

知识图谱信息抽取在智能审单业务中的应用实践

- 姓名 : 刘占亮
- 公司 : 一览群智
- Title : 技术副总裁

百闻
是AI



7月3日-4日

CSDN

目录

- 一. 从感知智能到认知智能
- 二. 知识图谱简要介绍
- 三. 知识图谱构建技术
- 四. 知识图谱行业应用——智能审单系统



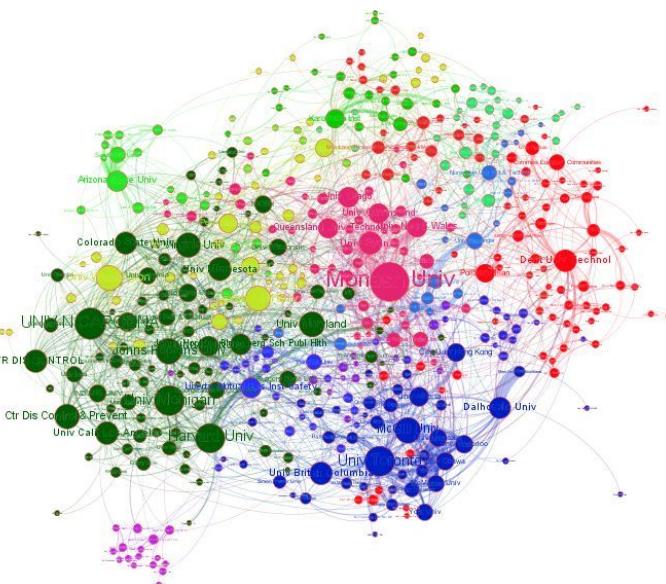
7月3日-4日

CSDN

从感知智能到认知智能



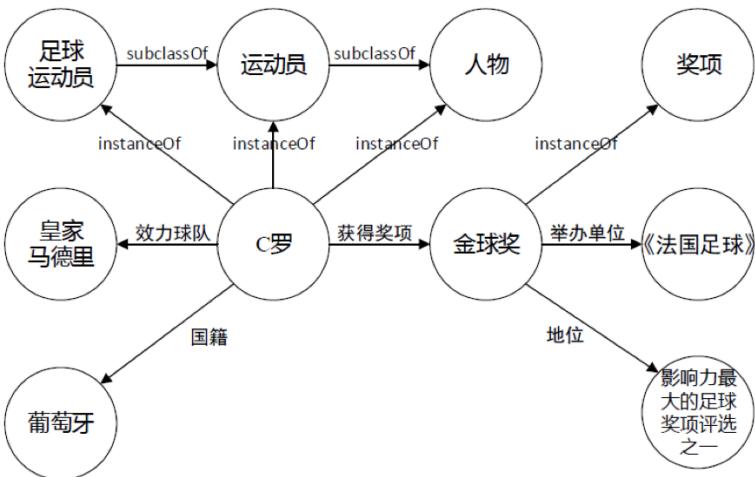
知识图谱简要介绍



知识图谱的概念是由谷歌公司在2012年5月17日提出的，谷歌公司以此为基础构建下一代智能化搜索引擎，知识图谱技术创造出一种全新的信息检索模式，为解决信息检索问题提供了新的思路。本质上，知识图谱是一种揭示实体之间关系的语义网络，可以对现实世界的事物及其相互关系进行形式化地描述。现在的知识图谱已被用来泛指各种大规模的知识库。



知识图谱简介介绍



知识图谱本质上是一种语义网络，将客观的经验沉淀在巨大的网络中，结点代表实体(entity)或者概念(concept)，边 (edge) 代表实体/概念之间的语义关系。知识图谱富含实体、概念、属性、关系等信息，使得机器理解与解释成为可能。成熟的图数据库如neo4j，Dgraph，JanusGraph等可以用来存储知识图谱。



