

【ICSE2016 论文介绍】使用 Rust 语言开发 Servo 浏览器引擎

2016-06-01 软件工程研究与实践

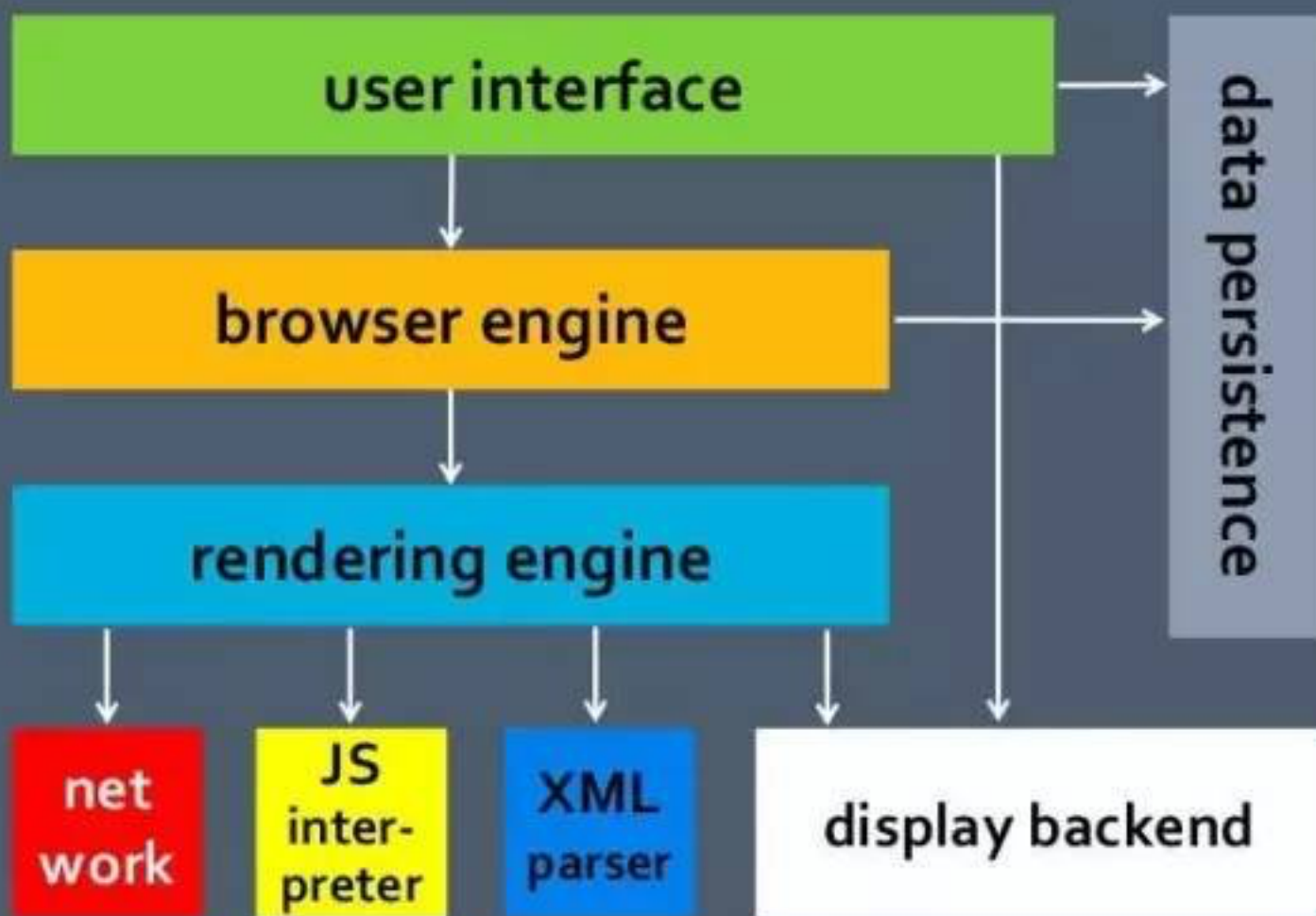
【研究背景与动机】

现代 Web 应用的规模和复杂度与日俱增，不亚于本地应用，这给 Web 浏览器平台带来了功能和性能上的挑战，也随之导致 Web 浏览器的代码日益复杂而庞大。Web 浏览器的核心称为『引擎』，如 IE 的 Trident/Spartan、Chrome、Safari 的 Blink/Webkit 以及 Firefox 的 Gecko 引擎，这些浏览器引擎均由百万行级别的 C++ 代码组成。

