Óíèâåðñàëüíàÿ ñèñòåìà ðàçìåòêè òåêñòà ATE-2

Àâòîð Ëåâêñåé Èãîðåâè÷ Çîáíèí
18.07.2008 ã.
Ïîñëåäíÿÿ îáíîâëåíèå 20.07.2008 ã.

Òåêñò ïå÷àòíîãî èçäàíèÿ â ôîðìàòå PDF

Ñèñòåìà ñèíòàêñè÷åñêîé ðàçìåòêè äðåâíåðóññêèõ òåêñòîâ, ñóùåñòâîâàâøàÿ â Èíñòèòóòå ðóññêîãî ÿçûêà èì. Â. Â. Âèíîãðàäîâà òåêñòà íàáîð ïðèçíàêîâ èç çàäàííîãî ñïèñêà, à òàêæå ñòðîèòü çàïðîñû [Çîáíèí 2005: 44–47]. Íåñìîòðÿ íà òî, ÷òî ñ åå ïîìîùüþ

– íåëüçÿ áûëî îáúåäèíèòü â åäèíûé ôðàãìåíò îòäåëüíî ðàñïîëîæåííûå ÷àñòè òåêñòà;

– íåâîçìîæíî áûëî óêàçàòü ðîëü îòäåëüíûõ ÷àñòåé ôðàãìåíòà âî âñåì ôðàãìåíòå; ñèñòåìà íå ìîãëà óñòàíàâëèâàòü ñâÿçè ì

Ó ïîëüçîâàòåëåé ñèñòåìû íå áûëî ÷åòêîãî ñîãëàñèÿ îòíîñèòåëüíî ãðàíèö âûäåëÿåìûõ ôðàãìåíòîâ. Â èòîãå â ðàçíûõ ÷àñòÿõ òåêñò ñóùåñòâèòåëüíî âî ôðàãìåíòå ‘ñóáñòàíòèâ àòðèáóò’).

Ñèñòåìà áûëà íàïèñàíà íà Microsoft Access è èìåëà íå î÷åíü óäîáíûé èíòåðôåéñ.

Áûëî çàìå÷åíî, ÷òî îñîáåííîñòüþ ñèíòàêñèñà äðåâíåðóññêîãî ÿçûêà ÿâëÿåòñÿ ñóùåñòâîâàíèå «íåëèíåéíûõ» ñèíòàêñè÷åñêèõ îòíî ðåàëüíûå ñèíòàêñè÷åñêèå îòíîøåíèÿ â òåêñòå, ïîýòîìó òðåáîâàëîñü ðåàëèçîâàòü ñåòåâóþ ñòðóêòóðó äàííûõ.

Íà ðàçðàáîò÷èêîâ ñèñòåìû áîëüøîå âëèÿíèå îêàçàëà ñèñòåìà «Ìàíóñêðèïò», ñîçäàííàÿ ãðóïïîé ðàçðàáîò÷èêîâ èç ÈæÃÒÓ è ÓäÃÓ 234–270]. Â Èíñòèòóòå ðóññêîãî ÿçûêà áûëà ïðåäïðèíÿòà ïîïûòêà ÷àñòè÷íî ðåàëèçîâàòü ìîäåëü ñèíòàêñè÷åñêîé ðàçìåòêè ñëèøêîì ãðîìîçäêîé. Âîçíèêàëî ìíîãî âîïðîñîâ î òîì, êàê ïðàâèëüíî íàäî ðàçìå÷àòü äàííûé ððàãìåíò, ÷òî ñâèäåòåëüñòâîâàëî î íåîäíîçíà

Â õîäå ýêñïåðèìåíòàëüíîé ðàáîòû áûëî ïðèíÿòî ðåøåíèå î òîì, ÷òî óêàçàíèå êîìïîíåíòîâ ñâÿçåé è èõ ñâîéñòâ äîëæíî ïðåäøåñòâîâàòü ñâÿçè, à çàïðîñû ê ñèñòåìå ñòðîèòü êàê óñëîâèÿ íà êîìáèíàöèþ òàêèõ áèíàðíûõ ñâÿçåé. Íî ïîñëåäíèé ïîäõîä ÿâëÿåòñÿ ïðîòèâîðå÷èâûì è ïîçäíî âîçèêíåò âîïðîñ î ñîâìåñòèìîñòè ñ äðóãèìè ñèñòåìàìè, òðåáóåòñÿ ñîçäàíèå îòäåëüíîãî ìîäóëÿ çàïðîñîâ.

