

Èçâëå÷åíèå è àíàëèç äàò ïðîèçâåäåíèé â êîðíóñå öèòàò ïíëàéí-ñëîâàðÿ

Àâòîð Àíäðåé Àìàòïëüââè÷ Êðèæàíâñèé
05.08.2012 ÿ

05.08.2012 á.

Linnexäääää iäääääääää 03.10.2012 ä.

Òåêñò â ôîðìàòå PDF (262.18 kB)

Òåêñò â ôîðìàòå PDF_En (2.07 MB)

Summary. Quantitative evaluation of quotations in the Russian Wiktionary was performed with the use of the developed Wiktionary parser. It was found that the number of quotes in the dictionary grows fast (51.5 thousands in 2011, 62 thousands in 2012). These quotes were extracted and stored to the database of the machine-readable dictionary. The tables of the relational database of the machine-readable dictionary related to the quotations were designed. The histogram of distribution of quotations of literary works created in different years was built.

Åååäåíèå.

Đàcàèòèå èlìiñüpòåðíûò òåðñïéïâéè íåðñïä-èeï íiyâéäíéå ñëíâåðåé íiâîåò òeïà íièäé-ñïðåâî-íéêîâ, à ñïçäåíéè èlòiðûò lïæåò ïðéíýò ò-àñòèå ñøðíéèé èdóå åóáöñùò çàèiòåðåññâåííûò íiñüçïtåòåðåééå. lï ñòðåâíåíép ñ ðòåäæòeíñííé èåéñòéíåðåðåééå íiäíåíáy ïðåâíéçåòéý ðååíòû, ñ íåñíé ñòðåðíû, íåðñïä-èåâåò lïñññíåññûò lïðåèòñòåñòå (åñññíéèé ðåii ðååíòû, åíçñíæññòü íåñòåéååíéy è ðåååéòeòðååíéy ñëíâåðíûò ñòðåðåé ía èþpáñ ýòåïå ññòåâéåééåy), ñ ãððååíé íåíåðòæèååò ñòùåñòååííûé íåäññòåòòé, èlòiðûé çåééþ-àåòñý à åñññíéé ååðñïòñòe íiyâéäíéy èåéóí èåé à èñññéüçóåííí iàòåðèåéå, òåé è à ó ñåéñòå ñåññíå ñëíâåðy. lïðåñòååéýåñý, -òi ááííáy iðñåéåìå à ñèéò ñåñéé åèòðåéüññòe çàññéóæèååò ññíåíåí áíèòìåíéy è íiñåò ñåñéé åòåðåíå iòòåí ñíçäåíéy ññåòèåéüñûò lïðåñòåíi, ññóùåñòåéýþùèò åòåðåíåò-áññéé è áíåéèç íièäé-ñëíâåðy ía èþpáñ ýòåíå áåñí ñíçäåíéy. Å íåññíüåé ñòðåðüå ía íðéíåðå áíåéèçå èéëéþñòðåòéåññíå iàòåðèåéå, íðåäññòååéåííííå à «Đóññéíi Áéèéñèíåðå», åóäöò iàìå-åíû è íðíñññòåñòåííííå ìåéñòåñå áíçñíæñûå íiññòåñå è ðååðåíép íacåñåííé íðíñññåíû.

Âèêèñëîâàðü ýòî

Äîñòîèíñòâàè

Âèèëñèïâàðý ÿâèÿåðñý áíèüòíé íáú, í è ðàçííâðàçèå ëåêñèëîâðåòè÷åññèëõ äàííûõ. Âðàáíòàõ [Edèæäííâñèé, 2011à; Meyer and Gurevych, 2012] ííèàçàí, þòí íí íáú, íó èíòîðìàöèè íåìåðöèé Âèèëñèïâàðü ñííñòàâè íñ òåçåóðóñàìè GermaNet è OpenThesaurus, à Áíàééñèéñèé Âèèëñèïâàðü áâæå íðåâîñòäèò íáú, í àáííûõ WordNet. Ñëíàâðý, ïðàâèíèííä íí íáú, íó è ðàçííâðàçèþ ëåêñèëîâðåòè÷åññèëõ äàííûõ íñ Ðóññèèì Âèèëñèïâàð, í è íàðíäÿùåäñý â ìòéðûòñ äññòóïå, íà äàííûé ííàíó, íí-âèäèìíó, íåò.

