

知识图谱中的推理浅介

目录

什么是推理？

推理的应用

相关技术导引：OWL中的推理

一个人物图谱的例子

什么是推理？

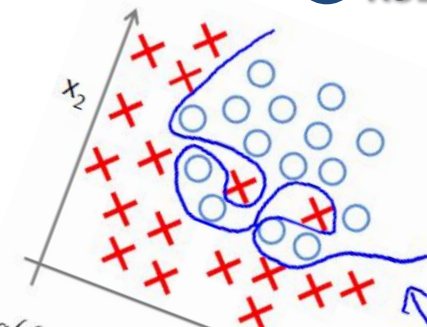
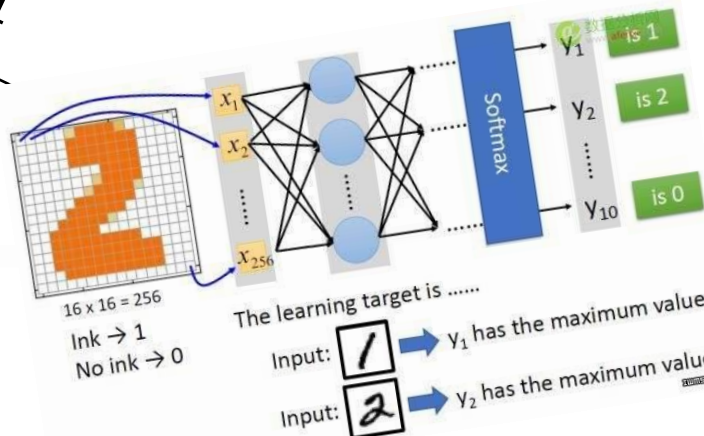


Caffe

今天来听live，
且听得很认真，

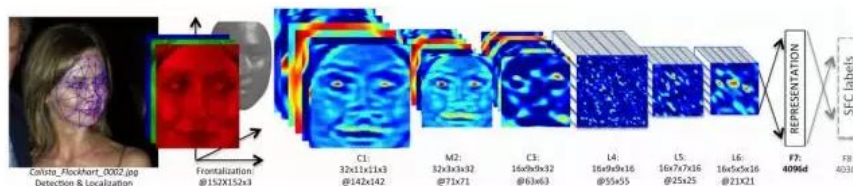
收获太多了！

演绎推理



$$g(\theta_0 + \theta_1 x_1 + \theta_2 x_1^2 + \theta_3 x_1^2 x_2 + \theta_4 x_1^2 x_2^2 + \theta_5 x_1^2 x_2^3 + \theta_6 x_1^3 x_2 + \dots)$$

"Over-fit"



off the
convex path

归纳推理



推理的应用



IBM Watson



无人驾驶

百度 专家系统 深度学习 (Deep Learning) 无人车 修改

如何看待百度无人车，三千多个场景，一万多个if？ 修改

[图片] [图片] 百度无人车的内部人员说，百度无人车，有三千多个场景，一万多个if，巨坑无比，并且还在不断增加中，这不是专家系统吗？这样的无人车能上路吗？大家敢坐吗？...显示全部

direct_declarator

IDENTIFIER

C的文法产生式

|direct_declarator' [' ']'

|direct_declarator ' [' constant_expression ']'

| IDENTIFIER ' (' parameter_list ')'

| IDENTIFIER ' (' ')'

|direct_declarator ' , ' identifier_list

编译器

```
desktop_cleaner.pl - 记事本
文件(F)  编辑(E)  格式(O)  查看(V)  帮助(H)

%
% created at 2015-8-13
% author: Zhangquan Zhou
%
% a prolog system for managing desktop files:
% there are following types of operations:
% 1. save_as
% 2. delete
% 3. move_to
% 4. rename
% 5. make_dir
%
save_as(type=survey) :- file_end_with('pdf'),file_start_with('paper')
delete :- unmove_over('1y')
move_to(dest=statistic) :- (file_end_with('xls') or file_end_with('xlsx'))
move_to(dest=media) :- (file_end_with('mp4') or file_end_with('mkv') or file_end_with('rmvb'))
rename(prefix='tmp') :- file_end_with('txt')

make_dir(dest=tmp_files) :- not_exist(dest=tmp_files)
move_to(dest=tmp_files) :- unmove_over('1y')
```

