



**Московский государственный университет
имени М.В. Ломоносова**



ФАКУЛЬТЕТ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И КИБЕРНЕТИКИ

Знания в автоматической обработке текстов

Зверев Дмитрий, 210-я группа ВМК МГУ

План работы

Посмотреть представление для слов русского языка в BabelNet, а также синонимы, отношения и переводы для них (при наличии). Проанализировать 6 слов (по 2 существительных, прилагательных и глагола). Сделать выводы, а также сравнить полученные результаты с RuWordNet.

Анализ слов

1. Для начала будем рассматривать слова, являющиеся существительными. Рассмотрим, к примеру, слово “наушники” с использованием BabelNet.

The screenshot shows the BabelNet search results for the Russian noun 'наушники'. The search bar at the top contains 'наушники', with 'Russian' selected as the language. Below the search bar, there are three main entries:

- наушники** (имя существительное): Definition - Наушники – два телефона с оголовьем, предназначенные для подключения к радиоэлектронным аппаратам. Examples: **наушники** | **наушники** | **наушники**. Below the definition are three small thumbnail images corresponding to these examples.
- защитные наушники** (имя существительное): Definition - Защитные наушники – устройства для покрытия ушей и защиты от шума и других внешних воздействий. Examples: **защитные наушники** | **наушники**. Below the definition is one small thumbnail image.
- головная гарнитура** (имя существительное): Definition - Головная гарнитура – аксессуар систем связи, представляющий собой комбинацию головных телефонов с микрофоном и, в от... Examples: **наушники** | **наушники**. Below the definition is one small thumbnail image.

Нас будет интересовать первое значение.

The screenshot compares the BabelNet and Wikipedia definitions for 'наушники'.

BabelNet Definition: Наушники – два телефона с оголовьем, предназначенные для подключения к радиоэлектронным аппаратам. This is linked to the 'наушники' entry in the BabelNet network diagram.

Wikipedia Definition: Наушники представляют собой пару небольших по размеру головных телефонов или блока наушников[2], надеваемых на голову или вставляемых прямиком в ушные каналы. Наушники часто применяются в быту и в профессиональной деятельности для речевой связи и прослушивания музыки и проч., для компьютерных игр, когда необходима мобильность или звукоизоляция от окружающего пространства. Наушники с присоединенным к ним микрофоном образуют дополненную гарнитуру. Иногда в профессиональной деятельности вместо пары используется один отдельный наушник, он называется монитор.

A purple arrow points from the BabelNet definition to the corresponding section in the Wikipedia article.

Wikidata Item: The Wikidata item for 'headphone' (Q186819) includes the BabelNet definition and links to other languages like Russian and Tatar. The table below shows the Wikidata item details:

Language	Label	Description	Also known as
English	headphone	small loudspeakers worn on the head for listening to music or for communications purposes	circumaural headphones over-ear headphones supra-aural headphones on-ear headphones headphones
Russian	наушники	два телефона с оголовьем, предназначенные для подключения к бытовым радиоэлектронным аппаратам	наушник головные телефоны аудионаушники
Tatar	көлакчыннар	No description defined	көлакчын аудиоколачыннар
Bashkir	No label defined	No description defined	

Система выдает для данного слова 2 определения, которые ссылаются на ресурсы Wikipedia и Wikidata, а также синонимы “головной телефон”, “дужка”. При этом мы можем видеть изображение, соответствующее слову “наушники”.

Рассмотрим вкладку **СВЯЗИ**.

ОПРЕДЕЛЕНИЯ	СВЯЗИ	ИСТОЧНИКИ
Russian > Больше языков ...		
IS A	электроакустического преобразователя • Звуковое оборудование • громкоговоритель • аксессуар в одежде • устройство	
PART OF	телефонная трубка • головная гарнитура	
HAS KIND	EN earplug • телефонную трубку • шумоподавления наушники • EN Koss Porta Pro • EN Sennheiser HE-1 +3 relations	
HAS INSTANCE	EN crystal earpiece • EN Xbox 360 Wireless Headset • EN Samsung Gear IconX • EN Samsung Gear IconX • EN EarPods +2 relations	
DIFFERENT FROM	наушники • телефонная трубка • головная гарнитура • внутриканальные наушники • излучатели и приёмники звука	
USES	излучатели и приёмники звука	

