++ return statements, as it does not exit ot break the loop.

These sequences of items are concatenated, in the order of the incoming tuples, and the obtained sequence is returned by the FLWOR expression.

We are now giving practical examples with a hint on how it maps to SQL.

For Clauses

For clauses allow iteration on a sequence.

For each incoming tuple, the expression in the for clause is evaluated to a sequence. Each item in this sequence is in turn bound to the for variable. A tuple is hence produced for each incoming tuple, and for each item in the sequence produced by the for clause for this tuple.

For example, the following for clause:

```
for $x in 1 to 3
```

produces the following stream of tuples (note how JSON syntax is so convenient that it can be used to represent these tuples):

```
{ "$x" : 1 }
{ "$x" : 2 }
{ "$x" : 3 }
```

The order in which items are bound by the for clause can be relaxed with unordered expressions, as described later in this section.

The following query, using a for and a return clause, is the counterpart of SQL's "SELECT display_name FROM answers". $x is bound in turn to each item in the answers collection.

Example 10.1. A for clause.

```
for $x in collection("answers")
return $x.owner.display_name
```

For clause expressions are composable, there can be several of them.

Example 10.2. Two for clauses.

```
for $x in ( 1, 2, 3 )
for $y in ( 1, 2, 3 )
return 10 * $x + $y
```

Example 10.3. A for clause with two variables.

```
for $x in ( 1, 2, 3 ), $y in ( 1, 2, 3 )
return 10 * $x + $y
```

A for variable is visible to subsequent bindings.

Example 10.4. A for clause.

```
for $x in ( [ 1, 2, 3 ], [ 4, 5, 6 ], [ 7, 8, 9 ] ), $y in $x()
return $y,
for $x in collection("faq"), $y in $x.tags()
return { "id" : $x.question_id, "tag" : $y }
```

It is also possible to bind the position of the current item in the sequence to a variable.

Example 10.5. A for clause.

```
for $x at $position in collection("answers")
return { "old id" : $x.answer_id, "new id" : $position }
```

JSONiq supports joins. For example, the counterpart of "SELECT q.title AS question, q.question_id FROM faq q JOIN answers a ON q.question_id = a.question_id" is:

Example 10.6. A regular join

```
for $question in collection("faq"),
    $answer in collection("answers")
    [ $$.question_id eq $question.question_id ]
return { "question" : $question.title,
         "answer score" : $answer.score }
```

Note how JSONiq handles semi-structured data in a flexible way.

Outer joins are also possible with "allowing empty", i.e., output will also be produced if there is no matching movie for a captain. The following query is the counterpart of "SELECT q.title AS question, q.question_id FROM faq q LEFT JOIN answers a ON q.question_id = a.question_id".

Example 10.7. An outer join

```
for $question in collection("faq"),
    $answer allowing empty in collection("answers")
```

```
    [ $$.question_id eq $question.question_id ]
return { "question" : $question.title,
         "answer score" : $answer.score }
```

Where Clauses

Where clauses are used for filtering (selection operator in the relational algebra).

For each incoming tuple, the expression in the where clause is evaluated to a boolean (possibly converting an atomic to a boolean). if this boolean is true, the tuple is forwarded to the next clause, otherwise it is dropped.

The following query corresponds to "SELECT q.title as question, q.question_id as id FROM faq WHERE CONTAINS(question, 'NoSQL')".

Example 10.8. A where clause.

```
for $question in collection("faq")
where contains($question.title, "NoSQL")
return { "question" : $question.title, "id" : $question.question_id }
```

JSONiq can do joins with where clauses, too:

Example 10.9. A join

```
  for $question in collection("faq"),
      $answer in collection("answers")
  where $$.question_id eq $question.question_id
  return { "question" : $question.title,
           "answer score" : $answer.score }
```

Order Clauses

Order clauses are for reordering tuples.

For each incoming tuple, the expression in the where clause is evaluated to an atomic. The tuples are then sorted based on the atomics they are associated with, and then forwarded to the next clause.

