

Đàññīçíàâàíèå òåêñòå ðóêïïèñíûô è ñòàðïïå÷àòíûô êíèå íà iññîâå íåéðñåòåâûô òåðñïïæäé

Àåòð Ñâðñäé Èâàïíâè Ëðíèåíé

26.06.2008 ã.

Íñëäääíâá íññâäé 15.07.2008 ã.

Òåçèñû á ôiðìàòå RTF (47.7 kB 2008-07-15 16:54:34) Òåçèñû á ôiðìàòå PDF (54.75 kB 2008-07-15 16:55:05)

Ñðåäè ïðíäéå ìåðåâåä

èñòðè÷åññéö èñòï÷íèéå à ìàøèï÷èòàåíûé ôîðìàò é íàèáîéå ñëíæíû è òðóäåìèéè
îòññèòñý ñíçääíèå åâðñéé ðóêïïèñíûô è ñòàðïïå÷àòíûô êíèå á ôîðìàòå
ýéåéòïïííâ òåêñòå. xóíáû ïíé÷èòü ïíëíñòåíûé ýéåéòïïíûé òåêñòå, êåé íðåâèéé,
èññéüçóåòñý åäí ðó÷íé èññéüçóåòñý íáåíð n ïíññúþ ñíåöèåéüíûô ðåäåéòïðîå. Á
çíà÷èòåéüíé íåðå óéàçåíûå ðóðäåíñòå ñäýçåíû ñ òåí, ÷òí ñéàíèòåâåíèå ñ
íññéååòþüé òåññíçíàâåíèå òåêñâå ïðåà òåêñòå íå áååò óäåéåòåâîðéòåéüíûô
ðåçóéüòåòåíâå.

Íäííé èç íðè÷éí ýóïáí ýäéÿåòñý íåñòåíäåðòíñòü è
ðàçííáðàçéå ðøðéòåíâ èñòï÷íèéå ëðéäéåíâ, íàéè÷éå íäëñòï÷íûô çíàéíâ,
íâééååö, íðíñòí ïíåðåòíñòé íáïèñåíé, «çàòåðòíñòé», ííàðéé è èíûå áåôåéòû,
çàòðóäíþþüéå òåññíçíàâåíèå ñèíâíééå.

Íðåäñòåâéÿåòñý, ÷òí ðåçàéå ðøðéòåíâ õéàçåííé çàäå÷é ííæåò ëåæàòû á
íëíññéòé ñíçääíèý ïíáðéàééçéòåííûô íðíäðåííûô íðíäóéòå íà iññîâå íàèáîéå
íåðñíåéòåéåíûô åèäåå ñíåðåíåíûô èíòðíàòéííûô òåðñíééé. Á íàññüåå åðåïý è
òåéåíâù íòíññýòñý òåðñíéíâé, íðè ñíçääíèé èíòòðûô èññéüçóþöñý íðéåíû è íåòíðä
èñéóññòåâííâ ñíåðåééåéòå.

Ñ õåéüþ èññéüçíàâéé òåðñíééé èñéóññòåâííâ ëíðåééåéòå
äéý ðåçåíéý íðíäéåíâ ìåðåâåíâ á ìàøèï÷èòàåíûé ôîðìàò ñòàðñíû è
ñòàðñíå÷àòíûô êíèå ñíðòðåíéé è ëàáîðàòíðè èñòïðè÷åññéé è ííèòé÷åññéé
éíòíðíàòéé èñòïðèéí-ííéòíéíâ-åññéí ñâéóéüòå ëåéåäåòû íðéééåííé
íàòåíàòéé è èíòíðíàòéé íàðíññéí åíéâåðñèòå ìà-àéé ñíâéñòíûå ðàáòû,
ëíá-ííé õåéüþ èíòòðûô ííæåò ñòåòü ñíçääíèå ñíåöèåééçéòåííâ íðíäðåííâ
éííæåééå.