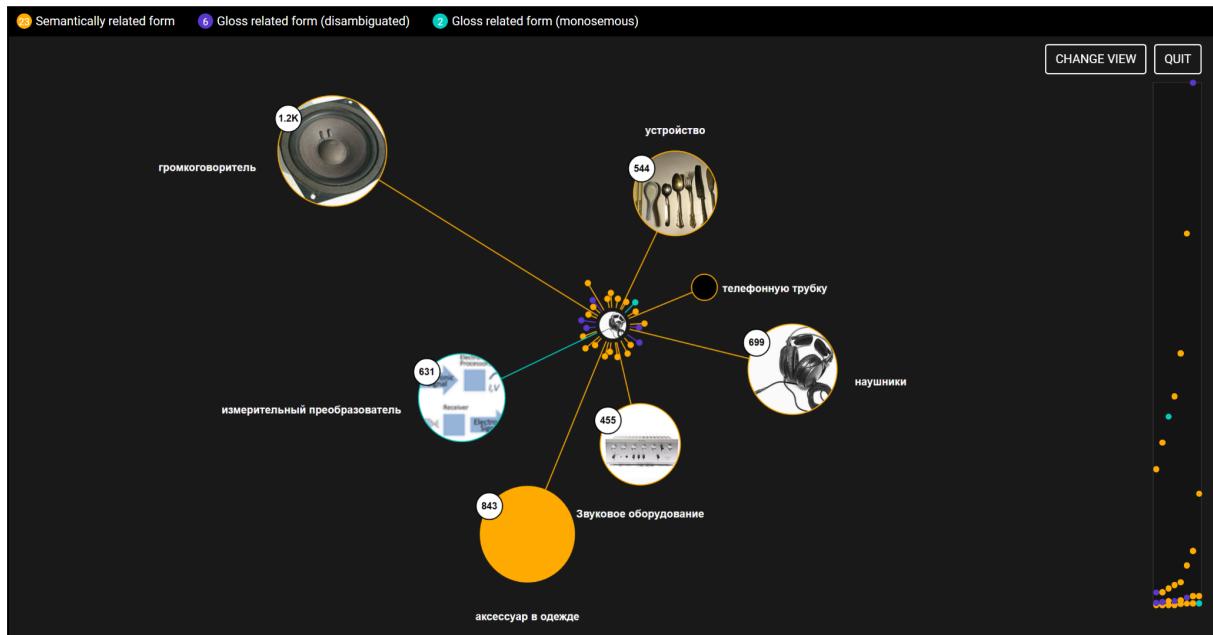
Здесь говорится, что такое “наушники” и для чего они используются. Кроме того, отсюда следует:

- слово “наушники” является частью понятий “телефонная трубка” и “головная гарнитура”;
- существует множество типов объекта “наушники”, к примеру, “earplug” (беруши);
- существуют такие экземпляры наушников, как “Xbox 360 Wireless Headset”, “Samsung Gear IconX” и другие.

Во вкладке **источники** содержится информация об источниках данных для данного слова:

ОПРЕДЕЛЕНИЯ	СВЯЗИ	ИСТОЧНИКИ
Russian > Больше языков ...		
W Wikipedia	RU наушники	
Q1 Wikidata	RU наушники	
Q1 Wiktionary	RU головной телефон, дужка, наушники	
OmegaWiki	RU наушники	
Wikipedia Redirections	RU Аудионаушники, головные телефоны, наушник, Наушники (аудиоустройство)	
Wikidata Alias	RU головные телефоны, Головные телефоны	
WordNet Translations	RU наушник, наушники	
Wikipedia Translations	RU наушники, наушников	

Также во вкладке **Explore network** мы можем детально увидеть графы знаний для текущего синсета. В данной сети можно наблюдать множество связанных со словом “наушники” понятий.



Аналогично проанализируем теперь слово “*подушка*”.

The screenshot shows the BabelNet search results for the word 'подушка'. The search bar contains 'подушка' and 'Russian'. The results are listed below:

- RU подушка** имя существительное
 - Подушка – постельная принадлежность, предназначена для поддержки головы лежащего человека во время сна.
bn:00062414n | Понятие
 - Подушка – река в России, протекает в Томской области.
bn:15692632n | Именованная сущность
 - EN A small cushion that is used for decorative purposes
bn:00077096n | Понятие

Нас будет интересовать первое значение.

Страница BabelNet для слова "подушка":

- ОПРЕДЕЛЕНИЯ:** Подушка – постельная принадлежность, предназначена для поддержки головы лежащего человека во время сна.
- СВЯЗИ:**
 - IS A: диванная подушка
 - PART OF: постельное бельё
 - HAS KIND: кровати подушку, валик
 - HAS INSTANCE: Dyna
 - DERIVATION: pillow
 - DESCRIBED BY SOURCE: Научная энциклопедия Отто Гутцзинь түшүү цээчэн
 - DIFFERENT FROM: диванная подушка
- ИСТОЧНИКИ:** Wikipedia, Wikidata

Страница Википедии о слове "Подушка":

Страница Wikidata о слове "pillow":

Система выдает для данного слова 2 определения, которые ссылаются на ресурсы Wikipedia и Wikidata.

Перейдем ко вкладке **СВЯЗИ**.

