Evaluating Functional Programming for Maintainable REST APIs

A thesis presented
by
Marc Coquand
to
The Department of Computer Science

for a degree of
Master of Science in Engineering
in the subject of
Interaction Technology and Design

Umeå University
Supervised by Anders Broberg
July 29, 2019

Abstract

This study's goal is to investigate the use of functional programming as a measure for reducing software defects and how it affects readability. It does so by looking at a specific case and introducing a REST framework called Cause which is completely type safe and helps creating type safe REST frameworks.

Acknowledgements

TODO I want to thank myself for my amazing work

Contents

1 Introduction: Software paradigms and complexity **4**

2 Background **6**
2.1 Introduction to REST servers 6
2.1.1 Implementation concerns for REST apis 8
2.2 Architecture 8
2.2.1 Unit testing 9
2.2.2 Integration testing 9
2.2.3 End-2-End Tests 9
2.2.4 Challenges 10
2.3 SOLID principles 11
2.4 Readability in code 12

3 Theory **13**
3.1 Concepts from Functional Programming 13
3.1.1 ADTs: Sum types and product types 14
3.1.2 GADT 15
3.1.3 Type classes 16
3.1.4 Functors and Contravariant Functors 17
3.1.5 Brief introduction to Monads for side effects 17
3.2 SOLID principles with functional programming 18
3.2.1 Single Responsibility Principle 18
3.2.2 Liskov Substitution Principle 19
3.2.3 Dependency Inversion Principle 19
3.2.4 Interface Segregation Principle 20
3.2.5 Open/Closed principles 21
3.3 Functional servers 24

4 Method **28**
4.1 Constructing the server in Express and Cause 28
4.2 Evaluating maintainability 29
4.2.1 Evaluating readability through code reviews 29
4.2.2 Evaluating the answers 30

5 Results **32**
5.1 Evaluating adherence to SOLID 32
5.1.1 Functional solution 32
5.1.2 Imperative solution 33
5.2 Interviews 33
5.2.1 Raw results 33
5.3 Bias 33
5.4 Limitations 33
5.4.1 Improvements to implementation 33
5.5 future work 34
5.5.1 Relations to cardinality 34
5.6 ReasonML REST implementation 36
5.7 NodeJS REST implementation 36

Chapter 1

Introduction: Software paradigms and complexity

Different schools of thoughts have different approaches when it comes to building applications. There is one that is the traditional, object oriented, procedural way of doing it. Then there is a contender, a functional approach, as an alternative way to build applications. Early programming languages were based around procedure calls. Procedures are the series of steps the computer needs to perform the computations desired. [1] These programming languages were turing complete. A programming language is Turing complete if is able to calculate everything that is calculatable. [2] Object-oriented programming is used to make code more reusable and structured. Java, amongst most languages, is a popular language that is turing complete, object-oriented and imperative. [3]

Functional programming originates from 1936 from Lambda calculus. Lambda calculus is a theory for functions and evolved from Church and Curry to create an alternative foundation for mathematics. [4] Even though Turing machines and Lambda calculus developed separately, the church-turing thesis proves that any computational problem that is solvable with Lambda Calculus is also solveable for Turing machines and vice versa. [5] What this means in practice is any program written in the functional paradigm can be written in the procedural, object-oriented paradigm. However it does not mean that the paradigms are the same as one solution might be very complex in a procedural language and simple in a functional language and vice versa as long as they are turing-complete.

As software engineers, one factor of concern is software quality. Software quality can be divided into two different subparts: software functional quality and software structural quality. Software functional quality reflects how well our system conforms

to given functional requirements or specification and the degree of which we produce correct software. To check that the software is correct, software engineers create tests. [6] In order to create tests, the engineer employs various patterns in the code to make the code easier to test. One approach is functional programming. However Functional programming is not the most popular approach.

Software structural quality refers to how well the software adheres to non-functional requirements such as robustness and maintainability. [6] Some of the maintainability aspects, such as readability, is hard to measure quantitatively. By performing semi-structured interviews, it is possible to investigate how well the code is understood. Since functional programming is not the most popular approach to software engineering today, it is worth taking into consideration how employing functional programming might affect the readability and maintainability of the software. If no one understands the code, how can they be expected to maintain the software?

The goal of this thesis is therefore to investigate how readability with Functional programming in a common area of software development, servers. Or more specifically, REST servers.

Chapter 2

Background

This chapter will introduce REST servers and how software quality is ensured today in the industry. It will also establish the challenges that arises.

2.1 Introduction to REST servers

Servers are applications that provide functionality for other programs or devices, called clients. Services are servers that allow sharing data or resources among clients or to perform a computation.

REST (Representational State Transfer) is a software architecture style that is used to construct web services. A so called RESTful web service allow requesting systems to access and manipulate textual representations of web services by using a set of stateless operations. The architectual constraints of REST are as follows:

Client - Server Architecture Separate the concerns between user interface concerns and data storage concerns.

Statelessness Each request contains all the information neccessary to perform a request. State can be handled by cookies on the user side or by using databases. The server itself contains no state.

Cacheability As on the World Wide Web, clients and intermediaries can cache responses. Responses must therefore, implicitly or explicitly, define themselves as cacheable or not to prevent clients from getting stale or inappropriate data in response to further requests. Well-managed caching partially or completely eliminates some client–server interactions, further improving scalability and performance.

Layered system A client can not tell if it is connected to an end server or some intermeditary server.

Code on demand Servers can send functionality of a client via exectuable code such as javascript. This can be used to send the frontend for example.

