

Формализация Словаря Академии Российской

Подтележникова Е.Н.,
доцент, к.ф.н., ВГУ

1. Параметрический анализ лексики

Факторы адекватности словарного описания: время, пространство, размер словаря, его тип и концепция авторов.

Параметрический анализ – время.

Лексическая нуклеология – большое и малое параметрические ядра лексики.

Параметрический вес лексемы словаря:

- употребительность, косвенно оцениваемая по длине слова;
- синтагматика - богатство сочетаемости;
- парадигматика - богатство синонимии;
- эпидигматика - многозначность.

2. Словарь Академии Российской

Словарь Академии Российской (1789 г.)

Первый академический толковый словарь
русского языка

Содержит 43357 слов.

Инновационный характер словаря:

- формирование словника,
- структура словаря в целом и отдельной словарной статьи,
- ориентирование на словопроизводство.

3. Особенности создания базы данных САР

Плохое качество печати и результатов сканирования.

FineReader

Afterscan

Python

Microsoft Excel

Результат розпознавання в FineReader

Со'еззаЗоьеннъл, ная, ное, ОЗеззаЗокенЗ,
йна, чено, прил. просшонар. Свобод-
ный хлопотъ, попечений, труда. ЯЪ«.
ъпвЗ діло, сталЗ талеръ совсемЗ о.
оеззаЗогеиЗ.

ОЗАБОЧИВАЮ, еіпъ, ббтилъ, бочу, ЧИ-
вашь, ббгаишь. гл. д. проси. Налагаю
па себя, или на другаго заботу, попе-
. ченіе о чемъ; причиняю заботу. Уіз*
лишне самЗ сесГя хлопотами озаЗо-
тилЗ. Уіомнп, сто лтеЗлозаЗогкваю
ділоліЗ важнимЗ.

ОзаЗо гиваюсь, слоился, бочу ся, тташь~
ся, шиться. гл. возвр. Рачительно за-

Результат работы AfterScan

АНБАРЪ, р'а. с : йі Зри'А мблрѣ. " АНГЕЛЪ, сѣтѣв) па. с. м. Греч: **знача-**уррА Вѣстникѣ. 1) Вѣ священномѣ писаніи *Тогда послетѣ Ангелы своя.* Марк: хш. ау. *РТосланѣ бысть Ангелѣ ■ Равр'илѣ: Куі. 2,7. 2.) Вѣ про-етор: им моего Ангела.* 3) Человѣкъ святаго и отмѣннаго житія, который **называется** также *Ангелѣ во плоти, Ангелѣ / подобный.* 4) Вещественное **изобра-**женіе Ангела. *Зіаписатѣ > изваять Ангела.*

```
#!/usr/bin/python
import smtplib
import base64
import os
import sys
import xlswriter
import xlwt
import datetime
import MySQLdb
from pyh import *
from email.MIMEMultipart import MIMEMultipart
from email.MIMEText import MIMEText
```



```
db =
    MySQLdb.connect("192.168.1.118","stp","stp","S
TP")
cursor = db.cursor()
query = ("""select * from stp_automation_output""")
cursor.execute(query)
myresults = cursor.fetchall()
workbook = xlwt.Workbook()
worksheet = workbook.add_sheet("My Sheet")
#date_format =
    workbook.add_format({'num_format': 'd mmmm
yyyy'})
```

```
bold = workbook.add_format({'bold': 1})
worksheet.write('A1','Sno',bold)
worksheet.write('B1','function_name',bold)
worksheet.write('C1','input1',bold)
worksheet.write('D1','input2',bold)
worksheet.write('E1','input3',bold)
worksheet.write('F1','Expected_output',bold)
worksheet.write('G1','Actual_output',bold)
row = 1
col = 0
workbook.save()
```

База данных САР

	A	B	C	D	E	F	G	H	I	J
1	№	Лемма Современная	Лемма Ст.	Пометы	Номер зна	Толковани	Пример			
2	1	Аба	аба	с : ж.		1 бълое тол	Под аба положить			
3	2	Абинный	Абинный	прил.		1 Изъ абы :	Абинный кофтанъ			
4	3	Абасы	Абасы	м: множ		1 Персидсѣя	Много абасовъ			
5	4	Абіе	Абіе	Нар : Сл.		1 Скоро	Совсем абіе			
6	5	Абшить	Абшить	с: м.		1 Писменно	Абшить по собственному желанию			
7	6	Авва	Авва	с: м.		1 Отець	Авва Доровей			
8	7	Августъ	Августъ	с : м.		1 Август	Теплый Августъ			
9	8	Авдотька	Авдотька	с: ж.		1 Имя птици	Маленькая авдотька			
10	9	Авось,Авосълибо	Авось,Аво	Наръч,сл		1 Может бы	Авось разживця			
11	10	Агаръ	Агаръ	с: м.		1 Родъ мал	Зри куликъ			

Результаты

1. Подбор оптимальных инструментов формализации САР, их перспективность для параметрического анализа.
2. Создание базы данных словаря.
3. Ориентирование САР на словопроизводство, что проявляется в приведении каждого деривата в отдельной словарной статье, потребует применения других средств формализации для использования материала САР в сопоставительной лексикографии.

Спасибо за внимание!