

(سوال 1)

(الف)

Pioneer:

'bitcoin': 0.016491403697765646, 'self': 0.016580020985077142, 'focus':  
0.019457795296241212, 'pioneer': 0.02674949362692034, 'mine': 0.029204534890197222,  
'cnw': 0.02924393790586201, 'toronto': 0.029256251094770664, '388toronto':  
0.029278969421280658, 'corp': 0.02929313930440453, 'hut': 0.029444921236442254

Trade:

'avax': 0.020548397787744335, '4th': 0.020762630076087113, 'pm': 0.02076411511026999,  
'ecash': 0.021421106194513387, 'xec': 0.021421106194513387, 'today':  
0.02159381076848921, 'dollar': 0.02269240730921413, 'write': 0.023483389394908537,  
'largest': 0.023950539071871023, 'trade': 0.03316580067651429

(ب)

Pioneer:

'tsx': 0.027953508639571426, 'self': 0.034362825851774526, 'focus': 0.0399537674961746,  
'pioneer': 0.05388411236716642, 'mine': 0.056953123914254694, 'toronto':  
0.06059402664128192, '388toronto': 0.0624733253174246, 'cnw': 0.0628794176925582,  
'corp': 0.06354758097133309, 'hut': 0.06409520834987868

Trade:

'higher': 0.02806502632191169, 'lead': 0.02846366841682767, 'rang': 0.029032145422186375,  
'morn': 0.02907388602732166, 'pm': 0.029662097209626676, 'eastern':  
0.030132190045250307, 'narrow': 0.03214132905543969, 'today': 0.032645816333121594,  
'4th': 0.033176650978054154, 'dollar': 0.0343001554860945, 'largest': 0.03597153674960071,  
'write': 0.03827838185521724, 'xec': 0.03889999986000717, 'ecash': 0.03889999986000717,  
'trade': 0.055969227539940186

(پ) نتایج خیلی فرق نکرد. احتمالاً در مقایسه هر یک از کلمات ذکر شده با یک کلمه عام، خیلی با هم در اسناد مشابه دیده نشده اند. از طرفی در مرحله پیش پردازش، stop word ها را حذف کرده ایم و ریشه کلمات را در مجموعه واژگان قرار داده ایم.

Trade:

dollar': 0.05552084642300459, 'coin': 0.06048329525197056, 'cryptocurr': '  
0.06545388904107328, 'pm': 0.06894911037882615, 'worth': 0.0731495819516308, 'day':  
0.07336183337639311, 'here': 0.0745873421366863, 'market': 0.07799486593130288, 'last':  
0.08176580006033166, 'exchang': 0.08630537039079036, 'trade': 0.6074119918520828

Pioneer:

nasdaq': 0.09246268554573162, 'comput': 0.0964785952304399, 'bring': '  
0.09949202076055363, 'infrastructur': 0.10112304770499853, 'innov': 0.10565565883144357,  
'cnw': 0.11158775383699637, 'focus': 0.11167071766576714, 'america':  
0.11424425903051019, 'self': 0.11424425903051019, 'corp': 0.11563582743956882, 'highlight':  
0.11659554172604235, '388toronto': 0.11733009467231073, 'tsx': 0.11951242085412335,  
'hut': 0.1253190688161852, 'pioneer': 0.14969963504130485

در جانشینی کلماتی در رتبه بالا آمده اند که در موقعیت مشابه احتمالا بجای trade و pioneer می توانند در جمله قرار گیرند (مانند dog و cat در اسلاید که ربطی به هم ندارند به صورت تکی اما در جملات مختلف بجای هم می توانند قرار گیرند). اما در هم نشینی کلماتی در رتبه بالا قرار میگیرند که از نظر موضوعی به کلمه های مدنظر (در اینجا trade و pioneer) نزدیک اند (مثلا exchang و market و worth که در بالا و برای هم نشینی برای trade در رتبه بالا قرار گرفته است)

Postings list for (" information ")

1111111110111100110/100/11111000000

$$(1111100110)_b = 998 \quad (10)_b = 2 \quad (100000)_b = 32$$

Inverted list for " information " =  $d_{998}, d_{1000}, d_{1032}$

Posting list for " retrieval "

111111111100000000111/0/111101111

0 -> 1

$$(10000000111)_b = 1031 \quad (11111)_b = 31$$

Inverted list of " retrieval " =  $d_{1031}, d_{1032}, d_{1063}$

( الف )

q = information AND NOT retrieval

Inverted list for " information " - Inverted list of " retrieval " =

$$[998, 1000, 1032] - [1031, 1032, 1063] = [998, 1000]$$

( ب )

$$\delta_{code}(998) = 1110001111100110$$

$$\delta_{code}(1000) = 1110001111101000$$

$$\delta_{code}(1032) = 11100100000001000$$

$$\delta_{code}(1031) = 11100100000000111$$

$$\delta_{code}(1063) = 11100100000100111$$

سوال 3

بله برقرار است.

$$S(Q, D) = \left[ \sum_{\substack{W \in Q \\ W \in D}} C(W, Q) \times \ln \left( 1 + \frac{C(W, D)}{\mu_P(W)C} \right) \right] + |Q| \ln \left( \frac{\mu}{\mu + |D|} \right)$$