Uniform interface The interface of a RESTful server consists of four components. The request must specify how it would like the resource to be represented; that can for example be as JSON, XML or HTTP which are not the servers internal representation. Servers internal representation is therefore separated. When the client holds a representation of the resource and metadata it has enough information to manipulate or delete the resource. Also the REST server has to, in it's response, specify how the representation for the resource. This is done using Media type. Some common media types are JSON, HTML and XML.

A typical HTTP request on a restful server consists of one of the verbs: GET, POST, DELETE, PATCH and PUT. They are used as follows:

GET Fetches a resource from the server. Does not perform any mutation.

POST Update or modify a resource.

PUT Modify or create a resource.

DELETE Remove a resource from the server.

PATCH Changes a resource.

A request will specify a header "Content-Type" which contains the media representation of the request content. For example if the new resource is represented as Json then content-type will be "application/json". It also specifies a header "Accept" which informs which type of representation it would like to have, for example Html or Json.

A request will also contain a route for the resource it is requesting. These requests can also have optional parameters called query parameters. In the request route:

```
1 /api/books?author=Mary&published=1995
```

the ? informs that the request contains query parameters which are optional. In the example above it specfies that the request wants to access the books resource with the parameters author as Mary and published as 1995.

When a request has been done the server responds with a status code that explains the result of the request. The full list of status codes and their descriptions can be found here: `https://en.wikipedia.org/wiki/List_of_HTTP_status_codes`

2.1.1 Implementation concerns for REST apis

A REST api has to concern themselves with the following:

- Ensure that the response has the correct status code.
- Ensure that the correct representation is sent to the client.
- Parse the route and extract it's parameters.
- Parse the query and extract it's parameters.
- Handle errors if the route or query are badly formatted.
- Generate the correct response body containing all the resources needed.

Every type of error has a specific status code, these need to be set correctly.

2.2 Architecture

When developing large scale server applications, often the requirements are as follows:

- There is a team of developers
- New team members must get productive quickly
- The system must be continuously developed and adapt to new requirements
- The system needs to be continuously tested
- System must be able to adapt to new and emerging frameworks

Two different approaches to developing these large scale applications are microservice and monolithic systems. The monolithic system comprises of one big "top-down" architecture that dictates what the program should do. This is simple to develop using some IDE and deploying simply requires deploying some files to the runtime.

As the system starts to grow the large monolithic system becomes harder to understand as the size doubles. As a result, development typically slows down. Since there are no boundaries, modularity tends to break down and the IDE becomes slower over time, making it harder to replace parts as needed. Since redeploying requires the entire application to be replaced and tests becomes slower; the developer becomes less

productive as a result. Since all code is written in the same environment introducing new technology becomes harder.

In a microservice architecture the program comprises of small entities that each have their own responsibility. There can be one service for metrics, one that interacts with the database and one that takes care of frontend. This decomposition allows the developers to easier understand parts of the system, scale into autonomous teams, IDE becomes faster since codebases are smaller, faults become easier to understand as they each break in isolation. Also long-term commitment to one stack becomes less and it becomes easier to introduce a new stack.

The issue with microservices is that when scaling the complexity becomes harder to predict. While testing one system in isolation is easier testing the entire system with all parts together becomes harder. Thus there are different types of tests that are conducted: unit tests, integration tests and E2E-tests.

2.2.1 Unit testing

Unit testing is a testing method where the individual units of code and operating procedures are tested to see if they are fit for use. A unit is informally the smallest testable part of the application. To deal with units dependence one can use method stubs, mock objects and fakes to test in isolation. The goal of unit testing is to isolate each part of the programs and ensure that the individual parts are correct. It also allows for easier refactoring since it ensures that the individual parts still satisfy their part of the application.

To create effective unit tests it's important that it's easy to mock examples. This is usually hindered if the code is dependant on some state since previous states might affect future states.

2.2.2 Integration testing

Whereas unit testing validates that the individual parts work in isolation; integration tests make sure that the modules work when combined. The purpose is to expose faults that occurs when the modules interact with each other.

2.2.3 End-2-End Tests

An End-2-End test (also known as E2E test) is a test that tests an entire passage through the program, testing multiple components on the way. This sometimes requires setting up an emulated environment mock environment with fake variables.

2.2.4 Challenges

When writing unit tests that depend on some environment, for example fetching a user from some database, it can be difficult to test without simulating the environment itself. In such cases one can use dependency injections and mock the environment with fake data. Dependency injection is a method that substitutes environment calls and returns data instead. The issue with unit tests is that even if a feature works well in isolation it does not imply that it will work well when composed with other functions.

The challenge in integration and E2E-tests comes with simulating the entire environments. Given a server connected to some file storage and a database it requires setting up a local simulation of that environment to run the tests. This results in slower execution time for tests and also requires work setting up the environment. Thus it ends up being costly. Also the bigger the space that is being tested the less close the test is to actually finding the error, thus the test ends up finding some error but it can be hard to track it down.

Thus to mitigate these issues the correct architecture needs to be created to make it easier to test. However if there is nothing forcing the programmer to develop software in this way it creates the possibility for the programmer to "cheat" and create software that is not maintainable.

To mitigate the programmer from making mistakes, some languages feature a type system. The type system is a compiler check that ensures that the allowed values are entered. Different strengths exist between various programming languages with some featuring higher-kinded types (types of types) and other constructs.

