

Белорусизация cuneiform

Юрий Адамов

АКА

begemotv2718

16 сентября 2014 г.

Системы OCR

Проприетарные

- ▶ FineReader

Open source

- ▶ Tesseract
- ▶ Cuneiform
- ▶ Claraocr
- ▶ gocr
- ▶ .. ?

Системы OCR

Проприетарные

- ▶ FineReader

Open source

- ▶ Tesseract
- ▶ Cuneiform
- ▶ Claraocr
- ▶ gocr
- ▶ .. ?

Cuneiform

поддерживает русский, болгарский, сербский, украинский, казахский:.

Как добавлять язык

Основная часть

тривиальна (добавление языка во всевозможные списки, алфавит кириллический, большая часть сделана для украинского)

Ї

код копируется с русского Й

Однако!

Есть интересная часть – словари.

Словарь

Программа для генерации словарей

в самом пакете отсутствует → Самописная

Формат словарей `suneiform`

таблица основ (дерево) + таблица окончаний

нужны слова с таблицей окончаний

Подготовка базы слов

Исходный материал

- ▶ Словарь для ispell
- ▶ Словарь для hunspell (via Ihor Hrachyshka)
- ▶ *Корпус текстов + linguistica*
- ▶ Корпус текстов + (stemka, mystem)

Залатыя кроплі



Хата, у якой пэўны час жыў Лявон Луцкевіч зь сям'ёю, стаяла пасярод густога лесу. Зрэшты, яе ўсе так і звалі – “Лясная хатка”. А збудаваў яе муж Людзьвікі Сівіцкай, больш вядомай як Зоська Верас. (Яе дачка Галіна й стала спадарожніцай жыцця Лявона Антонавіча).

Усё было тут добра: цішыня, якую толькі дапаўняла цвырканьне ды шчэбет птушак, наўкола была некранутая прырода, – працуй, думай, а

Образец относительно хороший

Залатыя

кроплі

Хата, у якой пэйны час жыў Лявон Луцкевіч зь сям'ею, стаяла пасярод густога лесу. Зрэшты, яе ўсе так і звалі в Ё" "Лясная хатка". А збудаваў яе муж Людзьвікі Сівіцкай, больш вядомай як Зоська Верас. (Яе дачка Галіна й стала спадарожніцай жыцця Лявона Антонавіча). Усе было тут добра: цішыня, якую толькі дапаўняла цвырканьне ды шчэбет птушак, наўкола была нескранутая прырода, в Ё" працуй, думай, а стаміўся в Ё" шчыруй у сваім гародчыку, дзе поўна было розных дзівосаў. Аб цывілізацыі нагадваў толькі далекі водгалас цягнікоў... І цяжка было паверыць, што зусім побач, за гарою, пачынаецца вялікі старажытны горад в Ё" Вільня.

Але жыццё ч "Лясной хатцы"ўскладняла адна праблема в Ё" не было студні. Як ні шукалі ваду, дзе ні сьвідравалі в Ё" не знаходзілі жылу. А без вады в Ё" якое жыццё. І даводзілася неяк выкручвацца в Ё" пітную ваду прывозіць з гораду, а для гаспадарчых мэтаў зьбіраць па кропельцы ўсялякую: і

Образец похуже

Л 86

БОК 908 (474.5)

Фотаздымкі

В.Харчаікі', В. Савяікова, В. Бальчуц)са,

А. Кун чуса, М. Сакалаускаса.

ІЯВМ 9986-9228-2-8

© Луцкевіч Лявон, 1993

Выданьне кнігі сыпалася магчымае дзякуючы фінансавым
дапамозе

спі. Ялііны Кахаіоускаюі, спі. Алены Сіуко-Парасоіене,

спн. Гсіііны Войцік, Фуідацыі П.Крэчэускага.

Где искать

[begemotv2718.github.com/cuneiform-by](https://github.com/begemotv2718/cuneiform-by)