

О.Ю. Савина

ИНТЕГРАЦИЯ ИНСТРУМЕНТОВ ЛИНГВИСТИЧЕСКОГО ПОИСКА В ЛИНГВОДИДАКТИЧЕСКИХ ЦЕЛЯХ

*Лингвисты, преподаватели, учителя и студенты-германисты осуществляют поиск лексикографической информации к немецкоязычным текстам с помощью традиционных и сетевых словарей и баз данных. Для повышения эффективности поиска в сетевых источниках информации автор предлагает к использованию новый программный продукт для генерации немецкоязычного тексториентированного лексического минимума *Vokabelliste*. В статье представлены основные характеристики программы.*

Ключевые слова: интеграция лингвистического поиска, немецкоязычный тексториентированный лексический минимум.

Поиск лексикографической информации к текстам в сетевых словарях и лингвистических корпусах осуществляется по разным алгоритмам в зависимости от его целей. Накопленная разнородная информация зачастую требует детальной ручной обработки. Для автоматизации процесса поиска и форматирования лексико-графической информации (для немецкого языка) была разработана программа *Vokabelliste* [1]. В данной статье охарактеризованы целевые группы пользователей программы, описаны источники лексикографической информации и подробно представлен функционал предлагаемого программного продукта.

В области лингводидактики можно выделить три целевые группы пользователей, стремящихся к получению углублённой лингвистической информации, это преподаватели специального вуза, школьные учителя и студенты, изучающие немецкий язык как специальность.

Пользователь *преподаватель вуза* преподаёт теоретические лингвистические дисциплины студентам-германистам и осуществляет поиск при подготовке практического (справочного) материала для лингвистического анализа текста. Например, для подбора информации о терминах, заимствованиях, метафорах, метонимиях, культурно-специфической лексике, территориально окрашенной лексике, архаизмах, неологизмах; для выбора лексики при построении семантического / функционального поля; для сбора

синонимов к отдельным лексемам текста; для подбора примеров употребления слов; для анализа коллокаций (на основе сведений о частотности); для анализа грамматических категорий слова, и др.

Пользователь *школьный учитель* обучает детей и подростков немецкому как иностранному языку (на разных уровнях) и использует лингвистические сетевые ресурсы при подготовке справочного материала к тексту для чтения. Например, для сбора новой для учеников лексики и ее переводов; для подбора примеров употребления слов в минимальном контексте, для выбора грамматических парадигм слов (имеющих сложные грамматические формы); для составления лексических парадигм (синонимы, антонимы, семантические поля); для поиска транскрипции слов, трудных для произношения и др.

Пользователь *студент* изучает немецкий как специальность (первый или второй иностранный язык) и осуществляет сбор справочного материала для чтения текста. Например, для перевода незнакомых слов; для поиска толкований значений слов; для поиска грамматических парадигм сложных слов; для подбора синонимических рядов, антонимических пар, семантических полей, коллокаций и типичного левого / правого окружения; для проверки произношения незнакомых / сложных слов и др.

Для решения вышеперечисленных поисковых задач наряду с традиционными словарями часто используются электронные (сетевые) словари, например: толковый интернет-словарь современного немецкого языка *DWDS* [2], сетевая лексическая информационная система немецкого языка *OWID* [3], онлайн-словарь *Duden* [4], поисковая лексическая система *Wortschatz* университета Лейпцига [5], грамматическая информационная система *GRAMMIS* [6], база данных Немецкие словари и грамматика *CANOO* [7], многоязычный онлайн-словарь *PONS* [8] и др. Кроме того, часть подобных задач может успешно решаться с помощью лингвистических корпусов, таких как Мангеймский корпус немецкого языка [9], Архив устного немецкого языка [10] и др.

Несомненное преимущество использования для лингвистического поиска именно онлайн-ресурсов – актуальность информации (обычно все подобные ресурсы обновляются в режиме реального времени). Наиболее авторитетные онлайн-словари и корпусы содержат исключительно аутентичные примеры функционирования

лексических единиц, что также свидетельствует об их высоком качестве.

