


专业实验I（计算机实习）

A pencil and a ruler are positioned diagonally across the right side of the slide. The pencil is sharpened and its tip is near the ruler. The ruler has markings for centimeters, with the numbers 9 and 10 clearly visible.

授课老师：万能

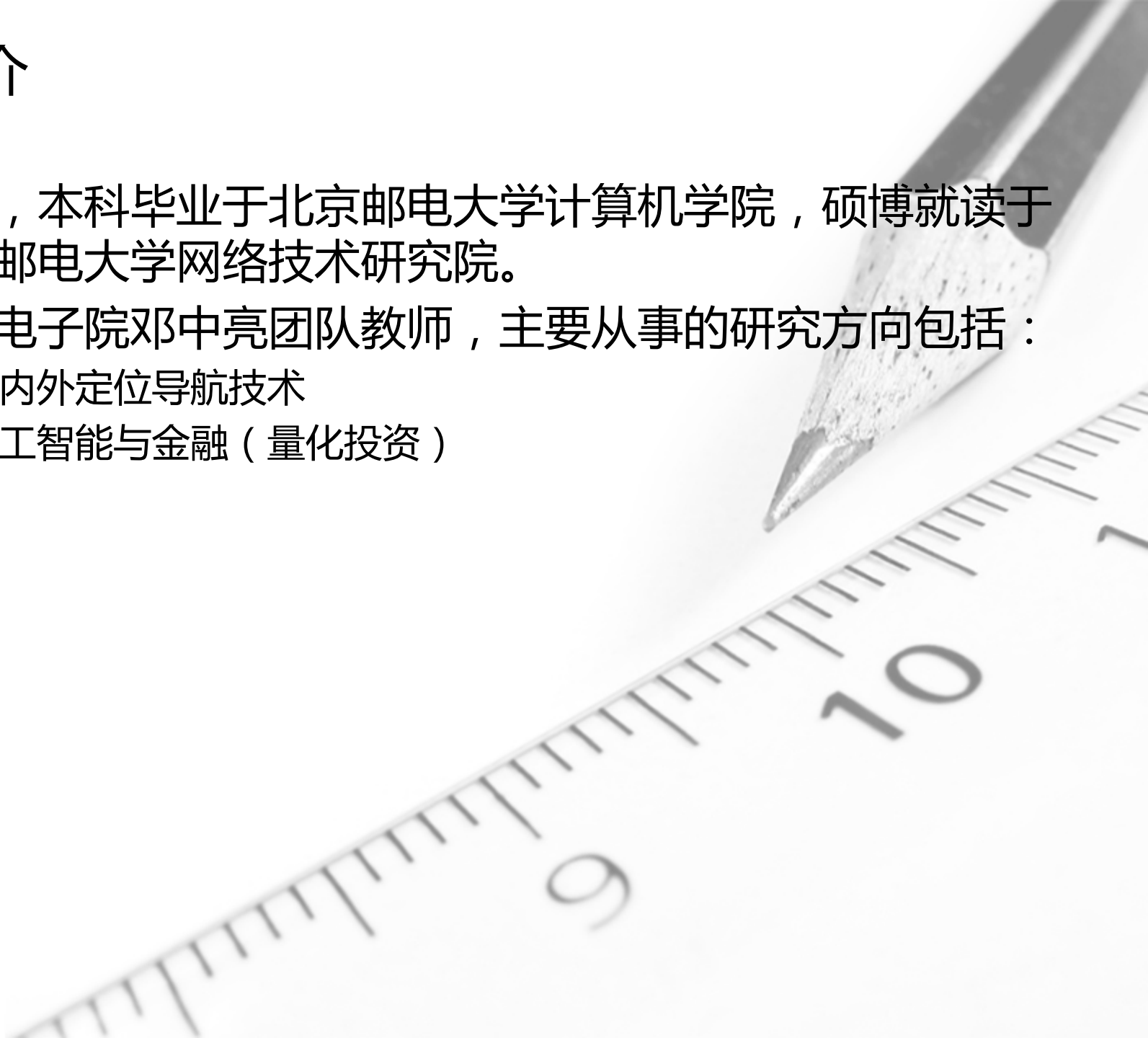
邮箱：buptwn@qq.com（实验报告）

