

A R A N E A : ΝΑΙΙΑΕΝΟΑΙ ΙΕΕΕΕΑΔΑΙΥΟ ΑΑΑ-ΕΙΘΙΟΝΙΑ

© Αεααει]δ Ααίει (Vladimir Benko). Νετααεεγ, Αδαοηηεααα. Ειηοεοοο
γυειγαιεγ ει. Επαταεοα Οοοδα Νετααοειε
Αεαααεε ιαοε, ΠΙΑΝΕΙ
εαοααδα ιιγαιγυ-ιιε ε ιαεεοεουοοδιε ειιοιεεαοεε Οιεααδηεοαοα ει. Βια
Εταιηεται α Αδαοηηεααα

Εαεοεε ιδααηοααγο ιοταεο
ηαιαεηοαα ααα-ειθιοηια Aranea αεγ γυειτα εηηεγυαααιυο ε/εεε
ιδαηαααιυο α ηετααοεεο οιεααδηεοαοαο, ειθιουα ιδααιαγαι-αιυ αεγ ιδαηααααεγ
οεεηεταε-αηεεο ε οδαηηεαοηεταε-αηεεο ιδααιαοια, α οαεαα ε αεγ εεηαεηοε-αηεεο
εηηεααηααεε.

Ααα-ειθιοη ιδααηοααεγαο ηιαιε ηηαυε
αεα εεηαεηοε-αηεηαι ειθιοηα, ειθιουε ηηααι ιοδαη ηηοαηαηιε ααδογεε οαεηοια
εγ ειοαδτααο ιδε ηηηε ααοηαοεγδταααιυο ιθιοααοδ, ειθιουα ια εαοο ηηαααεγπο
γυε ε ειαεθιαεο ιααεηυιυο ααα ηοδαηεο, οααεγπο οααειυ, γεαηαου ιααεααοεε,
ηηυεεε ε δαεεαο (δ.ι. boilerplate),
ηηουαηοαεγπο οδαηηοηιαοεη ια οαεηο, οεεουοδαοεη, ηηιαεεγαοεη ε ααοηεεαοεη
ηηεο-αηιυο αηεοιαοια, ειθιουα αοαοη ηηαι ιαδααηαοου οδααεοειηιυε
ειηοοηιαοιαε ειθιοηηιε εεηαεηοεεε (οιεαηεγαοεγ, ιεθοηηεηοαεηε-αηεαγ ε
ηεηοαεηε-αηεαγ αηηοαοεγ) ε αηααδεου α ηηεηηαοη ειθιοηηοη ηεηοαηο. Ηηααηεα
ααα-ειθιοηα ια οηεηεη ιαηηαι ααοααεα, ηη ιδαααα αηααη ααη δαγαιδ ηηαο αουου
αααα ια ηδγαιε αηεηοα οδααεοειηιυο ειθιοηηια.

Ιδε ηηααηεε αηαο ειθιοηηια αυεα
ηδεηαηα ιαεηαεηαγ ιαοηεηαεγ ε ιαεηαηα ιααηδ ιδαααηιυο ειηοοηιαοια: SpiderLing, Onion, Unitok [1] ε TreeTagger [2]. Α εα-α
ηηεηηαηε ηεηοαηη εηηεγυαοαοηγ NoSketchEngine [3] (open-ource) εεε SketchEngine [4] (ηεαοηαγ). Ο ειθιοηηια ιαααηεγ ια «ιαεοο
(εαοεηηεη) γυεα ιαηηα-αηυαη γυε ε οηεα δαγαιδ ειθιοηηα, ιαηδεηαδ AraneumAnglicumMinus,
AraneumRussicumMaius, ε ο. ι. Α ιαηοηηυαα αδαηγ ηαιαεηοαη ηηααδεεο
18 ειθιοηηια ια 14 γυεαα α ααοο δαγαιδαο ε αηα ειθιοηηη αιηοοηιυ α ααηηεαοηη
δααεηα ια ειθιοηηηη ηηοαεα ιοηαεοα [5].

