

Èääåïëñäëÿ è òåõíèêà ðàçìåòêè â Òåõíïëñäëè ñìåðàíñäî íàáîðà

Àâòîð Àëåêñàíäð Âèêòîðîâè÷ Êîâàëåíèí
27.06.2008 ã

27.06.2008 á.

Íñëåäíåå íáííâëåíèå 16.07.2008 å.

Óâcêñú â ôïðìàòå RTF (57.69 kB 2008-07-13 21:38:16) Óâcêñú â ôïðìàòå PDF (122.14 kB 2008-07-13 21:37:33)

Òåõíïëîãèÿ

Аëй ðаàïòú ñ òàéëìè òåêñòàïè ñðåáóàðñý ðåðåíèå ñò, õ áçàèïññâýçàíûò çààä:

1) hīcāiaéa áaeíie ieaó-óidú áey iðaañoaáéaíey á iaiii ðaaéea ðaçíiaí ðiaá ðoðaaíiaíóia; 2) iæðiæááieá hiiñiaá ðaçáeáíey ðaaéenòa ía ðoðaaíiaíóu; 3) iðaañiñðaaéaíey iñeüçiaðaaðep niiñiaá iñeñiaíey eíøaðiðaaðaaðeë ðoðaaíiaíóa.

Öåëü äàííé

1. Óåðìèíû.

Óâæñò iù ïïèìàâi êâè çàôèëñéðiâàíóþ iññéââiâà-òâëüíñòü çíâéîâ. Óâêñò
ïiâââðââðâòñý èíóâðiðâðâòâðè -âðiââðâëi
èëè ââðiâðòiñ ñ êiíèðâðâðiíé iðâðèò- -âñéñéíé èëè èññéââiâàðâëüñéíé òâëüþ. Êâæäiâ èññéââiâàíéâ
ïiðâðââðñý ìà èíóâðiâðöþ, ñâýçáííóþ ñ êiíèðâðòiùè ìâñòâiè â òâðñòå, è âñýéâÿ
ñiâiá ñiññiâà èíóâðiðâðâòâðè íóæâââð-ñý â óéâçáiè èâñòâ òâéíé ñiâiú. Ní-âî-êói-íñòü
óéâçáiè ìà ìâñòâ ã òâðñòâ, ñâýçáííû ñ êiíèðâðiíé èíóâðiðâðâòâðèâé, iù ìâçüâââi ðâçìâðâëi. Òâéèì ìâðâçií,
ñ òâðñòâ ñiæââð áûòû ñâýçáíí ìiñâð ðâçìâðiè, ìòââð-âðþuëò ðâçíû ñâëyì.

Âàæíûìè

Iðèìàðàìè ðàçìåðòè è ýâéëþþñòý ðàçáèáíèå òàðéñòà íà ñïñòàðåéëý-þùèå áâî
ôðåäàíåíóù è îòðàæáíèå ñòðóéòòðû òàðéñòà á áèää èåðàðòèè. Ñäýçü
èåðàðòèè ñ òàðéñòì íáóñèíâéåíà áçäéÿäíí íà òàðéñò (ðäéëþ èíøåðíðåòàöèè), òàê ÷òî
ðàçíúâ ðàçìåðòèè íâíííà òàðéñòà ííåòò ííååððæèàðòü ðàçíúâ èåðàðòèè. Íî á íàúâí
ñéó-àå ðàçìåðòè – ýòí íà ðàçáèáíèå è íá èåðàðòèý, à òïëüêí íááíð óêàçàíéè íà
íåñòà á òàðéñòå.

