**作业2 - 概率图模型练习**

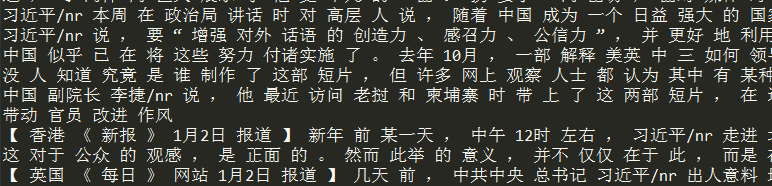
**任务2：中文人名识别**

**人名识别** 是命名实体识别的一部分，是将中文序列中的人名识别出来。

**要求：**用最大熵或条件随机场模型实现中文人名识别。

**数据资源：**语料来自1998年1月份人民日报语料。

**语料格式：**



其中，人名已显式标注。

**语料使用方法：**

按照8:2比例，将数据划分为训练数据和测试数据；训练数据部分用来训练模型参数，测试部分用来验证模型效果。

**模型验证：**

采用交叉验证方式，**指标：**准确率、召回率、F1值。

**模型说明：**

推荐使用条件随机场（CRF），

**使用方式可选择：**

（1）自己编码实现（鼓励有能力、感兴趣的同学采用这种方式）；

（2）利用开源工具包：

**条件随机场工具包：CRF++ （见目录“CRF++-0.58”）**

**作业截止日期：2017-11-17**