电话/微信：13810020981



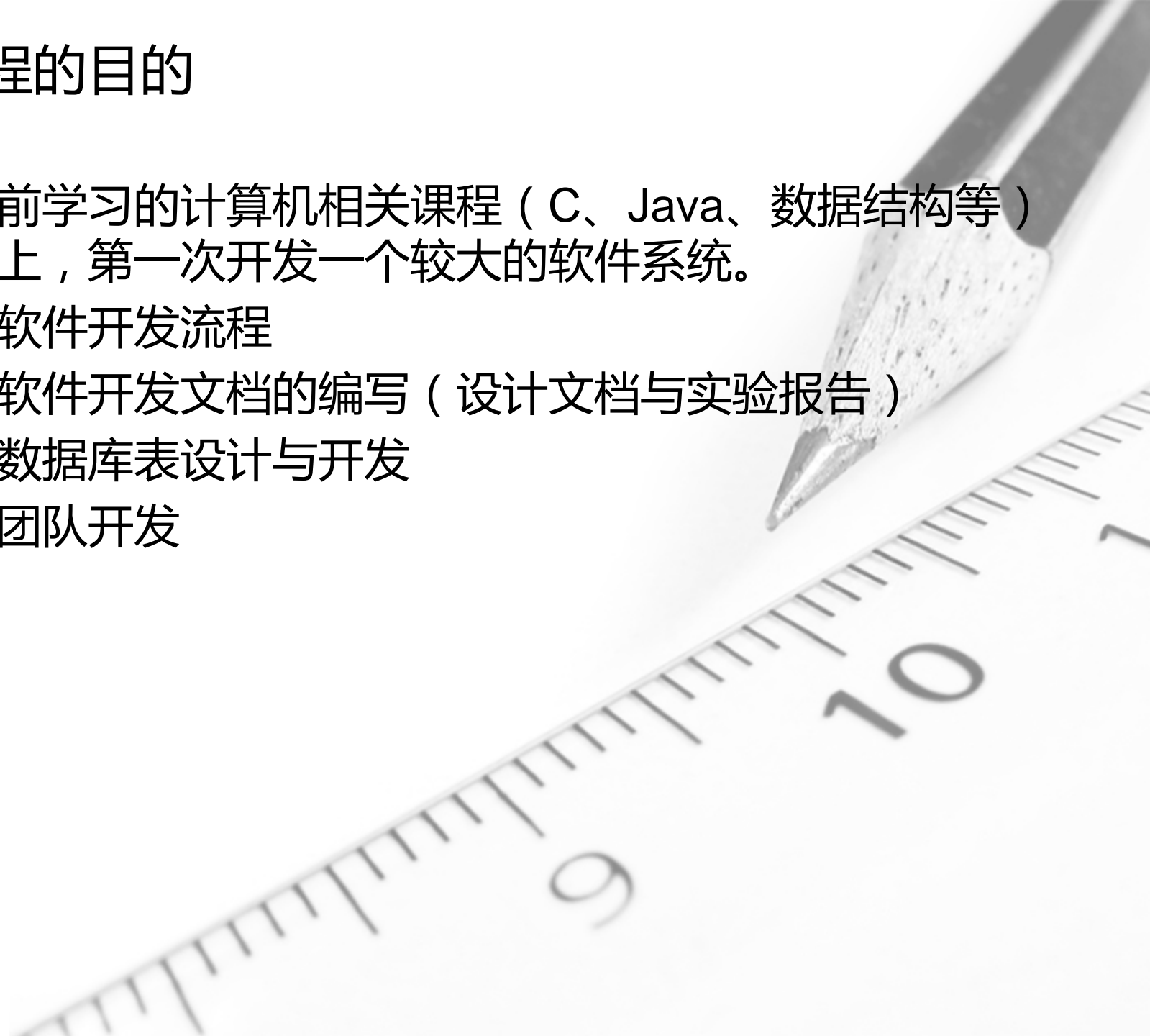
老师简介

- 万能，本科毕业于北京邮电大学计算机学院，硕博就读于北京邮电大学网络技术研究院。
- 现为电子院邓中亮团队教师，主要从事的研究方向包括：
 - 室内外定位导航技术
 - 人工智能与金融（量化投资）



本门课程的目的

- 在之前学习的计算机相关课程（C、Java、数据结构等）基础上，第一次开发一个较大的软件系统。
- 学会软件开发流程
- 学会软件开发文档的编写（设计文档与实验报告）
- 学会数据库表设计与开发
- 学习团队开发