ÍÍ
ñííñíáó ñâýçè ñ òåéñòíí ðàcìåòêè äåéýöñý íà äåà ééàññà: áíåøíéå è áíóòðåíéå

Âíåøíyy

Ðàçìåòêà –

Yóí iàáîð nññûéè ià yéâlâiðù òâéñòà, ððâíèlùé iòäâéüí ìò òâéñòà. Òâéàÿ ðâçìàðòéà iàééò-øèì iáðâçii ññòâðòñòâðòå ìâçââéñèíñòé iáúâéòà (òâéñòà) ìò áâî ïïèñàíèÿ, iâññiâ-èâàÿ iâçââé-ñèíñòù ðâçìàðòé. Íðèíâðií áíâáøíâé ðâçìàðòéè ïðæåò áûòù iâðâð-âíü iâñò á ïññéââà-òâéüíñòé cíâéñâ, áûòðæáííûé ìñâðâíè ñéíâ òâéñòà, áóéâ ñéíâà.

Âíóòðåíýý

Đà Nẵng –

ðàéîé ñíññíá ðàçìåðòéè, íðè éþòðíði lâñòði á ðàéñòðà ìò-ìà-÷àðòñý áiâññáéàì íðýíí á ñíñéäáñâðàðåéüíñòð cíàéñâ ðàéñòðà ñíñðöèàéüíûðò lâðòñé – ýéâiáðóñâ ýçûñéà ðàçìåðòéè. Íðè ýòði ðâéñòð óæáá íá ýâëÿðòñý íåèçíáññûð íáúáðòñi, à èçíáñýðòñý ñ ãíññâðééñéàì êàæäé ñíññé ðàçìåðòéè. Áiâñòðà ñ áíñððåññé ðàçìåðòéè óáñññí ðàññíàðòðéàðòü è íåýâíóþ ðàçìåðòé – àéæéñûð è íåññéæññûð ýéâiáññòð ñéññéñññí ñíññé ðàññíàðòéè.

Âíåøíÿ

ðaçìàðòéà èìåðàò áâà íáóðíñòåà: 1) äëÿ á, ðääàéðòéðíàíèÿ íáðíñòåàíèíí ñíäöéèåéüííà íðíàðàíííà íáññíà÷áíèà (NÍÍ), 2) ðääàéðòéðíàíèà ñàïñíà òâéñòà áéèéþò íà áñà èíàþùéåñý ðaçìàðòéè, á òíì ÷èñéà ñíçàáííúá íáçàâéñèíùíè èññéåáàðåðåéýíè è áíá NÍÍ. Òái íá íáíå, áíåðííýjy ðaçìàðòéà èñííëüçóåðñý á íðåéðòééå, ðàéé êéæé á, áâæ-íùí áññòåéíñò-áíí ýäéëþþñý lïäååðæéå à lïäæåñòåà íáçàâéñè-íùó ðaçìàðòé è íáèç-íáí-ííñòü (çàùéù, ííñòü) ñàìèò òâéñòå à íéðíóñà. Èçååñòåí è lïäåñíä, íðé èíòåðíí áíåðííýjy è áíóðåðåííýjy ðaçìàðòéè íáðå-áíäýøñý äðóå á äðóåà – äëÿ ðääàéðòéðíàíèÿ áñà áíåðíèå ðaçìàðòéè áíññøñý á ðåéñò, ñòåííâýñú áíóðåðåííè, à áíóð-ðåí-íéå ðaçìàðòéè äëÿ ððåíáíéÿ è íéðåðíðåðåðòéè ñííâà íáðååáíäýøñý áíí áíåðííþþ ðíðíó. Á èíðíóñà lïäåò èñííëüçåðåòüñý è ðaçìàðòéè íáññòå ëéèññíà, òái áíëåå ÷òí áíåðííýjy ðaçìàðòéà lïäæåò ñ óñíäåñíí lïéðåðåòüñý íà ýäíóþ è íáÿäíóþ áíóðåííþþ ðaçìàðòé.

