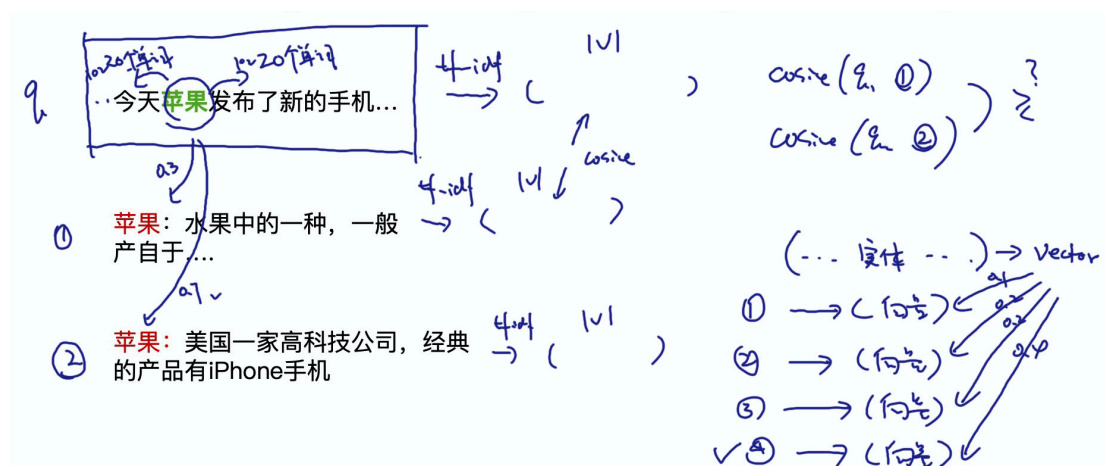
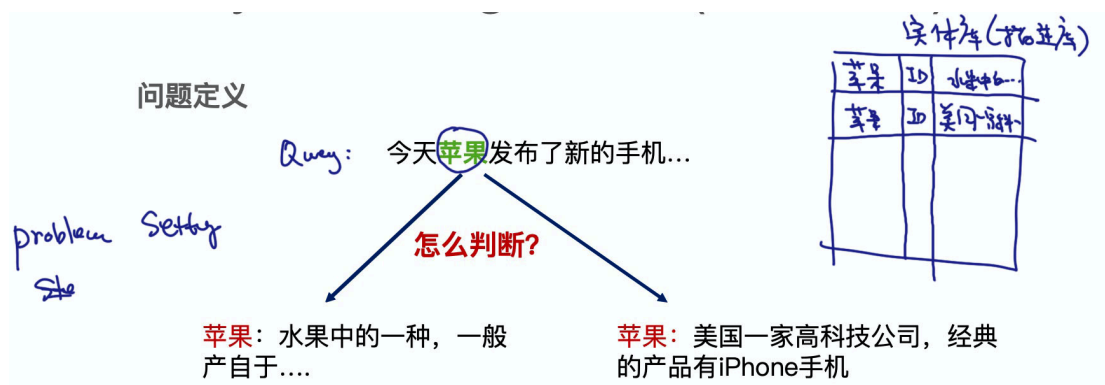


