

# Áaèòíðíay íiáaëü íðáañòàáèáíey òáèñòíáíé eíóíðíàöèè

Áaóíð Nóaíeñéaa Áañéüááè- Íí-áíia  
18.07.2008 á.  
Íñéaaíáá íáííáéaíéá 20.07.2008 á.

Óáèñò íá-àòííáí èçááíey á óíðíàòá PDF

Áaaaáíeá

Øeðíéíá íðeíáíáíeá ñðáañòá áú-èñéèòáéüííé òáóíéèè á ðaçèè-íúð íáèàñòýð çíáíéé ñííðíáíæááòñý áúñòðúí ðíñòíí íáúáííá íáðá èíóíðíàöèè.

Íñíáóþ áeóóáéüííñòú íðeíáðáòááò ðaçðááíòeá íáòíáíá èçáéá-áíey è óíðíeðíáíey ííáúð çíáíéé, íáíáóíáeíúð áey ðáøáíey eííéðáóí eíííeáeñííá ñeñòáíííá eñííeüçíááíeá ðaçèè-íúð eéíááèñòè-áñéèð ííáóíáíá è íáòíáíá eñéóññòááíííá eíóáèèéáèòá, íáíðááéáííúð íá eáþ, çæíæáííóþ ááðíðíí.

Íñííáííé çááá-áé, áíçíeèàþúáé íðe ðááíòá ñ ííeííòáeñòíáííe áaçàíe ááííúð, ýáeyáòñý çááá-à ííeñéá áíeóíáíóíá íí èð ñíááðæáíeþ ííeñéíáííe íàøeíáíe á Éíòáðíáòá, çà-àñòóþ íá íááñíá-èááþò áááéááòííáí áúáíðá eíóíðíàöèè íí çáíðííó ííeüçíááòáéý.

Ííeüçíááòáéü íá áñáááá ííæáò òí-íí ñóíðíóèèðíááòú ííeñéíáííe çáíðíí íá ííeó-áíeá eíóíðíàöèè, eíðíðay áíó íáíáóíáeíá. Áíeáá òíáí, Óðóáííñòè, ñáyçáííúá ñ ðáøáíeáí ýóíé çááá-è, çæèþ-àþòñý á íííáííáðaçèè áíçííæíúð óíðí áúðáæáíeý íáííé è óíé æá èááè, íúñéè.

Á ááííé ñòáòúá ðaññíàòðeáapòñý íáeíóíðúá ííáóíáí è ðáøáíeþ óéaçáííúð íðíáéáí. Íñííáííá áíeíáíeá óááéáíí íáòíááí ááèòíðííáí íðá

## 1. Íáçíð íáòíáíá ááèòíðííáí íðáañòàáèáíey òáèñòíáí

Á eííóá 80 ò áííáíá á ðááíòáð Næéóííá [Salton et al. 1994] áúeá íðááeíæáíá ááèòíðíay ííáaëü eáè áeüðáðíáòeáá eáèñè-áñéííó áá ñíáéòðá ñeíá è ñíòááòñòááííí ááèòíðá á eáèñè-áñéíí íðíñòðáíñòáá. Á íðíóáññá ííeñéá -àñòíóíúé íðòðáò çáíðííá ðaññíàòðeááá ðáéáááíóíúá áíeóíáíóú.

Á áíeáá íðíááeíóóúð ááèòíðííúð ííáaëýð ðaçíáðííñòú íðíñòðáíñòáá ñíeðáúááòñý íáðáñíúááíeáí íàèáíeáá ðañíðíñòðáíííúð èèè ðá

Αναφορικά με την αλληλεπίδραση των λέξεων και την σημασία τους, ο Salton (1989) εισήγαγε τον αλγόριθμο Latent Semantic Indexing (LSI). Η βασική ιδέα του LSI είναι η Singular Value Decomposition (SVD) της μήτρας των λέξεων.

Ο αλγόριθμος LSI βασίζεται στην ανάλυση της μήτρας των λέξεων με τη μέθοδο της SVD, η οποία διασπά την μήτρα σε τρεις μήτρες: μια μήτρα λέξεων, μια μήτρα εννοιών και μια μήτρα κειμένων.