Nâîè íåóäîáñòâà

2. Đà Nẵng

à ÒÑí – áíóòðääíïý. Å iðéiöéëü ñëéëäüâåæéëñü iíä åéèëýéæáì
 ñòðåíäéåíëý: à) ê éäéíïíé ðíñòë ðàçìåðòëë (èñòñíäúíü òåéñò ãíéæåí iñòåâåòùñý
 èåéé iíäéíï ëö-ðå ÷éòåâåíü); á) ê ííñæñòðåâåí iñòë ðàçìåðòëë å íäíí òåéñòå
 (ðåâåí÷åý ïñòå òåéñòå ãíéæåíà iñòåðòüñý åäééñòðåâåííé äëý ðàçíüõ çåäå÷); á) ê
 íåçåâåéñèíñòë ýòëë ðàçìåðòëë (ííâåý åíññèíàÿ ðàçìåðòëå íå åíéæåíà åéèëýòü íà
 èíòåðïðåòåòëëþ ðàíååå áíåñ, ííñòë); á) ê ïðñòòðå ìíñåáíëý èíòåðïðåòåòëë
 ðàçìåðòëë (ííññåíèå åíéæåíí áúòòå åíñòóíü ííñòë).

À. Èàêîíè÷íñòü

Íáåñïå÷èåàåòñý ñëåäóþùèìè ñðåäñòåàìè.

Âñ-âòñðûñ, õíöy yéâìåíòù ðäçìåòèè ìòêðûñâàþò ñîáñòâåííúå ôðàäìåíòù òåêñòà, ê íèì íå ïðåäüýâëÿðòñý ñòðñâå òðåäâàíèå èìåòü ìàðíûé (“çàêðûñâàþùèé”) yéâìåíò (÷òí íå ìàðøåð ëññëü-çí-âàòù ìàðíñòù îðè íåññëüèñòè, îñâåððæèâàÿ á, îðè ñïñòàíèè íòðåðòèè).

Ýòè
Óâðåáíéý áâéèàþò ðàçìàðóèè â óâéñòå ðâîðåòè÷âñéè èéèìàëüíùè: íàðéà ñòàâèòñý òàì è òïëüêè òàì, áäå áîçíèéàåò íåíáõíæìñòú â ñíåíá èíòåðíðå-òàöèè, íå óñìàòðèåàåìàÿ èç äðóãéø íàðéåðíâ. Â ðåçóëüòàðå íåéîòí-ðûå çàäà÷è ëíâóò ðåòàðüñý ááç áíâñáíéý áíííéòéòåëüííé ðàçìàðóèè â óåéå èíàþüéâñý èíòíñà.

Á. Ííñáññóðåññíññöü
ðàçìåòðîê íááññá-èáååðòñý áíçííæíññöüþ íåðåäå êåæ-äúì ííñå÷åííúì óððåäåíåíóìí
âñòåååéòù ííñûé, ñeoóæåáíüé óððåäåíåíó ñí ñåtåé iåð-éíté. Óæééå óððåäåíåíóù
ñíéíðåéññé-åññéé íå ìøëè-àþþöñý ìò óððåäåíåíóíå èñññííñííå ðåéññòå (è ííñåóó íåññòå èþþáíå
éíøè-åññóåí éíôíðåòéé, ñíååððæåòåëüííé äey ííñååðæéååíñííñííå åáíííé ðàçìåòðéíté
åññíåéòå). Iðè ýòíí, åññéé iåòéå óððåäåíåíó íå iðèéååííå íééåéññííå çíå-åíéy iðè
ííññåíéé íéóåðíðåòåòéé, óí ååññó óððåäåíåíó (à íå óíëüéí iåðééå) éäññíðéðåòåññý.
Óæééí íáðåçíí, äíáååééåíéå á òåðéñò ííñûó iåðóíé íå åééëýåò íå óæå ññóùåññóþþéå èíð
íé-åññ íå ñéàçáíí.

Â. Íàçàâèñèïñòö ðàçìåòö ðâàëëçóåòñÿ, èöñìå ôéàçàííîé ñïññíàíñòè íàðàíèçìà èíøåð-ïðå-òàöèè
èáíñðèðñàòö “-óæèå” ôðàäìåíòû, öàì, -òi íí (íàðàíèç) äñóñêàò è íðîèçâîëüíå
çàäåâàíèå ðòðàéòñâèè ñàïíäí ðàçáèåíèÿ íà ôðàäìåí-òû, íïçâíèÿ ï÷-òàòü
ñòúåñòåíüìè äéÿ ðàçáèåíèÿ öíèüéí íóæíûå íàòèè.