Α ηοεε-εα ηο αη-εηηεοοαεηυιυο εεηαεηοηα,
ειθιουα ιαδαααοηααπο ειθιοηηηια ααηηια α ιαεαοηηη δααεηα, ηηοαεηυια ηεγυααοαεε
ειθιοηηια ιαη-ηη αεηαοαδαηηααηηια α ηηεηεαο ειηεδαοηιυο ηδεηαδηα ηεηα, ηη-αοαηεε ε
ηεηοαεηε-αηεεο ηοδοεοοδ δαηηα-αοαηηιυο α αεαα ειηεηααηηα, -αηοηοηιυο ηηεηηα ε
ηθιοεεαε ια γεδαηα. Ο-εοουαγ δαγαιδου ηηαδααηηιυο ειθιοηηια ηοαηαο ηηηηη, -οη
γοδαεοεαηηηου ε οαηηηηου ηηεηηαηε γαεγαοηγ η-αηη αααηηηη οαεοηηηη αεγ αηγεηε
δααηου η ειθιοηηηη.

Ναηεηαδ ηηηαγυαη ιδαεοεεα δααηου η
ηαιαεηοαηη ειθιοηηηηη ηηεηηαηηη ηεηοαηη NoSketch
Engine [6]ε Sketch Engine [7],
ηδεηααεααηεη ε ηαηηη εο-οεη ιεδα ειηοοηιαοιαη αεγ δααηου ηη «ηααδδοαηεηοεε»
ειθιοηηαηε (δαγαιδηη α ααηγοεε ιεεεεαδαηα οιεαηηα). Ιαα ηεηοαηη αυεε ηηααηη α Εααηδαοηδεε ιαδααηοεε αηοαηοααηηηη γυεα
Οαεοεηοαοα ειθιθιαοεεε Οιεααδηεοαοα ει. Ιαηαδεεα α Αδηη, ηδε-αη οοηεοεε
ααηηεαοηηε (open-source) ηεηοαηηη NoSketch
Engine γαεγποηγ ηηηηεαηηηηη οοηεοεε Sketch Engine ε αεεη-απο α ηηηηη
ιαυαηα δααηοο ηη ηηεηηαηε ηεηα (Word List)
ε ειηεηηααηηηηηη, ο. α. ηηεηε ηη ηεηαηοηηηα, εαηηα, ηη-αοαηεε ε ηηδοηηεηοαεηε-αηεηε
ιαδεα α δαγαιδ ηηηαεηαοεγ ια γυεα CQL
(Corpus Query Language), ε οηεα

âı-èñéáíèá êíèèíèàòèé íà áàçà ñòàòèñòè-âñèèò íàð ñí-âòàáííòè (T-score, MI, MI3, log likelihood, min. sensitivity è logDice). Ñèñòàíà ðàáíòààò á ðæèìá ñáðááð/èèèáíò, ááá íà ñáðááðá òðáíçòñý áñá àáííúá è íñóáñòáèççòñý ïèñèíáúá ïáðáòèè è ïèççíáàòáèú ðàáíòààò ñ èèèáíòí -âðáç ááá-èíòáðòáèí ïðè ïííè ñòáíáàðòííáí áðáòçáðà.

İèàòíáç ñèñòàíà Sketch Engine ñíááðæèò èðííá áñáò óóíèèé NoSketch Engine òðè ñóàñòááííúò ðáñèèðáíèç – èíèèíèá ïðíòèèè (ñèáò-è) ïíñòðíáííúá íà áàçà ïèççíáàòáèúñíèíè ñèáò-áðáíàòèèè, àèñòðèáóòèáíúé òàçàòðòñ è óóíèèè ñðááíáíèç ñèáò-áé äèç äáóó èáèñè-âñèèò ááèíèò. Áñá çòè óóíèèè ðàáíòàðò ñ àáííúè áú-èñèáííúè çàðáíáá, -òí ááèááò ñèñòáíó í-áíú áúñòðíè è óáíííè. Ñèñòàíà ïðáííòàáèççòñý á èèáá ñáðáèñà (ïíáíèñèè) íà ñáðááðáò èíííáíèè Lexical Computing, íà èíòíðúò ððáíçòñý èíðíóñú áíèáá -áí íà 80 ççúèáò, áèèç-àç 15-ìèèèèáðáíúé èíðíóñ ðòññèíáí ççúèá.

