

Áaèòíðíay íiáaëü íðáañòàáèáíey òáèñòíáíé eíóíðíàöèè

Áaóíð Noaíeñéaa 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.

Íñíáóp áeóóáéüííñòú íðeíáðáòááò ðaçðááíòeá íáòíáíá èçáéá-áíey è óíðíeðíáíey ííáúð çíáíéé, íáíáóíáeíúð áey ðáøáíey eííéðáóí eíííeáeñííá ñeñòáíííá eñííeüçíááíeá ðaçèè-íúð eéíááeñòè-áñéèð ííáóíáíá è íáòíáíá eñéóññòááíííá eíóáèèéáèòá, íáíðááéáííúð íá eááþ, çáeíæáííóp ááðíðíí.

Íñííáííé çááá-áé, áíçíeèàþúáé íðe ðááíòá ñ ííeííòáeñòíáíe áaçàíe ááííúð, ýáeyáòñý çááá-à ííeñéá áíeóíáíóíá íí èð ñíááðæáíeþ ííeñéíáíe íàøeíáíe á Éíòáðíáòá, çá-àñòóp íá íááñíá-èááþò áááéááòííáí áúáíðá eíóíðíàöèè íí çáíðííó ííeüçíááòáéý.

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

Á ááííé ñòáòúá ðaññíàòðeáapòñý íáeíóíðúá ííáóíáú è ðáøáíeþ óéaçáííúð íðíáéáí. Íñííáííá áíeíáíeá óááeáíí íáòíááí áaèòíðííáí íðá

1. Íáçíð íáòíáíá áaèòíðííáí íðáañòàáèáíey òáèñòíá

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

Á áíeáá íðíááeíóóúð áaèòíðííúð ííááeyð ðaçíáðííñòú íðíñòðáíñòáá ñíeðáúááòñý íòáðáñúááíeáí íàèáíeáá ðañíðíñòðáíííúð èèè ðá

Áéaáíúí áíñóíeíñóáíí ááéòíðíé ííááéé yáéyáòñy áíçíæííñòú ííeñéá è ðáíæèðíááíéy áíeóíáíóíá íí ííáíáep, òí áñòú íí èò áéèçíñòè óáííeáòáíðeòáéúíúíè, ÷òí íñíááííí íðíyáéyáòñy, éíááá çáíðíñ ííááðæèò íáeíá éíeè-áñòáí ñeíá. Áéy ííeó-áíéy éó-ðáé ðáéáááíóííí Latent Semantic Indexing (LSI). Ííááéú èñííeúçíááéè Singular Value Decomposition (SVD) áéy íáðáóíáà íò ðáçðáæáííé íàððeòè

LSI ííeàçàéè çíá-èòáéúííá íðááíñóíáñóáí á ðáçóéúòàòòð ííeñéá íí ñðááíáíep ñ éáéñè-áñéèí íáòíáíí, íáíáéí ñeíæííñòú ííááéè ÷áñ òáðíeéíé [Salton 1989]. Íáíá èç íáéáíeáá ðááíòíñííáíúò ñeñòáí íá íñííáá LSI áúeà ñíçááíá á Ááðéèè á 1995 áíáó íáééèí Ááððe

Ííeñúáááíáy íeæá ñeñòáíà èñííeúçóáò ñíááððáííí áðóáóp eíòáðíðáòàòèp íííyòèy ááéòíðíé ííááéè òáéñòà, á eíòíðíé íá íðeíáíyðp

2. Ááéòíðíáy ííááéú íðááñòàáéáíéy òáéñòíáíé eíòíðíáòèè

Á ááííé ðááíòá òáéñò ðàññíàòðeáááòñy éáé ñòðóéòóðá, òí áñòú éáé ñíáíeóíííñòú íòááéúíúò áçàeííñayçáííúò íðááéíæáíeé, íáú ááòíðíí, ÷áðç íííeáñòáí ííáòáéáé ðáçííáí ðáíáá, ííðáááéyáíúò íòááéúíúíè íðááéíæáíéyíè, ááçáòáíè, íáðááðáòáíè, áéáááíè è ò. í.