Áûëî íàéäåíî ðåøåíèå, óñòðàíÿþùåå âñå ïåðå÷èñëåííûå íåäîñòàòêè è ñîõðàíÿþùåå âñå äîñòèíñòâà. Íà îñíîâå íîâîé ïðåäëàãàåìîé ñèíòàêñè÷åñêîé, ìîðôîëîãè÷åñêîé è ïðî÷èõ ðàçìåòîê òåêñòà. Ñôîðìóëèðóåì îñíîâíûå âîçìîæíîñòè, ðåàëèçîâàííûå â íîâîé ñèñòåìå ñêîâêàõ ìû áóäåì óêàçûâàòü àëüòåðíàòèâíûå íàçâàíèÿ, ïðèøåäøèå èç îáúåêòíî-îðèåíòèðîâàííîãî ïðîãðàììèðîâàíèÿ.

Ñèñòåìà ïîñòðîåíà íà îñíîâå «ãèáêèõ» ìàêåòîâ. Ïîëüçîâàòåëü èìååò âîçìîæíîñòü ñàìîñòîÿòåëüíî çàäàâàòü ñòðóêòóðó ðàçìåòêè, îï
íà åãî îñíîâå óæå îòîáðàæàåòñÿ è âåäåòñÿ ñàìà ðàçìåòêà.

Ñèñòåìà èñïîëüçóåò îáúåêòíî-îðèåíòèðîâàííûé ïîäõîä. Ìàêåò (ñòðóêòóðà ðàçìåòêè) ïðåäñòàâëÿåò ñîáîé íàáîð øàáëîíîâ (êëàññîâ)
ïîëÿìè, ýëåìåíòàìè øàáëîíà) èëè íàáîðàìè òàêèõ êîìïîíåíò. Êàæäàÿ êîìïîíåíòà øàáëîíà — ýòî ñíîâà íåêîòîðûé øàáëîí (èëè âñòî
òåêñòîâóþ êîìïîíåíòó ‘ñîäåðæàíèå’ (íîâà ìîæåò âû÷èñëÿòüñÿ ïî ïîðääëåíííûì ïðàâèëàì ÷åðåç ñîäåðæàíèå êîìïîíåíòîâ),
ñ óêàçàíèåì óæå ñóùåñòâóþùèõ åäèíèö â êà÷åñòâå ýëåìåíòîâ äàííîãî îáúåêòà øàáëîíà. Êðîìå èíôîðìàöèè î òèïàõ ñâîèõ êîìïîíåíòîâ
ïåðâîãî ïîðÿäêà, êîòîðûå äîëæíû áûòü èñòèíûìè äëÿ ëþáîãî îáúåêòà øàáëîíà. Ìîæíî ïðîâåñòè ñëåäóþùóþ àíàëîãèþ ñ ðåëÿöèîííûìè
âîçíèêàþò ñâÿçè «îäèí êî ìíîãèì» è «ìíîãî êî ìíîãèì»). Îãðàíè÷åíèÿ íàïîìèíàþò óñëîâèÿ öåëîñòíîñòè áàçû äàííûõ. Îäíàêî âîçìîæíîñòè îá
ñóùåñòâåííûì îòëè÷èåì ÿâëÿåòñÿ âîçìîæíîñòü íàñëåäîâàíèÿ îäíèõ øàáëîíîâ èç äðóãèõ (òî åñòü øàáëîí-íàñëåäíèê ïðèíàäëåæèò âñ
øàáëîíîâ. Êðîìå òîãî, ìîæíî çàäàòü ïðàâèëà ñîðòèðîâêè è ãðóïïèðîâêè ýëåìåíòîâ â íàáîðå êîìïîíåíò.

Â ñèñòåìå íåò ñòðóêòóðíûõ ðàçëè÷èé ìåæäó ýëåìåíòàìè ðàçìåòêè è çàïðîñàìè ê áàçå äàííûõ. Ëþáîé çàïðîñ ìîæåò áûòü ïðåäñòàâ
çàïðîñ — ýòî âñåãäà âûáîðêà ýêçåìïëÿðîâ ïîðääëåííîãî øàáëîíà. Îäíàêî, â îòëè÷èå îò ïðîöåññà ðàçìåòêè, ýëåìåíòû, âûáèðà

Îãðàíè÷åíèÿ, çàäàâàåìûå â øàáëîíàõ, äîëæíû áûòü ïðåäíàçíà÷åíû íå òîëüêî äëÿ ïîääåðæêè êîððåêòíîñòè, íî è äëÿ àâòîìàòèçàöèè
øàáëîíîâ, îáÿçàòåëüíûìè êîìïîíåíòàìè êîòîðûõ ýòè îáúåêòû ìîãóò ÿâëÿòüñÿ (ñ ñîáëþäåíèåì îãðàíè÷åíèé íà ñâîéñòâà ýòèõ îáúåêòîâ)
çíà÷åíèÿ äîïîëíèòåëüíûõ ñâîéñòâ (àòðèáóòîâ) ýòîãî îáúåêòà. Ñïèñîê äîïóñòèìûõ çíà÷åíèé êàæäîãî îáîáÿçàòåëüíîãî ñâîéñòâà ìîæåò
àâòîìàòè÷åñêè.

