



IT大咖说
知识共享平台

OCR图像识别算法行业应用分析

胡丰

袋鼠云高级算法专家



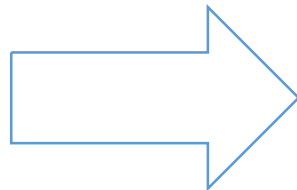
AICUG

目录

CONTENTS

- 背景概述
- 分析方法
- 行业应用场景分析
- 难点分析
- 袋鼠云案例-医院费用清单识别



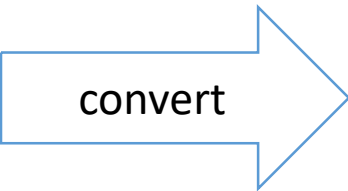




复旦大学附属中山医院病人住院费用汇总表

三十大厦4楼 门诊 住院部(2017-07-05 14:00) 2017-07-24 14:21

项目	名称	规格	单位	数量	金额
药品费小计:	诺和灵注射液 A/D	40U:30	支	10.0	100.00
药品费小计:	丹参注射液(诺和)I	250ml:15	瓶	1.0	60.00
药品费小计:					400.00
材料费小计:	注射用水		L	1.00	0.00
材料费小计:	注射用水		L	1.00	0.00
材料费小计:	葡萄糖		100g	1.00	48.00
材料费小计:	葡萄糖氯化钠		L	1.00	0.00
材料费小计:	葡萄糖		100g	1.00	48.00
材料费小计:	小瓶装		L	1.00	11.00
材料费小计:	葡萄糖注射液		L	2.00	4.00
材料费小计:					100.00
材料费小计:	注射液	250ml:15	瓶	1.00	48.00
材料费小计:					100.00



目录

CONTENTS

背景概述

分析方法

行业应用场景分析

难点分析

袋鼠云案例-医院费用清单识别



IT大咖说
知识共享平台



How OCR work---- General method



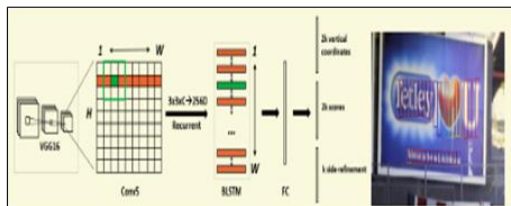
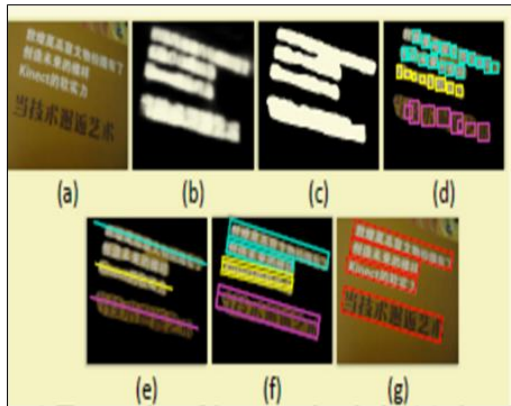
- Image input
- Image pre-processing
- Image segmentation
- Layout analysis
- Normalise
- Character recognition
- NLP Post-processing