浏览器引擎的功能

浏览器引擎的这么多代码实现哪些功能呢？它们主要包括：

- HTML 和 CSS 分析 (Parsing)：一个 URL 标识要加载的 HTML 资源。浏览器引擎对 HTML 资源进行分析，将其转换成 DOM（Document Object Model）树。浏览器引擎需要能处理分析 HTML 资源时遇到的错误并进行错误恢复；需要能在遇到 script 元素时，暂停对 HTML 资源的分析直至执行完所包含的 JavaScript 脚本；需要能通过投机地执行记号流扫描和资源预取，来减少资源加载时的延迟等 [WLZC11]。
- 样式计算 (Styling)：在构造 DOM 之后，浏览器会利用 HTML 及其链接的 CSS 文件中的样式表来计算样式信息，构建 flow tree 等数据结构。Flow tree 是 Servo 中引入的名词，用来描述 DOM 元素在页面中的布局。
- 布局 (Layout)：将 flow tree 中的抽象元素具体化为实际的图形、文本等，构建 display list 数据结构。
- 渲染 (Rendering)：将 display list 中的元素一层层渲染在内存缓冲区（这里简称缓存）或者直接到图形装置等。
- 合成 (Compositing)：对缓存或图形装置等的各个 " 图层 " 进行变换、组合和优化，得到最终呈现给用户的内容。
- 脚本执行 (Scripting)：JavaScript 脚本可以在分析、布局、绘制、显示的各个阶段中被执行，这些脚本在执行时可能会修改 DOM 树，从而要求重新计算布局并绘制。浏览器期望最小化这种重新计算的开销。



www.html5rocks.com/en/tutorials/internals/howbrowserswork/

Servo项目的研发动机

Servo 项目的动机是什么呢？事实上，这些年来对主流浏览器引擎的开发和维护暴露了一些软件工程上的问题：

- 无法有效利用多处理器架构，尤其是在智能手机等移动设备上 [MTK+12, CFMO+13]。
- Gecko 中约 50% 的安全漏洞均为内存相关，如悬空指针、数组越界、整数溢出等。即便是经验丰富的 C++ 编程团队加上使用最好的静态分析工具也难以排除这些错误。
- 由于 Web 的交互性越来越强，这些引擎绝大多数采用串行处理的架构，导致难以在不损失交互性的情况下加入新的特性。
- 用 C++ 编写的引擎代码库维护成本高，并且培训编程新手的成本高。

【Rust语言】

Rust 语言（<https://www.rust-lang.org/>）是针对上述需求量身打造的系统级编程语言，其设计主要受 C 和 ML 语言族的启发，它允许开发人员细粒度地控制内存布局和预测性能。与 C 程序不同的是，Rust 程序缺省地是内存安全的，只允许不安全的操作出现在显式声明的 `unsafe` 块中。

Rust 提供静态的强类型系统和精确的内存管理。在默认情况下，可以防止以下内存问题出现：

- 悬空指针（Dangling pointer）

- 数据竞争（Data race）
- 整数溢出（Integer overflow）
- 缓存区溢出（Buffer overflow）
- 迭代器失效（Iterator invalidation）

具体来说，这些是如何做到的呢？ Rust 中有『所有权（Ownership）』（<http://doc.rust-lang.org/book/ownership.html>）和『借用（Borrowing）』（<http://doc.rust-lang.org/book/references-and-borrowing.html>）这两个基本概念。基于所有权的类型系统是一种 *affine type* 类型系统。 Rust 中的所有权模型受 Singularity OS [HLA+05] 的所有权模型、以及 Cyclone 语言 [GMJ+02] 和 MLKit [TB98] 中的基于区域 (region) 的内存管理系统的影响。