Íåðíññéé íàó-ííé ñéóññòåâííâ èíðåééåéòå [íåðíññéé 2008] íàéíéåí çíà÷èòåéüíûé ííûò ðàçðàáñòé è áíååðåíé
èíðåééåéòåéüíûô èíòíðíàòéííûô ñèñòåí, ñðåäè êòòðûô íàééåí âåñòåíàòþ
ñèñòåí, íðåäíàçíà-áíûå äéý ðåññíçíàâéé ñèíæåíû á òåðñíé-åññéí íòíðåéé
òåéñòå (íå-åòéé, çåäðýçíåíûô, ñ ííæåñòåí)
ííàðíé, íàéñòåíûô ñ èññéüçíàâéé ñàçíûô ðøðéòåíâ), òåééé, êåé áåòííæéüíûå
ííåðíñâ çíàéé, ííàéññé òåðñíé-åññéé ãééåðíñòåòéé. Á íññîâå òåééé ñèñòåí, êåé
íðåæéé, ííæåò íàéðíñòåðåâûå òåðñííéíâé, ííàéññéíûå áééåðéòíàé íðåæåðéòåéüííé
íåðåáñòéé èçíáðåæåéé (éííòðåñòéðåíâéå, áéíàðéçåöé, óäæéåíéå ííòðííéé
òåðåñòéé è äö.) è ñðåäñòåâåíè äíðåññíçíàâéé, áééþ÷àþüéé íðòðåðåòé-åññéé,
ñèíðåéñé-åññéé è ñåíàéòé-åññéé áééü áíàééçá.

Á íàññüåå åðåïý ñ õåéüþ àïðíäàòéé
èååéé íðèíâåíèå íåòíâå áèñéóññòåâííâ èíðåééåéòå äéý ðåññíçíàâéé ñèíâíéé
òåéñòå ñòàðñíû è ñòàðñíå-åòíûô êíèå ñíçääíí ãéíñòðåòéííûé íðíòíðéé, á
íññîâå êíòíðíâí çàééåéí íåðññíòðíí ñèíèñòé ñòðóéòåðû ñ ñèäíñéé
åééòéåàòéííûé ñòéé. Á êå-åññòåâ áíàéé ñíæåñòåâ áúéé èññéüçíàâíû

Øðéòòù ñòàðñéèåýíñééé íèñüìáíññòè. Íñéå ìáó÷åíéý íà áôïä íåðñåòðñíà
íñäåâæñü ñèìâíéü èç ññòåðòñòåóþùåñ òåññòà, íðè÷åì íåéñòå ðòðùå èç íèõ áúëè
èñéææåíù íñåðñàìè, à òåéæå áûññéåíù ðøéòòì, íðéè÷åþùèññý ìò òíññ, êîòîðñé áûë
èññééçñàì á ìáó÷åþùåì íñæåññòå.

Êàé íñéàçàëè ýéññåðèìåíù, íåéðñíàÿ ñåòü óññåðøí
ðåññíçìåâæå äî 80 % ñèìâíéñ
äðåâíññ åòåññòà.

Íñé÷åíññå ðåçóöüòåòù áàþò ññíññåíèå áññåðøöù í
íåðññåéòèåññòè èññééçñåíéý íñåññåí ìåéðñåðååññåí ïñäóey äëý ññçääíéý
ñèñòåíù ðåññíçìåâæåíéý òåññòå ðóéññéñò è ñòàðñå÷åðñò ëíèñ. Ñóùññòåóþò
ðååéññå ãíçññéòè ïñåññòè ðåññíçìåâæåíññò ìåðàçñå, íðè
ýññ íå òíëüñ ìòåéññå ñèìâíéñ, íí è èõ ññ-åðåññéé, à òñ ÷èññéå ñ
íåññòå ñíñèåíè, çà ñ-åò íñåññå ìåéññå ïòéìåéññò ìåðåéñéåí ìåéðñíññò
ññååé, à òåéæå áññééñéåíè èçíåðåæåíéé. Å áåéññåéøåí ðåññéñéèåðññý àåòññéèåðñò
äåññññòðåøéññé ïðñòòè, äåññåÿ áññ äí õðññéå ëíññåð÷åññéé ïðññåðñòå.

Ñieññé åèòåðåòóðü

Íåðññéå 2008 –; Íåðññéåÿ íåó÷íàÿ
øêéèå èññéññòåññåí ñíññéé ñíññééåò è åå èíññåòøéññéå ïðññéòò / È. Í. ßññéøéèé, Å. Å. Áññæåðü, Ñ. Í. Áóðæéí è äð. ; ïññ åðåä. È. Í. ßññéøéèé
«Ðååóöýðíàÿ è ñåñòè-åññéåÿ äéíññéé», 2008. – 75 ñ.

OCR of manuscripts and early printed
books using neural networks

Sergey I. Kornienko, F. M. Cherepanov,
Leonid N. Yasnitsky

Perm State University, Perm, Russia

This paper describes the possibilities for using
artificial intelligence technologies, in particular the neural networks, for recognition
of handwritten and early printed texts. Early results using the technology are described.