It is possible to combine the type system with design patterns to force the developer to create the right thing. Such constructs that exists are monads, which can be used to force the separation between pure and impure functionality. Later chapters will introduce a REST framework named Cause, which has been created to force the developer to create REST compliant servers.

However these patterns make heavy use of functional programming. Functional programming as a software paradigm is not popular, with the preferred software paradigm being Object-oriented programming. Thus this thesis aims to investigate the effects of introducing functional constructions to programmers with little familiarity with functional programming when it comes to understandability. Understandability is important to reduce the learning time for programmers and cut down learning costs.

2.3 SOLID principles

A poorly written system can lead to rotten design. Martin Robert, a software engineer, claims that there are four big indicators of rotten design. Rotten design also leads to problems that were established in Chapter 2, such as making easy unit tests. Thus a system should avoid the following.

Rigidity is the tendency for software to be difficult to change. This makes it difficult to change non-critical parts of the software and what can seem like a quick change takes a long time.

Fragility is when the software tends to break when doing simple changes. It makes the software difficult to maintain, with each fix introducing new errors.

Immobility is when it is impossible to reuse software from other projects in the new project. So engineers discover that, even though they need the same module that was in another project, too much work is required to decouple and separate the desirable parts.

Viscosity comes in two forms: the viscosity of the environment and the viscosity of the design. When making changes to code there are often multiple solutions. Some solutions preserve the design of the system and some are "hacks". The engineer can therefore easily implement an unmaintainable solution. The long compile times affect engineers and makes them attempt to make changes that do not cause long compile times. This leads to viscosity in the environment.

To avoid creating rotten designs, Martin Robert proposes the SOLID guideline. SOLID mnemonic for five design principles to make software more maintainable, flexible and understandable. The SOLID guidelines are:

Single responsibility principle Here, responsibility means "reason to change". Modules and classes should have one reason to change and no more.

Open/Closed principle States we should write our modules to be extended without modification of the original source code.

Liskov substitution principle Given a base class and an derived class derive, the user of a base class should be able to use the derived class and the program should function properly.

Interface segregation principle No client should be forced to depend on methods it does not use. The general idea is that you want to split big interfaces to smaller, specific ones.

Dependency inversion principle A strategy to avoid making our source code dependent on specific implementations is by using this principle. This allows us, if we depend on one third-party module, to swap that module for another one should we need to. This can be done by creating an abstract interface and then instance that interface with a class that calls the third-party operations. [7]

Using a SOLID architecture helps make programs that are not as dependent on the environments, making them easier to test (swapping the production environment to a test environment becomes trivial). When investigating the testability, it is important to look at programs that are written in such a way that all parts are easy to test. Thus choosing a SOLID architecture for programs will allow making more testable software. These concepts were designed for Object-oriented programming but can be translated to functional programming which will be demonstrated after introducing the functional concepts needed.

2.4 Readability in code

When evaluating the readability of code, companies can use Code reviews. A code review is an activity in which humans check how well the code can be understood by reading it. Thus similarity it can be used to evaluate how well users without experience with functional code understand functional programs.

By creating an semi-structured interview, where the programmers is asked open questions about how the code works it can give insights about the defects of the software and if there is something fundamental about functional programming that makes it harder to understand.

So in summary: adherence to SOLID principles can be used to avoid rotten design and in order to evaluate the readability a code review can be conducted. Thus a good REST framework should be understood by other programmers and follow SOLID principles.

Chapter 3

Theory

Based on the challenges outlined in Chapter 2, the goal now becomes to construct a maintainable library for REST apis. The approach in this thesis is by using constructs from Functional programming. This chapter will introduce the fundamentals of functional programming to then move on and use that to construct a server library with automatic error handling and type safety.

3.1 Concepts from Functional Programming

While different definitions exist of what Functional programming means, here functional programming is a paradigm that uses of pure functions, decoupling state from logic and immutable data.

Purity When a function is pure it means that calling a function with the same arguments will always return the same value and that it does not mutate any value. For example, given $f(x) = 2 \cdot x$, then $f(2)$ will always return 4. It follows then that an impure functions is either dependant on some state or mutates state in some way. For example, given $g(x) = currenttime \cdot x$, $g(5)$ will yield a different value depending on what time it is called. This makes it dependant on some state of the world. Or given $x = 0$, $h() = x + 1$. Then $h()$ will yield $x = 1$ and $(h \circ h)()$ will yield $x = 2$, making it impure. [8]

Immutable data by deafult Immutable data is data that after initialization can not change. This means if we initialize a record, `abc = {a: 1, b: 2, c: 3}` then `abc.a := 4` is an illegal operation. Immutable data, along with purity, ensures that no data can be mutated unless it is specifically created as mutable

data. Mutable data is an easy source of bug because it can cause two different functions to modify the same value, leading to unexpected results.

Higher-order functions Higher-order functions are functions which either return a function or take one or more functions as arguments. A function $twice : (a \to a) \to (a \to a)$, $twice\ f = f \circ f$, takes a function as an argument and returns a new function which performs given function twice on the argument. So for example $addOne = (+)1$, $addTwo = twiceaddOne$.

Decoupling state from logic Even if functional programs emphasise purity applications still need to deal with state somehow. For example a server would need to interact with a database. Functional programs solve this by separating pure functions and effectful functions. Effects are observable interactions with the environment, such as database access or printing a message. While various strategies exist, like Functional Reactive Programming[1], Dialogs[2] or uniqueness types[3], the one used in Haskell (the language used in this thesis to construct the programs) is the IO monad. For the uninitiated, one can think of Monads as a way to note which functions are pure and which are effectful and managing the way they intermingle. It enables handling errors and state.[4].

