

РУСКО-ЧЕШКА ЛЕКСИКАЛНА БАЗА ДАННИ

Руско-чешката електронна лексикална база данни е създадена от колектив лексикографи от Секцията за славянска лингвистика и лексикография към Славянския институт на Чешката академия на науките (ЧАН). Работата по проекта започна през 2007 г. в рамките на грант, отпуснат от ЧАН с название *Комплексно съпоставително описание на лексиката на руски и чешки език с отчитане на нейната динамика* (2007–2011). Работата върху речниковата база данни продължи и след приключване на гранта в края на 2011 г. В началото на 2014 г. базата данни е пусната в пробна експлоатация. Качена е в интернет на адрес <http://slovník.slu.cas.cz/slovník/main/index.html> и е осигурен свободен безплатен достъп за потребители. В настоящия момент са готови всички букви освен К, Л и П, които ще бъдат допълнени. Завършената до този момент част включва около 50 000 ключови думи (речникови статии), като се предвижда готовата база данни да съдържа около 75 000 статии.

Целта на руско-чешката лексикална база данни е да служи като електронен ресурс, който да се използва на първо място като източник за описание на актуалното състояние на руската лексика в сравнение с чешката, и на второ място – като голям електронен руско-чешки речник, отразяващ актуална и употребявана лексика. По този начин базата данни съчетава характеристики на класически двуезичен речник и възможностите за извличане на информация от ресурси в електронен формат. Тук ще се съсредоточим върху особеностите на разглежданата база данни, които я отличават от речник в книжен формат, и по-точно – върху нейните предимства в сравнение с класически речник. Поради тази причина микроструктурата на речника не е предмет на това кратко описание на руско-чешката лексикална база данни.

1. Технически параметри

Базата данни се управлява от системата PostgreSQL. За редактиране на речниковите статии в нея бе създаден софтуер, основан на архитектура Client-Server. Това дава възможност на авторите да работят в интернет от всяко място. Софтуерът се състои от компонента Application Server, като се използва технология Apache Tomcat и Java Servlet. Интерфейсът е създаден чрез Google Web Toolkit.

2. Словник

Словникът на лексикалната база данни е подбран от следните руски тълковни речници: *Толковый словарь русского языка* под редакцията на Ожегов (1999), *Толковый словарь русского языка* под редакцията на Шведова (2008) и *Толковый словарь русского языка начала XXI века* под редакцията на Скляревская (2006), съдържащ най-актуалната лексика. В речника се включват и

лексикални единици от двуезичния *Русско-чешский словарь* в два тома под редакцията на Венцовская. Словникът е допълнен и с неологичен материал от *Русско-чешский и чешско-русский словарь неологизмов*, създаден в края на 90-те години също в Секцията за славянска лингвистика и лексикология към Славянския институт на ЧАН, а също и с материали, които са ексцерпирани от авторите на базата данни.

При създаване на словника на руско-чешката лексикална база данни авторите се ръководиха от стремежа тя да предоставя информация за актуална лексика на руски (и следователно и на чешки) език в най-широкия смисъл и да отразява нейната динамика. С оглед на целта да се създаде цялостна картина на лексиката на двата езика в словника е включена не само общоупотребима книжовна лексика, но и определено количество думи от други слоеве на езика – сленгови, некнижовни или диалектни. Освен актуална лексика в лексикалната база данни присъстват и някои остарели или архаични думи, които потребителите биха могли да срещнат например в художествената литература. Част от тези думи обозначават реалии, които вече не съществуват, като в тази категория попадат и редица названия на обекти и понятия от не толкова далечното минало, най-вече свързани със социалистическата действителност. Такъв тип лексика се придружава от съкращението *dř.* (което означава ‘по-рано’). В базата данни са включени и термини от редица специални области (както общоупотребими, така и извън широка употреба), като се избягва описанието им в подробности и дълбочина, присъщо за терминологичен речник. При създаването на словника бе обърнато особено внимание на включването на нови думи в него, тъй като неологията и неографията имат дълга традиция в Славянския институт на ЧАН. Словникът е допълнен не само с неологизми, но и с лексеми, които присъстват отдавна в езика, но не са отразявани в съществуващите преди речници или са отразявани само като част от словообразователното гнездо на мотивиращата дума. Тук се включват названията на жени, образувани от названия за мъже от категориите *nomina agentis* и *nomina actoris*, които в руския език често носят разговорен характер (напр. *адвокатесса*); отглаголни съществителни (напр. *жонглирование*); абстрактни съществителни (напр. *заблаговременность*); наречия, образувани от прилагателни със суфикс *-о* (напр. *мирно*).