Η βασική ιδέα του LSI είναι η Singular Value Decomposition (SVD) της μήτρας των λέξεων.

## 2. Αλληλεπίδραση των λέξεων και σημασιολογία

Η βασική ιδέα του LSI είναι η Singular Value Decomposition (SVD) της μήτρας των λέξεων.

Η βασική ιδέα του LSI είναι η Singular Value Decomposition (SVD) της μήτρας των λέξεων.

$$G = \{G_1, G_2, \dots, G_i, \dots, G_n\}$$

$$V_g = \{V_{g1}, V_{g2}, \dots, V_{gi}, \dots, V_{gn}\}$$

Η βασική ιδέα του LSI είναι η Singular Value Decomposition (SVD) της μήτρας των λέξεων.

Η βασική ιδέα του LSI είναι η Singular Value Decomposition (SVD) της μήτρας των λέξεων.

$$A_i = \{A_{i1}, A_{i2}, \dots, A_{ij}, \dots, A_{im}\}$$

$$V_{ai} = \{V_{ai1}, V_{ai2}, \dots, V_{aij}, \dots, V_{aim}\}$$

ἄἄἄ Ἀἰ &mdash; ἰῆῆῆῆῆῆῆ ἰ ἰῆ ἄῆῆῆῆῆ; Ἀἰἰ &mdash; ἰ ἰῆ ἄῆῆῆῆ, ἰ = 1 &hellip; ἠ; ἂἰ &mdash; ἰῆῆῆῆῆῆ ἄῆῆῆῆῆ ἄῆῆῆῆ.

ἰῆῆῆῆῆῆῆῆῆῆ ἰῆῆῆῆῆῆῆῆῆῆ ἄῆῆῆῆῆῆῆ ἰῆῆῆῆῆῆῆ ἄῆῆῆῆῆῆῆ ἰῆῆῆῆῆῆῆῆ ἄῆῆῆῆῆῆῆῆ ἄῆῆῆῆῆῆῆ ἄῆῆῆῆῆῆῆ ἄῆῆῆῆῆῆῆ:

$P_{ij} = \{P_{ij1}, P_{ij2}, \dots, P_{ijh}, \dots, P_{ijk}\}$

$V_{\delta ij} = \{V_{\delta ij1}, V_{\delta ij2}, \dots, V_{\delta ijh}, \dots, V_{\delta ijk}\},$

ἄἄἄ  $\Delta_{ij}$  &mdash; ἰῆῆῆῆῆῆῆ ἰῆῆῆῆῆῆῆῆῆῆ ἰ ἰῆ ἄῆῆῆῆῆῆ ἰ ἰῆ ἄῆῆῆῆῆῆῆ ἰ ἰῆ ἄῆῆῆῆῆῆῆῆ ἰ ἰῆ ἄῆῆῆῆῆῆῆῆ, ἰ = 1 &hellip; ἠ  
 $V_{p\delta ij}$  &mdash; ἄῆῆῆῆῆῆῆ ἰ ἰῆ ἄῆῆῆῆῆῆῆῆ ἰ ἰῆ ἄῆῆῆῆῆῆῆῆ ἰ ἰῆ ἄῆῆῆῆῆῆῆῆ.

Ἐῆ ἰῆῆῆῆῆῆῆῆῆῆ ἰῆῆῆῆῆῆῆῆῆῆ, ἰῆῆῆῆῆῆῆῆῆῆ ἰῆῆῆῆῆῆῆῆῆῆ (ῆῆῆῆῆῆῆῆῆῆ) ῆῆῆῆῆῆ ἰῆῆῆῆῆῆῆῆῆῆ ἰ ἰῆῆῆῆῆῆῆῆῆῆ ἰῆῆῆῆῆῆῆῆῆῆ ἰῆῆῆῆῆῆῆῆῆῆ.