As a strategy to further separate state and logic, one can construct a three-layered architecture, called the three layer Haskell cake. Here, the strategy is that one implements simple effectful functions, containing no logic as a base layer. Then on a second layer one implements an interface that implements a pure solution and one effectful solution. Then on the third layer one implements the logic of the program in pure code.

So while no exact definition of Functional programming exist, this thesis defines it as making functions pure and inheritance being based around functionality rather than attributes.

More advanced constructs also exists for functional programming that need to be introduced for constructing a maintainable rest library.

3.1.1 ADTs: Sum types and product types

A type is in Haskell a *set* of possible values that a given data can have. This can be *int*, *char* and custom defined types. A *sum type*, *Algebraic data type (ADT)* or *union*

[1]Read more: `en.wikipedia.org/wiki/Functional_reactive_programming`
[2]Read more: `stackoverflow.com/questions/17002119/haskell-pre-monadic-i-o`
[3]Read more: `https://en.wikipedia.org/wiki/Clean_(programming_language)`
[4]This is simplified as Monads are notoriously difficult to explain.

type is a type which is the sum of types, meaning that it can be one of those it's given types. For example the type `type IntChar = Int | Char` is either an Int or a Char. A useful application for sum types are enums such as `type Color = Red | Green | Blue`, meaning that a value of type Color is either red, green or blue. A sum type can be used to model data which may or may not have a value, by introducing the Maybe type: `type Maybe value = Just value | Nothing`.A product type is a type which is the product of types, for example `type User = User Name Email`. Informally, a product type can be likened to a record in Javascript. This allows us to model computations that might fail. For example given $sqrt(x) = \sqrt{x}$, $x \in \mathbb{Z}$ then $sqrt(-1)$ is undefined and would cause Haskell to crash. Instead by introducing a function `safeSqrt`, where `safeSqrt x = if x > 0 then Just (sqrt x) else Nothing`, the program can force the developer to handle the special case of negative numbers.

3.1.2 GADT

a GADT is a *generalized abstract data type*. They specify, depending on the input, what the output should be of that type. GADT enables implementing *domain-specific languages* (DSL). A DSL is a language with a limited scope for specific applications. For example a parsing library or a calculator.

```
1  data Calculator where
2    Number : Int -> Calculator Int
3    Add : Calculator a -> Calculator b -> Calculator c
4    Multiply : Calculator a -> Calculator b -> Calculator c
```

Figure 3.1: A Calculator GADT with two operations add and multiply.

```
1  mathExpression = (Number 5 `Add` Number 3) `Multiply` (
       Number 4 `Add` Number 3)
```

Figure 3.2: A mathematical expression constructed using the GADT in figure 3.1

Figure 3.1 defines a GADT for a calculator. The calculator can do two operations, add and multiply. This allows us to construct mathematical expressions. The expression in Figure 3.2 can translates to $(5+3)*(4+3)$ by defining a way to evaluate the expression. Figure 3.3 defines an evaluation for the program.

```
1          evaluate :: Calculator -> Int
2          evaluate (Add expr1 expr2) = evaluate expr1 + evaluate expr2
3          evaluate (Multiply expr1 expr2) = evaluate expr1 + evaluate
               expr2
4          evaluate (Number i) = i
```

Figure 3.3: Evaluator for the calculator

By separating how the expression from it's evaluation, the expression can be reused for different purposes. For example a calculator that should for different platforms just needs to implement different evaluators and can be sure that logic will be the same for all platforms.

3.1.3 Type classes

A type class is a construct that allows for ad hoc polymorphism. This allows to create constraints to type variables in parametrically polymorphic types. In English, that means that it allows creating interfaces that must be implemented for the types. For example the equality type class, defined in Figure 3.4

```
1          class Eq a where
2            (==) :: a -> a -> Bool
3            (/=) :: a -> a -> Bool
```

Figure 3.4: Equality type class in Haskell.

By defining an Equality type class one can create general functions that can be used for anything that is "equalable". For example Figure 3.5 is a function that prints a text if two items are equal. This function can be used for floats, ints, tuples and everything else that implements the `Eq` type class. Other uses for type classes is Num which implements numeric operations for floats and integers. This is useful for implementing the MTL technique which will allow us to implement the Interpreter pattern which will be described in the following sections.

```
printIfEqual :: Eq a => a -> a -> IO ()
printIfEqual a b =
    if a == b then
        putStrLn "They are equal"
    else
        putStrLn "They are not equal"
```

Figure 3.5: A function that prints a text if the two items are equal.

3.1.4 Functors and Contravariant Functors

The Functor type class defines a function $map : (a \to b) \to m\ a \to m\ b$. So every type that can be mapped over is a Functor. Examples of this are lists, where map morphs every value in the list from a to b. Another example is for Maybe, defined in 3.1.1. A Functor for Maybe checks if the value is $Just\ a$, if so it morphs that value to $Just\ b$, otherwise it returns $Nothing$.

Not every type with a type parameter is a Functor. For example the type $Predicate\ a = a \to Boolean$, is a function that when given some value a returns a boolean. This type can not be a Functor due to the type parameter being the *input* of the function. When the type parameter of the type is the input, it is said to be in negative position and the type is *contravariant*. When the type parameter is the output of a function, it is said to be in positive position and the type is covariant. A type can be a Functor only if it is covariant.