必选实验

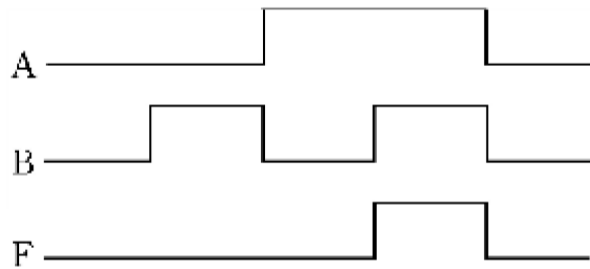
- 数字波形生成

- 给出数字序列A、B的值和逻辑函数F的表达式（与、或、非、异或、同或等），生成A、B和F的波形图。
注：A、B序列长度最大值不小于16。

- 例：A：0 0 1 1

- B：0 1 0 1

- F = A & B 可生成如下波形



可选实验一：多个小题（全做）

- 见WORD文档



可选实验二：基于爬虫的数据分析系统

- 人员数量：2个人协作完成
 - 1位同学负责爬虫编写，并将数据存入数据库
 - 1位同学负责从数据库读取数据进行可视化分析
- 爬虫要求
 - 开发一个爬虫，每天自动爬取东方财富网的数据
 - 将爬取到的数据存入本地数据库
- 数据源
 - 股东户数 爬取所有股票的融资融券金额及其历史数据：
<http://data.eastmoney.com/rzrq/>
- 爬取技巧
 - 爬取某个股票每天的融资融券余额和每天收盘股价（半年内数据），修改后头的股票代码即可
 - 融资融券金额: <http://data.eastmoney.com/rzrq/stock/600816.html>
 - 注意：只有部分股票支持融资融券

实验二：基于爬虫的数据分析系统

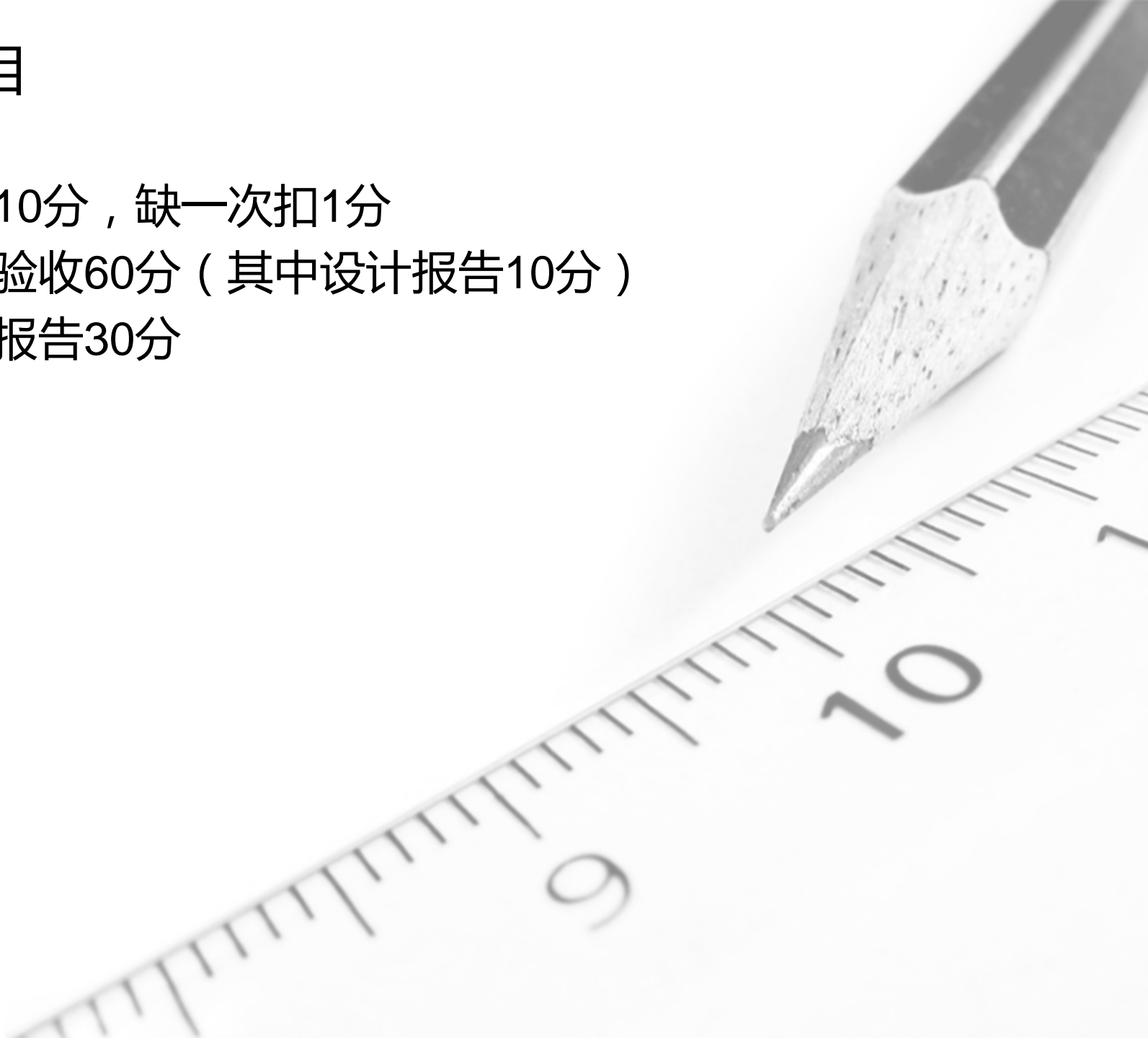
- 数据可视化分析（融资数据）
 - 画出半年来的融资、股价双曲线图
 - 查询功能：根据根据下列条件进行组合查询：
 - 时间（过去1天、5天、10天、自选时间段）
 - 融资增减比例（增加、减少、增加/减少10%以上，增加/减少20%以上、增加/减少30%以上）
 - 股价涨跌幅（上涨、下跌、上涨30%以上、上涨20%以上、上涨10%以上、下跌30%以上、下跌20%以上、下跌10%以上）
 - 点击个股，在同一个图上显示 股价与个股波动曲线。在曲线上标明波动幅度。