Ἐῆῆ ῆῆῆῆῆῆῆῆ, ἠῆῆῆῆῆῆῆῆ ἰ ἰῆῆῆῆῆῆῆῆῆῆ ῆῆῆῆῆῆῆῆῆῆ ἰῆῆῆῆῆῆῆῆῆῆ ἰῆῆῆῆῆῆῆῆῆῆ ἰῆῆῆῆῆῆῆῆῆῆ (ἰῆῆῆῆῆῆῆῆῆῆ ἰῆῆῆῆῆῆῆῆῆῆ ῆῆῆῆῆῆῆῆῆῆ ἰῆῆῆῆῆῆῆῆῆῆ ῆῆῆῆῆῆῆῆῆῆ ῆῆῆῆῆῆῆῆῆῆ ῆῆῆῆῆῆῆῆῆῆ).

Ἄῆῆῆῆῆῆῆ ἰῆῆῆῆῆῆῆ. ἰῆῆῆῆῆῆῆ. ἰῆῆῆῆῆῆῆῆ ἰῆῆῆῆῆῆῆῆ ἰῆῆῆῆῆῆῆῆ. (Ἀῆῆῆῆῆῆῆ).

Ἐῆῆῆ ἰῆῆῆῆῆῆῆῆ ἰῆῆῆῆῆῆῆῆῆῆ ἰῆῆῆῆῆῆῆῆῆῆ ἰῆῆῆῆῆῆῆῆῆῆ ἰῆῆῆῆῆῆῆῆῆῆ ἰῆῆῆῆῆῆῆῆῆῆ ἰῆῆῆῆῆῆῆῆῆῆ ἰῆῆῆῆῆῆῆῆῆῆ ἰῆῆῆῆῆῆῆῆῆῆ.

ἰῆῆῆῆῆῆῆῆῆῆ (ἰῆῆῆῆῆῆῆῆ ἰῆῆῆῆῆῆῆῆῆῆ).

ἰῆῆῆῆῆῆῆῆῆῆ ἰῆῆῆῆῆῆῆῆῆῆῆῆ, ἰῆῆῆῆῆῆῆῆῆῆῆῆ ἰῆῆῆῆῆῆῆῆῆῆῆῆ, ἰῆῆῆῆῆῆῆῆῆῆῆῆ ἰῆῆῆῆῆῆῆῆῆῆῆῆ ἰῆῆῆῆῆῆῆῆῆῆῆῆ.

Éàè óæá íòíá÷àéíńú áúðá, èàæáíá íðááéíæáíéá íáńáò á ńááá íðáááéáíúé ńíúńé, çàèèááúáááíúé ááòíðí, è íááńíá÷èáááò íðíá. Á íáúáí ńéó÷áá èàæáíá íðááéíæáíéá èíááò ńíòááòńòáópúéé ááèòíð òáéè. Óàèèí íáðàçí, òáéńò íæíí íðáááéèòú èàè ńòðóéò.

Íðááéíæáíáý ìíááüü ááèòíðà òáéè íæáò áúòú íðááńòááéáíá á àèáá òðáó èííííáíò:

Vbegin &mdash; íà÷àèüíáý òáéü, áúðàæááíáý ÷áðáç íà÷àèüíúé ááèòíð ń çàááíúíè èíðáéíàòàìè;

Vend èííá÷íáý òáéü, áúðàæááíáý ÷áðáç èííá÷íúé ááèòíð ń çàááíúíè èíðáéíàòàìè;

Z áèá ńáyçè ìæáó íà÷àèüíúí ááèòíðí è èííá÷íúí ááèòíðí .

Á èà÷áńòáá èíðáéíàò ááèòíðà ííáóò áúńòóíàòú íòááéüíúá ńéíáá, ííýòèý, èíáííúá áðóííú, íòááéüíúá íðááéíæáíéý, ńíúńéíáúá áðó.

Ííńéíèüéò èìáòńý òðè ńíńòááéýòúéá, èíòíðúá íðáááéýòò ááèòíð, òí ńíòááòńòááííí áéý ííńéááópúááí àíàèèçà íàìè áúéè áúááéá.

1) íðíńòíé ááèòíð: = ( ) èèè V = ( );

2) íóéááíé ááèòíð: = (Ø);

3) ííéíúé ááèòíð: = ( , ) ńí ńáyçüò Z;

4) íóńòíé ááèòíð: = ( , ) ááç ńáyçè Z;

5) ēāāúé āāēòíð: = ();

6) íðāāúé āāēòíð: = ().