Contravariant Functor type class define a function $contramap : (a \to b) \to m\ b \to m\ a$. These are useful for defining how the value should be *consumed*. So for example a $type\ encoder = a \to encoded$, defines an encoder. The contravariant functor would allow transforming the encoder into intermediate value.

3.1.5 Brief introduction to Monads for side effects

Monads[5] are a way to sequence computations that might fail while automating away boilerplate code. Figure 3.6 shows how Monads are implemented as a typeclass in Haskell. It implements the function `return`, the function bind `(>>=)`, the function sequence `(>>)` which is bind whilst ignoring the prior argument and `fail` which handles crashes.

[5] `en.wikipedia.org/wiki/Monad_(functional_programming)`

```haskell
class Monad m where
    return :: a -> m a
    (>>=) :: m a -> (a -> m b) -> m b
    (>>) :: m a -> m b -> m b
    fail :: String -> m a
    fail msg = error msg
```

Figure 3.6: Monad type class in Haskell.

Informally, Monads are as a design pattern that allows us to sequence different computations. Without them the developer would have to explicitly check if a computation has failed. For example, given the function $unsafeSqrtLog = sqrt \circ log$, then $unsafeSqrtLog(-1)$ would throw an error since log and $sqrt$ are undefined for -1. Section 3.1.1 showed how the `Maybe value` type could be used to create a safe computation `safeSqrt`. To sequence that computation with a function `safeLog`, the user would have to manually check that `safeSqrt` returned a value `Just result` and not `Nothing`. Monads allows sequencing these computations without explicitly writing this check, so composing `safeSqrt` and `safeLog` using bind becomes `safeSqrtLog n = safeSqrt n >>= safeLog`. The same idea applies for effectful computations such as fetching data from a database.

3.2 SOLID principles with functional programming

Chapter 2 established the SOLID principles and how they can be used as indicators to ensure that software is maintainable. SOLID is originally used for Object-oriented programming and thus the concepts must be translated to functional programming.

3.2.1 Single Responsibility Principle

A function takes a single input and produces a single output. If file structure is centered around the morphisms of a single type then the responsibility of a file is to morph that type into some other value. Thus it keeps the modules focused and simple and would ensure that the single responsibility principle is held. It can also be thought of as "One function modifies one thing". So in summary, a program follows the Single Responsibility Principle if

1. Each functions performs only a morphism, which is guaranteed if the function is pure.

2. The file does not contain functions that do not contain any of the types declared within that file.

3.2.2 Liskov Substitution Principle

Liskov's Substitution Principle states how reasoning about subtyping among objects should be done. Since objects do not exist for functional programs some translating is needed of those concepts. The formal requirements of Liskov's Substitution Principle are as follows:

- Contravariance of method arguments should be in the subtype.
- Covariance of method arguments in the subtype.
- No new exceptions should be thrown by each subtype, except where those exceptions are themselves subtypes of exceptions thrown by the supertype.

Let's break each of these down to see how they translate in functional programming:

Contravariance of method arguments should be in the subtype. In functional programming, given a type which is contravariant. In order to morph that into a subtype the function contramap must be used. Recall that $contramap : (a \to b) \to t\ b \to t\ a$. For a function $from \to to$ to exist $from$ must be a subtype of to! Thus this is guaranteed.

Covariance of method arguments in the subtype. Same logic as for contravariance applies for covariance. In order to morph a subtype using map there must be a function $subtype \to type$.

No new exceptions should be thrown Depending on how the program implements exception it is impossible for a function to throw a new exception that have not been declared in it's return type.

3.2.3 Dependency Inversion Principle

Dependency Inversion Principle states that the logic should not depend on it's environment. To achieve that in functional programming the environment can be abstracted and taken as parameters of the program. For instance given the program readNPrint in Figure 3.7, this program depends on the computer IO, making it difficult to extend it to different environments, such as databases.

```
readNPrint : IO ()
readNPrint = readLine >>= putStrLn
```

Figure 3.7: A program that reads input from the computer and then prints it.

Instead, Figure 3.8 shows how the parameters are abstracted and readNPrint is a higher order function instead that takes some function that can generate a string and some function that can print a string.

```
readNPrint : (IO String) -> (String -> IO ()) -> IO ()
readNPrint reader printer = reader >>= printer

-- and then later
consoleIO : IO ()
consoleIO = readNPrint readLine putStrLn
```

Figure 3.8: A program that reads input from the computer and then prints it, where the logic is separated from it's environment.

This way, the dependencies can be mocked and replaced with different ones. So if we later want to create a *applicationIO* we can reuse *readNPrint* with the functions for printing in the application and reading input from the application.

Now for a REST api library, it means that the logic should not depend on it's environment means that the specification of the REST api should not depend on the server implementation. In other words, it should be trivial to port the server logic to another runtime if needed. To do this, GADTs can be used to separate the expression from it's evaluation. So the REST api is simply described as instructions of a GADT.