Çíà÷åíèÿ ñëîâ â îíëàéí-ñëîâàðå

Þíðið í áæðaþóðu ðeð óáðaðið er eða óðaðið. Þíðið er óðaðið. Æðið er óðaðið.

Ӧәәүп һіңдіккінде өзіншілдікке қарастырылады. Өзіншілдікке қарастырылады. Өзіншілдікке қарастырылады.

Áððéòåðéòðå áàçû ãàííûõ êîðïóñà öèòàò. Áàçà ãàííûõ êîðïóñà öèòàò ýâëÿåòñÿ ÷àñòüþ ðåëÿöèíííé áàçû ãàííûõ ìàøèí÷èòåðåííâ ñeiâàðÿ, ïðåñòàâëåííé â ðàáîòå [Éðèæàíâñêé, 2010].

òåêñò öèòàòû,

ïåðåâîä íà ðóññêèé ýçûê;

òðàíñêðèïöèÿ öèòàòû;

èíôîðìàöèÿ î öèòàòå:

íàçâàíèå
ïðîèçâåääåíèÿ;

àâòîð ïðîèçâåäåíèÿ;

èçääíèå äëý öèòàò
èc æóðíàëüíûô, ãàçåòíûô ñòàòåé;

äàòà ñîçääàíèÿ
ïðîècâåäåíèÿ:

êîðïóñ òåêñòîâ,
îòêóäà áçÿòà öèòàòà

Đèñ. 1. Òàáëèöû è ôòíøåíèÿ â áàçå
äàííüö êòðíóñà öèòàò

ìàøèíî÷èòàåíäî ñëîâàðý

Yēñïåðèìåíòû.

Â iððääüüäóùåé ðääíòå [Êðëæàííâññéé, 2011á] á õíäå ýéñíäðèíàíòå áúëè iðíàíæéèçèðíâàíù áàííùå èí, í ààðòðíà öèòàò. Â ýòíé ðääíòå ýéñíäðèíàíòú ñâýçàíù n äàòàìè iðíèçâåäåíéé. Â òàáéèöó ñ ãíàìè (quot_year) ññðåíýþðòñý äåà åíäà íà÷àéè è íêíí÷àíéý ñíçääàíéý èèòåðàòóðíñíí iðíèçâåäåíéý. Åññéè iðíèçâåäåíéå áúëèí ñíçääàíí á òà÷àíéå íäííñí åíäà, óí íàá ííéý ñíäåðæàò íäéíàéíàíñí ÷èñéí. xéñéíí óíèéàëüíûò íàð (åíä íà÷àéè åíä íêíí÷àíéý) ñíñòàâééè 862.

Äëÿ åû÷èñëåíèÿ ÷èñëà öèòàò îí åîääì

âúññéíýéñý íáðtâ áññâð öèòàò è, áññéè á öèòàòà óéàçáì ãíä ñíçääíéý iðiçääáâáiéý XXXX, eëéâí äèáïäçíí á íáññéüêí eåò XXXX –; YYYY, òí ÷eññí öèòàò äéý ýòèo eåò óáâéè–éáâåòñý íá ååéíéoo. íaïðèìåð, á Ðóññéíí Áééèññéâåðå á ñéíâàðíûñ ñòàöüýô «Áíçðíæäáâíéâ» è «òåíñâåàðü» á öèòàòàð óéàçáì 1927 á. è 1945–; 1955 áã. ñíñòåâåðñòåâáíí:

Íí rääetääää ja lääetää õõdääetääeööü, ðòi íí ðäeë õõdäöö ðäeoööäo. A. E. läñðööäöäe, «Äïeööäöäe, Äeäaäää», 1945–1955 a.

