Ýëåêòðîííîå êðèòè÷åñêîå èçäàíèå ñðåäíåâåêîâîãî òåêñòà: èíñòðóìåíòàëüíûå ñðåäñòâà âèç

Àâòîð Ñåðãåé Âèêòîðîâè÷ Äóáîâöåâ
25.09.2009 ã.
Ïîñëåäíÿÿ îáíîâëåíèå 01.10.2009 ã.

Òðàäèöèîííîå ïå÷àòíîå êðèòè÷åñêîå èçäàíèå ñðåäíåâåêîâîãî ïàìÿòíèêà ïèñüìåííîñòè ïðåäñòàâëÿåò ñîáîé íàó÷íóþ ïóáëèêàöèþ, â êîòîðîé îñîáåííîñòè òåêñòà ïðåäñòàâëåíû åãî âàðèàíòàìè â äîøåäøèõ äî íàñòîÿùåãî âðåìåíè ñïèñêàõ.

Íàöåëåííîå íà ñîîîñòàâëåíèå âàðèàíòîâ, èçäàíèå ñòðóêòóðèðîâàíî òàêèì îáðàçîì, ÷òî ïîçâîëÿåò óâèäåòü ðàçëè÷èÿ ìåæäó ðóêîïèñÿìè, âîçíèêøèå â ðåçóëüòàòå ïðàâêè, ðåäàêòèðîâàíèÿ, âñòàâîê, óòðàò, âî âðåìÿ ïåðåïèñûâàíèÿ, â êîíå÷íîì ñ÷åòå - èññëåäîâàòü èñòîðèþ òåêñòà, òåêñòîëîãè÷åñêèå, ëèíãâèñòè÷åñêèå è èíûå ïðèçíàêè êàæäîãî êîíêðåòíîãî ñïèñêà. À ïðè íàëè÷èè òåêñòà íà ðàçíûõ ÿçûêàõ - óñòàíîâèòü çíà÷èìîå äëÿ èñòîðèêà, ëèíãâèñòà, êóëüòóðîëîãà ñîîòíîøåíèå ïåðâîîäà è îðèãèíàëà.

Ñóùåñòâóþùèå êðèòè÷åñêèå èçäàíèÿ, íåñìîòðÿ íà çíà÷èòåëüíóþ âàðèàòèâíîñòü â ñïîñîáàõ ïîäà÷è ìàòåðèàëà, óñòîé÷èâû â îñíîâíûõ ñâîèõ ÷åðòàõ: íàëè÷èå îäíîãî îñíîâíîãî òåêñòà, ïîäâåäåíèå ðàçíî÷òåíèé ïî îòíîøåíèþ äëÿ ýòîãî ñïèñêà, âêëþ÷åíèå ïðèëîæåíèé â âèäå ñïðàâî÷íèêîâ, óêàçàòåëåé, êîìåíòàðèåâ, à òàêæå, â ñëó÷àå ïåðâîïàäíîñòè òåêñòà, îðèãèíàëà íà äðóãîì ÿçûêå.

Â òî æå âðåìÿ, åñëè ó ÷èòàòåëÿ ñóùåñòâóåò íåîáõîäèìîñòü âûáîðêè ëèíãâèñòè÷åñêèõ äàííûõ, íå ïðåäñòàâëåííûõ â ÿâíîì âèäå ñïèñêà, ïå÷àòíàÿ ôîðìà ïóáëèêàöèè òðåáóåò äîïîëíèòåëüíîé ðàáîòû ïî îòáîðó è ïðåãðóïïèðîâêå ìàòåðèàëà.

Â ðàìêàõ ïðîåêòà «Ìàíóñêðèïò» äâà ãîäà íàçàä áûëà íà÷àòà ðàáîòà ïî ñîçäàíèþ ýëåêòðîííîãî àíàëîãà ïå÷àòíîãî êðèòè÷åñêîãî èçäàíèÿ (äàëåå - ÝÊÈ), êîòîðîå äîëæíî ïðåäîñòàâëÿòü ÷èòàòåëþ (ïîëüçîâàòåëþ) è òå âîçìîæíîñòè çíàêîìñòâà è ðàáîòû ñ òåêñòîì, êîòîðûå äàåò òðàäèöèîííàÿ ïå÷àòíàÿ ïóáëèêàöèÿ, è òàêèå, êîòîðûå â íåé îòñóòñòâóþò. Òðåáîâàíèÿ ê ÝÊÈ áûëè èçëîæåíû â ðàáîòå [Áàðàíîâ-Ãóòèêîâ 2008].