Like for ordering comparisons, null values are always considered the smallest.

The following query is the counterpart of SQL's "SELECT a.display_name, a.score FROM answers a ORDER BY a.display_name".

Example 10.10. An order by clause.

```
for $answer in collection("answers")
order by $answer.owner.display_name
return { "owner" : $answer.owner.display_name, "score" : $answer.score }
```

Multiple sorting criteria can be given - they are treated with the semantics of a lexicographic order (the most important criterion being the first one).

Example 10.11. An order by clause.

```
for $answer in collection("answers")
order by $answer.owner.display_name, $answer.score
return { "owner" : $answer.owner.display_name, "score" : $answer.score }
```

It can be specified, for each criterion, whether the order is ascending or descending. Empty sequences are allowed and it can be chosen whether to put them first or last.

Example 10.12. An order by clause.

```
for $answer in collection("answers")
order by $answer.owner.display_name descending empty greatest, $answer.score as
return { "owner" : $answer.owner.display_name, "score" : $answer.score }
```

An error is raised if the expression does not evaluate to an atomic or the empty sequence.

Group Clauses

Grouping is also supported, like in SQL.

For each incoming tuple, the expression in the group clause is evaluated to an atomic. The value of this atomic is called a grouping key. The incoming tuples are then grouped according to the grouping key -- one group for each value of the grouping key.

For each group, a tuple is output, in which the grouping variable is bound to the group's key.

Example 10.13. An group by clause.

```
for $answer in collection("answers")
group by $question := $answer.question_id
return { "question" : $question }
```

As for the other (non-grouping) variables, their values within one group are all concatenated, keeping the same name. Aggregations can be done on these variables.

The following query is equivalent to "SELECT question_id, COUNT(*) FROM answers GROUP BY question_id".

Example 10.14. A group by clause using count aggregation.

```
for $answer in collection("answers")
group by $question := $answer.question_id
return { "question" : $question, "count" : count($answer) }
```

The following query is equivalent to "SELECT question_id, AVG(score) FROM answers GROUP BY question_id".

Example 10.15. A group by clause using average aggregation.

```
for $answer in collection("answers")
group by $question := $answer.question_id
return { "question" : $question, "average score" : avg($answer.score) }
```

JSONiq's group by is more flexible than SQL and is fully composable.

Example 10.16. A group by by clause with a nested expression.

```
for $answer in collection("answers")
group by $question := $answer.question_id
return { "question" : $question, "scores" : [ $answer.score ]  }
```

Unlike SQL, JSONiq does not need a having clause, because a where clause works perfectly after grouping as well.

The following query is the counterpart of "SELECT question_id, COUNT(*) FROM answers GROUP BY question_id HAVING COUNT(*) > 1"

Example 10.17. A group by clause with a post-grouping condition.

```
for $answer in collection("answers")
group by $question := $answer.question_id
where count($answer) gt 1
return { "question" : $question, "count" : count($answer)  }
```

Let Clauses

Let bindings can be used to define aliases for any sequence, for convenience.

For each incoming tuple, the expression in the let clause is evaluated to a sequence. A binding is added from this sequence to the let variable in each tuple. A tuple is hence produced for each incoming tuple.

Example 10.18. An let clause.

```
for $answer in collection("answers")
let $qid := $answer.question_id
group by $question := $qid
let $count := count($answer)
where $count gt 1
return { "question" : $question, "count" : $count  }
```

Note that it is perfectly fine to reuse a variable name and hide a variable binding.

Example 10.19. An let clause reusing the same variable name.

```
for $answer in collection("answers")
let $qid := $answer.question_id
group by $qid
let $count := count($answer)
where $count gt 1
```

```
let $count := size(
  collection("faq")[ $$.question_id eq $qid ].tags
)
return { "question" : collection("faq")[$$.question_id := $qid].title,
         "count" : $count }
```

Count Clauses

For each incoming tuple, a binding from the position of this tuple in the tuple stream to the count variable is added. The new tuple is then forwarded to the next clause.