Íáõîä 26596 öèòàò, â êîòîðûô
óêàçàíà äàòà, iïçâïëë ïïñòðèöü ñèåäóþùóþ ãèñòîäðàìíó (ðèñ. 2), iïèàçûâàþùóþ çàâèñèïñòü ÷èñëà öèòàò â iíëàéí-ñèîâàðå
îò ãïäà iðèçâäääíéÿ â äèàïçíá 1750–2012 ãä.

Iêê ÷èñëà òèòàò à 2000 à ã.,
 âåðöýòíí, iñæíí íáúýñíéòü ìòíññéòåëüíí áîëüøèí êïëë÷åñòâíí ãàçåò è æóðíàëéâ,
 äíñòðöííûò à lâòèñíáëüíí êïðlóñâ ðóññéíâí ýçûêà (ÍÉÐß) çà ýòíò iåðèíä. Å äáíííí
 ñéó÷åå ñéåäåò ðòéåíòèðíàòöñý íà êïðlóñ íÉÐß, ò.é. îääàåéÿþùåå ÷èñëíí òèòàò
 Ðóññéíâí Äéèéñéíâàðý åçvôí èc íåâíí [ÍÉðèæâíññéé, 2011á].

Äëÿ ïíèìàíèÿ ìòíñèòååëüíí áîëüøíâí
 -èñèà öèòàò íà ðèñ.
 2 å 1830 å – 1880 å ïíðîáóåí ìðíàíàëèçèôðíâàòü
 áåéëäà ñàïùòí -àñòíòí öèòèðòåíùòí å Äèéèñëíàðäà íèñàðååëå

Đèñ. 2. Çàâèñèìñòü ÷èñëà
öèòàò à ïíèàéí-ñèïâàðå ìò ãîñà ïðîèçâåäåíèÿ

1

1 â êîëííéå «Àâôïð» ïðêâåäåíí èìý
ïèñàòåäéÿ ñ ìàéñèìàëüíû ÷ èñëèí òèòàò à Âèéèñëíâàðå. Âòïðàÿ è òðåòüÿ êëííè
ïíèàçûñàòþ ñòðåìèòåëüíûé ðîñò ðàçìàðå ñëíâàðÿ ï ÷ èñëó òèòàò äéÿ ýòëõ ïèñàòåäåé
çà 2011 è 2012 åã.

Íñíâíé

èñòî÷èê öèòàò äëÿ Äèêèñëâàðÿ ÍÊÐß, îöñþäà ñòîëáåö «Íóáëèàöèè â ÍÊÐß» (ãäåû

Íåðâíé è ïñëäääíåé ïóáëëàöèè ääíííá àâòòðà, ñóùñòåóþùåé â êòðïóñå). È äëÿ ýòïâí æä íåðèïäà ïäñ-èòàíí íáùåå ÷èñëî öèòàò â Âèëèñëïâàðå. Íñëäääíéé ñòiëååö ïíêàçûååö ïðïöåíííå ìòíøåíéå ÷èñëà öèòàò àâòòðà (òðåòèé ñòiëååö) è ÷èñëö áñåö öèòàò â Âèëèñëïâàðå çà òîò íåðèïä, çà êòòðûé ïóáëëàöèè àâòòðà äíñòóííû â ÍEDB (ïðåäññëäääíéé ñòiëååö).

Ñóììàðíå

÷èñëî öèòàò ñâìè íåðâûõ ïèñàòåëåé çà 1815–1910 áã. Õàâíí 22429, ÷òî ñîñòååëýåò 20,5%, ð.å. íäíó ïýòóþ ìò áñåö öèòàò Âèëèñëïâàðý çà ýòîò íåðèïä. Ååðýòíí, áññéàÿ öèòëðóåññòü ýòèõ ïèñàòåëåé è íáúýñíÿåò íèé íà ðèñ.

2 á 1830–1880 á áñäû.

