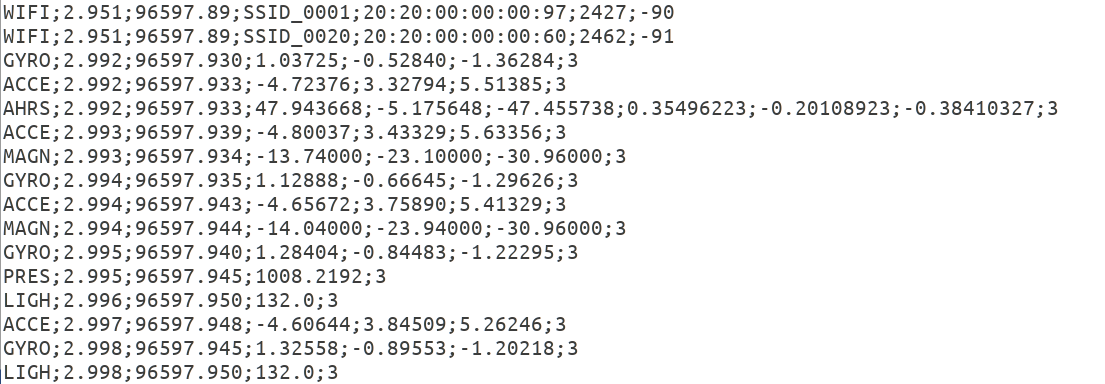
# ATR实验室研一软件培训结业测试题

现有一Android手机采集的传感器数据集，数据集以文本格式存储。现需要你编写一个C/C++代码对其进行预处理，以提取数据。

1. **输入数据集格式说明：**

该文本文件由多行组成，文件包括注释和实际数据。以“%”开头行为注释。非“%”开始的行为有效数据。每一行有效数据由多个由“;”分隔的域组成。第一个域是4个字母的数据类型（它包括ACCE、GYRO、MAGN、PRES、LIGH、AHRS、GNSS、WIFI、SOUN），第二、三个域是时间，再接着长度不定的实际数据。

数据示例：



ACCE、GYRO、数据采集频率均为200HZ，MAGN、AHRS为100HZ，PRES、LIGH约为5HZ，GNSS约为1HZ，WIFI约为0.25HZ。

1. **输出：**

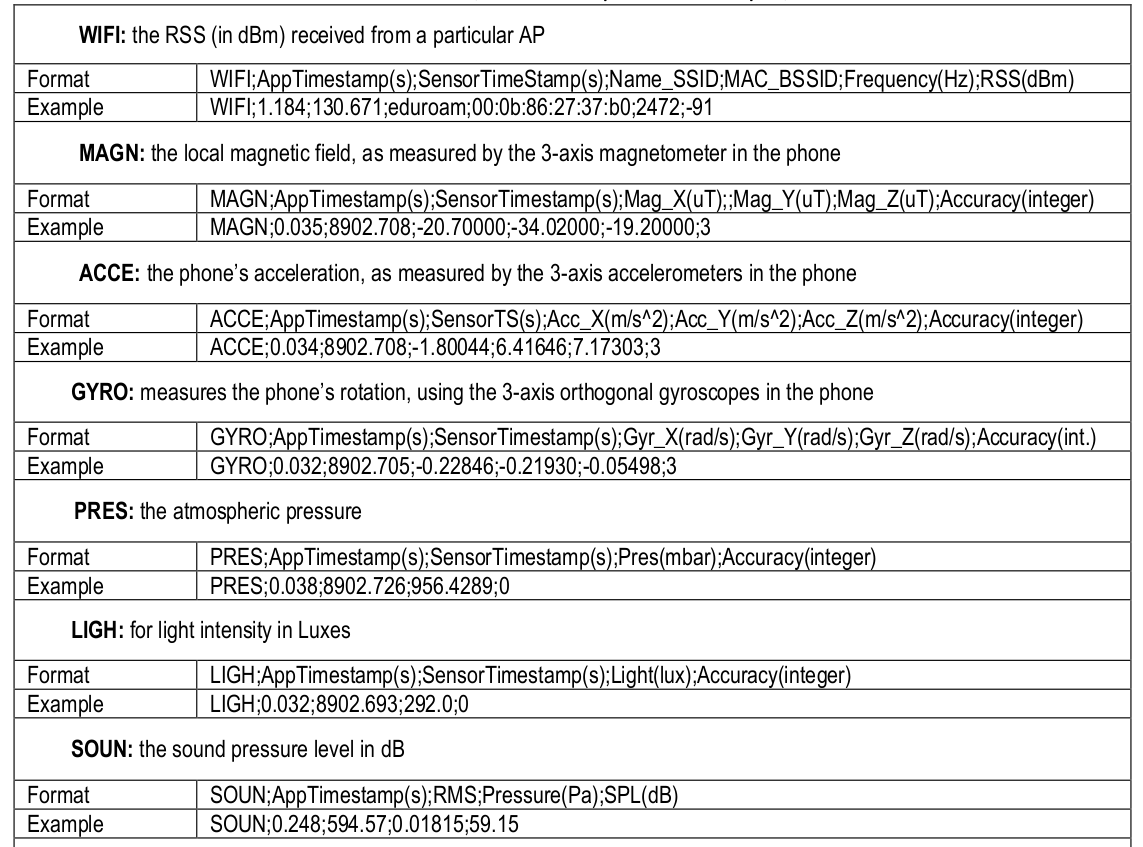
输出imu.txt,wifi.txt,gnss.txt共3个文件，各文件格式说明参见附录