## 1. Entity disambiguation 实体消歧

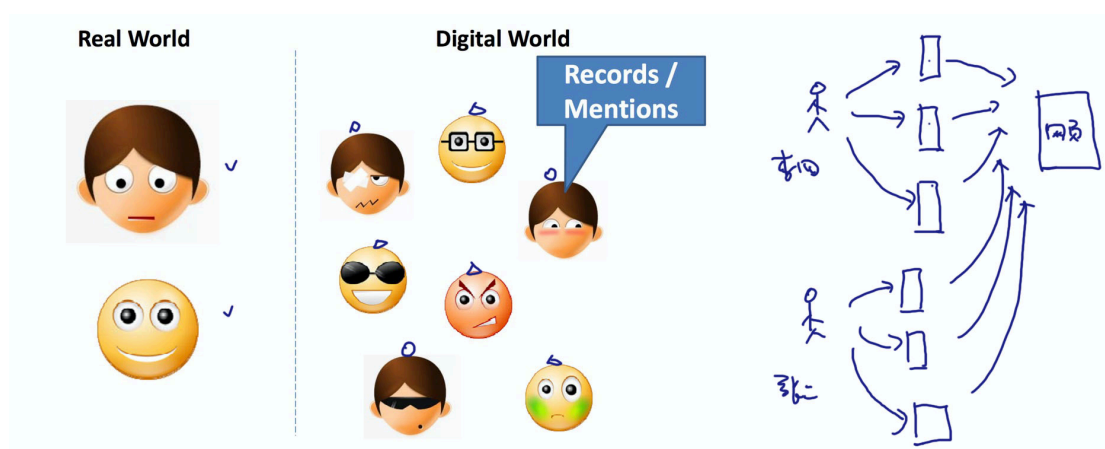
本质：一个词在不同的上下文中所表达的含义不太一样。

解决方法：找两个词的相似度。

问题：如何转换成向量。



## 2. entity resolution 实体统一



### 2.1 算法

第一种：

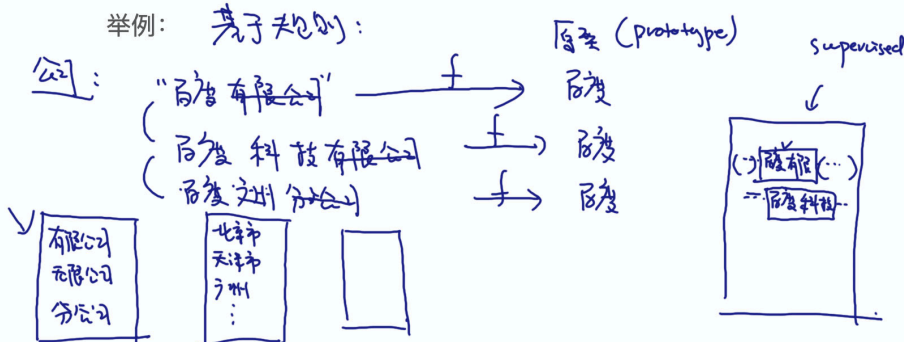
问题定义：给定两个实体，判断是否是指向同一个实体？

举例：给定两个实体，字符串  $str_1, str_2$   
同候体：

第一种方法： $\text{sim}(\text{str}_1, \text{str}_2)$   
Edit Distance

第二种：

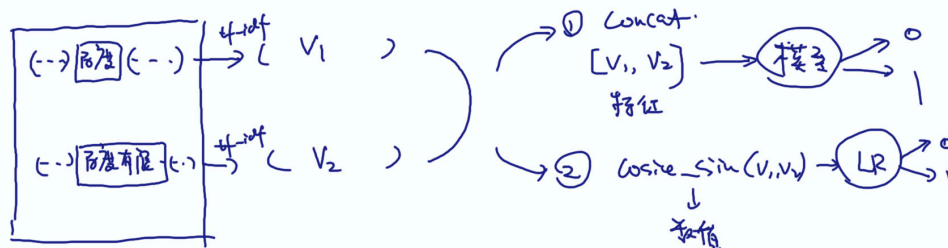
举例：基于规则：



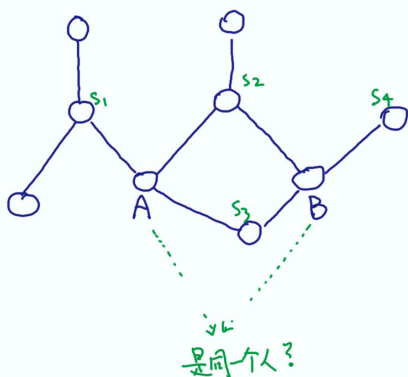
第三种：

③ supervised

新特征： $\text{EditDist}(\text{百度}, \text{百度有限})$



## 2.2 基于图的实体统一



$V_A = (\text{age, salary, job} \dots \dots \text{3, } S_1, S_2, S_3 \dots \dots)$   
个体特征 关系相关特征

$V_B = (\text{age, salary, job} \dots \dots \text{3, } S_2, S_3, S_4 \dots \dots)$   
个体特征

$\text{Sim}(V_A, V_B) \geq \text{Threshold} = 0.3/0.4$

if yes: merge

otherwise: keep it!

### 3. co-reference resolution 指代消解

( ) 张三 (没有去上班), 因为生病了。昨天, 李四陪他去了医院。  
好消息是, 他已经好了。

① 最简单的方法

② Supervised  
learning

(i) Collect Data

(i)  $(张三, A) \rightarrow 1$   
 $(李四, A) \rightarrow 0$   
 $(张三, B) \rightarrow 1$   
 $(李四, B) \rightarrow 0$  } 样本

考虑  $(张三, A) \rightarrow 1$

↓ ↓

( 向量 )  $\xrightarrow{f}$  1