一、问题：① 4个组件指代ln部分分别是什么

② WordNet semantic tree 如何构建

③ 3.2中为Si与Sij比较还是为规化

二、方法：

1° 根据相似度建立一个候选集C（string上ln相似度）

Ci in $O_k^{(m)}$, Cj in $O_l^{(n)}$

思路：

① 计算每一个Ci与Cj的相似度 Sij（m×n个）

② * Si指Ci与O_l in相似度，且若 Sij > α，则

$$S_i = \sum_{j=0}^{n} S_{ij}$$

③ 若 $S_i > \beta$，则 Ci插入候选集C

相似度ln求法：edit-distance based 和 word-net based 2种方法结合。

求concepts之间ln相似度：

① edit-distance based

|sopk3| 指词wi转化为wj的操作次数

|wi| 词长

$$e\_d(w_i, w_j) = \frac{|sopk3|}{\max(|w_i|, |w_j|)}$$

$$S_{ij} = \frac{1}{1 + e\_d(i,j)}$$

② word_Net based（word_Net semantic tree）

V：表示vi和vj的第一个共同祖先

vi, vj 指 Wordi, Wordj 在 WordNet tree in 节点

count(vi) 指以vi为根节点的子树中节点的个数

total 整棵树的总节点数

$$p(v_i) = \frac{count(v_i)}{total}$$

$$S_{ij} = \frac{2 \times \log p(v)}{\log p(v_i) + \log(v_j)}$$

若①或②方法中有 Sij=1，则 Sij=1；否则 $S_{ij} = \frac{① + ②}{2}$

求concept与O_l in相似度：

求Ci与O_l in similarity 人 $S_i = \sum_{j=0}^{n} S_{ij}$；最后若 $S_i > \beta$，则Ci放入候选集C

若 Sij > α，则

将Si正则化，即除以 max(Si)

2° 基于候选集C，根据相关性构造本体，（语义上的相似度）指concepts对O_l

relevance和influence相统一：

相关性度两个因素ln影响，前文ln Si和influence；influence 指concepts展现ln是Ci的邻近结点对O_l的 相似度和

Influence会随着distance↗而↘，符合高斯函数ln特性。（从图知x离(0,0)↗，则f(x)↘）

构建本体：

① |Ci-Cj| 指Ci与Cj in 距离，即最短路径（在本体图中）

Cj是Ci的一个邻近点，⇒若 |Ci-Cj| ≤ 2：

$$\varphi_j(C_i) = S_j \times e^{-|C_i - C_j|^2} \quad (i \neq j)$$

（Ci, Cj 均属于候选集C）

$$\therefore \varphi(C_i) = \sum \varphi_j(C_i)$$

② Ci对O_l的 relevance 为：

$$r_i = S_i \times \varphi(C_i)$$

③ 若 $r_i < \gamma$，则将Ci从C中移除

此时得到本体所需的concepts

④ 将C中剩下的concepts构成有向连接图。

{gij}，且gij是O_h的极大连通子图。(sub-graph)

D(Ol) 表示 O_l 的degree

$$D(O_l) = \frac{\text{边}}{\text{点}}$$

Dgij 表示每一个gij顶点的度=出+入度

|D(Ol) - Dgij| > T，则去掉gi的顶点

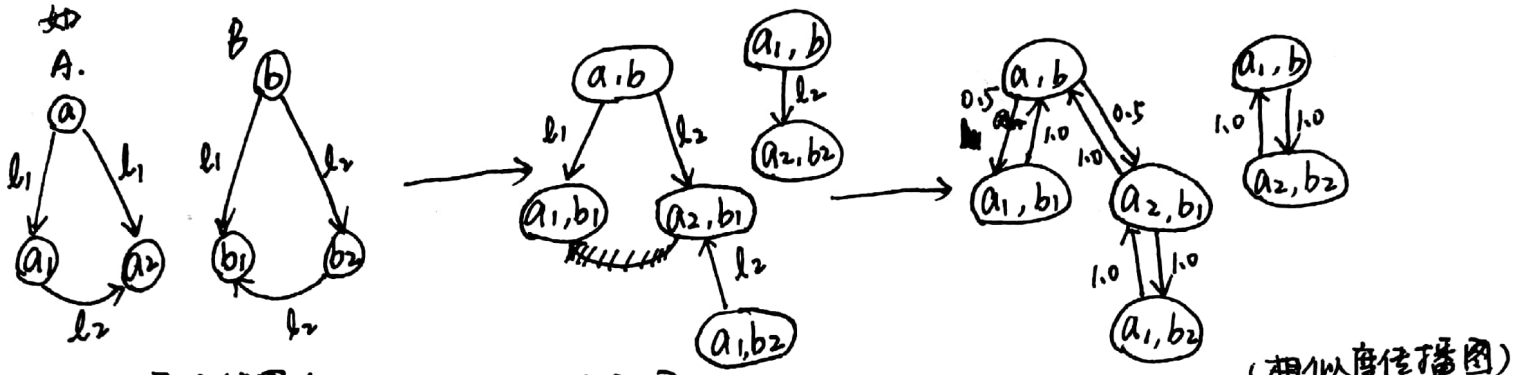
∵degree差别大，图的结构相差也大

即不相似（或↓）

剩下的图gij为子本体。

3° 匹配 O_1 和子本体 ⇒ 用 similarity flooding 方法

① 将两个本体构建一个有向图，即图 2.



原本体图 1　　连通图 2　　(相似度传播图) Induced propagation 图 3

l_1, l_2 指 relation，若本体 A 和本体 B 的点间拥有同样的 relation，则变成图 2

② 对于图 2 的每一条边 $\langle v_i, v_j \rangle$ 都有权重 w_{ij}，且 $w_{ij} = \frac{1}{v_i \text{出度}}$

③ s_i^0 表示顶点 v_i 的初始相似度，即第 0 次迭代的相似度.

$\sum_j s_j^n \times w_{ij}$ 指 v_i 的直系邻点的第 n 次迭代相似度 × 权重 (从 v_i 离开)

$\sum_j s_j^n \times w_{ji}$ 指 所有 (进入 v_i)

$\therefore s_i^{n+1} = s_i^n + \sum_j s_j^n \times w_{ij} + \sum_j s_j^n \times w_{ji}$

直到 $s_i^{n+1} - s_i^n \leq \alpha$，则输出 $M(O_1, O_n)$

∵ 有可能 $s_i^{n+1} - s_i^n > \alpha$，则去掉