3. Работа с лексикалната база данни

Работен език, чрез който се използва лексикалната базата данни, е чешкият, но при преминаване от демоверсия към пълна експлоатация инструкциите за работа и упътването в базата ще бъдат преведени и на руски. Преди да започнем описанието на начина на използване на базата данни, е необходимо да направим уточнението, че се работи винаги само с една определена (избрана) буква. Следва се този начин на работа, защото поради обема на базата данни използването ѝ би ставало бавно, ако трябва да се зарежда цялата база. Ако е нужно да се търсят думи, започващи с различни букви, се налага да се зареждат последователно различните букви. Смяната на буквите се извършва чрез кликване върху желаната буква в списъка в горния ляв ъгъл. За удобство данните, посочени в заявката за търсене, остават същите и при промяна на буквата, в която се търси, поради което не е необходимо да се въвеждат отново.

4. Търсене на дума

Лексикалната база данни представлява съчетание на класически речник с новите възможности, основани на използване на компютърни технологии. Предполагаме, че тя ще бъде употребявана предимно във функция на двуезичен речник. Търсената дума може да се намери в лексикалната база данни по два начина. При първия начин след влизане в базата данни (чрез избор на надпис VSTUP) се кликва върху команда VYHLEDÁVÁNÍ (Търси) и след това се избира команда Vyhledávání (Търси). В отворения прозорец се маркира опцията klíčové slovo (ключова дума) и začíná na (започва с) и в прозорчето след този избор се вписва исканата дума, след което се избира Set (вж. фиг. 1).

След такава заявка в лявата страна на екрана се появява списък с думи, започващи с търсената дума (ако има такива). При по-кратки думи е възможно да се появи цяла поредица от думи, започващи с посочената комбинация от букви. Този начин на търсене в редица случаи дава информация и за лексеми, сродни на търсената. Колкото по-дълги са търсените лексеми, толкова по-кратък е списъкът, който се появява. От списъка се избира интересуваща ни дума и след маркирането ѝ се отваря нов прозорец с нейната речникова статия.

RUSKO-ČESKÁ SLOVNÍKOVÁ DATABÁZE Poslat připomínky

ÚVOD ZPRACOVÁNÍ HESEL VYHLEDÁVÁNÍ SEZNAM ZKRATEK

Klíčové slovo začíná na
 obsahuje
 končí na

Indexy: ADJ.1

V textu:

Set

А Б В Г Д Е Ж З И Й К Л
М Н О П Р С Т У Ф Х Ц Ч Ш
Щ Ъ Ы Ь Э Ю Я

а I
а II
а III
а IV
абажур
абазинец
абазинна
абазинский
аббат
аббатиса
аббатский
аббатство
аббревиатура
аббревиатурный
абerrационный
абerrация
абьц
абиссианец
абиссинка
абиссинский
абитуриент
абитуриентка
абитуриентский
абонемент
абонементный
абонент
абонентка
абонентный
абонентский
абонировать
абордаж
абордажный
абориген
аборигенка
аборигенный
аборт
абортивность
абортивный
Сейкел: 2031

Фиг. 1. Заявка за търсене на думата **рай**

Речникова статия на думата **рай**:

рай^{M91}, рая *m*

1. *náb.* ráj, nebo *s*; как в раю jako v ráji.