Á ñáíp í-āðāāú āāēòíðú è íīāóò ñíñòíýòú èç íīāāāēòíðíā, èāē íòāāēúíúð ñáíñòíýòāēúíúð ÷-āñòāé ēíðāēíāò, íðēíāāēāæāúēð ò

Éāæāy ēíðāēíāò à èíāāò ñāíē àòðēāóòú atr. Àòðēāóòāìē íīāóò yāēýòñy āðāíáííúā ēēē íðíñòðáíñòāííúā òāðāēòāðēñòēē ēíñ

Ñíñòāā ēíðāēíāò āāēòíðā íðāāāēýāòñy ñēíæííñòúþ ííñòðíáíēý íðāāēíæáíēý. Á íáúāí ñēó÷āā ñíñā÷ēíáíííñòú íòāāēúíúð ÷-āñòāé í

ðāññíòðēì íðēíáííēā íñēñáííē āúøā āāēòíðíē íīāāē è íā ēííēðāòííí íðēíāðā.

Íðíáíāēēçēðóáì ñēāáópùāā íðāāēíæáíēā.

Áí āñā āðāíāíā ēþāē ñòāēēēāþòñy ñ íáíēì è òāìē æā íðíāēāíāìē ýēíííēēē.

Āáííā íðāāēíæáíēā íīæāò áúòú íðāāñòāāēāí ā āāēòíðíē ðíðíā:  $\forall p(x_1; y_1) \text{ ñí ñāýçùþ āēāā } z_1$ , ēēē ā óíðíūáííē ðíðíā  $\forall p(x_1; y_1) ($

āāā ēíðāēíāòā  $x_1 = \{ēþāē\}$ ;

ēíðāēíāòā  $y_1 = \{íðíāēāíú ýēíííēēē\}$ ;

ñâyçü âèää z1 = (ñòàèèèèâàpòñý).

Íðè ýòì àòðéáóòàìè êĩĩðäèíàòù ò1 ýäèýpòñý atrx = (âĩ âñá âðáíáíà), à àòðéáóòàìè êĩĩðäèíàòù y1 ýäèýpòñý atrý = (íáíè è òá æá)

Íà ñĩĩáá ááèòíðĩĩáĩ ïðááñòááèéáíèý ïĩáóò áúòù ðáðáíú íáéíòíðúá ïðíáéáíú íáðááíòèè òáèñòíáíé èíòíðíàòèè, â ÷-àñòíĩñè:

&ndash; ñíèðáùáíèá íáúáìà èñòíáííé èíòíðíàòèè äèý áúĩĩéíáíèý ïðíóááóð áíàèèçá òáèñòà è òíðíèðíááíèý ñèñòáì è áàç çíáíèé;

&ndash; ñèíòáç òáèñòà ñ èñĩĩèùçíááíèáì èíòíðíàòèè, èçáèáèááííé èç áàç çíáíèé.

Â ñèááópùáì ïàðááðáòá ðáññíàòðèááàòñý ááĩíáòðè-áñèáý èíòáðíðáòàòèý òáóííéíáèè ááèòíðĩĩáĩ ïðááñòááèéáíèý òáèñòà.

3. Íðèíáíáíèá òáóííéíáèè ááèòíðĩĩáĩ ïðááñòááèéáíèý ïðè áíàèèçá è ñèíòáçá òáèñòíáíé èíòíðíàòèè

Ðáññíòðáííáý áúðá ááèòíðíáý ïĩááèù ïðááñòááèéáíèý òáèñòíáíé èíòíðíàòèè ïĩæáò áúòù èñĩĩèùçíááíèá ïðè áíàèèçá è ñèíòáçá òáèñòà

Íà ðèñ. 1 ïðááñòááèéáíèá òíðíùáííáý èíòáðíðáòàòèý ááèòíðĩĩáĩ ïðááñòááèéáíèý òáèñòà á ïðíñòðáíñòáá òðáó êĩĩðäèíàò ò, ç, atr.