Example 10.20. A count clause.

```
for $question in collection("faq")
order by size($question.tags)
count $count
return { "id" : $count, "faq" : $question.title }
```

Map Operator

JSONiq provides a shortcut for a for-return construct, automatically binding each item in the left-hand-side sequence to the context item.

Example 10.21. A simple map

```
(1 to 10) ! ($$ * 2)
```

Example 10.22. An equivalent query

```
for $i in 1 to 10
return $i * 2
```

Composing FLWOR Expressions

Like all other expressions, FLWOR expressions can be composed. In the following example, a FLWOR is nested in an existential quantifier, nested in a FLWOR, nested in a function call, nested in a FLWOR, nested in an array constructor:

Example 10.23. Nested FLWORs

```
[
  for $answer in collection("answers")
  let $oid := $answer.owner.user_id
  where count(
      for $question in collection("faq")
      where some $other-answer in collection("answers")
                [$$.question_id eq $question.question_id
                 and
```

```
                 $$.owner.user_id eq $oid]
            satisfies
            $other-answer.score gt $answer.score
       return $question
  ) ge 2
  return $answer.owner.display_name
]
```

Ordered and Unordered Expressions

By default, the order in which a for clause binds its items is important.

This behaviour can be relaxed in order give the optimizer more leeway. An unordered expression relaxes ordering by for clauses within its operand scope:

Example 10.24. An unordered expression.

```
unordered {
  for $answer in collection("answers")
  where $answer.score ge 4
  return $answer
}
```

An ordered expression can be used to reactivate ordering behaviour in a subscope.

Example 10.25. An ordered expression.

```
unordered {
  for $question in collection("faq")
  where exists(
    ordered {
      for $answer at $i in collection("answers")
      where $i eq 5
      where $answer.question_id eq $question.question_id
      return $answer
    }
  )
  return $question
}
```

Chapter 11. Expressions Dealing With Types

We have already introduced the sequence type syntax. It is now time to introduce the expressions that deal with types.

Instance-of Expressions

A quick glimpse on this expression was already given. An instance expression can be used to tell whether a JSONiq value matches a given sequence type, like in Java.

Example 11.1. Instance of expression.

```
1 instance of integer,
1 instance of string,
"foo" instance of string,
{ "foo" : "bar" } instance of object,
({ "foo" : "bar" }, { "bar" : "foo" }) instance of json-item+,
[ 1, 2, 3 ] instance of array?,
() instance of ()
```

Treat Expressions

A treat expression just forwards its operand value, but only after checking that a JSONiq value matches a given sequence type. If it is not the case, an error is raised.

Example 11.2. Treat as expression.

```
1 treat as integer,
"foo" treat as string,
{ "foo" : "bar" } treat as object,
({ "foo" : "bar" }, { "bar" : "foo" }) treat as json-item+,
[ 1, 2, 3 ] treat as array?,
() treat as ()
```

Example 11.3. Treat as expression (failing).

```
1 treat as string
```

Castable Expressions

A castable expression checks whether a JSONiq value can be cast to a given atomic type and returns true or false accordingly. It can be used before actually casting to that type.

The question mark allows for an empty sequence.

Example 11.4. Castable as expression.

```
"1" castable as integer,
"foo" castable as integer,
"2013-04-02" castable as date,
() castable as date,
("2013-04-02", "2013-04-03") castable as date,
() castable as date?
```

Cast Expressions

A cast expression casts a (single) JSONiq value to a given atomic type. The resulting value is annotated with this type.

Also here, the question mark allows for an empty sequence. An error is raised if the cast is unsuccessful.

Example 11.5. Cast as expression.

```
"1" cast as integer,
"2013-04-02" cast as date,
() cast as date?,
"2013-04-02" cast as date?
```

Example 11.6. Cast as expression (failing).

```
("2013-04-02", "2013-04-03") cast as date,
"foo" cast as integer,
() cast as date
```

Typeswitch Expressions

A typeswitch expressions tests if the value resulting from the first operand matches a given list of types. The expression corresponding to the first matching case is finally evaluated. If there is no match, the expression in the default clause is evaluated.