采用SWI Prolog编写

桌面文件管理系统

推理的应用



条件选股公式编辑器

公式名称: TTIG 密码保护: ☐ 公式类型: 其他类型 确定 取消 另存为

公式描述: 月周选股MACD和均线 公式版本: 0

参数1-4: 参数5-8 | 参数9-12 | 参数13-16

参数	最小	最大	缺省
1			
2			
3			
4			

编辑操作 插入函数 插入资源 引入公式 测试公式

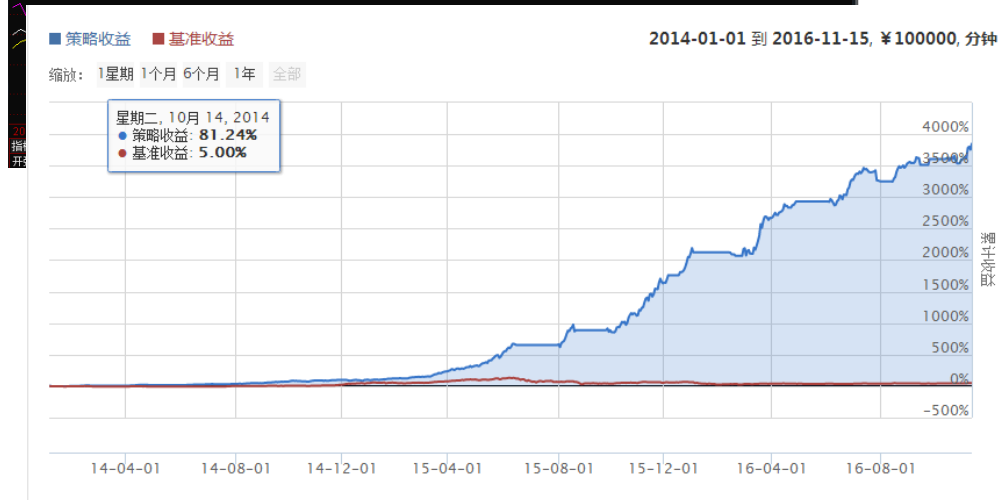
```

MACD:=EMA(CLOSE,12)-EMA(CLOSE,26)-EMA(EMA(CLOSE,12)-EMA(CLOSE,26),9);
MACD>REF(MACD,1)
AND MA(CLOSE,10)>MA(CLOSE,20)
AND MA(CLOSE,5)>MA(CLOSE,20);
    
```

MACD取值: 收盘价的12日指数移动平均-收盘价的26日指数移动平均-收盘价的12日指数移动平均-收盘价的26日指数移动平均的9日指数移动平均
 MACD>1日前的MACD AND 收盘价的10日简单移动平均 > 收盘价的20日简单移动平均 AND 收盘价的5日简单移动平均 > 收盘价的20日简单移动平均

动态翻译 测试结果 参数精灵 用法注释

选股器
(通达信)



100.00	14:59	3282.12	2.12亿
80.00	14:59	3282.14	2.16亿
60.00	14:59	3281.87	1.35亿
40.00	14:59	3282.53	1.85亿
20.00	14:59	3281.80	2.30亿
0.00	15:00	3281.54	9627万
	15:00	3282.74	1.55亿
	15:00	3282.94	2301万
+ - 0	15:01	3281.87	0.0

买 价 细 势 联 值 主 筹

小盘股轮动策略：20条左右规则

两年收益35倍



相关技术导引

■ RDF/OWL

RDF以及OWL是W3C推荐的本体描述框架以及语言。后面例子基于标准的RDF/OWL语法。具体介绍见：<https://www.w3.org/2001/sw/wiki/OWL>。