2. ráj; летом в лесу ~ v létě je v lese ráj *n.* překrásně; земной ~ ráj na zemi; *nespis.* иди к богу в ~! padej, vypadni

♦ *přísl.* рад бы в ~, да грехи не пускают rád by, ale nejde to; *přísl. hovor. hanl.* на чужом горбу в ~ въехать хочет šplhá vzhůru po cizích hřbetech, dostává se k lepšímu cizí práci

Вторият начин на търсене на определена дума е чрез скролиране в списъка с думите от лявата страна на екрана, като се кликва върху избраната лексема. Ударението на думите не е от значение при търсенето на съответната дума, тъй като то не се отбелязва при лексемите в списъка. В речниковите статии обаче то се отбелязва в целия руски текст.

5. Улеснения за потребителя

Базата данни предлага някои улеснения за потребителя, които са възможни благодарение на компютърната обработка на материала.

А) Едно от тях са граматичните индекси. Индексирани са думи от флективните части на речта, по-точно съществителни и прилагателни имена, местоимения и числителни. Индексирани са също така и наречията, които се степенуват. В настоящия момент още не са индексирани глаголите, като това е планирано да се извърши в следващия етап от работата по базата данни. Разработката на граматичните типове е много подробна с цел предоставяне на повече инфор-

RUSKO-ČESKÁ SLOVNÍKOVÁ DATABÁZE

ÚVOD ZPRACOVÁNÍ HESEL VYHLEDÁVÁNÍ SEZNAM ZKRATEK

фактор
факторие
факторисей
фактория
фактура
фактурный
факультатив
факультативность
факультативный
факультет
факультетский
факт
факва I
факвага II
факвага III
факвагистский
факвад
факвадль
факваерист
факваеристика
факваеристика
факваический
факвалос
факваификат
факваификатор
факваификаторский
факваификаторша
факваификационный
факваификация
факваифицированный
факваифицировать
факваистарт
факваич
факваический
факваический
факваичевание
факваичевать
факваичет
факваичетный
факваичка
факваичковный
факваичковик
факваичковская
Сейкел 1140

фактура^{F1}

1. (стекло, бумага) Тур: комната за textura, kresba.

2. (картинка)

3. экон. тѣл

	Sg	Pl
N	комната, дорога	комнаты, дороги
G	комнаты, дороги	комнат
D	комнате	комнатам
A	комнату	комнаты, дороги живущим
L	о комнате	о комнатах
I	комнатой	комнатами

Фиг. 2. Граматична таблица към индекс F1

мация за обикновените потребители. В базата данни се използват 314 индекса (без глаголните индекси), които препращат към граматични таблици. По-точно граматичната таблица с парадигмата на съответната дума се визуализира на екрана след кликуване върху индекса, като не е необходимо да се излиза от прозореца с разглежданата речникова статия (вж. фиг. 2).

При някои ключови думи се посочват два индекса. Това се извършва при прилагателни, които може да се степенуват (больш | ой^{ADJ1.COMP20}). В такъв случай единият индекс се отнася към типа склонение, а другият – към типа степенуване. Два индекса се посочват и при ключови думи с различен тип склонение в отделни техни значения, като при отделните значения е даден само релевантният индекс. Вж. например речниковата статия на лексемата *сын*:

сын^{M24,M1}, -а *m*

1. *mn.* ~овья^{M24} (*старший*) *syn*; родной *сын* *vlastní syn*; приёмный ~ *adoptovaný n adoptivní syn* ♦ *bibl.* блудный ~ *marnotratný syn*.

2. *mn.* ~ы^{M1} *kniž.* (*родины, революции, нашего века*) *syn*

При съчетаване на две думи, една от които не се скланя, индексът се посочва при склоняемата дума, например:

иван-чай^{M3}, иван-чая *m vrbovka ž* ■ *bot.* *vrbovka úzkolistá*.