3.2.4 Interface Segregation Principle

Interface Segragation Principle states that no client should be forced to depend on methods it does not use. This translates to, in Functional programming, that the smallest set of data should be used for each function to work. Recall earlier that types can be thought of as sets. Recall also that the cardinality of a set is the amount of possible values that set can have. If the cardinality of a type is higher than expected it allows introducing illegal states. For example, $type\ Color = Blue|Green|Red$ has

a cardinality of 3 (since it can either be Blue, Green or Red) whereas Fig. 3.9 has a cardinality of $2 \cdot 2 \cdot 2 = 8$ meaning that it has 5 states that are impossible! By choosing the right data structure it lowers the amount of possible values that are possible. So Interface Segragation Principle in Functional programming states that a function should not be able to produce values it does not use.

```
1          type Color = { Blue: Bool, Red: Bool, Green: Bool}
```

Figure 3.9: Product type Color with cardinality too high

Observe that in Fig 3.10, the type `IUserRepo` has two operations, one which is not needed by the `getUserEndpoint` (why would that function need to storeUser?). Thus the function is capable of having more values than it should be. This means it breaks the Interface segregation principle.

```
1          data IUserRepo =  {
2              getUser : Id -> IO User,
3              storeUser : User -> Id -> IO ()
4          }
5
6          -- Later on
7          getUserEndpoint : IUserRepo -> Request -> Response
8                  -- ...
```

Figure 3.10: Normal interface for operations

3.2.5 Open/Closed principles

Open/Closed principle states that software entities (classes, modules, functions, etc.) should be open for extension, but closed for modification.

The OCP is an advice on how to write modules in such a way that we have backwards compatibility and so that if extra functionality is needed, the modifier does not need to look at the class in order to make modifications. So if a class has some new requirements you do not need to modify the source code but can instead extend the superclass.

When this principle is applied into Functional programming, we run into a well known problem called the expression problem. The expression problem states that

"The goal is to define a datatype by cases, where one can add new cases to the datatype and new functions over the datatype, without recompiling existing code and while retaining static type safety (e.g., no casts)."

(ADD_REFERENCE `http://homepages.inf.ed.ac.uk/wadler/papers/expression/expression.txt`)

The similarity with expression problem and OCP is that you want to be able to extend the program (add new cases to the datatype) without recompiling existing code. Object-oriented programming uses classes that should be open for extension and closed for modification. In functional programming, when this principle is applied, new cases to datatype should be possible and new functions. OCP exists because modifying battle tested code is dangerous and might cause regressions. Thus a preferable solution is to extend the previous code instead.

Example: Creating a OCP compliant paint programming

In a paint program, various shapes should be possible to paint: circles, squares, stars and custom shapes. It should also have a custom menu depending on the shape, a circle should be able to set the radius, a square the area and stars the diameter.

A functional approach is to create a sum type of the shape, seen in Fig 3.11. To add more shapes, the original source code would need to be modified. This means that Open/Closed principle is not being followed. It can cause a lot of trouble down the line, one function is acceptable but what if we had thousands of functions that depended on shape. Adding one shape would mean changing thousands of lines of code scattered all over the place.

```
1                type Shape
2           = Star size
3                  | Custom [vector]
4                  | Circle radius curvature
5                  | Square size
6
7                render : Shape -> IO ()
8                render shape =
9                        case shape of
10                               Star size = Star.render(size)
11               ...
12
13               -- Do the same thing
14               renderMenu : Shape -> IO ()
15               renderMenu = ...
```

Figure 3.11: A sum type of shapes

A different approach is by using type classes and contravariance. In order to render shapes, there needs some general format which we can use to render them. Let's assume we have some function $render : SetVector- > IO()$ for rendering. This is great because we know that any shape can be represented as a set of vectors in the end. Let us define $typeRenderablea = Renderable(a- > SetVector)$. Now it becomes possible to define a render function $render : Renderablea- > a- > IO()$, that works for all shapes. Shapes can be made in separation now by contramapping properties, seen in fig 3.12.

```
1                type Renderable a = Renderable (a -> Set Vector)
2                instance Contravariant a => Renderable a where
3            contramap cf b = \a -> b $ cf a
4
5                type Circle = {radius: Int}
6        circle : Renderable Circle
7        circle = circleToVector . radius -- circleToVector turns it
             to vector
8
9        setRadiusFactor :  Int -> Renderable Circle -> Renderable
             Circle
10       setRadiusFactor factor = contramap ({radius = factor})
11
12               type Custom = {scale : Int, shape : Set Vector}
13               custom : Renderable Custom
14               custom = scale * shape
15
16               addVertex : Vector -> Renderable Custom ->
                     Renderable Custom
17               addVertex vertex = contramap (Set.union vertex)
```

Figure 3.12: A contravariant approach to shapes

Contravariance forces adherence to a certain interface but leaves it open to extension, in spirit to the Open/Closed Principle. A separate part of the code can contain Renderable Square, Renderable Star without modifying the original code. Thus, in Functional programming, contravariance and type classes can enable OCP compliant code that solves the expression problem.

3.3 Functional servers

A *server* is a function of type $Request \rightarrow Responseencoded$. Request is a product of the path, media type, accept header, content type header and a body. The path needs to be parsed to figure out which endpoint to use. Based on the accept header and content type header the correct encoders and decoders also need to be set.

Based on the REST api description outlined in chapter 2, we can define a *type specification a b* as a GADT which constructs a small DSL for a REST api. For $specification$ there exists a function $makeServer : \forall a\ a.\ Specification\ a\ a \rightarrow (Request \rightarrow Responseencoded)$. $makeServer$ works as a parser, it extracts the parameters from the request based on the specification and feeds them into a handler

function, which is also constructed from the specification. Specification encodes the response based on the request's accept header and the included encoder in the specification.

The same *specification* value can also be used afterwards to generate documentation, client functions for generating correct requests.

A verb is either get, post, put, patch or delete, represented as a sum type $type\ Verb = Get \mid Post \mid Put \mid Patch \mid Delete$. Content types are harder to represent as a sum type due to there being so many and are thus represented as a string instead.

Request and Response

A Request is a product type consisting of a url, queries, accept header (Media), headers (a list of string tuples), accept media, verb, encodded body and an encoding. The correct encoder can be deduced from the accept header.