■ 描述逻辑

描述逻辑 (Description Logic) 是基于对象的知识表示的形式化，也叫概念表示语言或术语逻辑，是本体语言的逻辑基础。具体介绍见[1]。

■ Datalog

Datalog是面向知识库和数据库设计的逻辑语言，表达能力和OWL相当。后面例子中的规则基于Datalog语法。具体介绍见[2]。目前支持RDF/OWL语法以及Datalog推理的代表推理机有OWLim，Jena，RDFox等[3]。

不做深入介绍，感兴趣的朋友欢迎交流讨论。

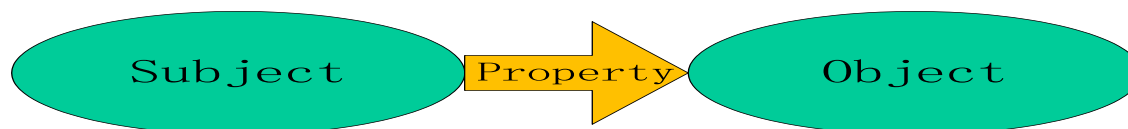


■ RDF/OWL

RDF以及OWL是W3C推荐的本体描述框架以及语言。后面例子基于标准的RDF/OWL语法。具体介绍见：<https://www.w3.org/2001/sw/wiki/OWL>。

■ 语法：

RDF/XML语法，三元组；



■ 逻辑基础：描述逻辑

描述逻辑 (Description Logic) 是基于对象的知识表示的形式化，也叫概念表示语言或术语逻辑。是一阶谓词逻辑的一个可判定子集。具体介绍见[1]。

因为本讲关注逻辑推理，因此我们聚焦于OWL的逻辑基础，即描述逻辑。

分类(Classification)

分类：针对Tbox的推理，计算新的概念包含关系。

$\text{Mother} \sqsubseteq \text{Women}$ $\text{Women} \sqsubseteq \text{Person}$ \longrightarrow $\text{Mother} \sqsubseteq \text{Person}$

注意：这里的分类和机器学习中分类的区别。

分类的例子

苹果由富达和黑石投资。

$\text{Apple} \sqsubseteq \exists \text{beInvestedBy.}(\text{Fidelity} \sqcap \text{BlackStone})$

借助富达融资的公司都是创新企业。

$\exists \text{beFundedBy.Fidelity} \sqsubseteq \text{InnovativeCompanies}$

借助黑石融资的公司都是创新企业。

$\exists \text{beFundedBy.BlackStone} \sqsubseteq \text{InnovativeCompanies}$

$\text{beInvestedBy} \sqsubseteq \text{beFundedBy}$

投资即是帮助融资。

$\text{Apple} \sqsubseteq \exists \text{beInvestedBy.Fidelity}$

苹果由富达投资。

$\text{Apple} \sqsubseteq \exists \text{beFundedBy.Fidelity}$

苹果由黑石投资。

$\text{Apple} \sqsubseteq \text{InnovativeCompanies}$

苹果是创新企业。

富达基金和黑石基金都喜欢投资高新科技公司。

物化(materialization)

物化即计算属于某个概念或关系的所有实例的集合。在知识库中检测实例，即检测**C^I**包含于**D^I**是否在所有的解释中成立；在Tbox中检测，即检测**C^I**包含于**D^I**是否在Tbox T中的所有解释中成立。

Mother(Helen)
Mother \sqsubseteq Women



Women(Helen)

has_son(Helen,Jack)
has_son \sqsubseteq has_child



has_child(Helen,Jack)

物化(materialization)的一个例子：

一个兼并重组（可以是业务兼并，不是收购）套利策略：

策略思想：与大盘股公司兼并重组的上市企业有很高的预期收益。

$\exists \text{ merge.BigCapital} \sqsubseteq \text{ValueSecurity}$

定义什么是大盘股（按照策略自己调整）：

上证50和沪深300指数中的标的属于大盘股。

$\text{SZ50} \sqsubseteq \text{BigCapital}, \text{HS300} \sqsubseteq \text{BigCapital}, \text{SZ180} \sqsubseteq \text{HS300}$