Ако са склоняеми и двете думи, двата индекса се намират след втората дума, например:

имя-отчество^{N26-N1}, имени-отчества *s*

Б) Друг вид улеснение за потребителя представляват линковете от една речникова статия към друга. Така например при глаголи от несвършен вид има линк към глаголи от свършен вид и обратно. При умалителните съществителни се дава линк към основната дума, а при отглаголни съществителни – към мотивиращия глагол, например:

бомбардировк | а^{FS}, -и *ž voj.* *bombardování s*, (*артиллерийская*) *ostřelování s*; точечная ~а *bodové bombardování*; - *srov.* бомбардировать

6. Типове заявки

Както вече се каза, лексикалната база данни съчетава характеристиките на класически двуезичен речник с възможностите за извличане на информация, основани на компютърната обработка на данните. Тези възможности са свързани на първо място с използване на определен филтър в заявката за търсене.

Базата данни предлага три възможности за заявка чрез определен филтър: по ключова дума, по индекс и в текст (в рамките на речниковата статия). Отново е необходимо да се напомни, че при заявка чрез определен филтър се търси не в цялата база данни, а само в рамките на една буква. За получаване на общ резултат е необходимо да се направи съответна заявка за всички отделни букви една по една и след това така получените данни може да се съберат в един файл.

6.1. При търсене **по ключова дума** има три критерия за избор: (ключовата дума) *започва с*, *съдържа* и *свършва с*. Избира се една от тези възможности и се въвеждат желаните букви, за чийто брой няма ограничения.

Търсене по ключова дума и по критерий *започва с*

По този начин се търсят както съответни лексеми, така и поредица от думи с един и същ префикс или префиксоид. Във връзка с втората възможност трябва

да се отбележи, че в базата данни са включени ограничен брой думи, формирани чрез т. нар. серийно словообразуване.

Търсене по ключова дума и по критерий свършва с

Този начин на търсене е полезен, когато е необходим подбор на думи с един и същи суфикс или при сложни думи с еднакъв втори компонент. Например при търсене по този начин на лексемата *кино* има намерени две думи: *арт-кино* и *стереокино*.

6.2. При търсене **по индекс** се избира от категории в списъка с индекси. Названието на индекса се състои от съкращение на латинското название на част на речта или граматическа категория (например: F – *femininum* – съществително женски род, M – *maskulinum* – съществително мъжки род, COMP – *comparativ* – индекс за степенуване на прилагателни имена и наречия и пореден номер на таблицата.

6.3. Широк спектър от възможности предоставя търсенето **в текст** чрез определена заявка. По принцип по този начин може да се намери каквато и да е дума от речникова статия, но най-полезно е търсенето по граматична и стилистична характеристика, по специална сфера на употреба, по окончание или по друга характеристика като например „поговорка“. В резултат на пълнотекстово търсене от базата данни може да се извлекат всички термини от една специална област (при търсене в базата данни буква по буква, както вече се посочи), всички книжни или остарели думи, всички думи с експресивна отсянка или всички думи от определена част на речта (например всички глаголи от свършен/несвършен вид). Възможно е също чрез такъв тип търсене да се извлекат лексеми, употребени в безлични изречения. Друга възможност е получаването на списък с всички думи с определено падежно окончание (напр. с окончание *-а* за генитив на мъжкородови съществителни имена).

Типовете заявки за търсене (по ключова дума, по индекс, по текст) може да се комбинират, но не и да се повтарят. Например не може да се търсят наречия с разговорен характер, а само наречия или само разговорни думи като цяло.

* * *

Тук се очертаха някои от възможностите за работа с руско-чешката лексикална база данни. При довършителната работа върху този лексикален ресурс предстои да се завърши редактирането на речниковите статии на трите букви, които все още не са напълно готови. Следващият етап от работата ще включва някои подобрения от формална гледна точка като например превод на работните инструкции на руски или създаване на версия за достъп от мобилни телефони. Предвижда се освен това лексикографска обработка на вече готовите речникови статии, свързана с въвеждане на синоними и антоними към ключовата (заглавната) дума или с разширяване на техните редове там, където те вече са посочени. Следващият етап от работата до известна степен ще зависи и от резултатите от пробната експлоатация на базата данни като демоверсия.

Лексикалната база данни е ценно работно пространство за лингвисти, като едновременно с това тя представлява двуезичен руско-чешки речник с голям обем и в електронен вид, който е лесно достъпен, удобен за работа и полезен за широк кръг ползватели.