2020
ProCon
万人开发者大会

7月3日-4日

CSDN

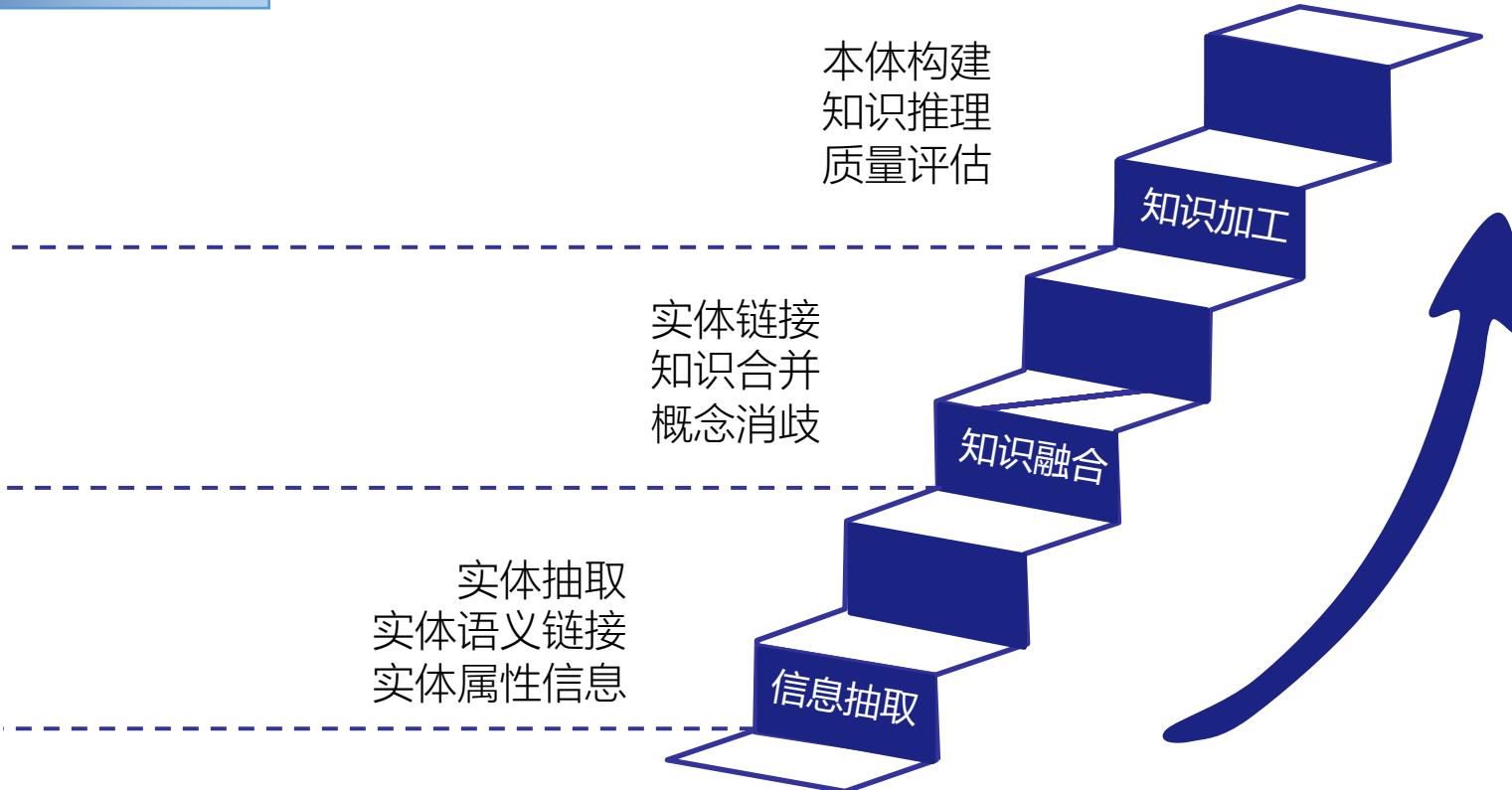
知识图谱使能认知智能



7月3日-4日

CSDN

知识图谱构建技术



深度学习的发展加速了知识图谱的构建



7月3日-4日

CSDN



2020
ProCon
万人开发者大会

7月3日-4日

CSDN

知识图谱构建技术

知识图谱整合不同种类的信息，形成一个关系网络。智策平台集成知识图谱平台，通过多规融合的知识图谱构建技术，从多源异构的数据中抽取实体信息，链接和融合实体，推理补全属性，识别语义并建立关系，并最终将知识存储于知识图谱数据库中，形成多种关系网络。

