# 什么是

诸君在爬取网页或者处理网页数据时，会经常遇见格式为json的文本文件，过往我们如果想从中提取重要信息，往往是用上一串很复杂的正则表达式来匹配。每每此时，小编就会想起Jamie Zawinski的一句名言“如果你要用正则表达式处理问题，那么现在你就有了两个问题”——将自己从正则表达式里解放出来吧——至少在遇到的文件格式是json时。本期将介绍一个如何让stata读取和输出json文件的命令：jsonio

http://5b0988e595225.cdn.sohucs.com/images/20170901/1e8cc74529c0460eb1040c6494e53de1.png

一．正本清源——什么是json

在回答什么是json以前，我们先思考这样一个问题：我们如何描述一份非结构化的信息，比方说小编与朋友们出去吃小龙虾，我们需要描述那天天气如何，去了几个人，吃了什么东西，每位小伙伴觉得口味怎样，类似这样的问题有很多，我们应该如何规范地记录它们？

计算机已经给出了它的答案，所有的非结构化信息可以用两种句式进行概括：①什么东西有什么 ②什么东西是什么

比如这次吃小龙虾，那么可以写成

吃小龙虾：{

天气：晴天，

成员：[小编，白富美，高富帅]，

花费：233元，

菜品：[好吃的，好吃的，还是好吃的]，

}

如此一来，所有此类信息都可以用此模板记录下来，我们说：吃小龙虾这个词汇对应的大括号的内容，是“吃小龙虾”有什么，而天气这个词汇对应的“晴天”代表天气是什么。

这种格式的数据就是json

Json是一种轻量级的数据交换格式，人好认，计算机也好解析，它的全称虽然是Java Object Notation，但可以完全脱离java使用，是理想的数据存取语言。在上述的例子中，所有的信息本质上都以键值对的形式记录了下来，如：天气是什么，花费是什么，而诸多键值对会被一个花括号涵盖住，我们称花括号及里面的内容为一个对象，即一个对象中包含多个键值对，json的灵活之处就在于，虽然键值对的“键”（例如：“天气”、“成员”）必须为字符串，但是键值对的“值”（“晴天”“233元”）不仅可以是数值，字符串，甚至也可以是一个对象，子子孙孙无穷匮也，无论多么复杂的信息集合也可以合适地表现出来。