Example 11.7. Typeswitch expression.

```
typeswitch("foo")
  case integer return "integer"
  case string return "string"
  case object return "object"
  default return "other"
```

In each clause, it is possible to bind the value of the first operand to a variable.

Example 11.8. Typeswitch expression.

```
typeswitch("foo")
  case $i as integer return $i + 1
  case $s as string return $s || "foo"
  case $o as object return [ $o ]
```

```
  default $d return $d
```

The vertical bar can be used to allow several types in the same case clause.

Example 11.9. Typeswitch expression.

```
typeswitch("foo")
  case $a as integer | string return { "integer or string" : $a }
  case $o as object return [ $o ]
  default $d return $d
```

Part IV. Prolog, Modules and Functions

Table of Contents

Chapter 12. Prologs

This section introduces prologs, which allows declaring functions and global variables that can then be used in the main query. A prolog also allows setting some default behaviour.

The prolog appears before the main query and is optional. It can contain setters and module imports, followed by function and variable declarations.

Module imports are explained in the next chapter.

Setters.

Setters allow to specify a default behaviour for various aspects of the language.

Default Ordering Mode

This specifies the default behaviour of for clauses, i.e., if they bind tuples in the order in which items occur in the binding sequence. It can be overriden with ordered and unordered expressions.

Example 12.1. A default ordering setter.

```
declare ordering unordered;
for $answer in collection("answers")
return {
  "owner" : $answer.owner.display_name,
  "score" : $answer.score
}
```

Default Ordering Behaviour For Empty Sequences

This specifies whether empty sequences come first or last in an ordering clause. It can be overriden by the corresponding directives in such clauses.

Example 12.2. A default ordering for empty sequences.

```
declare default order empty least;
for $x in ({ "foo" : "bar" }, {})
order by $x.foo
return $x
```

Default Decimal Format

This specifies a default decimal format for the builtin function format-number().

Example 12.3. A default ordering setter.

```
declare default decimal-format decimal-separator = ","
                               grouping-separator = " ";
format-number(12345.67890, "# ###,##")
```

Namespaces

Variables and functions live in namespaces that are URIs -- the semantics is similar to that of C++ namespaces. For convenience, namespaces are associated with a much shorter alias, and this alias can be used as a prefix to a variable or a function.

Until now, we only dealt with main queries. In main queries, the namespace alias *local:* is predefined so that global variables and functions that are local to the main query can use this alias, for example *local:myvariable* or *local:myfunction()*. This alias is associated with a namespace, but which namespace it is is not relevant for writing queries.

For variables, the alias is optional -- variables not prefixed with an alias live in no namespace.

For functions, the absence of alias is only allowed for builtin functions. Builtin functions live in their own special namespace.

Other namespaces and aliases can be defined as well with imported library modules. This is defined in the next chapter.

Global Variables

Variables can be declared global. Global variables are declared in the prolog.

Example 12.4. Global variable.

```
declare variable $obj := { "foo" : "bar" };
declare variable $numbers := (1, 2, 3, 4, 5);
$obj,
[ $numbers ]
```

You can specify a type for a variable. If the type does not match, an error is raised. Types will be explained later. In general, you do not need to worry too much about variable types except if you want to make sure that what you bind to a variable is really what you want. In most cases, the engine will take care of types for you.

Example 12.5. Global variable with a type.

```
declare variable $obj as object := { "foo" : "bar" };
$obj
```

An external variable allows you to pass a value from the outside environment, which can be very useful. Each implementation can choose their own way of passing a value to an external variable. A default value for an external variable can also be supplied in case none is provided outside.

Example 12.6. An external global variable with a default value.

```
declare variable $obj external := { "foo" : "bar" };
$obj
```

In these examples, global variables have no prefix. They can also be prefixed with the predefined alias *local:*, but them they must be prefixed both in the declaration and when used.