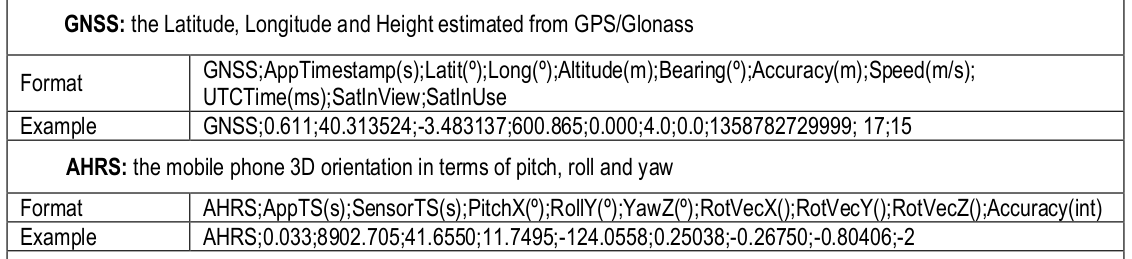
1. **处理要求：**
2. 从文件读取输入文件的每一行，丢弃注释数据，对有效数据，根据分割符“;”进行分割，放入一个二维字符数组char buf[M][N]，其中M为分割出的数据域最大值，N为字符串最大长度。
3. 根据读取到的数据类型，调用不同的子函数（分为imu，wifi,gnss三个处理函数）对buf中的数据进行处理。
4. 三个处理函数对buf进行处理后将结果缓存，每100条数据写一次文件。
5. 仅对这三个输出文件所需的数据进行处理，其他数据不处理
6. **其他要求**
7. 测试时间9:00-17:00
8. 在代码编写过程中需使用git管理代码，每小时提交一次git记录（无论编写或修改多少代码）,最终提交需将git提交记录一并提交。
9. 代码应有合理注释，并注意代码风格
10. 主要处理流程参考代码（matlab）

https://gitlab.com/getsensordatasuite/getsensordata\_matlab/-/blob/master/ReadLogFile21.m

**附录：**

**输入文件详细格式说明：**





需注意的是WIFI输入数据有两种格式，其中Frequency(Hz)为可选数据。具体应检查文件开始时的WIFI数据格式说明。

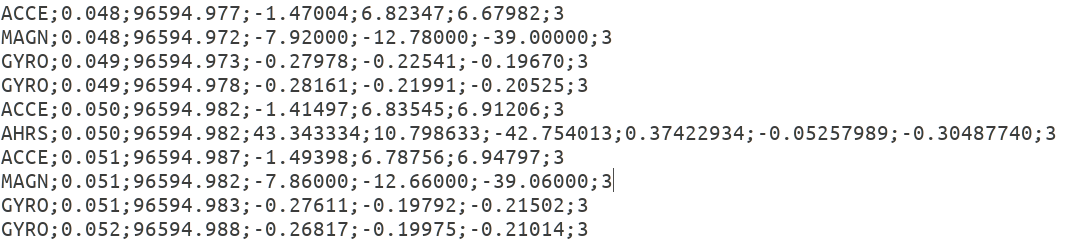
**输出文件格式说明**

* imu.txt:每行一条数据，格式及要求如下

TIME,Acc\_x,Acc\_y,Acc\_z,Gyr\_x,Gyr\_y,Gyr\_z,Mag\_x,Mag\_y,Mag\_z,pitchX,RollY,YawZ,Pres

