产品供应关系抽取进展

## 任务目标

从上市公司年报中找出该公司的客户和客户购买的产品。

## 信息来源

信息来自上市公司年报，具体来说分为两部分，一部分是纯文本信息，另一部分是格式化的信息。

由于格式化的信息抽取较为简单，目前主要研究是从纯文本信息中抽取供应关系。

## 整体流程图

整体处理流程如下图：

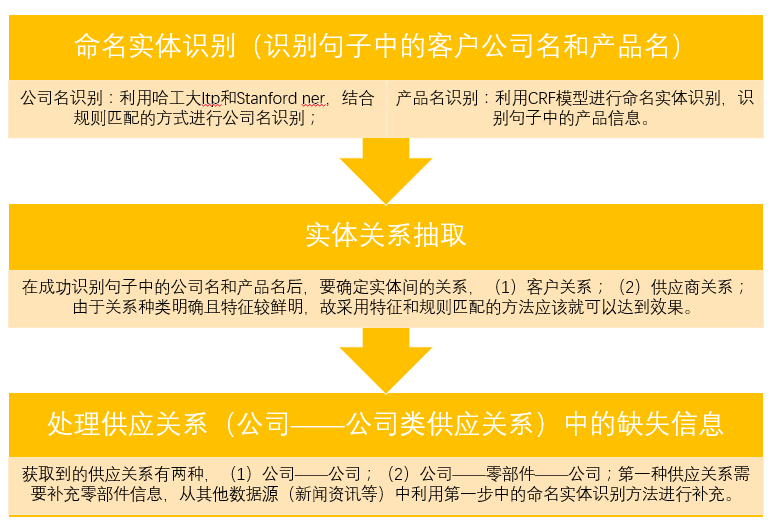
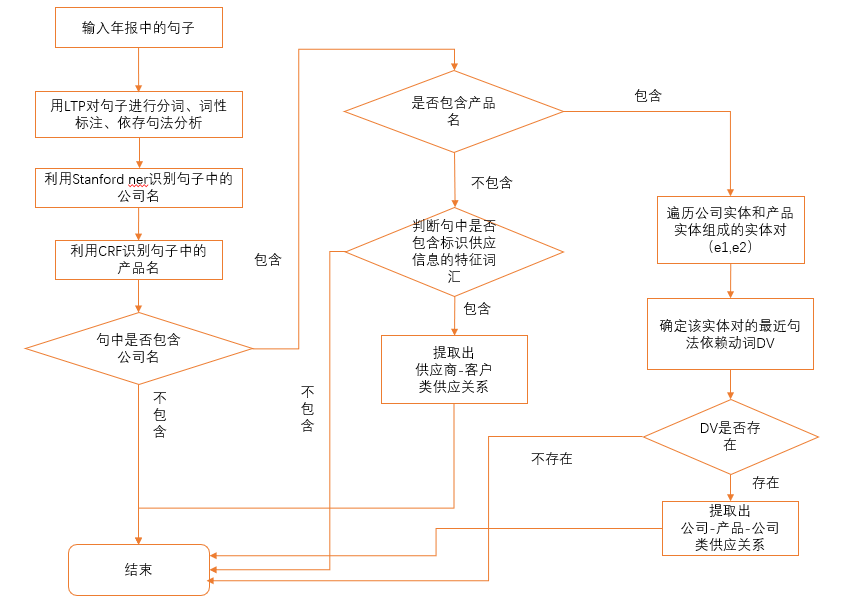


图1 整体流程图

目前已经完成了第一部分命名实体识别和第二部分实体关系抽取的任务。其中命名实体识别主要采用了机器学习的方法，第二部分实体关系抽取采用的是句法分析的方法，具体依赖的是哈工大的自然语言处理工具LTP。

## 目前处理流程



## 测试实例

从哪个公司年报中抽取的供应关系，其中的供应商就是哪个公司。

（1）例一

原句：LCM、触控业务加大对研发的投入力度，重点客户推进有力，集约化管理进步明显，年内拓展了小米、OPPO等新客户。

公司名：[小米, OPPO]

产品名：[LCM, 触控业务, 集约化]

供应关系：[小米-供应商-触控业务] [OPPO-供应商-触控业务]

[小米-供应商-LCM] [小米-供应商-触控业务]

（2）例二

原句：客户涵盖了包括百得、博世、 牧田、创科等在内的主要电动工具厂商。

公司名: [百, 博世, 牧田, 创科]

产品名：[]

供应关系：[百-供应商] [博世-供应商] [牧田-供应商] [创科-供应商]