时间计划

- 本课程共12周，每周2次课，每次课都必须签到，考勤分共10分，缺一次扣一分，不得代签。
- 第1-2周，根据题目要求，查找资料并学习，完成课题方案设计，第3周上课时提交设计报告一份。设计报告占10分。设计报告要求：
 - 拟采用的开发平台、开发语言、系统架构等
 - 数据库表设计
- 第3周，上台阐述设计思路
- 第3-11周，编程并调试
- 第2-12周，根据老师的测试用例进行测试验收，有问题则修改问题，完成实验报告并提交。

评分科目

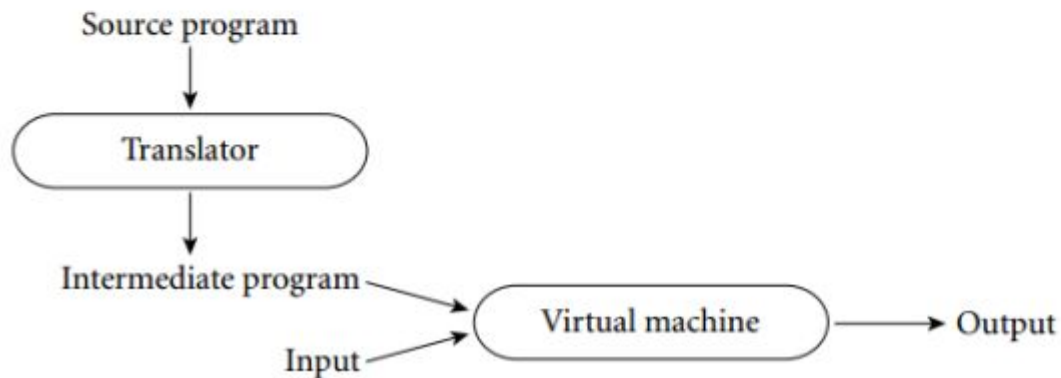
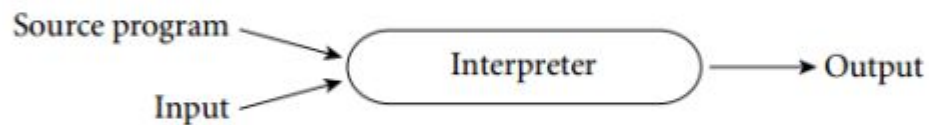
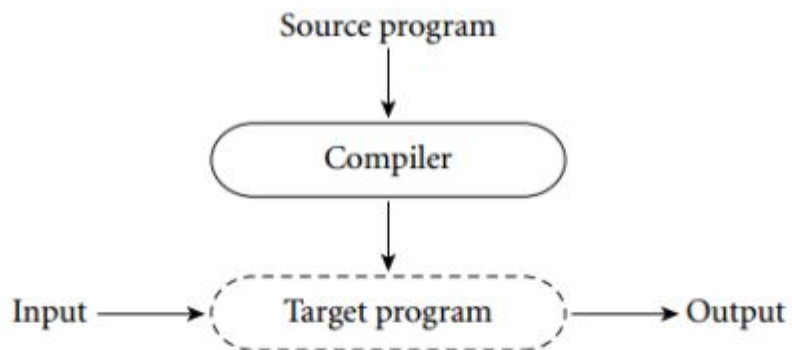
- 考勤10分，缺一次扣1分
- 项目验收60分（其中设计报告10分）
- 实验报告30分