Òàáë. 1. Ñàìûå ïñóëýðíûå
àâòòðû öèòàò â Âèëèñëïâàðå

1

Àâòòð

×èñëî öèòàò (2011)

×èñëî öèòàò (2012)

Íóáëëàöèè â ÍEDB

Åñååí öèòàò â Âèëèñëïâàðå

(çà ýòîò íåðèïä)

%

(2012)

1

À.Í. xåõîâ

716

931

1880–1904

4704

19,8%

2

Ë.Í. Òíëñòíé

529

710

1852–1910

14954

4,8%

3

À.Ñ. Íóøêèí

520

627

1815–1836

3217

19,5%

4

Ô.I. Äiñóïâñêèé

500

776

1846–1881

11853

6,6%

5

È.Ñ. Òóðãåíââ

457

697

1846–1882

12012

5,8%

6

Í.Â. Ñiâñëü

321

473

1831–1847

4511

10,5%

7

I.N. Ëåñêîâ

245

386

1862–1894

9039

4,3%

8

I.À. Áóëàéíâ

207

267

1920–1940

10049

2,7%

9

Áðàòüý Ñòðóâöêå

171

225

1964–1979

5699

4,0%

10

Â.I. Àñòàôüåâ

142

199

1967–2001

16327

1,2%

Đèñ. 3. Ðèñòáñðàìà ðèñëà öèòàò â
ííëàéí-ñëîâàðå ïï ãïäàì è ãïäû òâîð÷åñêíé

àêòèâíñòè íàèáíëåå öèòèðóåìûõ â Đóññéíí Ðèéèñëîâàðå íèñàòåëåé

Íñòà, òñý
îòêðûòûí áíñðîñ çà ñ÷, ò êàéèõ àâòîðîâ ïíëó÷èëñý òàéîé ìèé íà ðèñ.
2, äëÿùèéñý ñ ñåðåäéíû 1920 õ äî 1940 ãïäà?

Dàñiðåääåëåíèå

її հայեա. Ա օինա յենիածելայուա ասու իյօ-հայ Ծանօթահայեա հենիա օքօաօ ա հենիածին ո հոածի ի հայեա ն XVII ի XXI, ասա իյա XXI հայեա էիածոնի ա հեաօ ահա ն 2000 ի 2012 ահը բ-էօածի (ծեն 4).

11æii

Íðíðaðeðöü, ðóði éðaæðuúé ñeðaðóþùéé áðáé íðaðaðòàðaðéái á öðeðaðaðoð áíðéðað íiðéíí, -ðáí íðaðaðuðaðoðéé. Áðaðiyðií, ýða òaíðaðiðey ñiðoðaðiðey, ð.é. íðaðaðuð 12 eðað ýðiðií áðaðað óðæð aðaþò 10% ìð aðaðaðií -ðeññéað öðeðaðoð á ñeðaðað.

Đèñ. 4. Đàñïõåääëåíèå ÷èñëå
öèòàò à ñëèåàðíûô ñòàòüýô

ii ååêàì ñ XVII
äî XXI åâ.

Jáú

Êtõõõñää õèòàò. Ñ iiiiñüþ iðiñäðàiiû, èçäéâéâþùåé ääííúå èç Ðóññêiäi Äèéèñëiàðöy [Èdõeæäiññèé, 2010], iñ ääííúi Äèéèñëiàðöy ià 25.03.2012 áuë iññöðiáí êiðiõñ èç 62 õùñý÷ õèòàò (â 2011 á. áuëi 51.5 õùñý÷). Íðe ýòi 52 õùñý÷-e õèòàò (84% ìo âññääiñ ÷eññää õèòàò) èeééþñòðeðóþò Ðóññêeå ñeññää (â 2011 áuëi 82%).

