

(Clojure (for the masses))

(author “Tero Kadenius” “Tarmo Aidantausta”)

(date “27.03.2011”)

Contents

1	Introduction	3
1.1	Dialect of Lisp	3
1.2	Dynamic typing	3
1.3	Functional programming	3
2	Introduction to Clojure syntax	3
2.1	Lispy syntax	4
2.1.1	Parentheses, parentheses, parentheses	4
2.1.2	Lists	5
2.1.3	Prefix vs. infix notation	5
2.1.4	Defining functions	5
3	Functional vs. imperative programming	6
3.1	Imperative programming	6
3.2	Object oriented programming	6
3.3	Functional programming	7
3.3.1	Functions as first class objects	7
3.3.2	Higher-order functions	8
3.4	Differences	9
3.5	Critique	9
4	Closer look at Clojure	10
4.1	Syntax	10
4.1.1	Reader	10
4.1.2	Symbols	10
4.1.3	Literals	10
4.1.4	Lists	10
4.1.5	Vectors	11
4.1.6	Maps	11
4.1.7	Sets	11
4.2	Macros	11
4.3	Evaluation	11
4.4	Read-Eval-Print-Loop	11
4.5	Data structures	11
4.6	Data types	11
4.7	Control structures	11
5	Concurrency/parallelism in Clojure	11
5.1	From serial to parallel computing	11
5.2	Problems caused by imperative programming paradigm	12
5.3	Simple things should be simple	12
5.4	Reference types	12
5.5	Software transactional memory system (STM)	12
	References	13

1 Introduction

This seminar paper introduces Clojure, a dialect of Lisp, which is a dynamically typed functional programming language hosted on virtual machines like JVM and CLR.

The key point that makes Clojure stand out in the pool of new and fancy programming languages is the fact it has been developed from the ground up with multithreaded programming in mind. *Immutable data structures*, support for *software transactional memory* (STM) with the combination of *functional programming* paradigm give Clojure good tools working with concurrency. [30]

Even though concurrency had an emphasis in designing Clojure it is still a general purpose language with interoperability features to leverage the functionality and the ecosystem of the host platform, so it's possible to reuse already existing code and libraries done for the host platform. [26, 25]

The paper gives first a short tour to very basics of Clojure and dives into further details of the language and its features.

1.1 Dialect of Lisp

```
(= 2 (+ 1 1))
```

“Lisp is a family of computer programming languages based on formal functional calculus. Lisp (for "List Processing Language") stores and manipulates programs in the same manner as any other data, making it well suited for "meta-programming" applications.” [1]

“Lisp has jokingly been called "the most intelligent way to misuse a computer". I think that description is a great compliment because it transmits the full flavor of liberation: it has assisted a number of our most gifted fellow humans in thinking previously impossible thoughts.” [22]

1.2 Dynamic typing

“In a *dynamically typed* language, values have fixed types, but variables and expressions have no fixed types.” [36]

1.3 Functional programming

“In functional programming, the model of computation is the application of functions to arguments.” [36]

2 Introduction to Clojure syntax

To make it easier to read this paper, parts of the syntax of Clojure are introduced here with simple shortly explained examples. There are a lot of concepts that

are not explained but will be explained later in the article.

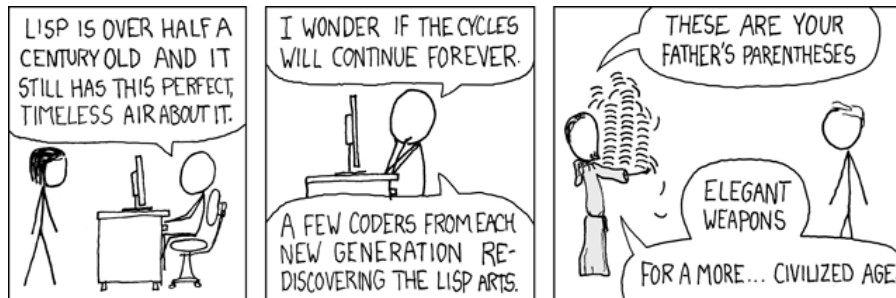
2.1 Lispy syntax



“At once, just like they said, I felt a great enlightenment. I saw the naked structure of Lisp code unfold before me.” [32]

Lisps, like Clojure, don’t have a lot of syntax in the traditional sense of syntax, compared to languages like Java, C or C++. The syntax is minimized to mainly into defining lists with parenthesis and the different rules of evaluations of those lists.