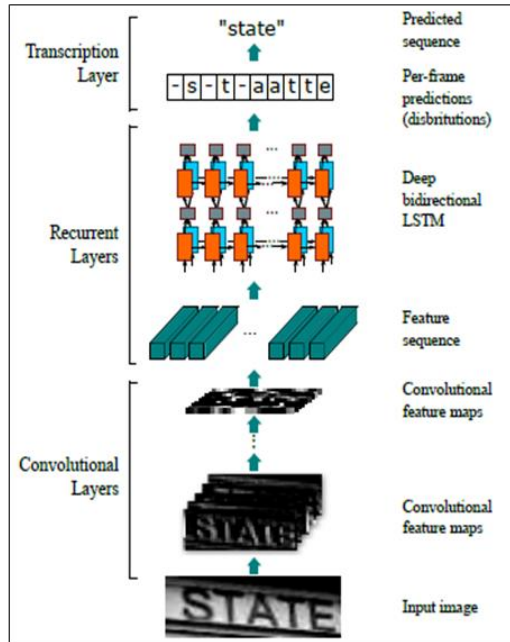
How OCR work---- Deep learning



Char detection



Character recognition



目录

CONTENTS

背景概述

分析方法

行业应用场景分析

难点分析

袋鼠云案例-医院费用清单识别



IT大咖说
知识共享平台



应用场景--车牌识别



IT大咖说
知识共享平台

- **电子警察**
主要安装于交叉路口附近，可以对过往的每辆车抓拍（卡口型电警），也可以抓拍违章车辆（闯红灯、违章变道、逆行等）
- **视频卡口**
检测往来车辆，抓拍每一辆车，识别包括车牌，车身颜色，车速等信息
- **智能相机**
与线圈、雷达等产品配合，线圈雷达负责捕获过往车辆，智能相机识别抓拍张图片的车牌



场景图像捕获



车牌检测、定位



车牌分割、字符切分



车牌字符识别



应用场景--证件OCR识别



最开始是基于PC的，近几年开始向移动端发展，主要有android，ios平台的SDK，目前成熟的有身份证识别，行驶证识别，驾驶证识别，护照识别等。

定制化程度高，识别率高

应用场景--文档OCR识别

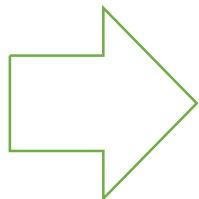


打印日期: 2017-08-02 08:44 202017-08-03 14:53:56

ICD9	项目名称	单位	单价	数量	医保金额	医保金额	医保金额	医保支付	总金额
	甲上链霉素注射液 30mg*10支/盒	支	1.36	1	1.36	0.00	0.00	1.36	
	甲硝唑注射液(多剂量) 500mg*10支/盒	支	0.42	30	12.60	0.00	0.00	12.60	
	甲硝唑注射液(多剂量) 500mg*10支/盒	支	0.42	60	25.20	0.00	0.00	25.20	
	甲硝唑注射液(多剂量) 500mg*10支/盒	支	2.38	1	2.38	0.00	0.00	2.38	
	▲乙脑灭活疫苗(乙脑灭活疫苗) 0.5ml*1支/盒	盒	94.13	1	94.13	0.00	0.00	94.13	
	乙脑灭活疫苗(乙脑灭活疫苗) 0.5ml*1支/盒	盒	93.04	1	93.04	0.00	0.00	93.04	
	甲肝灭活疫苗(甲肝灭活疫苗) 0.5ml*1支/盒	盒	3.50	1	3.50	0.00	0.00	3.50	
	甲肝灭活疫苗(甲肝灭活疫苗) 0.5ml*1支/盒	盒	4.49	3	13.47	0.00	0.00	13.47	
	甲肝灭活疫苗(甲肝灭活疫苗) 0.5ml*1支/盒	盒	3.44	2	6.82	0.00	0.00	6.82	
	甲肝灭活疫苗(甲肝灭活疫苗) 0.5ml*1支/盒	盒	3.31	1	3.31	0.00	0.00	3.31	
	乙脑灭活疫苗(乙脑灭活疫苗) 0.5ml*1支/盒	支	10.79	7	75.53	0.00	0.00	75.53	
	乙脑灭活疫苗(乙脑灭活疫苗) 0.5ml*1支/盒	支	331.09	1	331.09	0.00	0.00	331.09	
	乙脑灭活疫苗(乙脑灭活疫苗) 0.5ml*1支/盒	支	448.70	1	448.70	0.00	0.00	448.70	
	乙脑灭活疫苗(乙脑灭活疫苗) 0.5ml*1支/盒	支	120.01	0.00	0.00	0.00	0.00	120.01	
	乙脑灭活疫苗(乙脑灭活疫苗) 0.5ml*1支/盒	支	6.32	0.00	0.00	0.00	0.00	6.32	
	乙脑灭活疫苗(乙脑灭活疫苗) 0.5ml*1支/盒	支	32.11	0.00	0.00	0.00	0.00	32.11	
15101001	肌内注射 单次	次	3.00	1	3.00	0.00	0.00	3.00	
15101003	静脉注射 单次	次	12.00	1	12.00	0.00	0.00	12.00	
15101004	静脉注射 单次	次	30.00	1	30.00	0.00	0.00	30.00	
15101001	肌内注射 单次	次	10.00	4	40.00	0.00	0.00	40.00	
	乙脑灭活疫苗(乙脑灭活疫苗) 0.5ml*1支/盒	支	85.00	0.00	0.00	0.00	0.00	85.00	
10100022	血常规 每升	升	3.00	8	24.00	0.00	0.00	24.00	
	血常规 每升	升	24.00	0.00	0.00	0.00	0.00	24.00	
10100006	血常规 每升	升	25.00	1.5	37.50	0.00	0.00	37.50	
	血常规 每升	升	37.50	0.00	0.00	0.00	0.00	37.50	
10100002	血常规 每升	升	35.00	1.5	52.50	0.00	0.00	52.50	
	血常规 每升	升	52.50	0.00	0.00	0.00	0.00	52.50	
13101001	血常规 每升	升	28.00	2	56.00	0.00	0.00	56.00	
13101002	血常规 每升	升	10.00	1	10.00	0.00	0.00	10.00	
	血常规 每升	升	66.00	0.00	0.00	0.00	0.00	66.00	
06	一次性无菌手术衣(灭菌) 0.5ml*1支/盒	支	2.31	1	2.31	0.00	0.00	2.31	
	一次性无菌手术衣(灭菌) 0.5ml*1支/盒	支	19.80	0.00	0.00	0.00	0.00	19.80	
0411	血常规 每升	升	17.00	1	17.00	0.00	0.00	17.00	
	血常规 每升	升	24.53	1	24.53	0.00	0.00	24.53	
06	血常规 每升	升	53.90	0.00	0.00	0.00	0.00	53.90	
	血常规 每升	升	181.11	0.00	0.00	0.00	0.00	181.11	
	血常规 每升	升	176.03	0.00	0.00	0.00	0.00	176.03	