选股目标：找出兼并重组策略下所有高预期公司：

使用OWL的物化推理

物化(materialization) 一个例子：

一个兼并重组套利策略：

$\exists \text{ merge.BigCapital} \sqsubseteq \text{ValueSecurity}$

$\text{SZ50} \sqsubseteq \text{BigCapital}, \text{HS300} \sqsubseteq \text{BigCapital}, \text{SZ180} \sqsubseteq \text{HS300}$

选股目标: 找出兼并重组策略下所有高预期公司：

$\text{merge}(\text{SZ300377}, \text{SH600570})$ 赢时胜和恒生电子在区块链方面有业务兼并
 $\text{SZ180}(\text{SH600570})$ 恒生电子是上证180的成分股

推理：

$\text{HS300}(\text{SH600570}), \text{BigCapital}(\text{SH600570}), \text{ValueSecurity}(\text{SZ300377})$

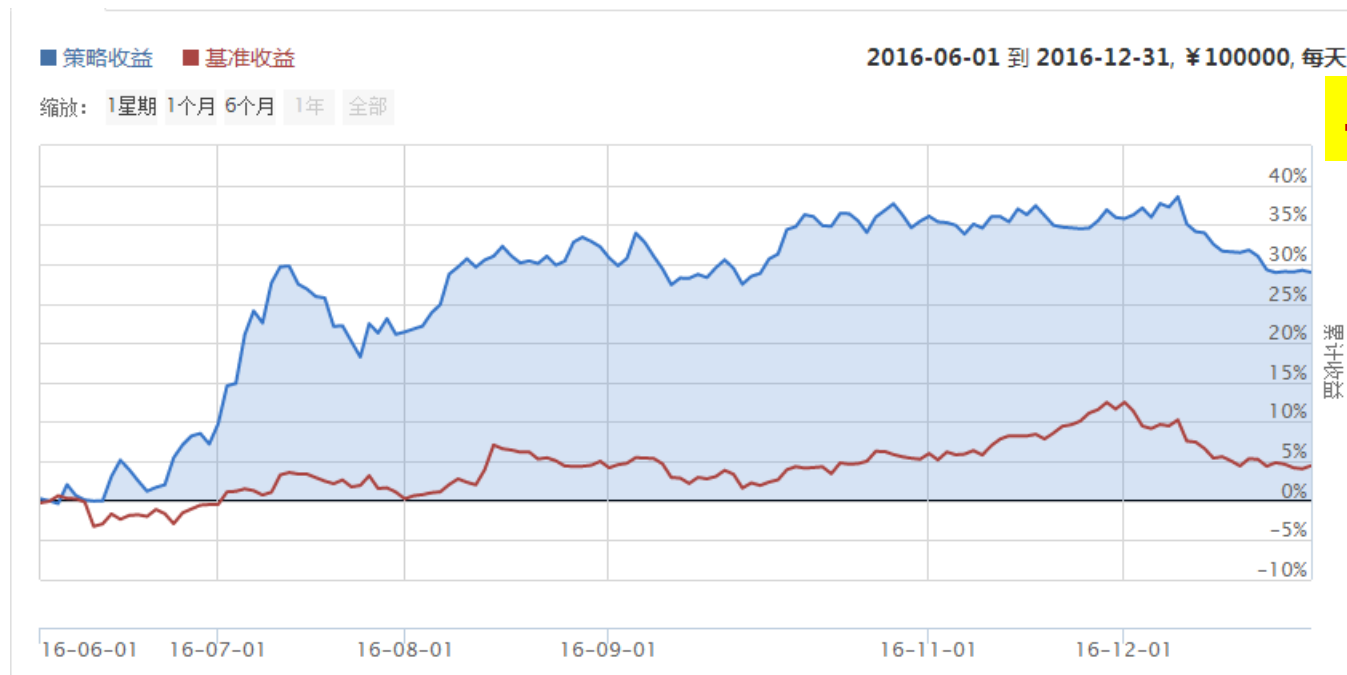
结论：SZ300377赢时胜在短期内是一家高收益公司。
这本质上用基于消息面的套利，推理机可以完成复杂股票筛选的过程。

