目录

目录

[Scala 学习笔记 1](#_Toc464850043)

[概述 1](#_Toc464850044)

[Intellij IDEA 2](#_Toc464850045)

[入门 3](#_Toc464850046)

[语法 3](#_Toc464850047)

[基本类型 4](#_Toc464850048)

[语句 4](#_Toc464850049)

[面向对象化编程 4](#_Toc464850050)

[应用场景 5](#_Toc464850051)

# Scala 学习笔记

**Joshua Conero**

## 概述

日期： 2016年10月17日 星期一

参考网站：<http://www.scala-lang.org/>

语言含义/特性：[WHATE IS SCALA](http://scala-lang.org/what-is-scala.html) ?

**scalability**-可扩展性，

**Scalable Language**-可伸缩语言

**Object-Oriented** - 面向对象

every value is an object and every operation is a method-call

The language supports advanced component architectures through classes and traits: **Class-Trait-Object**

**Functional**- 函数式

Scala is also a full-blown functional language

**Seamless Java Interop**- 与Java无缝对接

Java libraries, frameworks and tools are all available. Build tools like ant or **maven**, IDEs like **Eclipse**, **IntelliJ**, or **Netbeans**, frameworks like **Spring** or **Hibernate** all work seamlessly with Scala. Scala runs on all common **JVMs** and also on **Android**.

**Functions are Objects**- 函数即对象

The function type is just a regular class. The algebraic data types found in languages such as **Haskell**, **F#** or **ML** are modelled in Scala as class hierarchies. Pattern matching is possible over arbitrary classes

To some, Scala feels like a scripting language - 像脚本语言

can play with it by typing one-line expressions and observing the results

can also rely on it for large mission critical systems, as many companies, including Twitter, LinkedIn, or Intel do.

**Future-Proof** - 未来

makes use of concurrent and synchronous processing, parallel utilization of multiple cores, and distributed processing in the cloud.

Its functional nature makes it easier to write safe and performant multi-threaded code.