『所有权』与『借用』

所有权资源管理类似于 C++ 中的 RAI (resource acquisition is initialization) 设计。 Rust 中的每个值（或称资源）都只能被单个变量所独有，所有权本身可以被转移但不能被分享。对数据的引用（Reference）通过『借用』实现。借用默认是只读的（类似 C++ 的 `const` 引用），需要在类型签名上标记 `mut` 来提供可写引用。而且， Rust 的借用规则检查保证了可以有多个只读借用、但是最多只能有一个可写引用，从而解决了数据竞争的问题，同时产生了更多的编译优化空间。一个简单的例子如下所示：


```

// The single value we'll work with.
let val = Box::foo(0);

// Move the value in `val` to a new location
let moved_val = val;
// `val` is no longer accessible.
// The following will not compile.
// println(``{}``, val);

// Borrow and share the value immutably
// several times. This is done in a new
// block to limit the scope of the borrows.
{
    let shared1 = &moved_val;
    let shared2 = &moved_val;
    println!(``{}``, shared1);
    println!(``{}``, shared2);
    // Borrows are released at the end
    // of the block.
}

```

Figure 4: Ownership and borrowing.

其他的语言功能如 泛型（<http://doc.rust-lang.org/stable/book/generics.html>）、基于 trait 的多态（<http://doc.rust-lang.org/stable/book/traits.html>）、ADT（Algebraic Data Types）类型（<http://doc.rust-lang.org/stable/book/enums.html>）、模式匹配（<http://doc.rust-lang.org/stable/book/patterns.html>）、闭包（<http://doc.rust-lang.org/stable/book/closures.html>）等大大提高了语言的表达力、易用性、易读性。

并发支持

Rust 以库的形式提供各种 并发抽象（<https://doc.rust-lang.org/book/concurrency.html>），如线程、锁、消息队列、原子操作等均有完善的标准库支持。

trait 和 trait bound（即对多态加静态限制）的引入大大地提高了并发构造的可用性。如下代

码所示，只有实现了 Send marker trait 的数据结构才能在线程间传递，这样，实现了原子引用计数的 Arc 智能指针才能作为消息的一部分，而普通的 Rc 指针出现在消息中就会被编译器拒绝。

```
fn main() {
    let (tx, rx) = channel();

    // let shared = Rc::new(1); // Wrong
    // error: the trait `core::marker::Send`
    //       is not implemented for the
    //       type `alloc::rc::Rc<i32>`

    let shared = Arc::new(1); // Correct
    thread::spawn(move || {
        tx.send(*shared).unwrap();
    });
}
```

【 Rust 在 Servo 上的实践与验证 】

在性能上，初步的测试表明 Servo 在某些任务上比 Gecko 性能好很多（如表 1）：

Site	Gecko	Servo 1 thread	Servo 4 threads
Reddit	250	100	55
CNN	105	50	35

Table 1: Performance of Servo against Mozilla’s Gecko rendering engine on the layout portion of some common sites. Times are in milliseconds, where lower numbers are better.

下面重点讨论在实际使用 Rust 开发 Servo 的过程中所得到的观察。

Rust的语法

Rust 提供的现代语言构造（结构类型和基于 Algebraic Data Types 的枚举类型以及模式匹配等）更精炼、有效地表达了原来在 C++ 需要为很多类创建头文件和对应的实现文件才能表达的抽象。而且，基于模式匹配的静态分发效率显著高于 C++ 的虚函数调用。Rust 没有像 Cyclone [GMJ+02] 那样，为了代码可移植性而刻意和 C++ 等保持语法一致，避免了复杂化新语言本身，使其设计更为简洁一致。

编译策略

Rust 的泛型类似于 C++ 的模板，会在编译时根据具体类型特化泛型函数，即产生多个类似的代码。这种方法虽然会使编译产生的代码规模变大，但是可以避免在运行时根据类型进行动态分发的开销，从而能提高程序的运行效率。

Rust 的编译单元的规模相对较大，且受全程序编译的限制。其编译单元称为 *crate*，它可以由数百个提供命名空间和抽象的模块（*module*）组成。一个 *crate* 中模块之间的依赖关系可以有环。

这种大规模的编译单元会使编译变慢，也会削弱并行编译代码的能力。但是，这种做法能使程序员编写 Rust 代码的速度相当于其编写串行 C++ 代码的速度，而不要求 Servo 的开发者都成为编译器的专家。目前，Servo 本身超过两百个 *crates*，涉及上千个模块。