2.1.1 Parentheses, parentheses, parentheses



“These are your father’s parentheses. Elegant weapons, for a more... civilized age.” [33]

As Lisps are about lists - the name LISP derives from "LIST Processing" [31], and lists in Lisps are about parentheses. This means that in Clojure reading and writing parentheses is inevitable. A lot has been said about the amount of parentheses in Lisps, both good and bad.

“Lisp has all the visual appeal of oatmeal with fingernail clippings mixed in.” [35]

“...and here I thought it was LotsofInfernalStupidParentheses. My mistake; I must have just been in a worse mood. ;->” [34]

2.1.2 Lists

In Clojure, as in all the Lisps, the lists have special syntax to them which is due to the homoiconicity - code is data and data is code. This might seem confusing at first but the rules are quite simple although they overload the definition of lists a bit.

Below we can see a list which evaluates to a function call to `+` with 1, 2, and 3 as parameters to that function.

```
(+ 1 2 3)
```

Below is a definition of list with numbers. Quote in the beginning of the line tells the reader that it should treat the following list just as data.

```
'(+ 1 2 3)
```

Below we can see how a list can be also constructed with a function call to a list.

```
(list + 1 2 3)
```

2.1.3 Prefix vs. infix notation

“Polish notation, also known as prefix notation, is a form of notation for logic, arithmetic, and algebra. Its distinguishing feature is that it places operators to the left of their operands.”[18]

```
+ 1 2
```

“Infix notation is the common arithmetic and logical formula notation, in which operators are written infix-style between the operands they act on (e.g. `2 + 2`).” [13]

```
1 + 2
```

Clojure, as all Lisps, uses prefix notation in contrast to infix notation is which used in languages like C, C++ and Java. [18, 13]

2.1.4 Defining functions

Functions are first-class objects in Clojure and there is more than one way of defining them.[24]

```
(def hai (fn [] "Ou hai!"))
```

Above we define a function that returns string “Ou hai!” by using *macros* called *fn* and *def*. *fn* that creates the function takes names and the default values for parameters inside `[]` as the first parameter and the body of the function as the second parameter.[23] A macro called *def* binds a the name with a value.[23]

You can also define functions with a *macro* called *defn*. [24]

```
(defn hai-u [u]
  (str "Hai, " u))
```

That *macro* takes the name, optionally a document string and attribute map, parameters and the function body as parameters.[23]

You can also use a Clojures dispatch macro to create a function. [27]

```
(def hai-u2 #(str "Hai, " %1 " and " %2))
```

3 Functional vs. imperative programming

3.1 Imperative programming

“Imperative programming is a programming paradigm that describes computation in terms of statements that change a program state.”[12]

Nearly all machine code implementations are written in imperative style. The contents of the memory holds the state and machine language instructions modify it. Higher-level imperative languages have more advanced features like variables and complex statements, but the basic idea remains the same. [12]

Here is a small snippet of imperative code. It has a notion of state (a), which is mutated. In addition, an IO operation is performed.

```
int a = 3;
int b = 4;
if (a < b) {
    a++;
}
print(a);
```

3.2 Object oriented programming

"The set of values of the attributes of a particular object is called its state. The object consists of state and the behavior that's defined in the object's classes." [16]

Object oriented programming provides a feature called encapsulation. Encapsulation prevents users of the object from directly modifying the data that forms the state by providing operations (methods) for doing it. This is done in order to ensure the validity of the internal state of the object. [6]

In other words, at its heart, object oriented programming tends to be imperative. The paradigm itself doesn't enforce it, but that is usually the case. [28]

An example that demonstrates the imperative nature of OO code:

```

class Foo {
    int a = 3;
    int b = 4;

    increment() {
        if (a < b) {
            a++;
        }
    }

    print() {
        print(a);
    }
    ....
}

```

What happens here is identical to the imperative code example except that in this case the data (a and b) and the operations mutating the state (a++) or causing other side effects (print(a)) are encapsulated inside the object.

increment() is an instruction for modifying the state of the object. The result of increment() may vary in different points of time depending on the state of the object.

3.3 Functional programming

“Functional programming has its roots in mathematics. Instead of providing instructions for modifying the state of the program, functional programming emphasizes the application of functions and avoid state and mutable data in general. “[9]

3.3.1 Functions as first class objects

The notion of a function is not unique to functional programming languages. However, functional languages have what is called *first class functions*. This means that functions have a central role in the code, much like objects do in OO languages. Functions can be stored to data structures and the use of *higher-order functions* is common. [7] The objective of having no side effects manifests itself in *pure functions*.