К недостаткам можно отнести необходимость многократного и повторного обращения к одному или нескольким ресурсам, если поиск включает не один поисковый запрос, а несколько (например, при работе с текстом). Собранная таким образом информация обычно довольно разнородна и требует кропотливой ручной обработки – сортировки по типу информации, а также приведения к единому формату.

Стремление нивелировать названные недостатки с помощью одного поискового инструмента, интегрирующего лингвистический поиск по разным немецкоязычным сетевым ресурсам, привело к созданию *Программы генерации текстоориентированного лексического минимума (Recherchetool zur Erstellung von textbezogenen Vokabellisten)*, или *Vokabelliste*.

Программа *Vokabelliste* предназначена для автоматизации процессов сбора лингвистической информации из немецкоязычных лингвистических баз данных, функционирующих в режиме онлайн. Она позволяет оптимизировать трудоёмкий процесс дидактизации текстов для чтения, что способствует повышению эффективности труда учителя / преподавателя немецкого языка. Пользователями программы могут быть также студенты, изучающие немецкий язык.

Традиционно процесс сбора лингвистической информации с целью дидактизации иноязычного текста включает следующие шаги:

- 1) выбор лексем, подлежащих включению в лексический минимум, из текста для чтения;
- 2) отбор лингвистической информации по каждой из лексических единиц из разных баз данных;
- 3) приведение собранной информации в дидактически удобный вид.

При использовании программы *Vokabelliste* пользователь выполняет вручную только первый шаг: вводит текст в рабочее поле, отмечает в нем лексические единицы, для которых должна быть собрана лингвистическая информация, и указывает необходимые типы лингвистической информации.

Два следующих шага программа осуществляет автоматически: выполняет запрос к одной / нескольким базам данных по каждой из лексических единиц, маркированных в исходном тексте, копирует

отмеченные типы информации, соединяет их в единый документ и приводит в стандартный вид (рис. 1).

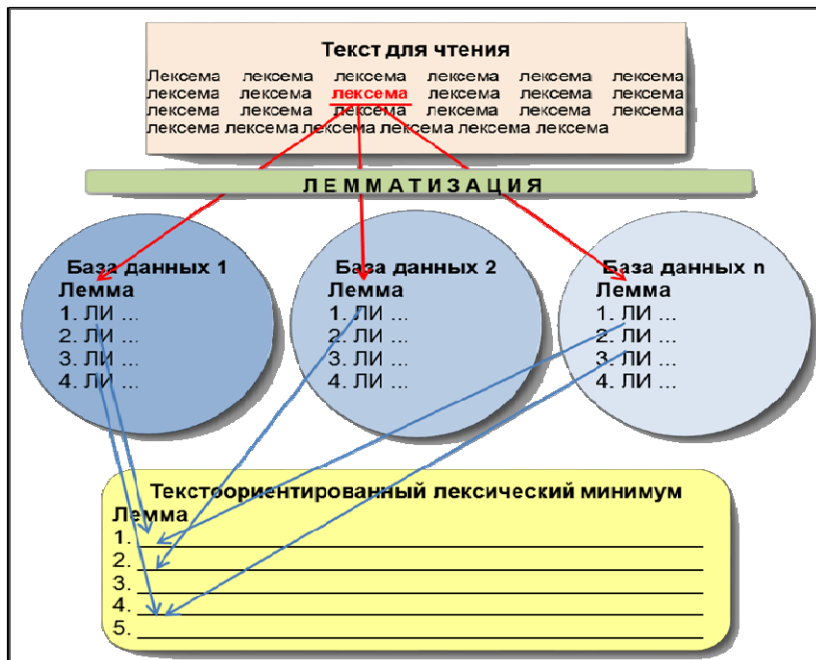


Рис. 1. Логическая модель функционирования программы *Vokabelliste*
(ЛИ – лингвистическая информация)

Программа обеспечивает возможность выполнения следующих функций:

- 1) импортирование текстового файла;
- 2) выделение и отбор лексем в импортированном тексте;
- 3) выбор типов лингвистической информации из предлагаемых базами данных в качестве опций (отдельно для каждой лексемы);
- 4) направление запроса к соответствующим базам данных;
- 5) отбор соответствующих типов лингвистической информации;
- 6) накопление лингвистической информации в одном текстовом документе;

7) приведение собранной информации к унифицированному виду.

