

Âíçìíæíñòè ïðèìåíåíèÿ óíèâåðñàëüííé ñèñòåìû ñèíøàêñè÷åñêíé ðàçìåòèè òåêñòà ObjectATI

Àâòîð Àíáà Âÿ÷åñëåââíà Ñàðàõðîà  
25.06.2008 á.  
Íñëåäíå Íàííàëåíèå 14.07.2008 á.

Óâcèñû à ôîðìàòå RTF (43.45 kB 2008-07-14 21:00:56) Óâcèñû à ôîðìàòå PDF (108.12 kB 2008-07-14 21:00:21)

Ñèñòåìà

Ծանոթագրական պատճենի մասին տեսական օրինակը հետևյալն է:

Íàøà

ନେଣେବୀଳା ଯାୟାଦିନ୍ୟ ତାଙ୍କୁ ଶେରାଏଇବେ ଏ ନିର୍ମାଣିତ ପାଇଁ ହେଉଥିଲା । ଲୋକଙ୍କ ନିର୍ମାଣିତ ପାଇଁ କାହାର ଜାଗରୂକତା ନାହିଁ । କାହାର ଜାଗରୂକତା ନାହିଁ ।

Ñèíòàêñè÷åñêèé

Àíáæéç, êâè èçâåñòíí, ííæåò ïðîâíäèòöñý íí ãðóííà è íí çâåèñèíñòý. Äëÿ óíâí ÷òíåú ðàçáèðåòü òåèñò íí çâåèñèíñòý, ííæí íðíñòí ñíçääòü ñíèñíê åñåð íåíáòíäèíüò òèííà ñèíòåèñè-åññéèò ñäýçåé è íá-÷åòü ñíçääàåòü («ñééèåèåàòü») ñèíòåèñè-åññéèå íáúååèòü, ñíñòíÿùèå èç íàðû ñèíâíòí &ndash; ååðøèíííé è ííä-÷éíáííé åé.

jiajii

çàíèìàòöñý àíàëèçí ïí áðóïïàì, ñíàéðàÿ öí÷íí òàèèå áæ ñíéíøàéñè÷åññéèå íáúåéòù,  
íí íá òíëüéî èç ñeiäíòí, íí è èç ñàìëö ñíéíøàéñè÷åññéèö íáúåéòù. Áëÿ ýòïäí òæå íaiáðïäèí ïðèìåíàíèå ìåðàíèçìà òàè  
íàçûåàåíùò íàëñòðíàé. íàëñòðíéå &ndash; ýóí ïíèñùåàåííà ïíëüçìàòåðéàí iííæñòåí  
ñeiäíòí è ñíéíøàéñè÷åññéèö íáúåéòù, íáéäàäþùèö ïíðåäåéäííùè ñaiéñòåàìè.  
íàïðèìàð, íàëñòðíéå «Ñéàçòåííà» áïèæíà áâéëþ÷åòù á ñíáÿ è íäíó ñíéíòí, è ñíéíøàéñè÷åññéèé íáúåéò, ñíñòïýùèé èç áëèàäíèå  
è çíàíàòåðéüíí ÷åñòè, ÷òíåù è òí, è áðóäíå, äííñòðè, ïíæíí áúéí ñíåäéíèòù ñ  
ïíäéäæàùèí äëÿ ñíçääíéÿ ííäíí íáúåéò &ndash; ñâýçé «ííäéäæàùåå&ndash;ñéàçòåííà».

lɪæʃɪ,

âîðîñ-åì, è êîíàéèåðîñâòù îàà ìåòîñä ñèíøåèñè-÷åññîåï áìàéèçà, êàè ýòi áåëåàåì ïü: íåéîòîñðûå ñèíøåèñè-÷åññèå áðòïïû ñîâèðåàåì öåéèèñî, íí äëÿ áîëüøèñòåà ñëó÷-ååâ íåðåàé-÷èååñïñ ìïñõðååíéåì ñâýçè íåæäó áåðøðéíàè.

Jäjàêî

åñèè iïðôîïeïäè ÷ åñêàÿ èíôîïàöèÿ î ñëïâîïîïàö äëÿ ðàçáèðàåìïäî òåêñòà

‘Idèñóòñòâòåò (òàê, äëÿ ðàçáéèðâàíûò íàìe «Í÷åëù» è Èéàâññéîé èäòïíèñè ñïðôïëíâèÿ óæå çàðâíåå ñääéàíá à ãðõáíé ñèñòåíå), òî ïðë ñèíóàéñè÷åññéîé ðàçìåòéå (÷òî ïí çàâéñèïñòÿ, ÷òî ïí ãðõííàì) åå ïíæíí èñííëüçíàòù, òåì ñàìùí ýòó ðàçìåòéó õíðîñüay èéè, ïí èðâéíåé íåðå, ñíèæäÿ ååðíýòíñòü îøéåéè.