Â íàñòîÿùåå âðåìÿ íà
ïîðòàëå «Ìàíóñêðèïò: ñëàâÿíñêîì ïèñüìåííîå íàñëåäèå» äåìîíñòðèðóåòñÿ ïåðâàÿ
âåðñèÿ ÝÊÈ (URL: http://manuscripts.ru/mns/cred.cred_test), ïðåäñòàâëÿþùàÿ òåêñò
ñëóæåáíîé ìèíåè íà ìàé ïî ñëàâÿíñêèì ñïèñêàì XI-XIV ââ.

Ñ ñàìîãî íà÷àëà àâòîðû
(ïîñòàíîâêà çàäà÷è è ëèíãâèñòè÷åñêàÿ ïîäãîòîâêà äàííûõ -
Â. À. Áàðàíîâ, ðàçðàáîòêà è ñîçäàíèå âåá-ìîäóëåé -
Ñ. Â. Äóáîâöåâ) ïðè ðàçðàáîòêå ïîëüçîâàòåëüñêèõ èíòåðôåéñîâ îðèåíòèðîâàëèñü
íà òðàäèöèè ïå÷àòíûõ êðèòè÷åñêèõ èçäàíèé (ðèñ. 1). Â òî æå âðåìÿ âîñïðîèçâåäåíèå
ñòðàíèöû ïå÷àòíîãî èçäàíèÿ áûëî ëèøü îäíîé èç çàäà÷ ïðè ðàçðàáîòêå èíòåðôåéñà. Îñíîâíîé
öåëüþ ÿâëÿåòñÿ ïðåäîñòàâëåíèå ïîëüçîâàòåëþ ðàçâèòîãî èíñòðóìåíòàðèÿ äëÿ îòáîðà
(âûáîðêè) ìàòåðèàëà, äëÿ åãî ãðóïïèðîâêè (óïîðÿäî÷åíèÿ) è äëÿ âèçóàëèçàöèè íà
ýêðàíå.

Ðèñ. 1: Äåìîíñòðàöèÿ ðàçíî÷òåíèé

Îñíîâíûå îñîáåííîñòè ïîäãîòîâëåííîãî
ÝÊÈ. ÝÊÈ, ðåàëèçîâàííîå êàê ðÿä âåá-èíòåðôåéñîâ äëÿ âûáîðêè, ãðóïïèðîâêè è
âèçóàëèçàöèè äàííûõ, äîñòóïíî ÷åðåç Èíòåðíåò (òðåáóåòñÿ ðåãèñòðàöèÿ íà ïîðòàëå
«Ìàíóñêðèïò: ñëàâÿíñêîå ïèñüìåííîå íàñëåäèå). Ôîðìû çàïðîñà ïîçâîëÿþò âàðüèðîâàòü
ôîðìó âèçóàëèçàöèè ìàòåðèàëà: âûáèðàòü êîëè÷åñòâî ðóêîïèñåé è ïîðåäåëÿòü îñíîâóþ
ñðåäè íèõ, óêàçûâàòü åäèíèöû ñðàâíåíèÿ è ôîðìû èõ ðàñïîëîæåíèÿ, ãðóïïèðîâêè, - ïîëó÷àòü
â êà÷åñòâå ðåçóëüòàòà âûáîðêó ðàçíîãî íáúåìà è êîëè÷åñòâåííûõ õàðàêòåðèñòèê. Íåñêîëüêî
ïàðàìåòðîâ, îïðåäåëÿåìûå ïîëüçîâàòåëåì ïðè ïîäãîòîâêå çàïðîñà, ïîçâîëÿþò î÷åíü
ãèáêî íàñòðîèòü âûâîðêó è ïîëó÷èòü òàêîé åå âèä, êîòîðûé è ïî ñîäåðæàíèþ, è ïî
ôîðìå óäîâëåòâîðÿåò ïîñòàâëåííé èññëäîâàòåëüñêîé çàäà÷å.

