

ÈNÍÍÉÜÇÎÀÀÍÈÅ TEI È RDF Å ÈÍÓÎÐÌÀÖÈÍÍ-ÍÍÈÑÈÎÀÍÉ ÑÈÑ-ÒÀÌÅ «ÐÓÑÑÈÀß ÈÈÒÅÐÀÇ

Àâòìð Àðòåì Âèêòîðíâè÷ Àíäðååâ
27.08.2012 å.
Íñéååíå Íáíâéåíéå 27.08.2012 å

Summary. The conceptual data model of the information retrieval system «Russian Literature of XVIIIth century» challenges efficient indexing. The system is based on a set of TEI P4 conformant documents where available search criteria are derived from the mark-up. Initial relational approach proved to be very inadequate and was replaced by a RDF-based system. Standard RDF schemata are used as much as possible but are too coarse-grained in many cases, so a custom schema was devised.

Èîòîðiàòéíí-ìèñêîâàÿ ñèñòàìà «Дóññêàÿ èëòåðàòóðà XVIII åâéà» íáðñíà-÷éâàò ìèñêè òâéñòíà è óðâàäíåòíà òâéñòíà ë î ðýäó íà ïàðàïåòóðû è ó.í. [Àíåðåâà è äð., 2009] Îñííàíé ÈÍÑ ýâëÿåöny èîòíñò òâéñòíà, ðàçìå-÷áíûõ à ñíòåâòñòåè è ñ ðåêíàòéÿè TEI P ýâëÿåöny òi, ÷òi íááíð ïèñêîâûõ íàðàïåòðíà íá çàääí ãéñòíà, à ïðäåäéÿåöny èñííëüçóåíûè óåâàïè TEI.

Íåññìòðý íà ìòíïñèòåëüíí íåáîëüøé íáúåí êîðïóñà, áïïðîñ ýôôåêòèåííé íðääàíèçàöèè èíååðòèðîâàííûò èíäåêñïâ áñòàë óæå íà ðàííèõ ñèñòåíû áûÿâèè ñèåäóþùèå ñèàåûâà íåñòà:

òðåâáâáéíéà ðíðæéëçæðöè ñíäåéëè íáéëåúðaaðo ñåðåíé-áíéý íá èñííüçóáíóþ ðàçíåðöéó, ééåí íðéåíäðéò è áðåðñ-óð áðííçäéè ñôðåíá.

ǣéíáè̄é̄-åñéäȳ ïðéòíäáá íáááðá íáðáááððíáá íéíðí ní-åðáááðñȳ n íáááðíäéíñðóüp̄ èíáðöü áðáñðóéóþ níðááíó ÁÁ

Íáéæñòöþ ãäéñòåèý ïäðæíåðöðä lïäæðó áúñöü êäé ðäéñò, ðäé è ïðiñéçäñëüñé áäí ðöðääñäåíò, ÷òí ðäéæðå lëiñòi áùñðäçèíí à ðäéÿöøëíííé

Íáííáéåíéå Ííèñéíáúö èíäåéñíá íéàçûåàåòñý íåòðèåéàëüííé çàäà÷åé.

Íþjóðið íñéða ðýðaa yéñiðaðeíàðiða áuði iðeýði ðáðøðaíeyð nñéðaðiðaòu iðaðiðaó, íaðað-áíiðið a [Tummarello et al., 2005] e ðáðaðnñòðaâðeyðu iðeó-ðeòu RDF-ðáðaðnñòðaâðeyða, iðeíñiðoð yéâðaâðeíðiða TEI-ðaçíðøða, à òiðeüði ñaðið-ðeòu áiçíðæíñiðu áuñòðiðið iðaðiðaðeyða.

ଲୁ ନୀଳାଧରେ ହୁ ପାଇଁ କାହାର କାହାର

Dublin

Core Terms äë́ý ïðääññöàäéáíéý áéáééíáðáðåðé-ññééíáí ïíñéñáíéý è ñòððéóðóðéðíáíéý ðáéññòðíá

OAF

äëÿ ïðåæñòàâëåíèÿ èìåí ñíáñòâåííûõ

Äëÿ íåéîòïðûõ ìàðàìàðòðíà, à íåðâóþ î÷åðâäü ñòðèòíàðä÷åñêèò, áïòíàûõ RDF-ýéàíàðòðíà, êàæåòðñÿ, íà ñòùàñòðåðåò àïíàúà, à äëÿ íåñòðåðåíûõ (íàïðèìàð, eìàíà èñòîðè÷åñêèò ëëò vs eìàíà ìèòîéñâ÷åñêèå). N äðóàíè ñòðïíû, ìòíøàíèå dcterms:isPartOf, åñòåñòðåðåíà ïñíàû ÷åñòåé òàéñòà, íàïðèìàð, yìeåðàòðíà. Íñýòíî íàìè áûëà ðàçðåáàòàíà ñíáñòåíàÿ RDF-ñòðàìà, (a) ðåàéèçóþùàÿ íåäñòðåðåíà, íàðàòåðçûé RDFS, íñïèéñüéò èñïíñéúçíàìèå OWL à íàðóàí ñòðàòðíà ìòðåñòðåðåíà, à ïðèéñüéò è õåðåéè÷åíèþ íàðåñòðåðåíà.

èàé âñòðíáííúé SGML-ìáðñåð (ðåàëëçóþùèé ïäíñæáñòáí iñéíñáí SGML, äñòàòòí÷íá äëý íáðàáîòèè äîéóìåíòâ TEI P4), òàé è ðàç

Áëàñïäàðíñòè

Ðàáîòà áûíñéíáíà íðè ôèíàíñíâíé
íñäåðæéá ÐÔÔÈ, åðàíò 1–07–00493à.

Ñiñíê
ëèòåðàòóðû

Àíäðååâ è äð.,
2009 Àíäðååâ Á. Á., Áóôàðèéí Í. Á., ìàòåååå Á. Í., ëíñàðååà Á. Í. Ñ. ðàçðàáîòèå íñéíé òåìðåòè÷åñéíé íñäåéè ðåìðåçåíòàöèè èñòòñíá, 2009. N. 303–310

Tummarello et
al., 2005 Tummarello G., Morbidoni
C., Pierazzo E. Toward textual encoding based on RDF // Proceedings of the
9th ICCC International Conference on Electronic Publishing. Leuven-Heverlee, 2005. Pp. 57–64.