编程语言选什么？

- **低级语言（机器语言、汇编语言）、高级语言**
- **编译型语言**：程序执行之前，需要一个专门的编译过程，把程序编译成为机器语言的文件，比如exe文件，以后要运行的话就不用重新翻译了，直接使用编译的结果就行了（exe文件），因为翻译只做了一次，运行时不需要翻译，所以编译型语言的程序执行效率高，需要内存少。缺点是依赖编译器，跨平台性（移植性）差，系统安全性也较差。开发调试复杂。
 - C/C++、Pascal、Swift
- **解释型语言**：程序不需要编译，程序在运行时才翻译成机器语言，每执行一次都要翻译一次。因此效率比较低。但移植性好。
 - Python、R、C#、Ruby、Perl、JavaScript、PHP、ASP
- **Java**：Java程序需要进行编译，但并不会生成特定平台的二进制机器码，它编译后生成的是一种与平台无关的字节码文件（.class,移植性好），这种字节码自然不能被平台直接执行，运行时需要由解释器（JVM, JAVA虚拟机）解释成相应平台的二进制机器码文件。现在也有直接编译成机器码的编译器了。
- 对效率要求高的系统，比如开发操作系统、大型应用程序、数据库系统等时都采用编译型语言。由于松散的安全性和平台依赖性，编译型语言不太适合开发基于Web的应用。