Интерфейс программы существует в двух языковых версиях – на немецком языке (по умолчанию) и на русском языке. При запуске программы открывается главное (первичное) окно (рис. 2). В нем расположены все функциональные элементы.

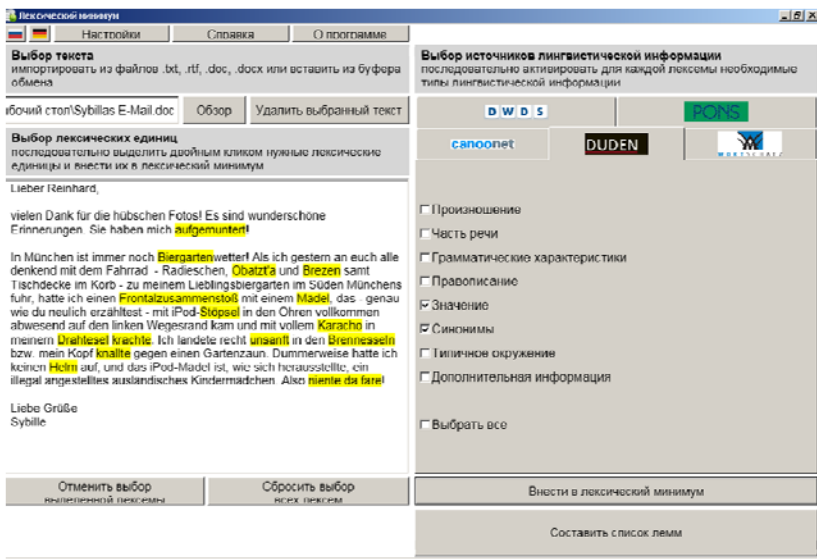


Рис. 2. Интерфейс программного продукта для автоматизированной генерации лексического минимума (русскоязычная версия)

Основные технические характеристики компьютера, при которых функционирует программа:

- операционная система Windows XP и выше,
- предустановленный текстовый редактор MS Word.

Пользователь программы должен обладать элементарными навыками пользователя ПК, иметь общее представление об используемых программой лингвистических сетевых базах данных,

понимать, какие типы информации необходимы ему для включения в генерируемый лексический минимум, отдавать себе отчет в возможном возникновении лингвистических трудностей при отборе лингвистической информации (по причине многозначности и омонимии) и осознавать необходимость проверки готового тексториентированного лексического минимума.

Входные данные (немецкоязычные тексты) загружаются в программу в виде текстового документа или через буфер обмена. Отмеченные пользователем лексемы направляются в виде запроса в лингвистические базы данных. Программа ведет отбор информации указанных типов и собирает ее в единый документ. Выходные данные (собранный лингвистическая информация) организованы в виде текстового документа, сгенерированного в текстовом редакторе MS Word. Программа автоматически приводит лексический минимум в стандартный формат. На этом поисковый запрос считается выполненным.

Семантическими особенностями лексических единиц (омонимия, полисемия) может быть обусловлено то, что результат будет содержать избыточную информацию. В таком случае он может быть откорректирован пользователем вручную. Трудоёмкость ручной работы напрямую зависит от цели и количества собранной информации – лишнее можно просто удалить из результата (текстового документа).

В готовом лексическом минимуме информация упорядочена по типу: орфографические и фонетические особенности (произношение, правописание); грамматическая информация (часть речи, грамматические характеристики, формы слова); семантическая информация (значение); примеры употребления (примеры, соотносенные со значениями, типичные примеры); парадигматические связи (в том числе синонимы); коллокации (типичное левое и правое окружение); перевод на русский язык. Источники информации указаны в виде логотипа соответствующей базы данных (рис. 3).