Example 12.7. An external global variable with the local: alias.

```
  declare variable $local:obj external := { "foo" : "bar" };
  $local:obj
```

Global variables that are imported from other modules are prefixed with the alias associated with the imported module, as will be explained in the next chapter.

User-Defined Functions

You can define your own functions in the prolog.

Unlike variables, user-defined functions must be prefixed, because unprefixed functions are the builtin functions.

In the prolog of a main query, these user-defined functions must be prefixed with the predefined alias *local:*, both in the declaration and when called.

Remember that types are optional, and if you do not specify any, *item** is assumed, both for parameters and for the return type.

Example 12.8. Some user-defined functions.

```
declare function local:say-hello-1($x)
{
  "Hello, " || $x || "!"
};

declare function local:say-hello-2($x as string)
{
  "Hello, " || $x || "!"
};

declare function local:say-hello-3($x as string) as string
{
  "Hello, " || $x || "!"
};

local:say-hello-1("Mister Spock"),
local:say-hello-2("Mister Spock"),
local:say-hello-3("Mister Spock")
```

If you do specify types, an error is raised in case of a mismatch

Example 12.9. A type mismatch for a user-defined function.

```
declare function local:say-hello($x as string)
{
  "Hello, " || $x || "!"
};

local:say-hello(1)
```

Chapter 13. Modules

You can group functions and variables in separate units, called library modules.

Up to now, everything we encountered were main modules, i.e., a prolog followed by a main query.

A library module does not contain any query - just functions and variables that can be imported by other modules.

A library module must be assigned to a namespace. For convenience, this namespace is bound to an alias in the module declaration. All variables and functions in a library module must be prefixed with this alias.

Example 13.1. A library module.

```
module namespace my = "http://www.example.com/my-module";

declare variable $my:variable := { "foo" : "bar" };
declare variable $my:n := 42;

declare function my:function($i as integer)
{
  $i * $i
};
```

Once you have defined a library module, you can import it in any other module (library or main). An alias must be given to the module namespace (my). Variables and functions from that module can be accessed by prefixing their names with this alias. The alias may be different than the internal alias defined in the imported module -- only the namespace really matters.

Example 13.2. An importing main module

```
import module namespace other = "http://www.example.com/my-module";
other:function($other:n)
```

An engine may come with a number of builtin library modules. For example, there is the standardized math module.

Example 13.3. The math module

```
import module namespace math =
    "http://www.w3.org/2005/xpath-functions/math";
math:pi(),
math:pow(2, 30)
```

Chapter 14. Function Library

JSONiq provides a rich set of builtin functions. We now introduce them, mostly by giving examples of usage.

JSON specific functions.

keys

This function returns the keys of an object in an implementation-dependent order.

```
keys($o as item*) as string*
```

Example 14.1. The keys function on an object.

```
keys({ "foo" : 1, "bar" : 2 })
```

Example 14.2. The keys function on a more general sequence.

```
keys(({ "foo" : 1, "bar" : 2 }, [ 1, 2 ], "foo", { "a" : 1, "b" : 2 }))
```

members

This functions returns all values in an array.

```
members($a as item*) as item*
```

Example 14.3. The members function on an array.

```
members([ "mercury", "venus", "earth", "mars" ])
```

Example 14.4. The members function on a general sequence.

```
members(([ "mercury", "venus", "earth", "mars" ]), "foo", { "foo" : "bar" }, [
```

parse-json

This function parses its first parameter (a string) as JSON, and returns the resulting object or array (or a sequence thereof).

```
parse-json($arg as string?) as json-item*
```

```
parse-json($arg as string?, $options as object) as json-item*
```

The object optionally supplied as the second parameter may contain additional options, in this case "jsoniq-multiple-top-level-items" which indicates where multiple objects and arrays are to be parsed.

If parsing is not successful, an error is raised.

