

正则测验1：

1. 校验密码强度
- 密码的强度必须是包含大小写字母和数字的组合， 不能使用特殊字符， 长度在8–10之间。
2. 校验中文
- 字符串仅能是中文。
3. 由数字、26个英文字母或下划线组成的字符串
4. 校验E-Mail 地址

同密码一样，下面是E-mail地址合规性的正则检查语句。

5. 校验身份证号码
6. 校验日期
- “yyyy-mm-dd“ 格式的日期校验， 考虑平闰年。
7. 校验金额

金额校验，精确到2位小数。

8. 校验手机号
- 国内 13、15、18开头的手机号正则表达式。（可根据目前国内收集号扩展前两位开头号码）
9. 提取URL链接
- 下面的这个表达式可以筛选出一段文本中的URL。
10. 文件路径及扩展名校验
- 验证windows下文件路径和扩展名（egs:.txt文件）
11. 文本清洗
- 清洗文本中的标点符号

正则测验2：

文本归一化，将阿拉伯数字转汉字

分类	原文	归一化
计量单位	这块黄金重达324.75克	这块黄金重达三百二十四点七五克
时间	她出生于86年8月18日，她弟弟出生于1995年3月1日	她出生于八六年八月十八日 她弟弟出生于一九九五年三月一日
编号	电影中梁朝伟扮演的陈永仁的编号27149	电影中梁朝伟扮演的陈永仁的编号二七一四九
分数	现场有7/12的观众投出了赞成票	现场有十二分之七的观众投出了赞成票
金额	随便来几个价格12块5， 34.5元， 20.1万	随便来几个价格十二块五 三十四点五元 二十点一万
百分比	明天有62%的概率降雨	明天有百分之六十二的概率降雨
电话	这是固话0421-33441122 这是手机+86 18544139121	这是固话零四二一三三四四一一二二 这是手机八六一八五四四一三九一二一