Language	Label	Description	Also known as
English	pillow	cushion for the head	Pillows
Russian	подушка	Постельная принадлежность	Подушка
Tatar	No label defined	No description defined	
Bashkir	No label defined	No description defined	

Из предоставленной информации следует:

- “подушка” является частью объекта “постельное бельё”;
- у объекта “подушка” есть следующие типы: “кровати подушки” (вероятно, имелась в виду классическая подушка, и неточная формулировка связана с WordNet Translations), “валик”, “orthopedic

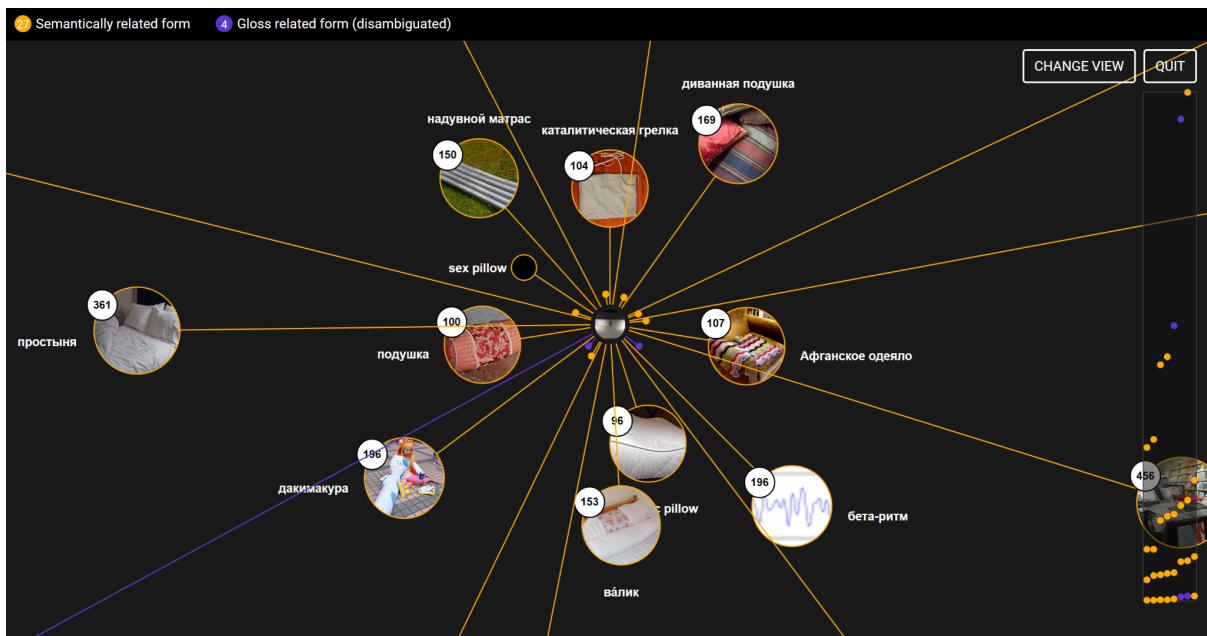
pillow" (ортопедическая подушка), "sex pillow" (секс-подушка), "дакимакура" и другие;

- конкретный экземпляр объекта "подушка" – "Dyna".

Источники данных:

ОПРЕДЕЛЕНИЯ	СВЯЗИ	ИСТОЧНИКИ
Russian > Больше языков ...		
Wikipedia		
RU подушка		
Wikidata		
RU подушка		
Wikipedia Redirections		
RU наволочка, наперник		
WordNet Translations		
RU подушка, подушки, подушку		
Wikipedia Translations		
RU подушки		

Аналогично, во вкладке **Explore network** мы можем видеть графы знаний для текущего синсета.



2. Теперь рассмотрим слова, являющиеся прилагательными.
Возьмем, к примеру, слово “**злой**”.

The screenshot shows the BabelNet 4.5.3 interface. In the search bar, 'злой' is entered. Below it, the first result is 'RU | злой имя существительное' (Noun). It includes a thumbnail of a person, a definition 'Злой (англ.)', and links to 'Именная сущность' and 'English'. The second result is 'RU | злой прилагательное' (Adjective), also with a thumbnail, a definition 'EN | Feeling or showing anger', and links to 'гнёвный • разгневанный' and 'языческий • злостный'.

Нас будет интересовать второе значение.

This screenshot shows the same BabelNet interface for the adjective 'злой'. The search bar still has 'злой'. The main result is 'RU | злой' with its definition 'Feeling or showing anger'. To the right is a network visualization with the text 'EXPLORE NETWORK'. Below the main result are sections for 'СВЯЗИ' (Relations) and 'ИСТОЧНИКИ' (Sources). Under 'СВЯЗИ', there are buttons for 'Russian' and 'Больше языков ...'. Under 'ИСТОЧНИКИ', there are buttons for 'WordNet 3.0' and 'WordNet 4.0'.

Итак, определение для данного слова – “*Feeling or showing anger*” (т.е. чувство или проявление гнева). Источник – WordNet.