Example 14.5. Parsing a JSON document

```
parse-json("{ \"foo\" : \"bar\" }"),
parse-json("{ \"foo\" : \"bar\" }{ \"foo\" : \"bar\" }",
           { "jsoniq-multiple-top-level-items" : true })
```

size

This function returns the size of the supplied array.

```
size($a as array) as integer
```

Example 14.6. Retrieving the size of an array.

```
size([1 to 10])
```

accumulate

This function dynamically builds an object, like the {| |} syntax, except that it does not throw an error upon pair collision. Instead, it accumulates them into an array.

```
accumulate($arg as object*) as object
```

Example 14.7. Accumulating objects.

```
accumulate(({ "foo" : "bar", "bar" : "foo" }, { "foo" : "other" }))
```

descendant-objects

This function returns all objects contained within a JSON item, regardless of depth.

```
descendant-objects($arg as json-item) as object*
```

Example 14.8. Retrieving the descendant objects.

```
import module namespace lib = "http://www.jsoniq.org/function-library";
lib:descendant-objects({
  "foo" : { "foo2" : "bar2" },
  "bar" : [ 1, 2, { "foo3" : "bar3" } ]
})
```

descendant-pairs

This function returns all descendant pairs of an object.

```
descendant-pairs($o as object) as object*
```

Example 14.9. Retrieving all descendant pairs.

```
let $o :=
{
  "first" : 1,
  "second" : {
    "first" : "a",
    "second" : "b"
  }
}
return descendant-pairs($o)
```

flatten

This function recursively "flattens" the array.

```
declare function flatten($a as array) as item*
{
  for $value in $a()
  return typeswitch ($value)
         case array return flatten($value)
         default return $value
};
```

intersect

This function returns the intersection of two objects, and aggregates values corresponding to the same name into an array.

```
declare function intersect($o as object*)
{
  {|
    let $common-keys := keys(head($o))[ every $object in tail($o)
                                        satisfies keys($object) = $$ ]
    for $key in $common-keys
    let $values := $o.$key
    return
      if (count($values) eq 1)
      then { $key : $values }
      else { $key : [ $values ] }
  |}
};
```

project

This function projects an object by filtering its pairs.

```
declare function project($o as object, $s as string*) as object
{
  {|
    for $key in $s
    let $value := $o.$key
    where exists ($value)
```

```
    return { $key : $value }
  |}
};
```

Example 14.10. Projecting an object 1

```
let $o := {
  "Captain" : "Kirk",
  "First Officer" : "Spock",
  "Engineer" : "Scott"
  }
return project($o, ("Captain", "First Officer"))
```

Example 14.11. Projecting an object 2

```
let $o := {
  "Captain" : "Kirk",
  "First Officer" : "Spock",
  "Engineer" : "Scott"
  }
return project($o, "XQuery Evangelist")
```

values

This functions returns all values in an Object.

```
declare function values($i as object) as item* {
  keys($i) ! $i.($$)
};
```

encode-for-roundtrip

This function encodes any sequence of items, even containing non-JSON types, to a sequence of JSON items that can be serialized as pure JSON, in a way that it can be parsed and decoded back using decode-from-roundtrip. JSON features are left intact, while atomic items annotated with a non-JSON type are converted to objects embedding all necessary information.

```
encode-for-roundtrip($items as item*) as json-item*
```

decode-from-roundtrip

This function decodes a sequence previously encoded with encode-for-roundtrip.

```
decode-from-roundtrip($items as json-item*) as item*
```

Functions taken from XQuery

- Access to the external environment: collection#1
- Function to turn atomics into booleans for use in two-valued logics: boolean#1