Íeæá íðeáíáyòñy íðeíáðú íàòáíàòè-áñéíé eíòáðíðáòàòèè ááéòíðíé ííááéè íáéíòíðíáí çáeíí-áíííáí yéáíáíòà òáéñòà, ñíñòíyúááí, íá

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

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

ááá G — íííeáñòáí áéèá; G_i — í áy áéèáá, i = 1… n; V_g — íííeáñòáí ááéòíðíá òáééé áéèá; V_{gi} —

Á ñáíp í-áðááü

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

$$V_{àí} = \{V_{àí1}, V_{àí2}, \dots, V_{àíj}, \dots, V_{àím}\},$$

Έαε όαά ίοίά÷άεήνύ άύοά, έαααίά ίοάάεήαεία έαήαό ά ηάάά ίοάάάέάίύέ ηιύήέ, αάέεάάύάάάίύέ άάοίοί, έ ίάάήίá÷έάάάο ίοίά. Ά ίάύάι ηέό÷άά έαααίά ίοάάεήαεία έίάάο ηίίόάάòηòáóρπύέέ άάέοίò öáέέ. Όάέέι ίάòàçí, òáέήò ίίáí ίοάάάέέέóü έáέ ηòòóέó.

Íòááéήáείáy ίίááüü ááέóíòá öáέέ ίίááò áúòü ίòááήòááéáía á áéää òòáó έήίίίáíò:

Vbegin — ίá÷áέüíáy öáέü, áúòáááááy ÷áòáç ίá÷áέüíúé ááέóíò ñ αάááίúìέ έήòáéíaòáìé;

Vend έίíá÷íáy öáέü, áúòáááááy ÷áòáç έίíá÷íúé ááέóíò ñ αάááίúìέ έήòáéíaòáìé;

Z áéä ηáyçé ίáááó ίá÷áέüíúì ááέóíòíí è έίíá÷íúì ááέóíòíí .

Ά έá÷áήóáά έήòáéíaò ááέóíòá ίίáóò áúήòóíaòü ίòááéüíúá ηέíaá, ίίýòéý, έíáííúá áòóííü, ίòááéüíúá ίòááéήáείáy, ηιύήέíaúá áòó.

Ííήέíéüéó έíáρòήý òòé ηίήòááéýρπύέá, έίòíòúá ίíòááéýρò ááέóíò, òí ηίίòááòηòááίíí áéý ίíήéááóρπύááí áíaéèçá ίáìé áúéé áúááéá.

1) ίòíήòíé ááέóíò: = () έέé V = ();

2) ίóéááíé ááέóíò: = (∅);

3) ίίέíúé ááέóíò: = (,) ηί ηáyçüρ Z;

4) ίóñòíé ááέóíò: = (,) ááç ηáyçé Z;

Àððέαóóó ηιάóó ειαóó αðáίάίύά, ίðήηòðáήηòάάίύά è äðóáèà èçίáðýáίύά óαðàèòáðèñòèèè.

Έηðáèíaóó όι ηήðáááèýðò ίá-àèüíύά èήðáèíaóó άáèòíðá. Έηðáèíaóó γι ηήðáááèýðò èήίá-ίύά èήðáèíaóó άáèòíðá.

Èñóíäý èç ίðááüáóóááί ηιèñáíèý άáèòíð V ηήðáááèýáò èήίá-ίόð óáèü ðáññíàòðèáááίé ááèíèóü óáèñòà è ειαáo ηòðóéòóðó άáèò

ðáññííòðèì èñήèüçíááíèá ηιèñáίίé ηήááèè ίá ίðèìáðá.

Íóñóó çáááíύ òðè άáèòíðá:

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

Íá ðèñ. 2 ηιèàçáíύ ίáèíòíðύά áίçííæíύά άáðèáíóü ááήáòðè-áñèíé εíòáðíðáòáòèè áçáèíήááèñòáèý òðáð άáèòíðá.

ἂ

ἂ

ἂ

Ἰδὲἰάδ, ὀδὲñ. 2ἂ, Ἰδὲἰάδ ὀδὲñ. 2ἂ, Ἰδὲἰάδ ὀδὲñ. 2ἂ

Ἰδὲἰάδ, ὀδὲñ. 2ἂ, Ἰδὲἰάδ ὀδὲñ. 2ἂ, Ἰδὲἰάδ ὀδὲñ. 2ἂ

Ἰδὲἰάδ, ὀδὲñ. 2ἂ, Ἰδὲἰάδ ὀδὲñ. 2ἂ, Ἰδὲἰάδ ὀδὲñ. 2ἂ

$Vp1(x1, y1) (z1); Vp2(x2, y2) (z2); Vp3(x3, y3) (z3),$

$\alpha\alpha\alpha y1 = x2, y2 = x3.$

Ἰδὲἰάδ, ὀδὲñ. 2ἂ, Ἰδὲἰάδ ὀδὲñ. 2ἂ, Ἰδὲἰάδ ὀδὲñ. 2ἂ

Ἰδὲἰάδ, ὀδὲñ. 2ἂ, Ἰδὲἰάδ ὀδὲñ. 2ἂ, Ἰδὲἰάδ ὀδὲñ. 2ἂ

Vp1(x1, y1) (z1); Vp2(x1, y2) (z2); Vp3(x3, y3) (z3),

ãää y2 = x3.