Äëy

Ýóðáí íóæíí ðàééæå á èñtíëüçíàòòù íàðóàíéç íàäñòððíåé, ðàéé êåéé íí lïçáîëýåò çáääååàòö ñóñéíåèý íá lïððóíëíåè-÷åññéèå ñâíéñòåà ñéíåíòðíù, lïçáîëýþþùéå áé èåðåòö lïððåååéäåííþ ñéíóåéñè-÷åññéóþ ðíëü. lïðåäííëíæéì, æéý ñéíåíòðíù, áðíäýùåé á íàðñòðíééó Ñéàçóåííå, lïæíí lïðåíòðéòü, -ðí lïà áíéæåíá lïððåñòðååéýòü ñíåéé èé-÷íúé áééåáíé. lïæíí ðàééæå çáääååàòö ñóñéíåèý íá áðíæååíéý ðåéé ñéíóåéñè-÷åññéé áðóííü á íàðñòðíééó (éåéé lïðååéé, ýòí áúååþþ ñóñéíåèý íá åå áåððééíó). Ðàéé, lïæíí ñòðíóééòðíåàòü, -ðí lïà áéý ñíñòååííå Ñéàçóåííå áåððøéíé áééæåíá áúòù áééåáíé áúòé á èé-÷ííé ðíðíå, ðíòý áíçííæíú óñéíåèý è íá ðíëüéí íá áåððøéíó. Áéý óåíáñòåà lïëüçíàòååééå lïðé lïëñåíéè lïðáíü ñóñéíåéé lïððåñlïéåååðñý áåââñòè áéý ñéíðåéñè-÷åññéíå íáúååéò lïíþþéå áú-÷ññéýåíü lïëéé, ò.å. íàðåíéçìá, áúââíäýùåáí çíà-÷åíéý íàðåíåòððíå ñâíííå ñéíóåéñè-÷åññéíå íáúååéò á èç lïððåååéííü lïðåíåòððíå áåí lïëéé.

Íñèñûâàÿ

óñëèïàéý íà áôñïæääíèå ñëíäîòòðì è ñëíòàéñè-÷áññèèö íáúåéòíâ á òïðåäääéäííà ííïæåñòåí, íù ïïæái èñïíüçïåàöü ýòí ííïæåñòåí íå òïëüéí äëý ñïçäàíéý ëíåûö ñëíòàéñè-÷áññèèö íáúåéòíâ, íí è äëý áûçïåà èç áàçû ääííûö ñïèñéå áñåö íáúåéòíâ, óäíåéåòåíòýþùèö ñòîòíóèèðíàáííû íàé è óñëèäéý (íàéðèíàð, ñïçäàöü íàéñòðíééó «Ñíñòàáííà ñêàçóáííà ñ íóëåâíé ñâýçéíé» è ñòðåçó æå áçäéýíööü íå áñå ëíäíáíûå ñêàçóåííûå).

Òàêæå

Ííæíí çäääöö íá ðíëüéí ñóéñäéy áðííæääíéy ñéíåíöödöù eëe ñéíøåéñè÷åñééíí íáúåéòå  
â íáäñöödöééó, íí è íåðäíé÷åíéy ñáíííí ñéíøåéñè÷åñééíí íáúåéòå &ndash; óñéíåéy  
ñí÷åðäíéy ñáíéñöå â íáíí áðíäýüèo ñéíåíöödöì. Ñéàæåí, äëy ïíäéåæäúååí è  
ñéàçóåíííí íáíååðéåååöñý ñíäéåññíåíéå ìí eëöö ã ñéååöþúåí åéäå: åñéé eëöö  
ñéàçóåíííí íáðåíá, òí ïíäéåæäúååí eëéåí ðíäçå ñòðíèò à íåðåííì eëööå, eëéåí  
íðååñöååéyå ñíáíé ñéíøåéñè÷åñééé ìíëü, è ò. ä.

Âñå ýòî

íciáà-àáò, -òòí òàé èéàé iú ðáñáííéàáàí ííðòóíéíáè-÷áñéíé ðáçìáòòéíé è ñéñòåííé ííéñàíéý íðíñòåééòò ñéíòàéñé-÷áñééò ïðáàééë ýçùñéá (íññíáíúò ïðáàééë ñíñéàññíàáíéý è óíðáàéáíéý, ñâýçáíúò ñ ííðòíéíáè-÷áñééò íáéééíí ñéñàíòíòí, à òàéæá ñ ííðýäéíí ñéíàí, ðí òæáá áííðáçáí íáíúòá ááðíýòíñòö