Â îñíîâå ïîäãîòîâëåííîãî
â íàñòîÿùåå âðåìÿ ÝÊÈ - ïîëíîòåêñòîâàÿ áàçà äàííûõ íåñêîëüêèõ ñïèñêîâ ìàéñêîé
ñëóæåáíîé èíåè, êîòîðàÿ ðåàëèçîâàíà íà ïëàòôîðìå ÑÓÁÄ Oracle, ïîçâîëÿþùåé ëåãêî ïðîèçâîäèòü
íåîáõîäèìóþ îáðàáîòêó äàííûõ ñ ïîñëäóþùèì âûâîäîì îò÷åòà â âèäå âåá-ñòðàíèöû.
Â îñíîâå îò÷åòà âñåõ ðåàëèçîâàííûõ ðåæèìîâ ëåæèò HTML-òàáëèöà, êîòîðàÿ çàïîëíÿåòñÿ ñâåðõó
âíèç è ñëåâà íàïðàâî. Çàïîëíåíèå â èíîé ïîñëääîâàòåëüíîñòè âîçìîæíî, íî
çàòðóäíåíî è, íà íàø âçãëÿä, ïðîòèâîðå÷èò ñàìîé èäåîëîãèè òàáëèöû. Â ñâÿçè ñ
ýòèì è íåêîòîðûìè äóõãèì ôàêòîðàìè îñëä ðÿäà ýêñïåðèìåíòîâ áûëî ðåøåíî
ðàçäåëèòü ôîðìèðîâàíèå îò÷åòà íà íåñêîëüêî ýòàïîâ.

Íà ïåðâîì ýòàïå ñòðîèòñÿ
ñòðóêòóðà äàííûõ, ñîäåðæàùàÿ òî, ÷òî ìîæåò áûòü èñïîëüçîâàíî íà ïîñëääóþùèõ ýòàïàõ.
Ýòî ïîçâîëÿåò åäèíîæäû âûáðàííûå äàííûå èñïîëüçîâàòü ìíîãîêðàòíî, íå çàïðàøèâàÿ

èõ êàæäûé ðàç èç áàçû äàííûõ ïðè íåîáõîäèìîñòè. Òàêæå íà ïåðâîì ýòàïå ôîðìèðóåòñÿ âèäèìîå ïðåäñòàâëåíèå ñëîæíûõ åäèíèö (íàïðèìåð, ñëîæíûõ ñëîâîôîðì è ñèíòàêñè÷åñêèõ åäèíèö), èíòåãðèðóåìûõ êàê îäíî öåëîå. Äëÿ åäèíèö ñðàâíèâàåìûõ ðóêîïèñåé íà ýòîé ñòàäèè ïîðåäåëÿþòñÿ ñîîòâåòñòâóþùèå èì åäèíèöû ðóêîïèñè, âûáðàííîé â êà÷åñòâå îñíîâíîé. Äëÿ óñòàíîâêè òàêèõ ñâÿçåé èñïîëüçóåòñÿ ñïåöèàëüíàÿ ñòðóêòóðà - ïðîòîòåêñò (ïðîòîòåêñò - ñîçäàííûé àâòîðîì ýëåêòðîííîãî èçäàíèÿ òåêñò, åäèíèöàì êîòîðîãî â ðóêîïèñÿõ ñîîòâåòñòâóþò òåêñòîâûå ïðåöåäåíòû). Íóæíî îòìåòèòü, ÷òî ñâÿçè ìåæäó îáðàáàòûâàåìûìè åäèíèöàìè ìîãóò ïðîõîäèòü ÷åðåç ðàçíîå ÷èñëî óðîâíåé èåðàðõèè, êàê, íàïðèìåð, â ñëó÷àå ïðîñòûõ è ñëîæíûõ åäèíèö, â êîòîðûõ è â òåêñòå, è â ïðîòîòåêñòå ÷èñëî «øàãîâ» îò åäèíèöû äî åäèíèöû ìîæåò áûòü ðàçíûì. È äëÿ òîãî, ÷òîáû èçáåæàòü ïðèâÿçêè ê îäííîó êîíêðåòíîó âàðèàíòó, ôóíêöèè, èñïîëüçóåìûå äëÿ îïðåäåëåíèÿ ñâÿçåé, áûëè ðåàëèçîâàíû êàê ðåêóðñèâíûå.