A Response is a product type consisting of a body, status code, content type Media, headers (a list of string tuples) and an *encoder for the body*. An encoded response instead contains an encoded body. These are separated because the body will be constructed separately from the encoder.

```
1   type Response a = {
2     code: StatusCode,
3     headers: Header ,
4     contentType: MediaType,
5     body: a -> encoded,
6     encoding: Encoding.t,
7   };
```

A server of parameter b is a type that takes a request and transforms it into a response of parameter b. I.E. $server : Request \rightarrow Response\ encoded$.

Of course, depending on the Request a different Response should be returned. So we need to introduce a *Router* that can, depending on the request, give a different response.

A Router is a parser that checks the content of the request and then modifies the response and applies the parameters of the request to a handler. We separate how the response should be encoded from the handler to ensure inversion of control. By making the implementation of the handler not depend on it's environment, we make it possible to test effects (E.G. adding a resource to a database, doing system calls) separate from the implementation of the specification.

In order to do so, we need to create a small language for creating specifications. This is done by using GADTs. Recall that a GADT is a sum type where you can

specify the return types. So the GADT for a specification becomes as follows:

```
1    type router 'input 'output =
2      Top : router ('a, 'a) ('a, 'a)
3      Exact : string -> router ('a, 'b) ('a, 'b)
4      Custom : (string -> Maybe a) -> router ('a -> 'b, 'c) ('b, 'c
           )
5      Query : string
6              -> (string -> Maybe 'a)
7              -> router (Maybe 'a -> 'b, 'c) ('b, 'c)
8      Slash : router ('a, 'b) ('c, 'd)
9              -> router ('c, 'd) ('e, 'f)
10             -> router ('a, 'b) ('e, 'f)
11     Integer : router (int -> 'a, 'b) ('a, 'b)
12     Method : HttpMethod -> router ('a, 'b) ('a, 'b)
13     Accept : [(MediaType, 'b -> string)] -> router ('a, string) (
           'a, 'b)
14     ContentType : [((MediaType, string -> Maybe 'a)]
15                   -> router ('a -> 'b, 'c) ('b, 'c)
16     Map : StatusCode
17         -> 'a
18         -> router ('a, string) (Result.t('c), 'c)
19         -> router ('(b, string) -> string) (Result (Response.
               content string))
20     OneOf :
21           [router ('(b, string) -> string) (Result (Response.
                 content string))]
22           -> router ((b, string) -> string) Request;
```

This looks complicated, but it essentially describes the operations of a REST api and routing. Top, Exact, Custom, Integer are used to match that URI is correct, for example an endpoint `/user/:id` is created by doing `Slash (Exact "user") Integer`. `Slash` connects operations together. Afterwards `Map` is used to map a side effect to the specification and `OneOf` is used to take a list of mapped routers and combine them into one. The infix operator for Slash is `(>-) = Slash`. So a valid specification can be `get = Method GET >- Exact "user" >- Integer`.

Router is simply an *instructions* for how it wants to be evaluated. That means it can be used to generate documentation for every endpoint, by constructing a function `document : router ((b, string) -> string) Request -> String`, the REST API library can automatically generate the documentation for us!

So since a Router is just instructions, we can construct `evaluate : router((b, string) -> string) -> (Request -> Response encoded)`, I.E. the router specification creates a server! This thesis will omit most of the details in regards to how this is implemented. The only thing to note is that `Map` will feed the parameters

from the request into a handler a. This means that, for example, a database function which fetches users will automatically not have to, in any way, concern itself with how that user is then encoded or what status message to return. This means that the server library *forces* the user to implement inversion of control, single responsibility principle and interface segregation principle. It forces inversion of control, because the evaluation of the specification is not dependent on the handler (an evaluator can just ignore it). It forces interface segregation interface principle because the handler is only fed the exact values, nothing more, for it to work and it can not return more than what the specification specifies that it needs to encode the value. Lastly it forces single responsibility principle by forcing the user to construct a separate function for fetching the user.

Chapter 4

Method

To evaluate if the functional approach to creating servers is more maintainable than existing solutions, a comparative study will be done. A popular library for developing server applications is by using an unopiniated solution using Express, which is a good candidate to compare to Cause.

Express is an unopiniated server framework written for Node.js for Javascript. That a framework is unopiniated means that it does not force you to architecture your code in any specific way.

4.1 Constructing the server in Express and Cause

To measure the maintainability, a comparison will be made by comparing a correct construction of an idiomatic server both made in Cause and the popular framework for Node Express. They will feature similar functionality which is a library api with the endpoints:

- `GET ‘‘api/books?released=int&author=string’’` Get a list of books and optionally ask for a specific author or a book from a specific year
- `DELETE ‘‘api/books/:id’’` Delete a book with a specified ID.
- `POST ‘‘api/books/:id’’ OR ‘‘api/books/’’` Create a new book or override a specific book

The server will also make use of a hashmap for database connection.

The accepted content types will be `application/json` and `www-url-formencoded` for all endpoints and the displayable content-types are `text/plain` and `application/json`.

Both implementations will handle all of the error cases. They will also be written in an idiomatic way, that is they will not take the challenges outlined in Chapter 2 into consideration; the only requirement is that they compile.