内存安全

Servo 从开发以来从未遇到过悬空指针问题。不过，目前 Servo 还包含不安全的部分。第一是 Servo 跨语言使用了 C++ 编写的 JavaScript 引擎 SpiderMonkey。尽管利用 Rust 的类型系统已把 Rust 和 JavaScript 引擎之间的接口做了很多固化，但是依然有误用 JavaScript 接口的可能；此外，Rust 的静态安全保障对 C++ 代码部分并不起任何作用。另外一个不安全的部分是由提供多种遍历的 API 的一些高性能数据结构的实现而引起的，比如双向链表包含指向前一元素的后向指针来辅助反方向的遍历。这些结构不满足 Rust 假设 " 每个元素只有一个所有者 " 的所有权模型。

语言互操作性

得益于 Rust 对 C FFI (Foreign Function Interface) 的优秀支持，Servo 可以直接使用大量的第三方库。比如，前面提到的 JavaScript 引擎，还有图像渲染、多媒体解码等。另一方面，Rust 代码还可以被 C 程序调用，比如目前正在将 Firefox 浏览器中 URL parser 等组件换成 Rust 版本。

不过，目前在语言互操作上还有两方面的局限性。首先，Rust 不支持变长的参数，即 *vargs*；其次，Rust 没有办法直接使用 C++ 接口，必须通过一层 C 代码来包装，这会造成性能损失。

基础库

Servo 在 Rust 通用标准库上定制了一些专用库，比如进程间通信 *ipc-channel*。

模块化

Servo 有效利用了 Rust 的模块系统和 Cargo 包管理器，不仅仅将浏览器按功能模块化为几个独立的组件，而且每个组件都依赖一些为特定目的（比如 HTML parser）开发的相对独立的库。相比以往大型工程在自己的代码仓库里维护所有库，这种所谓 *polyrepo* 的设计有效地提高了代码的可重用性，降低了维护成本；同时， Servo 项目以外的开发者如果用到这些小型库，其反馈也可以促进 Servo 项目本身。

宏

Rust 支持声明式、基于模式匹配 [KW87] 的 hygienic 宏系统。 Servo 自己就定义了超过一百个宏，这些宏的使用在某些地方近似于 DSL (Domain Specific Language)，如在 HTML tokenizer 中大量使用宏使得 tokenizer 代码简洁易懂。

Rust 还支持更为强大的编译插件功能，可以直接在用户代码中操纵语法树，从而实现在编译期间构造哈希表、自动生成 GC 追踪代码、代码检查（ linter ）等。

新贡献者

值得一提的是， Servo 平均每个星期都会有五个新的贡献者。这说明了 Servo 相比于同类的大型系统级软件项目更容易上手、维护。

【未来的工作和相关的项目】

开放的问题

目前还有很多开放性问题没有得到答案：

1. JIT 代码的正确性： JavaScript 引擎为了提高效率会使用 JIT (Just-in-time) 动态编译策略，如何保证 JIT 优化后的代码相对于环境和垃圾收集器的正确性、有效性，是一个有待解决的问题。
2. 不安全部分的检查： 同样为了效率， Rust 代码中仍有标记为不安全（ unsafe ）的代码。不过，虽然标准 Rust 编译器可能无法推导出这部分代码的正确性，但是用一些其他工具和标记来确保一些基本的安全性（ property ）也是很有用的。
3. 增量计算（ Incremental computation ）： 由于 JavaScript 的存在，网页的内容是动态的。如何减少冗余、重复计算而只进行影响最终结果的那部分计算、渲染，这是对浏览器性能有很大影响的问题。

相关的浏览器研究工作

1. 高通研究院的 ZOOMM 浏览器项目 [CFMO+13]： 在具有多核处理器的移动设备上提高并行度，从而提高可交互性（ interactivity ）。
2. 加州大学伯克利分校 Ras Rodik 组的并行浏览器项目： 提高布局（ layout ）计算的并行程度 [MB10]。

【结论】

总体来说， Servo 团队在较短时间、少量专职开发者供职的条件下开发出这样一个高性能、产品级的浏览器引擎，充分证明了 Rust 语言和相关软件工程理念的有效性。尤其对于偏系统、偏底层，对内存和正确性有比较严苛要求的应用场景， Rust 语言是一个非常好的选择。

【作者简介】

论文作者为 Mozilla Research 的 Brian Anderson 等人。

本导读的作者为中国科学技术大学计算机学院 2013 级本科生张震以及指导老师张昱副教授。目前张震同学正在 Google GSoC 项目赞助下参与 Servo 浏览器引擎的开发。