【自下而上、自上而下】





2020
ProCon
万人开发者大会

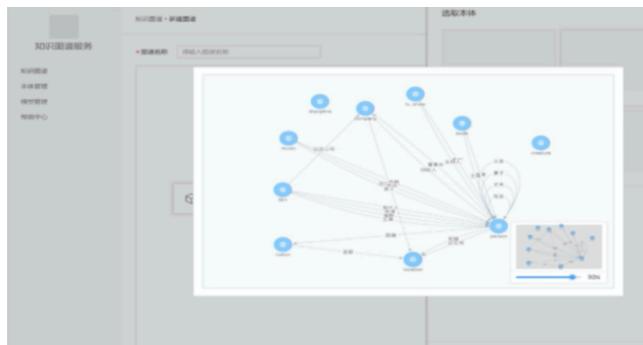
7月3日-4日

CSDN

知识图谱构建技术

——图谱构建流程工具

——可视化本体构建工具



——图谱编辑管理工具

——三元组数据标注工具

知识图谱分类

DBpedia
Freebase
WordNet
ConceptNet
CN-DBpedia

通用知识图谱

金融知识图谱
反欺诈知识图谱
企业知识图谱
商品知识图谱
图书馆知识图谱

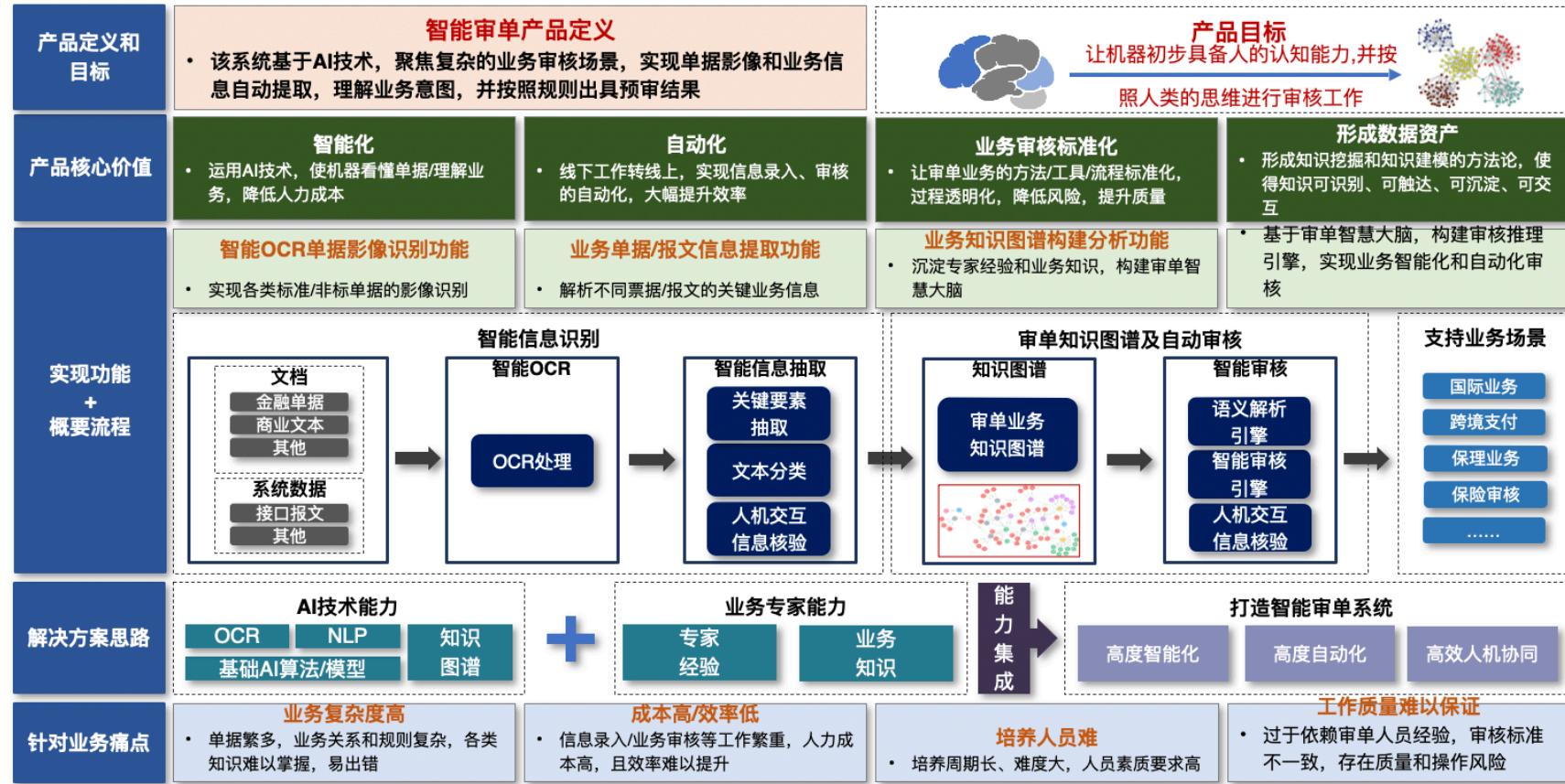
行业知识图谱



7月3日-4日

CSDN

知识图谱行业应用——智能审单系统



2020
ProCon
万人开发者大会

7月3日-4日

CSDN

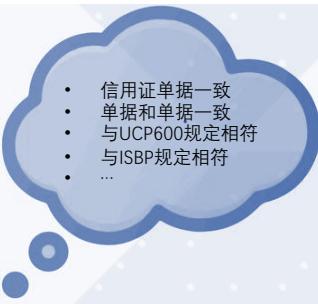
知识图谱行业应用——智能审单系统

国际结算

信用证/汇票/发票…



审核通过



跨境支付

汇票/申请书…



审核通过

单据和单据一致
与国家条例一致
符合外汇局规定

保理业务

发票/原产地证/质检证…



审核通过

贸易单据合规
担保材料合规
应收账款合规