Kindermädchen	
Orthographische und phonetische Besonderheiten Орфографические и фонетические особенности	
Aussprache Произношение	Betonung: Kindermädchen Lautschrift: ['kindɛmɛ:tʃən]
DUDEN	
Grammatische Informationen Грамматическая информация	
Wortformen Формы слова	Wortklasse: Nomen Genitiv Singular: Kindermädchens Nominativ Plural: Kindermädchen Artikel: das Flexionsklasse: s/- Besonderheiten: n-Tilgung im Dativ Plural
canoonet	
Semantische Informationen Семантическая информация	
Bedeutung Значение	[jüngere] weibliche Person, die in einer Familie zur Betreuung der Kinder angestellt ist
DUDEN	
Bedeutung Значение	(veraltend) in einer Familie fest angestelltes Mädchen, das die Kinder betreut das Kindermädchen spielte mit den Kindern
D W D S	
Paradigmatische Beziehungen Парадигматические связи	
Paradigmatische Beziehungen Парадигматические связи	Synonymgruppe: Kinderfrau, Kindermädchen, Nanny, Tagesmutter
D W D S	
Synonyme Синонимы	Amme, Kinderpflegerin, Nanny; (veraltend) Kinderfräulein, Nurse; (veraltet) Bonne; (früher) Gouvernante
DUDEN	
Kollokationen Коллокации	
Typische linke Nachbarn Типичное левое окружение	als (317,55), ein (162,61), das (145,36), einem (95,86), dem (94,22), ehemaliges (92,11), philippinische (84,63), philippinisches (70,88), Das (63,98), stachelige (59,4), ehemalige (58,19), philippinischen (45,89), früheres (45,47), neues (43,79)
Übersetzung ins Russische Перевод на русский язык	
Übersetzung ins Russische Перевод на русский язык	няня
CSDB	

Рис. 3. Пример результата поиска лингвистической информации к одной лемме (при запросе было активировано 8 опций в пяти базах данных)

Информационное обеспечение функционирования программы включает следующие немецкоязычные сетевые лингвистические базы данных:

- 1) онлайн-словарь *Duden* [4];
- 2) толковый интернет-словарь современного немецкого языка *DWDS* (*D*igitales *W*örterbuch der *d*eutschen *S*prache des 20. *J*h.) [2];
- 3) поисковая лексическая система *Wortschatz* (*D*eutscher *W*ortschatz *P*ortal) [5];
- 4) грамматическая сетевая база данных *CANOO* (*D*eutsche *W*örterbücher und *G*rammatik *C*ANOO) [7];
- 5) многоязычный онлайн-словарь *PONS* (немецко-русское направление перевода) [8].

Процесс генерации лексического минимума может потребовать разного количества времени в зависимости от количества отобранных лексем и активированных опций (т.е. требуемого количества лингвистической информации по каждой лексеме), а также от свойств сетевого подключения.

Проведённая оценка эффективности программы по скорости выполнения поискового запроса на базе 600 наблюдений по коэффициенту линейной корреляции Пирсона показала, что скорость выполнения поискового запроса (т.е. количество единиц, обработанных программой за единицу времени) тем выше, чем больше лексических единиц в поисковом запросе. При этом скорость в любом случае выше скорости сбора информации человеком.

SWOT-анализ¹ [11] прототипа программы и перспектив его внедрения в систему автоматизированного рабочего места преподавателя-лингвиста (см. таблицу) показал, что *Программа генерации тексториентированного лексического минимума Vokabelliste* может быть использована в нескольких направлениях:

- автоматизация сбора лингвистической информации лингвистами / преподавателями при проведении лингвистических исследований;
- автоматизация сбора лингвистической информации учителями немецкого языка в лингводидактических целях;
- автоматизация сбора лингвистической информации студентами в рамках самостоятельной работы над немецкоязычным текстом.

¹ SWOT = Strengths (сильные стороны), Weaknesses (слабые стороны), Opportunities (возможности) и Threats (угрозы).