【参考文献】

[CFMO+13] Cascaval, C., S. Fowler, P. Montesinos-Ortego, W. Piekarski, M. Reshadi, B. Robatmili, M. Weber, and V. Bhavsar. ZOOMM: A parallel web browser engine for multicore mobile devices. In Proceedings of the 18th ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming, PPOPP '13, Shenzhen, China, 2013. ACM, pp. 271–280

[HLA+05] Hunt, G., J. R. Larus, M. Abadi, M. Aiken, P. Barham, M. Fahndrich, C. Hawblitzel, O. Hodson, S. Levi, N. Murphy, B. Steensgaard, D. Tarditi, T. Wobber, and B. D. Zill. An Overview of the Singularity Project. Technical Report MSR-TR-2005-135, Microsoft Research, October 2005.

[GMJ+02] Grossman, D., G. Morrisett, T. Jim, M. Hicks, Y. Wang, and J. Cheney. Region-based memory management in Cyclone. In Proceedings of the ACM SIGPLAN 2002 Conference on Programming Language Design and Implementation, PLDI '02, Berlin, Germany, 2002. ACM, pp. 282–293.

[KW87] Kohlbecker, E. E. and M. Wand. Macro-by-example: Deriving syntactic transformations from their specifications. In Proceedings of the 14th ACM

更多精彩文章

刘德科：调控一整年 还有哪些城市是「漏网之鱼」？

机电汕头的新生们，这里有一封你的信件，请你签收！

【硝烟铸英雄·战火显神威】鑫鸿维实业 | 鸿维建材公司四月启动大会

卓正内控-图说内控五要素

【平安十一刊】北戴河沙滩||作者：闲云野鹤（天津）

【用益资管日报】厉害了！这家信托要把手中的股票陕天然气全部卖完，马上获利8000万

【风采志愿者】师生携手志愿暖冬 幸福福泉共建文明

当场身亡！朔州接送孩子的都看看吧！

女子酒醉错上隔壁小王床.....

刁吉祥：治百病的民间土方（五）

【瑞隆商贸】朗丽椰子卷

周鸿祎学坏了，360间隔两周再发新机，其中套路让雷军都不好招架

因为这个部位没长好，杨幂怎么整都不如刘亦菲？

我爸妈又不和我们住，凭什么买房也要出20万？

路噪诊断指南：3个指南助你甄别行车途中爱车的噪声问题

宁夏检察机关依法对张湧决定逮捕 依法对吴先志、路彦国涉嫌受贿案提起公诉

深龙社工英才资助金要发放啦，请注意查收！

北京一自建房发生火灾致5死8伤，电动车火灾预防知识你应该知道！

第812期【正道教研】池永贵：建构主义教学理论在作文教学中的运用研究（二）

情人之间，不是因为钱、也不是因为美、而是.....

社保卡好消息，不看亏大了

阳光海风白沙滩，这个假期你计划去美丽的海岛了吗？沙滩是谁造的？知道真相的我眼泪掉

下来.....

南阳民警客车上救下了130多只苍鹭和画眉 | 男子为给儿子娶媳妇不惜代价，却说无钱养活老母 | “唱响白河”宛城区专场今晚邀你欣赏

野云说钱：中国花钱谱（7）

会有一个你，温暖我的生命

你老公连射九日那么猛

二女共侍一夫，睡了6年后竟....真不害臊！真人真事！！

从说话方式，轻松看穿一个人

不要总拿佛法的镜子照别人

酒吧勾女，保你夜夜打炮！

中药香囊免费领——榆林中医药文化节第一波福利！

【店庆遇见12.12】疯狂再续！不容错过！

人，永远是相互的

【温馨回顾】硅烷浸渍及其在混凝土结构中的应用

副省长郭生练：打造永不落幕的《黄四姐》

关于市场上假借我公司名义进行产品招商、宣传的声明

【戏曲推荐】河南曲剧电影《卷席筒》全集视频

如果中国没有他们.....