Function is considered pure if:

1. “The function always evaluates the same result value given the same argument value(s). The function result value cannot depend on any hidden information or state that may change as program execution proceeds or between different executions of the program, nor can it depend on any external input from I/O devices. [3]

2. “Evaluation of the result does not cause any semantically observable side effect or output, such as mutation of mutable objects or output to I/O devices.” [3]

Using pure functions has several benefits:

1. Pure expression can be removed without affecting other expressions if the result of the pure expression is not used. [9]
2. *Referential transparency*. An expression can be replaced with its value without causing changes to the program. The output is always the same with the same input. [19]
3. If a pure function does not depend on the result of another pure function, they can be performed in any order. Ie. they are thread-safe and can be run in parallel without typical concurrency issues. [9]
4. Lack of side effects guaranteed by the language, provides opportunities for compiler optimizations. [9]

3.3.2 Higher-order functions

Higher-order function is a function that either takes one or more functions as parameters or returns one as a value. [11] Well-known examples of higher-order functions are *map* and *fold* [8]. “Map is the name of a higher-order function that applies a given function element-wise to a list of elements and returns a list of results.” [14]. Fold is a “function that iterate an arbitrary function over a data structure in some order and build up a return value”. [8]

- Doubling the value of every element in a list using map:

```
(map (fn [x] (* 2 x)) '(1 2 3))
=> ((* 2 1) (* 2 2) (* 2 3)) => (2 4 6)
infix notation equivalent: (2 * 1) (2 * 2)(2 * 3)
```

- In Clojure fold is called reduce. A trivial example for calculating the sum of the elements in a list:

```
(reduce + '(1 2 3))
=> (+ (+ 1 2) 3) => 6
infix notation equivalent: (1+2) + 3
```

Partial function application and currying

Higher-order functions enable an interesting feature where a new function can be generated based on another function. *Partial function application* is a technique which produces a function in which one or more of the arguments of the original function are fixed. *Currying* resembles partial function application. The difference is that in currying each (curried) function takes only a single argument and

produces a function which takes one argument less than its predecessor. Ie. currying produces a chain of functions, whereas with partial function application arbitrary number of functions can be fixed at once. [5]

Out of the box, Clojure supports partial function application but not currying.

A simple example where a function that adds 2 to its argument is applied to a list of elements:

```
• (map(partial + 2) '(1 2 3))

=> ((+ 2 1)(+ 2 2)(+ 2 3))=> (3 5 7)
infix notation equivalent: ((2 + 1)(2 + 2)(2 + 3))
```

3.4 Differences

Characteristic	Imperative approach	Functional approach
Programmer focus	How to perform tasks (algorithms) and how to track changes in state.	What information is desired and what transformations are required.
State changes	Important.	Non-existent.
Order of execution	Important.	Low importance.
Primary flow control	Loops, conditionals, and function (method) calls.	Function calls, including recursion.
Primary manipulation unit	Instances of structures or classes.	Functions as first-class objects and data collections.

[10]

3.5 Critique

The proponents of functional programming claim that imperative programming is fundamentally broken - especially in a multi-threaded environment. First of all, there is an argument that the world doesn't function in a way imperative programming models it. When dealing with mutable state, the "world" has to stop in order it be examined or changed. This becomes a major problem when bringing concurrent programming to the picture. [28]

OO programming suffers from the same problems as imperative programming. To quote Rich Hickey, the creator of Clojure: "Encapsulation just means: I'm in charge of this spaghetti code." Ie. encapsulation doesn't change the fact that OO is usually based on mutable state. It just tries to prevent the user of object's interface from seeing it (the (imperative) spaghetti code).

4 Closer look at Clojure

Now that the paradigm of functional programming has been introduced, some of the details of Clojures features and terminology is explained.

4.1 Syntax

“Clojure is a homoiconic language, which is a fancy term describing the fact that Clojure programs are represented by Clojure data structures.” [27]

Clojure syntax is built on *symbolic expressions*, *S-expressions*, that are list based data structures. [20, 27, 4] In addition to lists also symbols, literals, vectors, maps and sets make up the syntax and are parsed by the Clojure reader. [27]

4.1.1 Reader

The reader parses the textual presentation of the Clojure code to data structures.

```
(doc read)
```

Then it creates the form of that same data structure that the compiler will see. Clojure compiler compiles the code, data structures, to host platform bytecode. This bytecode is then executed by the host platform virtual machine. [29]