The screenshot shows the WordNet Search - 3.1 interface. The search term 'angry' is entered in the 'Word to search for:' field. Below it, 'Display Options' and 'Key' are set to their defaults. The 'Adjective' section lists three definitions:

- S: (adj) angry (*feeling or showing anger*) "angry at the weather"; "angry customers"; "an angry silence"; "sending angry letters to the papers"
- S: (adj) angry, furious, raging, tempestuous, wild ((of the elements) as if showing violent anger) "angry clouds on the horizon"; "furious winds"; "the raging sea"
- S: (adj) angry (severely inflamed and painful) "an angry sore"

Синонимами для понятия “злой” являются следующие слова: “гневный”, “разгневанный”, “рассерженный”; антонимом – слово “*unangry*” (т.е. незлой, беззаботный); схожими словами – “*aggravated*” (раздраженный), “*angered*” (разгневанный) и другие.

Во вкладке **источники** видим источники предоставленных данных.

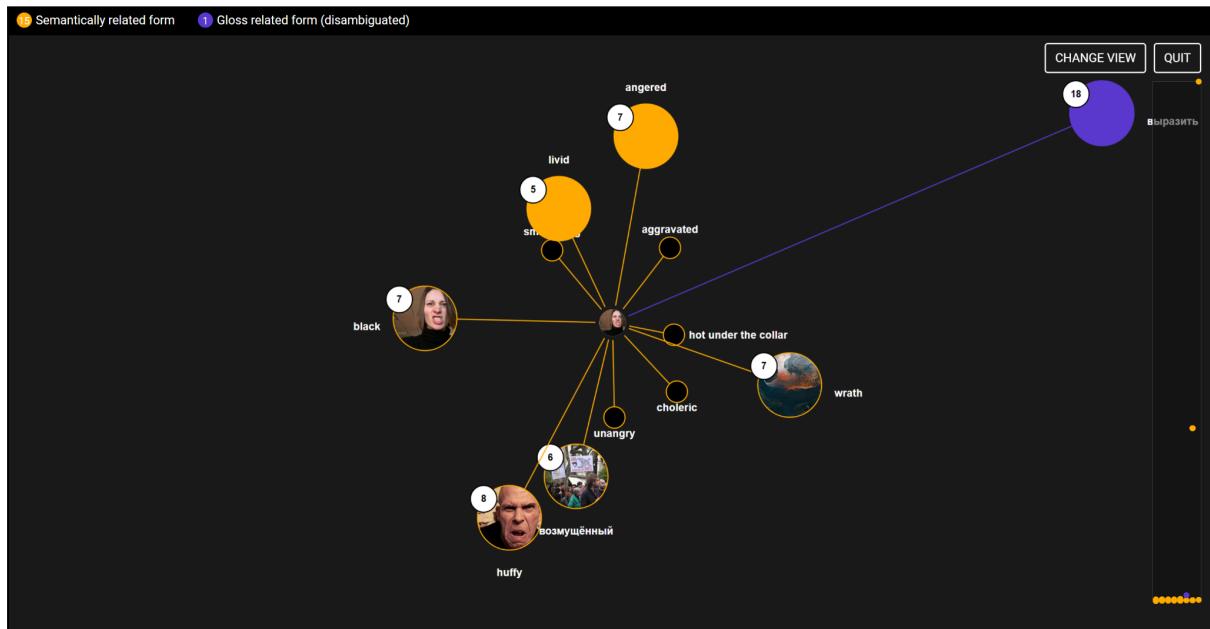
СВЯЗИ ИСТОЧНИКИ

Russian > Больше языков ...

Wiktionary RU гнёвный, злой, разгнёванный, рассёрженный, сердитый

OmegaWiki RU злой, разгневанный, сердитый

Во вкладке **Explore network** можем видеть графы знаний для текущего синсета.



Рассмотрим теперь слово “чёрный”, которое является прилагательным.

☰ BabelNet 5.3 Login | Предпочтения

чёрный Russian Перевести на... Q 78 Noun 8 Verb 1 Adj. 6 Adv. ▾

RU Чёрный прилагательное

EN Being of the achromatic color of maximum darkness; having little or no hue owing to absorption of almost all incident light

вороной bn:00098480a Понятие

BabelNet v5.3

Login | Предпочтения

чёрный

Russian | Перевести на... |

bn:00098480a | Прилагательное | Понятие

RU чёрный /a/ · вороной

EN Being of the achromatic color of maximum darkness; having little or no hue owing to absorption of almost all incident light WordNet 3.0

СВЯЗИ **ИСТОЧНИКИ**

Russian > Больше языков ...

ALSO SEE EN dark

ANTONYM белый

ATTRIBUTE яркость

DERIVATION чёрный

SAPIENZA NLP |

лицензия на ресурс |

Определение для данного слова также предоставлено сервисом WordNet.