水墨清华公益联盟五周岁 联欢

中国人牙齿健康率不足1% 口腔癌发病率激增

安全管理与“多米诺骨牌效应”

教师节，开个吐槽大会！一起来扒一扒那些年老师说过的“谎言”

央视曝光！全是假货，没有一件真品！广州人竟还在疯狂买！

景色|周末没有好去处？那就看这里。

张克 | 被締途虏获的“亚军”卡友中的“意见领袖”

知产榜样 | “鱼水情深”，劳模精神的时代诠释

加多宝：移动互联网+，作业本事件，商人思维 | 批判

【行业分析】特色小镇如何规范？要“三生融合”、“三位一体”

重磅推出！我们玉田的颜值因为这首诗美爆了

初中历史考试解题技巧大全

「成片」《稚心·巧手》

高血压肾损害人预防加重病情的方法

未来出行新选择 衣宝廉院士带你认识氢燃料电池汽车

邱老师分享：治腰疼、颈椎疼、膝盖疼，一病一绝招！太珍贵了~

乐游直通车10月12日星期四

2017-11-18推荐 瑞士花园 精装 满2年 带车位

昨日市场|美股三大股指全线收涨 科技股普遍上扬

呱呱的干冷天气今晚驾到！低至10℃！文末有惊喜哟！！！

【卡米熊】不是顾客难满足， 是你套路不够深！三分钟学会母婴店十四式经典套路

「型车魅影」提醒！你的手机壁纸要换了！ 1112期

吃得山珍味，不识人间烟

【正一瑜伽】瑜伽疗愈必修课

优考案例|关注过程 强化沟通 一分之超“踩”重大！

中级执法资格考试模拟题（三）参考答案

【泰州德立】用心陪伴——大众进口汽车“双十二”主题活动日，多重好礼享不停！

中大“新”声音 | 医学院：传承中大医科优良传统，建设基础临床融合的高水平学院

中润易达物流求贤榜

青州市档案馆、市民活动中心开工啦！将成青州新地标！

明代陆粲小楷《浒溪草堂记》，秀劲流美！

盘点民国时的奇葩书名翻译，莎士比亚听了都想打人

精华！高中政历地知识结构大梳理！让你轻松多考40分！

【十九大精神进江陵“微解读”】新时代全面依法治国的基本遵循

给孩子有温度的玩具 一针一线都是爱

睡梦珂专卖店、新款软床系列、床垫系列

潮流生活 | 全球首批无人驾驶公交深圳上路了！你敢坐吗？

有钱，来抢，20元！

【埋在土里的神奇“雪莲果”】挖出来就可以吃！水嫩多汁！煲汤更养人，降血脂调肠胃！看

上一眼，就想吃！29.9元9斤，现挖现发！

世纪之最！一艘集装箱船上藏了近6吨毒品，价值2.1亿欧元！【附视频】

金庸笔下，只羡鸳鸯不羡仙的奇女子

市委“两学一做”学习教育督导组莅临谢家集区孙庙乡检查指导工作

美女内急找不到女厕,结果..

【今日考讯 |2017广东惠州市财政局选调公务员2人公告】新仕途教育小军师公务员遴选考试

资讯推送

铜川一幼“分享好家风·传递正能量”

【揭秘】眼科医院配镜真的贵吗？为你揭秘不为人知的“潜规则”

这个暑假，给你家宝贝一个“上天”的机会！

【2017重点工作督查】全县重点工作督查组现场督查城川等6乡镇重点工作完成情况

他说他要吊打杨开...

2017兰州新区管委会办公室面向全国公开选调5名干部公告

威职荣誉榜 | 历数2017年全国职业院校技能大赛上，威职学子获过的奖

子若强于我，要钱做什么；子若不如我，留钱做什么（父母必读）

历届“湖洋经济文化促进会”珍贵照片

测量量子有啥用途

优质作品版权是数字文创产业的稀缺资源

奶粉配方注册制实施细则出炉 洋奶粉加上紧箍咒

【股市】巴疯特对2017.03.03的预测

【财务核算管理系统小知识】无法给一个批中新增凭证

阿佳组合《爱是一束光》MV欣赏

瓮安准备要回老家修房子的安逸啦，3o万的定制别墅都要出来啦！

关注你的速度

屋顶竟被设计成空中泳池，极简纯粹的别墅美到令人窒息！

关注 | 好客青城人在福州推介壮美内蒙古

关于 ICSE2016 论文 介绍 使用 语言 开发 浏览器 引擎 的文章 - 360doc