符合保理业务实操
规范

保险理赔

保险单/索赔清单/检验证明…



审核通过

单据和单据一致
与保险条例一致
与理赔实操一致
符合理赔规范

以及其他应用场景。。

百川
启智
AI

2020
ProCon
万人开发者大会

7月3日-4日

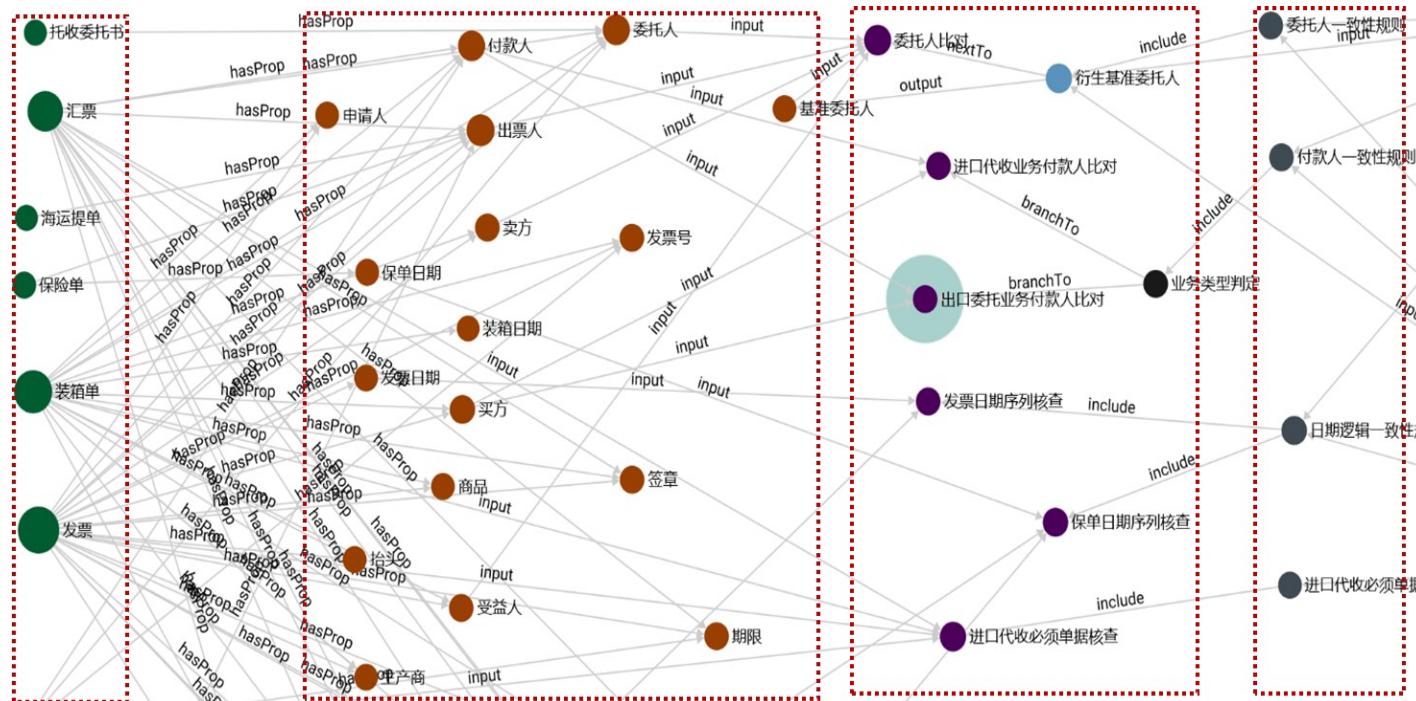
CSDN

知识图谱行业应用——智能审单系统

- 构建基础知识图谱、审核规则

知识图谱、单据知识图谱：

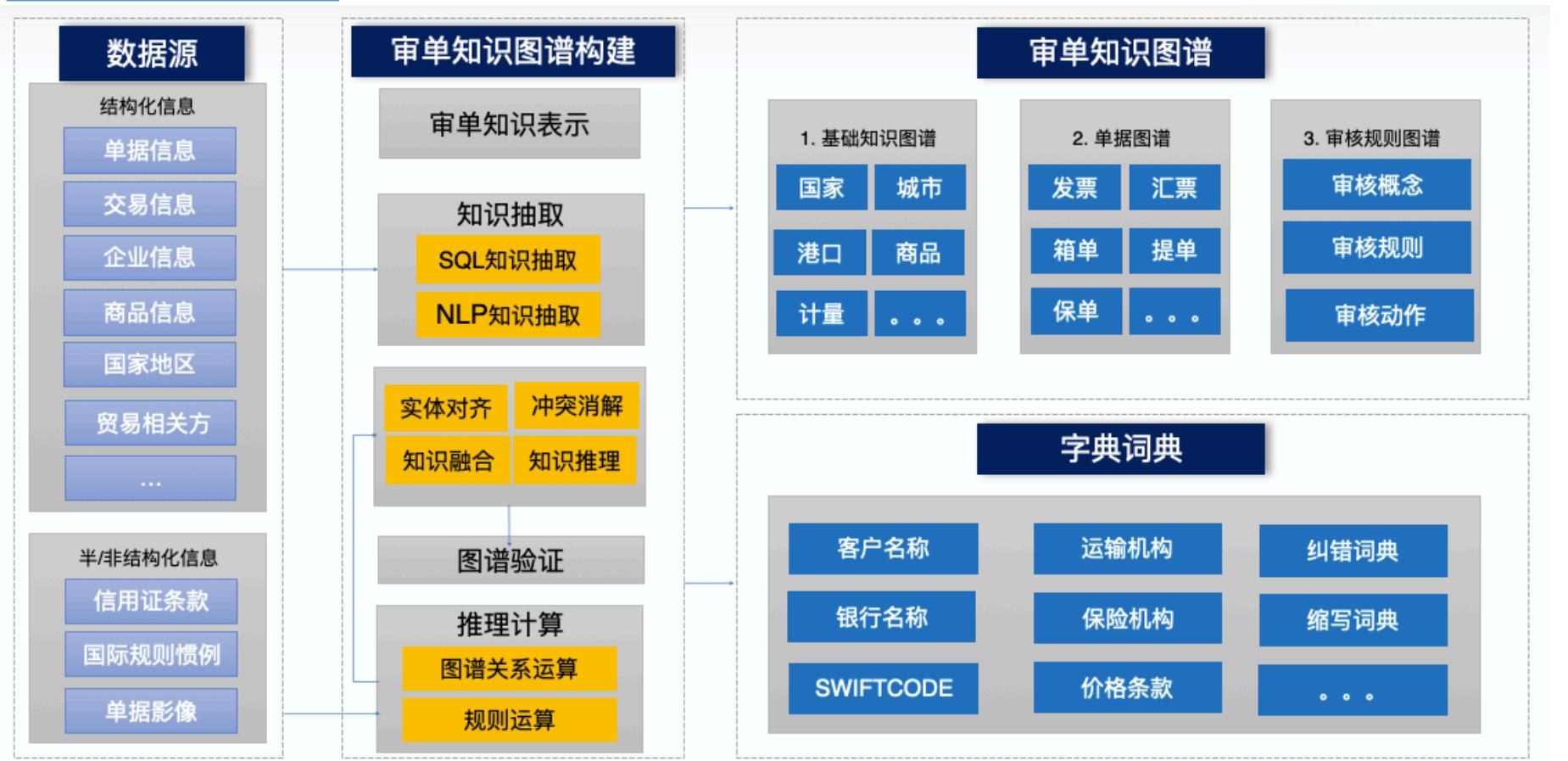
- 使用图谱固化审核规则；
- 使用图遍历方式来读取审核规则；
- 确定的遍历算法、确定的属性应对变化的审核要求。



7月3日-4日

CSDN

知识图谱行业应用——智能审单系统



2020
ProCon
万人开发者大会

7月3日-4日

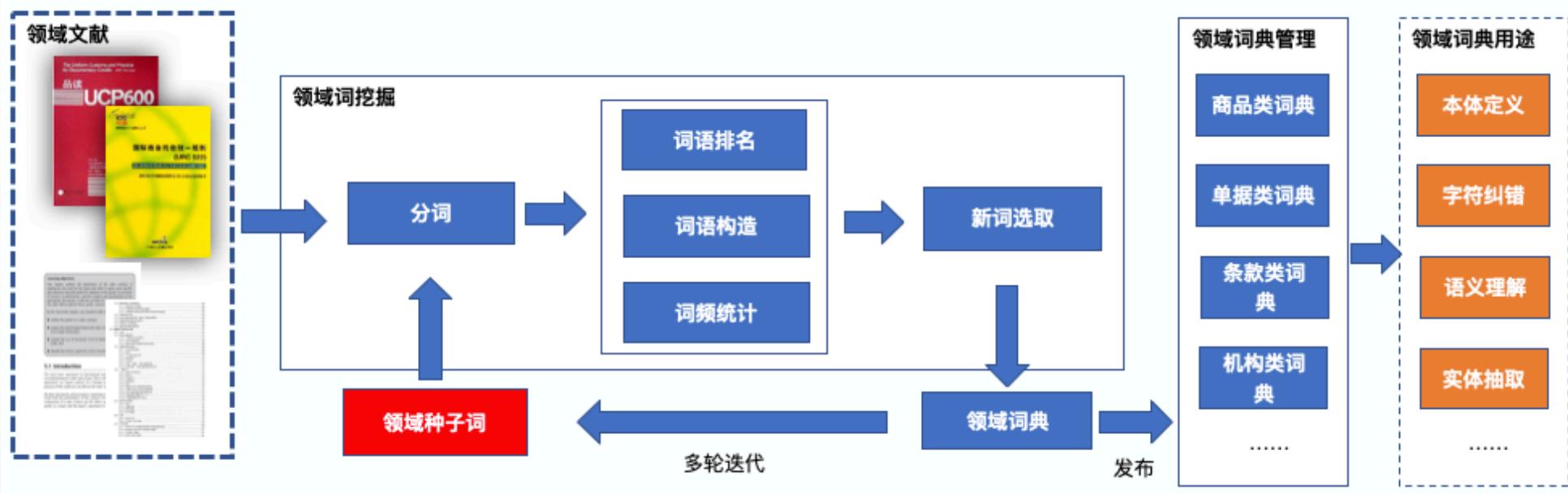
CSDN

知识图谱行业应用——智能审单系统

1. 基础知识图谱构建流程



知识图谱构建和应用过程中，本体定义、字符纠错、知识抽取等多个环节都需要有领域字典和词典的支持。领域无关的词典构建方法，支持从原始语料中有效挖掘高质量领域术语，形成领域词典。