打印日期: 2017-08-03 15:02:15

编码	名称	规格	价格
110900001e	三级医院(等级病种) -ICU		48.0000
110200005c	住院费(三级医院)		25.0000
120400001	肌肉注射		3.0000
120400001	皮下注射		3.0000
120400002	静脉注射		5.0000
120400004	动脉加压注射		7.0000
120400006	静脉输液		12.0000
120600004	小换药		14.0000
121100001	一般物理降温		3.0000
120100004	二级护理		28.0000
330100001	局部浸润麻醉		8.0000
220301002	浅表器官彩色多普勒超声检查(体表)		60.0000
220302012	临床操作的彩色多普勒超声引导(体表)		60.0000
220600004	心脏彩色多普勒超声		130.0000
220600010	左心功能测定		50.0000
220700004	组织多普勒成像(TDI)		30.0000
230400010a	PET/CT全身检查		7000.0000
270300001	穿膜组织活检检查与诊断		150.0000
270300004	骨髓组织活检检查与诊断		140.0000
270500001	特殊染色及酶组织化学染色诊断		80.0000
270500002	免疫组织化学染色诊断		120.0000
270500002	免疫组织化学染色诊断(第9只开始)		60.0000
270500002	免疫组织化学染色诊断(全自动操作)		30.0000
270500002	免疫组织化学染色诊断(全自动操作)		60.0000
270700003a	脱氧核糖核酸(DNA)测序术		600.0000
310401010	声导抗测定		40.0000
310403006	电子鼻咽喉镜检查		150.0000
310800001	骨髓穿刺术		100.0000
311400003b	皮肤活检术(头面部活检或深部活检)		550.0000
250101007	★血常规细胞形态学检查		10.0000
250101008b	★红细胞沉降率测定(ESR) (仪器法)		10.0000
250101019c	★血细胞分析(三类及以上)		20.0000
250102024	★尿沉渣定量		20.0000
250102035b	★尿液分析(尿9类及以上)		10.0000
250103001a	★粪便常规(手工法)		3.0000
250103002b	★凝血试验(免疫学方法)		10.0000
250201001	骨髓涂片细胞学检查		120.0000
250201007	骨髓特殊染色及酶组织化学染色检查		40.0000
250203020b	★血浆凝血酶原时间测定(PT) (仪器)		15.0000
250203025b	★活化部分凝血活酶时间测定(APT) (仪器)		15.0000
250203026b	★血纤蛋白原含量测定(仪器法)		10.0000
250203035b	★凝血酶时间测定(TT) (仪器法)		15.0000