相关词汇其他： REPL and IDE

Scala是纯面向对象的，每个值都是一个对象

Scala的设计目的是要和两种主流面向对象编程语言Java和C#实现无缝互操作，这两种主流语言都非纯面向对象。

Scala始于2001年，由**洛桑联邦理工学院**(EPFL)的编程方法实验室研发。

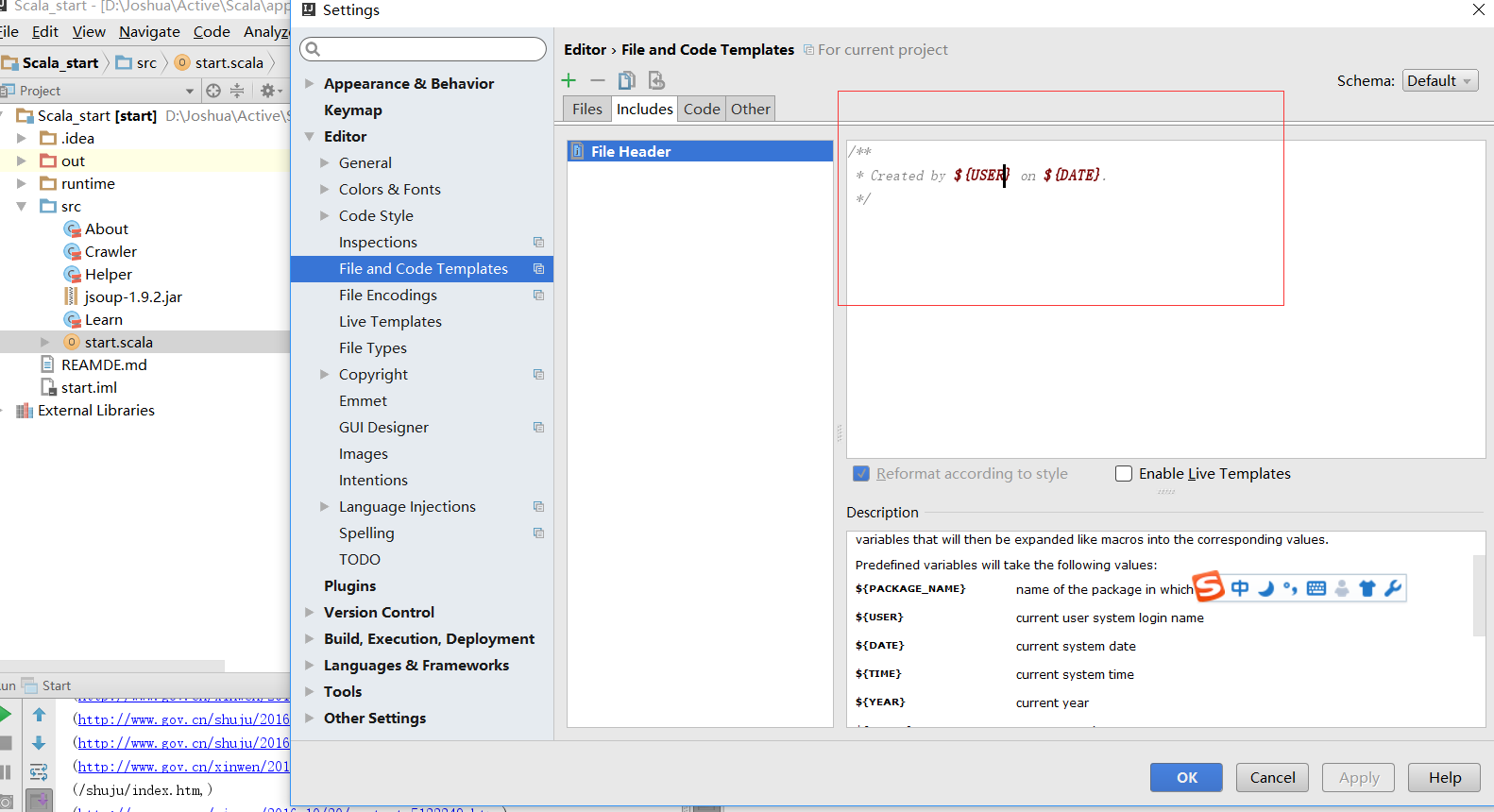
可运行JVM/.NET

### Intellij IDEA

网站: <http://www.jetbrains.com/idea/>

设置代码头注释格式:

1. 在IDEA环境中去设置，File->Settings->File and Code Templates(IDEA14)->Include->File Header:修改${USER}为自定义名字



\*来源: [http://www.tuicool.com/articles/QRfYZrB](%20http:/www.tuicool.com/articles/QRfYZrB)

## 入门

1.支持";"或者“空”结尾

### 语法

**基本类型定义**

**函数：**

**def** point(x:Int,y:Int): Int ={

}

说明： point函数返回 Int类型数据

**定义变量**：

var num = 1

常量：

val PI = 3.14

词法的两种模式：Scala模式与XML模式。

关键字：

**abstract case catch class def**

**do else extends false final**

**finally for forSome if implicit**

**import lazy match new null**

**object override package private protected**

**requires return sealed super this**

**throw trait try true type**

**val var while with yield**

?**柯里化的函数**

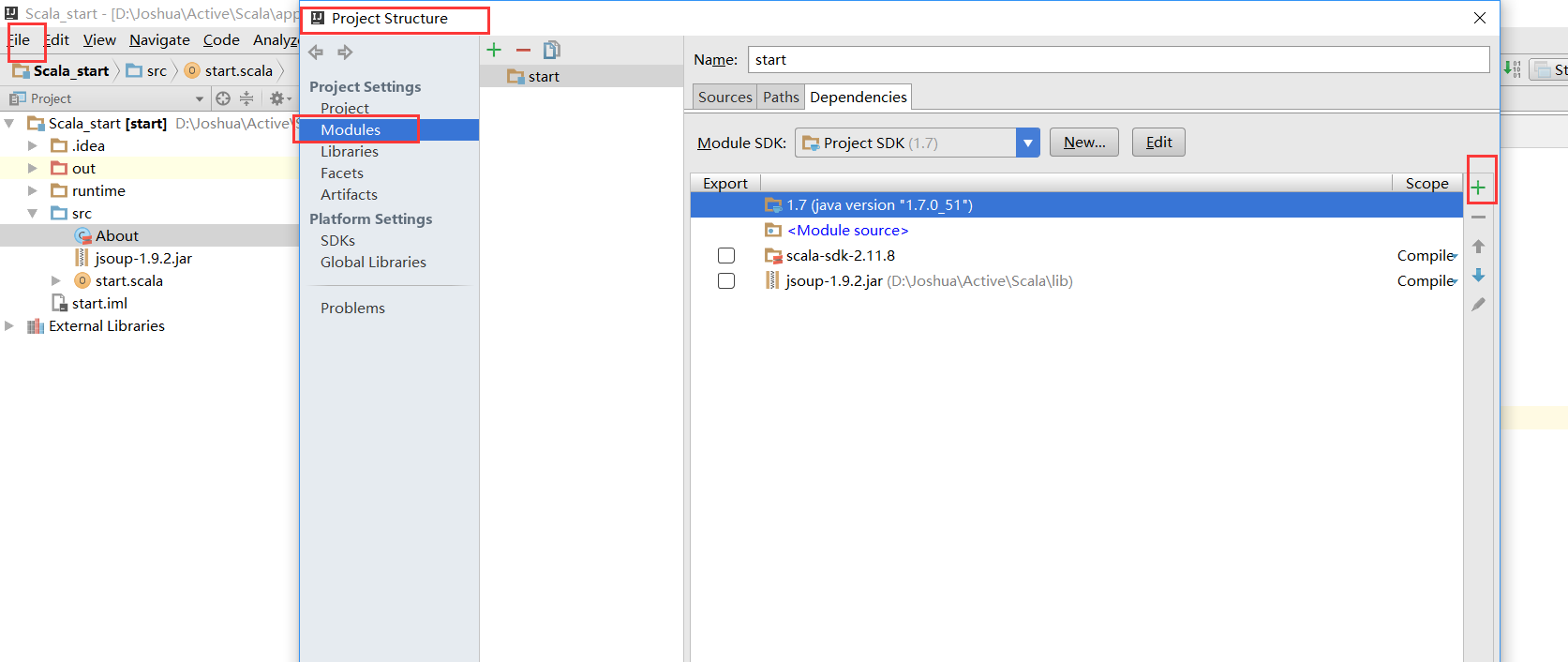
**绑定**：命名在局部定义与声明，继承，import子句，package子句中存在，这些可以统称为绑定。