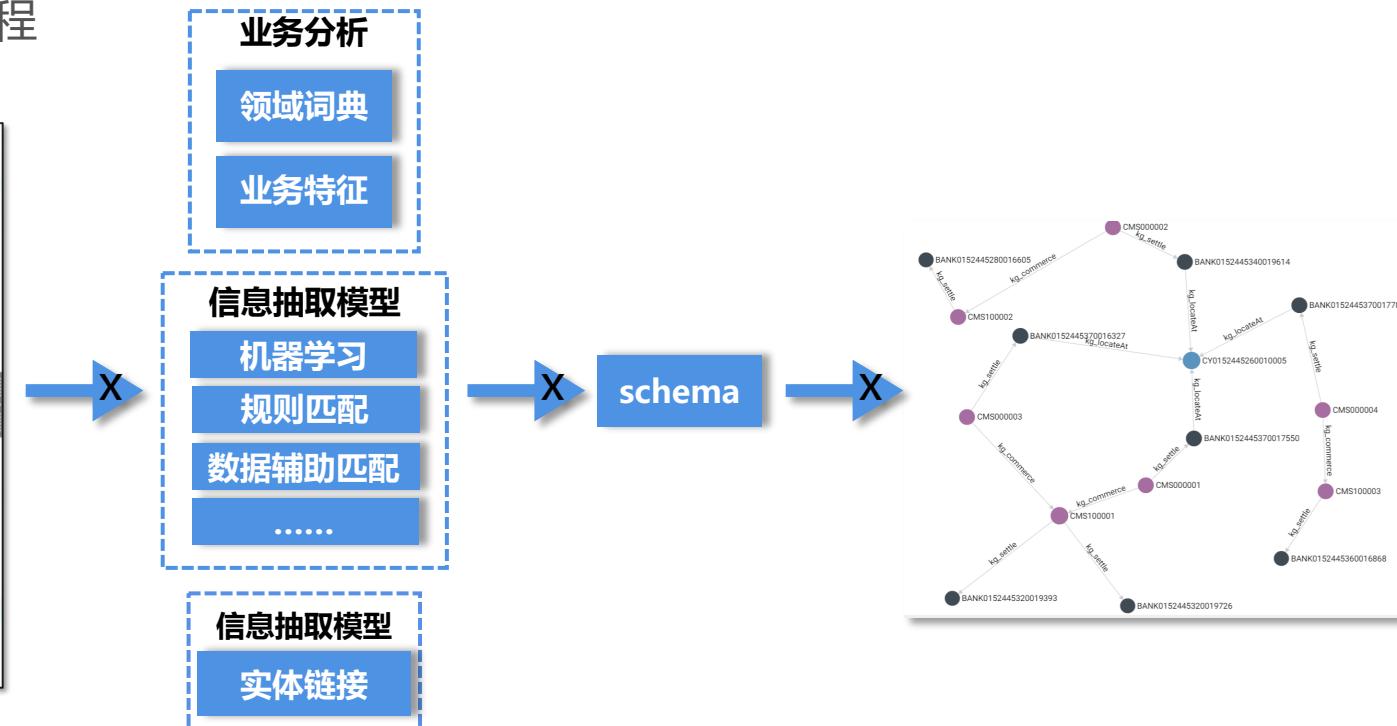
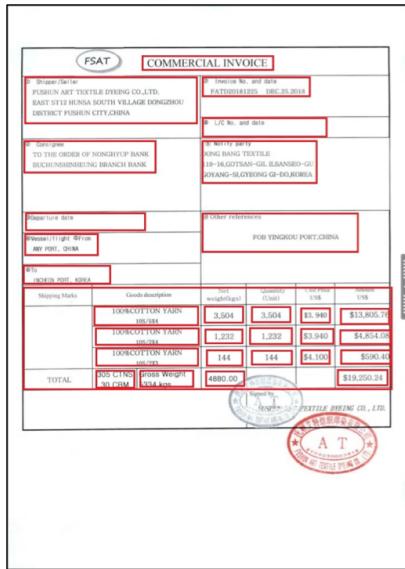


7月3日-4日

CSDN

知识图谱行业应用——智能审单系统

2. 单据知识图谱构建流程



单据知识图谱是智能审单的核心技术，类似于“审核专家”理解“的能力，承担对单据规范、条款的理解，结合单据内容进行单据审核的能力。并可利用知识图谱实例，对NLP能力进行补偿。