- Raising errors: error#0, error#1, error#2, error#3.
- Functions on numeric values: abs#1, ceilingabs#1, floorabs#1, roundabs#1, round-half-to-even#1
- Parsing numbers: number#0, number#1
- Formatting integers: format-integer#2, format-integer#3
- Formatting numbers: format-number#2, format-number#3
- Trigonometric and exponential functions: pi#0, exp#1, exp10#1, log#1, log10#1, pow#2, sqrt#1, sin#1, cos#1, tan#1, asin#1, acos#1, atan#1, atan2#1
- Functions to assemble and disassemble strings: codepoints-to-string#1, string-to-codepoints#1
- Comparison of strings: compare#2, compare#3, codepoint-equal#2
- Functions on string values: concat#2, string-join#1, string-join#2, substring#2, substring#3, string-length#0, string-length#1, normalize-space#0, normalize-space#1, normalize-unicode#1, normalize-unicode#2, upper-case#1, lower-case#1, translate#3
- Functions based on substring matching: contains#2, contains#3, starts-with#2, starts-with#3, ends-with#2, ends-with#3, substring-before#2, substring-before#3, substring-after#2, substring-after#3
- String functions that use regular expressions: matches#2, matches#3, replace#3, replace#4, tokenize#2, tokenize#3
- Functions that manipulate URIs: resolve-uri#1, resolve-uri#2, encode-for-uri#1, iri-to-uri#1, escape-html-uri#1
- General functions on sequences: empty#1, exists#1, head#1, tail#1, insert-before#3, remove#2, reverse#1, subsequence#2, subsequence#3, unordered#1
- Function that compare values in sequences: distinct-values#1, distinct-values#2, index-of#2, index-of#3, deep-equal#2. deep-equal#3
- Functions that test the cardinality of sequences: zero-or-one#1, one-or-more#1, exactly-one#1
- Aggregate functions: count#1, avg#1, max#1, min#1, sum#1
- Serializing functions: serialize#1 (unary)
- Context information: current-dateTime#1, current-date#1, current-time#1, implicit-timezone#1, default-collation#1
- Constructor functions: for all builtin types, with the name of the builtin type and unary. Equivalent to a cast expression.

Chapter 15. Equality and identity

As in most language, one can distinguish between physical equality and logical equality.

Atomics can only be compared logically. Their physically identity is totally opaque to you.

Example 15.1. Logical comparison of two atomics

```
1 eq 1
```

Example 15.2. Logical comparison of two atomics

```
1 eq 2
```

Example 15.3. Logical comparison of two atomics

```
"foo" eq "bar"
```

Example 15.4. Logical comparison of two atomics

```
"foo" ne "bar"
```

Two objects or arrays can be tested for logical equality as well, using deep-equal(), which performs a recursive comparison.

Example 15.5. Logical comparison of two JSON items

```
deep-equal({ "foo" : "bar" }, { "foo" : "bar" })
```

Example 15.6. Logical comparison of two JSON items

```
deep-equal({ "foo" : "bar" }, { "bar" : "foo" })
```

The physical identity of objects and arrays is not exposed to the user in the core JSONiq language itself. Some library modules might be able to reveal it, though.

Chapter 16. Notes

Sequences vs. Arrays

Even though JSON supports arrays, JSONiq uses a different construct as its first class citizens: sequences. Any value returned by or passed to an expression is a sequence.

The main difference between sequences and arrays is that sequences are completely flat, meaning they cannot contain other sequences.

Since sequences are flat, expressions of the JSONiq language just concatenate them to form bigger sequences.

This is crucial to allow streaming results, for example through an HTTP session.

Example 16.1. Flat sequences

```
( (1, 2), (3, 4) )
```

Arrays on the other side can contain nested arrays, like in JSON.

Example 16.2. Nesting arrays

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

Many expressions return single items - actually, they really return a singleton sequence, but a singleton sequence of one item is considered the same as this item.

Example 16.3. Singleton sequences

```
1 + 1
```

This is different for arrays: a singleton array is distinct from its unique member, like in JSON.

Example 16.4. Singleton sequences

```
[ 1 + 1 ]
```

An array is a single item. A (non-singleton) sequence is not. This can be observed by counting the number of items in a sequence.

Example 16.5. count() on an array

```
count([ 1, "foo", [ 1, 2, 3, 4 ], { "foo" : "bar" } ])
```

Example 16.6. count() on a sequence

```
count( ( 1, "foo", [ 1, 2, 3, 4 ], { "foo" : "bar" } ) )
```

Other than that, arrays and sequences can contain exactly the same members (atomics, arrays, objects).