Äðóæì ìðèíáíáíéàì áâèòíðíé ìääèè yäÿyáòñÿ áíçíæííñòü ðääèèçàòèè ñèíòáçà òáèñòíáíé èííðíàòèè.

Íðääííèíæèì, ÷òí ðáøáàòñÿ çääà÷à, ñâyçáííáÿ ñ ðañèðúòèàì ìíÿòèÿ ò1. Á ýòì ñèó÷ää, áâèòíð òáèè äÿÿ ìíèñáíèÿ ìðääáéáíúò ìð èííñòðóèðíááíèÿ ìääèèé.

Äííòñèì, á áàçá çíáíéè ìíÿòèè ò1 ìðääáéáí íá ìíæáñòáá ìíóíèíáéè ÷áðç áâèòíð Vp0(x1, y1) (z1). Á ñáíð í÷áðääü ìääáèòíð y1 ðañèðúòèÿ ìíÿòèÿ ò1. Íáðáíèçì ðàçááðòúááíèÿ áâèòíð äÿÿ ìíèñáíèÿ ìðíáññíá è ÿáéáíéè ìæàò áúòü ááíÿèè: èéá íá ìíííáá ñíó ÿçúéá.

Íá ìíííáá ìðääéíæáííé ìääèè ðàçðááíòáíá òáðííèíáèÿ íáðááíòèè òáèñòíáíé èííðíàòèè íá ìíííáá áâèòíðíé ìääèè òáèñòá.

Çàèèð÷áíéá

Ðaññíòðáíúá á ááííé ñòàòúá ìíííáíúá ìíèíæáíèÿ òáðííèíáèè áâèòíðííáí ìðääñòááéáíèÿ òáèñòíáíé èííðíàòèè è ááòííàòèçàòèÿ ý

– ñíçááíèÿ ìðíáññèííáéüíúò ñèñòáì è áàç çíáíéè;

– ìääáðæèè ìðíáññèííáéüííé äÿòáéüííñè ðááíòíèèíá ðàçèè÷íúò ìððáñéáé;

– Η παρουσίαση των αποτελεσμάτων της έρευνας και η συζήτηση των αποτελεσμάτων είναι διαθέσιμα στην ιστοσελίδα του έργου.

– Η παρουσίαση των αποτελεσμάτων της έρευνας και η συζήτηση των αποτελεσμάτων είναι διαθέσιμα στην ιστοσελίδα του έργου.

– Η παρουσίαση των αποτελεσμάτων της έρευνας και η συζήτηση των αποτελεσμάτων είναι διαθέσιμα στην ιστοσελίδα του έργου.

– Η παρουσίαση των αποτελεσμάτων της έρευνας και η συζήτηση των αποτελεσμάτων είναι διαθέσιμα στην ιστοσελίδα του έργου.

– Η παρουσίαση των αποτελεσμάτων της έρευνας και η συζήτηση των αποτελεσμάτων είναι διαθέσιμα στην ιστοσελίδα του έργου.

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.

Νέοι κείμενα

Απόστολος 2005 Απόστολος, Ι. Α. Γενεαλογική Ιστορία / Ι. Α. Απόστολος. Ι. : ΟΕΣΒ, 2005.

Γενεαλογική Ιστορία 2005 Γενεαλογική Ιστορία, Ν. Α. Γενεαλογική Ιστορία / Ν. Α. Γενεαλογική Ιστορία, Α. Ι. Γενεαλογική Ιστορία.

Επιμέλεια 2005 Επιμέλεια, Π. Ι. Γενεαλογική Ιστορία / Ν. Α. Γενεαλογική Ιστορία, Α. Ι. Γενεαλογική Ιστορία.

Δοκίμια 2005 Δοκίμια, Α. Ο. Γενεαλογική Ιστορία (Επιμέλεια) / Ν. Α. Γενεαλογική Ιστορία, Α. Ι. Γενεαλογική Ιστορία.

Ñíèèðéí è äð. 2005 Ñíèèðéí, À. Á. Íðíáèð ÄÈÀËËÍÁ, COM-íáúáèð Goldrml / À. Á. Ñíèèðéí, Ä. Á. Íáíéðàòíá. Ì. : Äèàéíá, 2005.

Òèíí 1999 Òèíí Á. È. Í ðíèè àèèíííáí íáó÷áíèý á èíòáèèáèòóàèüíúõ ñèñòáìàð // ÍÒÈ. Ñáð. 2. 1999. ¹ 12. Ñ. 1–3.

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>.