WordNet Search - 3.1
- [WordNet home page](#) - [Glossary](#) - [Help](#)

Word to search for: Search WordNet

Display Options: Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations
Display options for sense: (gloss) "an example sentence"

Noun

- S. (n) black, blackness, inkiness (the quality or state of the achromatic color of least lightness (bearing the least resemblance to white))
- S. (n) total darkness, lightlessness, blackness, pitch blackness, black (total absence of light) "they fumbled around in total darkness"; "in the black of night"
- S. (n) Black, Joseph Black (British chemist who identified carbon dioxide and who formulated the concepts of specific heat and latent heat (1728-1799))
- S. (n) Black, Shirley Temple Black, Shirley Temple (popular child actress of the 1930's (born in 1928))
- S. (n) Black, African-American, Negro, Negroid ((000))
- S. (n) black ((board games) the darker pieces)
- S. (n) black (black clothing (worn as a sign of mourning)) "the widow wore black"

Verb

- S. (v) blacken, melanize, melanise, black (make or become black) "The smoke blackened the ceiling"; "The ceiling blackened"

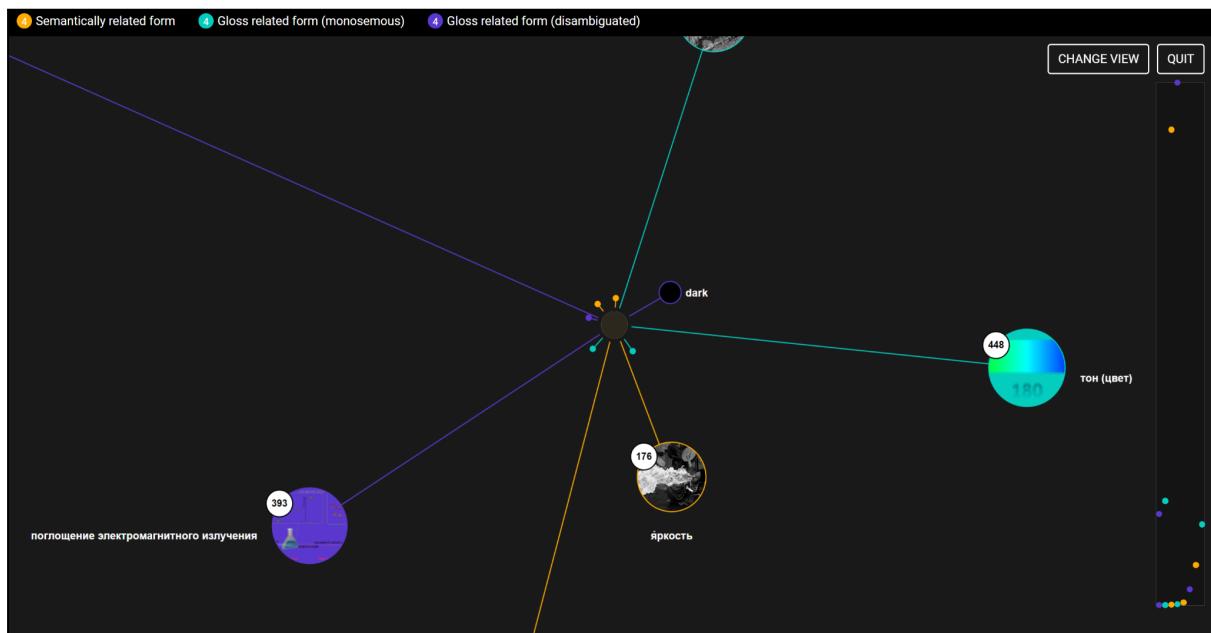
Adjective

- S. (adj) black (being of the achromatic color of maximum darkness; having little or no hue owing to absorption of almost all incident light) "black leather jackets"; "as black as coal"; "rich black soil"
- S. (adj) black (of or belonging to a racial group especially of sub-Saharan African origin) "a great people--a black people--...injected new meaning and dignity into the veins of civilization". Martin Luther King Jr.
- S. (adj) black (marked by anger or resentment or hostility) "black looks"; "black words"
- S. (adj) black, bleak, dim (offering little or no hope) "the future looked black"; "prospects were bleak"; "Life in the Aran Islands has always been bleak and difficult". J.M.Syng;
- S. (adj) black, dark, sinister (stemming from evil characteristics or forces; wicked or dishonorable) "black deeds"; "a black lie"; "his black heart has concocted yet another black deed"; "Darth Vader of the dark side"; "a dark purpose"; "dark undercurrents of ethnic hostility"; "the scheme of some sinister intelligence bent on punishing him"- Thomas Hardy
- S. (adj) black, calamitous, disastrous, fatal, fateful ((of events) having extremely unfortunate or dire consequences; bringing ruin) "the stock market crashed on Black Friday"; "calamitous defeat"; "the battle was a disastrous end to a disastrous campaign"; "such doctrines, if true, would be absolutely fatal to my theory". Charles Darwin, "it is fatal to enter any war without the will to win it"- Douglas MacArthur; "a fatal error"
- S. (adj) black, blackened ((of the face) made black especially as with suffused blood) "a face black with fury"