The two solutions are in the appendix, with the ReasonML implementation in section 5.6 and the Node solution in section 5.7

4.2 Evaluating maintainability

The aspects that to be evaluated when measuring maintainability were discussed in Chapter 2. This study will focus on evaluating the following criteria:

- Testability
- Extendability
- Readability
- Error-proneness

Chapter 3 established the SOLID guidelines in Functional programming as guidelines for maintainable software. However these guidelines do not state anything about the readability of the software. Thus two different methods will be needed to be used to evaluate the readability and to measure the testability, extendability and error-proneness.

To evaluate the testability, error-proneness and extendability, a comparison between the two solutions, one in ReasonML and the other in Nodejs, and it's adherence to the SOLID principle can be done since SOLID was translated to Functional programming in Chapter 3, where deductive reasoning can be used to find if the solution follows each criteria.

4.2.1 Evaluating readability through code reviews

Code reviews, also known as peer reviews, is an activity where a human evaluates the program to check for defects, finding better solutions and find readability aspects.

To measure the readability of the REST library, a semi-structured code review is conducted on five different people with varying knowledge of REST apis and functional programming.

Semi-structured interviews

Semi-structured interviews diverges from a structured interview which has a set amount of questions. In a semi-structured interview the interview is open and allows for new ideas to enter the discussion.

Semi-structured interviews are used to gather focused qualitative data. It is useful for finding specific insights in regards to the readability of the code and provides insights as to the code can actually be understood by the general user.

To conduct an semi-structured interview, the interview should avoid leading questions and use open-ended questions to get descriptive answers rather than yes or no answers.

The questions that will be asked are presented below.

- What is your experience with RESTful APIs?
- What is your experience with Express?
- What is your experience with ReasonML?
- After being presented the code api, can you explain what it does?
- Which media types does the endpoint post accept?
- What is the uri of DELETE
- Which media types representations can the endpoint show?
- Can you demonstrate how you would extend the api and add a new endpoint for a PUT request.
- Looking at the javascript api, can you explain what it does?
- Which media types does the endpoint get accept?
- Which content type and accept does post have?

4.2.2 Evaluating the answers

After performing the interviews conclusions can be made by interpreting the answers to conclude if the code is readable or not. If the code is readable the users being interviewed should be able to explain to the author what the code does.

In order to reduce the bias in the experiments each user will be shown a different code base first. So the 3 users will be shown the implementation in ReasonML and 2 users will be shown the implementation in Express. This also shows gives insights as to the reason the users do not understand the solution.

So in summary, the way each aspect of maintainability will be evaluated in both solutions by the following:

Testability Evaulated by comparing the number of dependencies that need to be mocked.

Extendability Evaluated by comparing to SOLID principles.

Readability Evaluated by comparing to SOLID principles.

Error-proneness Evaluated by SOLID principles and the interviews where we ask to extend the solution with a PUT request.

Afterwards from there a discussion can be had about the strengths and weaknesses of both solutions and the impacts of maintainability by using functional programming for developing REST servers.

Chapter 5

Results

With the method outlined in the previous chapter, the interview was performed on 4 different people, which is a bit less than the recommended by Norman group of five people due to difficulty finding enough users (ADD_REFERENCE `https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/`).

5.1 Evaluating adherence to SOLID

5.1.1 Functional solution

Single Responsibility Principle

Recall that the Single Responsibility Principle for functional programming states that all modules should revolve around one type.

Open/Closed Principle

Liskov Segregation Principle

Interface Segregation Principle

Dependency Inversion Principle

5.1.2 Imperative solution

Single Responsibility Principle

Open/Closed Principle

Liskov Segregation Principle

Interface Segregation Principle

Dependency Inversion Principle

5.2 Interviews

5.2.1 Raw results

Person 1

Person 2

Person 3

Person 4

5.3 Bias

5.4 Limitations

TODO

5.4.1 Improvements to implementation

TODO

5.5 future work

5.5.1 Relations to cardinality

TODO

Bibliography

[1] T. Ishida, Y. Sasaki, and Y. Fukuhara, "Use of procedural programming languages for controlling production systems," in *[1991] Proceedings. The Seventh IEEE Conference on Artificial Intelligence Application*, vol. i, pp. 71–75, Feb 1991.

[2] A. M. Turing, "On computable numbers, with an application to the entscheidungsproblem," *Proceedings of the London Mathematical Society*, vol. s2-42, no. 1, pp. 230–265, 1937.

[3] G. S. G. B. A. B. James Gosling, Bill Joy, "The java language specification," Febuary 2015. `https://docs.oracle.com/javase/specs/jls/se8/jls8.pdf`, accessed Feb 2018.

[4] D. A. Turner, "Some history of functional programming languages," in *Proceedings of the 2012 Conference on Trends in Functional Programming - Volume 7829*, TFP 2012, (New York, NY, USA), pp. 1–20, Springer-Verlag New York, Inc., 2013.

[5] B. J. Copeland, "The church-turing thesis," in *The Stanford Encyclopedia of Philosophy* (E. N. Zalta, ed.), Metaphysics Research Lab, Stanford University, winter 2017 ed., 2017.

[6] R. S. Pressman, *Software Engineering: A Practitioner's Approach*. New York, NY, USA: McGraw-Hill, Inc., 6 ed., 2005.

[7] M. C. Robert, "Design principles and design patterns," September 2015. originally `objectmentor.com`, archived at `https://fi.ort.edu.uy/innovaportal/file/2032/1/design_principles.pdf`, accessed Feb 2018.

[8] "Pure function," Aug 2018. `https://en.wikipedia.org/wiki/Pure_function`, accessed Feb 2019.

Appendix

5.6 ReasonML REST implementation

5.7 NodeJS REST implementation