OWL本体的推理



物化(materialization) 一个例子：

推理机选股后立马建仓，效果如何？



半年25%收益

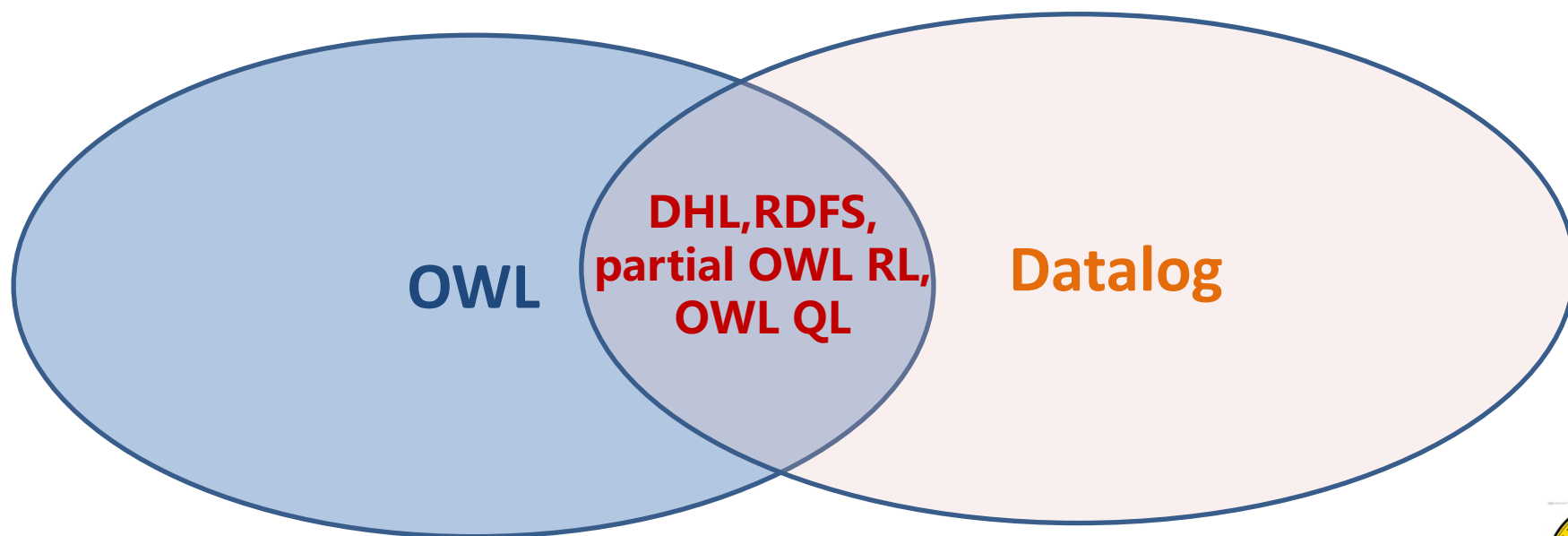
推理任务可以更加复杂和精确。

可以想象：证券分析师几天的分析结果，推理机只需要几秒钟就得到。



Datalog语言

- 面向知识库和数据库设计的逻辑语言，表达能力和OWL相当。具体介绍见[2]。
- 能够方便地撰写规则，实现推理。
- 有一些工具已经实现了OWL和Datalog的统一。目前支持RDF/OWL语法以及Datalog推理的代表推理机有OWLim，Jena，RDFox等[3]。