Ã. Iðññòòà
Íièñàíèÿ èíøåðíðåðàòèè ðàçìàòòèè íáññiâ-èâààðöñÿ òâî, ÷òî ñ òî-÷èè çðåíèÿ
ðääàé-òî-ðå èíðïóñà ííà ñíññòòèò à âûïèññâàíèè à íðñòòì òåêñòò-âî ôâééà
(“;ñòééââîò òåééâ”;) äéÿ èåæäíé íåðèè ñíññòðåâééíèÿ áé õåíí-÷èè ñèlåíèå è
éíòòðüå ñèåäåòåò íðèíàíèòü è ôðåäâíåíòó. Íðåíàðàçíâíèÿ, í êíò-ðûò èà, ò ðå-÷ü, óéâçûâà
ñâñìèò èíåíàí, çà êíòòðù ííæåò ñòíÿòú èéáí íðåíðåííàÿ ÿòíéöéÿ, èéáí
íåíðåðàíííí íðåíàðàçíâíèå, çàääââåâííà íåðå-÷-íáí íðñòòûò èéëè ðåêòðñèâíûò
çàíàí à ñíåöèàëüíî èåññòòâíé ôîðå. Néíòåêñè-åññèè èíåíà òåð ëèëè äðóåèò
íðåíàðàçíâíèé íå òåçèè-÷-åþòñÿ, òâè ÷òî
ðääàéòòð èíðïóñà ííæåò íí-ñâñâíò íåðåíðåâéëþòù èåèéèå-òî èc ñèñòåííûò óóíéöéè.

Ілімнің тарихи мәдениеті

Ілімнің тарихи мәдениетінің негізгі салынудардың бірі - қоғамдық мәдениеттің өзінен жаңынан орталық рөлдеуі. Оның негізгі мәдениеттік миссиясы - қоғамдық мәдениеттің өзінен жаңынан орталық рөлдеуі. Оның негізгі мәдениеттік миссиясы - қоғамдық мәдениеттің өзінен жаңынан орталық рөлдеуі. Оның негізгі мәдениеттік миссиясы - қоғамдық мәдениеттің өзінен жаңынан орталық рөлдеуі.

3. Ілімнің тарихи мәдениеті

Ілімнің тарихи мәдениетінің негізгі салынудардың бірі - қоғамдық мәдениеттің өзінен жаңынан орталық рөлдеуі. Оның негізгі мәдениеттік миссиясы - қоғамдық мәдениеттің өзінен жаңынан орталық рөлдеуі.

Ілімнің тарихи мәдениеті

Ілімнің тарихи мәдениетінің негізгі салынудардың бірі - қоғамдық мәдениеттің өзінен жаңынан орталық рөлдеуі. Оның негізгі мәдениеттік миссиясы - қоғамдық мәдениеттің өзінен жаңынан орталық рөлдеуі.

Оңтүстік-Азияның тарихи мәдениеті

Оңтүстік-Азияның тарихи мәдениетінің негізгі салынудардың бірі - қоғамдық мәдениеттің өзінен жаңынан орталық рөлдеуі.

Оңтүстік-Азияның тарихи мәдениетінің негізгі салынудардың бірі - қоғамдық мәдениеттің өзінен жаңынан орталық рөлдеуі.

Оңтүстік-Азияның тарихи мәдениетінің негізгі салынудардың бірі - қоғамдық мәдениеттің өзінен жаңынан орталық рөлдеуі.

Ideology and technique for encoding using the mixed scripts technique

Alexander V. Kovalenin

A. P. Ershov Institute of Informatics Systems of the Siberian Branch of
the Russian Academy of Science, Novosibirsk, Russia

This paper discusses common problems in text encoding and proposes their solution using a “mixed scripts technique” (MST), a way of working with texts containing parts which need different interpretation. MST’s way of text markup, together with a procedure of markup interpretation, is given in comparison with other solutions like XML. A special form of non-programmed user-defined transformations is shown, which can become common for carrying out certain kinds of scientific research.