编程语言选什么？



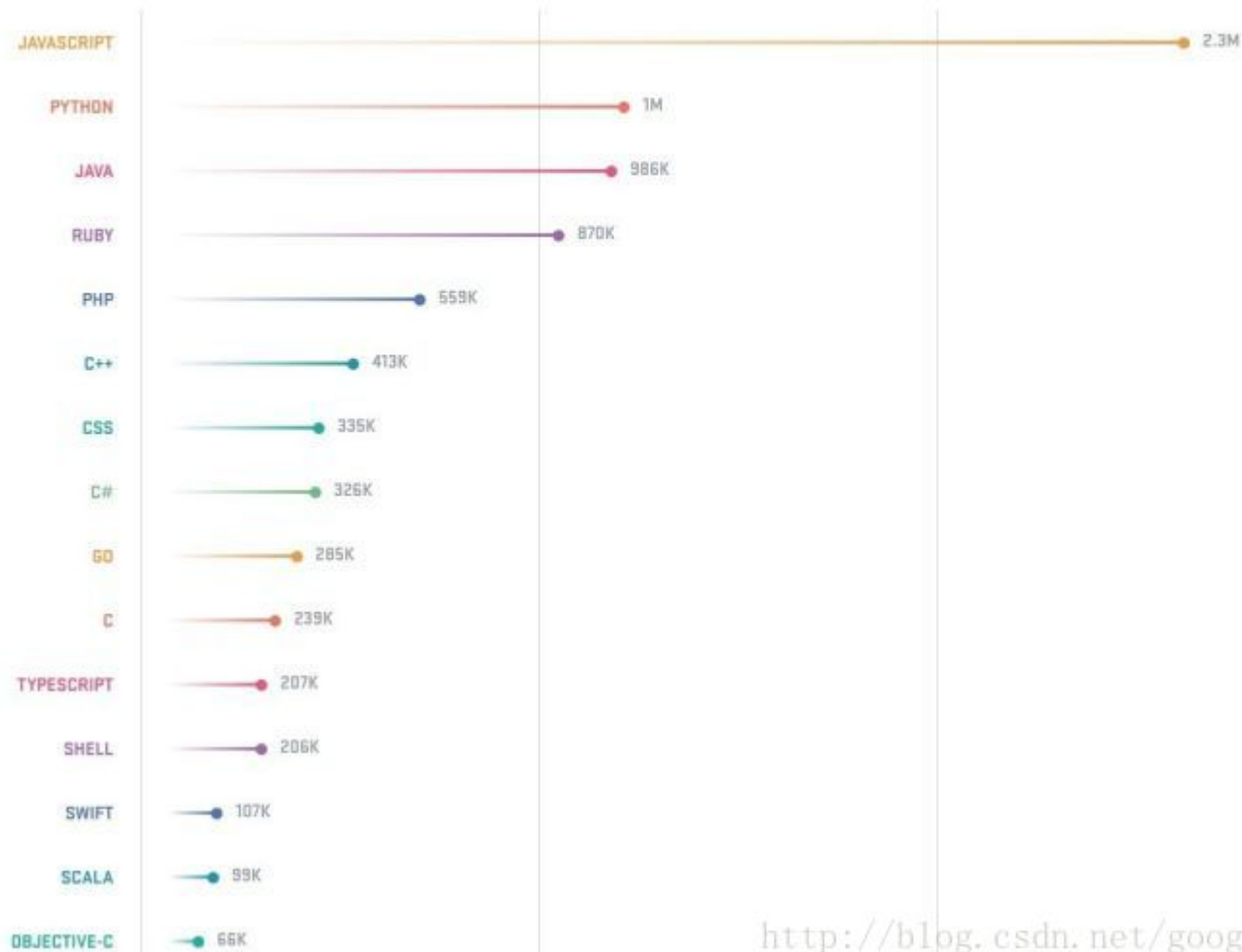


编程语言选什么？

The fifteen most popular languages on GitHub

by opened pull request

GitHub is home to open source projects written in 337 unique programming languages—but especially JavaScript.





GitHub

- 上图是根据2017年全年提交的 PR（ Pull Request ）数量来排名的。
- GitHub的流程
 - Fork（ 远端复制别人的仓库 ）
 - Clone（ 本地做一个分支，修改其中bug，或者补充功能 ）
 - 原作者如果觉PR给原仓库，让原作者看到你的修改
 - 得你改得好，他就会merge到项目中去
- 从0开始学习GitHub电子书：
 - <https://pan.baidu.com/s/1gfMQrtx> 密码: vvht



IEEE 2017年编程语言排名

Language Rank	Types	Spectrum Ranking
1. Python	 	100.0
2. C	  	99.7
3. Java	  	99.4
4. C++	  	97.2
5. C#	  	88.6
6. R		88.1
7. JavaScript	 	85.5
8. PHP		81.4
9. Go	 	76.1
10. Swift	 	75.3
11. Arduino		73.0
12. Ruby	 	72.4
13. Assembly		72.1
14. Scala	 	68.3
15. Matlab		68.0

Python

- 一种面向对象的解释型语言。Python具有丰富和强大的库。它常被昵称为胶水语言，能够把用其他语言制作的各种模块（尤其是C/C++）很轻松地联结在一起。
 - 如使用Python快速生成程序的原型，然后对其中有特别要求的部分，用更合适的语言改写，比如3D游戏中的图形渲染模块，性能要求特别高，就可以用C/C++重写，而后封装为Python可以调用的扩展类库。
- Python 有2和3两个版本，2将在2020年1月停止服务
- 数据分析/人工智能常用编程配置：Anaconda+Pycharm
 - Anaconda是一个开源的Python发行版本，其包含了conda（一个开源的软件包管理系统和环境管理系统）、Python等180多个科学包及其依赖项。
 - Pycharm是目前最流行的Python编程环境
 - 注册北邮edu.cn信箱，可以申请Pycharm专业版授权
 - 安装流程：
<https://blog.csdn.net/yggaoeecs/article/details/78378938>

数据库选什么？

The DB-Engines Ranking ranks database management systems according to their popularity. The ranking is updated monthly.

