

12  
 ??<sup>3</sup>  
 Wikipedia.org1  
 Facebook.comAmazon.com  
 Youtube.comTwitter.com  
 qq.com  
 Taobao.com

4  
 ??  
 5  
 Big<sup>67</sup>  
 Linked  
 Noisy  
 Unstructured<sup>8</sup>  
 Incomplete

9  
 10  
 11  
 12  
 1314  
 15  
 $(e_i, a_{ij}, s_{ijkl}, h_k, t_l)$   
 $e_i a_{ij} h_k t_l s_{ijkl}$   
 1  
 2  
 3  
 ??  
 f6  
 1718  
 ??

$$SD = \sum_{w \in D} S_w * weight(w) * modifier(w) \sum weight(w)$$

- (1)  
 $S_w weight(w) modifier(w)$   
 $PMI(x, y) = \log_2 F(x, y) F(x) F(y)$   
 (2)  
 $F(x, y) F(x) x x$

$$PMIIR(x) = \sum_{p \in pWords} PMI(x, p) - \sum_{n \in nWords} PMI(x, n)$$

(3)  
 $pWords nWords$   
 $DS\{(d, s) | d \in$   
 $D, s \in$   
 $S\}$

- $g : D \rightarrow S, g(d) = \arg \max_S f(d, s)$   
 (4)  
 $g f f f$   
 ??  
 $D$   
 $\{(f, s_f) | rep(f, D) > \rho_f, s_f = agg(S, f)\}$   
 (5)  
 $frep(f, D) s_f agg(S, f) f$   
 ??  
 ??  
 ??  
 ??