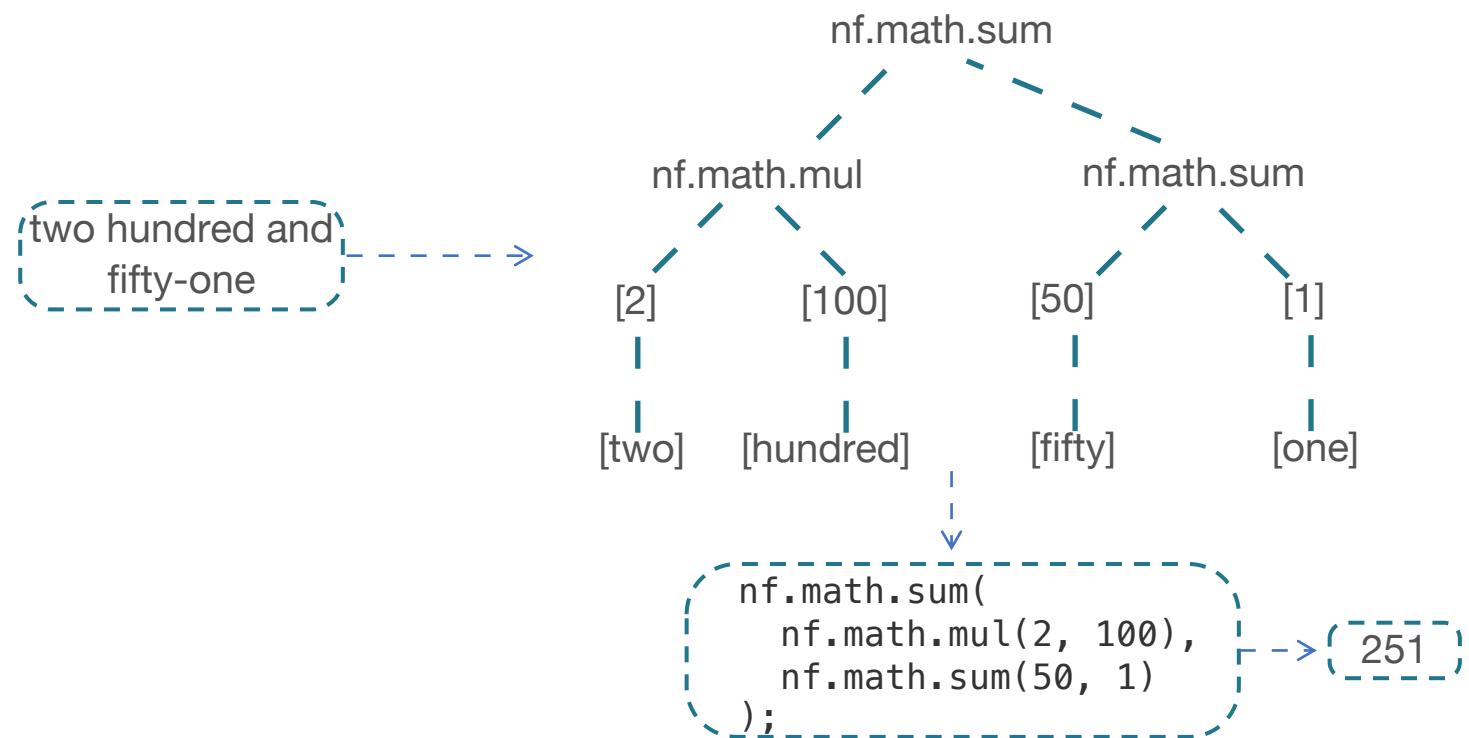


2020
ProCon
万人开发者大会

7月3日-4日

CSDN

知识图谱行业应用——智能审单系统



2020
ProCon
万人开发者大会
7月3日-4日

CSDN

知识图谱行业应用——智能审单系统

中文数字语法

```
1 <cn_unit>="一" {nf.I(1)} | "二" {nf.I(2)} | "九" {nf.I(9)} | "壹" {nf.I(1)} | "貳" {nf.I(2)} ...;
2 <cn_zero>="零" {nf.I(0)} | "〇" {nf.I(0)};
3 <cn_digit>=<cn_unit>{nf.I($1)} <cn_zero>{nf.I($1)} | <digits>{nf.I($1)};
4 <numbers><cn_digit>{nf.I($1)} <cn_digits> <numbers>{nf.util.concat($1, $2)};
5 <cn_e1>="+" {nf.I(10)} | "拾" {nf.I(10)}; <cn_e2>="百" {nf.I(100)} | "佰" {nf.I(100)};
6 <cn_e3>="千" {nf.I(1000)} | "仟" {nf.I(1000)}; <cn_e4>="万" {nf.I(10000)};
7 <cn_e8>="亿" {nf.I(10000000)} | "万万" {nf.I(100000000)};
8
9 <cn_eis>=<cn_e1>{nf.I($1)} | <cn_e1> <cn_unit>{nf.math.sum($1,$2)}
10    | <cn_unit> <cn_e1>{nf.math.mul($1,$2)}
11    | <cn_units> <cn_e1> <cn_units>{nf.math.sum(nf.math.mul($1,$2),$3)}
12    | <cn_units>{nf.I($1)} | <cn_zero>{nf.I($1)};
13 <cn_e2s>=<cn_unit> <cn_e2>{nf.math.mul($1,$2)} | <cn_e1s>{nf.I($1)}
14    | <cn_units> <cn_e1> <cn_units>{nf.math.sum(nf.math.mul($1,$2),nf.math.mul(10,$3))};
15    | <cn_units> <cn_e2> <cn_e1s>{nf.math.sum(nf.math.mul($1, $2), $3)}
16    | <cn_units> <cn_e2> <cn_zero> <cn_e1s>{nf.math.sum(nf.math.mul($1, $2), $4)};
17 <cn_e3s>=<cn_unit> <cn_e3>{nf.math.mul($1,$2)} | <cn_e2s>{nf.I($1)}
18    | <cn_units> <cn_e3> <cn_units>{nf.math.sum(nf.math.mul($1,$2),nf.math.mul(100,$3))};
19    | <cn_units> <cn_e3> <cn_e2s>{nf.math.sum(nf.math.mul($1,$2),$3)}
20    | <cn_units> <cn_e3> <cn_zero> <cn_e2s>{nf.math.sum(nf.math.mul($1,$2),$4)};
21 <cn_e4s>=<cn_e3s> <cn_e4>{nf.math.mul($1,$2)} | <cn_e3s>{nf.I($1)}
22    | <cn_e3s> <cn_e4> <cn_units>{nf.math.sum(nf.math.mul($1,$2),nf.math.mul(1000,$3))};
23    | <cn_e3s> <cn_e4> <cn_e3s>{nf.math.sum(nf.math.mul($1,$2),$3)}
24    | <cn_e3s> <cn_e4> <cn_zero> <cn_e3s>{nf.math.sum(nf.math.mul($1,$2),$4)};
25 <cn_e8s>=<cn_e4s> <cn_e8>{nf.math.mul($1,$2)} | <cn_e4s>{nf.I($1)}
26    | <cn_e4s> <cn_e8> <cn_units>{nf.math.sum(nf.math.mul($1,$2),nf.math.mul(1000000,$3))};
27    | <cn_e4s> <cn_e8> <cn_e4s>{nf.math.sum(nf.math.mul($1,$2),$3)}
28    | <cn_e4s> <cn_e8> <cn_zero> <cn_e4s>{nf.math.sum(nf.math.mul($1,$2),$4)};
29
30 <cn_decimal>="点" <numbers>{nf.math.decimal($2)};
31 <number>=<cn_e8s>{nf.I($1)} | <cn_decimal>{nf.I($1)} | <cn_e8s> <cn_decimal>{nf.math.sum($1,$2)};
32 <number>=<cn_e2>{nf.I($1)} | <cn_e3>{nf.I($1)} | <cn_e4> {nf.I($1)} | <cn_e8>{nf.I($1)};
```

英文数字语法

```
1 <unit>="zero" {nf.I(0)} | "one" {nf.I(1)} | "two" {nf.I(2)} | "three" {nf.I(3)} ...
2 <ten>="ten" {nf.I(10)} | "eleven" {nf.I(11)} | "twelve" {nf.I(12)} ... | "nineteen" {nf.I(19)};
3 <tens>="twenty" {nf.I(20)} | "thirty" {nf.I(30)} | "forty" {nf.I(40)} ... | "ninety" {nf.I(90)};
4
5 <e1>=<unit>{nf.I($1)} | <ten>{nf.I($1)} | <tens>{nf.I($1)};
6    | <tens> <unit>{nf.math.sum($1,$2)}
7    | <tens> "-" <unit>{nf.math.sum($1,$3)};
8 <e2>="hundred" {nf.I(100)}; <e3>="thousand" {nf.I(1000)};
9 <e6>="million" {nf.I(1000000)}; <e9>="billion" {nf.I(1000000000)};
10
11 <e2s>=<e1>{nf.I($1)} | <e2> {nf.I($1)} | <e1> <e2>{nf.math.mul($1,$2)}
12    | <e1> <e2> <e1> {nf.math.sum(nf.math.mul($1,$2),$3)}
13    | <e1> <e2> "and" <e1> {nf.math.sum(nf.math.mul($1,$2),$4)};
14
15 <e3s>=<e2s>{nf.I($1)} | <e3>{nf.I($1)} | <e2s> <e3>{nf.math.mul($1,$2)}
16    | <e2s> <e3> <e2s> {nf.math.sum(nf.math.mul($1,$2),$3)}
17    | <e2s> <e3> "and" <e2s> {nf.math.sum(nf.math.mul($1,$2),$4)};
18
19 <e6s>=<e3s>{nf.I($1)} | <e6>{nf.math.mul($1,$2)}
20    | <e3s> <e6> <e3s> {nf.math.sum(nf.math.mul($1,$2),$3)}
21    | <e3s> <e6> "and" <e3s> {nf.math.sum(nf.math.mul($1,$2),$4)};
22
23 <e9s>=<e6s>{nf.I($1)} | <e9>{nf.I($1)} | <e6s> <e9>{nf.math.mul($1,$2)}
24    | <e6s> <e9> <e6s> {nf.math.sum(nf.math.mul($1,$2),$3)}
25    | <e6s> <e9> "and" <e6s> {nf.math.sum(nf.math.mul($1,$2),$4)};
26
27 <en_decimal> = "point" <numbers> {nf.math.decimal($2)};
28 <en_number> = <e9s> {nf.I($1)};
29
30 <number> = <en_number> {nf.I($1)} | <en_decimal> {nf.I($1)}
31    | <en_number> <en_decimal> {nf.math.sum($1,$2)}
32    | <en_number> "and" <en_decimal> {nf.math.sum($1,$3)};
```