342 systems in ranking, May 2018

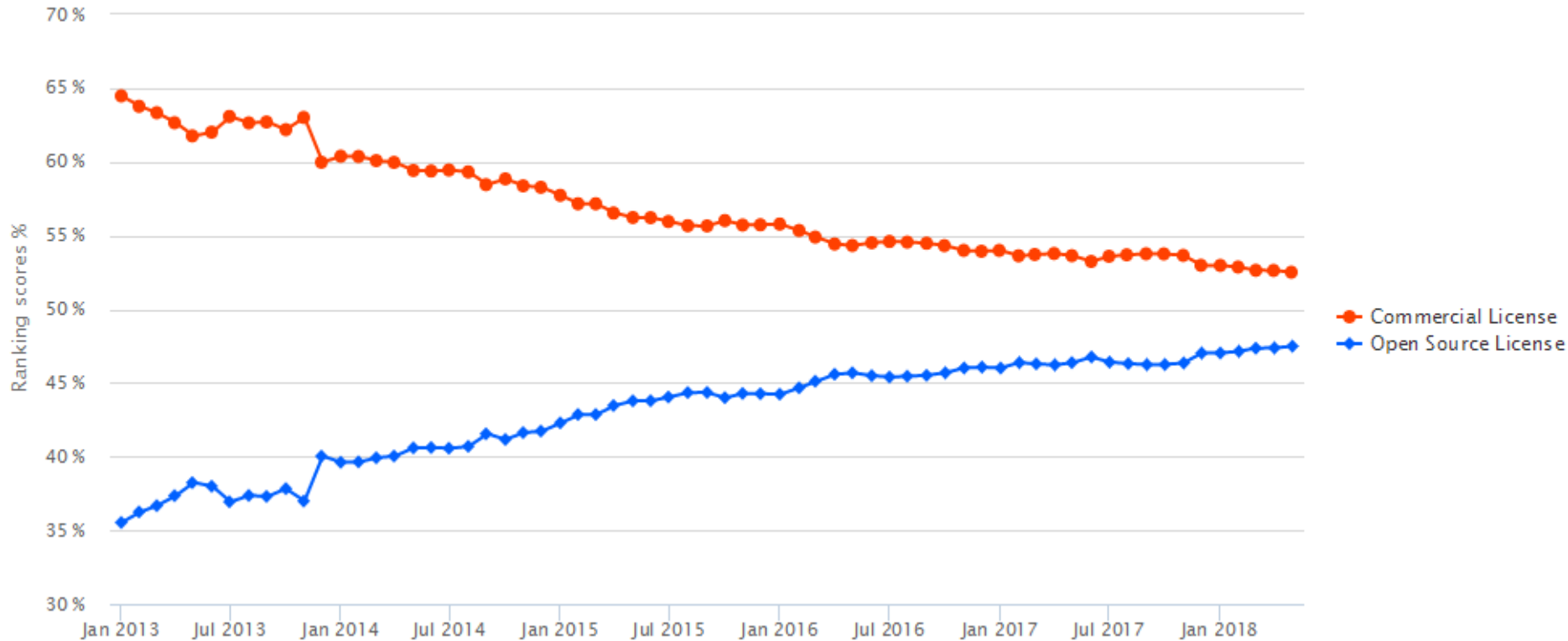
Rank			DBMS	Database Model	Score		
May 2018	Apr 2018	May 2017			May 2018	Apr 2018	May 2017
1.	1.	1.	Oracle +	Relational DBMS	1290.42	+0.63	-63.90
2.	2.	2.	MySQL +	Relational DBMS	1223.34	-3.06	-116.69
3.	3.	3.	Microsoft SQL Server +	Relational DBMS	1085.84	-9.67	-127.96
4.	4.	4.	PostgreSQL +	Relational DBMS	400.90	+5.43	+34.99
5.	5.	5.	MongoDB +	Document store	342.11	+0.70	+10.53
6.	6.	6.	DB2 +	Relational DBMS	185.61	-3.34	-3.23
7.	↑ 9.	↑ 9.	Redis +	Key-value store	135.35	+5.24	+17.90
8.	↓ 7.	↓ 7.	Microsoft Access	Relational DBMS	133.11	+0.89	+3.24
9.	↓ 8.	↑ 11.	Elasticsearch +	Search engine	130.44	-0.92	+21.62
10.	10.	↓ 8.	Cassandra +	Wide column store	117.83	-1.26	-5.28