- 医院名称
- 患者姓名
- 住院日期
- 药品名 (报销种类)
- 金额
- 数量
- 价格
- 总价



目录

CONTENTS

背景概述

分析方法

行业应用场景分析

难点分析

袋鼠云案例-医院费用清单识别



IT大咖说
知识共享平台



难点:非人为因素----难以避免



图像模糊、光照不均、背景干扰、排版布局、多语种混合...

难点:人为因素、道高一尺魔高一丈



【“开车”等词或将纳入网络媒体信息敏感词库】

据国家互联网文化信息部发布的通告称,近年来例如“开车”“老司机”“秋名山”等看似普通的词汇已经逐渐被网络不法分子用于传播淫秽色情信息。为了杜绝此类现象继续泛滥,增加侦查打击力度,还给广大网民安全洁净的网络环境,从2017年2月8日起,诸如“开车”“老司机”“秋名山”等一共41个词汇将被纳入敏感信息词库。对此,网友看法不一,有网友认为,坚决支持扫黄打非行动,共建美好网络环境;也有网友指出,被添加的许多词汇都是生活日常用语,被加入敏感词库后会对正常的网络交流造成严重不便。专家解读:2017,我们该如何抵御“开车”。

♥点♥我♥看♥好♥玩♥
内♥容♥请♥关♥注♥涨♥
♥姿♥势♥微♥信♥公♥
众♥号♥, 原本正常的一句话加上几颗小爱心就立马变得不一样了, 归结原因都要感谢那群伟大的生理教育工作者兼评论家——WB卖片党

- 打广告时很隐晦, 比如: 在贴吧不会说成卖片, 说成菜农售卖日本种子, 懂的人自然会懂
- 广告时间尽量要在夜晚

现代银行工业
Modern Business Trade Industry

当前金融票据诈骗 犯罪的特点及防范对策分析

周纪忠 李峰华¹

(1. 安徽江淮律师事务所, 安徽合肥 230001; 2. 安徽工业大学, 安徽 合肥 230021)

【摘 要】近几年票据诈骗案件频发, 涉案金额巨大, 给金融行业带来巨大损失, 给社会造成恶劣影响。本文通过分析当前票据诈骗案件的特点, 结合票据法的相关规定, 从票据诈骗犯罪的构成要件、犯罪主体、主观方面、客观方面、客体等方面进行了分析, 并对如何防范票据诈骗提出了建议。

关键词: 票据; 诈骗; 防范

中图分类号: D912.18

文献标识码: A

文章编号: 1672-3338(2019)16-0250-02

2002年10月, 2004年1月, 全国人大常委会相继颁布《刑法修正案(四)》和《刑法修正案(五)》, 对票据诈骗罪进行了修订。截至2014年11月30日, 全国法院共受理票据诈骗案件102件, 涉案金额达2.5亿元, 涉案被告人120人。2015年1月1日起, 最高人民法院和司法部联合发布《关于依法妥善审理票据纠纷案件促进国家纸质票据应用规范电子票据行为的指导意见》, 对票据纠纷案件的审理进行了规范和指导。