2020
ProCon
万人开发者大会

7月3日-4日

CSDN

知识图谱行业应用——智能审单系统

3. 审核图谱构建：经过专家归纳总结后，预设节点审核动作

核查算子类型	算子	功能	衍生算子类型	算子	功能
二元对比	字符一致性函数	字符完全相同	金额衍生	金额按比例上浮函数	币种不变 金额 * (1 + n%)
	公司一致性函数	公司名称+国家相同		金额按比例下浮函数	币种不变 金额 * (1 - n%)
	日期序列一致性函数	日期按从小到大序, <=统一日期格式 : yyyyMMdd		截止日计算函数	日期 + 天数
	金额一致性函数	币种+金额数字相同	条件算子类型		
	金额序列一致性函数	币种相同, 且金额数字按从小到大序, <=			
	单据要素取值范围函数	单据实体属性上下文中存在, 且值存在于指定范围内			
	商品一致性函数	商品产地、价格条款、商品名称、计量单位、数量、单价及总量相同			
	商品不矛盾函数	商品产地、价格条款相同 商品名称有包含性 总量合计相同		分支表达式条件函数	执行branchTo边上逻辑表达式, 表达式示例： \${Draft.docs} not in ['SIGHT']
	港口地理范围函数	港口位于指定地理区域或港口范围之内			
	机场地理范围函数	机场位于指定地理区域或机场范围之内		是否SWIFT地址	判断输入值是否为SWIFT地址
一元计算	单据要素存在性函数	单据实体属性上下文中存在, 且值非空			

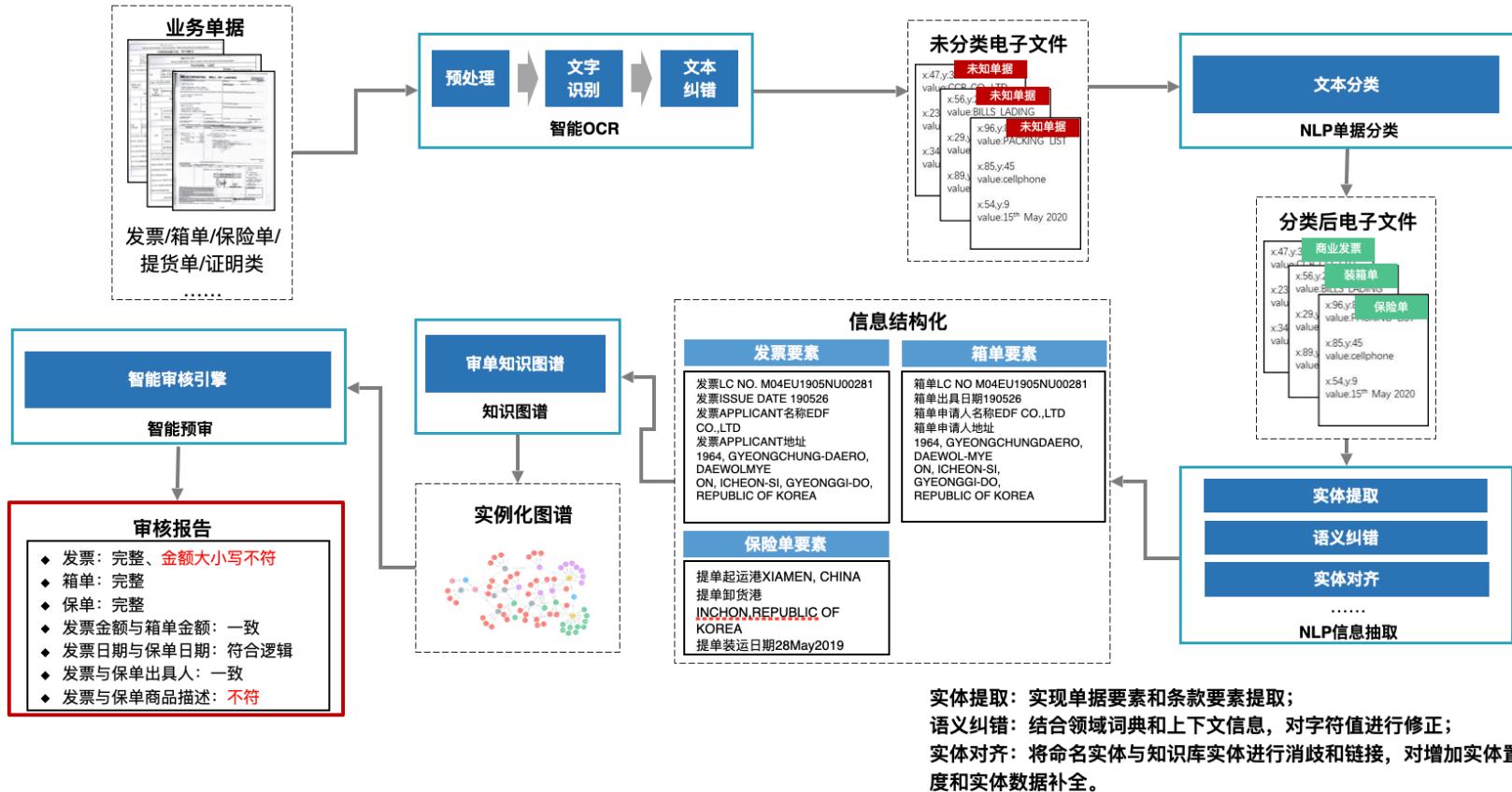


2020
ProCon
万人开发者大会

7月3日-4日

CSDN

知识图谱行业应用——智能审单系统



7月3日-4日

CSDN

知识图谱行业应用——智能审单系统

单据知识标准化定义



单据审核规则统一表示



审单规则引擎集中应用

采用兼容国际标准的本体语言进行单据，如商业发票、汇票、海运提单等各类单据知识本体定义，具有很高的扩展和复用性。可适用于信用证审单、外汇汇出汇款审单等多个业务领域。