SWOT-анализ программы *Vokabelliste*

Сильные стороны	Слабые стороны
<ul style="list-style-type: none"> • экономит время пользователя на поиск лингвистической информации; • позволяет вести целенаправленный отбор информации одного типа из разных баз данных; • позволяет наглядно и эффективно сравнивать полученную информацию одного типа из разных источников; • не требует инсталляции; • имеет двуязычный интерфейс; • позволяет импортировать разные текстовые типы файлов 	<ul style="list-style-type: none"> • имеет ограниченное количество подключенных баз данных; • не позволяет оценить объем информации до ее сбора и сохранения; • имеет традиционный интерфейс; • использует несовершенный лемматизатор (который, однако, позволяет себя корректировать); • существует только на носителе; • работает только на компьютерах с предустановленным MS Word; • не справляется с многозначностью и омонимией (выдает все существующие в базах данных значения лексем, включенных в поисковый запрос); • не позволяет вести поиск аналитических лексических форм (существительное с артиклем, глаголы и прилагательные в сложных формах, отделяемые приставки, сложные слова, устойчивые словосочетания); • может включать в результаты поиска нечёткие совпадения (онлайн-словарь PONS)
Возможности	Угрозы
<ul style="list-style-type: none"> • использование при подготовке к чтению / анализу текста; • использование для выполнения поискового запроса без текста; • применение для различных целевых групп (преподаватели, учителя, студенты); • применение в научных и / или лингводидактических целях 	<ul style="list-style-type: none"> • наличие в загрузочной папке большого количества файлов (что может привести в замешательство неопытного пользователя); • возможный отказ в выполнении поискового запроса при изменении параметров доступа владельцами используемых баз данных

Разработанная *Программа генерации текстоориентированного лексического минимума Vokabelliste* повышает эффективность труда преподавателя, поскольку позволяет интегрировать лингвистическую информацию различных сетевых баз данных. Таким

образом, программа может стать одним из элементов автоматизированного рабочего места лингвиста за счет автоматизации рутинных процессов, связанных с обучением.

Литература

1. *Программа* генерации текстоориентированного лексического минимума «Vokabelliste»: Свидетельство о государственной регистрации программы для ЭВМ № 2015619379 / О.Ю. Савина; заявитель и правообладатель Тюменский государственный университет. Заявл. 13.07.2015; гос. регистр. 01.09.2015. Бюл. № 2015619379, опублик. 20.09.2015.

2. *Digitales Wörterbuch der deutschen Sprache DWDS*. – URL: <http://www.dwds.de> (дата обращения: 16.07.2016).

3. *OWID*. Online-Wortschatz-Informationssystem Deutsch. – URL: <http://www.owid.de/> (дата обращения: 16.07.2016).

4. *Duden-Online*. Wörterbuch. Verlag Bibliographisches Institut GmbH. – URL: <http://www.duden.de> (дата обращения: 16.07.2016).

5. *Deutscher Wortschatz Portal*. Universität Leipzig. – URL: <http://www.wortschatz.uni-leipzig.de>. (дата обращения: 16.07.2016).

6. *Grammis 2.0*. Das grammatische informationssystem des Instituts für deutsche Sprache (IDS). – URL: <http://hypermedia.ids-mannheim.de/> (дата обращения: 16.07.2016).

7. *Deutsche Wörterbücher und Grammatik CANOO*. – URL: <http://www.canoo.net> (дата обращения: 16.07.2016).

8. *PONS Online-Wörterbuch*. – URL: <http://de.pons.com>. (дата обращения: 16.07.2016).

9. *Deutsches Referenzkorpus. DeReKo*. – URL: <http://www1.ids-mannheim.de/kl/projekte/korpora.html> (дата обращения: 16.07.2016).

10. *Archiv für Gesprochenes Deutsch*. – URL: <http://agd.ids-mannheim.de/index.shtml> (дата обращения: 16.07.2016).

11. *Майсак О.С.* SWOT-анализ: объект, факторы, стратегии. Проблема поиска связей между факторами // Прикаспийский журнал: управление и высокие технологии. – 2013. – № 1. – С. 151–157.

INTEGRATION OF LINGUISTIC SEARCH TOOLS FOR DIDACTIC PURPOSES

Voprosy leksikografii – Russian Journal of Lexicography, 2016, 2(10), pp. 55–66.

DOI: 10.17223/22274200/10/4

Olga Yu. Savina, Tyumen State University (Tyumen, Russian Federation). E-mail: osavina@gmail.com

Keywords: integration of linguistic search, text-based basic learner dictionary for the German language.

The paper presents a new search tool that integrates linguistic search in various network resources in German.