Ðèñ. 1. Ááèòíðĩĩá ïðááñòááèéáíèá òáèñòà

Íà ïðááñòááèéáíèá ðèñòíéá ïĩèàçáíú òðè ááèòíðá Vp1(x1, y1) (z1); Vp2(x2, y2) (z2); Vp3(x3, y3) (z3) è èò ïðíáèèèý íà ïèĩñéíñòù x, ïèĩñéíñòè: (x, y), (x, atr), (y, atr).

Àððέαóóó ηίαóò ειαóó αðáιαίίία, ίðήηòðαιήòάαίίία è äðóαέα εçιαðýαίία óαðàέòáðεήòέε.

Έηðαείαóó όι ηήðάάέýðò ία-άέüίία έηðαείαóó άάέοίðá. Έηðαείαóó ýι ηήðάάέýðò έίία-ίία έηðαείαóó άάέοίðá.

Èñοίáý εç ίðάáüáóóάάί ηιέηαιέý άάέοίð V ηήðάάέýáò έίία-ίόð óáεü ðáññιαòðεάάαιίέ άάέιέóü óáέηòá è ειαáo ηòðóέóóðó άάέó

ðáññηίòðέì èñήέüçíááίέα ηιέηαιίίέ ηήάάέè ίá ίðέìάðá.

Íόηóó çáááíü òðè άάέοίðá:

Vp1(x1, y1) (z1); Vp2(x1, y2) (z2); Vp3(x1, y3) (z3).

Íá ðεñ. 2 ηιέαçáíü ίάέιòίðüά áηçηίæίüά άáðεαίóü ááηίαòðε-áñέίέ είòáðήðáóáòέè áçαέηήάάέηòáέý òðáð άάέοίðίá.

à

á

â

### Παράδειγμα 2. Άσκηση με ασκήσεις σχετικά με τον άριστο χώρο

Έστω, δίδονται οι σημεία  $A(0, 2)$ ,  $B(1, 1)$ ,  $C(1, 0)$  στο επίπεδο  $Ox_1Ox_2$ . Η ευθεία  $AC$  ορίζεται από την εξίσωση  $x_1 + x_2 - 1 = 0$ . Ο άριστος χώρος ορίζεται από την εξίσωση  $x_1 + 2x_2 - 3 = 0$ . Να βρεθεί ο άριστος χώρος που περιέχει το τρίγωνο  $ABC$ .

Για τον άριστο χώρο  $Z$  δίδονται οι σημεία  $A(0, 2)$ ,  $B(1, 1)$ ,  $C(1, 0)$  στο επίπεδο  $Ox_1Ox_2$ . Η ευθεία  $AC$  ορίζεται από την εξίσωση  $x_1 + x_2 - 1 = 0$ . Ο άριστος χώρος ορίζεται από την εξίσωση  $x_1 + 2x_2 - 3 = 0$ . Να βρεθεί ο άριστος χώρος που περιέχει το τρίγωνο  $ABC$ .

$$V_1(x_1, y_1) (z_1); V_2(x_2, y_2) (z_2); V_3(x_3, y_3) (z_3),$$

$$\text{όπου } y_1 = x_2, y_2 = x_3.$$

Ο άριστος χώρος που περιέχει το τρίγωνο  $ABC$  ορίζεται από την εξίσωση  $x_1 + 2x_2 - 3 = 0$ . Να βρεθεί ο άριστος χώρος που περιέχει το τρίγωνο  $ABC$ .

### Άσκηση με ασκήσεις σχετικά με τον άριστο χώρο



$Vp_1(x_1, y_1) (z_1); Vp_2(x_1, y_2) (z_2); Vp_3(x_3, y_3) (z_3),$

$\alpha \alpha \alpha y_2 = x_3.$

Απόδειξη: Η απόδειξη είναι απλή και βασίζεται στην αρχή της επαγωγής. Η απόδειξη είναι η εξής:

Εάν  $x_1 = 0$ , τότε  $y_1 = 0$  και  $z_1 = 0$ . Η απόδειξη είναι η εξής:

Εάν  $x_1 = 1$ , τότε  $y_1 = 1$  και  $z_1 = 1$ . Η απόδειξη είναι η εξής:

Εάν  $x_1 = 2$ , τότε  $y_1 = 2$  και  $z_1 = 2$ . Η απόδειξη είναι η εξής:

Εάν  $x_1 = 3$ , τότε  $y_1 = 3$  και  $z_1 = 3$ . Η απόδειξη είναι η εξής:

Εάν  $x_1 = 4$ , τότε  $y_1 = 4$  και  $z_1 = 4$ . Η απόδειξη είναι η εξής:

Εάν  $x_1 = 5$ , τότε  $y_1 = 5$  και  $z_1 = 5$ . Η απόδειξη είναι η εξής:

Εάν  $x_1 = 6$ , τότε  $y_1 = 6$  και  $z_1 = 6$ . Η απόδειξη είναι η εξής:

– Η παρούσα εργασία εστιάζει στην αξιολόγηση των μεθόδων που χρησιμοποιούνται για την εξαγωγή της σημασίας των λέξεων

– Η αξιολόγηση γίνεται με βάση την αρχική σημασία των λέξεων και το ποσοστό της επίδρασης της λέξης στην ολική σημασία της φράσης

– Η μεθοδολογία που ακολουθείται είναι η ανάλυση της σημασίας των λέξεων με βάση την αρχική τους σημασία και το ποσοστό της επίδρασης της λέξης στην ολική σημασία της φράσης

– Η αξιολόγηση γίνεται με βάση την αρχική σημασία των λέξεων και το ποσοστό της επίδρασης της λέξης στην ολική σημασία της φράσης

– Η ανάλυση γίνεται με βάση την αρχική σημασία των λέξεων και το ποσοστό της επίδρασης της λέξης στην ολική σημασία της φράσης

## Summary

The paper considers the approaches associated with the vector representation of the textual information. The particularity of the approach under consideration is in the determination of the goal functions of separate sentences and representation of them in the form of some local vectors on which basis a global vector is built that determines the semantic component of the text on the whole. Various aspects of application of the proposed approach are considered.

Ημερομηνία: 2020

Απόδοση: 2005. Ημερομηνία, Γ. Α. Ημερομηνία: 2005. Ημερομηνία. Διεύθυνση: ΟΕΔΒ, 2005.

Ημερομηνία: 2005. Ημερομηνία, Γ. Α. Ημερομηνία: 2005. Ημερομηνία: 2005. Ημερομηνία: 2005. Ημερομηνία: 2005. Ημερομηνία: 2005. Ημερομηνία: 2005.

Ημερομηνία: 2005. Ημερομηνία, Γ. Α. Ημερομηνία: 2005. Ημερομηνία: 2005. Ημερομηνία: 2005. Ημερομηνία: 2005. Ημερομηνία: 2005. Ημερομηνία: 2005.

Ημερομηνία: 2005. Ημερομηνία, Γ. Α. Ημερομηνία: 2005. Ημερομηνία: 2005. Ημερομηνία: 2005. Ημερομηνία: 2005. Ημερομηνία: 2005. Ημερομηνία: 2005.

Ñíèèðèí è äð. 2005 Ñíèèðèí, À. Ã. Íðíáèð ÄÈÀÈÈÍÁ, COM-íáúáèð Goldrml / À. Ã. Ñíèèðèí, À. Á. Íáíéðàòíá. Ì. : Äèàèíá, 2005.

Òèíí 1999 Òèíí Á. È. Í ðíèè ìàøèííáí íáó÷áíèý á èíòáèèáèòóàèüíúõ ñèñòáìàõ // ÍÒÈ. Ñáð. 2. 1999. <sup>1</sup> 12. Ñ. 1&ndash;3.

Salton 1989 G. Salton. Automatic Text Processing. Addison-Wesley Publishing Company, Inc., Reading, MA, 1989.

Salton et al. 1994 G. Salton, J. Allan, and C. Buckley. Automatic structuring and retrieval of large text files. Communications of the ACM, 37(2), February 1994.

Todd et al. Todd A. Letsche and Michael W. Berry. Large-Scale Information Retrieval with Latent Semantic Indexing. URL: <http://www.cs.utk.edu/~berry/sc95/sc95.html>.