# Ðàçðàáîòêà è èñïîëüçîâàíèå ôîðìàòà mnsXML äëÿ ïîäãîòîâêè, îáðàáîòêè è îáìåíà ýëåêòðîííû

Àâòîð Âèòàëèé Àëåêñàíäðîâè÷ ÐÎÌàíåíêî
30.06.2008 ã.
Ïîñëåäíåå îáíîâëåíèå 27.07.2008 ã.

Òåçèñû â ôîðìàòå RTF (35.49 kB 2008-07-15 23:07:43) Òåçèñû â ôîðìàòå PDF (46.06 kB 2008-07-15 23:07:17)

Â ìèðå ñóùåñòâóåò óæå íåñêîëüêî ýëåêòðîííûõ ïîëíîòåêñòîâûõ êîëëåêöèé äðåâíèõ ñëàâÿíñêèõ ïèñüìåííûõ ïàìÿòíèêîâ, ïîäãîòî
êîëëåêòèâà èëè îòäåëüíîãî ó÷åíîãî, ïîääåðæèâàþùèõ êîëëåêöèþ, òàê è ïóòåì âçàèìîâûãîäíîãî îáìåíà ýëåêòðîííûìè íàáîðàìè.
Òðóäîåìêîñòü ïåðâîãî ïóòè – ïåðâè÷íûé ââîä òåêñòà è åãî ñâåðêà – î÷åíü âåëèêà è òðåáóåò ïðèâëå÷åíèÿ âûñîêî
îáðàáîòêè è âèçóàëèçàöèè îäíîãî è òîãî æå ìàòåðèàëà (òåêñòîâ, ïåðå÷íåé, âûáîðîê), íî òðåáóåò ñîçäàíèÿ ôîðìàòîâ îáìåíà äàííûìè
ñîçäàíèþ ôîðìàòà îáìåíà ýëåêòðîííûìè òåêñòàìè íà îñíîâå óíèâåðñàëüíîãî ÿçûêà ðàçìåòêè XML è ôîðìàòà ðàçìåòêè òåêñòîâ TEI (T
ñâåäåíèé î òåêñòîëîãè÷åñêèõ è ëèíãâèñòè÷åñêèõ åäèíèöàõ òåêñòîâ è ðóêîïèñåé â èíôîðìàöèîííî-àíàëèòè÷åñêóþ ñèñòåìó (ÈÀÑ)
äðóãèìè ïîëíîòåêñòîâûìè êîëëåêöèÿìè. Îñíîâíûå õàðàêòåðèñòèêè ôîðìàòà mnsXML: - íåáîëüø÷åíèå öèêëà çàãðóçêè-âûãðóçêè
åãî ðàçìåòêè, áíàëèòè÷åñêèì îïèñàíèåì è ññûëêàìè íà ñèîíâåðè, - âîçìîæíîñòü âûãðóçêè, çàãðóçêè è ððàíåíèÿ ñëîâàðåé, - ðàí
ñâÿçè ñ èçìåíåíèåì ôîðìàòîâ âíåøíèõ êîëëåêöèé. Ïðåäïîëàãàåòñÿ íåñêîëüêî ñõåì èñïîëüçîâàíèÿ ôîðìàòà mnxXML. 1. Çàãðóçêà
âêëþ÷àåò â ñåáÿ êîäèðîâêó ñèìâîëîâ òåêñòà è ñòðóêòóðó òåêñòà â âèäå îäíîé èëè íåñêîëüêèõ ðàçìåòîê. Îñíîâíóþ òðóäíîñòü ïðè
ïîçâîëèò çíà÷èòåëüíî ñíèçèòü òðóäîçàòðàòû ïî çàãðóçêâ íîâûõ òåêñòîâ â ÈÀÑ «Ìàíóñêðèïò». Äîïîëíèòåëüíî ïîÿâëÿåòñÿ âîçìîæíîñò
çàãðóçêè â áàçó äàííûõ «Ìàíóñêðèïò» òåêñòîâ Ì. Â. Ëîìîíîñîâà, ïðåäâàðèòåëüíî â ôîðìàòå ASCII, ðàçìå÷åííûõ äëÿ ñ
áûòü ïðåääàíû â äóãèå êîëëåêöèè èëè çàãðóæåíû ïîâòîðíî. Ñîçäàíèå ðåçåðâíûõ êîïèé òåêñòîâ ïîçâîëÿåò ïðè íåîáõîäèìîñòè âåðí
èñïîëüçîâàòü òàêæå äëÿ âûãðóçêè èç ÈÀÑ «Ìàíóñêðèïò» òåêñòîâ Ì. Â. Ëîìîíîñîâà äëÿ ïðåääà÷è èõ äóãèì êîëëåêòèâàì. 3. Îáìåí
êîëëåêöèþ) è äâóñòîðîííèì (ñîâìåñòíàÿ ðàáîòà íàä òåêñòîì). Îäíîé èç ïðîáëåì ïðè îáìåíå òåêñòàìè ìæäó êîëëåêöèÿìè ÿâëÿåòñÿ è
ñêâîçíîé óíèêàëüíûé íîìåð äëÿ âñåé êîëëåêöèè, ñîñòàâíîé êëþ÷ èç íîìåðà òåêñòà è íîìåðà åäèíèöû â òåêñòå è äð.), ÷òî äåëàåò íå
êîëëåêöèþ. Äóãàÿ ïðîáëåìà – ñîãëàñîâàíèå èçìåíåíèé, âíîñèìûõ â ðàçíûå êîïèè îäíîãî òåêñòà. Îññåå èçìåíåíèÿ òåêñòà â ó
ñëèÿíèÿ èçìåíåíèé. Â òàêîé ñèòóàöèè îïòèìàëüíûì áóäåò áëîêèðîâàíèå èçìåíåíèé òåêñòà â èñõîäíîé êîëëåêöèè, ïîêà îí ïðåäàâàí äëÿ
ñóùåñòâîâàâøåé ðàíåå åäèíèöû ïðîñõîäèò îñîñòî îáíîâëåíèå åå àòðèáóòîâ, òî äëÿ íîâîé åäèíèöû, ñîçäàííé â õîäå ðåäàêòèðîâàíè
âîññòàíîâëåíèåì ñâÿçåé ñî ñëîâàðÿìè. Îáìåí äíàëèòè÷åñêèì äàííûìè î òåêñòå îñîáûõ ñëîæíîñòåé íå ïðåäñòàâëÿåò, åñëè äëÿ äíà
ñòóêòóðû äíàëèòè÷åñêèõ äàííûõ â ôîðìàòå TEI ïîçâîëÿåò âçÿòü åãî çà îñíîâó. Îáìåí ñëîâàðÿìè è ññûëêàìè íà íèõ ÿâëÿåòñÿ íàè
ñëîâàðíûõ åäèíèö â ðàçíûõ êîëëåêöèÿõ), è òîëüêî ïîòîì âîçìîæåí íàìåí òåêñòàìè ñ ñîõðàíåíèåì ññûëîê íà ñëîâàðíûå åäèíèöû. Áëàã
12147â («Áîëüøîé êîðïóñ ðóññêîãî ÿçûêà XVIII â.»). The development and use of mnsXML for preparing, processing, and exchange of electronic texts Vitaliy A. Pomanenko, Alexey N. Zhelonkin Udmurtia State University, Izhevsk, Russia
The usage of mnsXML as a storage format for electronic text for the Manuscript information-analytical system and for interchange with other electronic full-text collections of ancient texts is described.