4.1.2 Symbols

“Symbols begin with a non-numeric character and can contain alphanumeric characters and *, +, !, -, _, and ?.” [27]

```
(def ines "a symbol called ines")
```

def is a macro that takes a symbol as a parameter and then gives that symbol a value if one is given. Here it defines a symbol called ines in the current namespace with the value “a symbol called ines”. [23]

4.1.3 Literals

```
'((strings are literals as are numbers and characters (1 2 3) (\a \b \c))  
( and of course booleans and and keywords true :keyword))
```

4.1.4 Lists

```
(list "can contain anything you want" 1 2 3 \a \b \c :keyword '())
```

If you don't use ' or quote Clojure will try to use the first cell off a list as a function call. So if you want to just express data, use ' or quote.

'(any of these won't be evaluated)

4.1.5 Vectors

(vector '[can contain anything too 1 2 3 \a \b \c :keyword])

4.1.6 Maps

(hash-map :key value key :value :map {:can contain :maps too}))

4.1.7 Sets

(hash-set "can contain sets" #{:a :b} #{:b :c} "and anything unique" 1 2 3 \a \b)

4.2 Macros

<http://clojure.org/macros>

4.3 Evaluation

<http://clojure.org/evaluation>

4.4 Read-Eval-Print-Loop

<http://clojure.org/dynamic>

4.5 Data structures

http://clojure.org/data_structures

4.6 Data types

<http://clojure.org/datatypes>

4.7 Control structures

Here be something about loop-recur http://clojure.org/functional_programming
and do [http://clojure.org/special_forms#Special%20Forms-\(do%20exprs*\)](http://clojure.org/special_forms#Special%20Forms-(do%20exprs*)), http://clojuredocs.org/clojure_c
and Programming Clojure - Flow Control page 69-75,

5 Concurrency/parallelism in Clojure

5.1 From serial to parallel computing

For many years the speed of majority of computer programs could be improved by upgrading the hardware on which the program was run. Ie. “Frequency scaling was the dominant reason for improvements in computer performance from the mid-1980s until 2004.” [17] Therefore, the standard practice has been to write software for serial computation meaning that: [21]

- Software is run on a single computer on a single processor (core)
- Only one instruction may execute at any moment in time.
- Instructions are executed one after another.

Moore's law states that the number of transistors that can be placed in an integrated circuit doubles approximately every two years. The trend started in 1958 and is expected to continue until 2015 or 2020 or later. [15] The extra transistors cannot be used for increasing the frequency of the microprocessor, but they can be used for adding new processor cores for parallel computing [17].

5.2 Problems caused by imperative programming paradigm

As stated earlier, by avoiding mutable state there is no need for locking and threads cannot interfere with each other. Let's take a look at two common concurrency issues.

- “*A deadlock* occurs when two threads each lock a different variable at the same time and then try to lock the variable that the other thread already locked.” [2]
- “*Race condition* occurs when two threads access a shared variable at the same time.” [2]

As we can see these issues are not necessarily related to the underlying problem we are trying to solve. They are mere implementation details.

5.3 Simple things should be simple

Clojure has concurrency features built-in. Immutable data structures and pure functions enable trivially easy parallel processing of functions as long as there is no data dependency between them.

A simple example of easy parallel execution of functions using `pmap`, a parallel map implementation:

```
(pmap (fn [x] (+ x 2)) '(1 2 3))
```

There is no need for explicitly spawning new threads nor is there fear for race conditions or deadlocks since the function does not rely on external state or mutate anything.