2016年1月, 最高人民法院发布《关于审理票据纠纷案件若干问题的规定》, 对票据纠纷案件的审理进行了规范和指导。2016年11月, 最高人民法院发布《关于审理票据纠纷案件若干问题的规定》, 对票据纠纷案件的审理进行了规范和指导。

2017年1月, 最高人民法院发布《关于审理票据纠纷案件若干问题的规定》, 对票据纠纷案件的审理进行了规范和指导。

2018年1月, 最高人民法院发布《关于审理票据纠纷案件若干问题的规定》, 对票据纠纷案件的审理进行了规范和指导。

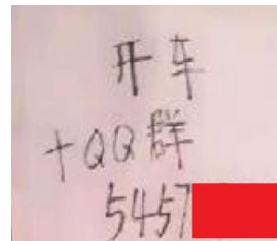
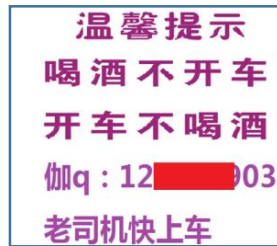
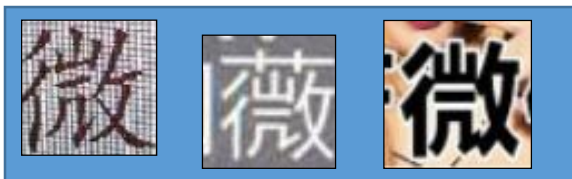
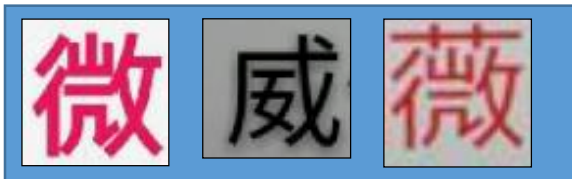
2019年1月, 最高人民法院发布《关于审理票据纠纷案件若干问题的规定》, 对票据纠纷案件的审理进行了规范和指导。



难点:人为因素、道高一尺魔高一丈



IT大咖说
知识共享平台



AICUG人工智能技术沙龙 (杭州站)

目录

CONTENTS

背景概述

分析方法

行业应用场景分析

难点分析

袋鼠云案例-医院费用清单识别



IT大咖说
知识共享平台





政策层面

贯彻落实“最多跑一次”政策要求，社保管理中心从基层经办人员业务培训到系统平台功能优化等多个维度，积极推进“一窗受理、集成服务”等改革事项落地，逐步提升群众满意度



行业刚需

当前异地就医人员必须先行与医院进行全额费用结算，然后将医疗过程中产生的相关单据送到各地社保中心进行审核后才能报销。造成有限的社保经办资源与庞大的受众服务之间的矛盾越发突出。