Â
 Đóng nhẽ ãi Áèéèñéñéâàðå äëÿ 23,8 òùñý÷ öèòàò (38,35% ìò áññåö öèòàò) áûë óèàçà èñòî÷íèê, èç êòîïðñâi ëëó÷âáìà äàííàÿ öèòàòà (â 2011 á. áûëi 17 òùñý÷ öèòàò ñ èñòî÷íèéàïè, ô.â. 33%). Áæáâiúi èñòî÷íèéi ÿäéÿâðñý lâðééíàëüíûé êîðíóñ ðóññéñéâi ýçûéà, íà êòîïðûé ñññüëåðñý 94,15% öèòàò ñ èñòî÷íèéàïè.

Càêëb÷åíèå.

Áëàãäàðíñòè

Đà Nẵng

Ôóíàìáíòàëüíûô èññëåäíàíéè ïðåçèäéòíà ÐÁÍ «Èíòåëéåêòóàëüíûâ èíôîðíàëöèííûâ
òåðííëíæè, íàðíàòåðè-åññéïà ïðåçèäéòíàíéè, ñèñòåííûé áíàëéç è àâòíàòèçàöëÿ» è
ïðåçèäéòà 1.2.2 ïðíàðäííû ïÍÈÒ ÐÁÍ «Èíòåëéåêòóàëüíûâ èíôîðíàëöèííûâ
òåðííëíæè, ñèñòåííûé áíàëéç è àâòíàòèçàöëÿ». Ñíàñèáí
íèéíëàþ Ôåñëÿ çà íëíàòåðííà íåñóæäáíéà ýéñíàðèíàòëüííè ÷àñòè ðàáíòû.

Ñïèñîê ëèòåðàòóðû

Ãðèøèíà

Коцебу, 2005 — Аðеøéìа, А. А., йеóíäýи А. А. йаðñïàéòèâû ðаçâèòëÿ йаðеííàëüííâî ëîðïöñà ðóññëîâî ýçûêа // йаðеííàëüíúé ëîðïöñ ðóññëîâî ýçûêа. І.: Еíäðе, 2005. <http://www.ruscorpora.ru/corpora-biblio.html>.

Êðèæàíâñêèé.

2010 — Êðèæàíññèé Á.Á. Íðåíáðàçñâàíèå ñòðóêòóðû ñëîâàðñíé ñòàòüè Áèéèñëíàðöý â òàáëèöû è ìòíøåíèÿ ðåëýöèííé áàçû äàííû // Íðåíðèíò. 2010. <http://scipeople.com/publication/100231/>.

Êðèæàíññèé,

2011à — Êðèæàíññèé Á.Á. Êîëè-åñòâåííûé àíàëèç ëåêñèèè àíãëèéñëâi ýçûêà â àèéèñëíàðöý è Wordnet // Òðóäû NIÈÈÐÀÍ. 2011. Âüi. 19. N. 87–101. <http://scipeople.com/publication/106012/>

Êðèæàíññèé,

2011á — Êðèæàíññèé Á.Á. Íðåíéà èñíëüçñâàíèÿ êîðñíñâ è ýëåêòðñíûõ

âèáëèòòåâ â Ðóññêî Æèéèñëíàðå // Òðóäû ïåæäóíàðñííé êíîôåðåíöèè

«Êîðñíàÿ ëëíâèñòèâ–2011». Niñá., 2011. C. 217–222. <http://scipeople.com/publication/102432/>.

Ëåäâí,âà,

2008 — Ëåäâí,âà Á.Á. Ëåêñèéñðàôèÿ ñâðâàíññâi ðóññêâi ýçûêà. Íðàêòèêóì: Ó÷åá. ñññâèå. I., 2008.

Meyer and Gurevych,

2012 — Meyer C.M.,
Gurevych I..

Wiktionary: a new rival for expert-built lexicons? Exploring the
possibilities of collaborative lexicography // Electronic Lexicography.

Oxford: Oxford University Press. 2012. (â íå÷àòè). http://www.informatik.tu-darmstadt.de/fileadmin/user_upload/Group_UKP/publikationen/2011/oup-elex2012-meyer-wiktionary.pdf.

[1]

Ni. <http://code.google.com/p/wikokit/>