äííóñòèòü íøéáéò íðé áíáééçá. Néíòàéñé-÷áñééé íáúáéò íá ñíçääñòñöý, áñéé åáí ÷áñòè íá áíéáóò á ñíñòåòñòåòþùéå íáññòðíééé (ñéàáéáí, ñóùåñòåéòåéüííå á êíññåáííí íá ááééåò áóááò òðåéòíáàòüñý ñéñòåííéé èéé ííäééæàùåá) èéé áñéé ýòé ÷áñòè áíéáóò, íí íäðåíé-÷áíéý ñàíííáí íáúáéòå íá ííçäíéýò ñíçääòü ñéíòàéñé-÷áñééé íáúáéò (áñéé, íáíðéíàò, èíý ñòòéò á èíáíéòåéüííí íááéæå, íí íá ñíñéàññíàáíí ñ èé-íúíí áéäáííéíí íí èéóó, èç íéò íá ííéó-÷éòñý ñíçääòü íáúáéò «Näýçü ííäééæàùåá&ndash;ñéäçóåííá»).

A

áðóáðúàì áëÿ ñèñðòàìù íæðíéððóåðñý ñíçääíéå íåðåáíèçà íåððåñ÷åðà áíàëèçà íí çåâéñèíñòþì á áíàëèç íí ãððíìà, ó.å.  
ñíçääíéå íåðåáíèçà ñáïðà ãððíìù íí áåððøèíå (íàïðèìåð, áñååñ íðîñòåñ îðåäëéíæðåéý íí åññ íñéàçóåññó).

Íàêîëåíàÿ

èíðîðìàöéý ï íðñþòåéðèò ñèíðàëéñè-÷-åññéèò íðåâéëæò ýçûéà ííæåò ííçáééòðü á áæäüúíáéðóál ñääéæðòù ñèíðàëéñè-÷-åññéèò ñáð ííðøåáðòíàðé-÷-åññéè, ð.å., èñííëüçöý èíñþòðóéðòð íáúåéðóíâ, ñðåçò ðåçå-÷-àðòù ííðååéëåííúå ðóðåäíåíòù òåéñòà ííñòðíóéèòðíåáííúå íðåäééàí.

Íàêííåö,  
òàè êàè íåéñòåíà, ñ êòîðûè íû èìåàí äåëî, íðåäñòàâëýþò ñíáíé  
ïåðåâíäû, ñèñòåíà ObjectATE íðåäóñíàòðèâàåò è ñðåäñòåà ïìèñaiéý ññòåâòñòåèé  
íàæåò íðéæíàéí è íåðåâíäû, ó.á. èññéåíäàíéý ðåðåéòåðå íåðåâíäà÷åñéíé äåýóåéüíñòè  
ñíçäàòåéäé íaiýòíèå. Áàçà ààííûò íi «Í÷åéå» ñíàðæèò íàðåéëåéüíûé åðå÷åñéèé  
òåéñò è ñïèñíè åðå÷åñéèò äåéñåí, ÷òî íïçâîéýåò ðàçìåò÷ééò ñíçäàâàòü íñíáûå  
íàúåéò ëéíåâèñò÷åñéíäà áíàéèçà, òàè íàçûâàíûò ðåðåâíäàòü íñíáûå  
íñíòåâòñòåèå íåíé èëè íåñéíèüéèì ñëàâýíñéèì ñëíàíòíàí íáíó èëè íåñéíèüéî  
ñéíàíòíàí íðèæíàéå, à òàéæå, åñèè íóæíí, íåéñåíû íðèæíàéå.

Ñièñíê  
ëèòåðàòóðû

Çíáíèí è äð. 2006 &ndash; Çíáíèí,  
Á. È. Óíèåâðñàéüíàÿ ñèñòåíà ðàçìåòéè òåéñòà ATE-2  
/ Á. È. Çíáíèí, Á. Á. Iàðéåéíà // Ñíàðåíàííûå  
èíôîðíàðèííûå òåðñíèíèè è íèñüíàííà íàñéåäèå: îò åðåâíéò ðóéíèñåé ê  
ýéâéòðííûì òåéñòàí : íàòåðèàéü íàæäóíàð. íàó÷í. êíîò., Èæåâñê, 13-17 èþëÿ 2006 á. &ndash; Èæåâñê, 2006. &ndash; Ñ. 51&ndash;

Possibilities for applying  
ObjectATE, a universal system for syntactic markup of text

Anna V. Sakharova

Vinogradov Institute of the Russian  
Language of the Russian Academy of Sciences, Moscow, Russia

The linguistic opportunities provided by Object  
ATE, a universal syntactic annotation system, are described. The system allows  
the user to perform morphological and syntactic analysis of ancient texts. In  
the future a semi-automatic parser of Old Russian may be created. In cases  
where the text analyzed is a translation, the system will describe the degree  
of conformity of the translation to the original.