- S. (adj) black, pitch-black, pitch-dark (extremely dark) "a black moonless night"; "through the pitch-black woods"; "it was pitch-dark in the cellar"
- S. (adj) black, grim, mordant (harshly ironic or sinister) "black humor"; "a grim joke"; "grim laughter"; "fun ranging from slapstick clowning ... to savage mordant wit"
- S. (adj) black ((of intelligence operations) deliberately misleading) "black propaganda"
- S. (adj) bootleg, black, black-market, contraband, smuggled (distributed or sold illicitly) "the black economy pays no taxes"
- S. (adj) black, disgraceful, ignominious, inglorious, opprobrious, shameful ((used of conduct or character) deserving or bringing disgrace or shame) "Man...has written one of his blackest records as a destroyer on the oceanic islands". Rachel Carson; "an ignominious retreat"; "inglorious defeat"; "an opprobrious monument to human greed"; "a shameful display of cowardice"
- S. (adj) black ((of coffee) without cream or sugar) "with feet black from playing outdoors"; "his shirt was black within an hour"

Синонимом для данного понятия является слово “вороной”, антонимом – “белый”. Также здесь указано, характеристикой чего является понятие “чёрный”, а именно – “яркости”.

Во вкладке **источники** можно видеть источники предоставленных данных (в данном случае это сервисы Wiktionary и OmegaWiki).

Аналогично, во вкладке **Explore network** можем наблюдать графы знаний для текущего синсета.



3. Рассмотрим слова, являющиеся глаголами.

Проанализируем слово “**смотреть**”.

The image shows two side-by-side search results for the word "смотреть". On the left is the BabelNet interface, which provides a detailed semantic network visualization and a list of related words like "глазеть", "прыгать", and "глядеть". On the right is the WordNet Search interface, which shows the definition for the noun sense of "stare" as "(a fixed look with eyes open wide)" and provides examples for the verb sense, such as "gaze, stare (look at with fixed eyes)" and "stare (fixate one's eyes)". A purple arrow points from the BabelNet results towards the WordNet results, highlighting the difference in source information.

Как можно заметить, источником определения является сервис WordNet.

Синонимы – “глазеть”, “пристально смотреть”, “пялиться”, “таращиться”, “уставиться”.

Типы для данного понятия – “*stare down*”, “*stargaze*”, “*mug*”.

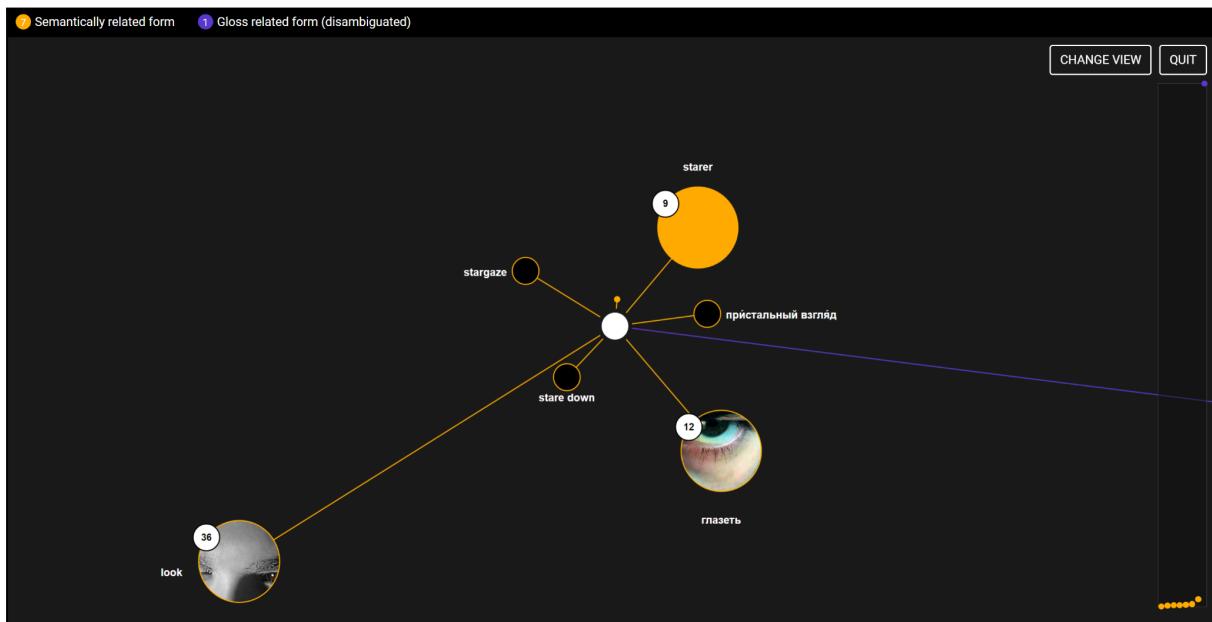


Источники предоставленных данных:

The screenshot shows a Russian language interface with two main sections:

- СВЯЗИ** (Connections):
 - Russian** > **Больше языков ...**
 - Wiktionary**: RU глазеть, пристально смотреть, пялиться, таращиться, устáвиться
 - WordNet Translations**: RU смотреть
- ИСТОЧНИКИ** (Sources): This section is currently empty.