Example 16.7. Members of an array

```
[ 1, "foo", [ 1, 2, 3, 4 ], { "foo" : "bar" } ]
```

Example 16.8. Members of an sequence

```
( 1, "foo", [ 1, 2, 3, 4 ], { "foo" : "bar" } )
```

Arrays can be converted to sequences, and vice-versa.

Example 16.9. Converting an array to a sequence

```
[ 1, "foo", [ 1, 2, 3, 4 ], { "foo" : "bar" } ]()
```

Example 16.10. Converting a sequence to an array

```
[ ( 1, "foo", [ 1, 2, 3, 4 ], { "foo" : "bar" } ) ]
```

Null vs. empty sequence

Null and the empty sequence are two different concepts.

Null is an item (an atomic value), and can be a member of an array or of a sequence, or the value associated with a key in an object. Sequences cannot, as they represent the absence of any item.

Example 16.11. Null values in an array

```
[ null, 1, null, 2 ]
```

Example 16.12. Null values in an object

```
{ "foo" : null }
```

Example 16.13. Null values in a sequence

```
(null, 1, null, 2)
```

If an empty sequence is found as an object value, it is automatically converted to null.

Example 16.14. Automatic conversion to null.

```
{ "foo" : () }
```

In an arithmetic opration or a comparison, if an operand is an empty sequence, an empty sequence is returned. If an operand is a null, an error is raised except for equality and inequality.

Example 16.15. Empty sequence in an arithmetic operation.

```
() + 2
```

Example 16.16. Null in an arithmetic operation.

```
null + 2
```

Example 16.17. Null and empty sequence in an arithmetic operation.

```
null + ()
```

Example 16.18. Empty sequence in a comparison.

```
() eq 2
```

Example 16.19. Null in a comparison.

```
null eq 2
```

Example 16.20. Null in a comparison.

```
null lt 2
```

Example 16.21. Null and the empty sequence in a comparison.

```
null eq ()
```

Example 16.22. Null and the empty sequence in a comparison.

```
null lt ()
```

Chapter 17. Open Issues

The JSON update syntax [http://jsoniq.org/docs/JSONiqExtensionToXQuery/html/section-json-updates.html] was not integrated yet into the core language. This is planned, and the syntax will be simplified (no json keyword, dot lookup allowed here as well).

The semantics for the JSON serialization method [http://jsoniq.org/docs/JSONiqExtensionToXQuery/html/section-json-serialization.html] is the same as in the JSONiq Extension to XQuery. It is still under discussion how to escape special characters with the Text output method.

Appendix A. Revision History

Revision History

Revision 1.0.7	Monday May 27th, 2013	Ghislain Fourny `<g@28.io>`

Relaxed array and object navigation to make it more XPath-like. No more cast is done. Lookup on mixed sequences of items (with an implicit iteration) are allowed. The empty sequence is returned in case of mismatch, i.e. in case of (i) integer lookup on object/atomic or (ii) string lookup on array/atomic.

Revision 1.0.6	Friday May 24th, 2013	Ghislain Fourny <g@28.io>

Reran all queries through a newer Zorba version (trunk 11475).

Revision 1.0.5	Thursday May 2nd, 2013	Ghislain Fourny <g@28.io>

Reran all queries through a newer Zorba version (trunk revision 11426).

Corrected some typos, like superfluous prefixes in function library sample queries.

Revision 1.0.4	Monday April 8th, 2013	Ghislain Fourny <g@28.io>

Allowed the context item to appear unparenthesized in an object lookup expression.

Added the function parse-json()

Revision 1.0.3	Thursday April 4th, 2013	Ghislain Fourny <g@28.io>

Fixed some typos.

Revision 1.0.2	Thursday April 4th, 2013	Ghislain Fourny <g@28.io>

Null are comparable with other atomic items and are considered the smallest.

Added link to semantics for all builtin functions inherited from XQuery.

Aggregated some examples.

Revision 1.0.1	Tuesday April 2nd, 2013	Ghislain Fourny <g@28.io>

Initial commit.