Ñèñòåìà èìååò âîçìîæíîñòü ïðåîáðàçîâàíèÿ îäíèõ ìàêåòîâ â äðóãèå. Òåì ñàìûì èñ÷åçàåò âîïðîñ î ñîâìåñòèìîñòè ôîðìàòîâ. Êàæäàÿ
ïðåîáðàçîâàíèÿ øàáëîíîâ îäíîãî ìàêåòà â øàáëîíû äðóãîãî.

Ñèñòåìà èìååò óäîáíûé èíòåðôåéñ. Â ðåçóëüòàòå âðåìÿ, çàòðà÷èâàåìîå ïîëüçîâàòåëåì íà ðàçìåòêó òåêñòà, ñóùåñòâåííî ñîêðàù

Øàáëîí íàïîìèíàåò ìíîãîîêíîïðóíó ñâÿçü èç ñèñòåìû «Ìàíóñêðèïò», îäíàêî, â îòëè÷èå îò ñâÿçè, îí èìåòò ñîáñòâåííûé íàáîð ñâîéñòâ è î
ñìîäåëèðîâàòü â ðàìêàõ îîâîãî ïîäõîäà. Îíæíî ñîçäàòü ìàêåòû è äëÿ ìîäåëè áèíàðíûõ ñâÿçåé, è äëÿ ïðåäûäóùåé ìîäåëè ðàçìåòêè ñè

Î÷åðåäíàÿ âåðñèÿ íîâîé ñèñòåìû ATE 2 (Ancient Text Editor 2) ðåàëèçîâàíà íà ïëàòôîðìå Microsoft .NET ñ èñïîëüçîâàíèåì ðåëÿöè
Microsoft SQL Server.

Summary

We present a new version of the Ancient Text Editor (ATE 2) which is to replace the previous version of the syntactical editor used in Vinogradov Institute for Russian Language of RAS. It relies on the object-oriented concepts and allows the user creation of his own markup templates. These templates define the rules of creating, managing, sorting, grouping and viewing the objects. The system can transform data from one set of templates into another. The program is based on the Microsoft .NET technology and relational database management systems.

Ñïèñîê ëèòåðàòóðû

Áàðàíîâ è äð. 2003a Áàðàíîâ Â. À. Ñïåöèàëèçèðîâàííé òåêñòîâûé ðåäàêòîð «Ìàíóñêðèïò» Ñèñòåìû îáðàáîòêè äðåâíèõ ðóêîïèñ «Èñòîðèÿ è êîìïüþòåð». — 2003. — ¹ 31. Ñ. 159–165. (http://manuscripts.ru/mns/docs/AIK2003b.pdf).

Áàðàíîâ è äð. 2003b Áàðàíîâ, Â. À. Ýëåêòðîííûå èçäàíèÿ äðåâíèõ ïèñüìåííûõ ïàìÿòíèêîâ è òåõíîëîãèÿ ñîçäàíèÿ ïîëíîòåêñòîâûõ áàç èñòîðè÷åñêîé èíôîðìàòèêè : òð. VIII êîíô. Àññîöèàöèè «Èñòîðèÿ è êîìïüþòåð» / ïîä ðåä. Ë. È. Áîðîäêèíà, Â. Í. Âëàäèìèðî-âà. &mda (http://manuscripts.ru/mns/docs/AIK2003.pdf).

Çîáíèí 2005 Çîáíèí, À. È. Êîðïóñ äðåâíåðóññêèõ ïåðåâîäîâ XI–XII âåêîâ: ðåçóëüòàòû è ïåðñïåêòèâû / À. È. Çîáíèí, À. À. 2005. — Ñåð.2. — ¹ 3. — Ñ. 44–47.

Êàãàðëèöêèé è äð. 2003 Êàãàðëèöêèé, Þ. Â. Ïðîáëåìû ñîçäàíèÿ ýëåêòðîííîãî êîðïóñà ïåðåâîäîâûõ ïàìÿòíèêîâ äðåâíåðóññêîé ïèñü ëèíãâèñòèêà â Ðîññèè. — 2003. — Âûï. 10. — Ñåð. 2.