Èííèðááíñ òíðíàòà KWIC äèç ñèíáíòíðíú «ííáíñèáèðñèèè»

İðááíñòíðííúá
èíèèíèáòú ñèíáíòíðíú «ííáíñèáèðñèèè»

Äèñòðèáóòèáíúé òàçàáð äèç èáííú
«ííáíñèáèðñèèè»

çàñòíòíáç àèñòðèáóòèç èáííú
«ííáíñèáèðñèèè» ïí TLD

İðíáðáííá
éóðñá

«A r a n e a : Ñáíáèñòáí ìèèèèáðáíúò
ááá-èíðíóñíá»

(10 -áñíá èáèèèííúò çáíçòèè è ïðáèèèèóííá,
16-20 íçáðç 2015 á.)

Èáèèèç 1. Ááááíèá. Èíááèñòè-âñèèè èíðíóñ èáè èñòí-íèè
èíòíðíàòèè í ççúèá

İñííáíúá ïíçòèç: àèáú çèáèððííúò
èíèèáèèèè òáèñòíá

Èñòíðèç ñíçááíèç èíðíóñíá, ááíáðáòèè, ïðèíáíáíèç

2. Ἰατροῦ ἰατροῦ CQL

Ἰατροῦ ἰατροῦ ἰατροῦ

Ἰατροῦ ἰατροῦ CQL:

ἰατροῦ ἰατροῦ ἰατροῦ ἰατροῦ ἰατροῦ

Ἰατροῦ ἰατροῦ, ἰατροῦ ἰατροῦ ἰατροῦ

Ἰατροῦ ἰατροῦ ἰατροῦ ἰατροῦ ἰατροῦ

B. Ἰατροῦ

ἰατροῦ ἰατροῦ ἰατροῦ ἰατροῦ SketchEngine

Ἰατροῦ ἰατροῦ 1: Ἰατροῦ ἰατροῦ Sketch Engine (<https://www.sketchengine.co.uk/>)

Ἰατροῦ ἰατροῦ ἰατροῦ (ἰατροῦ)

Ἰατροῦ ἰατροῦ: ἰατροῦ ἰατροῦ ἰατροῦ ἰατροῦ

Ἰατροῦ ἰατροῦ ἰατροῦ ἰατροῦ Aranea

Ἰατροῦ ἰατροῦ ἰατροῦ (Sketch-diff)

ἰατροῦ ἰατροῦ ἰατροῦ ἰατροῦ

Ἰατροῦ ἰατροῦ ἰατροῦ

Ἰατροῦ

2: Ἰατροῦ ἰατροῦ ἰατροῦ Sketch

Engine

Ἰατροῦ ἰατροῦ TenTen

Ἰατροῦ ἰατροῦ ἰατροῦ

Ἰατροῦ ἰατροῦ ἰατροῦ ἰατροῦ ἰατροῦ: Corpus

Architect ἰατροῦ WebBootCaT

Ἰατροῦ ἰατροῦ ἰατροῦ

Ἰατροῦ

Ἰατροῦ ἰατροῦ ἰατροῦ ἰατροῦ ἰατροῦ ἰατροῦ ἰατροῦ ἰατροῦ

ἰατροῦ ἰατροῦ ἰατροῦ ἰατροῦ ἰατροῦ ἰατροῦ ἰατροῦ Aranea (<http://ucts.uniba.sk/>) ἰατροῦ ἰατροῦ 3-ῶ

ἰατροῦ ἰατροῦ ἰατροῦ ἰατροῦ Sketch

Engine.

Webography

<https://savba.academia.edu/VladimirBenko>

http://ucts.uniba.sk/aranea_about/

[1]
<http://corpus.tools/>

[2]
<http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>

[3]
<http://nlp.fi.muni.cz/trac/noske>

[4]
<https://www.sketchengine.co.uk/>

[5]
http://ucts.uniba.sk/aranea_about/

[6]<http://nlp.fi.muni.cz/trac/noske>

[7]<https://www.sketchengine.co.uk/>

[8] Nĩáéò ĩđĩáêòà Aranea: http://ucts.uniba.sk/aranea_about/

Êĩđĩóńĩúá ĩđòàëû: Aranea: <http://ucts.uniba.sk/>
; <http://ella.juls.savba.sk/>