Datalog推理的例子：

一个兼并重组套利策略：

ValueSecurity(x) :- merge(x,y),BigCapital(y)

SZ50(x) :- BigCapital(x)

HS300(x) :- BigCapital(x)

SZ180(x) :- HS300(x)

规则集

merge(SZ300377, SH600570)

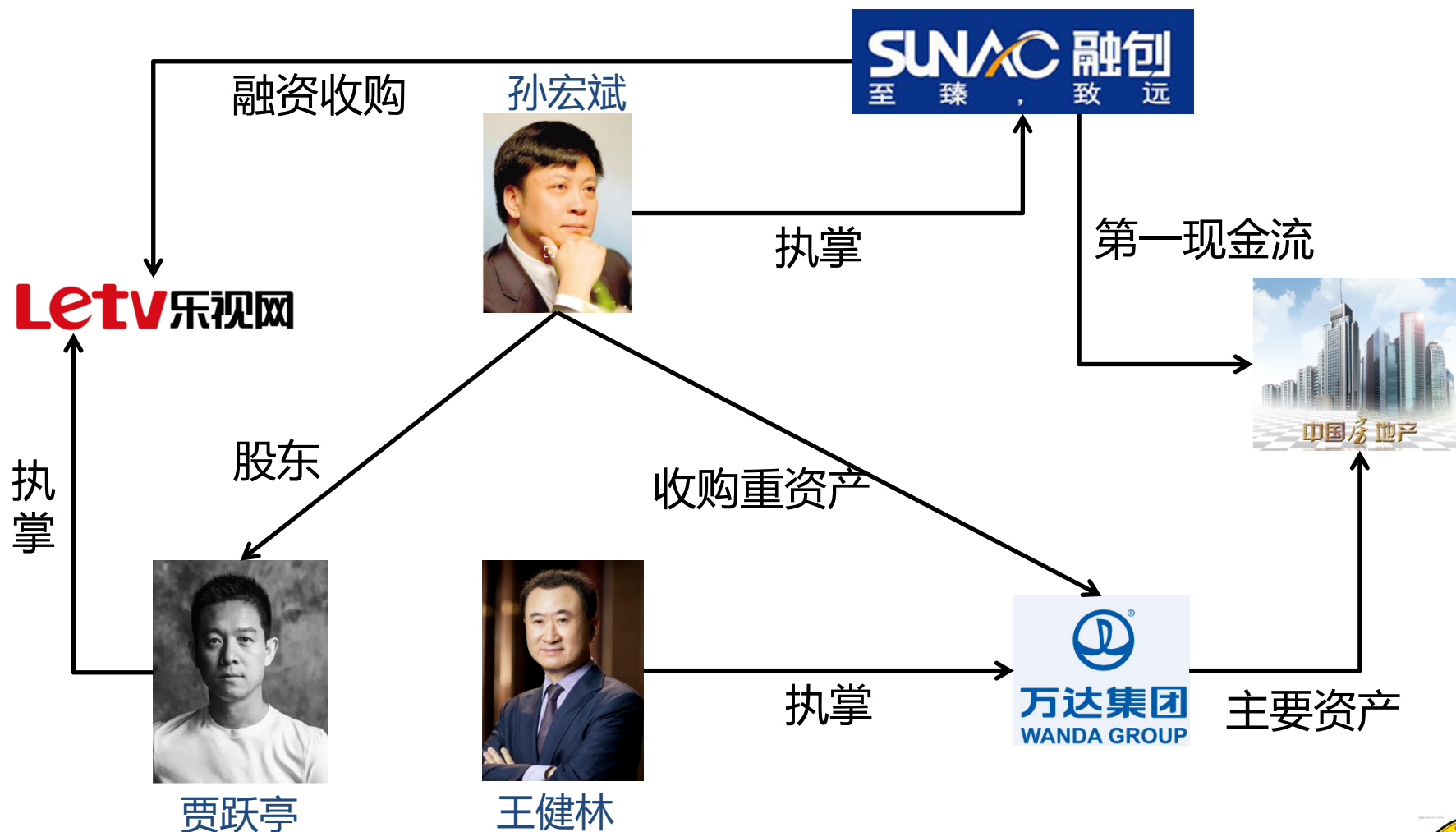
SZ180(SH600570)