其中TIME为该条数据SensorTimestamp(每条数据第三栏)与最小SensorTimestamp的差值，单位为毫秒（输入数据时间单位为秒）。此处最小SensorTimestamp为类型为ACCE、GYRO、MAGN、PRES、AHRS的输入数据的最小值。

例如数据集最开始有如下数据



这些数据中SensorTimestamp的最小值为96594.972，对于数据

ACCE;0.058;96595.002;-1.79086;6.58405;6.69898;3

其换算出的输出时间TIME=(96595.002-96594.972)\*1000=30

ACCE、GYRO、数据采集频率均为200HZ，MAGN、AHRS为100HZ，PRES约为5HZ，由于数据采集频率不同，应做插值处理（采用线性插值），将所有数据频率均变为200HZ。若ACCE、GYRO、MAGN、AHRS、PRES的时间戳不同，则将它们进行对齐处理，将最接近的数据拼在一起，且该条数据的TIME为所有拼接数据的最小值。若有数据缺失则应使用前后数据进行线性插值补齐，若缺失数据为采集开始时的数据，则丢弃，直到所有数据均已出现。

例如对于数据

1. ACCE;0.050;96594.982;-1.41497;6.83545;6.91206;3
2. AHRS;0.050;96594.982;43.343334;10.798633;-42.754013;0.37422934;-0.05257989;-0.30487740;3
3. MAGN;0.051;96594.982;-7.86000;-12.66000;-39.06000;3
4. GYRO;0.051;96594.983;-0.27611;-0.19792;-0.21502;3
5. PRES;0.071;96595.033;1008.2056;3
6. AHRS;0.074;96595.042;42.485546;10.190693;-43.268139;0.36598542;-0.05610375;-0.31235692;3
7. MAGN;0.075;96595.042;-8.10000;-11.52000;-39.84000;3
8. ACCE;0.074;96595.047;-1.16598;6.71813;6.81629;3
9. GYRO;0.075;96595.048;-0.36591;-0.18143;-0.09652;3
10. AHRS;0.077;96595.051;42.315960;10.063487;-43.273022;0.36437577;-0.05654010;-0.31310943;3
11. MAGN;0.078;96595.052;-8.28000;-11.82000;-39.66000;3
12. PRES;0.235;96595.213;1008.1990;3

前4条数据应丢弃，因第5条数据为首次出现，此时还没有采集到Press数据。对于8、9两条数据则在合并时将8，9拼接，这些数据中第9条GYRO数据与8条数据时间相差1ms，因而将这些数据拼接在一起（它与第其他数据时间差距更大）由于该时刻缺少AHRS、MAG和PRES数据，应使用插值的方法补全数据。对于AHRS则用6、10两条数据插值得到，MAG数据则用7、11插值，PRES用5、12插值。

* wifi.txt

TIME,MAC\_D,RSS

其中TIME与前一文件一致。MAC\_D为整数，它从MAC\_BSSID地址的16进制字符串转换为整数得到，例如

MAC\_BSSID：20:20:00:00:00:23将其转换为整数MAC\_D=(2\*32\*256+2\*32)\*256\*256\*256\*256+2\*16+3

需注意的是WIFI输入数据有两种格式，应检查输入文件的格式说明，自动识别处理。

* gnss.txt

该文件对原始数据时间处理后复制到gnss.txt，格式为

CTIME,Latit,Long,Altitude,Bearing,Accuracy,Speed,SatInView,SatInUse

其中CTIME为AppTimestamp与文件第一条数据的差值，单位为秒。需要注意的是，对于GNSS数据，若两条数据SensorTimeStamp一样，仅有AppTimestamp不同，则同重复数据，应仅保留第一次出现的数据。