Conventional using of online resources for linguistic search (network dictionaries, linguistic corpora and other databases) has its advantages and disadvantages: relevance of the information, authentic examples of the use of lexical units, but at the same time the information collected this way is usually quite heterogeneous and requires time-consuming laborious manual processing: sorting by the type of information and text formatting. The desire to neutralize the disadvantages mentioned above has led to the development of *Vokabelliste*, a lexical analyzer program for a text-based generation of a basic learner dictionary (for the German language).

The purpose of the article is to present the main characteristics of the developed program. The material of the study is based on the following German language network databases: the Duden Online Dictionary, the Digital Dictionary of the German Language, Leipzig Corpora Collection, Canoonet online language services, the multilingual online-dictionary PONS.

The methodological basis of the research was modelling (theoretical and application level). As a result, the functional prototype of the *Vokabelliste* program was developed.

Target groups of the program users are professors and schoolteachers of German, as well as students of German. To compile the basic learner dictionary using this program, the user manually performs only the choice of tokens from the text to be read, to be included in the basic learner dictionary, and specifies the necessary types of linguistic information. Next, the program automatically makes queries into databases for each of the lexical units, combines the marked types of information in a single document and provides a standard display of search results.

In the final version of the basic learner dictionary the information is ordered by type: spelling and phonetic features; information on grammar; information on semantics; usage examples; paradigmatic relations; collocations; translation into Russian.

The *Vokabelliste* program developed for the generation of a text-based basic learner dictionary improves efficiency (findings confirmed statistically), because it allows to automatically integrate linguistic information of various databases. Thus, the program can be one of the elements of an automated workplace of a linguist through the automation of routine processes.

References

1. Savina, O.Yu. (2015) *The Vokabelliste program of text-oriented lexical minimum generation*: Certificate of state registration of the computer program 2015619379; applicant and rights holder Tyumen State University. Applied on 13 July 2015; State registration on 01 September 2015. Bulletin 2015619379, published 20 September 2015.
2. DWDS. (n.d.) *Digitales Wörterbuch der deutschen Sprache* [Digital dictionary of the German language]. [Online]. Available from: <http://www.dwds.de>. (Accessed: 16th July 2016).
3. OWID. (n.d.) *Online-Wortschatz-Informationen system Deutsch* [Online German vocabulary information system]. [Online]. Available from: <http://www.owid.de/>. (Accessed: 16th July 2016).
4. Duden-Online. (n.d.) *Wörterbuch* [Dictionary]. Verlag Bibliographisches Institut GmbH. [Online]. Available from: <http://www.duden.de>. (Accessed: 16th July 2016).

5. Leipzig University. (n.d.) *Deutscher Wortschatz Portal* [German vocabulary portal]. [Online]. Available from: <http://www.wortschatz.uni-leipzig.de>. (Accessed: 16th July 2016).
6. Grammis 2.0. (n.d.) *Das grammatische informationssystem des Instituts für deutsche Sprache (IDS)* [The grammatical information system of the Institute for German Language (IDS)]. [Online]. Available from: <http://hypermedia.ids-mannheim.de/>. (Accessed: 16th July 2016).
7. CANOO. (n.d.) *Deutsche Wörterbücher und Grammatik* [German dictionary and grammar]. [Online]. Available from: <http://www.canoo.net>. (Accessed: 16th July 2016).
8. PONS. (n.d.) *Online-Wörterbuch* [Online dictionary]. [Online]. Available from: <http://de.pons.com>. (Accessed: 16th July 2016).
9. IDS. (c. 2016) *Deutsches Referenzkorpus. DeReKo* [The German reference corpus. DeReKo]. [Online]. Available from: <http://www1.ids-mannheim.de/kl/projekte/korpora.html>. (Accessed: 16th July 2016).
10. *Archiv für Gesprochenes Deutsch* [Archive for Spoken German]. [Online]. Available from: <http://agd.ids-mannheim.de/index.shtml>. (Accessed: 16th July 2016).
11. Maysak, O.S. (2013) SWOT-analiz: ob'ekt, faktory, strategii. Problema poiska svyazey mezhdru faktorami [SWOT-analysis: object, factors, strategies. The problem of finding connections between factors]. *Prikaspiyskiy zhurnal: upravlenie i vysokie tekhnologii*. 1. pp. 151–157.