（3）例三

原句：公司通信业务主营产品天线和滤波器是基站建设的核心组件，主要客户包括华为、爱立信等全球领先的移动通信网络设备制造商。

公司名：[华为, 爱立信]

产品名：[产品天线, 滤波器, 核心组件, 移动通信网络设备制造商]

供应关系：[华为-供应商-产品天线] [华为-供应商-滤波器]

[爱立信-供应商-产品天线] [爱立信-供应商-滤波器]

(4) 例四

目前公司已拥有LED封装、LED背光、触控面板、模组贴合、结构件的全产业链完整产能，拥有联想、JDI、小米、华为等行业知名客户以及具备一定技术基础的研发团队。

公司名：[联想, JDI, 小米, 华为]

产品名：[LED封装, LED背光, 触控面板, 模组贴合, 结构件, 完整产能]

供应关系：空

（5）例五

得益于公司的市场布局及大客户战略，伴随着OPPO、VIVO、华为等客户市场份额的增长，2016年公司在金属外观件等产品领域取得了高速的增长。

公司名：[OPPO, VIVO, 华为]

产品名：[金属外观件等产品领域]

供应关系：[OPPO-供应商-金属外观件等产品领域]

[VIVO-供应商-金属外观件等产品领域]

[华为-供应商-金属外观件等产品领域]

（6）例六

京东方已成为半导体显示领域世界顶级供货商，与包括三星、LG、海信、康佳、联想、戴尔、惠普等在内的国内外知名客户保持了长期、可持续的合作，是众多国际一线品牌的第一供应商。

公司名：[三星, LG, 海信, 康佳, 联想, 戴尔, 惠普]

产品名：[半导体显示领域]

供应关系：[空]

（7）例七

车辆减震产品获得了华晨汽车等工厂认可并获得整车悬置件的订单，并有十多个车型的减震产品正在开发研制，汽车密封件整车业务全面推进，并逐步进入上汽荣威、上汽大通、一汽轿车、一汽大众、广汽三菱、北汽及新能源车的整车密封供系统， 为公司的持续发展打下了坚实基础。

公司名：[华晨汽车, 上汽荣威, 上汽大通, 一汽轿车, 一汽大众, 广汽三菱, 北汽, 新能源车, 整车, 打下]

产品名：[车辆减震产品, 悬置件, 汽车密封件整车业务, 的, 密封供系统]

供应关系：

[华晨汽车-供应商-车辆减震产品]

[上汽荣威-供应商-车辆减震产品]

[上汽大通-供应商-车辆减震产品]

[一汽轿车-供应商-车辆减震产品]

[一汽大众-供应商-车辆减震产品]

[广汽三菱-供应商-车辆减震产品]

[北汽-供应商-车辆减震产品]

[新能源车-供应商-车辆减震产品]

[整车-供应商-车辆减震产品]

[打下-供应商-车辆减震产品]

## 待改进的地方

有些包含句子既包含公司名也包含产品名，在实际操作中由于公司和产品之前没有动词相关联，直接把这类语句略过了，但实际上是包含有价值信息的，例如上面的例6，虽然不能明确的知道给客户供应了哪些产品，但是至少知道客户有些。接下来要把这一类句子中的信息利用起来。

Loading classifier from classifiers/chinese.misc.distsim.crf.ser.gz ... done [4.7 sec].

[小米, OPPO]

[LCM, 触控业务, 集约化]

小米 测试公司 触控业务

OPPO 测试公司 触控业务

[百, 博世, 牧田, 创科]

[]

百 测试公司

博世 测试公司

牧田 测试公司

创科 测试公司

[华为, 爱立信]

[产品天线, 滤波器, 核心组件, 移动通信网络设备制造商]

华为 测试公司 产品天线

华为 测试公司 滤波器

爱立信 测试公司 产品天线

爱立信 测试公司 滤波器

[联想, JDI, 小米, 华为]

[LED封装, LED背光, 触控面板, 模组贴合, 结构件, 完整产能]

[OPPO, VIVO, 华为]

[金属外观件等产品领域]

OPPO 测试公司 金属外观件等产品领域

VIVO 测试公司 金属外观件等产品领域

华为 测试公司 金属外观件等产品领域

[三星, LG, 海信, 康佳, 联想, 戴尔, 惠普]

[半导体显示领域]

[华晨汽车, 上汽荣威, 上汽大通, 一汽轿车, 一汽大众, 广汽三菱, 北汽, 新能源车, 整车, 打下]

[车辆减震产品, 悬置件, 汽车密封件整车业务, 的, 密封供系统]