Так же, как и для других слов, во вкладке **Explore network** находятся графы знаний для текущего синсета.



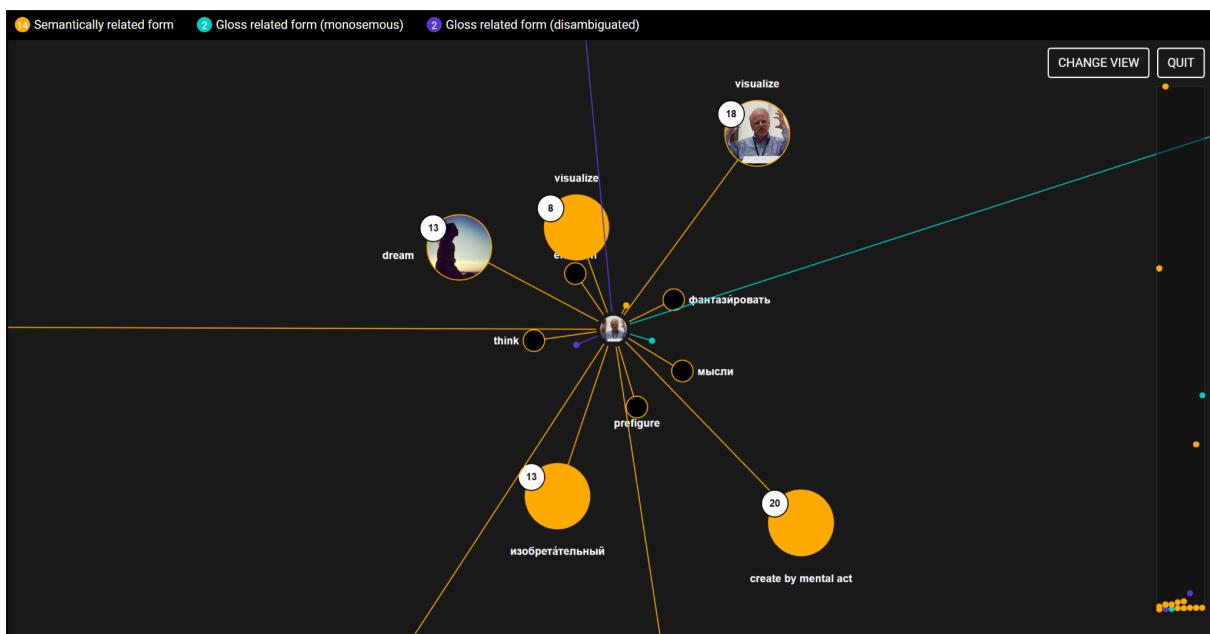
Рассмотрим также глагол “*представлять*”.

Рассмотрим определение для данного слова. Оно взято из WordNet.

Синонимы для данного слова – “воображать”, “вообразить”, “обдумывать”. Определение для данного слова определяется как “*create by mental act*”, т.е. “создавать мыслями”. Типы для понятия “представлять” – “*dream*”, “*visualize*”, “*envision*” (мечтать, визуализировать, воображать соответственно).

Ниже указаны источники информации и сопутствующие ссылки.

Аналогично, во вкладке **Explore network** можем видеть графы знаний для текущего синсета.



Сравнение с RuWordNet

Последовательно исследуем каждое проанализированное слово в системе RuWordNet.

Для слова “**наушники**” информации не найдено.

Рассмотрим слово “**подушка**”. Для него в системе RuWordNet найдено 2 значения.

The screenshot shows the RuWordNet search interface with the query 'подушка' entered. It displays two main semantic types for 'подушка': 'ПОДУШКА' (cushion) and 'Подушка 2' (pillow). The left panel for 'ПОДУШКА' lists various related terms like 'ПОДУШЕЧКА', 'ПОДУШКА ДЛЯ СНА', and 'ПОДУШЕЧНЫЙ'. The right panel for 'Подушка 2' provides detailed definitions, including its use as a 'soft pillow for use on a bed' and its association with 'ПОСТЕЛЬНАЯ ПРИНАДЛЕЖНОСТЬ'. Both panels include links to their respective WordNet entries and interlingual indices.