业务领域单据知识在单据知识图谱中不断沉淀和统一应用。

针对业务差异性，各领域对单据的关注点和审核逻辑存在不同，提供统一的单据审核规则知识表示框架，支持在单据知识本体和其他知识本体的基础上将审核规则知识显性化。

业务领域审单知识集中存储在审单知识图谱中，进行统一维护和管理。

在单据知识标准化定义和单据审核规则统一表示的基础上，实现各类业务单据的概念推理、关系推定和审核策略的自动化执行，并支持规则触发策略、动作算子等自定义扩展和配置，适配多业务领域特性需求。

业务领域审单规则逻辑集中执行应用。

举例：ISBP745 汇票 Para B9

规则：当信用证仅以银行的SWIFT地址表示汇票付款人时，汇票可以相同的SWIFT地址或该银行的全称显示付款人。

知识表示



实例图谱生成



审单规则执行



2020
ProCon
万人开发者大会

7月3日-4日

CSDN

知识图谱行业应用——智能审单系统



单据分类校验



信息抽取校验



审核结果校验

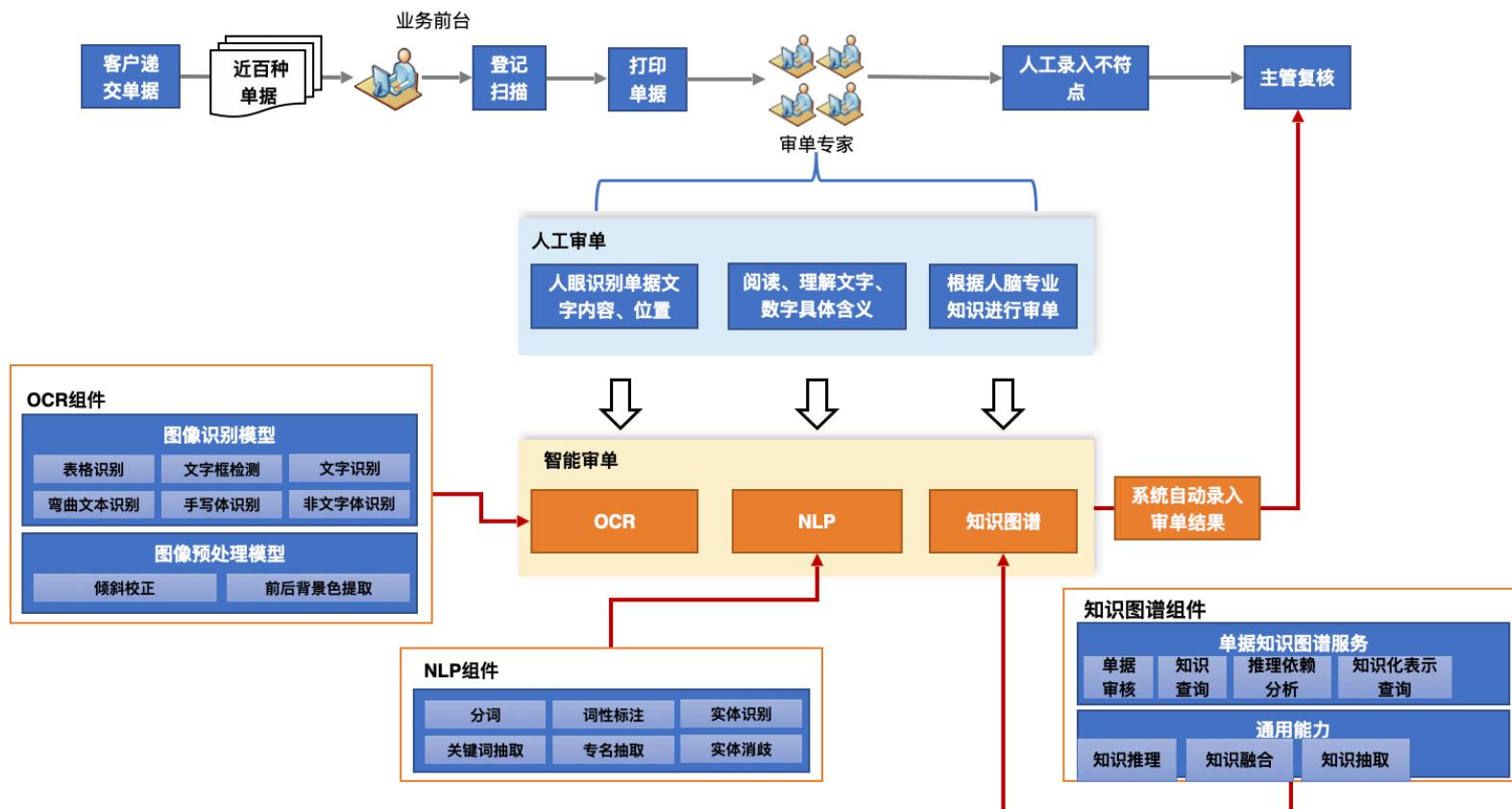


2020
ProCon
万人开发者大会

7月3日-4日

CSDN

知识图谱行业应用——智能审单系统



百万
AI

2020
ProCon
万人开发者大会
7月3日-4日

CSDN

百万人学AI【招聘时刻】——“职”等你来

- 职位名称：NLP算法工程师
- 职位描述：1. 参与核心政务智能信息助手、对话系统等方向的研发，具体工作包括但不限于语言模型、文本匹配、情感分析、知识图谱、问答、对话等；2. 参与自然语言处理和信息检索领域相关的数据抽取、模型建立、模型开发、模型优化、算法评估。
- 职位要求：1. 对自然语言处理开源库/深度学习开源框架非常熟悉，能够快速使用相关工具，熟悉Java/Python/C++/Golang等其中一种或若干种编程语言，熟悉Linux环境，有开源库/开源框架使用经验者优先；2. 计算机及计算机相关专业，硕士及以上学历，985、211高校优先考虑，具有自然语言处理/深度学习/机器学习/数据挖掘/统计分析相关背景；3. 具有较强的学习和研究能力，英语熟练，能够读懂英文技术资料，具有良好的沟通能力，和良好的团队合作精神。
- 简历投递：hr@elensdata.com



7月3日-4日

CSDN