事实集

推理结果

HS300(SH600570), BigCapital(SH600570), ValueSecurity(SZ300377)

人物知识图谱上的推理和查询



构建本体：



<http://www.example.org/finance/kse#孙宏斌>

URI

finance:孙宏斌

命名空间, namespace



<http://www.example.org/finance/kse#融创中国>

finance :融创中国



<http://www.example.org/finance/kse#地产事业>

finance :地产事业

三元组

finance :融创中国 rdf:type finance :地产事业

本体形式：

finance :融创中国 rdf:type finance :地产事业

finance :孙宏斌 finance :control finance :融创中国

finance :贾跃亭 finance :control finance :乐视网

finance :孙宏斌 finance :hold_share finance :乐视网

finance :王健林 finance:control finance :万达集团

finance :万达集团 finance :main_income finance :地产事业

finance :融创中国 finance:acquire finance :乐视网

finance :融创中国 finance:acquire finance :万达集团

撰写规则：

两条自定义Horn推理规则：

- 1) 执掌一家公司就一定是这家公司的股东；
- 2) 某人同时是两家公司的股东，那么这两家公司一定有关联交易；

Datalog推理规则

```
hold_share(X, Y) : - control(X, Y)
conn_trans(Y,Z) : - hold_share(X, Y), hold_share(X, Z)
```

本体语境下的推理规则

```
finance:hold_share(X, Y) : - finance:control(X, Y)
finance:conn_trans(Y,Z) : - finance:hold_share(X, Y), finance:hold_share(X, Z)
```

为了方便起见，我们还是采用上面的形式。

推理：

finance :融创中国 rdf:type finance :地产事业
finance :孙宏斌 finance :control finance :融创中国
finance :贾跃亭 finance :control finance :乐视网
finance :孙宏斌 finance :hold_share finance :乐视网
finance :王健林 finance :control finance :万达集团
finance :万达集团 finance :main_income finance :地产事业
finance :融创中国 finance :acquire finance :乐视网
finance :融创中国 finance :acquire finance :万达集团

hold_share(X, Y) : - control(X, Y)
conn_trans(Y,Z) : - hold_share(X, Y), hold_share(X, Z)

hold_share(孙宏斌, 融创中国)
conn_trans(乐视网, 融创中国)

查询：**SPARQL: 本体的SQL语言。**

finance :融创中国 rdf:type finance :地产事业
finance :孙宏斌 finance :control finance :融创中国
finance :贾跃亭 finance :control finance :乐视网
finance :孙宏斌 finance :hold_share finance :乐视网
finance :王健林 finance :control finance :万达集团
finance :万达集团 finance :main_income finance :地产事业
finance :融创中国 finance :acquire finance :乐视网
finance :融创中国 finance :acquire finance :万达集团

SELECT ?P ?X

WHERE {

?P finance:control ?c .

?c finance:acquire ?X .

}

查询结果

? P	?X
孙宏斌	乐视网
孙宏斌	万达集团

参考文献

- [1] Baader, F.; Calvanese, D.; McGuinness, D. L.; Nardi, D. & Patel-Schneider, P. F. The Description Logic Handbook: Theory, Implementation and Applications. *Cambridge University Press*, **2003**.

- [2] Abiteboul, S.; Hull, R. & Vianu, V. Foundations of Databases. *Addison-Wesley*, **1995**.

- [3] Motik, B.; Nenov, Y.; Piro, R.; Horrocks, I. & Olteanu, D. Parallel Materialisation of Datalog Programs in Centralised, Main-Memory RDF Systems. In *Proc. of AAAI*, **2014**, 129-137.