Íà âòîðîì ýòàïå åäèíèöû ñëîâàðÿ
ôðàãìåíòîâ, íà îñíîâå êîòîðîãî ïîðåäåëÿþòñÿ ñâÿçè ìåæäó ñîîòâåòñòâóþùèìè äðóã äðóãó ôðàãìåíòàìè ðóêîïèñåé, îáðàáàòûâàþòñÿ óæå ïîî÷åðåäîî. Ïîðÿäîê èõ îáðàáîòêè îïðåäåëÿåòñÿ ïîðÿäêîì ñëåäîâàíèÿ ôðàãìåíòîâ â îñíîâíîé ðóêîïèñè. Íà ýòîì øàãå ïðè ïîìîùè ñïåöèàëüíî ðàçðàáîòàííîãî àëãîðèòìà îïðåäåëÿåòñÿ îïòèìàëüíîå ðàñïîëîæåíèå åäèíèö èñõîäÿ èç ïðîñòîãî ïðàâèëà: âñå åäèíèöû, ïðèâÿçàííûå ê îäííé åäèíèöå ïðîòîòåêñòà, ïðè âèçóàëèçàöèè äîëæíû íàõîäèòüñÿ â îäííé ñòðîêå òàáëèöû. Åñëè æå ïî êàêèì-èèáî ïðè÷èíàì ýòî ñäåëàòü íå óäàåòñÿ, òî ñîîòâåòñòâèå ïîêàçûâàåòñÿ ãðàôè÷åñêè, â âèäå ëèíèé, ñîåäèíÿþùèõ ñîîòâåòñòâóþùèå ñëîâîôîðìû. Ãëàâíàÿ èç ýòèõ ïðè÷èí - ðàçëè÷èå ïîðÿäêà ñëåäîâàíèÿ åäèíèö â ðóêîïèñÿõ. Â òîì ñëó÷àå, åñëè â êà÷åñòâå åäèíèöû ñðàâíåíèÿ âûñòóïàåò ôðàãìåíò ðóêîïèñè, òî àíàëîãè÷íûé àëãîðèòì èñïîëüçóåòñÿ äëÿ óñòàíîâëåíèÿ ñîîòâåòñòâèé ìåæäó ôðàãìåíòàìè íà îñíîâå ñëîâàðÿ ôðàãìåíòîâ (ðèñ. 2).

Ðèñ. 2: Âèçóàëèçàöèÿ ñîîñòàâà
ðóêîïèñåé è ïîðÿäêà ñëåäîâàíèÿ ôðàãìåíòîâ

Îòäåëüíîé çàäà÷åé ñòàëà è
îáðàáîòêà ïîâòîðÿþùèõñÿ â îäííé ðóêîïèñè ôðàãìåíòîâ, ñîäåðæàùèõ îäèí è òîò æå òåêñò. Íàëè÷èå ïîâòîðÿþùèõñÿ ôðàãìåíòîâ âëå÷åò çà ñîáîé ïîÿâëåíèå íåñêîëüêèõ äèíèö äîêóìåíòà, ïîðèâÿçàííûõ ê îäííîó ôðàãìåíòó ñëîâàðÿ. Çàäà÷à âèçóàëèçàöèè òàêîé ñèòóàöèè áûëà óñïåøíî ðåøåíà ïðè ïîìîùè äîïîëíèòåëüíûõ îêîí, ñîäåðæàùèõ èíôîðìàöèþ, îòñóòñòâóþù ê îñíîâíîì îêíå-îò÷åòà. Ðåøåíèå ïîçâîëèëî îòêàçàòüñÿ îò ïîèñêà ïóòåé ñîàìåùåíèÿ â îäííé ïðîöåäóðå äâóõ ðàçëè÷íûõ âàðèàíòîâ âûâîäà - áåç ïîâòîðÿþùèõñÿ ôðàãìåíòîâ è ñ íèìè, à ðåàëèçîâàòü êàæäûé èç íèõ ïî îääëüíîñòè, è êàê ñëåäñòâèå - ïðåäîñòàâèòü ïîëüçîâàòåëþ ïîíÿòíûé èíòåðôåéñ, ãäå íà îêíà ñ äîïîëíèòåëüíîé èíôîðìàöèåé äàþòñÿ ññûëêè â îñíîâíîì îêíå-îò÷åòà (ðèñ. 3).

Ðèñ. 3: Ðàçëè÷íûå ðåæèìû âèçóàëèçàöèè
ïîâòîðÿþùèõñÿ ôðàãìåíòîâ

Â ïîñëåäíåì ýòàïå
ðåøàëàñü çàäà÷à ãðàôè÷åñêîãî ïîêàçà ñîîòâåòñòâèé.

