

Α ηαηόγυαά αδαίγ ιύ ινούαηαέγυαί αδαίιαδè-αηέορ δααίαδèο ιδααηααεαίιύο á είδηόηά æεòèè. Δααααίοαί οίδηαο ιδααηααεαέα είαδèιδααοέγ ιηαεúιύο οεοδ ααεηεò ιò οίαι, è èαείè -αηòè δά-è ιοίηεοηý ηεíαι.

Á είδηόηά οαέηòò æεòèè ιδααηααεαίύ άααæáú — á οαέηοίαιι οίδηαοά è οίδηαοά δάααεοίδη Word. Á οαέηοίαιι οαέεá ιδ áεáá: áúíηίύά αóεáú á εδóαεúο ηεíáεαò áηα-áεáíú íá ηáíá ιáηοί á ηεíáá ιι ηιúηέο, ιιούάιύ áεοáιόιύά αίαεè. Ιηίαúιè ηεíáιεáιè οáεηòà, áίηεíηáι η ιηεáé; ιθεáí-ίύά íáιèηáιέγ. Íá ιηίááιεè οáεηò-áíαι óáεεá ηίααααοηý ηεíáιόεααοáεú. Οáεηò á δάααεοίδη V áúíηίύά αóεáú ααίεαοò ηáíá ιáηοί íáá ηòδíεíè, ιι áíáθáιáι áεáó γοίò οáεηò ιδεáεèæααοηý è οáεηòó δóεíηεè, ιòεè-áγηú ιò íá

Íáιè áúεí ιδεéγοι δάθáιεá ηάεεαòú ιιόáεèéíαιάιύά οáεηòò æεòèè áηòóιίύιè áεγ áíáθáιέο ιηεúαίáαòáεá è ιδααηααεòú εò íá ηá οáεé — á οίδηαò XML, εííáδòàοέγ ιηούαηαέγáοηý áαòíáδè-áηεè.

Á οίδηαεúιι ιεáíá XML-δααίαδèè εíδηόηά ιηίáúáααòηý íá íáæáοíáδíáιύο ιίδηαò ιοίδηεáιέγ ýεáεòδíιύò εααáιεé οáεηòà, á -αηοίη

Á ιηίáó ηòδóεòóδú æεòèγ εáè ýεáεòδíιίαι áίεοίáíòά ιηεíæáíú οίδηαεúιύά οαδáεòάδèηòεèè δóεíηεè: δααéáεá οáεηòà íá èèη (íáòεè) íá-áεá/είίòά èèηòà, είεíεè, ηòδíεè η ηιíòáαòηòáορúáé ιοíáδáòéáé.

Íδααηααεγý οáεηò á ýεáεòδíιιι οίδηαòά, ιηεí áúάδáοú íáεí εα ááοò ιóοáé: íáεηεíáεúιí οί-ίι áηηίδηεαíáεòú áεá δóεíηεηίηáι οáá áοίδηέ ιóου. Οί-ίι óáè æá, εíááá οίδηαεúιίá -εáίáίεá οáεηòà (íá ηòδíεè è èèηòú) íá ηíáíááááò ηí ηιúηεíáúι (íá ηεíáá), ιú áηáááá οáεééíι, ιι ιδè γοίι ιοíá-áαòηý ιáηοί, ááá ιδíδηεò áδáίεòά ηòδíεè.

Ááδèèèèδíááιύά XML-íδααηααεáíέγ æεòèéιύò οáεηòíá αóáóò á ááεúíáεøáι áηηεíáíú ιíδóιεíáè-áηεíè δααίáδèé: δαααáιòáι

Νίηεíè èèάδáòóδú

Áááδèíá è áδ. 1990 Áááδèíá, Ν. Á. Ιαúé δónηεíè ááεíáδáòεè XVI á.: Íιúò áαòíáδè-áηεíáι áíáεεα / Ν. Á. Áááδèíá, È. Á. Αα-á

Áááδèíá è áδ. 1993 Áááδèíá, Ν. Á. Èáεηεéá è ηεíáíáδααίáíεá á δónηεíè ááεíáδáòè-áηεíè èèάδáòóδú XVI á.: Íιúò áαòíáδè

Академия è äð. 1996 Академия, Н. А. Еёёёёё è иðîéíåý â ðîññéé äåéåðå-åñéé èòðððððð XVI ä. / Н. А. Академия, È. А.

Summary

The Corpus of Russian hagiographic texts of the 16–17th centuries at present comprises 52 texts or 500 000 word-tokens, represented in 2 formats: as text files and Microsoft Word files; the texts are provided with a word form index. 10 texts have been published; they are available to Internet users in the PDF and XML formats. The work is under way to provide all texts with the morphological information. Automatic normalization of varying Church-Slavonic spelling is another important task.

[1] Å 2000–2006 ää. иððððððððððð: Åèòèå Èèðèèå Äåéçäðñéåé, Åèòèå Äååññåéå Ñåððñéåé, Åèòèå Àéééé Ñèééé Äèééé Äèðððððð è Äðððððð Ìåðððð-ñééé, Åèòèå Èðððððð Èñåññééé, Åèòèå Ìååå Ìåééé è Ñðððð Ìðððððð.

[2] Ìð ðàçèè-ääè èòèè è иéðððð, ñðåýýääñý ìåä äúñññéé áóèåé.

[3] Ìæäóéèððððð èíñððððð ì äððððððð ìðð ÿæèððððððð ðàçèððððð èåññéé: The Text Encoding Initiative [Ýæèððððððððð ðåññððð]. Çåäè. ñ ÿèðððð.