行业瓶颈

需要花费大量的人力来进行处理，使医保业务专家陷入事务性工作中，同时不均匀的医保单据报送量与经办资源及其处理能力之间的矛盾越发突出



公平公正

全程手工录入与审核，不但录入的准确率无法保证，且审核过程中人为因素较多，“公平”原则不好落实



人工审核

手工方式下单据的复查、检索面临巨大的困难



数据质量

由于手工分割，一般情况下分类会比较粗放，数据细节丢失严重，不利于数据沉淀



社保行业票据智能识别方案



前台综合受理、后
台分类审批



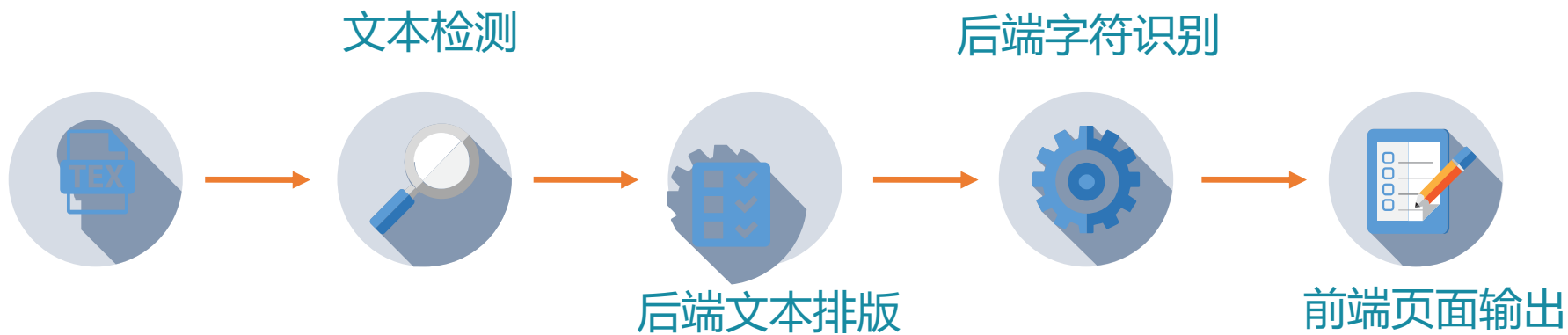
有效降低经办审核
业务的周期



提升基层经办
业务效能



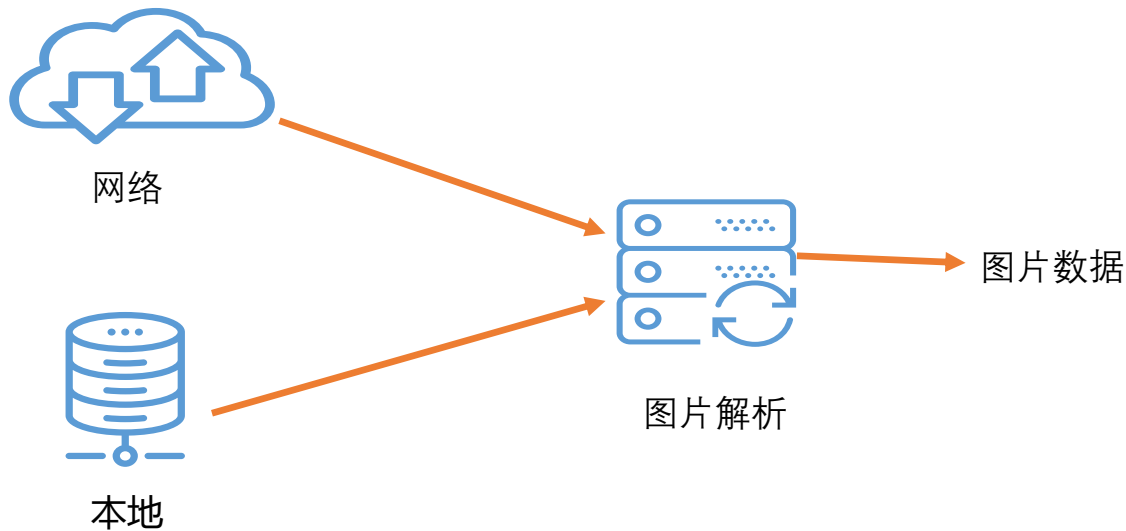
进一步提升社保公
共服务整体水平

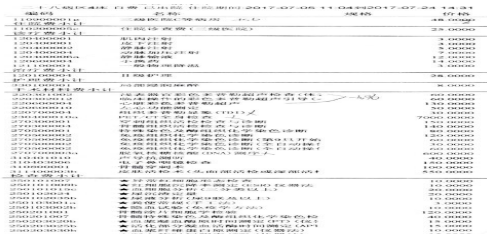


图像获取



从本地路径或网络端接收图片，读取解析图片数据供后续文本检测、分析识别用





文字行检测

250304006a 镁测定 ※*1/

文字段检测

250304006a 镁测定 ※*1/

文本段连接

250304006a 镁测定 ※*1/



编码	名称	规格	价格
110000001g	三级医院C类病房(一、二、三)		48.0000
110200005a	住院诊查费(三级医院)		25.0000
120400001	肌肉注射		3.0000
120400001	皮下注射		3.0000
120400002	静脉注射		5.0000
120400004	动脉加压注射		7.0000
120400006a	静脉输液		12.0000
120600004	小手术		14.0000
121100001	一般物理降温		3.0000
120100004	二级护理		28.0000
330100001	局部浸润麻醉		8.0000
220301002	浅表器官彩色多普勒超声检查(体)		60.0000
220302012	临床操作的彩色多普勒超声引导(C)		60.0000
220600004	心脏彩色多普勒超声		130.0000
220600010	左心功能测定		30.0000
220700004	组织多普勒显象(TDI)		30.0000
230400010a	PET-CT全身检查		7000.0000
270300001	窄带组织活检检查与诊断		150.0000
270300004	骨髓组织活检检查与诊断		140.0000
270500001	特殊染色及酶组织化学染色诊断		80.0000
270500002	免疫组织化学染色诊断		120.0000
270500002	免疫组织化学染色诊断(第9只开始)		60.0000
270500002	免疫组织化学染色诊断(全自动操作)		30.0000
270500002	免疫组织化学染色诊断(全自动操作)		60.0000
270700003a	脱氧核糖核酸(DNA)测序		600.0000
310401010	声导抗测定		40.0000
310403006	电子鼻咽镜检查		150.0000
311000001	骨髓穿刺术		100.0000
311400003b	皮肤活检术(头面部活检或深部活检)		550.0000
250101007	★异型红细胞形态检查		10.0000
250101008b	★红细胞沉降率测定(ESR)仪器法		10.0000
250101013c	★血细胞分析(三分类以上)		20.0000
250102024	★尿沉渣定量		20.0000
250102035b	★尿液分析(尿9联及以上)		10.0000
250103001a	★粪便常规(手工法)		3.0000