Äëÿ ðåàëèçàöèè áûëî
ðåøåíî èñïîëüçîâàòü ÿçûê ðàçìåòêè ìàñøòàáèðóåìîé âåêòîðíîé ãðàôèêè (SVG), òàê êàê âåêòîðíûå èçîáðàæåíèÿ ïðè
ìàñøòàáèðîâàíèè íå òåðÿþò êà÷åñòâà. Ïðè ïîìîùè ýòîãî èíñòðóìåíòà íà óæå
ñóùåñòâóþùèõ ñòðàíèöó íàêëàäûâàþñÿ ëèíèè, ñîåäèíÿþùèå ñîîòâåòñòâóþùèå äðóã
äðóãó åäèíèöû. ×òîáû èçáåæàòü èçáûòî÷íîñòè, ãîðèçîíòàëüíûå ëèíèè íå ïîêàçûâàþòñÿ.
Íà äàííûé ìîìåíò ðåàëèçîâàíî äâà ñïîñîáà âûâîäà ëèíèé: ñîåäèíåíèå åäèíèöû
îñíîâíîé ðóêîïèñè ñ ñîîòâåòñòâóþùèì åäèíèöàìè ñðàâíèâàåìûõ è ñîåäèíåíèå åäèíèöû
ñ ñîîòâåòñòâóþùåé â ñëåäóþùåì ñòîëáöå, ÷òî ïîçâîëÿåò ñôîðìèðîâàòü «ïóòü»
ïåðåìåùåíèÿ åäèíèöû.

Èç-çà òîãî, ÷òî ñàì ïî
ñåáå áðàóçåð íå èìååò íåîáõîäèìûõ äëÿ ðåøåíèÿ äàííîé çàäà÷è ñðåäñòâ
èíòåðïðåòàöèè, äëÿ ðàáîòû ñ ñîçäàííûì ÝÊÈ íåîáõîäèìî óñòàíîâèòü ñîîòâåòñòâóþùèé
áåñïëàòíûé ïëàãèí. Äëÿ óäîáñòâà íàõîæäåíèÿ ñîîòâåòñòâóþùèõ äðóã äðóãó åäèíèö â
ðàçëè÷íûõ îáëàñòÿõ îêíà îò÷åòà ñ ïîìùüþ èíñòðóìåíòà Javascript ðåàëèçîâàíà ïîäñâåòêà ñâÿçàííûõ îáúåêòîâ
è âîçìîæíîñòü ôèêñàöèè âûäåëåíèÿ (ðèñ. 4).

Ðèñ. 4: Âèçóàëèçàöèÿ ñîñòàâà
ôðàãìåíòîâ è ñëåäîâàíèÿ â íèõ èíîãâèñòè÷åñêèõ äèíîâè

Â íàñòîÿùèé ìîìåíò ðàçðàáîòàííûé
èíñòðóìåíò èñïîëüçóåòñÿ äëÿ ïîäãîòîâêè åùå äâóõ ýëåêòðîííûõ ïóáëèêàöèé (ÝÊÈ
«Åâàíãåëèå», ÝÊÈ «Ðóññêèå ëåòîïèñè»), êîòîðûå ïîñëå îêîí÷àíèÿ ëèíãâèñòè÷åñêîé
ðàáîòû íàä íèì áóäóò ïðåäñòàâëåíû íà ïîðòàëå «Ìàíóñêðèïò: ñëàâÿíñêîå
ïèñüìåííîå íàñëåäèå».

Áëàãîäàðíîñòè

Ðàáîòà âûïîëíÿåòñÿ ïðè
ôèíàíñîâîé ïîääåðæêå ÐÃÍÔ (ïðîåêò ¹ 07-04-00369à).

Ëèòåðàòóðà

Áàðàíîâ-Ãîóòèêîâ 2008 - Áàðàíîâ, Â. À. Ýëåêòðîííîå
êðèòè÷åñêîå èçäàíèå ñðåäíåâåêîâîãî òåêñòà: ïîñòàíîâêà çàäà÷è, îñíîâíûå
òðåáîâàíèÿ è èíñòðóìåíòàëüíàÿ ïîäãîòîâêà / Â. À. Áàðàíîâ, Ð. Ì. Ãîóòèêîâ //
Ñîâðåìåííûå èíôîðìàöèîííûå òåõíîëîãèè è ïèñüìåííîå íàñëåäèå: îò äðåâíèõ òåêñòîâ
ê ýëåêòðîííûì áèáëèîòåêàì [Òåêñò] : ìàòåðèàëû ìåæäóíàð. íàó÷. êîíô. (Êàçàíü,
26-30 àâãóñòà 2008 ã.) / îòâ. ðåä. Â. À. Áàðàíîâ,
Â. Ä. Ñîêîâüåâ. - Êàçàíü : Èçä-âî ÊÃÓ, 2008. -
Ñ 37-44.