商用数据库vs开源数据库

60

800

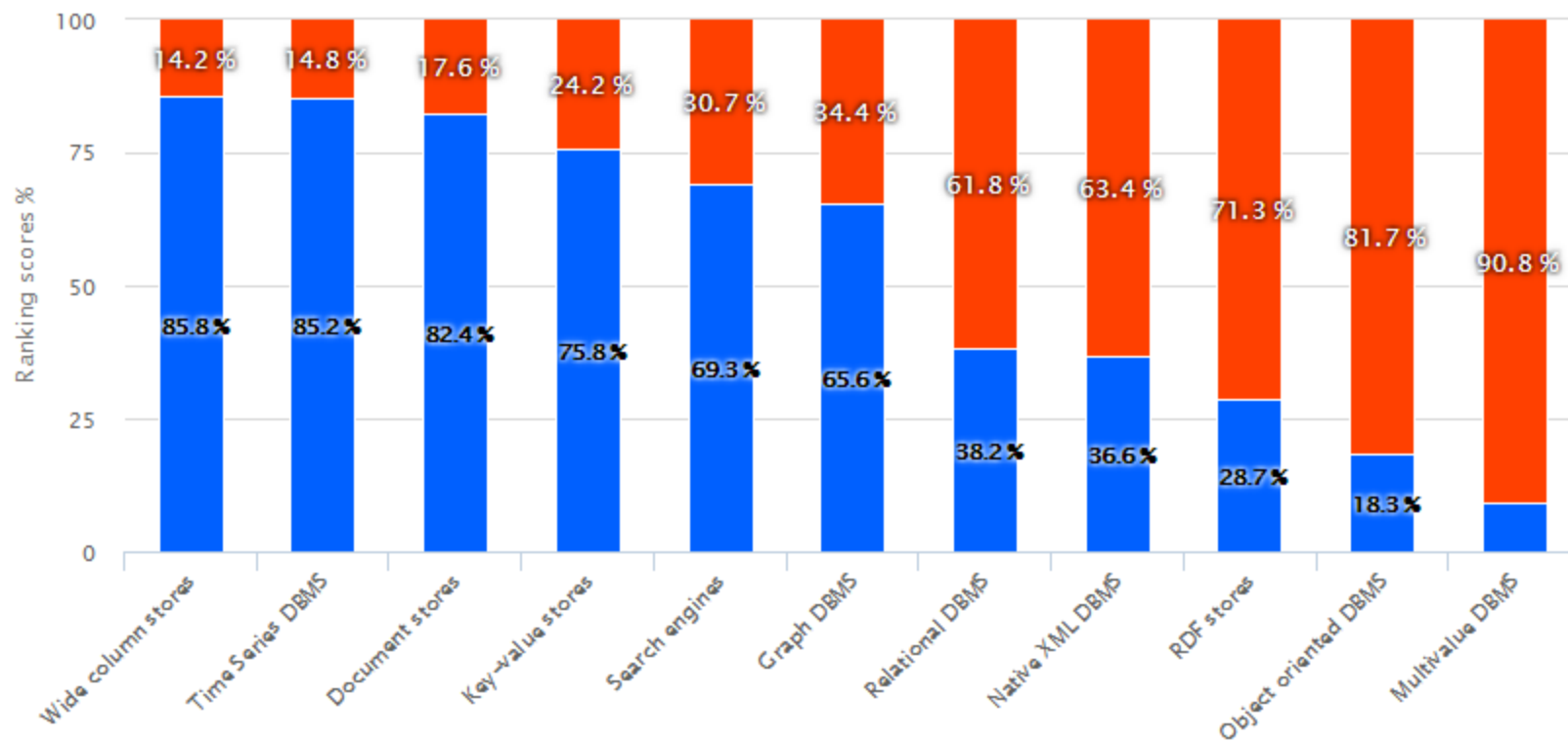
Popularity trend



© 2018, DB-Engines.com

商用数据库vs开源数据库

Popularity broken down by database model, May 2018



商用数据库vs开源数据库

The top 5 commercial systems, May 2018

Rank	System	Score	Overall Rank
1.	Oracle	1290	1.
2.	Microsoft SQL Server	1086	3.
3.	DB2	186	6.
4.	Microsoft Access	133	8.
5.	Teradata	74	12.

The top 5 open source systems, May 2018

Rank	System	Score	Overall Rank
1.	MySQL	1223	2.
2.	PostgreSQL	401	4.
3.	MongoDB	342	5.
4.	Redis	135	7.
5.	Elasticsearch	130	9.

三大数据库对比

- MySQL：多平台。免费（某些应用收费）。容易学习。性能很好，但扩展性查。
- SQL Server：只支持windows，收费，对.net支持很好。容易学习。用户连接过多时，性能变差。无开放性，安全性较差。
- Oracle：所有主流平台上运行。采用完全开放策略。价格昂贵，建议适用于超大型项目，小项目没必要用它。数据库容量可以无限大。较难学习。安全级别高。
- 在数据量少时，连接速度MySQL>SQL Server>Oracle
- 千万级数据时，连接速度SQL Server>Oracle>MySQL
- 当达到亿级时，Oracle速度最快。
- 并发性能：Oracle最好，MySQL次之，尤其是并发数量超过1000以后，Oracle并发性能优势明显。