**Java/Scala调用第三方Jar**:

1.**Intellij IDEA**

选择: >File>Project>Structrue>Modules>

选择:>Dependencies>左边 "+">1.Jars or ....



### 基本类型

### 语句

### 面向对象化编程

1. **private变量**：在Scala声明private变量会Scala编译器会自动生成get/set,在Scala中变量是需要初始化的，如果不声明private默认是public的

2. **trait**：在java中可以通过interface实现多重继承，在Scala中可以通过特征（trait）实现多重继承，不过与java不同的是，它可以定义自己的属性和实现方法体，在没有自己的实现方法体时可以认为它时java interface是等价的，在Scala中也是一般只能继承一个父类，可以通过多个with进行多重继承。

3. **Class**中的method的参数 默认 是 val，及不可修改类型

4.**object**:

Object 在Scala里面的引入可能就是因为scala没有静态的方法和字段，没有办法 实现类似C++的singleton模型所以引入object

Object由于上述原因，同类的差别在于没有状态，只有方法。换句话说成员中仅有val没有var，而且不能给object传递参数

Object定义并不表示声明了一个新的类型，定义一个变量是obj类型是错误的。

object对象类似于Java的静态类,它的成员、它的方法都默认是静态的。

构造函数：

Class 内可将游离的代码作为构造函数，级系统自己判断。

**类/构造函数:**

**class** Learn() {

**val** *u* = 0 // 属性  
 run()  
 **def** run():Unit = {  
 test() // this.test() 可省this.  
 run\_hack()  
 }

def test():Unit = {}

def run\_hack():Unit = {}

}

## 应用场景

#### 网络爬虫+数据分析+分析数据保存/ Scala\_start

概述：

1.java框架

2.Scala与java互操作

3.Scala数据分析以及保存处理

项目名称： **Scala\_start**

java爬虫框架介绍：

参考网站：

1. [给推荐几个github上优秀的java爬虫项目？](https://www.zhihu.com/question/31427895)

2.[网络爬虫的基本原理](http://www.cnblogs.com/wawlian/archive/2012/06/18/2553061.html)

[nutch](https://github.com/apache/nutch) apache下的开源爬虫程序，功能丰富，文档完整。有数据抓取解析以及存储的模块。

[Heritrix](https://github.com/internetarchive/heritrix3) 很早就有了，经历过很多次更新，使用的人比较多，功能齐全，文档完整，网上的资料也多。有自己的web管理控制台，包含了一个HTTP 服务器。操作者可以通过选择Crawler命令来操作控制台。

[crawler4j](https://github.com/yasserg/crawler4j)因为只拥有爬虫的核心功能，所以上手极为简单，几分钟就可以写一个多线程爬虫程序。

项目实际

一、引用库/jar

1. jsoup HTML 解析库(jsoup-1.9.2.jar)