Здесь так же, как и в системе BabelNet, можно видеть определения, в том числе из сервиса WordNet, но в более сокращенном виде.

В отличие от BabelNet, здесь указываются однокоренные слова, что дает более подробную информацию об исходном слове.

Также можно видеть значительно более обширное определение гиперонимов и гипонимов, причем написанные полностью на русском языке, в отличие от BableNet.

Еще одним отличием является то, что в данном сервисе нельзя визуально увидеть графы знаний, как это можно было наблюдать в BableNet, причем это относится ко всем словам.

Проанализируем слово “**злой**”.

ЗЛОЙ прилагательное

Злой злой, недобрый

Синсет

[1 ЗЛОЙ, НЕДОБРЫЙ](#)

[Понятие РуТез: злой, недобрый]

однокоренные слова

[1 ЗЛОБНЫЙ, ЗЛОБА, ЗЛОСТЬ, ЗЛОбНОСТЬ](#)

гипероним

[1 НЕДОВОЛЬНЫЙ](#)

[Понятие РуТез: недовольство, неудовольствие]

Здесь мы наблюдаем несколько меньший, по сравнению с BableNet, синсет, который состоит лишь из 2-х слов, и отсутствие конкретного определения. Тем не менее, здесь присутствует раздел, посвященный однокоренным словам и гиперонимам, чего нет в BableNet. Важным отличием является и то, что информация, предоставленная сервисом RuWordNet, идет полностью на русском языке с учетом его тонкостей, избавляя пользователя от необходимости перевода с английского языка, при котором может появиться некоторая расплывчатость в определении слова, т.к. зачастую слово при переводе с английского языка имеет не единственное значение.

Изучим слово “**черный**”.

Тезаурус русского языка RuWordNet

[черный] [Искать] [GB]

ЧЕРНЫЙ прилагательное

Черный черный цвет

Синсет

1 [ЧЕРНОВАТЫЙ](#), [ЧЕРНЫЙ](#)
[Понятие Рутез: черный цвет]

однокоренные слова

1 ЧЕРНОВАТЫЙ, ЧЕРНОТА, ПОЧЕРНELЫЙ, ПОЧЕРНЕНИЕ, ПОЧЕРНЕТЬ, ЧЕРНеть, ЧЕРНЕТЬ, ЧЕРНЕТЬСЯ, ЗАЧЕРНЕТЬ, ЧЕРНОВОЛОСЫЙ, ЧЕРНЕНИЕ, ЗАЧЕРНЕНИЕ, ЧЕРНИТЬ, ЗАЧЕРНИТЬ, НАЧЕРНИТЬ, ЧЕРНО-БЕЛЫЙ

Interlingual Index

1 [i2222 black.a.01 \(black\)](#)
[Определение Wordnet: being of the achromatic color of maximum darkness; having little or no hue owing to absorption of almost all incident light]

гипероним

1 [ЦВЕТОВОЙ](#)
[Понятие Рутез: цвет, окраска]

домен

1 [ОКРАСКА](#), [ОКРАШЕННОСТЬ](#), [ЦВЕТ](#)
[Понятие Рутез: цвет, окраска]

гипоним

1 [МАРЕНГО](#)
[Понятие Рутез: маренго]

частеречная синонимия

1 [ЧЕРНОТА](#), [ЧЕРНЫЙ ЦВЕТ](#)
[Понятие Рутез: черный цвет]

В отличие от BableNet, здесь можно видеть гораздо более обширное описание для данного слова, а именно:

- синсет;
- большое множество однокоренных слов;
- определение из WordNet;
- гипероним;
- домен;
- гипоним;
- частеречная синонимия.

Преимущество RuWordNet перед BableNet здесь очевидно.

Необходимо также сделать замечание, что слова, которые являются прилагательными, система BableNet очень часто определяет как существительное или не определяет вовсе, что является некорректным. Подобные ситуации можно наблюдать при исследовании слов “красивый”, “красный”, “телефонный”, “чистый”.

Перейдем к глаголам. Проанализируем слова “*смотреть*” и “*представлять*”.

Система определяет для данных слов множество значений, причем описание для каждого из них очень велико, и в данном отчете не приводится. Очевидно, эти слова описано гораздо более подробно на сервисе RuWordNet.

Выводы

При выполнении данного задания были изучены принципы работы сервиса BableNet. Были подробно исследованы представления для некоторых слов русского языка трех частей речи – существительного, прилагательного и глагола, а также их синонимы, отношения с другими словами и переводы. После подробного анализа каждого из слов было проведено сравнение сервисов BableNet и RuWordNet. Последний показал заметно более качественные результаты за счет того, что является русскоязычным и лучше учитывает тонкости русского языка, а также благодаря более структурированным и корректным данным. Тем не менее, не все проанализированные слова оказались на RuWordNet, поэтому данная система все еще нуждается в расширении набора слов.