5.4 Reference types

- atom, agent and ref
- identity and equality

5.5 Software transactional memory system (STM)

- refs require a transaction

References

- [1] Lisp programming language. Available from: http://en.wikiquote.org/wiki/Lisp_programming_language.
- [2] Description of race conditions and deadlocks. December 2006. Available from: <http://support.microsoft.com/kb/317723>.
- [3] Pure function. *Wikipedia*, 2010. Available from: http://en.wikipedia.org/wiki/Pure_function.
- [4] Clojure. *Wikipedia*, 2011. Available from: <http://en.wikipedia.org/wiki/Clojure>.
- [5] Currying. *Wikipedia*, March 2011. Available from: <http://en.wikipedia.org/wiki/Currying>.
- [6] Encapsulation (object-oriented programming). *Wikipedia*, 2011. Available from: http://en.wikipedia.org/wiki/Encapsulation_%28object-oriented_programming%29.
- [7] First-class function. *Wikipedia*, March 2011. Available from: http://en.wikipedia.org/wiki/First-class_function.
- [8] Fold (higher-order function). *Wikipedia*, March 2011. Available from: http://en.wikipedia.org/wiki/Fold_%28higher-order_function%29.
- [9] Functional programming. *Wikipedia*, 2011. Available from: http://en.wikipedia.org/wiki/Functional_programming.
- [10] Functional programming vs. imperative programming. web, 2011.
- [11] Higher-order function. *Wikipedia*, March 2011. Available from: http://en.wikipedia.org/wiki/Higher-order_function.
- [12] Imperative programming. *Wikipedia*, March 2011. Available from: http://en.wikipedia.org/wiki/Imperative_programming.
- [13] Infix notation. *Wikipedia*, 2011. Available from: http://en.wikipedia.org/wiki/Infix_notation.
- [14] Map (higher-order function). *Wikipedia*, March 2011. Available from: http://en.wikipedia.org/wiki/Map_%28higher-order_function%29.
- [15] Moore's law. *Wikipedia*, March 2011. Available from: http://en.wikipedia.org/wiki/Moore%27s_Law.
- [16] Object-oriented programming. *Wikipedia*, March 2011. Available from: http://en.wikipedia.org/wiki/Object-oriented_programming.
- [17] Parallel computing. *Wikipedia*, March 2011. Available from: http://en.wikipedia.org/wiki/Parallel_computing.

- [18] Polish notation. *Wikipedia*, 2011. Available from: http://simple.wikipedia.org/wiki/Prefix_notation.
- [19] Referential transparency (computer science). *Wikipedia*, 2011. Available from: http://en.wikipedia.org/wiki/Referential_transparency_%28computer_science%29.
- [20] S-expression. *Wikipedia*, 2011. Available from: <http://en.wikipedia.org/wiki/S-expression>.
- [21] B. Barney. Introduction to parallel computing. *Lawrence Livermore National Laboratory*. Available from: https://computing.llnl.gov/tutorials/parallel_comp.
- [22] E. W. Dijkstra. The humble programmer. *Commun. ACM*, 15:859–866, October 1972. Available from: <http://doi.acm.org/10.1145/355604.361591>, doi:<http://doi.acm.org/10.1145/355604.361591>.
- [23] R. Hickey. clojure.core - clojure v1.2 api documentation. *clojure.org*, 2010. Available from: <http://clojure.github.com/clojure/clojure.core-api.html>.
- [24] R. Hickey. Functional programming. *clojure.org*, 2010. Available from: http://clojure.org/functional_programming.
- [25] R. Hickey. Java interop. *clojure.org*, 2010. Available from: http://clojure.org/java_interop.
- [26] R. Hickey. Rationale. *clojure.org*, 2010. Available from: <http://clojure.org/rationale>.
- [27] R. Hickey. The reader. *clojure.org*, 2010. Available from: <http://clojure.org/reader>.
- [28] R. Hickey. Values and change - clojure’s approach to identity and state. *clojure.org*, 2010. Available from: <http://clojure.org/state>.
- [29] R. Hickey. Ahead-of-time compilation and class generation. *clojure.org*, 2011. Available from: <http://clojure.org/compilation>.
- [30] J. M. Kraus and H. A. Kestler. Multi-core parallelization in clojure: a case study. In *Proceedings of the 6th European Lisp Workshop*, ELW ’09, pages 8–17, New York, NY, USA, 2009. ACM. Available from: <http://doi.acm.org/10.1145/1562868.1562870>, doi:<http://doi.acm.org/10.1145/1562868.1562870>.
- [31] J. McCarthy. Recursive functions of symbolic expressions and their computation by machine, part i. *Commun. ACM*, 3:184–195, April 1960. Available from: <http://doi.acm.org/10.1145/367177.367199>, doi:<http://doi.acm.org/10.1145/367177.367199>.

- [32] R. Munroe. Lisp. web comic. Available from: <http://xkcd.com/224/>.
- [33] R. Munroe. Lisp cycles. web comic. Available from: <http://xkcd.com/297/>.
- [34] Unknown. Lost in a sea of parentheses. *c2.com wiki*, 2008. Available from: <http://c2.com/cgi/wiki?LostInaSeaofParentheses>.
- [35] L. Wall. Wherefore art, thou? *The Perl Journal*, 1, 1996. Available from: <http://www.linuxjournal.com/article/2070>.
- [36] W. Watt, David A. & Findlay. *Programming language design concepts*. John Wiley & Sons, Ltd., 2004. Available from: <http://books.google.com/books?id=vogP3P2L4tgC>.