 $D_1$ :

$$S_{new}(Q, D_1) = S_{old}(Q, D_1) + C(q_1, Q) \ln \left( 1 + \frac{C(q_1, D_1) + 1}{\mu_P(q_1|C)} \right) - C(q_1, Q) \ln \left( 1 + \frac{C(q_1, D_1)}{\mu_P(q_1|C)} \right)$$

 $D_2$ :

$$S_{new}(Q, D_2) = S_{old}(Q, D_1) + C(q_2, Q) \ln \left( 1 + \frac{1}{\mu_P(q_2|C)} \right)$$

$$S_{old}(Q, D_1) = S_{old}(Q, D_2)$$

$$S_{new}(Q, D_2) > S_{new}(Q, D_1) \rightarrow$$

$$S_{old}(Q, D_1) + \ln \left( 1 + \frac{1}{\mu_P(q_2|C)} \right) > S_{old}(Q, D_1) + C(q_1, Q) \ln \left( 1 + \frac{C(q_1, D_1) + 1}{\mu_P(q_1|C)} \right) - C(q_1, Q) \ln \left( 1 + \frac{C(q_1, D_1)}{\mu_P(q_1|C)} \right)$$

$$\rightarrow \frac{\mu_P(q_2|C) + 1}{\mu_P(q_2|C)} > \ln(\mu_P(q_1|C) + C(q_1, D_1) + 1) - \ln(\mu_P(q_1|C) + C(q_1, D_1) + 1) \rightarrow$$

$$\frac{\mu_P(q_2|C) + 1}{\mu_P(q_2|C)} > \frac{\mu_P(q_1|C) + C(q_1, D_1) + 1}{\mu_P(q_1|C) + C(q_1, D_1)} \rightarrow$$

$$\mu^2 p(q_2|C) p(q_1|C) + \mu_P(q_2|C) c(q_1, D_1) + \mu_P(q_1|C) + c(q_1, D_1) >$$

$$\mu^2 p(q_2|C) p(q_1|C) + \mu_P(q_2|C) c(q_1, D_1) + \mu_P(q_2|C) \rightarrow$$

$$\mu_P(q_1|C) + c(q_1, D_1) > \mu_P(q_2|C) \rightarrow c(q_1, D_1) > 0$$

$$td(q_1) = td(q_2) \rightarrow p(q_1|C) = p(q_2|C)$$

در اینجا فرض مسئله برقرار است .

سوال 3)

(ب)

$$S(Q, D) = \sum_{w \in Q \cap D} \frac{1 + \ln(1 + \ln(c(w, d)))}{1 - s + s \frac{|D|}{avdl}} \times c(w, Q) \times \ln \frac{N + 1}{df(W)}$$

$$td(q_1) = td(q_2) \Rightarrow df(q_1) = df(q_2)$$

$$\begin{aligned}
S_{new}(Q, D_1) &= S_{old}(Q, D_1) + \frac{1 + \ln(1 + \ln(c(q_1, D_1) + 1))}{1 - s + s \frac{|D_1|}{avdl}} \times c(q_1, Q) \times \ln \frac{N+1}{df(q_1)} \\
&\quad - \frac{1 + \ln(1 + \ln(c(q_1, D_1)))}{1 - s + s \frac{|D_1|}{avdl}} \times c(q_1, Q) \times \ln \frac{N+1}{df(q_1)} \\
&= S_{old}(Q, D_1) + \frac{1 + \ln(1 + \ln(2))}{1 - s + s} \times 1 \times \ln \frac{N+1}{df(q_1)} \\
&\quad - \frac{1 + \ln(1 + \ln(1))}{1 - s + s} \times 1 \times \ln \frac{N+1}{df(q_1)} \\
&= S_{old}(Q, D_1) + \ln \frac{N+1}{df(q_1)} (1 + \ln(1 + \ln(2)) - 1) \\
&= S_{old}(Q, D_1) + \ln(1 + \ln(2)) \times \ln \frac{N+1}{df(q_1)}
\end{aligned}$$

$$\begin{aligned}
S_{new}(Q, D_2) &= S_{old}(Q, D_2) + \frac{1 + \ln(1 + \ln(c(q_1, D_2)))}{1 - s + s \frac{|D_2|}{avdl}} \times c(q_2, Q) \times \ln \frac{N+1}{df(q_2)} - S_{old}(Q, D_2) \\
&\quad + \frac{1 + \ln(1 + \ln(1))}{1 - s + s \frac{|D_2|}{avdl}} \times 1 \times \ln \frac{N+1}{df(q_2)} = S_{old}(Q, D_2) + \frac{\ln \frac{N+1}{df(q_2)}}{1 - s + s \frac{|D_2|}{avdl}}
\end{aligned}$$

$$S_{new}(Q, D_2) > S_{old}(Q, D_1) \rightarrow$$

$$S_{old}(Q, D_2) + \frac{\ln \frac{N+1}{df(q_2)}}{1 - s + s \frac{|D_2|}{avdl}} > S_{old}(Q, D_1) + \ln(1 + \ln(2)) \times \ln \frac{N+1}{df(q_1)}$$

$$df(q_1) = df(q_2) \rightarrow \frac{1}{1 - s + s \frac{|D_2|}{avdl}} > \ln(1 + \ln(2)) \rightarrow$$

$$\ln(1 + \ln(2)) \times 1 - s + s \frac{|D_2|}{avdl} < 1 \rightarrow$$

$$|D_2| < \left( \frac{1}{\ln(1 + \ln(2))} + s - 1 \right) \times \frac{avdl}{s}$$