250103002b	★血涂片镜检(免疫学方法)		10.0000
250201001	骨髓涂片细胞学检查		120.0000
250201007	骨髓特殊染色及酶组织化学染色		40.0000
250203020b	★血常规血细胞时间测定(PT)(仪器)		15.0000
250203025b	★活化部分凝血活酶时间测定(APTT)		15.0000
250203030b	★血浆纤维蛋白原测定(仪器法)		10.0000
250203035b	★凝血酶时间测定(TT)(仪器法)		15.0000

文档图像



将检测得到文本行与其对应文档中的空间位置对应、同时去除直线、表格等干扰，得到文本的排版结果

编码	名称	规格	价格
110000001g	三级医院C类病房(一、二、三)		48.0000
110200005a	住院诊查费(三级医院)		25.0000
120400001	肌肉注射		3.0000
120400001	皮下注射		3.0000
120400002	静脉注射		5.0000
120400004	动脉加压注射		7.0000
120400006a	静脉输液		12.0000
120600004	小手术		14.0000
121100001	一般物理降温		3.0000
120100004	二级护理		28.0000
330100001	局部浸润麻醉		8.0000
220301002	浅表器官彩色多普勒超声检查(体)		60.0000
220302012	临床操作的彩色多普勒超声引导(C)		60.0000
220600004	心脏彩色多普勒超声		130.0000
220600010	左心功能测定		30.0000
220700004	组织多普勒显象(TDI)		30.0000
230400010a	PET-CT全身检查		7000.0000
270300001	窄带组织活检检查与诊断		150.0000
270300004	骨髓组织活检检查与诊断		140.0000
270500001	特殊染色及酶组织化学染色诊断		80.0000
270500002	免疫组织化学染色诊断		120.0000
270500002	免疫组织化学染色诊断(第9只开始)		60.0000
270500002	免疫组织化学染色诊断(全自动操作)		30.0000
270500002	免疫组织化学染色诊断(全自动操作)		60.0000
270700003a	脱氧核糖核酸(DNA)测序		600.0000
310401010	声导抗测定		40.0000
310403006	电子鼻咽镜检查		150.0000
311000001	骨髓穿刺术		100.0000
311400003b	皮肤活检术(头面部活检或深部活检)		550.0000
250101007	★异型红细胞形态检查		10.0000
250101008b	★红细胞沉降率测定(ESR)仪器法		10.0000
250101013c	★血细胞分析(三分类以上)		20.0000
250102024	★尿沉渣定量		20.0000
250102035b	★尿液分析(尿9联及以上)		10.0000
250103001a	★粪便常规(手工法)		3.0000
250103002b	★血涂片镜检(免疫学方法)		10.0000
250201001	骨髓涂片细胞学检查		120.0000
250201007	骨髓特殊染色及酶组织化学染色		40.0000
250203020b	★血常规血细胞时间测定(PT)(仪器)		15.0000
250203025b	★活化部分凝血活酶时间测定(APTT)		15.0000
250203030b	★血浆纤维蛋白原测定(仪器法)		10.0000
250203035b	★凝血酶时间测定(TT)(仪器法)		15.0000

文本检测、排版

整体流程效果



文档图片

文本检测

文本识别前端显示

结构化存储业务应用

获取文档图片

对文档进行文本检测及排版分析

将识别结果显示在前端页面上

识别结果结构化，依业务需求输出对应信息并保存到电子表格或社保系统中

复旦大学附属中山医院病人住院费用汇总表

编码	名称	规格	价格	数量	金额
110900001e	三级医院C等病房	床位	48.0000	19.5	936.00
住院费小计					936.00
110200005c	住院诊查费(三级医院)		25.0000	19.5	487.50
诊疗费小计					487.50
120400001	肌肉注射		3.0000	1	3.00
120400001	皮下注射		3.0000	1	3.00
120400002	静脉注射		5.0000	118	590.00
120400004	动脉加压注射		7.0000	1	7.00
120400006a	静脉输液		12.0000	105	1,260.00
120600004	小换药		14.0000	1	14.00
121100001	一般物理降温		3.0000	2	6.00
治疗费小计					1,883.00
120100004	二级护理		28.0000	19.5	546.00
护理费小计					546.00

编码	名称	价格	数量	金额
110900001g	三级医院C等病房b	48.0000	19.5	936.00
住院费小计				936.00
110200005c	住院诊查费(三级医院)	25.0000	19.5	487.50
诊疗费小计				487.50
120400001	肌肉注射	3.0000	1	3.00
120400001	皮下注射	3.0000	1	3.00
120400002	静脉注射	5.0000	118	590.00
120400004	动脉加压注射	7.0000	1	7.00
120400006a	静脉输液	12.0000	105	1,260.00
120600004	小换药	14.0000	1	14.00
121100001	一般物理降温	3.0000	2	6.00
治疗费小计				1,883.00
120100004	二级护理	28.0000	19.5	546.00
护理费小计				546.00

复旦大学附属中山医院病人住院费用汇总表

编码	名称	价格	数量	金额
110900001g	三级医院C等病房b	48.0000	19.5	936.00
住院费小计				936.00
110200005c	住院诊查费(三级医院)	25.0000	19.5	487.50
诊疗费小计				487.50
120400001	肌肉注射	3.0000	1	3.00
120400001	皮下注射	3.0000	1	3.00
120400002	静脉注射	5.0000	118	590.00
120400004	动脉加压注射	7.0000	1	7.00
120400006a	静脉输液	12.0000	105	1,260.00
120600004	小换药	14.0000	1	14.00
121100001	一般物理降温	3.0000	2	6.00
治疗费小计				1,883.00
120100004	二级护理	28.0000	19.5	546.00
护理费小计				546.00

医院名称	价格	数量	金额
复旦大学附属中山医院 病人住院费用清单			
三级医院C等病房b	48.0000	19.5	936.00
住院诊查费(三级医院)	25.0000	19.5	487.50
肌肉注射	3.0000	1	3.00
皮下注射	3.0000	1	3.00
静脉注射	5.0000	118	590.00
动脉加压注射	7.0000	1	7.00
静脉输液	12.0000	105	1,260.00
小换药	14.0000	1	14.00
一般物理降温	3.0000	2	6.00
二级护理	28.0000	19.5	546.00





让未来变成现在

DI Brings the Future Today

